

Transactions on Graph Data and Knowledge

Volume 1 | Issue 1 | December, 2023

Special Issue: Trends in Graph Data and Knowledge

^{Edited by} Aidan Hogan Ian Horrocks Andreas Hotho Lalana Kagal

TGDK, Vol. 1, Issue 1

https://www.dagstuhl.de/tgdk

TGDK Special Issue Editors

Aidan Hogan 回

DCC, Universidad de Chile, IMFD, Chile ahogan@dcc.uchile.cl

Ian Horrocks

University of Oxford, U.K. ian.horrocks@cs.ox.ac.uk

Andreas Hotho 回

Department of Informatics, University of Würzburg, Germany hotho@informatik.uni-wuerzburg.de

Lalana Kagal 回

Massachusetts Institute of Technology, Cambridge, MA, USA Ikagal@csail.mit.edu

ACM Classification 2012

Computing methodologies \rightarrow Knowledge representation and reasoning; Information systems \rightarrow Semantic web description languages; Information systems \rightarrow Graph-based database models; Computing methodologies \rightarrow Machine learning; Theory of computation \rightarrow Graph algorithms analysis; Mathematics of computing \rightarrow Graph theory

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany.

Publication date December, 2023

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://dnb.d-nb.de.

License

This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0): https: //creativecommons.org/licenses/by/4.0.



In brief, this license authorizes each and everybody to share (to copy, dis-

tribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors. *Digital Object Identifier* 10.4230/TGDK.1.1.0

Aims and Scope

Transactions on Graph Data and Knowledge (TGDK) is an Open Access journal that publishes original research articles and survey articles on graph-based abstractions for data and knowledge, and the techniques that such abstractions enable with respect to integration, querying, reasoning and learning. The scope of the journal thus intersects with areas such as Graph Algorithms, Graph Databases, Graph Representation Learning, Knowledge Graphs, Knowledge Representation, Linked Data and the Semantic Web. Also inscope for the journal is research investigating graphbased abstractions of data and knowledge in the context of Data Integration, Data Science, Information Extraction, Information Retrieval, Machine Learning, Natural Language Processing, and the Web.

The journal is Open Access without fees for readers or for authors (also known as Diamond Open Access).

Editors in Chief

- Aidan Hogan
- Ian Horrocks
- Andreas Hotho
- 🔳 Lalana Kagal

Editorial Office Schloss Dagstuhl – Leibniz-Zentrum für Informatik TGDK, Editorial Office Oktavie-Allee, 66687 Wadern, Germany tgdk@dagstuhl.de

https://www.dagstuhl.de/tgdk

Contents

Preface

Transactions on Graph Data an	d Knowledge		
Aidan Hogan, Ian Horrocks,	Andreas Hotho,	and Lalana Kagal	 1:1-1:4

Trends in Graph Data and Knowledge

 Large Language Models and Knowledge Graphs: Opportunities and Challenges Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and 	
Damien Graux	2:1-2:38
Knowledge Engineering Using Large Language Models Bradley P. Allen, Lise Stork, and Paul Groth	3:1–3:19
 Knowledge Graph Embeddings: Open Challenges and Opportunities Russa Biswas, Lucie-Aimée Kaffee, Michael Cochez, Stefania Dumbrava, Theis E. Jendal, Matteo Lissandrini, Vanessa Lopez, Eneldo Loza Mencía, Heiko Paulheim, Harald Sack, Edlira Kalemi Vakaj, and Gerard de Melo 	4:1-4:32
Knowledge Graphs for the Life Sciences: Recent Developments, Challenges and	
Opportunities Jiaoyan Chen, Hang Dong, Janna Hastings, Ernesto Jiménez-Ruiz, Vanessa López, Pierre Monnin, Catia Pesquita, Petr Škoda, and Valentina Tamma	5:1-5:33
Towards Ordinal Data Science Gerd Stumme, Dominik Dürrschnabel, and Tom Hanika	6:1-6:39
Rule Learning over Knowledge Graphs: A Review Hong Wu, Zhe Wang, Kewen Wang, Pouya Ghiasnezhad Omran, and Jiangmeng Li	7:1-7:23
Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges Claudia d'Amato, Louis Mahon, Pierre Monnin, and Giorgos Stamou	8:1-8:35
Trust, Accountability, and Autonomy in Knowledge Graph-Based AI for Self-Determination Luis-Daniel Ibáñez, John Domingue, Sabrina Kirrane, Oshani Seneviratne, Aisling Third, and Maria-Esther Vidal	9:1-9:32
Multilingual Knowledge Graphs and Low-Resource Languages: A Review Lucie-Aimée Kaffee, Russa Biswas, C. Maria Keet, Edlira Kalemi Vakaj, and Gerard de Melo	10:1-10:19
How Does Knowledge Evolve in Open Knowledge Graphs? Axel Polleres, Romana Pernisch, Angela Bonifati, Daniele Dell'Aglio, Daniil Dobriy, Stefania Dumbrava, Lorena Etcheverry, Nicolas Ferranti, Katja Hose, Ernesto Jiménez-Ruiz, Matteo Lissandrini, Ansgar Scherp, Riccardo Tommasini, and Johannes Wachs	11:1-11:59
Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 0, pp. 0:v-0:x Special Issue: Trends in Graph Data and Knowledge. Editors: Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal	



Transactions on Graph Data and Knowledge TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

0:vi Contents

Structural Summarization of Semantic Graphs Using Quotients Ansgar Scherp, David Richerby, Till Blume, Michael Cochez, and Jannik Rau	12:1–12:25
Autonomy in the Age of Knowledge Graphs: Vision and Challenges	
Jean-Paul Calbimonte, Andrei Ciortea, Timotheus Kampik, Simon Mayer,	
Terry R. Payne, Valentina Tamma, and Antoine Zimmermann	13:1–13:22

List of Authors

Bradley P. Allen (3) University of Amsterdam, Amsterdam, The Netherlands

Russa Biswas (0 (2, 4, 10) Hasso-Plattner Institute, Potsdam, Germany

Till Blume (12) Ernst and Young Research, Berlin, Germany

Angela Bonifati (D) (2, 11) Lyon 1 University, CNRS, IUF, France

Jean-Paul Calbimonte (D) (13) University of Applied Sciences and Arts Western Switzerland HES-SO, Sierre, Switzerland

Jiaoyan Chen (2, 5) The University of Manchester, United Kingdom; University of Oxford, United Kingdom

Andrei Ciortea (D) (13) Institute of Computer Science, University of St.Gallen, Switzerland

Michael Cochez (4, 12) Vrije Universiteit Amsterdam, The Netherlands; Elsevier Discovery Lab, Netherlands

Claudia d'Amato (0) University of Bari, Italy

Gerard de Melo 💿 (2, 4, 10) Hasso-Plattner Institute, Potsdam, Germany

Daniele Dell'Aglio (11) Aalborg University, Denmark

Stefan Dietze (2) GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany; Heinrich-Heine-Universität Düsseldorf, Germany

Daniil Dobriy (11) Vienna University of Economics and Business, Austria

John Domingue (9) Knowledge Media Institute, The Open University, Milton Keynes, UK

Hang Dong (5) Department of Computer Science, University of Oxford, UK

Mauro Dragoni (D) (2) Fondazione Bruno Kessler, Trento, Italy Stefania Dumbrava ^(D) (4, 11) ENSIIE, France

Dominik Dürrschnabel (6) Knowledge & Data Engineering Group, Research Center for Information System Design &, Department of Electrical Engineering and Computer Science, University of Kassel, Germany

Lorena Etcheverry (D) (11) Universidad de la República, Uruguay

Nicolas Ferranti (D) (11) Vienna University of Economics and Business, Austria

Damien Graux (2) Edinburgh Research Centre, CSI, Huawei Technologies UK, United Kingdom

Paul Groth (D) (3) University of Amsterdam, Amsterdam, The Netherlands

Tom Hanika (D) (6) Institute of Computer Science, University of Hildesheim, Germany; Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Germany

Janna Hastings (5) Institute for Implementation Science in Health Care, University of Zurich, Switzerland; School of Medicine, University of St. Gallen, Switzerland

Aidan Hogan 💿 (1) DCC, Universidad de Chile, IMFD, Chile

Ian Horrocks (D) (1) University of Oxford, U.K

Katja Hose (11) TU Wien, Austria

Andreas Hotho (D) (1) Department of Informatics, University of Würzburg, Germany

Luis-Daniel Ibáñez (9) Department of Electronics and Computer Science, University of Southampton, UK

Hajira Jabeen (D) (2) GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

Theis E. Jendal (4) Aalborg University, Denmark

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 0, pp. 0:vii-0:x Special Issue: Trends in Graph Data and Knowledge. Editors: Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal Transactions on Graph Data and Knowledge

 TGDK
 Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Ernesto Jiménez-Ruiz (5, 11) City, University of London, UK; SIRIUS, University of Oslo, Norway

Lucie-Aimée Kaffee <a>[b] (4, 10) Hasso-Plattner-Institut, Potsdam, Germany

Lalana Kagal ^(D) (1) Massachusetts Institute of Technology, Cambridge, MA, USA

Jan-Christoph Kalo (2) University of Amsterdam, The Netherlands

Timotheus Kampik (D) (13) Department of Computing Science, Umeå University, Sweden; SAP Signavio, Berlin, Germany

C. Maria Keet (10) University of Cape Town, South Africa

Sabrina Kirrane (9) Institute for Information Systems & New Media, Vienna University of Economics and Business, Austria

Jiangmeng Li (7) Institute of Software, Chinese Academy of Sciences, China

Matteo Lissandrini (D) (2, 4, 11) Aalborg University, Denmark

Vanessa Lopez (1) IBM Research Dublin, Ireland

Vanessa López (5) IBM Research Europe, Dublin, Ireland

Louis Mahon (b) (8) Oxford University, United Kingdom

Simon Mayer (13) Institute of Computer Science, University of St.Gallen, Switzerland

Eneldo Loza Mencía 💿 (4) research.eneldo.net, Frankfurt, Germany

Pierre Monnin (5, 8) Université Côte d'Azur, Inria, CNRS, I3S, France

Janna Omeliyanenko (D) (2) University of Würzburg, Germany

Pouya Ghiasnezhad Omran ^(D) (7) School of Computing, The Australian National University, Australia

Jeff Z. Pan (¹⁰) (2) The University of Edinburgh, United Kingdom Heiko Paulheim (4) Universität Mannheim, Germany

Terry R. Payne (D) (13) Department of Computer Science, University of Liverpool, United Kingdom

Romana Pernisch (D) (11) Vrije Universiteit Amsterdam, The Netherlands; Discovery Lab, Elsevier, Amsterdam, The Netherlands

Catia Pesquita (5) LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Axel Polleres (11) Vienna University of Economics and Business, Austria; Complexity Science Hub Vienna, Austria

Jannik Rau (12) Ulm University, Germany

Simon Razniewski (¹⁰) (2) Bosch Center for AI, Renningen, Germany

David Richerby (12) University of Essex, UK

Harald Sack (D) (4) FIZ Karlsruhe, Germany; Karlsruhe Institute of Technology, AIFB, Germany

Ansgar Scherp (D) (11, 12) Ulm University, Germany

Oshani Seneviratne (0) Department of Computer Science, Rensselaer Polytechnic Institute, USA

Sneha Singhania 🔎 (2) Max Planck Institute for Informatics, Saarbrücken, Germany

Giorgos Stamou (0) (8) National Technical University of Athens, Greece

Lise Stork (0) (3) Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Gerd Stumme () (6) Knowledge & Data Engineering Group, Research Center for Information System Design &, Department of Electrical Engineering and Computer Science, University of Kassel, Germany

Valentina Tamma (0, 13) Department of Computer Science, University of Liverpool, UK

Authors

Aisling Third ^(D) (9) Knowledge Media Institute, The Open University, Milton Keynes, UK

Riccardo Tommasini 💿 (11) INSA Lyon, CNRS LIRIS, France

Edlira Vakaj 💿 (2) Birmingham City University, United Kingdom

Edlira Kalemi Vakaj ^(D) (4, 10) Birmingham City University, UK

Maria-Esther Vidal (D) (9) Leibniz University of Hannover, Germany; TIB-Leibniz Information Centre of Science and Technology, Hannover, Germany; L3S Research Centre, Hannover, Germany

Johannes Wachs (D) (11) Corvinus University of Budapest, Hungary; Centre for Economic and Regional Studies, Hungary; Complexity Science Hub Vienna, Austria

Kewen Wang (7) School of Information and Communication Technology, Griffith University, Australia

Zhe Wang (0) (7) School of Information and Communication Technology, Griffith University, Australia

Hong Wu (D) (7) School of Information and Communication Technology, Griffith University, Australia

Wen Zhang (2) Zhejiang University, China

Antoine Zimmermann (D) (13) Mines Saint-Etienne, Université Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, France

Petr Škoda (5) Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague, Czechia

Transactions on Graph Data and Knowledge

Aidan Hogan 🖂 🏠 🝺

DCC, Universidad de Chile, IMFD, Chile

Ian Horrocks 🖂 🏠 💿 University of Oxford, U.K.

AndreasHotho 🖂 🏠 💿 Department of Informatics, University of Würzburg, Germany

Lalana Kagal 🖂 🏠 💿

Massachusetts Institute of Technology, Cambridge, MA, USA

— Abstract -

Transactions on Graph Data and Knowledge (TGDK) is a new journal publishing peer-reviewed research on graph-based abstractions for data and knowledge, as well as the techniques, theories, applications and results that arise in this setting. TGDK is a community-run, Diamond Open Access journal, meaning that papers are published openly on the Web without fees for authors or read-

ers. In this preface, we provide some brief remarks about the rationale and goals of the new journal, followed by an introduction to its inaugural issue, entitled "Trends in Graph Data and Knowledge", which collects together 12 diverse vision, position and survey papers on the types of research topics that exemplify the scope of this new journal.

2012 ACM Subject Classification Computing methodologies \rightarrow Knowledge representation and reasoning; Information systems \rightarrow Semantic web description languages; Information systems \rightarrow Graph-based database models; Computing methodologies \rightarrow Machine learning; Theory of computation \rightarrow Graph algorithms analysis; Mathematics of computing \rightarrow Graph theory

Keywords and phrases Graphs, Data, Knowledge

Digital Object Identifier 10.4230/TGDK.1.1.1

Category Preface

Acknowledgements We warmly thank Dagstuhl Publishing for their continued collaboration, the Semantic Web Science Association (SWSA) for their support, our colleagues on the SWSA Task Force who helped to plan this new journal, as well as our Advisory and Editorial Boards for their contributions towards getting the journal up and running and ensuring its continued operation and development.

Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1

Transactions on Graph Data and Knowledge

Transactions on Graph Data and Knowledge (TGDK) is a new Open Access journal publishing research contributions on the use of graph-based abstractions for representing data and knowledge. Such abstractions enable a wide variety of techniques to be applied for integrating, querying, reasoning over and learning from diverse data and knowledge at large scale. As such, the journal draws together expertise from a variety of research communities of Computer Science – including Graph Databases, Graph Representation Learning, Graph Theory, Knowledge Graphs, Knowledge Representation and the Semantic Web – on the topic of graph-based data and knowledge.

Herein we provide an overview of the scope, goals and history of TGDK, and then introduce the First Issue of the journal, entitled "Trends in Graph Data and Knowledge".



© Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal;

licensed under Creative Commons License CC-BY 4.0 Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 1, pp. 1:1-1:4

Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1:2 Transactions on Graph Data and Knowledge

1.1 Scope of the journal

Recent years have seen a resurgence in interest in the use of graphs for representing data and knowledge, with, for example, the topic of Knowledge Graphs gaining increasing traction in industry and academia alike. Advances in other areas of Computer Science are further reflected in this field; for example, key advances in the area of Deep Learning have led to the emergence of powerful new techniques in the area of Graph Representation Learning.

Such advances increasingly involve a confluence of diverse techniques from seemingly disparate paradigms. Within the field of Artificial Intelligence, new research results and directions continue to emerge on how best to synthesise the symbolic, deductive and logical foundations of Knowledge Representation with the neural, inductive and connectionist underpinnings of Machine Learning for better harnessing graph-based data and knowledge at large scale; advances are being achieved, for example, on combining knowledge graph embeddings and graph neural networks with formal semantics, ontologies and logic. We can then ask how such techniques might interplay with Graph Data Management to enable novel forms of queries over not only explicit data at large scale, but also implicit knowledge: what query languages, optimisations, notions of schema, etc., will form the basis of the graph databases of the future. We can wonder what insights are to be gained from Graph Algorithms & Theory in this setting, where, for example, graph centrality measures, graph isomorphism, spectral graph theory, tree decompositions, etc., have already led to key practical and theoretical advances in various related subareas of Graph Data and Knowledge, and will continue to bolster new insights. In the Semantic Web area, we see such techniques being increasingly drawn together in order to better structure and leverage the content of the Web itself, giving rise to a massive and unprecedented decentralised repository of graph-based knowledge, composed of comprehensive knowledge graphs such as Wikidata, complex ontologies such as those capturing shared knowledge in the Biomedical domain, and a great many Linked Datasets such as those published by governments for the public good.

Graph Data and Knowledge can also play a role in other areas of computer science, leading to advances in Data Science, Information Retrieval, Natural Language Processing, and more besides. More and more use-cases for Graph Data and Knowledge continue to emerge in our society, facilitating new discoveries in scientific disciplines, more transparency in governance, smarter applications for users, better data management in industries, along with many other advances.

All of these techniques, theories, systems, applications and areas interact in complex ways, but have at their foundation a common starting point: the use of graphs to represent data and knowledge. How they will continue to interact lies at the core of TGDK's scope, and our aim is to publish some of the key research contributions that will shape the area's future in the journal. These contributions may range from novel theoretical results – stemming from areas such as Database Theory, Graph Theory, Knowledge Representation, Logic and Ordinal Data Science – to results from systems-oriented, empirical and/or applied research – including Graph Database Systems, Graph Representation Learning, Knowledge Graphs, Linked Data, and more besides. Works that combine both theoretical and empirical contributions will be particularly welcome.

TGDK currently solicits and publishes two types of articles as part of its regular call: research articles and survey articles. Research articles present novel research contributions that advance the state-of-the-art in their respective area, and are judged based on novelty, relevance, potential impact, technical soundness, reproducibility and clarity. Survey articles present a novel, systematic and comprehensive synthesis of published research works within a chosen scope, and are judged based on novelty, relevance, scope, completeness and clarity. Aside from research and survey articles, other types of submissions are occasionally welcomed as part of a Special Issue. This Special Issue, the first of TGDK, presents a selection of vision, position and survey papers. We are also planning a Special Issue for 2024 that will solicit resources papers.

1.2 Goals of the journal

TGDK has the following stated goals and principles:

- **To publish high-quality articles:** TGDK publishes articles relating to research on graph data and knowledge. All articles undergo a rigorous peer-review process run by the Editors-in-Chief and the Editorial Board consisting of experienced researchers of international standing, ensuring a high standard regarding novelty, relevance, impact, technical soundness, reproducibility and clarity. We aim to have TGDK indexed as widely as possible as soon as possible, with an Impact Factor competing with the best journals in the area.
- **To publish under Diamond OA:** TGDK provides Open Access under the "Diamond" (sometimes known as "Platinum") model, meaning that no fees are charged to authors or to readers. This is enabled by the financial support of scientific organisations, such as the Semantic Web Science Association (SWSA), that cover the low fees required by Dagstuhl Publishing.
- To publish online in a dependable, reputable manner: Dagstuhl Publishing provides to TGDK a formal publication mechanism that includes archiving by the *Deutsche Nationalbibliothek* (German National Library). The journal will have an ISSN, and each article will be assigned a DOI and a URN. All articles are indexed on DBLP and we will push for them to be indexed as widely and as soon as possible, e.g., in Scopus, Web of Science collections, etc.
- **To publish rather than control content:** All articles at TGDK will be published under a nonexclusive publishing agreement under the CC-BY 4.0 licence. Authors will retain the copyright of their own works. Meta-data relating to articles will be published under CC0, facilitating their inclusion and use as part of open collections of scholarly knowledge.
- **To build a community:** TGDK adopts a distinctive scope that combines graphs with data and knowledge. Though existing journals have overlapping scope with TGDK, we believe this particular focus on the combination of graphs with data and/or knowledge to be unique amongst journals. Our goal then is that TGDK crystallise a community specifically around this topic, one composed of a cross-section of established communities such as Artificial Intelligence, Databases, Graph Algorithms & Theory, Knowledge Representation, the Semantic Web, etc.
- To be community-run: TGDK is run by the community in a non-profit manner. It puts the needs of its research community first. Editors-in-Chief (EiCs) will serve terms of a maximum of four years, and each year a new EiC will be elected from the Editorial Board. The Editorial Board will represent the community in all of its diversity, and participate in strategic decisions. New community members will be regularly invited to form part of the Editorial Board.
- **To operate transparently:** Though some aspects of the academic process such as reviewer identities during single-blind review are not subject to transparency, our goal is to be as transparent as possible regarding the operation of TGDK, in terms of scope, criteria, review process, budget, procedures for selecting new EiCs, etc. We will also make meta-data about the journal its articles, authors, etc. publicly available in structured formats.
- To innovate in how best to serve the research community: At TGDK, we are open to discussing and exploring novel opportunities to improve the publication process for the community, which may involve, for example, initiatives to improve the review process, reduce review times, incentivise reviewers, improve reproducibility, publish supplementary material, foster good practices in terms of meta-data, enable post-publication discussion, etc.
- **To become an exemplary non-profit journal:** Together with Dagstuhl Publishing, we aim for TGDK to show that journals run by the research community, for the research community, offer a promising and we argue *better* alternative to journals run by, for example, commercial academic publishers. In particular, TGDK will endeavour to inspire and support other journals to follow a similar path to that exemplified by TGDK and other Diamond OA journals.

1:4 Transactions on Graph Data and Knowledge

1.3 History of the journal

The inception of the journal begins with Ian Horrocks, Andreas Hotho and Lalana Kagal, who were Editors-in-Chief of the Journal of Web Semantics (JWS) published by Elsevier. They opted to resign together from JWS at the end of 2022 in order to initiate a new journal that they hoped might better serve the community. They approached the Semantic Web Science Association (SWSA) – a non-profit organisation that manages scientific events in the Semantic Web area, and that is also affiliated with JWS – to discuss the possibility of SWSA supporting a new journal.

A SWSA Task Force comprising senior members of the Semantic Web community, including the three ex-EiCs of JWS, as well as Abraham Bernstein, Tim Finin, Chiara Ghidini, Markus Krötzsch and Axel Polleres, was set up to explore potential publishing models, scopes, titles, etc., for the new journal. A fourth Editor-in-Chief, Aidan Hogan, was invited to join the new initiative.

The Task Force decided that the new journal should have a broad scope so as to capture the growing interest on Knowledge Graphs as well as more established areas of Semantic Web research. Furthermore, the consensus was that the journal should be run in a non-profit manner by the community, should be Open Access, and should strive to be free of fees for authors and readers.

The new Editors-in-Chief began the process of establishing this new journal – creating a webpage with the title and scope, inviting the Editorial Board (with particular emphasis on inviting members from areas other than the Semantic Web) to join, soliciting seed funding from SWSA, contacting potential publishers, etc. – in early 2023. Planning for this First Issue began in April of 2023, with the Editors-in-Chief soliciting topics for articles, culminating in a first batch of papers being published in December 2023. More papers in this particular collection, entitled "Trends in Graph Data and Knowledge" will be published early in 2024.

2 Trends in Graph Data and Knowledge

As the Inaugural Editors-in-Chief, we are pleased to herein present the First Issue of this journal: a Special Issue entitled "Trends in Graph Data and Knowledge".

This issue includes a collection of 12 vision, position and survey papers relating to the past, present and future of research in the area of Graph Data and Knowledge. It draws together diverse perspectives from internationally-renowned experts – many of whom are members of the Editorial Board of TGDK – on what the future holds for this exciting area. All papers received at least three reviews and one meta-review from members of our Editorial Board. This collection of papers covers a diverse range of topics that discuss:

- the interplay between the area of Graph Data & Knowledge and the areas of Machine Learning, Natural Language Processing (particularly large language models) and Ordinal Data Science;
- how Graph Data & Knowledge can help to address challenges relating to autonomous agents, data self-determination, knowledge engineering, the life sciences, and low-resource languages;
- algorithms, analyses, concepts and techniques involving embeddings, evolution, rule learning, and summarisation in the context of Graph Data & Knowledge.

These articles exemplify the types of research topics that we hope to see TGDK publications address in the future. And in case the reader is inspired by one of these articles to follow up on a research direction it mentions, we would be pleased to receive submissions of their contributions.

Large Language Models and Knowledge Graphs: **Opportunities and Challenges**

Jeff Z. Pan 🖂 🏠 💿

The University of Edinburgh, United Kingdom

Jan-Christoph Kalo 🖂 🗅 University of Amsterdam, The Netherlands

Jiaoyan Chen 🖂 🖻 The University of Manchester, United Kingdom University of Oxford, United Kingdom

HajiraJabeen 🖂 回

GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

Wen Zhang 🖂 🕩 Zhejiang University, China

Russa Biswas 🖂 回 Hasso-Plattner Institute, Potsdam, Germany

Angela Bonifati 🖂 🗈 Lyon 1 University, CNRS, IUF, France

Mauro Dragoni 🖂 回 Fondazione Bruno Kessler, Trento, Italy

Simon Razniewski 🖂 🗈 Bosch Center for AI, Renningen, Germany

Sneha Singhania 🖂 回 Max Planck Institute for Informatics, Saarbrücken, Germany

Stefan Dietze 🖂 🖻 GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany Heinrich-Heine-Universität Düsseldorf, Germany

Janna Omeliyanenko 🖂 🗈 University of Würzburg, Germany

Matteo Lissandrini 🖂 🕩 Aalborg University, Denmark

Gerard de Melo ⊠© Hasso-Plattner Institute, Potsdam, Germany

Edlira Vakaj ⊠© Birmingham City University, United Kingdom

Damien Graux 🖂 🗅 Edinburgh Research Centre, CSI, Huawei Technologies UK, United Kingdom

— Abstract –

Large Language Models (LLMs) have taken Knowledge Representation – and the world – by storm. This inflection point marks a shift from explicit knowledge representation to a renewed focus on the hybrid representation of both explicit knowledge and parametric knowledge. In this position paper,

we will discuss some of the common debate points within the community on LLMs (parametric knowledge) and Knowledge Graphs (explicit knowledge) and speculate on opportunities and visions that the renewed focus brings, as well as related research topics and challenges.

2012 ACM Subject Classification General and reference \rightarrow Surveys and overviews; Computing methodologies \rightarrow Knowledge representation and reasoning; Computing methodologies \rightarrow Natural language processing

Keywords and phrases Large Language Models, Pre-trained Language Models, Knowledge Graphs, Ontology, Retrieval Augmented Language Models

Digital Object Identifier 10.4230/TGDK.1.1.2

Category Position

Related Version Full Version: https://arxiv.org/abs/2308.06374

Acknowledgements We would like thank Xiaoqi Han's helpful discussion and support when finalising the camera ready version of the paper.

Received 2023-06-26 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge



© Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux; licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 2, pp. 2:1-2:38 Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



2:2 LLMs and KGs: Opportunities and Challenges

1 Introduction

Large Language Models (LLMs) have taken Knowledge Representation (KR) – and the world – by storm, as they have demonstrated human-level performance on a vast spectrum of natural language tasks [25]. Following this, people are gradually starting to accept the possibility of having knowledge represented in the parameters by some language models, also known as *parametric knowledge*, which is the information encoded from training datasets into language models' internal model parameters. The arrival of LLMs announces the era of Knowledge Computing, in which the notion of reasoning within KR is broadened to many computation tasks based on various knowledge representations.

This is a big step for the field of Knowledge Representation. For a long time, people focused on explicit knowledge, such as those embedded in texts, sometimes also known as unstructured data, and those in a structured form, such as in databases and knowledge graphs (KGs) [141]. Historically, for a long time, humans used texts to pass down their knowledge from one generation to another, until around the 1960s, when researchers started to study knowledge representation for better natural language understanding and developed early systems, such as ELIZA [207] at the MIT. In the early 2000s, the Knowledge Representation and the Semantic Web communities worked together to standardize the widely used knowledge representation languages, such as RDF [139] and OWL [67], at web scale, using which the large-scale knowledge bases are then more widely known as KGs [141], due to their helpful graph structures, enabling both logical reasoning and graph-based learning.

This inflection point, with the arrival of LLMs, marks a paradigm shift from explicit knowledge representation to a renewed focus on the hybrid representation, sometimes also referred to as *semi-parametric knowledge*, of both explicit knowledge and parametric knowledge in neural networks (NNs). As a popular approach for explicit knowledge representation, KGs are now widely investigated for the combination with Transformer-based LLMs, including pre-trained masked language models (PLMs) like BERT [47] and RoBERTa [122], and more recent generative LLMs like the GPT series [24] and LLaMA [191]. Some works use LLMs to augment KGs for, e.g., knowledge extraction, KG construction, and refinement, while others use KGs to augment LLMs for, e.g., training and prompt learning, or knowledge augmentation. In this paper, considering both directions, LLMs for KGs and KGs for LLMs, we present a better understanding of the shift from explicit knowledge representation to a renewed focus on the hybrid representation of both explicit knowledge and parametric knowledge.

There has been some related work on LLMs and KGs. A related survey paper [235] presents a comprehensive review of using LLMs for KG construction and reasoning, while our work provides a more in-depth view of the inflection point, considering not only relational KGs but also KGs with ontologies as schemas, as well as other dimensions of structured knowledge, including tabular data [210] and numerical values [140]. Other works on the intersection of LLMs and KGs have a minor overlap with the topics covered in our paper; e.g., on studies using LLMs as KGs [5], on using KGs to augment LLMs [213], or on comparing GPT-4 with ChatGPT and SOTA fine-tuning methods on three knowledge-related tasks – entity, relation and event extraction, link prediction, and KG question answering [235]. Overall, none of these papers look into the implications of the inflection point with concrete applications. To this end, this paper summarizes the common debate points within the community, introduces the state-of-the-art for a comprehensive set of topics where KGs and LLMs are integrated, and further presents opportunities and challenges.

2 Terminology

Language models. The term "(large) language model (LLM)" covers quite a set of model families. Although the term already existed before the rise of neural models, it arguably rose to prominence with the publication of the BERT architecture [47], a Transformer-based neural architecture. The BERT model was typically used with parameters pre-trained on a large text corpus and fine-tuned on application specific data. BERT-style (masked) architectures have become a building block of many language-related tasks, where they can be used, for instance, for obtaining latent semantic representations of words or phrases, and are frequently employed in classification, similarity estimation, or retrieval tasks.

A second, slightly more recent paradigm of language models are autoregressive (generative) architectures like the GPT family. These models are suited for the generation of text from given prefixes, and, in contrast to BERT models, can (in principle) directly be used in downstream use cases. At the same time, they represent less mature architecture, with media headlines taken by closed commercial models like ChatGPT and GPT-4 [137], and open models like LLaMA [191], Vicuna [37], or Falcon [149] only recently entering the stage. Although fine-tuning is possible for some of them, autoregressive models are often used without adapting their parameters, e.g., via zero-shot or few-shot text-based prompts.

Where the distinction is important, we will explicitly state to which model family and prompt style we refer.

Explicit knowledge. Explicit knowledge includes both unstructured knowledge, such as text, images and videos, and structured knowledge, also referred to as symbolic knowledge, which is knowledge codified in a logical representation, such as knowledge graphs and databases. This has been the standard in the AI and KR communities for long and enables an intuitive understanding of assertions. Nevertheless, it has been known for long that not all knowledge is suited for explicit representations, and certain types of knowledge remain typically implicit.

Parametric knowledge. This is a kind of implicit knowledge. We use the term "parametric knowledge" to refer to the information encoded from training datasets into language models' internal model parameters. Most importantly, these parameters describe the weights of the neural network and thus consist of a set of numeric values.

Some parametric knowledge cannot be easily translated into symbolic knowledge, for instance, LLMs typically have solid "knowledge" on whether a textual phrase has positive sentiment or not, yet there is typically no lossless conversion into a symbolic formula. At the other end of the spectrum is the parametric knowledge that has been shown possible to be extracted into a crisp symbolic form, for example, "Who was the first person on the moon?".

As we will show later, parametric knowledge of the latter kind is important for many steps in KG construction and consolidation pipelines, while parametric knowledge of the former kind offers a direct supplement or alternative to existing KGs.

3 Common Debate Points within the Community

The usage of parametric and explicit knowledge together is a topic of debate in the Knowledge Computing community, with proponents and skeptics offering different perspectives. Below are some summaries of common points of contention.

2:4 LLMs and KGs: Opportunities and Challenges

Knowledge Representation and Reasoning. KGs offer a structured representation of knowledge with explicit relationships, enabling reasoning and inference [128]. Critics argue that parametric knowledge in NNs and LLMs relies on statistical patterns rather than true understanding and reasoning [14]. Proponents of LLMs like ChatGPT highlight their ability to generalize from large-scale text corpora, capturing a range of information, and excellent language understanding capabilities. On the one hand, LLMs could generate plausible but incorrect or nonsensical responses, such as hallucinations, due to a lack of explicit knowledge representation [163, 223]. There are also doubts on whether LLMs have the ability to learn directional entailments [114] or infer subsumption between concepts [75]. On the other hand, KGs can be costly to build. While LLMs can be expensive to train too, they can be readily usable to support many downstream applications, bringing AI from the backstage to center stage. To sum up, in comparison to the classic trade-off between expressiveness and decidability in knowledge representation, here KGs and LLMs have the trade-off between precision and recall considering using explicit and parametric knowledge in knowledge computing tasks.

High Precision Methods. The success of KGs can largely be attributed to their ability to provide factual information about entities with high accuracy. For instance, YAGO [180] asserts an accuracy rate exceeding 95%. Similarly, Google necessitates high accuracy in its KG for operational use, e.g., the semi-automatic construction method of Knowledge Vault was not utilized in production, partly due to its inability to reach the desired 99% accuracy in their benchmark [206]. Along this line of thought, many LLM-based methodologies for KG completion fail to reach these high-performance levels, as exemplified by the performance of BERT in [115, 193], and GPT-3, equipped with hundreds of billions of parameters [4]. This calls for novel high precision methods for KG construction based on LLMs.

Numerical Values. It is widely recognized that LLMs grapple with handling numerical values. Even straightforward arithmetic tasks can be a struggle for LMs, as highlighted in a study by Big-bench [179]. This shortcoming also extends to KG completion tasks [95]. Multiple LLMs have been evaluated on their ability to complete KGs using numerical facts from Wikidata [196], such as individuals' birth and death years. However, none of the tested models accurately predicted even a single year. This raises questions about the capability of current LLMs to correctly memorize numbers during pre-training in a way that enables them for subsequent use in KG completion. While LLMs like PaLM [41] demonstrate some proficiency in dealing with numbers, more commonly used smaller models seem ill-equipped for this task. The complexity escalates when considering the intricacies of metrics and diverse numbering formats and types. Currently, modifying LLMs to handle numerical values remains unresolved, making their utilization for numerical KG completion seem far from practical.

Long-tail Knowledge. One of the key research questions on LLMs for the Knowledge Computing community (and beyond) is how much knowledge LLMs remember [125]. Investigations indicate that LLMs' performance significantly deteriorates when dealing with random Wikidata facts, specifically those associated with long-tail entities, in comparison to popular entities, as evidenced in the PopQA dataset [125] and other datasets [152, 193]. This effect can be traced back to a causal relationship between the frequency of an entity's appearance in the pre-training corpus and the LLMs' capacity for memorization [53]. Even sizable LLMs face difficulties when trying to retain information about long-tail entities [97]. KGs inherently present an advantage over LLMs through their provision of knowledge about long-tail entities [95, 193] and thus can further help improve the recall for Knowledge Computing tasks.

Bias, Fairness and Beyond. Critics argue that LLMs can perpetuate and amplify biases present in the training data, leading to biased outputs. LLMs may generate biased or prejudiced responses if the training data contains stereotypes or discriminatory information [131, 108]. On the other hand, proponents argue that bias is not inherent to LLMs but reflects societal biases embedded in the data. They emphasize the importance of addressing bias in the training data and developing mitigation techniques [167, 154, 161]. A survey [17] argued that "bias" in Natural Language Processing (NLP) systems took various conceptualizations without being critically engaged by practitioners. KGs are also used in a plethora of downstream tasks, and social biases engraved in KG embeddings get propagated [68, 104]. Specifically, ontology creation, which generally comprises manual rules factored by opinions, motivations, and personal choices, is a source of bias [89, 51]. Also, automated pipelines for KG construction exhibit gender bias [127]. There are other similar concerns of LLMs beyond bias and fairness, including (but not limited to) copyright violation and misinformation. In general, due to the implicit nature of parametric knowledge, it is less straightforward to forget such toxic information from LLMs, compared to explicit knowledge.

Explainability and Interpretability. KGs are often preferred in scenarios where explainability and interpretability are crucial [31], as they explicitly represent relationships between entities and provide a structured knowledge representation. Skeptics of LLMs argue that these models lack transparency and interpretability, making it difficult to understand how they arrive at their answers or recommendations. Proponents of LLMs acknowledge the challenge of explainability but argue that recent research efforts [8, 88] are improving LLM's interpretability through techniques like attention mechanisms, model introspection. Some also argue that Chain-of-Thoughts (CoT) [204] can also improve the explainability of LLMs, although question decomposition and precisely answering sub-questions with LLMs are still far from being solved. Attribution evaluation and augmentation of LLMs with e.g., source paragraphs and sentences is another recent research topic for improving their explainability in question answering [18].

4 Opportunities and Visions

One of the key questions this paper needs to answer is, now with the emergence of parametric knowledge, what new opportunities do we have? Here are some of our thoughts on such new opportunities with the arrival of parametric knowledge and its potential integration with explicit knowledge.

- 1. Instant access to huge text corpora: As mentioned in the Introduction, for a long time, human beings passed down their knowledge in texts. Thus, a lot of knowledge these days are in textual form. Using LLMs gives access to extremely large text corpora at high speed, and recently even on consumer hardware [79]. This allows AI developers to avoid getting bogged down in previously critical challenges around data gathering, preparation, storage, and querying at scale. It also helps to reduce previously critical dependencies on the field of information retrieval.
- 2. Richer knowledge for many subtasks: Although the most prominent capabilities of LLMs, question answering and dialogue, are still under critical scrutiny, it should not be overlooked that LLMs have significantly advanced and simplified many traditional tasks of the knowledge engineering pipeline. Out-of-the-box, with fine-tuning on a few examples, or via few-shot prompting, LLMs have advanced many tasks such as dependency and structured parsing, entity recognition, and relation extraction. And just as errors propagate along a pipeline, so do improvements, thus enabling KG construction at unprecedented scale and quality. Furthermore, LLMs are readily usable for many downstream tasks beyond knowledge

2:6 LLMs and KGs: Opportunities and Challenges

engineering. By injecting explicit, and in particular structured, knowledge into LLMs, such as through retrieval augmented methods, one can make explicit knowledge more readily usable for such a wide range of downstream tasks, further realizing the vision of 'Knowledge is power'.

- **3. Even more advanced language understanding:** LLMs alone already significantly advanced the "understanding" of natural language, as evidenced by tasks like textual entailment, summarization, paraphrase detection and generation, etc. These capabilities are critical to making knowledge engineering robust to linguistic variance, typographic errors, redundancy, and other features of human-written, web-scraped, and other noisy forms of text. Now with potential novel approaches to combining parametric knowledge with explicit knowledge, it is possible to have even more advanced language understanding, not only for textual entailments, but also for other NLP tasks, such as summarization and consistent generation.
- 4. Compression entails consolidation: An important step in traditional knowledge engineering is the consolidation and aggregation of conflicting and concurring pieces of information, requiring often elaborate methods for consolidating observations from sentences, patterns, and constraints [173]. In LLM training, an aggregation occurs automatically. Although this step is not entirely understood, it brings the potential for outsourcing a major challenge in knowledge engineering.

With the above new opportunities brought by the combination of parametric and explicit knowledge, our vision is two-folded:

- In Explicit-Knowledge-First use cases, our vision is that LLMs will enable, advance, and simplify crucial steps in the knowledge engineering pipeline so much as to enable KGs at unprecedented scale, quality, and utility.
- In Parametric-Knowledge-First use cases, our vision is that KGs will improve, ground, and verify LLM generations so as to significantly increase reliability and trust in LLM usage.

Both visions are neither alternatives, nor does one build upon the other. Instead, we believe that classes of use cases will continue to exist side-by-side, some of which favor Explicit-Knowledge-First (scrutable) approaches, some of which favor Parametric-Knowledge-First (blackbox) approaches, with either of these having significant potential for benefiting from synergies of the two directions.

5 Key Research Topics and Related Challenges

With the opportunities and visions related to the availability of both parametric and explicit knowledge in place, in this section, we categorize, summarize, and present the recent developments in using LLMs and KGs under five different themes:

- 1. LLMs for KGs: Knowledge Extraction and Canonicalisation
- **2.** LLMs for KGs: KG Construction
- **3.** LLMs for KGs: Ontological Schema Construction
- 4. KGs for LLMs: Training and Augmenting LLMs

Figure 1 provides an overview. In Section 5.1, we start with the traditional KG construction from text and tables, then show in Sections 5.2 and 5.3 how assertions and ontological information can be extracted from LLMs. We then reverse perspective, and discuss in 5.4 how KGs can help and enhance LLMs in training and deployment.

5.1 LLMs for KGs: Knowledge Extraction and Canonicalisation

KG construction is a complex task that demands collecting and integrating information from a wide array of sources, encompassing structured, semi-structured, and unstructured data. Traditional methods often rely on specific modules designed to process each data type in isolation and struggle when the content is diverse and heterogeneous in structure. However, LLMs are powerful NLP

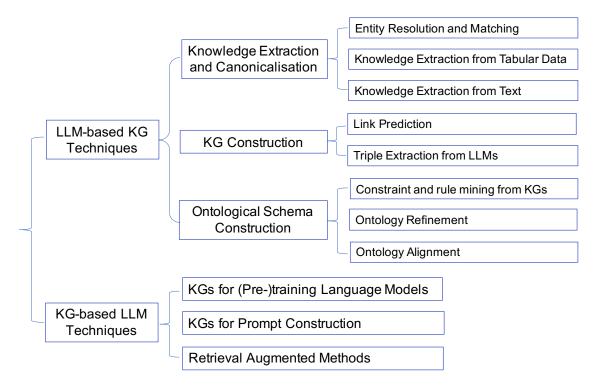


Figure 1 A Framework of Key Research Topics and Related Challenges.

models trained on a broad spectrum of information sources, making them well-suited for knowledge extraction tasks. This section presents work that uses LLMs for knowledge extraction from various sources.

5.1.1 Entity Resolution and Matching

Entity resolution (also known as entity matching, entity linking or entity alignment) is the process of linking pieces of information occurring in multiple heterogeneous datasets and referring to the same world entity [55, 59, 144]. Past research has focused on developing the methodologies and similarity measures among entities represented by flat structured data as well as semi-structured data. Recent benchmarks, among which a recent one [147], have been proposed to compare existing matching systems (including relational ones) on their accuracy, amount of corner cases and robustness to unseen matching pairs. A similar benchmark and survey tailored for knowledge graph entity alignment [225] focuses on the problem of alignment entities in different knowledge graphs using embedding-based techniques. One of the findings of the comparative analysis of existing systems is that semantic information (e.g. attribute triples and relation predicates) allow to achieve better matching accuracy.

Embedding-based entity alignment methods for KGs reduces the symbolic similarities between graph entities to a vector space in order to flatten the heterogeneity of graph components and facilitate reasoning [182]. Specifically, a total of 23 representative embedding alignment approaches are cross-compared in terms of performance but also are shown to require significant supervision in the labeling phase. Therefore, unsupervised methods and methods that cope with large-scale KGs are highly desirable in future research investigations.

CG-MuAlign [234] employs Graph Neural Networks (GNNs) to perform multi-type entity alignment, leverages the neighborhood information and generalizes to unlabeled types, and REA [148] which tackles the multilingual entity alignment problem by combining adversarial training with GNNs to cope with the problem of noisy labeled data provided as input.

2:8 LLMs and KGs: Opportunities and Challenges

LLMs are used in entity resolution and linking for KGs in multiple ways [7].

We envision the following important research challenges on the interplay between entity resolution and linking and LLMs.

- **Generation of labeled training data**, which is typically a resource-intensive and timeconsuming step, hindering the performance of entity alignment for KGs [94].
- Generation of corpora of entity matching rules, which will allow to increase the amount of rules available for these tasks on real-world knowledge graphs.
- Generative entity linking, is about using a generative language model performing entity linking as a sequence-to-sequence task [44, 93].

Similarly to how [170] employs Generative Adversarial Networks (GANs) to reduce the effort of labeling data, we argue that LLMs can provide labeled samples of KGs and control the performances of the aforementioned embedding-based approaches. Also, LLMs can help build robust and trustworthy corpora of entity matching rules, modulo the fact that a declarative formalism with a logical language L is defined in the graph setting. Training data for this logical language should be provided as input to LLMs, similarly to SQL statements that are readily available for consumption in text corpora. However, prompt engineering is needed in order to produce meaningful rule corpora for real-world large-scale KGs, such as DBpedia [9] and Wikidata [196]. Entity matching rule logs can be envisioned for these real-world large-scale KGs in a similar fashion as query logs for these KGs [19, 20].

Another interesting line of research work revolves around the alignment of factoid questions for building robust question answering (QA) systems. Existing approaches focus on the use of biencoders to efficiently perform both entity mention detection and entity linking in only one pass [110]. Matching of questions and predicate sequences on a knowledge base using both entity linking and deep convolutional neural networks has been investigated in prior work [218].

Other interesting directions to pursue for future developments leading to intertwining symbolic AI and statistical AI are as follows:

- Combination of entity linking approaches and embedding-based approaches with LLMs, which can lead to better integration of LLMs with knowledge reasoning and question answering [81].
- Marrying accuracy and interpretability, which will leverage the power of LLMs to generate graph data and rules and improve the accuracy of the entity linking processes but also enhance the interpretability of the processes.

5.1.2 Knowledge Extraction from Tabular Data

Extracting knowledge from tabular data like databases, Web tables and CSV files is a common way for KG construction. For tables whose semantics (meta information) are already known, heuristic rules can be defined and used to transform their data into KG facts. However, real-world tables often have unclear semantics with important meta information, such as table name and column header, not clearly defined. Meanwhile, the original data usually need to be retrieved, explored, integrated and curated, before expected knowledge can be extracted. In recent years, Transformer-based LMs have been investigated for processing tables, especially their textual contents. They can be applied to table vector representation as a foundation of other prediction tasks [194]. TURL [46] is a typical method of table representation learning that uses BERT [47] and has been applied in several tasks such as cell filling, column type annotation, and relation extraction. Similarly, RPT [188] uses BERT and GPT to pre-train a table representation model. Starmie [56] transforms columns into sequences using a template and fine-tunes BERT with a contrast learning framework using unionable and not unionable column pairs as samples. Among all the table processing tasks, semantic table annotation which matches table data to KG components (e.g., table column to KG class, table cell to KG entity, inter-column relationship to KG property) can be directly applied to extract knowledge for KG construction and population [121, 92]. There have been several attempts that use LLMs for these tasks. Doduo [181] serializes a table into a sequence of tokens and trains BERT for predicting column types and inter-column relationships. Korini et al. [103] prompts ChatGPT to annotate semantic column types. When task-specific samples for demonstration are minimal or absent, ChatGPT achieves a similar performance as the RoBERTa model. Although some attention has been given to utilizing LLMs for tabular data processing and KG construction, there is still much room for investigation.

Challenge and Opportunities.

- **Transforming table contents into sequences:** A table or a table element with its structured context needs to be transformed into a sequence before it can be fed into LLMs. Different transformation methods are required for different LLM utilization scenarios, such as fine-tuning LLMs, LLM inference with prompts, and instruction tuning of LLMs.
- Representing and utilizing non-textual tabular data: A table often contains not only long and short text, but also data of other types, like numbers and dates. There are still few works that consider these data.
- **Extracting tabular knowledge:** LLMs are mostly applied to process and understand tables but rarely applied to the final step of knowledge extraction. OntoGPT [27], which extracts instances from texts to populate an ontology using ChatGPT is known, but there are no counterparts for tables. Beyond instances, extracting relational facts is more challenging.

5.1.3 Knowledge Extraction from Text

Knowledge extraction from text generally entails automatic extraction of entities and associated relations, with traditional pipelines processing vast amounts of sentences and documents. This process enables the transformation of raw text into actionable knowledge, facilitating various applications such as information retrieval, recommendation systems, and KG construction. The language understanding capabilities of LLMs have enhanced this process.

For example,

- (1) Named Entity Recognition (NER) and Entity Linking, as mentioned in Section 5.1.1, involve identifying and categorizing named entities (such as persons, organizations, and locations) in text and linking (more in Section 5.2.1) them to KGs.
- (2) Relation extraction focuses on identifying and classifying relationships between entities, with LLMs that leverage zero-shot and few-shot in-context learning techniques [205, 111].
- (3) Event extraction aims to detect and classify events mentioned in the text, including their participants and attributes [197, 224].
- (4) Semantic Role Labeling (SRL) involves identifying the roles played by entities in a sentence, such as the subject, object and predicate [172, 230].

These approaches allow LLMs to extract information from text without needing a large amount of explicit training on specific domains, thereby improving their versatility and adaptability [44]. Moreover, LLMs have demonstrated proficiency in extracting knowledge from languages other than English, including low-resource languages, paving the way for cross-lingual knowledge extraction and enabling the utilization of LLMs in diverse linguistic contexts [106].

Furthermore, prompting LLMs has introduced new paradigms and possibilities in the field of NLP. LLMs can generate high-quality synthetic data, which can then be used to fine-tune smaller task-specific models. This approach, known as synthetic data generation, addresses the

2:10 LLMs and KGs: Opportunities and Challenges

challenge of limited training data availability and enhances the performance of models [94, 189]. Additionally, instruction tuning has emerged as a powerful technique where LLMs are trained on datasets described by explicit instructions, enabling more precise control and customization of their behavior for specific tasks [205, 201]. Also, for constructing domain-specific KGs, the stakes are higher, and hence scrutinizing the generated text (by experts) is necessary. However, it is still a step forward since human annotation is less expensive than human text generation.

Apart from the obvious substantial computational resource requirements for training and utilizing these LLM, there are various challenges, including the ones presented in Section 3.

Challenge and Opportunities.

- Efficient extraction from very long documents. Current LLMs are incapable of processing very long documents like novels in one shot. In this regard, modeling long-range dependencies and performing corpus-level information extraction can be further improved.
- High-coverage information extraction. Almost all extraction pipelines focus on high precision. However, high recall is ignored or under-explored [176]. Building knowledge extractors with high precision and high recall will result in a great leap toward building lifelong information extractors.

5.2 LLMs for KGs: Knowledge Graph Construction

We highlight the important role that LLMs play in improving KG construction, focusing on current trends, issues, and unanswered questions in this field. We start by discussing link prediction, which is a way to generate new facts given an existing KG. Next, we look at inductive link prediction, a method that predicts triples for unseen relationships. Our focus then shifts to a more recent method where triples are directly extracted from the parametric knowledge of an LLM.

As a conclusion of this section, we discuss the challenges of LLM-based methods for KG construction. These involve issues with long-tail entities, numerical values, and also the precision of these methods.

5.2.1 Link Prediction

Link prediction is about predicting a missing element of a triple given the other two elements. It includes head entity prediction (?, r, t), relation prediction (h, ?, t), and tail prediction (h, r, ?).

KG link prediction methods have mostly been studied for static snapshots of a KG. Many approaches, in fact, assume a training phase in which the current state of the KG is used to learn embeddings and other model parameters. Subsequently, many such models can only operate on entities for which an embedding was learned in the training phase. This leaves them incapable of predicting links for any previously unseen entities, such as newly added people or products. Inductive link prediction (ILP), in contrast, focuses on techniques that can predict links to new entities not originally contained in a KG. Furthermore, existing KG embeddingbased KG completion approaches frequently fail to leverage textual information and other literal information [65].

To overcome these challenges, current research focuses on incorporating textual information available in the KGs to improve the KG embeddings, boost performance in downstream tasks, and support ILP. The latent representation is learned from textual information using a variety of encoding models such as linear models, convolutional models, recurrent neural models, and LLMs and studied in [123, 142]. In this work, we focus only on LLM-based methods.

The LLM encoder BERT [47] is used in Pretrain-KGE [227] to generate initial entity embeddings from entity descriptions and relations. These embeddings are then fed into KG embedding models to generate final embeddings. MADLINK [15] uses SBERT to generate embeddings from entity descriptions, together with the entity embeddings obtained from structural information. KEPLER [202] offers a unified approach for KG embedding and pre-trained language representation, embedding text-enhanced knowledge and factual knowledge into LLMs. Nayyeri et al. [132] use LLMs to produce representations at word, sentence, and document levels, merging them with graph structure embeddings. Huang et al. [84] propose a model that combines LLMs with other modalities, such as vision, to generate a multi-model embedding space. CoDEx [3] uses a novel loss function driven by LLMs that helps KG embedding models estimate the likelihood of triples based on textual information. While these approaches can leverage the structural information contained in the graph, KG embeddings can not be directly used to predict unknown entities in the ILP setting. To still predict entities within ILP, existing works that combine text embeddings and traditional KG embeddings use only the text embeddings in the ILP setting [124] or apply similarity based heuristics to generate KG embeddings for unseen entities [198].

Instead of considering the graph structure, another line of research directly leverages LLMs for KG Completion. For example, KG-BERT [216] represents a triple as an ordered sequence of head text, including surface form, descriptions, and attributes, relation text, tail text separated with special tokens. KG-BERT [216] optimizes the BERT model on KGs, followed by KG-GPT2 [16] that fine-tunes the GPT-2 model. MTL-KGC [101] enhances the effectiveness of KG-BERT by combining prediction and relevance ranking tasks. PKGC evaluates triple validity by translating it into natural language sentences, while LLMs process these sentences for binary classification. Masked Language Models were introduced to encode KG text, with MEMKGC [38] predicting masked entities using the MEM classification model. Open world KGC [39] expands MEMKGC with modules Entity Description Prediction (EDP) and Incomplete Triple Prediction (ITP), focusing on predicting entities with given textual descriptions. StAR [199] uses Siamese-style textual encoders for text and a scoring module, while SimKGC [200] leverages a Siamese textual encoder. LP-BERT [112] is a hybrid KG completion method that combines MLM encoding for pre-training with LLM and separated encoding for fine-tuning, using a contrastive learning strategy. Also, LLMs such as GPT-3 and GPT-4 have the in-context learning capability which could adapt to new tasks through careful prompt design without fine-tuning the model parameters. Since LLMs are directly capable of predicting entities that are not yet contained in the KG, many works using direct LLM approaches also evaluate their models in the ILP setting [38, 39, 43, 199, 200, 202, 222].

Beyond the described approaches that construct and leverage embeddings for link prediction, LLMs may also be directly used in a prompting setting to find suitable links between entities, with existing methods described in detail in the following Section 5.2.2. As with the previous approaches that purely use LLMs, prompt-based approaches are directly applicable for ILP and are commonly evaluated in this setting [21, 90, 152, 174, 231]. LLMs make it significantly easier to jointly utilize structural and text information for link prediction, while there are still challenges.

Challenge and Opportunities.

- The generative language model which uses a decoder-only or encoder-decoder architecture cannot ensure that the generated result is already included in the KG. Also, one entity may have multiple natural language names. Thus it is hard to judge the correctness of the generated results from LLMs.
- Current link prediction models are mostly evaluated on ranking metrics, such as Hit@k and Mean Reciprocal Rank, which requires methods to give a ranked list of candidates. Considering that the number of candidate entities might be huge (e.g., over ten thousand), scoring each candidate entity in inference requires many model predictions which may be problematic for LLMs due to their large parameter size and high computational cost.

2:12 LLMs and KGs: Opportunities and Challenges

- The key research question of link prediction is how well a method could learn to infer new triples based on existing ones. LLMs are trained based on a massive corpus that might overlap with KGs such as Wikidata [196]. Thus it is not easy to distinguish whether the LLM completes the prediction by utilizing its memory or reasoning over existing facts.
- LLMs provide several benefits for ILP as demonstrated by the many existing approaches, but their application also has certain drawbacks. Since ILP entities may occur that are not contained within the underlying KG, the task is inherently reliant on auxiliary information. When leveraging LLMs for ILP, many studies focus on improving the extraction of knowledge from LLMs through prompt engineering, which is, by itself, a current popular research area that may directly benefit the domain.
- While prompting provides promising results, it requires well-designed strategies for predicting multi-token entities and out-of-vocabulary tokens. Further, if required knowledge is not captured by the LLM, e.g., when querying novel concepts that emerged after LLM pre-training, schemes to incorporate further knowledge into the LLM are required. Meanwhile, as feature selection, finding a suitable prompt also needs much searching with many experiments, which is costly especially for those recent commercial LLMs like GPT-4.
- Alternative methods rely on available high-quality textual descriptions of unknown entities. These methods generally inherit the limitations of classical link prediction, in that they require one-vs-all comparisons against all entity candidates during inference, which may become computationally prohibitive for large KGs [199, 200]. While there are already approaches that address this problem by applying efficient search heuristics, additional structural information, and pre-computation [44, 153, 199, 200], they involve additional memory overhead. As such, efficient strategies for obtaining predictions may provide a promising research direction.
- Further, the combination of these methods with classical link prediction models has already shown strong performance and is a promising trend followed by increasing numbers of works that further explore the incorporation of structural information into LLM-driven approaches.

Though challenges exist, opportunities lie in designing efficient and effective link prediction methods combining LLMs preserving the efficiency of traditional methods and robustness of language models.

5.2.2 LLMs for KGs: Triple Extraction from LLMs

Traditionally, retrieval and reasoning of relational knowledge have both relied on symbolic knowledge bases [60], that often are constructed using supervised extraction techniques applied to unstructured corpora, e.g. Web archives [220, 190]. More recently, self-supervised LLMs have been investigated for their ability to directly retrieve relational knowledge [233] from their parameters, e.g. through question answering, prompting through the use of cloze-style questions [76, 166] or statement scoring [183]. In this context, the ability of LLMs to retrieve, infer and generalize relational knowledge is seen as a crucial indicator of their capacity to understand and interpret natural language. Even though a range of terms are used in that context, e.g. fact or knowledge retrieval as well as knowledge inference, we refer to the task of accessing relational knowledge from LLM parameters as *knowledge retrieval*.

Benchmarks and Baselines. LAMA is the first benchmark dataset to evaluate knowledge retrieval in LLMs [152]. Related works show that knowledge retrieval through prompts is inconsistent with regard to paraphrasing [54, 76], with some types of information guiding LLMs towards more correct answers [26, 151, 35], while others are harmful to their performance [143, 100]. LLMs struggle to retrieve knowledge from low-frequency phenomena [162] and [86] argue that LLMs fail to express large varieties of knowledge when prompted in a zero-shot manner.

Zhong et al. [231] propose that the models' accuracy may originate from memorizing training data, not actually inferring knowledge. Similar to LAMA, the experiments on a more recent probing work KAMEL [95] confirm that LLMs are still far from the knowledge access capabilities of symbolic knowledge bases. The Knowledge Memorization, Identification and Reasoning test work KMIR [62] reveals that while LLMs struggle to robustly recall facts, their capacity to retain information is determined more by the number of parameters than the training methods, and while model compression can help preserve the memorization performance, it reduces the ability to identify and reason about the information in LLMs from transformer-based language models. Linzbach et al. [116] also present similar findings. LLMs are known to struggle with more complex reasoning tasks [83, 75]. Branco et al. [22] explore generalizability of common-sense reasoning capabilities and the impact of shortcuts in training data.

Biases in Triple Extraction Evaluation. LLMs may exhibit various types of biases; representation of the majority viewpoint being a common issue due to distributions prevalent within pretraining data [13], neglecting disagreements among multiple viewpoints (e.g. by majority voting) [42]. Prior works investigate individual factors (such as frequency) or LLM biases in other tasks [126], as well as knowledge retrieval [231].

With respect to the interpretation, reliability and generalizability of knowledge retrieval, several studies [22, 26] investigate whether LLMs actually learn transferable generalizations or only exploit incidental shortcuts in the data. [26] explore biases in three different knowledge retrieval paradigms, namely *prompt-based retrieval*, *case-based analogy*, *context-based inference*, finding that decent performance of existing knowledge retrieval baselines tends to be driven by biased prompts that overfit to artifacts in the data, guide the LLM towards correct entity types or unintentionally leak correct answers or additional constraints applicable to the correct answer. In a similar context, [50] discuss the shortcut learning behavior arising due to skewed training datasets, the model, or the fine-tuning process. [169] demonstrate an intriguing similarity between human cognitive biases and those exhibited by LLMs. Using insights from psychology, they analyze the learning and decision-making processes of black-box models to reveal their biases towards right-and-wrong for decision-making. Therefore, rigorous assessment of existing benchmark datasets is necessary for generalizable insights about knowledge retrieval and inference performance, and to facilitate efficient, unbiased knowledge retrieval from LLMs.

Prompt Engineering for Triple Extraction. Cao et al. [26] propose three paradigms for factual knowledge extraction from LLMs: prompt-based, case-based, and context-based. Results suggest the prompt-based retrieval is biased towards prompt structure. Prompt engineering [10] aims to create prompts that efficiently elicit desired responses from LLMs for a specific task. However, a limited number of manually created prompts only reveal a portion of the model's encoded knowledge [90], as the response can be influenced by the phrasing of the question. Thus, prompt engineering is a crucial part of knowledge retrieval from LLMs. LPAQA [90] uses an automated mining-based and paraphrasing-based method to generate diverse high-quality prompts, as well as ensemble methods to combine answers from different prompts. Automatic Prompt Engineer [233] uses LLM models like InstructGPT [138] and instruction induction [78] to generate instruction candidates which are then improved by proposing semantically similar instruction variants to achieve human-level performance. Zhou et al. [233] investigate the ability of LLMs, such as GPT-3, to generate high-quality prompts for a variety of tasks.Initial experiments on the role of syntax in knowledge retrieval [116] find a strong dependency on prompt structure and knowledge retrieval performance.

2:14 LLMs and KGs: Opportunities and Challenges

To summarize, prior works have shown that relational knowledge is captured by LLMs to a certain extent. However, there is still insufficient understanding of how performance differs across different kinds of knowledge or relations, for instance, commonsense knowledge compared to entity-centric encyclopedic facts or transversal versus hierarchical relations. In addition, several studies raise questions about LLMs capacity to infer knowledge beyond already-seen statements.

Challenges and Opportunities.

- **Entity Disambiguation:** Entity disambiguation is essential for KG construction to ensure unique identification of entities and to maintain consistency across the graph. However, when extracting facts from LLMs, entity disambiguation presents several challenges, since LLMs only operate on word token level. Hence, polysemy and homonymy make it difficult to determine the correct entity when a term has multiple meanings or is spelled the same as others but has different meanings. Also, the need to resolve co-references, where the same entity is mentioned in various ways within a text, further complicates the process. Moreover, the same piece of text can refer to different entities depending on the context, making it a significant challenge to correctly identify and classify the entities. Entities that were not present in the training data or are less common in general can be particularly hard to disambiguate. This can be a frequent issue with newer, less well-known, or very domain-specific entities. These complexities lead to major challenges that need to be addressed: enhancing disambiguation techniques to better handle long-tail entities; developing methods to better understand and utilize context in entity disambiguation; and improving co-reference resolution in such a way that it can be effectively incorporated into KG construction.
- **Long-tail Entities:** Existing LLMs still manifest a low level of precision on long-tail entities. Models may begin to generate incorrect information when they fail to memorize the right facts. The answers provided by these models often lack consistency. Incorrect correlations drawn from the pre-training corpus can lead to various biases in KG completion. Whether retrieval-augmented models serve as a viable solution to this problem remains uncertain, as does the potential necessity to adapt pre-training and fine-tuning processes to enhance model robustness in handling long-tail entities.
- High-Precision: LLMs face challenges in achieving high-precision predictions when performing knowledge retrieval [193]. A potential strategy to derive high-precision KGs from LLMs is to focus on model calibration. However, there are pressing challenges that remain unsolved. How can LLM training be adapted to prioritize high-precision learning? Can LLMs be used for validation purposes? These questions form the crux of the ongoing exploration in this field.
- **Provenance:** Extracting factual knowledge directly from LLMs does not provide provenance, the origin and credibility of the information, which presents multiple issues. Without provenance, verifying the accuracy of information becomes challenging, potentially leading to the spread of misinformation. Additionally, bias detection is hindered, as the lack of source information makes it difficult to account for potential biases in the data used for training. Provenance also provides critical context, without which information can be misunderstood or misapplied. Lastly, the absence of source information compromises model transparency, making it hard to evaluate the accountability of the LLMs.

5.3 LLMs for KGs: Ontological Schema Construction

A KG is often equipped with an ontological schema (including rules, constraints, and ontologies) for ensuring quality, enabling easier knowledge access, and supporting reasoning. At the same time, an independent ontology, which usually represents conceptual knowledge sometimes with

logics, can also be regarded as a KG. In this part, we describe the interaction between LLMs and approaches to learn ontological schemas as well techniques that can exploit LLMs to manage ontologies and rules.

5.3.1 Constraint and Rule Mining from KGs

The existing KGs are mostly obtained from large-scale data extraction pipelines, which are notoriously brittle and can introduce errors and inconsistencies in these graphs [48, 158]. Moreover, a KG is *never considered complete* since the closed world assumption does not hold [48, 146], i.e., it is not possible to conclude that a missing fact is false *unless it contradicts another existing fact*. Instead, we usually consider that in a KG it holds the open-world assumption, that is a missing fact is simply considered as *unknown*.

Practical applications impose high demands in terms of (semi-)automatic methods for data quality assessment and validation [102, 157, 2]. Since KGs contain huge amounts of data, it is not feasible to manually inspect and correct their errors. Therefore, a common approach is to instantiate rules and constraints that can be automatically enforced. These constraints express dependencies and conditions that the KG needs to satisfy at all times and that should not be violated by the introduction of new facts or their deletion. In KGs, rules and constraints can take the form of Graph Functional Dependencies [57], declarative first-order logic rules [61], entire ontologies [69], or *validating shapes* [102, 155]. Once a set of rules or constraints are instantiated, the next step is to either identify which entities or facts in the KG violate any of them, or employ them to delete erroneous information, or, finally, to employ them to deduce any missing information [58, 159].

Example 1. The following rules could apply for a subset of a graph describing people and their relationships:

- $r_1: \texttt{hasParent}(x, y) \Rightarrow \texttt{hasChild}(y, x);$
- r_2 : hasParent $(x, y) \Rightarrow \exists v_1, v_2 \in \mathbb{N} \mid hasBirthYear<math>(x, v_1) \land hasBirthYear(y, v_2);$
- r_3 :hasChild(x, y)AhasBirthYear (x, v_1) AhasBirthYear $(y, v_2) \Rightarrow v_1 < v_2;$

Where r_1 states that hasChild is the inverse equivalent relation of hasParent, r_2 states that for each person in the KGs for which we know the parent-child relationship we should know the birth year, and r_3 states that if y is a child of x then x should be born before y.

Nonetheless, a fundamental challenge is how to generate such rules and constraints. Specifying them manually is prohibitively difficult and expensive [2, 157]. On the one hand, the domain experts, who know the semantics for the dataset at hand, may not have the skill set or the background necessary to formally express those rules. Even when skilled, domain experts would require a substantial amount of manual work to exhaustively materialize a complete list of such rules [158]. Therefore, in the past decade, we have witnessed an increased interest in methods that can (semi-) automatically extract rules and constraints [158, 33]. In all those cases, the data in the KG is exploited as a *silver standard* [146]. That is, while the data are not perfect, we can still assume it is at least of reasonable quality. This approach works for rules that can find frequent positive examples. The challenge arises for rules that need to govern entities and facts that are rare, and for negative rules that need to prohibit some facts from appearing. To handle these cases, contextual information, e.g., documents and other user input provided in textual form, can be exploited in many different ways. In this context, LLMs bring a number of important opportunities:

2:16 LLMs and KGs: Opportunities and Challenges

- ability to extract contextual information from a given text, e.g., a document describing rules in natural language;
- information extraction from contextual data, like domain knowledge in textual form explicitly used during training;
- automatic generation of new rules via deductive chain-of-thought prompting;
- understanding rules vocabulary in context of a given domain description expressed in natural language, allowing to reconcile rules with synonyms and polysemy;
- providing explanations of rules or generating candidate validating or counterfactual examples.

Consider, for example, a rule where the relationships hasParent has been used, an LLM can be queries to identify whether hasParent between two entities refers more likely to the semantic meaning of being biologically descended or being owned or controlled. LLMs can be used to filter out candidate constraints, e.g., given their ability to mimic simple common sense reasoning, rules can be translated in the form of questions to the LLM, e.g., "if X is child of Y is it true that the birth year of Y comes always before the birth year of X?". Thus, an LLM can be used in support or replacement of a human annotator towards generated rules.

One of the most promising abilities of LLMs is parsing long texts. In companies and organizations, documents exist that contain reference governing information, e.g., procedures, regulations, and specifications. Here we see an untapped opportunity in parsing these documents in relation to the entities and predicates in the KG to extract constraints. Yet, the challenge arises in the fact that the LLM needs to use the correct vocabulary of entities and relations and the correct rule syntax. Recent advancements in program synthesis show a possible way to the challenge.

Finally, an even more fundamental challenge is that of transcending the usage of LLMs for NLP alone, and using them directly on large sets of facts within a KG. We can think of this setting as a special kind of multi-modal LLMs, where the KG is a specific modality. We have seen similar approaches in the past when word embedding approaches have been translated to knowledge graphs [165]. This brings to the question of whether it is possible to train LLMs to treat the task of rule generation as, for example, a graph embedding or summarization task [11, 28]. This would require then the ability to perform both inductive and abductive reasoning and treat rules as summaries from the set of facts in the KG.

A Rule Mining Case with ChatGPT. Let's have a look at the following real example asking ChatGPT to generate rules.

Example 2. We ask ChatGPT "Please list three rules that could help to infer one person's father. And output it in a first-order rule form." The following is the output we get:

"Certainly! Here are three rules expressed in first-order logic form that could help infer a person's father:

Rule 1: $\forall x, y \; (\texttt{child}(x, y) \land \texttt{male}(y) \rightarrow \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'child' relationship between x and y and y is male, then y is the father of x.

Rule 2: $\forall x, y \text{ (biologicalChild}(x, y) \rightarrow \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'biologicalChild' relationship between x and y, then y is the father of x.

Rule 3: $\forall x, y \; (\texttt{fatherOf}(y, x) \to \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'fatherOf' relationship between y and x, then y is the father of x.

These rules, expressed in first-order logic, demonstrate how the relationships between individuals can be used to infer the father of a person within a KG." From the above example, we can see that ChatGPT gives three rules that can possibly infer the father relationship. Rule 1 is correct which is consistent with the definition of father; Rule 2 is a rule with high probability; and Rule 3 is also correct since father and fatherOf are inverse relations. This demonstrates LLMs have rule knowledge and could express it in a formal form following instruction, thus they could serve as experts for automatic rule construction. To this end, however, we need to align the rules from the LLM to the existing knowledge in the KG, and feed the LLM with the existing knowledge and the context encoded in the KG. Both are consistent with discussion about the company and family examples.

Challenges and Opportunities.

- Retrieval augmented rule extraction and generation: LLMs can be given access to reference documents and KG and use this access to produce candidate rules informed by the textual content of the documents as well.
- **LLM guided rule explanation and validation:** LLMs can be used to identify from a list of generated rules those that are more likely to hold within a given context. Further, it can generate explanations on the meaning of the rule and its validity.

5.3.2 Ontology Refinement

Ontology refinement includes quite a few topics like knowledge completion (e.g., subsumption prediction, complex concept learning and new concept placement), erroneous knowledge detection and repair (e.g., inconsistency checking) and knowledge canonicalization (e.g., entity renaming). Besides formally represented knowledge, real-world ontologies, such as the widely used medical ontology SNOMED CT^1 and food ontology FoodOn², also include a lot of meta information defined by different annotation properties for usability, such as entity labels, synonyms and natural language definition. Taking the concept *obo:FOODON_00002809* in FoodOn as an example, it has not only formal knowledge such as named super concepts and logical restrictions, but also labels and synonyms (e.g., "edamame"), definitions (e.g., "Edamame is a preparation of immature soybean ..."), comments and so on. These meta information, especially the natural language text, further motivates people to use LLMs for ontology refinement.

For a refinement task, usually there are quite a few existing examples in the original ontology. Therefore, a straightforward solution, which has been adopted by most current methods, is fine-tuning a Pre-trained Language Model such as BERT together with an attached classifier. One typical method is BERTSubs [29] which is to predict the subsumption relationship between two named concepts, or between one named concept and one complex concept. It concatenates the corresponding texts of the two candidate concepts with special tokens as the input of a model composed of a pre-trained BERT and a logistic regression classifier, and fine-tunes the model with the existing subsumptions in the target ontology. For a named concept, the text could be either its name (label or synonym), or its name in combination with a textual description of its surrounding concepts; while for a complex concept, the text is its description (a.k.a. verbalisation). Another typical work is [119] which fine-tunes BERT and predicts the position to place in SNOMED CT for a new concept. Note that there are also some language model-based methods in taxonomy curation, such as [171] which fine-tunes BERT for taxonomy edge completion and GenTaxo [221] which fine-tunes a BERT variant named SciBERT for predicting positions that need new concepts. They can be directly applied or easily extended to refine an ontology's concept hierarchies.

¹ https://www.snomed.org/

² https://foodon.org/

2:18 LLMs and KGs: Opportunities and Challenges

Exploiting LLMs is a promising direction for ontology refinement, but it still needs much effort before they become practical tools. DeepOnto [73], which is a Python-based package that can support quite a few ontology engineering tasks, has already included some tools for ontology refinement and alignment using LLMs, but more development is needed to make it more accessible and to support generative LLMs like LLaMA and GPT-4. One obvious challenge is that those very recent generative LLMs have been rarely explored for ontology engineering. However, we think the following two research challenges are more fundamental.

Challenges and Opportunities.

- **Exploiting the graph structure and logic of an ontology together with its text**. Currently, LLM fine-tuning-based methods can well utilize the text of individual entities, but their other formal semantics are often ignored or not effectively incorporated. Besides fine-tuning with samples constructed by some templates, more LLM techniques such as prompt learning and instruction tuning could be considered.
- Combing symbolic reasoning with LLM inference. Symbolic reasoning, such as consistency checking in OWL ontologies, can still play a role e.g., in validating the knowledge inferred by LLMs. One aspect of incorporating symbolic reasoning is constructing samples for LLM fine-tuning and extracting prompts for LLM inference, while another aspect is the synergized framework [235] where LLM inference and symbolic reasoning work iteratively.

5.3.3 Ontology Alignment

The content of one single ontology is often incomplete and many real-world applications rely on cross-domain knowledge. Ontology alignment (a.k.a. ontology matching), which is to identify cross-ontology mappings between entities that have an equivalent, subsumption or membership relationship, thus becomes especially important for knowledge integration. The entity can be a concept (class), an individual (instance) or a property. Traditional systems (e.g., LogMap [91]) heavily rely on lexical matching and optionally use symbolic reasoning to remove mappings that lead to logical conflicts; while some recent methods combine these techniques with machine learning techniques like feature engineering, semantic embedding and distant supervision for better performance (e.g., LogMap-ML [30]). Especially, when the ontologies have a large ratio of assertions (large ABoxes) and the task is to discover equivalent individuals, ontology alignment is very close to the KG entity alignment task that has been widely investigated in recent years using KG embeddings [228]. As in ontology refinement, exploiting the textual information by applying LLMs is a promising direction for augmenting ontology alignment.

The study of LLM application in ontology alignment is similar to ontology refinement. Pretrained language models such as BERT have been applied via fine-tuning [133, 72]. BERTMap [72] is a typical system that has achieved state-of-the-art performance in many biomedical ontology alignment benchmarks. It fine-tunes a pre-trained LM with synonym pairs extracted from the original ontologies and the potentially given mappings, and combines the predicted concept equivalence scores with lexical matching scores and reasoning for the mappings. Those recent LLMs like GPT-4 have not been applied in ontology alignment, as far as we know, and the two fundamental research challenges mentioned in ontology refinement are applicable in ontology alignment. Besides, ontology alignment has another critical challenge.

Challenges and Opportunities.

Evaluating LLM-based ontology alignment systems. Novel evaluation protocols with new metrics are needed to fairly and efficiently compare LLM-based systems even with incomplete ground truth mappings [74]. Meanwhile, the semantics from the textual meta information and the LLM may be inconsistent with formal semantics defined in ontologies, and thus it is sometimes hard to determine whether a mapping by an LLM-based system is true or not.

5.4 KGs for LLMs: Training and Augmenting LLMs

In Sections 5.1 to 5.3, we discussed three different aspects on using LLMs for KGs. In this section, we look into the other direction, i.e., using KGs for LLMs. There are a few dimensions here. Firstly, KGs can be used as training data for LLMs. Secondly, triples in KGs can be used for prompt construction. Last but not least, KGs can be used as external knowledge in retrieval-augmented language models.

5.4.1 KGs for (Pre-)Training Language Models

KGs typically incorporate data derived from sources that are deemed highly reliable. The extracted data goes through a post-processing phase before it is certified by human evaluations. This vetted information from KGs has been incorporated into the pre-training corpus to complement natural language text, which on its own, can result in limited information coverage [216, 150, 1, 212].

Using factual knowledge from KGs to pre-train LLMs has also infused structured knowledge [130]. This integration of KGs with LLMs, along with efficient prompts, has made it convenient to inject world knowledge and incorporate new evolving information into language models [49]. Additionally, knowledge expressed in high-resource language KBs has been transferred into LMs tuned for low-resource languages [232, 118].

Furthermore, grounding knowledge from KGs to pre-train LMs has shown improvements in performance on generation and QA tasks [34, 164, 136]. In another approach, [192] proposed an interpretable neuro-symbolic KB, where the memory consists of vector representations of entities and relations from an existing KB. These representations are augmented to an LM during pre-training and fine-tuning, enabling the model to excel in knowledge-intensive QA tasks.

Challenges and Opportunities.

- High-quality knowledge injection: KGs can prove to be a valuable source of knowledge, fostering a more profound understanding of the relationships between entities within LLMs this is particularly true as existing KGs such as Wikidata [196] are committed to ensuring high precision. Moreover, KGs' ability to include dynamic updates with the latest information can provide significant advantages during pre-training in comparison to outdated textual corpus.
- Information duplication: KGs encapsulate direct and indirect relations, and one can induce the same information using multiple facts. Such a kind of complex linking between entities is generally hard to find in a single textual source. Therefore, leveraging KGs can enhance the LLM's understanding and reasoning capacities about popular as well as long-tail entities.

5.4.2 KGs for Prompt Construction

The attention received by the integration of KGs and LLMs has grown recently. On the one hand, there is the explored direction of prompting LLMs for collecting and distilling knowledge in order to make it available to the end-users. On the other hand, there is the less explored research direction where KGs are used in synergy with prompts in order to enhance LLMs with capabilities making them more effective and, at the same time, trustworthy. A number of studies have leveraged KGs to enrich and fine-tune prompt creation resulting in a significant increase in prompt quantity, quality, and diversity compared to manual approaches. KGs have been employed in single and in multi-turn scaffolding prompts at scale, powered by numerous traversal paths over KGs with low authoring cost while considering the meaningful learning patterns [107]. Other studies have investigated how incorporating explicit knowledge from external sources like KGs can help prompt engineering, especially by giving additional contexts (e.g., attributes, K-hop neighbors) of the entities in order to help the LLMs to generate better predictions [23]. Approaches like

2:20 LLMs and KGs: Opportunities and Challenges

KnowPrompt [35] use KGs to incorporate semantic and prior knowledge among relation labels into prompt-tuning for relation extraction, enhancing the prompt construction process and optimizing their representation with structured constraints. Certain studies have utilized LLMs and prompts in the task of reasoning over KGs [40], e.g., LARK uses entities and relations in queries to find pertinent sub-graph contexts within abstract KGs, and then, performs chain reasoning over these contexts using LLM prompts of decomposed logical queries outperforming previous state-of-the-art approaches by a significant margin.

Challenges and Opportunities. The current research in the field of KG utilization for prompt creation predominantly centers around LLMs, which are considered to have relatively lower efficacy compared to LLMs. LLMs present significant potential for advancing prompt creation methodologies in conjunction with KGs. We may summarize this perspective within the following four challenges:

- KGs can be employed to automatically extract and represent relevant knowledge to generate context-aware writing prompts. Analyze and understand the relationships between different writing prompts, enabling the generation of prompts that build upon each other.
- KGs can be combined with LLMs to facilitate the interactive and dynamic generation of prompts, adapting to user feedback and preferences in real-time. Furthermore, the use of KGs in prompt creation has opened up possibilities for explainability and interpretability. Since KGs provide a transparent representation of knowledge, the prompts generated from KGs can be easily traced back to their underlying sources.
- KGs can integrate into prompts the definitions of guards exploited during the generative task. Such guards may enhance the trustworthiness of the information generated by LLMs and make them more compliant with specific domain-wise or context-wise constraints.
- KGs can create prompts that ask questions (e.g., inferring missing relations in an incomplete KG) that trigger KG complex reasoning capabilities and intermediate reasoning steps.

The integration of KGs within the prompt construction activities will allow us to answer the following preparatory set of research questions. For each research question, we provide a link to the challenges mentioned above, aiming to identify appropriate research pathways.

- **RQ1:** How can KGs be integrated into existing prompts to enhance the effectiveness (including relevance, absence of bias, privacy-preservation) of the information extracted from LLMs? → C1, C2, C3.
- **RQ2:** How can KGs be exploited to drive the generative capabilities of LLMs in order to properly address the whole ethical constraints of AI-based solutions? \longrightarrow C2, C3.
- **RQ3:** What are the optimal approaches to generate KGs-based prompts that enhance reasoning capabilities? → C4.

Also, in the following three scenarios, the community would benefit from tackling the abovementioned challenges and research questions.

- **KGs for Hallucination Detection in LLMs:** The reliability of LLMs is greatly affected by the hallucination problem, where they generate inaccurate information. Despite attempts to address it, the issue of hallucination is likely to persist in the realm of LLMs for the foreseeable future. To aid in the detection of hallucinations, KGs-based prompting aims to offer reliable information that can serve as a foundation. By combining LLMs and KGs, researchers can develop a comprehensive prompt-based fact-checking model that can identify hallucinations in various domains.
- **KGs for Editing Knowledge in LLMs:** LLMs possess the ability to store extensive realworld knowledge, but they struggle to exploit prompts to update their internal knowledge to reflect real-world changes. KGs-based prompts offer a potential solution for modifying

knowledge in LLMs, but they are restricted to handling basic tuple-based knowledge in KGs. Indeed, even if the entire LLM undergoes re-training, the knowledge presented through prompts would likely be assimilated within the vast network structure of the LLM.

KGs for Black-box LLMs Knowledge Injection: While pre-training and knowledge editing methods can bring LLMs up to date with the latest information, they require access to the internal structures and parameters of LLMs. However, many cutting-edge large LLMs only offer APIs that allow users and developers to interact with them, keeping their internal workings hidden from the public. Consequently, traditional KG injection techniques that involve modifying LLM structures with additional knowledge fusion modules cannot be employed. One potential solution is to convert various types of knowledge into different text prompts. However, it remains an area of ongoing research to determine if these prompts can effectively adapt to new LLMs. Additionally, the approach of using KGs-based prompts is constrained by the length of input tokens accepted by LLMs. Therefore, the question of how to enable effective knowledge injection for black-box LLMs still remains unanswered.

5.4.3 Retrieval Augmented Methods

There are a few reasons for why retrieval augmented methods are necessary for LLMs to obtain external knowledge. One reason is to address the problem of knowledge cutoff, i.e., LLMs are not aware of the events that happened after their training. Also, although parametric knowledge would increase when the size of parameters increases, training LLMs is expensive; e.g., GPT-3 (175B parameters) costs \$4.6 million to train, and PaLM (540B parameters) costs \$17 million. In fact, research suggests that the obtained knowledge from such training is mainly about popular entities [125]. Furthermore, for domain specific applications, there might be some significant knowledge that is not yet in LLMs, including private and business critical knowledge that cannot be put into LLMs.

One idea to deal with the above lack of (updated) knowledge is to edit the knowledge in LLMs. Knowledge Editing methods assume that we know which entries are wrong and how to fix them. An obvious strategy is to retrain and fine-tune the model based on the modified data. However, apart from being costly, retraining cannot guarantee that erroneous data will be corrected. Another strategy is to develop a hyper-network to learn a parameter shift for the base model. De Cao et al.[45] trained a hyper-network, KnowledgeEditor, to modify a fact and used Kullback-Leibler (KL) divergence-constrained optimization to alleviate the side effect on other data/knowledge that should not be changed. However, many state of the art methods do not perform well when editing *multiple edits*, as they use the same strategy to process multiple edits and ignore the relation between different edit gradients, resulting in a "zero-sum" phenomenon, where the inter-gradient conflict will inevitably cause some data modifications to fail. Han et al. [71] design explicit and implicit multi-editor models to learn diverse editing strategies in terms of dynamic structure and dynamic parameters respectively, allowing to deal with the conflict data in an efficient end-to-end manner.

Alternatively, people could look into retrieve-generate architectures for building retrieval augmented generation models. In this setting, we assume that the external sources contain the needed knowledge for addressing users' requirements. Initially, most such methods use unstructured passages as external knowledge. RAG [109] outperforms DPR [99] by marginalizing the retrieval step to train the generator and retriever jointly with the supervision of the label answer. RAG tends to focus too much on label-relevant passages, rather than question-relevant passages and answers. Huang et al. [85] address this issue by incorporating rich answer encoding through Dense Knowledge Similarity (DKS) and Retriever as Answer Classifier (RAC). FiD [87] encodes the concatenation of the passages retrieved by pre-trained DPR and the original question separately,

2:22 LLMs and KGs: Opportunities and Challenges

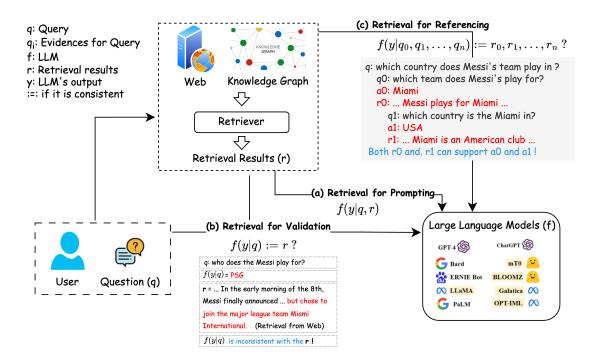


Figure 2 Different Tasks for Retrieval Augmented Methods.

and then fuses them with concatenation to the decoder. It is expected that structured knowledge will be the main source of external knowledge (cf. Figure 2), as passages often contain noise. Knowledge Graphs can be used directly as external knowledge. For example, KG-FiD [219] utilizes knowledge graphs to further enhance the retrieval quality by establishing the structural relationship among the retrieved passages. Another external source is databases. Vougiouklis et al. [195] propose an efficient NL2SQL approach, transforming questions in natural language into SQL over given database schemas, outperforming ChatGPT under the zero-shot setting while achieving about 10 times speedup over other state of the art NL2SQL engines.

As illustrated in Figure 2, it is important to note that there are different tasks for retrieval augmented methods to enhance large language models. In addition to the above mentioned retrieval augmented methods for prompting, there are also methods for other tasks. For example, Zhao et al. [229] propose the Verify-and-Edit framework, which makes use of retrieval augmented methods to validate texts generated by large language models for individual steps within CoT-style reasoning chains. Yang et al. [214] proposes LLMA accelerator to make use of retrieval augmented methods for speeding up decoding of large language models. Gao et al. [63] propose the RARR system, which retrieve evidence for texts generated by large language models for referencing purpose. Apart from retrieval augmentation for generations, retrieval augmented methods can be used for pre-training and fine-tuning [70] as well. Knowledge graphs are expected to play key roles in these tasks.

Challenges and Opportunities. To sum up, retrieval augmentation is a very promising direction. There are a few pressing challenges:

Unifying Knowledge Editing and Retrieval Augmentation: KGs can be used for editing knowledge in LLMs, while at the same time, KGs can also be used as external knowledge to assist LLMs in retrieval augmented methods. In fact, knowledge editing and retrieval

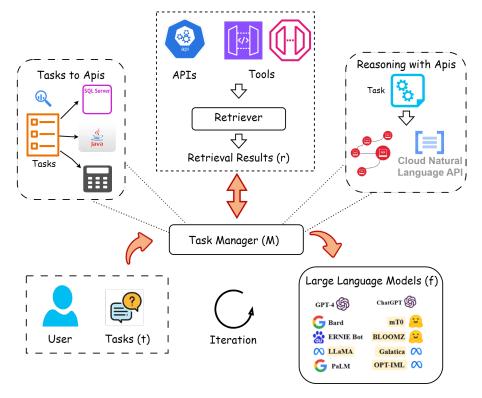


Figure 3 Tools and APIs as External Knowledge Sources.

augmentation is getting very close. For example, Mitchel et al. [129] proposed a Retrieval-Augmented Counterfactual Model (SERAC), which stores edits in an explicit memory for knowledge editing over LLMs. However, SERAC primarily focuses on batch editing, assuming that all editing data is known. Han et al. [177] propose a Retrieval Augmented Sequential Model Editing framework (RASE) that enhances editing generalisation and could work with different knowledge editors, supporting thousand's of edits.

- Semi-parametric LLMs: This direction is highly related to the topic of this position paper. The idea is to make use of explicit knowledge to augment LLMs. One of the key issues is to integrate different explicit knowledge [184], including unstructured ones, such as passages, and structured ones, such as KGs and databases, for augmenting LLMs.
- **Support of Complex Reasoning**: Can we go beyond simply retrieving explicit knowledge by enabling reasoning through retrieval augmented methods? BehnamGhader et al. [12] demonstrated with their experimental results that the similarity metric used by the retrievers is generally insufficient for reasoning tasks. Furthermore, LLMs do not take the complicated relations between statements into account, thus leading to poor reasoning performance.
- Attribution Evaluation and Revision: Can we retrieve relevant passages or structure knowledge for some generated text t for a given query q (or a dialog [52])? If it turns out that t is not appropriate for q, can we produce some revised text t' instead [64]?
- **Tool Use and Agent**: In a sense, retrieval augmented methods are mainly using a retriever to offer external knowledge. One step further is to use tools [168] and APIs [156] as external source of knowledge, as illustrated in Figure 3. There will be lots of room for Knowledge Graph related research in this space.

6 Some Thoughts on Applications

The integration of KGs and LLMs in a unified approach holds significant potential, as their combination mutually enhances and complements each other in a valuable manner. For instance, KGs provide very accurate and explicit knowledge, which is crucial for some applications i.e. healthcare, whereas LLMs have been criticized for their lack of factual knowledge leading to hallucinations and inaccurate facts. secondly, LLMs lack explainability instead, KGs given their symbolic reasoning ability, are able to generate interpretable results. On the other hand, KGs are difficult to construct from unstructured text and suffer from incompleteness therefore, LLMs could be utilized in addressing these challenges by text processing. Various applications have adopted this methodology of combining LLMs with KGs, such as healthcare assistants³, question answering systems [217] or ChatBots, and sustainability, among others.

6.1 Commonsense Knowledge

The majority of KGs capture facts of the sort one might encounter in an encyclopedia or in a relational database. However, commonsense knowledge is another important form of world knowledge for AI systems. For instance, we may wish for a KG to not only capture that the Congo rainforest lies in Central Africa, but also that tropical rainforests have significant rainfall and lush green vegetation. ConceptNet is the most well-known commonsense knowledge graph, developed using manual crowdsourcing along with automated refinement techniques [120]. However, crowdsourcing is very labor-intensive and costly, so alternative means of harvesting such knowledge have long been sought.

Commonsense Knowledge from LLMs. The first study to investigate extracting knowledge from a language model to the best of our knowledge was indeed one that targeted commonsense knowledge [185]. The authors mined commonsense triples such as hasProperty (apples, green) from the Google Web 1T n-gram data as well as from Microsoft's Web-scale smoothed language models [82]. This was later extended into a large-scale commonsense knowledge graph [187] that covered a range of different relations and became a part of the WebChild KG [186].

As both crowdsourcing and information extraction from text are likely to lead to incomplete knowledge, a key challenge is how to generalize beyond what has been collected. The WebBrain project explored neural knowledge graph completion [32] for better generalization. COMET, short for COMmonsEnse Transformer [21], and the improved COMET-ATOMIC 2020 [86], used existing data to fine-tune Transformer-based models. This line of work considers the original ConceptNet relations as well as reasoning-related knowledge pertaining to events, causes, and effects, e.g., what goals might have motivated a person A to leave an event without person B. Recently, [208] shows how common-sense triples could be extracted from an LLM and use through distillation to transfer knowledge into a smaller LM, outperforming the larger one. Overall, fine-tuned LLMs are found to outperform off-the-shelf LLMs, while also benefiting from the advances of the latter.

Challenges and Opportunities. Commonsense knowledge, in particular, is genuinely open-ended, such that it depends on a number of considerations whether it makes sense to attempt to materialize relevant knowledge beforehand or rather invoke a (possibly slow) LLM on the fly. Commonsense knowledge may also differ substantially between different cultures [134]. This also leads to the

³ https://neo4j.com/blog/doctor-ai-a-voice-chatbot-for-healthcare-powered-by-neo4j-and-aws/

question of what kinds of biases are acceptable. Finally, a long-term challenge is how to capture knowledge that is not easily expressed in language, e.g., how a robot ought to grasp different kinds of objects.

6.2 Digital Build Environment

In the domain of Build Environment, where it is vital to design and construct in a safe and sustainable way, a number of regulations and guidelines need to be met. Automated Compliance Code Checking has bloomed to support this, but still, in the past, it was quite challenging to interpret regulations and execute rules in 3D models automatically due to the ambiguity of the text, and the need for extensive expert knowledge for interpretation. Many applications now in this area have combined LLMs + KGs in addressing these challenges by following an LLMs enhanced KG approach [235]. LLMs interpret the text in the regulations and enhance a KG of rules, which are further serialized using dome Domain Specific Language [226, 203]. Interrogating 3D models modeled as graphs using the Linked Building Data approach is another challenge for the domain as it requires skill sets in query languages like SPARQL. LLMs are helping in understanding human language written questions and converting those into relevant query languages by bringing a new way of how domain experts interact and interrogate 3D models and their various forms. An example of this is AI Speckle ⁴.

6.3 Digital Healthcare

The Digital Healthcare sector holds immense potential for various possibilities concerning the adoption of LLMs, including the automation of clinical documentation, the synthesis of patient histories, and the identification of potential candidates for clinical trials. Although these advancements are remarkable, it is crucial to recognize the potential risks associated with employing LLMs in healthcare. Indeed, Digital Healthcare is one of the most critical application domains for the adoption of LLMs. The needs of the major stakeholders (i.e., physicians, healthcare providers, and policymakers) row against the paradigm behind the creation of LLMs. In particular, the two major significant risks related to the model's accuracy and the privacy concerns stemming from its usage.

Accuracy. Some demonstrations of LLMs have showcased impressive capabilities. Nevertheless, there have also been documented instances where LLMs have made mistakes or exhibited erratic behavior. In the Digital Healthcare sector, where patient safety is of utmost importance, it is crucial for healthcare organizations to comprehend the potential risks associated with LLM usage. When utilized to diagnose hypothetical patient cases, LLMs have exhibited accuracy at a level comparable to that of a third- or fourth-year medical student, albeit not reaching a professional's proficiency. Despite this high level of performance, LLMs have also been known to generate false information, invent sources, commit logical errors, and provide answers that are inappropriate or unethical. The integration of KGs would definitely enhance the capabilities of LLMs given the possibility of injecting domain-specific knowledge able to mitigate the issues mentioned above. Avoiding hallucinations and preserving ethics are definitely the two major aspects to which LLMs+KG may contribute in a significant manner.

⁴ https://speckle.systems/blog/ai-powered-conversations-with-speckle/

2:26 LLMs and KGs: Opportunities and Challenges

Privacy. One major concern with LLMs is that employing any third-party application necessitates the transmission of data to that party. When data, including protected health information (PHI), is managed by a covered entity like a hospital, it becomes subject to the regulations of the jurisdiction where the entity is located (e.g., GDPR). Furthermore, by sending PHI to additional third parties, organizations lose control over how that data will be handled. For instance, healthcare organizations are unable to determine the exact storage location of their data, whether it will be mixed with data from other organizations and utilized to train future language models, or what security measures are in place to safeguard the data. Healthcare organizations employing LLMs must recognize that their data are potentially more susceptible to breaches or misuse. The role of preserving private information may be played by KGs through the modeling of axioms defining which data may be shared and with who and how personal knowledge may be anonymized in order to be transmitted to possible external systems.

An alternative approach, prioritizing privacy, involves running an open-source LLM within the infrastructure of a healthcare organization. This way, it would be possible to directly work on the parameters of LLMs given the possibility of having control of the injected knowledge and, at the same time, ensuring that data are never shared with a third party. However, these open-source models are currently not as advanced or extensively trained as the more popular commercial systems (e.g., ChatGPT). Additionally, the effort to create the knowledge resources to inject and the expertise required to program and maintain an open-source LLM may not be readily available to many healthcare organizations.

The two risks described above lead to likewise challenges that must be tackled in order to make a significant step toward the adoption of these models within the clinical practice.

6.3.1 Domain Specific Content Search

Recently we have witnessed the success of models like GPT-4 [137] in a multitude of domain-specific NLP applications that involve multiple modalities and domain specific adaptations. In these cases they need to be able to access data stored in multiple modalities (e.g., pictures, diagrams, scientific images, PDF documents, or system logs) and an abundance of domain specific information that has not been accessed during training. In these use-cases, while LLMs are often treated as generative models, they can be easily adopted in search and reasoning tasks when integrated in specific tools and pipelines ⁵. Nonetheless, as seen in the past all these specialized domains can better benefit from the inclusion of symbolic knowledge in machine methods [6]. Here we focus on two prototypical applications: (1) semantic image and video search and (2) technical document understanding.

Recently, many methods, primarily based on deep learning models such as CLIP [160] and BLIP2 [113], achieved state-of-the-art performance on image retrieval tasks. These multi-modal models jointly learn vector embeddings for images and text, such that the embedding of the image should be close to that of the text that describes that image. Nonetheless, we have also seen increased interest in obtaining a more symbolic representation of the contents of an image [145, 105]. Datasets like Visual Genome [105] annotate images with *scene graphs*. A scene graph is a small KG that describes, with a structured formal graphical representation, the contents of an image in terms of objects (people, animals, items) as nodes connected via pairwise relationships (e.g., actions or positioning relationships) as edges. Therefore, Multimodal LLM can be trained to reason and exploit this additional representation offering an advanced ability to understand the contents of an image (or a video). Importantly, a scene graph node and edge can be annotated

⁵ For example https://haystack.deepset.ai/

(grounded) with features and positions from the image (e.g., relative size). This can be exploited in applications like image and video search and question answering. When entities in a scene graph are connected to a background taxonomy or KG, then questions that require forms of abstraction and computation, e.g., What is the genus of the largest animal in the picture? What is the average price of the car in the picture? Retrieve images depicting kids wearing vegan friendly items of clothing in the catalog.

On the other hand, the digitalization of domain specific documents, e.g., especially contracts, is enabling in-depth applications of machine intelligence to help humans more effectively perform time-consuming tasks. Among these, contract review costs humans substantial time, money, and attention (many law firms spend approximately 50% of their time reviewing contracts, costing hundreds of thousands of dollars) [77]. The Contract Understanding Atticus Dataset (CUAD) is a new dataset for legal contract review [77]. CUAD was created with legal experts and consisted of over 13,000 annotations. Recent advancements in generic pre-trained language models showed their power in some text-understanding tasks. Therefore, we have seen different domain specific applications of NLP and LLM methods [209, 215]. Yet, they do not have access to all relevant knowledge and are ill-suited for certain calculations [98]. This can be solved by exploiting external domain specific symbolic information, e.g., domain specific knowledge graphs [135, 117, 66], and by adding symbolic and reasoning capabilities to the LLMs [98]. This promising direction will lead to extending current state-of-the-art neuro-symbolic methods to combine the advantages of a graph representation when extracting symbolic knowledge from complex documents, the ability to understand long-form unstructured texts of LLMs, and the good properties of domain-specific code synthesis of LLMs to address advanced retrieval and question answering use-cases, e.g., How many weeks are left before the expiration of this contract? What are the conflicts between this liability clause and previous contracts my company signed? What is the risk of side-effects of this treatment for a patient with this given health record? Get the average yearly yield and spread for competitor companies of ACorp and their recent acquisitions in the EU market.

7 Outlook

In conclusion, the recent advances on large language models (LLMs) mark an important inflection point for knowledge graph (KG) research. While important questions on the ability to combine their strengths remain open, these offer exciting opportunities for future research. This paper has covered a number of relevant topics about KG and LLM. There are some others that may have not been widely investigated at the current stage but are worthy of mentioning, including LLM for multi-modal and temporal KG construction [36, 211], LLM for KG question answering [80], multilingualism of LLM and KG [178], KG for LLM instruction tuning and human alignment, KG for faster LLM inference and so on. The community is already rapidly adapting their research focus, with novel forums like the KBC-LM workshop [96], the LM-KBC challenge [175] and the LLM sub-track of OAEI Bio-ML⁶ arising, and resources massively shifting towards hybrid approaches to knowledge extraction, consolidation, and usage. We give out the following recommendations:

- 1. Don't throw out the KG with the paradigm shift: For a range of reliability or safety-critical applications, structured knowledge remains indispensable, and we have outlined many ways in which KGs and LLMs can fertilize each other. KGs are here to stay, do not just ditch them out of fashion.
- 2. Murder your (pipeline) darlings: LLMs have substantially advanced many tasks in the KG and ontology construction pipeline, and even made some tasks obsolete. Take critical care in examining even the most established pipeline components, and compare them continuously with the LLM-based state of the art.

⁶ https://www.cs.ox.ac.uk/isg/projects/ConCur/oaei/2023/index.html

2:28 LLMs and KGs: Opportunities and Challenges

- **3.** Stay curious, stay critical: LLMs are arguably the most impressive artifact of AI research of the past years. Nonetheless, there exist a magnitude of exaggerated claims and expectations in the public as well as in the research literature, and one should retain a healthy dose of critical reflection. In particular, a fundamental fix to the so-called problem of hallucinations is not in sight.
- 4. The past is over, let's begin the new journey: The advances triggered by LLMs have uprooted the field in an unprecedented manner, and enable to enter the field with significant shortcuts. There is no better time to start anew in fields related to Knowledge Computing, than now.

Although the direction of the present transformation is widely open, as researchers continue to explore the potentials and challenges of hybrid approaches, we can expect to see new breakthroughs in the representation and processing of knowledge, with far-reaching implications for fields ranging from Knowledge Computing to NLP, AI, and beyond.

- References -

- 1 Oshin Agarwal, Heming Ge, Siamak Shakeri, et al. Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pretraining. In NAACL, pages 3554–3565, jun 2021. doi:10.18653/V1/2021.NAACL-MAIN.278.
- 2 Naser Ahmadi, Viet-Phi Huynh, Vamsi Meduri, Stefano Ortona, and Paolo Papotti. Mining expressive rules in knowledge graphs. J. Data and Information Quality, 12(2), 2020. doi:10.1145/ 3371315.
- 3 Mirza Mohtashim Alam, Md Rashad Al Hasan Rony, Mojtaba Nayyeri, Karishma Mohiuddin, MST Mahfuja Akter, Sahar Vahdati, and Jens Lehmann. Language model guided knowledge graph embeddings. *IEEE Access*, 10:76008–76020, 2022. doi:10.1109/ACCESS.2022.3191666.
- 4 Dimitrios Alivanistos, Selene Báez Santamaría, Michael Cochez, Jan-Christoph Kalo, Emile van Krieken, and Thiviyan Thanapalasingam. Prompting as probing: Using language models for knowledge base construction, 2022. doi:10.48550/ ARXIV.2208.11057.
- 5 Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. A review on language models as knowledge bases. arXiv, 2022. doi:10.48550/ARXIV.2204.06031.
- 6 Mona Alshahrani, Mohammad Asif Khan, Omar Maddouri, Akira R Kinjo, Núria Queralt-Rosinach, and Robert Hoehndorf. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics*, 33(17):2723–2730, apr 2017. doi: 10.1093/BIOINFORMATICS/BTX275.
- 7 Sihem Amer-Yahia, Angela Bonifati, Lei Chen, Guoliang Li, Kyuseok Shim, Jianliang Xu, and Xiaochun Yang. From large language models to databases and back: A discussion on research and education. DASFAA, abs/2306.01388, 2023. doi:10.48550/ARXIV.2306.01388.
- 8 Alejandro Barredo Arrieta, Natalia Díaz Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, A. Barbado, Salvador García, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges

toward responsible ai. *Information Fusion*, 2020. doi:10.1016/J.INFFUS.2019.12.012.

- 9 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735, 2007. doi:10.1007/978-3-540-76298-0_52.
- 10 Stephen H Bach, Victor Sanh, Zheng-Xin Yong, Albert Webson, Colin Raffel, Nihal V Nayak, Abheesht Sharma, Taewoon Kim, M Saiful Bari, Thibault Fevry, et al. Promptsource: An integrated development environment and repository for natural language prompts. ACL, 2022. arXiv: 2202.01279, doi:10.48550/arXiv.2202.01279.
- 11 Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. From arguments to key points: Towards automatic argument summarization. In ACL, pages 4029–4039, 2020. doi:10.18653/V1/2020.ACL-MAIN.371.
- 12 Parishad BehnamGhader, Santiago Miret, and Siva Reddy. Can retriever-augmented language models reason? the blame game between the retriever and the language model. In *arXiv*, 2022. doi:10.48550/ARXIV.2212.09146.
- 13 Emily M Bender, Timnit Gebru, Angelina McMillan-Major, et al. On the dangers of stochastic parrots: Can language models be too big? In FAT, pages 610–623, 2021. doi:10.1145/ 3442188.3445922.
- 14 Emily M. Bender and Alexander Koller. Climbing towards NLU: On meaning, form, and understanding in the age of data. In ACL, 2020. doi:10.18653/V1/2020.ACL-MAIN.463.
- 15 Russa Biswas, Harald Sack, and Mehwish Alam. Madlink: Attentive multihop and entity descriptions for link prediction in knowledge graphs. SWJ, pages 1–24, 2022. doi:10.3233/SW-222960.
- 16 Russa Biswas, Radina Sofronova, Mehwish Alam, and Harald Sack. Contextual language models for knowledge graph completion. In *MLSMKG*, 2021. doi:10.34657/7668.
- 17 Su Lin Blodgett, Solon Barocas, Hal Daum'e, and Hanna M. Wallach. Language (technology) is power: A critical survey of "bias" in nlp. ACL, 2020. doi:10.18653/V1/2020.ACL-MAIN.485.

- 18 Bernd Bohnet, Vinh Q Tran, Pat Verga, Roee Aharoni, Daniel Andor, Livio Baldini Soares, Jacob Eisenstein, Kuzman Ganchev, Jonathan Herzig, Kai Hui, et al. Attributed question answering: Evaluation and modeling for attributed large language models. arXiv preprint arXiv:2212.08037, 2022. doi:10.48550/ARXIV.2212.08037.
- 19 Angela Bonifati, Wim Martens, and Thomas Timm. Navigating the maze of wikidata query logs. In WWW, pages 127–138, 2019. doi: 10.1145/3308558.3313472.
- 20 Angela Bonifati, Wim Martens, and Thomas Timm. An analytical study of large SPARQL query logs. VLDB J., 29(2-3):655–679, 2020. doi:10.1007/S00778-019-00558-9.
- 21 Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense transformers for automatic knowledge graph construction. In ACL, pages 4762–4779, 2019. doi:10.18653/V1/ P19-1470.
- 22 Ruben Branco, António Branco, João António Rodrigues, et al. Shortcutted commonsense: Data spuriousness in deep learning of commonsense reasoning. In *EMNLP*, pages 1504–1521, nov 2021. doi:10.18653/V1/2021.EMNLP-MAIN.113.
- 23 Ryan Brate, Minh Hoang Dang, Fabian Hoppe, Yuan He, Albert Meroño-Peñuela, and Vijay Sadashivaiah. Improving language model predictions via prompts enriched with knowledge graphs. In Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Virtual Conference, online, October 24, 2022, volume 3342 of CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3342/paper-3.pdf.
- 24 Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *neur-IPS*, 33:1877-1901, 2020. URL: https: //proceedings.neurips.cc/paper/2020/hash/ 1457c0d6bfcb4967418bfb8ac142f64a-Abstract. html.
- 25 Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of artificial general intelligence: Early experiments with gpt-4, 2023. doi:10.48550/ARXIV.2303.12712.
- 26 Boxi Cao, Hongyu Lin, Xianpei Han, Le Sun, Lingyong Yan, Meng Liao, Tong Xue, and Jin Xu. Knowledgeable or educated guess? revisiting language models as knowledge bases. In ACL, pages 1860–1874, aug 2021. doi:10.18653/V1/ 2021.ACL-LONG.146.
- 27 J Harry Caufield, Harshad Hegde, Vincent Emonet, Nomi L Harris, Marcin P Joachimiak, Nicolas Matentzoglu, HyeongSik Kim, Sierra AT Moxon, Justin T Reese, Melissa A Haendel, et al. Structured prompt interrogation and recursive extraction of semantics (spires): A method for populating

knowledge bases using zero-shot learning. *arXiv*, 2023. doi:10.48550/ARXIV.2304.02711.

- 28 Sejla Cebiric, François Goasdoué, Haridimos Kondylakis, Dimitris Kotzinos, Ioana Manolescu, Georgia Troullinou, and Mussab Zneika. Summarizing semantic graphs: a survey. VLDB J., 28(3):295– 327, 2019. doi:10.1007/S00778-018-0528-3.
- 29 Jiaoyan Chen, Yuan He, Yuxia Geng, Ernesto Jiménez-Ruiz, Hang Dong, and Ian Horrocks. Contextual semantic embeddings for ontology subsumption prediction. WWW, pages 1–23, 2023. doi:10.1007/S11280-023-01169-9.
- 30 Jiaoyan Chen, Ernesto Jiménez-Ruiz, Ian Horrocks, Denvar Antonyrajah, Ali Hadian, and Jaehun Lee. Augmenting ontology alignment by semantic embedding and distant supervision. In ESWC, pages 392–408, 2021. doi:10.1007/978-3-030-77385-4_23.
- 31 Jiaoyan Chen, Freddy Lecue, Jeff Z. Pan, Ian Horrocks, and Huajun Chen. Knowledge-based Transfer Learning Explanation. In *KR*, pages 349–358, 2018. URL: https://aaai.org/ocs/index.php/KR/KR18/paper/view/18054.
- 32 Jiaqiang Chen, Niket Tandon, Charles Darwis Hariman, and Gerard de Melo. WebBrain: Joint neural learning of large-scale commonsense knowledge. In *ISWC*, pages 102–118, 2016. URL: http: //gerard.demelo.org/webbrain/, doi:10.1007/ 978-3-319-46523-4_7.
- 33 Mingyang Chen, Wen Zhang, Yuxia Geng, Zezhong Xu, Jeff Z. Pan, and Huajun Chen. Generalizing to unseen elements: A survey on knowledge extrapolation for knowledge graphs. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China, pages 6574–6582. ijcai.org, 2023. doi:10.24963/IJCAI.2023/737.
- 34 Wenhu Chen, Yu Su, Xifeng Yan, and William Yang Wang. KGPT: Knowledge-grounded pretraining for data-to-text generation. In *EMNLP*, 2020. doi:10.18653/V1/2020.EMNLP-MAIN.697.
- 35 Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini, editors, WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 29, 2022, pages 2778–2788. ACM, 2022. doi:10.1145/3485447.3511998.
- 36 Zhuo Chen, Jiaoyan Chen, Wen Zhang, Lingbing Guo, Yin Fang, Yufeng Huang, Yuxia Geng, Jeff Z Pan, Wenting Song, and Huajun Chen. Meaformer: Multi-modal entity alignment transformer for meta modality hybrid. In ACM Multimedia, 2023. doi:10.1145/3581783.3611786.
- 37 Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, mar 2023. URL: https://lmsys.org/blog/ 2023-03-30-vicuna/.

- 38 Bonggeun Choi, Daesik Jang, and Youngjoong Ko. Mem-kgc: Masked entity model for knowledge graph completion with pre-trained language model. *IEEE Access*, 9, 2021. doi:10.1109/ACCESS.2021. 3113329.
- 39 Bonggeun Choi and Youngjoong Ko. Knowledge graph extension with a pre-trained language model via unified learning method. *Knowledge-Based Sys*tems, page 110245, 2023. doi:10.1016/J.KNOSYS. 2022.110245.
- 40 Nurendra Choudhary and Chandan K. Reddy. Complex logical reasoning over knowledge graphs using large language models. CoRR, abs/2305.01157, 2023. doi:10.48550/ARXIV.2305. 01157.
- 41 Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. Palm: Scaling language modeling with pathways, 2023. URL: http://jmlr.org/papers/v24/22-1144.html.
- 42 Aida Mostafazadeh Davani, Mark Díaz, and Vinodkumar Prabhakaran. Dealing with disagreements: Looking beyond the majority vote in subjective annotations. *TACL*, 10:92–110, 2022. doi:10.1162/TACL_A_00449.
- 43 Daniel Daza, Michael Cochez, and Paul Groth. Inductive Entity Representations from Text via Link Prediction. In WWW, pages 798–808, 2021. doi:10.1145/3442381.3450141.
- 44 N De Cao, G Izacard, S Riedel, and F Petroni. Autoregressive entity retrieval. In *ICLR 2021-*9th International Conference on Learning Representations, volume 2021. ICLR, 2020. URL: https://openreview.net/forum?id=5k8F6UU39V.
- 45 Nicola De Cao, Wilker Aziz, and Ivan Titov. Editing factual knowledge in language models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6491–6506, Online and Punta Cana, Dominican Republic, nov 2021. Association for Computational Linguistics. doi:10.18653/V1/2021.EMNLP-MAIN. 522.
- 46 Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. Turl: Table understanding through representation learning. ACM SIGMOD Re-

 $cord,\ 51(1){:}33{-}40,\ 2022.$ doi:10.1145/3542700. 3542709.

- 47 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. NAACL, 2019. doi:10.18653/V1/N19-1423.
- 48 Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, and Wei Zhang. From data fusion to knowledge fusion. VLDB, 7(10):881–892, jun 2014. doi:10.14778/2732951.2732962.
- 49 Cícero Nogueira dos Santos, Zhe Dong, Daniel Matthew Cer, John Nham, Siamak Shakeri, Jianmo Ni, and Yun-Hsuan Sung. Knowledge prompts: Injecting world knowledge into language models through soft prompts. ArXiv, 2022. doi: 10.48550/ARXIV.2210.04726.
- 50 Mengnan Du, Fengxiang He, Na Zou, et al. Shortcut learning of large language models in natural language understanding: A survey. arXiv, 2022. doi:10.48550/ARXIV.2208.11857.
- 51 Yupei Du, Qi Zheng, Yuanbin Wu, Man Lan, Yan Yang, and Meirong Ma. Understanding gender bias in knowledge base embeddings. In ACL, 2022. doi:10.18653/V1/2022.ACL-LONG.98.
- 52 N Dziri, H Rashkin, T Linzen, and D Reitter. Evaluating attribution in dialogue systems: The BEGIN benchmark. *TACL*, 2022. URL: https://transacl.org/ojs/index. php/tacl/article/view/3977.
- 53 Yanai Elazar, Nora Kassner, Shauli Ravfogel, Amir Feder, Abhilasha Ravichander, Marius Mosbach, Yonatan Belinkov, Hinrich Schütze, and Yoav Goldberg. Measuring causal effects of data statistics on language model's 'factual' predictions. arXiv, 2023. doi:10.48550/ARXIV.2207.14251.
- 54 Yanai Elazar, Nora Kassner, Shauli Ravfogel, Abhilasha Ravichander, Eduard Hovy, Hinrich Schütze, and Yoav Goldberg. Measuring and improving consistency in pretrained language models. *TACL*, 9, 2021. doi:10.1162/TACL_A_00410.
- 55 Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, and Vassilios S. Verykios. Duplicate record detection: A survey. *TKDE*, 19(1):1–16, 2007. doi:10.1109/TKDE.2007.250581.
- 56 Grace Fan, Jin Wang, Yuliang Li, Dan Zhang, and Renée Miller. Semantics-aware dataset discovery from data lakes with contextualized columnbased representation learning. *VLDB*, 2023. doi: 10.14778/3587136.3587146.
- 57 Wenfei Fan, Chunming Hu, Xueli Liu, and Ping Lu. Discovering graph functional dependencies. ACM Trans. Database Syst., 45(3), sep 2020. doi:10.1145/3397198.
- 58 Wenfei Fan, Ping Lu, Chao Tian, and Jingren Zhou. Deducing certain fixes to graphs. Proc. VLDB Endow., 12(7):752-765, mar 2019. doi: 10.14778/3317315.3317318.
- 59 I. P. Fellegi and A. B. Sunter. A theory for record linkage. *Journal of the American Statistical Association*, 64:1183–1210, 1969. doi:10.1080/ 01621459.1969.10501049.
- 60 Besnik Fetahu, Ujwal Gadiraju, and Stefan Dietze. Improving entity retrieval on structured data. In *ISWC*, pages 474–491, 2015. doi: 10.1007/978-3-319-25007-6_28.

- 61 Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. Amie: Association rule mining under incomplete evidence in ontological knowledge bases. In WWW, WWW '13, pages 413–422, 2013. doi:10.1145/2488388.2488425.
- 62 Daniel Gao, Yantao Jia, Lei Li, Chengzhen Fu, Zhicheng Dou, Hao Jiang, Xinyu Zhang, Lei Chen, and Zhao Cao. Kmir: A benchmark for evaluating knowledge memorization, identification and reasoning abilities of language models, 2022. doi:10.48550/arXiv.2202.13529.
- 63 Luyu Gao, Zhuyun Dai, Panupong Pasupat, Anthony Chen, Arun Tejasvi Chaganty, Yicheng Fan, Vincent Zhao, Ni Lao, Hongrae Lee, Da-Cheng Juan, and Kelvin Guu. RARR: Researching and revising what language models say, using language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 16477–16508, Toronto, Canada, jul 2023. Association for Computational Linguistics. doi:10.18653/V1/2023.ACL-LONG.910.
- 64 Luyu Gao, Zhuyun Dai, Panupong Pasupat, Anthony Chen, Arun Tejasvi Chaganty, Yicheng Fan, Vincent Zhao, Ni Lao, Hongrae Lee, Da-Cheng Juan, and Kelvin Guu. Rarr: Researching and revising what language models say, using language models. In ACL2023, 2023. doi: 10.18653/V1/2023.ACL-LONG.910.
- 65 Genet Asefa Gesese, Russa Biswas, Mehwish Alam, and Harald Sack. A survey on knowledge graph embeddings with literals: Which model links better literal-ly? *Semantic Web*, 12(4):617–647, 2021. doi:10.3233/SW-200404.
- 66 Agneta Ghose, Matteo Lissandrini, Emil Riis Hansen, and Bo Pedersen Weidema. A core ontology for modeling life cycle sustainability assessment on the semantic web. *Journal of Industrial Ecology*, 26(3):731–747, 2022. doi:10.1111/jiec. 13220.
- 67 Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter F. Patel-Schneider, and Ulrike Sattler. OWL 2: The next step for OWL. J. Web Semant, 6(4):309–322, 2008. doi:10.1016/J. WEBSEM.2008.05.001.
- 68 Paul Groth, Elena Paslaru Bontas Simperl, Marieke van Erp, and Denny Vrandecic. Knowledge graphs and their role in the knowledge engineering of the 21st century (dagstuhl seminar 22372). Dagstuhl Reports, 12:60–120, 2022. doi:10.4230/DAGREP.12.9.60.
- 69 OWL Working Group. OWL 2 Web Ontology Language Document Overview: W3C Recommendation, 2012. URL: https://www.w3.org/TR/ owl2-overview/.
- 70 Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. Retrieval augmented language model pre-training. In Hal Daumé III and Aarti Singh, editors, Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, pages 3929–3938. PMLR, 13–18 July 2020. URL: http://proceedings.mlr.press/ v119/guu20a.html.

- 71 Xiaoqi Han, Ru Li, Xiaoli Li, and Jeff Z. Pan. A Divide and Conquer Framework for Knowledge Editing. *Knowledge Based Systems*, 279:110826, 2023. doi:10.1016/j.knosys.2023.110826.
- 72 Yuan He, Jiaoyan Chen, Denvar Antonyrajah, and Ian Horrocks. Bertmap: a bert-based ontology alignment system. In AAAI, volume 36, pages 5684– 5691, 2022. doi:10.1609/AAAI.V36I5.20510.
- 73 Yuan He, Jiaoyan Chen, Hang Dong, Ian Horrocks, Carlo Allocca, Taehun Kim, and Brahmananda Sapkota. DeepOnto: A Python package for ontology engineering with deep learning. arXiv preprint arXiv:2307.03067, 2023. doi:10.48550/ ARXIV.2307.03067.
- 74 Yuan He, Jiaoyan Chen, Hang Dong, Ernesto Jiménez-Ruiz, Ali Hadian, and Ian Horrocks. Machine learning-friendly biomedical datasets for equivalence and subsumption ontology matching. In *ISWC*, pages 575–591, 2022. doi:10.1007/ 978-3-031-19433-7_33.
- 75 Yuan He, Jiaoyan Chen, Ernesto Jimenez-Ruiz, Hang Dong, and Ian Horrocks. Language model analysis for ontology subsumption inference. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3439–3453, Toronto, Canada, jul 2023. Association for Computational Linguistics. doi:10.18653/V1/2023. FINDINGS-ACL.213.
- 76 Benjamin Heinzerling and Kentaro Inui. Language models as knowledge bases: On entity representations, storage capacity, and paraphrased queries. In *EACL*, pages 1772–1791, 2021. doi: 10.18653/V1/2021.EACL-MAIN.153.
- 77 Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. CUAD: An expert-annotated NLP dataset for legal contract review. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021. URL: https://openreview.net/forum?id= 711Ygs3Bamw.
- 78 Or Honovich, Uri Shaham, Samuel R. Bowman, and Omer Levy. Instruction induction: From few examples to natural language task descriptions. In Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, pages 1935– 1952. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.108.
- 79 Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR* 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. URL: https://openreview.net/ forum?id=nZeVKeeFYf9.
- 80 Nan Hu, Yike Wu, Guilin Qi, Dehai Min, Jiaoyan Chen, Jeff Z. Pan, and Zafar Ali. An empirical study of pre-trained language models in simple knowledge graph question answering. World Wide Web (WWW), 26(5):2855–2886, 2023. doi:10. 1007/S11280-023-01166-Y.
- 81 Ziniu Hu, Yichong Xu, Wenhao Yu, Shuohang Wang, Ziyi Yang, Chenguang Zhu, Kai-Wei Chang,

and Yizhou Sun. Empowering language models with knowledge graph reasoning for open-domain question answering. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings* of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 9562–9581. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. EMNLP-MAIN.650.

- 82 Jian Huang, Jianfeng Gao, Jiangbo Miao, Xiaolong Li, Kuansan Wang, Fritz Behr, and C. Lee Giles. Exploring web scale language models for search query processing. In WWW, pages 451–460, 2010. doi:10.1145/1772690.1772737.
- 83 Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. *Findings of ACL*, 2023. doi:10.48550/ARXIV. 2212.10403.
- 84 Ningyuan Huang, Yash R Deshpande, Yibo Liu, Houda Alberts, Kyunghyun Cho, Clara Vania, and Iacer Calixto. Endowing language models with multimodal knowledge graph representations. arXiv, 2022. doi:10.48550/ARXIV.2206.13163.
- 85 Wenyu Huang, Mirella Lapata, Pavlos Vougiouklis, Nikos Papasarantopoulos, and Jeff Z. Pan. Retrieval Augmented Generation with Rich Answer Encoding. In *Proc. of IJCNLP-AACL 2023*, 2023.
- 86 Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. Comet-atomic 2020: On symbolic and neural commonsense knowledge graphs. In AAAI, 2021. doi:10.1609/AAAI.V3517.16792.
- 87 Gautier Izacard and Edouard Grave. Leveraging passage retrieval with generative models for open domain question answering. In Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty, editors, Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 – 23, 2021, pages 874–880. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.EACL-MAIN.74.
- 88 Sarthak Jain and Byron C. Wallace. Attention is not explanation. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), pages 3543– 3556. Association for Computational Linguistics, 2019. doi:10.18653/V1/N19-1357.
- 89 Krzysztof Janowicz, Bo Yan, Blake Regalia, Rui Zhu, and Gengchen Mai. Debiasing knowledge graphs: Why female presidents are not like female popes. In Marieke van Erp, Medha Atre, Vanessa López, Kavitha Srinivas, and Carolina Fortuna, editors, Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - 12th, 2018, volume 2180 of CEUR Workshop Proceedings. CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2180/ ISWC_2018_Outrageous_Ideas_paper_17.pdf.

- 90 Zhengbao Jiang, Frank F Xu, Jun Araki, et al. How can we know what language models know? TACL, 8:423–438, 2020. doi:10.1162/TACL_A_00324.
- 91 Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. Logmap: Logic-based and scalable ontology matching. In *ISWC*, pages 273–288, 2011. doi:10.1007/ 978-3-642-25073-6_18.
- 92 Ernesto Jiménez-Ruiz, Oktie Hassanzadeh, Vasilis Efthymiou, Jiaoyan Chen, and Kavitha Srinivas. Semtab 2019: Resources to benchmark tabular data to knowledge graph matching systems. In *ESWC*, pages 514–530, 2020. doi:10.1007/ 978-3-030-49461-2_30.
- 93 Martin Josifoski, Nicola De Cao, Maxime Peyrard, Fabio Petroni, and Robert West. Genie: Generative information extraction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4626–4643, 2022. doi:10.18653/V1/2022.NAACL-MAIN.342.
- 94 Martin Josifoski, Marija Sakota, Maxime Peyrard, and Robert West. Exploiting asymmetry for synthetic training data generation: Synthie and the case of information extraction. *ArXiv*, 2023. doi:10.48550/ARXIV.2303.04132.
- 95 Jan-Christoph Kalo and Leandra Fichtel. Kamel: Knowledge analysis with multitoken entities in language models. In AKBC, 2022. URL: https://www.akbc.ws/2022/assets/pdfs/ 15_kamel_knowledge_analysis_with_.pdf.
- 96 Jan-Christoph Kalo, Simon Razniewski, Sneha Singhania, and Jeff Z. Pan. LM-KBC: Knowledge base construction from pre-trained language models. *ISWC Challenges*, 2023.
- 97 Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. Large language models struggle to learn long-tail knowledge, 2023. doi:10.48550/ARXIV.2211.08411.
- 98 Ehud Karpas, Omri Abend, Yonatan Belinkov, Barak Lenz, Opher Lieber, Nir Ratner, Yoav Shoham, Hofit Bata, Yoav Levine, Kevin Leyton-Brown, Dor Muhlgay, Noam Rozen, Erez Schwartz, Gal Shachaf, Shai Shalev-Shwartz, Amnon Shashua, and Moshe Tennenholtz. MRKL systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning. CoRR, abs/2205.00445, 2022. doi:10.48550/ARXIV.2205. 00445.
- 99 Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen tau Yih. Dense passage retrieval for open-domain question answering. In *EMNLP*, pages 6769–6781, 2020. doi:10.18653/V1/2020. EMNLP-MAIN.550.
- 100 Nora Kassner and Hinrich Schütze. Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly. In ACL, 2020. doi:10.18653/V1/2020.ACL-MAIN.698.
- 101 Bosung Kim, Taesuk Hong, Youngjoong Ko, and Jungyun Seo. Multi-task learning for knowledge graph completion with pre-trained language models. In *COLING*, pages 1737–1743, 2020. doi: 10.18653/V1/2020.COLING-MAIN.153.
- 102 Holger Knublauch and Dimitris Kontokostas. Shapes constraint language (SHACL). Technical

report, W3C, jul 2017. URL: https://www.w3. org/TR/shacl/.

- 103 Keti Korini and Christian Bizer. Column type annotation using chatgpt. arXiv, 2023. doi: 10.48550/ARXIV.2306.00745.
- 104 Angelie Kraft and Ricardo Usbeck. The lifecycle of "facts": A survey of social bias in knowledge graphs. In AACL, 2022. URL: https: //aclanthology.org/2022.aacl-main.49.
- 105 Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *Int. J. Comput. Vis.*, 123(1):32–73, 2017. doi: 10.1007/S11263-016-0981-7.
- 106 Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning. ArXiv, 2023. doi:10.48550/ARXIV.2304.05613.
- 107 Yoonjoo Lee, John Joon Young Chung, Tae Soo Kim, Jean Y. Song, and Juho Kim. Promptiverse: Scalable generation of scaffolding prompts through human-ai hybrid knowledge graph annotation. In *CHI*, 2022. doi:10.1145/3491102.3502087.
- 108 Alina Leidinger and Richard Rogers. Which stereotypes are moderated and under-moderated in search engine autocompletion? In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2023, Chicago, IL, USA, June 12-15, 2023, pages 1049–1061. ACM, 2023. doi:10.1145/3593013.3594062.
- 109 Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In neurIPS, volume 33, 2020. URL: https: //proceedings.neurips.cc/paper/2020/hash/ 6b493230205f780e1bc26945df7481e5-Abstract. html.
- 110 Belinda Z. Li, Sewon Min, Srinivasan Iyer, Yashar Mehdad, and Wen-tau Yih. Efficient one-pass end-to-end entity linking for questions. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pages 6433–6441, 2020. doi:10.18653/V1/2020.EMNLP-MAIN.522.
- 111 Bo Li, Gexiang Fang, Yang Yang, Quansen Wang, Wei Ye, Wen Zhao, and Shikun Zhang. Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness. ArXiv, 2023. doi:10.48550/ARXIV.2304.11633.
- 112 Da Li, Ming Yi, and Yukai He. Lp-bert: Multitask pre-training knowledge graph bert for link prediction. arXiv, 2022. arXiv:2201.04843.
- 113 Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. BLIP-2: Bootstrapping language-image pretraining with frozen image encoders and large language models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt,

Sivan Sabato, and Jonathan Scarlett, editors, Proceedings of the 40th International Conference on Machine Learning, volume 202 of Proceedings of Machine Learning Research, pages 19730– 19742. PMLR, 23–29 July 2023. URL: https: //proceedings.mlr.press/v202/li23q.html.

- 114 Tianyi Li, Mohammad Javad Hosseini, Sabine Weber, and Mark Steedman. Language models are poor learners of directional inference. In Findings of the Association for Computational Linguistics: EMNLP 2022, pages 903–921, Abu Dhabi, United Arab Emirates, dec 2022. Association for Computational Linguistics. doi:10.18653/V1/ 2022.FINDINGS-EMNLP.64.
- 115 Tianyi Li, Wenyu Huang, Nikos Papasarantopoulos, Pavlos Vougiouklis, and Jeff Z. Pan. Taskspecific pre-training and prompt decomposition for knowledge graph population with language models. In *LM-KBC*, 2022. arXiv:2208.12539, doi:10.48550/ARXIV.2208.12539.
- 116 Stephan Linzbach, Tim Tressel, Laura Kallmeyer, Stefan Dietze, and Hajira Jabeen. Decoding prompt syntax: Analysing its impact on knowledge retrieval in large language models. In NLP4KGC, 2023. doi:10.1145/3543873.3587655.
- 117 Matteo Lissandrini, Davide Mottin, Themis Palpanas, Dimitra Papadimitriou, and Yannis Velegrakis. Unleashing the power of information graphs. *SIGMOD Rec.*, 43(4):21–26, 2014. doi:10.1145/2737817.2737822.
- 118 Fangyu Liu, Ivan Vulic, Anna Korhonen, and Nigel Collier. Learning domain-specialised representations for cross-lingual biomedical entity linking. In ACL, 2021. doi:10.18653/V1/2021.ACL-SHORT. 72.
- 119 Hao Liu, Yehoshua Perl, and James Geller. Concept placement using bert trained by transforming and summarizing biomedical ontology structure. Journal of Biomedical Informatics, 112:103607, 2020. doi:10.1016/J.JBI.2020. 103607.
- 120 Hugo Liu and Push Singh. Commonsense reasoning in and over natural language. In Knowledge-Based Intelligent Information and Engineering Systems, pages 293–306, 2004. doi:10.1007/ 978-3-540-30134-9_40.
- 121 Jixiong Liu, Yoan Chabot, Raphaël Troncy, Viet-Phi Huynh, Thomas Labbé, and Pierre Monnin. From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. J. Web Semant., 2022. doi:10.1016/J.WEBSEM. 2022.100761.
- 122 Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv, 2019. arXiv:1907.11692.
- 123 Fengyuan Lu, Peijin Cong, and Xinli Huang. Utilizing textual information in knowledge graph embedding: A survey of methods and applications. *IEEE Access*, 8:92072–92088, 2020. doi: 10.1109/ACCESS.2020.2995074.
- 124 Chaitanya Malaviya, Chandra Bhagavatula, Antoine Bosselut, and Yejin Choi. Commonsense knowledge base completion with structural and semantic

context. In AAAI, volume 34, pages 2925–2933, 2020. doi:10.1609/AAAI.V34I03.5684.

- 125 Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 9802–9822, Toronto, Canada, jul 2023. Association for Computational Linguistics. doi:10.18653/V1/2023. ACL-LONG.546.
- 126 Rui Mao, Qian Liu, Kai He, et al. The biases of pretrained language models: An empirical study on prompt-based sentiment analysis and emotion detection. *IEEE Transactions on Affective Computing*, pages 1–11, 2022. doi:10.1109/TAFFC.2022. 3204972.
- 127 Ninareh Mehrabi, Thamme Gowda, Fred Morstatter, Nanyun Peng, and Aram Galstyan. Man is to person as woman is to location: Measuring gender bias in named entity recognition. In Ujwal Gadiraju, editor, HT '20: 31st ACM Conference on Hypertext and Social Media, Virtual Event, USA, July 13-15, 2020, pages 231–232. ACM, 2020. doi:10.1145/3372923.3404804.
- 128 Chris Mellish and Jeff Z. Pan. Natural Language Directed Inference from Ontologies. In Artificial Intelligence Journal, 2008. doi:10.1016/J. ARTINT.2008.01.003.
- 129 Eric Mitchell, Charles Lin, Antoine Bosselut, Christopher D Manning, and Chelsea Finn. Memory-based model editing at scale. In *International Conference on Machine Learning*, pages 15817–15831. PMLR, 2022. URL: https://proceedings.mlr.press/v162/mitchell22a.html.
- 130 Fedor Moiseev, Zhe Dong, Enrique Alfonseca, and Martin Jaggi. Skill: Structured knowledge infusion for large language models. In NAACL, 2022. doi:10.18653/V1/2022.NAACL-MAIN.113.
- 131 Moin Nadeem, Anna Bethke, and Siva Reddy. Stereoset: Measuring stereotypical bias in pretrained language models. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, pages 5356–5371. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.ACL-LONG. 416.
- 132 Mojtaba Nayyeri, Zihao Wang, Mst Akter, Mirza Mohtashim Alam, Md Rashad Al Hasan Rony, Jens Lehmann, Steffen Staab, et al. Integrating knowledge graph embedding and pretrained language models in hypercomplex spaces. *arXiv*, 2022. doi:10.48550/ARXIV.2208.02743.
- 133 Sophie Neutel and Maaike HT de Boer. Towards automatic ontology alignment using bert. In AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, 2021. URL: https://ceur-ws.org/Vol-2846/paper28.pdf.
- 134 Tuan-Phong Nguyen, Simon Razniewski, Aparna Varde, and Gerhard Weikum. Extracting cultural

commonsense knowledge at scale. In WWW, pages 1907–1917, 2023. doi:10.1145/3543507.3583535.

- 135 Natalya Fridman Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM*, 62(8):36–43, 2019. doi:10.1145/3331166.
- 136 Barlas Oguz, Xilun Chen, Vladimir Karpukhin, Stan Peshterliev, Dmytro Okhonko, Michael Sejr Schlichtkrull, Sonal Gupta, Yashar Mehdad, and Scott Yih. Unik-qa: Unified representations of structured and unstructured knowledge for opendomain question answering. In Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz, editors, *Findings of the Association for Computational Linguistics: NAACL* 2022, Seattle, WA, United States, July 10-15, 2022, pages 1535–1546. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. FINDINGS-NAACL.115.
- 137 OpenAI. GPT-4 technical report, 2023. doi: 10.48550/ARXIV.2303.08774.
- 138 Long Ouyang, Jeff Wu, Xu Jiang, et al. Training language models to follow instructions with human feedback. neurIPS, 2022. URL: http://papers. nips.cc/paper_files/paper/2022/hash/ b1efde53be364a73914f58805a001731-Abstract-Conference.html.
- 139 Jeff Z. Pan. Resource Description Framework. In Handbook on Ontologies. IOS Press, 2009. doi: 10.1007/978-3-540-92673-3_3.
- 140 Jeff Z. Pan and Ian Horrocks. Web Ontology Reasoning with Datatype Groups. In ISWC, pages 47–63, 2003. doi:10.1007/978-3-540-39718-2_4.
- 141 Jeff Z. Pan, Guido Vetere, José Manuél Gómez-Pérez, and Honghan Wu. Exploiting linked data and knowledge graphs in large organisations. Springer International Publishing, 2017. doi:10.1007/978-3-319-45654-6.
- 142 Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. arXiv, 2023. doi:10.48550/ARXIV.2306.08302.
- 143 Lalchand Pandia and Allyson Ettinger. Sorting through the noise: Testing robustness of information processing in pre-trained language models. *EMNLP*, 2021. doi:10.48550/arXiv.2109.12393.
- 144 George Papadakis, Ekaterini Ioannou, Emanouil Thanos, and Themis Palpanas. The four generations of entity resolution. In Synthesis Lectures on Data Management, 2021. doi:10.2200/ S01067ED1V01Y202012DTM064.
- 145 Jae Sung Park, Chandra Bhagavatula, Roozbeh Mottaghi, Ali Farhadi, and Yejin Choi. Visualcomet: Reasoning about the dynamic context of a still image. In Computer Vision – ECCV 2020 – 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part V, volume 12350 of Lecture Notes in Computer Science, pages 508–524. Springer, 2020. doi:10.1007/ 978-3-030-58558-7_30.
- 146 Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web, 8(3):489–508, 2017. doi: 10.3233/SW-160218.

- 147 Ralph Peeters, Reng Chiz Der, and Christian Bizer. WDC products: A multi-dimensional entity matching benchmark. In *Proceedings 27th International Conference on Extending Database Technology*, *EDBT 2024, Paestum, Italy, March 25 – March* 28, pages 22–33, 2024. doi:10.48786/EDBT.2024. 03.
- 148 Shichao Pei, Lu Yu, Guoxian Yu, and Xiangliang Zhang. Rea: Robust cross-lingual entity alignment between knowledge graphs. In KDD, pages 2175–2184, 2020. doi:10.1145/3394486.3403268.
- 149 Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. The refinedweb dataset for falcon llm: Outperforming curated corpora with web data, and web data only, 2023. arXiv: 2306.01116, doi:10.48550/ARXIV.2306.01116.
- 150 Matthew E Peters, Mark Neumann, Robert L Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A Smith. Knowledge enhanced contextual word representations. *EMNLP*, 2019. arXiv:1909.04164.
- 151 Fabio Petroni, Patrick Lewis, Aleksandra Piktus, et al. How context affects language models' factual predictions. AKBC, 2020. doi:10.24432/C5201W.
- 152 Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 2463–2473, Hong Kong, China, nov 2019. Association for Computational Linguistics. doi:10.18653/V1/D19-1250.
- 153 Gabriele Picco, Marcos Martinez Galindo, Alberto Purpura, Leopold Fuchs, Vanessa Lopez, and Thanh Lam Hoang. Zshot: An open-source framework for zero-shot named entity recognition and relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 357–368, 2023. doi:10.18653/V1/2023.ACL-DEM0.34.
- 154 Barbara Plank. The "problem" of human label variation: On ground truth in data, modeling and evaluation. *EMNLP*, abs/2211.02570, 2022. doi:10.18653/V1/2022.EMNLP-MAIN.731.
- 155 Eric Prud'hommeaux, José Emilio Labra Gayo, and Harold R. Solbrig. Shape expressions: an RDF validation and transformation language. In Harald Sack, Agata Filipowska, Jens Lehmann, and Sebastian Hellmann, editors, Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2014, Leipzig, Germany, September 4-5, 2014, pages 32–40. ACM, 2014. doi:10.1145/2660517.2660523.
- 156 Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, and Maosong Sun. Toolllm: Facilitating large language models to master 16000+ real-world apis, 2023. doi:10.48550/ARXIV.2307.16789.

- 157 Kashif Rabbani, Matteo Lissandrini, and Katja Hose. SHACL and shex in the wild: A community survey on validating shapes generation and adoption. In Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini, editors, Companion of The Web Conference 2022, Virtual Event / Lyon, France, April 25 – 29, 2022, pages 260–263. ACM, 2022. doi:10.1145/3487553.3524253.
- 158 Kashif Rabbani, Matteo Lissandrini, and Katja Hose. Extraction of validating shapes from very large knowledge graphs. *Proc. VLDB En*dow., 16(5):1023–1032, jan 2023. doi:10.14778/ 3579075.3579078.
- 159 Kashif Rabbani, Matteo Lissandrini, and Katja Hose. SHACTOR: improving the quality of largescale knowledge graphs with validating shapes. In Sudipto Das, Ippokratis Pandis, K. Selçuk Candan, and Sihem Amer-Yahia, editors, Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023, pages 151–154. ACM, 2023. doi:10.1145/3555041.3589723.
- 160 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang, editors, Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 8748–8763. PMLR, 18–24 July 2021. URL: https://proceedings.mlr.press/ v139/radford21a.html.
- 161 Leonardo Ranaldi, Elena Sofia Ruzzetti, David A. Venditti, Dario Onorati, and Fabio Massimo Zanzotto. A trip towards fairness: Bias and debiasing in large language models. ArXiv, 2023. doi:10.48550/ARXIV.2305.13862.
- 162 Abhilasha Ravichander, Eduard Hovy, Kaheer Suleman, Adam Trischler, and Jackie Chi Kit Cheung. On the systematicity of probing contextualized word representations: The case of hypernymy in BERT. In Joint Conference on Lexical and Computational Semantics, pages 88– 102, 2020. URL: https://aclanthology.org/ 2020.starsem-1.10/.
- 163 Simon Razniewski, Andrew Yates, Nora Kassner, and Gerhard Weikum. Language models as or for knowledge bases. CoRR, abs/2110.04888, 2021. arXiv:2110.04888.
- 164 Leonardo F. R. Ribeiro, Martin Schmitt, Hinrich Schütze, and Iryna Gurevych. Investigating pretrained language models for graph-to-text generation. Workshop on Natural Language Processing for Conversational AI, abs/2007.08426, 2021. arXiv:2007.08426.
- 165 Petar Ristoski, Jessica Rosati, Tommaso Di Noia, Renato De Leone, and Heiko Paulheim. Rdf2vec: RDF graph embeddings and their applications. Semantic Web, 10(4):721–752, 2019. doi:10.3233/ SW-180317.
- 166 Devendra Sachan, Yuhao Zhang, Peng Qi, et al. Do syntax trees help pre-trained transformers ex-

tract information? In *EACL*, pages 2647–2661, apr 2021. doi:10.18653/V1/2021.EACL-MAIN.228.

- 167 Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. The risk of racial bias in hate speech detection. In ACL, 2019. doi:10.18653/V1/P19-1163.
- 168 Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. In arXiv, 2023. doi:10.48550/ARXIV.2302.04761.
- 169 Patrick Schramowski, Cigdem Turan, Nico Andersen, et al. Large pre-trained language models contain human-like biases of what is right and wrong to do. Nature Machine Intelligence, 4(3):258–268, 2022. doi:10.1038/S42256-022-00458-8.
- 170 Jingyu Shao, Qing Wang, Asiri Wijesinghe, and Erhard Rahm. Ergan: Generative adversarial networks for entity resolution. In *ICDM*, pages 1250– 1255, 2020. doi:10.1109/ICDM50108.2020.00158.
- 171 Jingchuan Shi, Jiaoyan Chen, Hang Dong, Ishita Khan, Lizzie Liang, Qunzhi Zhou, Zhe Wu, and Ian Horrocks. Subsumption prediction for e-commerce taxonomies. In *ESWC*, pages 244–261, 2023. doi:10.1007/978-3-031-33455-9_15.
- 172 Peng Shi and Jimmy J. Lin. Simple BERT models for relation extraction and semantic role labeling. *ArXiv*, abs/1904.05255, 2019. URL: http://arxiv. org/abs/1904.05255, arXiv:1904.05255.
- 173 Jaeho Shin, Sen Wu, Feiran Wang, Christopher De Sa, Ce Zhang, and Christopher Ré. Incremental knowledge base construction using deepdive. In VLDB, volume 8, page 1310, 2015. doi:10.14778/2809974.2809991.
- 174 Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. Auto-Prompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *EMNLP*, pages 4222–4235, 2020. doi:10.18653/ V1/2020.EMNLP-MAIN.346.
- 175 Sneha Singhania, Tuan-Phong Nguyen, and Simon Razniewski. LM-KBC: Knowledge base construction from pre-trained language models. *Semantic Web challenge*, 2022. URL: https://ceur-ws.org/ Vol-3274/paper1.pdf.
- 176 Sneha Singhania, Simon Razniewski, and Gerhard Weikum. Predicting Document Coverage for Relation Extraction. Transactions of the Association for Computational Linguistics, 10:207–223, mar 2022. doi:10.1162/TACL_A_00456.
- 177 Xiaoqi Han snf Ru Li, Hongye Tan, Yuanlong Wang, Qinghua Chai, and Jeff Z. Pan. Improving Sequential Model Editing with Fact Retrieval. In *Findings of EMNLP*, 2023.
- 178 Ran Song, Shizhu He, Shengxiang Gao, Li Cai, Kang Liu, Zhengtao Yu, and Jun Zhao. Multilingual knowledge graph completion from pretrained language models with knowledge constraints. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7709–7721, 2023. doi:10.18653/V1/2023.FINDINGS-ACL.488.
- 179 Aarohi Srivastava et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on*

Machine Learning Research, 2023. URL: https: //openreview.net/forum?id=uyTL5Bvosj.

- 180 Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web, WWW '07, pages 697–706, New York, NY, USA, 2007. Association for Computing Machinery. doi:10.1145/1242572. 1242667.
- 181 Yoshihiko Suhara, Jinfeng Li, Yuliang Li, Dan Zhang, Çağatay Demiralp, Chen Chen, and Wang-Chiew Tan. Annotating columns with pre-trained language models. In SIGMOD, pages 1493–1503, 2022. doi:10.1145/3514221.3517906.
- 182 Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. A benchmarking study of embeddingbased entity alignment for knowledge graphs. *VLDB*, 13(11):2326-2340, 2020. URL: http:// www.vldb.org/pvldb/vol13/p2326-sun.pdf.
- 183 Alexandre Tamborrino, Nicola Pellicano, Baptiste Pannier, et al. Pre-training is (almost) all you need: An application to commonsense reasoning. ACL, 2020. doi:10.48550/arXiv.2004.14074.
- 184 Wang-Chiew Tan, Yuliang Li, Pedro Rodriguez, Richard James, Xi Victoria Lin, Alon Halevy, and Scott Yih. Reimagining retrieval augmented language models for answering queries. In *Findings of ACL*, pages 6131–6146, 2023. doi: 10.18653/V1/2023.FINDINGS-ACL.382.
- 185 Niket Tandon and Gerard de Melo. Information extraction from web-scale n-gram data. In Web N-gram Workshop, volume 5803, pages 8-15, 2010. URL: http://citeseerx.ist.psu.edu/ viewdoc/summary?doi=10.1.1.365.2318.
- 186 Niket Tandon, Gerard de Melo, Fabian M. Suchanek, and Gerhard Weikum. Webchild: harvesting and organizing commonsense knowledge from the web. Proceedings of the 7th ACM international conference on Web search and data mining, 2014. doi:10.1145/2556195.2556245.
- 187 Niket Tandon, Gerard de Melo, and Gerhard Weikum. Deriving a Web-scale common sense fact database. In AAAI, pages 152– 157, 2011. URL: http://www.aaai.org/ocs/ index.php/AAAI/AAAI11/paper/view/3764, doi: 10.1609/AAAI.V25I1.7841.
- 188 Nan Tang, Ju Fan, Fangyi Li, Jianhong Tu, Xiaoyong Du, Guoliang Li, Sam Madden, and Mourad Ouzzani. Rpt: relational pre-trained transformer is almost all you need towards democratizing data preparation. *VLDB*, 2021. doi:10.14778/3457390. 3457391.
- 189 Ruixiang Tang, Xiaotian Han, Xiaoqian Jiang, and Xia Hu. Does synthetic data generation of LLMs help clinical text mining? ArXiv, 2023. doi:10.48550/ARXIV.2303.04360.
- 190 Nicolas Tempelmeier, Elena Demidova, and Stefan Dietze. Inferring missing categorical information in noisy and sparse web markup. In WWW, 2018. doi:10.1145/3178876.3186028.
- 191 Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and

efficient foundation language models. *arXiv*, 2023. doi:10.48550/ARXIV.2302.13971.

- 192 Pat Verga, Haitian Sun, Livio Baldini Soares, et al. Facts as experts: Adaptable and interpretable neural memory over symbolic knowledge. NAACL, 2021. arXiv:2007.00849.
- 193 Blerta Veseli, Sneha Singhania, Simon Razniewski, and Gerhard Weikum. Evaluating language models for knowledge base completion. In *The Semantic* Web – 20th International Conference, ESWC 2023, Hersonissos, Crete, Greece, May 28 – June 1, 2023, Proceedings, volume 13870 of Lecture Notes in Computer Science, pages 227–243. Springer, 2023. doi:10.1007/978-3-031-33455-9_14.
- 194 Liane Vogel, Benjamin Hilprecht, and Carsten Binnig. Towards foundation models for relational databases [vision paper]. TRL@NeurIPS2022, 2023. doi:10.48550/ARXIV.2305.15321.
- 195 Pavlos Vougiouklis, Nikos Papasarantopoulos, Danna Zheng, David Tuckey, Chenxin Diao, Zhili Shen, and Jeff Z. Pan. FastRAT: Fast and Efficient Cross-lingual Text-to-SQL Semantic Parsing. In Proc. of IJCNLP-AACL 2023, 2023.
- 196 Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. Commun. ACM, 57:78–85, sep 2014. doi:10.1145/2629489.
- 197 David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. Entity, relation, and event extraction with contextualized span representations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 5784–5789, Hong Kong, China, nov 2019. Association for Computational Linguistics. doi:10.18653/V1/D19-1585.
- 198 Bin Wang, Guangtao Wang, Jing Huang, Jiaxuan You, Jure Leskovec, and C.-C. Jay Kuo. Inductive learning on commonsense knowledge graph completion. In *Joint Conference on Neural Networks*, pages 1–8, 2021. doi:10.1109/IJCNN52387.2021. 9534355.
- 199 Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, Ying Wang, and Yi Chang. Structure-augmented text representation learning for efficient knowledge graph completion. In WWW, pages 1737–1748, 2021. doi:10.1145/3442381.3450043.
- 200 Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. SimKGC: Simple Contrastive Knowledge Graph Completion with Pre-trained Language Models. In ACL, pages 4281–4294, may 2022. doi:10.18653/V1/2022.ACL-LONG.295.
- 201 Xiao Wang, Wei Zhou, Can Zu, Han Xia, Tianze Chen, Yuan Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, Jihua Kang, J. Yang, Siyuan Li, and Chunsai Du. Instructuie: Multi-task instruction tuning for unified information extraction. *ArXiv*, 2023. doi:10.48550/ARXIV.2304.08085.
- 202 Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. Kepler: A unified model for knowledge embedding and pre-trained language representation. *TACL*, 9:176–194, 2021. doi:10.1162/TACL_ A_00360.

- 203 Xiyu Wang and Nora El-Gohary. Deep learningbased relation extraction and knowledge graphbased representation of construction safety requirements. Automation in Construction, 147:104696, 2023. doi:10.1016/j.autcon.2022.104696.
- 204 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Huai hsin Chi, F. Xia, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models. ArXiv, abs/2201.11903, 2022. arXiv:2201.11903.
- 205 Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, Yong Jiang, and Wenjuan Han. Zero-shot information extraction via chatting with chatgpt. ArXiv, 2023. doi:10.48550/ARXIV.2302.10205.
- 206 Gerhard Weikum, Luna Dong, Simon Razniewski, and Fabian M. Suchanek. Machine knowledge: Creation and curation of comprehensive knowledge bases. CoRR, abs/2009.11564, 2020. arXiv: 2009.11564.
- 207 Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 1966. doi:10.1145/365153.365168.
- 208 Peter West, Chandra Bhagavatula, Jack Hessel, Jena D Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, and Yejin Choi. Symbolic knowledge distillation: from general language models to commonsense models. arXiv preprint arXiv:2110.07178, 2021. doi:10.48550/ arXiv.2110.07178.
- 209 Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David S. Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance. *CoRR*, abs/2303.17564, 2023. doi: 10.48550/ARXIV.2303.17564.
- 210 Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In *IJCAI*, pages 5511–5519, 2018. doi:10.24963/IJCAI.2018/777.
- 211 Wenjie Xu, Ben Liu, Miao Peng, Xu Jia, and Min Peng. Pre-trained language model with prompts for temporal knowledge graph completion. In *Findings of ACL 2023*, 2023. doi:10.18653/V1/2023. FINDINGS-ACL.493.
- 212 Yan Xu, Mahdi Namazifar, Devamanyu Hazarika, Aishwarya Padmakumar, Yang Liu, and Dilek Hakkani-Tür. KILM: knowledge injection into encoder-decoder language models. In Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, pages 5013– 5035. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.275.
- 213 Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Chatgpt is not enough: Enhancing large language models with knowledge graphs for fact-aware language modeling. arXiv, 2023. doi:10.48550/ARXIV.2306.11489.

2:38 LLMs and KGs: Opportunities and Challenges

- 214 Nan Yang, Tao Ge, Liang Wang, Binxing Jiao, Daxin Jiang, Linjun Yang, Rangan Majumder, and Furu Wei. Inference with reference: Lossless acceleration of large language models. In arXiv, 2023. doi:10.48550/ARXIV.2304.04487.
- 215 Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E. Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B. Costa, Mona G. Flores, Ying Zhang, Tanja Magoc, Christopher A. Harle, Gloria Lipori, Duane A. Mitchell, William R. Hogan, Elizabeth A. Shenkman, Jiang Bian, and Yonghui Wu. A large language model for electronic health records. npj Digital Medicine, 5(1):194, 2022. doi:10.1038/ S41746-022-00742-2.
- 216 Liang Yao, Chengsheng Mao, and Yuan Luo. Kgbert: Bert for knowledge graph completion. arXiv, 2019. doi:10.48550/arXiv.1909.03193.
- 217 Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. NAACL, 2021. doi:10.18653/V1/2021.NAACL-MAIN.45.
- 218 Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. Semantic parsing via staged query graph generation: Question answering with knowledge base. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers, pages 1321–1331, 2015. doi:10.3115/V1/P15-1128.
- 219 Donghan Yu, Chenguang Zhu, Yuwei Fang, Wenhao Yu, Shuohang Wang, Yichong Xu, Xiang Ren, Yiming Yang, and Michael Zeng. Kg-fid: Infusing knowledge graph in fusion-in-decoder for open-domain question answering. In ACL, 2022. doi:10.18653/V1/2022.ACL-LONG.340.
- 220 Ran Yu, Ujwal Gadiraju, Besnik Fetahu, Oliver Lehmberg, Dominique Ritze, and Stefan Dietze. Knowmore – knowledge base augmentation with structured web markup. Semantic Web, 10(1):159– 180, 2019. doi:10.3233/SW-180304.
- 221 Qingkai Zeng, Jinfeng Lin, Wenhao Yu, Jane Cleland-Huang, and Meng Jiang. Enhancing taxonomy completion with concept generation via fusing relational representations. In KDD, pages 2104–2113, 2021. doi:10.1145/3447548.3467308.
- 222 Hanwen Zha, Zhiyu Chen, and Xifeng Yan. Inductive relation prediction by bert. In AAAI, volume 36, pages 5923–5931, 2022. doi:10.1609/ AAAI.V3615.20537.
- 223 Chaoning Zhang, Chenshuang Zhang, Sheng Zheng, Yu Qiao, Chenghao Li, Mengchun Zhang, Sumit Kumar Dam, Chu Myaet Thwal, Ye Lin Tun, Le Luang Huy, Donguk kim, Sung-Ho Bae, Lik-Hang Lee, Yang Yang, Heng Tao Shen, In-So Kweon, and Choong-Seon Hong. A complete survey on generative ai (aigc): Is chatgpt from gpt-4 to gpt-5 all you need? ArXiv, 2023. doi:10.48550/ARXIV.2303.11717.
- 224 Meiru Zhang, Yixuan Su, Zaiqiao Meng, Zihao Fu, and Nigel Collier. COFFEE: a contrastive oracle-free framework for event extraction. ArXiv,

abs/2303.14452, 2023. doi:10.48550/ARXIV.2303.14452.

- 225 Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi. A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning. VLDB J., 31(5):1143–1168, 2022. doi: 10.1007/S00778-022-00747-Z.
- 226 Ruichuan Zhang and Nora El-Gohary. Transformer-based approach for automated context-aware IFC-regulation semantic information alignment. *Automation in Construction*, 145, 2023. doi:10.1016/j.autcon.2022.104540.
- 227 Zhiyuan Zhang, Xiaoqian Liu, Yi Zhang, Qi Su, Xu Sun, and Bin He. Pretrain-kge: Learning knowledge representation from pretrained language models. In *EMNLP Findings*, 2020. doi: 10.18653/V1/2020.FINDINGS-EMNLP.25.
- 228 Ziheng Zhang, Hualuo Liu, Jiaoyan Chen, Xi Chen, Bo Liu, Yuejia Xiang, and Yefeng Zheng. An industry evaluation of embedding-based entity alignment. In *COLING*, pages 179–189, 2020. doi:10.18653/V1/2020.COLING-INDUSTRY.17.
- 229 Ruochen Zhao, Xingxuan Li, Shafiq Joty, Chengwei Qin, and Lidong Bing. Verify-and-edit: A knowledge-enhanced chain-of-thought framework. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 5823–5840, Toronto, Canada, jul 2023. Association for Computational Linguistics. doi: 10.18653/V1/2023.ACL-LONG.320.
- 230 Zexuan Zhong and Danqi Chen. A frustratingly easy approach for entity and relation extraction. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 50–61, Online, jun 2021. Association for Computational Linguistics. doi:10.18653/V1/2021.NAACL-MAIN.5.
- 231 Zexuan Zhong, Dan Friedman, and Danqi Chen. Factual probing is [mask]: Learning vs. learning to recall. In NAACL, 2021. doi:10.18653/V1/2021. NAACL-MAIN.398.
- 232 Wenxuan Zhou, Fangyu Liu, Ivan Vulic, Nigel Collier, and Muhao Chen. Prix-lm: Pretraining for multilingual knowledge base construction. In ACL, 2021. doi:10.48550/arXiv.2110.08443.
- 233 Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are humanlevel prompt engineers. In arXiv, 2023. doi: 10.48550/ARXIV.2211.01910.
- 234 Qi Zhu, Hao Wei, Bunyamin Sisman, Da Zheng, Christos Faloutsos, Xin Luna Dong, and Jiawei Han. Collective multi-type entity alignment between knowledge graphs. In WWW, pages 2241– 2252, 2020. doi:10.1145/3366423.3380289.
- 235 Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. arXiv, 2023. doi:10.48550/ARXIV.2305.13168.

Knowledge Engineering Using Large Language Models

Bradley P. Allen¹ \square \square

University of Amsterdam, Amsterdam, The Netherlands

Lise Stork 🖂 💿

Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Paul Groth 🖂 💿

University of Amsterdam, Amsterdam, The Netherlands

— Abstract -

Knowledge engineering is a discipline that focuses on the creation and maintenance of processes that generate and apply knowledge. Traditionally, knowledge engineering approaches have focused on knowledge expressed in formal languages. The emergence of large language models and their capabilities to effectively work with natural language, in its broadest sense, raises questions about the foundations

and practice of knowledge engineering. Here, we outline the potential role of LLMs in knowledge engineering, identifying two central directions: 1) creating hybrid neuro-symbolic knowledge systems; and 2) enabling knowledge engineering in natural language. Additionally, we formulate key open research questions to tackle these directions.

2012 ACM Subject Classification Computing methodologies \rightarrow Natural language processing; Computing methodologies \rightarrow Machine learning; Computing methodologies \rightarrow Philosophical/theoretical foundations of artificial intelligence; Software and its engineering \rightarrow Software development methods

Keywords and phrases knowledge engineering, large language models

Digital Object Identifier 10.4230/TGDK.1.1.3

Category Vision

Related Version Previous Version: https://doi.org/10.48550/arXiv.2310.00637

Funding Lise Stork: EU's Horizon Europe research and innovation programme, the MUHAI project (grant agreement no. 951846).

Paul Groth: EU's Horizon Europe research and innovation programme, the ENEXA project (grant Agreement no. 101070305).

Acknowledgements This work has benefited from Dagstuhl Seminar 22372 "Knowledge Graphs and Their Role in the Knowledge Engineering of the 21st Century." We also thank Frank van Harmelen for conversations on this topic.

Received 2023-06-30 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1 Introduction

Knowledge engineering (KE) is a discipline concerned with the development and maintenance of automated processes that generate and apply knowledge [4, 93]. Knowledge engineering rose to prominence in the nineteen-seventies, when Edward Feigenbaum and others became convinced that automating knowledge production through the application of research into artificial intelligence required a domain-specific focus [32]. From the mid-nineteen-seventies to the nineteen-eighties, knowledge engineering was mainly defined as the development of expert systems for automated decision-making. By the early nineteen-nineties, however, it became clear that the expert systems approach, given its dependence on manual knowledge acquisition and rule-based representation

© Bradley P. Allen, Lise Stork, and Paul Groth;

licensed under Creative Commons License CC-BY 4.0 Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 3, pp. 3:1-3:19

Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

¹ Corresponding author

3:2 Knowledge Engineering Using Large Language Models

of knowledge by highly skilled knowledge engineers, resulted in systems that were expensive to maintain and difficult to adapt to changing requirements or application contexts. Feigenbaum argued that, to be successful, future knowledge-based systems would need to be scalable, globally distributed, and interoperable [34].

The establishment of the World Wide Web and the emergence of Web architectural principles in the mid-nineteen-nineties provided a means to address these requirements. Tim Berners-Lee argued for a "Web of Data" based on linked data principles, standard ontologies, and data sharing protocols that established open standards for knowledge representation and delivery on and across the Web [11]. The subsequent twenty years witnessed the development of a globally federated open linked data "cloud" [13], the refinement of techniques for ontology engineering [51], and methodologies for the development of knowledge-based systems [86]. During the same period, increasing use of machine learning and natural language processing techniques led to new means of knowledge production through the automated extraction of knowledge from natural language documents and structured data sources [26, 68]. Internet-based businesses in particular found value in using such technologies to improve access to and discovery of Web content and data [43]. A consensus emerged around the use of knowledge graphs as the main approach to knowledge representation in the practice of knowledge engineering in both commercial and research arenas, providing easier reuse of knowledge across different tasks and a better developer experience for knowledge engineers [45].

More recently, the increase in the availability of graphical processing hardware for fast matrix arithmetic, and the exploitation of such hardware to drive concurrent innovations in neural network architectures at heretofore unseen scales [106], has led to a new set of possibilities for the production of knowledge using large language models (LLMs). LLMs are probabilistic models of natural language, trained on very large corpora of content, principally acquired from the Web. Similar to previous approaches to language modeling, given a sequence of tokens, LLMs predict a probable next sequence of tokens based on a learned probability distribution of such sequences. However, presumably due to the vast amount of content processed in learning and the large size and architecture of the neural networks involved, LLMs exhibit remarkable capabilities for natural language processing that far exceed earlier approaches [60].

These capabilities include the ability to do zero- or few-shot learning across domains [20], to generalize across tasks, including the ability to perform domain-independent question answering integrating large amounts of world knowledge [77], to generate text passages at human levels of fluency and coherence [28, 96], to deal gracefully with ambiguity and long-range dependencies in natural language [104], and to reduce or even eliminate the need for manual feature engineering [98]. LLMs also exhibit the ability to generate and interpret structured and semi-structured information, including programming language code [6, 100], tables [46, 53], and RDF metadata [106, 58, 7]. The generalization of language models (termed "foundation models" by some) to other modalities including images and audio have led to similarly significant advances in image understanding [23, 117], image generation [38, 79, 83], speech recognition, and text-to-speech generation [78, 105]. Such capabilities have prompted a significant amount of research and development activity demonstrating potential applications of LLMs [66, 84, 54]. However, the means of incorporating LLMs into structured, controllable, and repeatable approaches to developing and fielding such applications in production use are only just beginning to be considered in detail [73].

This paper engages with the question of how LLMs can be effectively employed in the context of knowledge engineering. We start by examining the different forms that knowledge can take, both as inputs for constructing knowledge systems and as outputs of such systems. We argue that the distinction between knowledge expressed in natural language (or other evolved, naturally occurring modalities such as images or video) and knowledge expressed in formal languages (for example, as knowledge graphs or rules), sheds light how LLMs can be brought to bear on the development of knowledge systems.

Based on this perspective, we then describe two potential paths forward. One approach involves treating LLMs as components within hybrid neuro-symbolic knowledge systems. The other approach treats LLMs and prompt engineering [57] as a standalone approach to knowledge engineering², using natural language as the primary representation of knowledge. We then enumerate a set of open research problems in the exploration of these paths. These problems aim to determine the feasibility of and potential approaches to using LLMs with existing KE methodologies, as well as the development of new KE methodologies centered around LLMs and prompt engineering.

2 Forms of knowledge and their engineering

In the history of the computational investigation of knowledge engineering, knowledge has been often treated primarily as symbolic expressions. However, as [39] noted, knowledge is actually encoded in a variety of media and forms, most notably in natural language (e.g. English) but also in images, video, or even spreadsheets. This fact becomes even more apparent when looking at institutional knowledge practices that have developed over centuries, for example, in the sciences or archives [44]. We now illustrate this point by describing the many ways in which knowledge manifests itself in the context of biodiversity informatics.

2.1 The multimodal richness of knowledge: an example from biodiversity sciences

The ultimate goal of biodiversity science is to understand species evolution, variation, and distribution, but finds applications in a variety of other fields such as climate science and policy. At its heart is the collection and observation of organisms, providing evidence for deductions about the natural world [59]. Such knowledge is inherently multimodal in nature, most commonly appearing in the form of images, physical objects, tree structures and sequences, i.e., molecular data.

Historically, organism sightings have been carefully logged in handwritten field diaries to describe species behavior and environmental conditions. Detailed drawings and later photographs were made to capture color, organs and other knowledge about an organism's traits used for identification, which is best conveyed visually but which is challenging to preserve in natural specimens. These manuscripts are housed, together with the physical zoological specimens and herbaria which they describe, in museums and collection facilities across the world. Both the multimodal nature of these knowledge sources as well as their distributed nature hamper knowledge integration and synthesis.

Metadata describes the specimen's provenance: where specimens were found, who found them, and provides an attempt at identifying the type of organism (such as the preserved squid specimen shown in Figure 1). Such knowledge is paramount, as it allows researchers to understand resources within the context in which they were produced, enabling researchers to carry out ecological studies such as distribution modeling over time.

For a systematic comparison of the multitude of resources available, the biodiversity sciences have had a long-standing tradition of developing information standards [67]. From Linnaeus' Systema naturae mid 18th century as well as his formal introduction of zoological nomenclature,

 $^{^2}$ As defined by [57], prompt engineering is finding the most appropriate prompt or input text to an LLM to have it solve a given task.

3:4 Knowledge Engineering Using Large Language Models

taxonomists have started categorizing natural specimens according to tree-like hierarchical structures. The process is challenging, given that biologists up until this day do not have a full picture of all living organisms on earth, and incomplete, naturally evolved and fuzzy knowledge is not easily systematized.



Figure 1 A specimen of the *Loligo vulgaris Lamarck*, 1798 species from the *Naturalis–Zoology and Geology* catalogues.^{*a*} Images free of known restrictions under copyright law (Public Domain Mark 1.0).

^a https://bioportal.naturalis.nl/nl/specimen/RMNH.MOL.5009890

The development of digital methods has opened up new pathways for comparison and analysis. Gene sequencing technology has led biologists to the genetic comparison of species, by the calculation of ancestry and construction of evolutionary tree structures in the study of phylogeny [50]. More importantly, digital methods allowed the transfer of analog resources, such as specimen collection scans [14] and metadata, to the digital world. Such techniques have furthered formalization and thereby interoperability of collected data through the use of Web standards, such as globally unique identifiers for species names [72] as well as shared vocabularies for data integration across collections [10]. The Global Biodiversity Information Facility (GBIF) and their data integration toolkit serves as a great example of such integration efforts [97, 81]. Currently, there is a large emphasis on linking up disparate digital resources in the creation of an interconnected network of digital collection objects on the Web, linked up with relevant ecological, environmental and other related data in support of machine actionability (i.e., the ability of computational systems to find, access, interoperate, and reuse data with minimal intervention) for an array of interdisciplinary tasks such as fact-based decision-making and forecasting [41]. Using data standards for describing and reasoning over collection data can aid researchers counter unwanted biases via transparency. However, making data comply with data standards can also lead to oversimplification or reinterpretation [71].

Machine learning and knowledge engineering strategies can help to (semi-)automatically extract and structure biodiversity knowledge according [102, 91], for instance using state-of-the-art computer vision or natural language processing techniques as well as crowd-sourcing platforms for the annotation of field diaries and other collection objects with formal language [92, 29]. Nevertheless, a bottleneck in the digitization of collections and their use for machine actionability is the amount of work and domain expertise required for the formalization of such knowledge, and the extraction from unstructured texts, images and videos. Historical resources, i.e. handwritten texts, pose an additional challenge, as they are exceptionally challenging to interpret within the current scientific paradigm [107].

The variety and usefulness of different forms of knowledge both natural and formal and the challenges they pose is not limited to the biodiversity domain as described above. We see the same diversity happening in law [82], medicine [16, 21] and even self-driving vehicles [9]. To summarize:

- domain knowledge is often best represented in a variety of modalities, i.e., images, taxonomies, or free text, each modality with its own data structure and characteristics which should be preserved, and no easy way of integrating, interfacing with or reasoning over multimodal knowledge in a federated way exists;
- provenance of data is paramount in understanding knowledge within the context in which it was produced;
- = fuzzy, incomplete, or complex knowledge is not easily systematized;
- using data standards for describing and reasoning over collection data can aid researchers counter unwanted biases via transparency;
- making data comply with data standards can lead to oversimplification or reinterpretation;
- the production of structured domain knowledge, for instance from images or free text, requires domain expertise, and is therefore labor intensive and costly;
- knowledge evolves, and knowledge-based systems are required to deal with updates in their knowledge bases.

2.2 KE as the transformation of knowledge expressed in natural language into knowledge expressed in a formal language

This sort of rich and complex array of modalities for the representation of knowledge has traditionally posed a challenge to knowledge engineers [33]. Much of the literature on knowledge engineering methodology has focused on the ways in which knowledge in these naturally-occurring forms can be recast into a structured symbolic representation, e.g., using methods of knowledge elicitation from subject matter experts [88], for instance by the formulation of competency questions for analysing application ontologies [12]. One way to think about this is as the process of expressing knowledge presented in a natural, humanly evolved language in a formally-defined language. This notion of the transformation of natural language into a formal language as a means of enabling effective reasoning has a deep history rooted in methodologies developed by analytical philosophers of the early twentieth century [24, 69], but dating even further back to Liebniz's *lingua rationalis* [35] and the thought of Ramón Lull [37]. Catarina Dutilh Novaes [69] has argued that formal languages enable reasoning that is less skewed by bias and held beliefs, an effect achieved through *de-semantification*, i.e., the process of replacing terms in a natural language with symbols that can be manipulated without interpretation using a system of rules of transformation. Coupled with sensorimotor manipulation of symbols in a notational system, people can reason in a manner that outstrips their abilities unaided by such a technology.

While Dutilh Novaes' analysis focuses on this idea of formal languages as a cognitive tool used by humans directly, e.g. through the manipulation of a system of notation using paper and pencil, she notes that this manipulation of symbols is the route to the mechanization of reasoning through computation. When externally manifested as a function executed by a machine through either interpretation by an inference engine, or through compilation into a machine-level language, this approach of formalization yields the benefits of reliability, greater speed and efficiency in reasoning.

This idea captures precisely the essence of the practice of knowledge engineering: Starting from sources of knowledge expressed in natural language and other modalities of human expression, through the process of formalization [51, 95], knowledge engineers create computational artifacts embodying this knowledge. These computational artifacts then enable us to reason using this knowledge in a predictable, efficient, and repeatable fashion. This is done either by proxy through the action of autonomous agents, or in the context of human-mediated decision-making processes.

3:6 Knowledge Engineering Using Large Language Models

2.3 LLMs as a general-purpose technology for transforming natural language into formal language

Until recently, there have been two ways in which this sort of formalization could be performed: through the manual authoring of symbolic/logical representations, e.g., as in the traditional notion of expert systems [34], or through the use of machine learning and natural language processing to extract such representations automatically from natural language text [61]. But what has become evident with the emergence of LLMs, with their capabilities for language learning and processing, is that they provide a new and powerful type of general purpose tool for mapping between natural language³ and formal language, as well as other modalities. LLMs have shown state-of-the-art performance on challenging NLP tasks such as relation extraction [5] or text abstraction/summarization [114], and have been used to translate between other modalities, such as images and text (called vision-language models [119, 77]) in computer vision tasks, or from natural language to code [113, 47], in which a pretrained task-agnostic language model can be zero-shot and few-shot transferred to perform a certain task [20, 52]. If one accepts the position that KE can be generally described as the process of transforming knowledge in natural language into knowledge in formal language, then it becomes clear that LLMs provide an advance in our ability to perform knowledge engineering tasks.

3 The use of LLMs in the practice of knowledge engineering: two scenarios

Given the above discussion, the natural question that arises is: what might be the utility and impact of the use of LLMs for the transformation of natural language into formal language, when applied in the context of the practice of knowledge engineering?

When LLMs emerged as a new technology in the mid-2010s, two views of the relationship between LLMs and knowledge bases (KBs) were put forward. One was the LLM can be a useful component for various processes that are part of a larger knowledge engineering workflow (i.e. "LMs for KBs" [3]); the other was that the LLM is a cognitive artifact that can be treated as a knowledge base in and of itself (i.e., "LMs as KBs" [75]). We exploit this dichotomy to formulate a pair of possible future scenarios for the use of LLMs in the practice of KE. One is to use LLMs as a technology for or tool in support of implementing knowledge tasks that have traditionally been built using older technologies such as rule bases and natural language processing (NLP). Another is to use LLMs to remove the need for knowledge task to be fluent in a formal language, i.e., by allowing knowledge for a given knowledge task to be expressed in natural language, and then using prompt engineering as the primary paradigm for the implementation of reasoning and learning. We now explore each of these scenarios in turn, and consider the open research problems that they raise.

3.1 LLMs as components or tools used in knowledge engineering

We illustrate the first scenario through reference to CommonKADS [86], a structured methodology that has been used by knowledge engineers since the early 2000's. CommonKADS is the refinement of an approach to providing a disciplined approach to the development of knowledge systems. This approach saw initial development in the nineteen-eighties as a reaction to both the ad-hoc nature of early expert systems development [111] and to the frequency of failures in the deployment of expert systems in an organizational context [34]. Stemming from early work on making expert

³ Again, we note that natural language should be read to include all modalities. Hence, the term "foundation model" [15] was coined to refer to LLMs.

systems development understandable and repeatable [42], CommonKADS is distinguished from methodologies more focused on ontology development (e.g., NeON [94], Kendall and McGuinness's "Ontology 101" framework [51], and Presutti's ontology design patterns [76]) in that it provides practical guidance for specification and implementation of knowledge systems components in a broader sense. It attempts to provide a synoptic guide to the full scope of activities involved in the practice of KE, and show how it relates to the activities of the organization in which that engineering is taking place. As such, in the context of this paper we can use it as a framework to explore for what tasks and in what ways LLMs can be used for KE.

Some tasks identified by CommonKADS as part of the KE process may remain largely unchanged by the use of LLMs. These include knowledge task identification and project organizational design. But others can involve the use of LLMs. LLMs can assist knowledge engineers and/or knowledge providers in the performance of knowledge engineering tasks. They can also be a means for the implementation of modules performing knowledge-intensive tasks. Examples of these uses include the following:

- Knowledge acquisition and elicitation. LLMs can be used to support knowledge acquisition and elicitation in a given domain of interest. Engineers can create prompts that target specific aspects of the domain, using the responses as a starting point for building the knowledge base. Dialogs between LLMs trained using such prompts and knowledge providers, the subject matter experts, can support the review, validation, and refinement of the acquired knowledge [8].
- Knowledge organization. LLMs can be used to organize the acquired knowledge into a coherent structure using natural language, making it easy to understand and update. Prompt engineering can be used to develop a set of prompts that extract formal language using the LLM, e.g., for text to graph generation [40] or vice versa [18, 2]. Moreover, LLMs are used for program synthesis [113, 47], the generation of metadata [56] or for fusing knowledge graphs [118].
- **Data augmentation.** LLMs can be used to generate synthetic training data to aid in testing the knowledge system by evaluating its performance on instances of the specific task [116].
- **Testing and refinement.** Feedback from subject matter experts and users can be used to prompt an LLM to refine the natural language knowledge base and improve the system's accuracy and efficiency through self-correction of prompts and tuning of the LLM model settings as needed to optimize the system's performance [110].
- Maintenance. LLMs can be used to monitor new information and trends, and to then propose new prompts integrating those updates into the knowledge base.

Consider the CommonKADS knowledge task hierarchy shown in Figure 2. Synthetic knowledgeintensive tasks, e.g. design or configuration, are amenable to generative approaches [109]; analytic knowledge-intensive tasks can involve LLM components within a hybrid neuro-symbolic knowledge system.

A shortcoming of using CommonKADS for our purposes, however, is that it predates the widespread use of machine learning and statistical natural language processing in KE. A number of architectural approaches have since been developed that extend the CommonKADS concepts of a knowledge-intensive task type hierarchy and knowledge module templates. These include modeling the fine-grained data flows and workflows associated with knowledge systems that combine components that ingest, clean, transform, aggregate and generate data, as well as generate and apply models built using machine learning [103, 19, 27, 31, 101]. These architectures are put forward as providing a general framework for composing heterogeneous tools for knowledge representation and inference into a single integrated hybrid neuro-symbolic system. The design pattern notations put forward in recent work [103, 101, 31] treat data, models, and symbolic representations as the inputs and outputs of components composed into a variety of knowledge

3:8 Knowledge Engineering Using Large Language Models

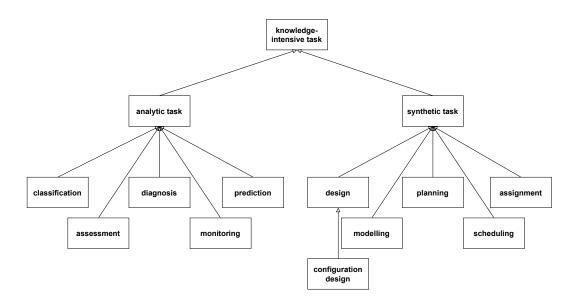


Figure 2 Hierarchy of knowledge-intensive task types from CommonKADS ([86], p.125).

system design patterns. Generalizing these into natural language and formal language inputs and outputs can provide a simple way to extend these design notations to accommodate both LLMs as well as a richer set of knowledge representations.

3.2 Knowledge engineering as prompt engineering

Given that LLMs enable knowledge modeling in natural language, it is conceivable that the programming of knowledge modules could take place entirely in natural language. Consider that prompt programming is "finding the most appropriate prompt to allow an LLM to solve a task" [57]. One can through this lens view knowledge engineering as the crafting of dialogues in which a subject matter expert (SME) arrives at a conclusion by considering the preceding context and argumentation [80, 109, 89, 60]. This framing of knowledge engineering as prompt engineering is the second scenario we wish to explore.

From the perspective of the CommonKADS knowledge-intensive task type hierarchy, this would involve a redefinition of the types and hierarchy to use LLMs and prompt programming design patterns, e.g. as described in [57]. Several aspects of this redefinition could include:

Natural language inference. LLMs can be used to build natural language inference engines that use the organized knowledge to perform the specific task by taking input queries and generate output using prompt engineering to guide the LLM towards generating accurate inferences, e.g. using zero- or few-shot chain-of-thought design patterns. The benefit here is that the gap between the knowledge engineer, knowledge provider (the subject matter expert) and the user is smaller since a translation to a formal language (the language of the engineer) is no longer required.

- Knowledge-intensive task execution through human/machine dialog. LLMs can be used to construct a conversational interface that allows users to interact with the knowledge system and receive task-specific support.
- **Testing and refinement through human/machine dialog.** Feedback from subject matter experts and users can be used to prompt an LLM to refine the natural language knowledge base and improve the system's accuracy and efficiency through self-correction of prompts and tuning of the LLM model settings as needed to optimize the system's performance.

One possible benefit of this approach would be that the barrier to adoption of knowledge engineering as a practice could be lowered significantly. Knowledge elicitation could be conducted entirely within natural language, meaning that subject matter experts without training in formal knowledge representations could perform these tasks directly. However, this approach assumes that predictable inference [101] using natural language is satisfactory. The propensity of current LLMs to "hallucinate", i.e., to confabulate facts, is an obstacle to the realization of this idea [48]. Multiple efforts have been devoted to the creation of prompt programming patterns that address this issue, ranging from chain-of-thought approaches [108] to retrieval-assisted generation, i.e. the augmentation of LLMs with authoritative document indexes and stores [84, 65]. Recent work [73] has described ways in which knowledge graphs as a formal language can be integrated with natural language and LLM-based language processing and reasoning to provide knowledge systems architectures that directly address this issue. [115] surveys work in this direction.

4 Open research questions

Using the scenarios outlined above, we can identify a number of open research questions to be addressed to realize either or both of these two possible approaches to the use of LLMs in knowledge engineering. These questions touch on three general areas: the impact of LLMs on the methodologies used to build knowledge systems, on the architectural design of knowledge systems incorporating and/or based on LLMs, and on the evaluation of such systems. For each of these open questions, we provide a link back to the biodiversity scenario discussed in Section 2.1 denoted by a \swarrow .

4.1 Methodology

4.1.1 How can knowledge engineering methodologies best be adapted to use LLMs?

How can we harmoniously meld the considerable body of work on knowledge engineering methodologies [51, 36, 76, 94, 87, 85, 90] with the new capabilities presented by LLMs?

Schreiber's conceptualization of knowledge engineering as the construction of different aspect models of human knowledge [86], as discussed above, offers a framework for further elaboration. The distinctive characteristics of LLMs, coupled with prompt engineering, present unique challenges and opportunities for building agents within a knowledge system, one that is consistent with the CommonKADS approach.

While the role definitions within KE methodologies might mostly remain the same, the skills required for knowledge engineers will need morphing to adapt to the LLM environment. This evolution of roles calls for an extensive investigation into what these new skills might look like, and how they can be cultivated. Additionally, the adaptability of the various knowledge-intensive task type hierarchies described by CommonKADS and its descendants in the literature on hybrid neuro-symbolic systems (e.g., as described in [19]) to accommodate LLMs is another fertile area for exploration.

LLM-based applications, likened to synthetic tasks within these knowledge engineering frameworks, raise compelling research questions regarding accuracy and the prevention of hallucinations. LLM-based applications have a lower bar to reach with respect to notions of accuracy and avoidance of hallucinations, but still must provide useful and reliable guidance to users and practitioners.

 \checkmark Connecting back to the biodiversity domain, answering these questions would provide guidance on the appropriate methodology to adopt when developing a new specimen curation and collection knowledge management system that needs to deal with multimodal assets like handwritten text or images.

3:10 Knowledge Engineering Using Large Language Models

4.1.2 How do principles of content and data management apply to prompt engineering?

Applying content and/or data management principles to collections of prompts and prompt templates, integral to work with LLMs, is an area ripe for exploration. Properly managing these resources could improve efficiency and guide the development of improved methodologies in knowledge engineering. This calls for a rigorous investigation of current data management practices, their applicability to LLMs, and potential areas of refinement. Ensuring the reproducibility of LLM engineering from a FAIR data standpoint [112] is a crucial yet complex challenge. Developing and validating practices and protocols that facilitate easy tracing and reproduction of LLM-based processes and outputs is central to this endeavor.

Addressing this challenge will aid researchers in applying LLM engineering in a FAIR way. Doing so is critical for biodiversity research and science in general where precision, reproducibility and provenance are key for knowledge discovery and research integrity.

4.1.3 What are the cognitive norms that govern the conduct of KE?

A crucial area of inquiry involves the identification and understanding of *cognitive norms*, as described by Menary [62], that govern the practice of knowledge engineering. Cognitive norms are established within a human community of practice as a way of governing the acceptable use of "external representational vehicles to complete a cognitive task" [63]. As the consumer adoption of LLM technology has progressed, we see a great deal of controversy about when and how it is appropriate to use, e.g. in the context of education or the authoring of research publications. Understanding how these norms shape the use of LLMs in this context is an under-explored field of study. By unraveling the interplay between these cognitive norms and LLM usage, we can gain valuable insights into the dynamics of knowledge engineering practices and possibly foster more effective and responsible uses of LLMs.

 \checkmark In the biodiversity sciences, this means understanding the cognitive norms specific to the domain, to understand how LLMs can be used in a way that respects the domain's practices and standards.

4.1.4 How do LLMs impact the labor economics of KE?

A related but distinct question pertains to the impact of LLMs on the economic costs associated with knowledge engineering. The introduction and application of LLMs in this field may significantly alter the economic landscape, either by driving costs down through automation and efficiency or by introducing new costs tied to system development, maintenance, and oversight. Thoroughly exploring these economic implications can shed light on the broader effects of integrating LLMs into knowledge engineering.

The realm of labor economics as it pertains to hybrid or *centaur* systems [1], is another area ripe for investigation. Understanding how the deployment of these systems influences labor distribution, skill requirements, and job roles could provide valuable input into the planning and implementation of such technologies. Additionally, it could reveal the potential societal and economic impacts of this technological evolution.

Developments for LLM-based KE can help mitigate labor of knowledge experts in the biodiversity sciences, for instance by the development of more efficient KE workflows for the digitization of museum specimens or manuscripts.

4.2 Architecture

4.2.1 How can hybrid neuro-symbolic architectural models incorporate LLMs?

Design patterns for hybrid neuro-symbolic systems, as described in [103], offer a structured approach to comprehend the flow of data within a knowledge system. Adapting this model to account for the differences between natural and formal language could significantly enhance our ability to trace and manage data within knowledge systems. A salient research question emerging from this scenario pertains to the actual process of integrating LLMs into knowledge engineering data processing flows [27]. Understanding the nuances of this process will involve a deep examination of the shifts in methodologies, practices, and the potential re-evaluations of existing knowledge engineering paradigms. The perspective of KE enabled by LLMs as focused on the transformation of natural language into formal language provides insights that can be used to improve the motivation for hybrid neuro-symbolic systems; e.g., [19] references [17] in using dual process theories of reasoning (i.e. the "System 1/System 2" model described in [49]) as a motivation for hybridization in knowledge systems, but more recent analyses [69, 64] cast doubt on the validity of such models, and point to more nuanced perspectives that provide a better grounding for the benefits of hybridization.

Addressing these questions would shed light on tasks for which hybridization using LLMs would prove favorable, e.g., image classification of species.

4.2.2 How can prompt engineering patterns support reasoning in natural language?

One fundamental question that arises is how prompt engineering patterns can be utilized to facilitate reasoning in natural language. Exploring this topic involves understanding the mechanics of these patterns and their implications on natural language processing capabilities of LLMs. This line of research could open new possibilities for enhancing the functionality and efficiency of these models.

A related inquiry concerns the structure, controllability, and repeatability of reasoning facilitated by LLMs. Examining ways to create structured, manageable, and reproducible reasoning processes within these models could significantly advance our capacity to handle complex knowledge engineering tasks and improve the reliability of LLMs.

The interaction of LLMs and approaches to reasoning based on probabilistic formalisms is also an underexplored area of research. A particularly evocative effort in this area is that described in [113], which describes the use of LLMs to transform natural language into programs in a probabilistic programming language, which can then be executed to support reasoning in a particular problem domain. We note that this work provides an excellent example of the knowledge engineering as the transformation of natural language into formal language perspective and of the impact of LLMs in advancing that perspective. Investigating how to automatically generate and assess other nuanced forms of knowledge within LLMs could lead to a more refined understanding of these models and their capabilities.

➡ Given that biodiversity knowledge is often best represented in a variety of modalities each with their own data structures and characteristics, research may explore how LLMs can act as natural language interfaces to such multimodal knowledge bases.

3:12 Knowledge Engineering Using Large Language Models

4.2.3 How can we manage bias, trust and control in LLMs using knowledge graphs?

Trust, control, and bias in LLMs, especially when these models leverage knowledge graphs, are critical areas to explore. Understanding how to detect, measure, and mitigate bias, as well as establish trust and exert control in these models, is an essential aspect of ensuring ethical and responsible use of LLMs. Furthermore, investigating methods to update facts in LLMs serving as knowledge graphs is a crucial area of research. Developing strategies for efficient and reliable fact updating could enhance the accuracy and usefulness of these models.

Another key question involves understanding how we can add provenance to statements produced by LLMs. This line of research could prove vital in tracking the origin of information within these models, thus enhancing their reliability and usability. It opens the door to more robust auditing and validation practices in the use of LLMs.

Addressing this challenge can help biodiversity researchers detect and mitigate biases, as use of LLMs might further exacerbate knowledge gaps, e.g., groups of individuals omitted from historical narratives in archival collections. Moreover, novel update mechanisms can aid researchers to reliably update facts or changing knowledge structures learned by LLMs, for instance when domain knowledge evolves.

4.2.4 Is extrinsic explanation sufficient?

A significant area of interest pertains to how we can effectively address the explainability of answers generated using LLMs [30]. This exploration requires a deep dive into the functioning of LLMs and the mechanisms that govern their responses to prompts. Developing a thorough understanding of these processes can aid in creating transparency and trust in LLMs, as well as fostering their effective use.

The need for explanation in LLMs also leads to the question of whether extrinsic explanation is sufficient for the purposes of justifying a knowledge system's reasoning, as argued in general for the intelligibility of knowledge systems by Cappelen and Devers [22], or if intrinsic explainability is a necessary requirement [55]. This question calls for a thoughtful exploration of the value and limitations of both extrinsic and intrinsic explanation methodologies, and their implications for the understanding and usage of LLMs. An exciting research avenue arises from the work of Tiddi [99], concerning explainability with formal languages. The exploration of this topic could reveal significant insights into how we can leverage formal languages to enhance the explainability of LLMs. This could pave the way for new methods to increase transparency and intelligibility in these models.

✓ In the sciences in general, answering these questions would aid explainability of LLM-generated answers via curated facts, increasing transparency and trust.

4.2.5 How can LLMs support the engineering of hybrid human/machine knowledge systems?

Another topic of interest involves exploring the potential of hybrid systems that combine human cognition with machine capabilities within a dialogical framework [64, 70]. As an exciting example of the possibilities for new approaches to human/machine collaboration in this vein, we point to the recent results reported by [74] on the creation of conversational agents that simulate goal-directed human conversation and collaboration on tasks. One can imagine coupling LLM-based agents with human interlocutors working collaboratively in this manner on specific knowledge-intensive tasks. Understanding how to develop these types of systems, and what their implications might

be for the practice of knowledge engineering presents a fertile research line. It requires the careful analysis of human-machine interaction, the study of system design principles, and the investigation of their potential impact.

4.3 Evaluation

4.3.1 How do we evaluate knowledge systems with LLM components?

The first point of interest involves the evaluation of knowledge-based systems, with a focus beyond just logic. This area calls for innovative methodologies to assess the system's capacity to manage and utilize knowledge efficiently, going beyond traditional logical evaluations. This topic of evaluation naturally extends to the question of how we evaluate ontologies and design patterns within knowledge engineering. Evaluating these aspects would require a deep dive into the structures and mechanisms underpinning these elements, potentially leading to the development of refined evaluation metrics and methodologies.

Interestingly, the long-standing paradigm of machine learning evaluation, relying on benchmarking against a standard train/test dataset, seems to falter in the era of LLMs [25]. This presents an intriguing challenge for researchers and engineers alike. It is quite possible that traditional methods may need to be significantly buttressed by methodologies and supporting tools for the direct human evaluation of knowledge system performance. This has implications concerning the cost and speed of evaluation processes, encouraging the rethink of current approaches to perhaps develop new strategies that balance accuracy, cost-effectiveness, and timeliness. Reimagining evaluation methodologies in this new context could provide transformative insights into how we can gain confidence in the reliability engineering of knowledge systems that use LLMs.

Developments in this direction may aid biodiversity researchers to get a better understanding of the real-world efficacy of employing knowledge-based systems with LLM components in their institutions. One can think of improving access to collections, knowledge discovery, or accuracy in describing institutional knowledge.

4.3.2 What is the relationship between evaluation and explainability?

Lastly, there is an inherent dependency of evaluation on effective solutions for explainability within knowledge systems. Understanding this relationship could help in the creation of more comprehensive evaluation models that take into account not only the performance of a system but also its explainability.

5 Summary

In this paper, we have advocated for a reconsideration of the practice and methodology of knowledge engineering in light of the emergence of LLMs. We argued that LLMs allow naturally-occurring and humanly-evolved means of conveying knowledge to be brought to bear in the automation of knowledge tasks. We described how this can enhance the engineering of hybrid neuro-symbolic knowledge systems, and how this can make knowledge engineering possible by people who do not necessarily have the experience of recasting natural language into formal, structured representation languages. Both of these possibilities will involve addressing a broad range of open questions, which we have attempted to outline above. Given the rapid pace of the development of this area of research, it is our earnest hope that the coming months and years will yield results shedding light on these questions.

— References -

- 1 Zeynep Akata, Dan Balliet, Maarten De Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, et al. A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(8):18–28, 2020. doi:10.1109/MC.2020.2996587.
- 2 Dimitrios Alivanistos, Selene Báez Santamaría, Michael Cochez, Jan Christoph Kalo, Emile van Krieken, and Thiviyan Thanapalasingam. Prompting as probing: Using language models for knowledge base construction. In Sneha Singhania, Tuan-Phong Nguyen, and Simon Razniewski, editors, *LM-KBC 2022 Knowledge Base Construction from Pre-trained Language Models 2022*, CEUR Workshop Proceedings, pages 11–34. CEUR-WS.org, 2022. doi:10.48550/ARXIV.2208.11057.
- 3 Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. A review on language models as knowledge bases. arXiv preprint arXiv:2204.06031, 2022. doi:10.48550/ ARXIV.2204.06031.
- 4 Bradley P Allen, Filip Ilievski, and Saurav Joshi. Identifying and consolidating knowledge engineering requirements. arXiv preprint arXiv:2306.15124, 2023. doi:10.48550/ARXIV. 2306.15124.
- 5 Christoph Alt, Marc Hübner, and Leonhard Hennig. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1388–1398, 2019. doi:10.18653/V1/P19-1134.
- 6 Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, et al. Program synthesis with large language models. arXiv preprint arXiv:2108.07732, 2021. arXiv:2108.07732.
- 7 Agnes Axelsson and Gabriel Skantze. Using large language models for zero-shot natural language generation from knowledge graphs. arXiv preprint arXiv:2307.07312, 2023. doi:10.48550/ ARXIV.2307.07312.
- 8 Stephen Bach, Victor Sanh, Zheng Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesht Sharma, Taewoon Kim, M Saiful Bari, Thibault Fevry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-david, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Fries, Maged Al-shaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir Radev, Mike Tian-jian Jiang, and Alexander Rush. PromptSource: An integrated development environment and repository for natural language prompts. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 93–104, Dublin, Ireland, may 2022. Association for Computational Linguistics. doi:10.18653/V1/2022.ACL-DEM0.9.
- 9 Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Ber-

riel, Thiago M Paixao, Filipe Mutz, et al. Selfdriving cars: A survey. *Expert Systems with Applications*, 165:113816, 2021. doi:10.1016/J.ESWA. 2020.113816.

- 10 Steve Baskauf, Roger Hyam, Stanley Blum, Robert A Morris, Jonathan Rees, Joel Sachs, Greg Whitbread, and John Wieczorek. Tdwg standards documentation specification. Technical report, Biodiversity Information Standards (TDWG), 2017. doi:10.3897/biss.3.35297.
- 11 Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):34-43, 2001. doi:10.1038/ scientificamerican0501-34.
- 12 Camila Bezerra, Fred Freitas, and Filipe Santana. Evaluating ontologies with competency questions. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 3, pages 284–285. IEEE, 2013. doi:10.1109/WI-IAT.2013. 199.
- 13 Christian Bizer, Tom Heath, Kingsley Idehen, and Tim Berners-Lee. Linked data on the web (ldow2008). In Proceedings of the 17th international conference on World Wide Web, pages 1265– 1266, 2008. doi:10.1145/1367497.1367760.
- 14 Vladimir Blagoderov, Ian J Kitching, Laurence Livermore, Thomas J Simonsen, and Vincent S Smith. No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys*, 209:133–146, 2012. doi:10.3897/zookeys.209. 3178.
- 15 Rishi Bommasani and et al. On the opportunities and risks of foundation models. CoRR, abs/2108.07258, 2021. arXiv:2108.07258.
- 16 Stephen Bonner, Ian P Barrett, Cheng Ye, Rowan Swiers, Ola Engkvist, Andreas Bender, Charles Tapley Hoyt, and William L Hamilton. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. Briefings in Bioinformatics, 23(6):bbac404, 2022. doi: 10.1093/BIB/BBAC404.
- 17 Grady Booch, Francesco Fabiano, Lior Horesh, Kiran Kate, Jonathan Lenchner, Nick Linck, Andreas Loreggia, Keerthiram Murgesan, Nicholas Mattei, Francesca Rossi, et al. Thinking fast and slow in ai. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 15042– 15046, 2021. doi:10.1609/AAAI.V35I17.17765.
- 18 Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense transformers for automatic knowledge graph construction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4762–4779, Florence, Italy, jul 2019. Association for Computational Linguistics. doi:10.18653/V1/ P19-1470.
- 19 Anna Breit, Laura Waltersdorfer, Fajar J Ekaputra, Marta Sabou, Andreas Ekelhart, Andreea Iana, Heiko Paulheim, Jan Portisch, Artem Revenko, Annette ten Teije, et al. Combining machine learning and semantic web: A systematic

mapping study. ACM Computing Surveys, 2023. doi:10.1145/3586163.

- 20 Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877-1901, 2020. URL: https://proceedings. neurips.cc/paper_files/paper/2020/file/ 1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.
- 21 Tiffany J. Callahan, Ignacio J. Tripodi, Harrison Pielke-Lombardo, and Lawrence E. Hunter. Knowledge-based biomedical data science. Annual Review of Biomedical Data Science, 3(1):23-41, 2020. doi: 10.1146/annurev-biodatasci-010820-091627.
- 22 Herman Cappelen and Josh Dever. Making AI intelligible: Philosophical foundations. Oxford University Press, 2021. doi:10.1093/oso/ 9780192894724.001.0001.
- 23 Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in selfsupervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pages 9650–9660, 2021. doi: 10.1109/ICCV48922.2021.00951.
- 24 André W Carus. Carnap and twentieth-century thought: Explication as enlightenment. Cambridge University Press, 2007. doi:10.1017/ cbo9780511487132.
- 25 Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Kaijie Zhu, Hao Chen, Linyi Yang, Xiaoy-uan Yi, Cunxiang Wang, Yidong Wang, et al. A survey on evaluation of large language models. arXiv preprint arXiv:2307.03109, 2023. doi: 10.48550/ARXIV.2307.03109.
- 26 Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. Journal of machine learning research, 12(ARTICLE):2493–2537, 2011. doi:10.5555/ 1953048.2078186.
- 27 Enrico Daga and Paul Groth. Data journeys: explaining ai workflows through abstraction. Semantic Web, Preprint:1-27, 2023. doi:10.3233/ sw-233407.
- 28 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018. doi: 10.48550/arXiv.1810.04805.
- 29 Chris Dijkshoorn, Mieke HR Leyssen, Archana Nottamkandath, Jasper Oosterman, Myriam C Traub, Lora Aroyo, Alessandro Bozzon, Wan J Fokkink, Geert-Jan Houben, Henrike Hovelmann, et al. Personalized nichesourcing: Acquisition of qualitative annotations from niche communities. In UMAP Workshops, 2013. URL: https://ceur-ws. org/Vol-997/patch2013_paper_13.pdf.
- 30 Filip Karlo Došilović, Mario Brčić, and Nikica Hlupić. Explainable artificial intelligence: A survey. In 2018 41st International convention on information and communication technology, electronics

and microelectronics (MIPRO), pages 0210–0215. IEEE, 2018. doi:10.23919/MIPR0.2018.8400040.

- 31 Fajar J Ekaputra, Majlinda Llugiqi, Marta Sabou, Andreas Ekelhart, Heiko Paulheim, Anna Breit, Artem Revenko, Laura Waltersdorfer, Kheir Eddine Farfar, and Sören Auer. Describing and organizing semantic web and machine learning systems in the swemls-kg. In *European Semantic* Web Conference, pages 372–389. Springer, 2023. doi:10.1007/978-3-031-33455-9_22.
- 32 Edward A Feigenbaum. The art of artificial intelligence: Themes and case studies of knowledge engineering. In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, volume 2. Boston, 1977. URL: http://ijcai.org/ Proceedings/77-2/Papers/092.pdf.
- 33 EDWARD A. FEIGENBAUM. Knowledge engineering. Annals of the New York Academy of Sciences, 426(1 Computer Cult):91-107, nov 1984. doi:10.1111/j.1749-6632.1984.tb16513.x.
- 34 Edward A Feigenbaum. A personal view of expert systems: Looking back and looking ahead. Knowledge Systems Laboratory, Department of Computer Science, Stanford ..., 1992. doi:10.1016/ 0957-4174(92)90004-c.
- 35 Dov M Gabbay and John Woods. The rise of modern logic: from Leibniz to Frege. Elsevier, 2004.
- 36 Aldo Gangemi and Valentina Presutti. Ontology design patterns. In *Handbook on ontologies*, pages 221–243. Springer, 2009. doi:10.1007/ 978-3-540-92673-3_10.
- 37 Clark Glymour, Kenneth M Ford, and Patrick J Hayes. Ramón lull and the infidels. AI Magazine, 19(2):136-136, 1998. doi:10.1609/AIMAG.V19I2. 1380.
- 38 Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. doi:10.1145/3422622.
- 39 Paul Groth, Aidan Hogan, Lise Stork, Katherine Thornton, and Vrandečić Denny. Knowledge graphs vs. other forms of knowledge representation. Dagstuhl Reports, 12(9):101–105, 2023. doi:10.4230/DagRep.12.9.60.
- 40 Qipeng Guo, Zhijing Jin, Xipeng Qiu, Weinan Zhang, David Wipf, and Zheng Zhang. Cyclegt: Unsupervised graph-to-text and text-to-graph generation via cycle training. arXiv preprint arXiv:2006.04702, 2020. doi:10.48550/arXiv. 2006.04702.
- 41 Alex R Hardisty, Elizabeth R Ellwood, Gil Nelson, Breda Zimkus, Jutta Buschbom, Wouter Addink, Richard K Rabeler, John Bates, Andrew Bentley, José AB Fortes, et al. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *BioScience*, 72(10):978–987, 2022. doi:10.1093/biosci/biac060.
- 42 Frederick Hayes-Roth, Donald A Waterman, and Douglas B Lenat. Building expert systems. Addison-Wesley Longman Publishing Co., Inc., 1983. doi:10.1017/s0263574700004069.
- 43 James Hendler, Fabien Gandon, and Dean Allemang. Semantic web for the working ontologist:

Effective modeling for linked data, RDFS, and OWL. Morgan & Claypool, 2020.

- 44 Birger Hjørland. What is knowledge organization (ko)? KO Knowledge Organization, 35(2-3):86–101, 2008. doi:10.5771/0943-7444-2008-2-3-86.
- 45 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. ACM Computing Surveys (CSUR), 54(4):1–37, 2021. doi:10.1007/ 978-3-031-01918-0.
- 46 Madelon Hulsebos, Kevin Hu, Michiel Bakker, Emanuel Zgraggen, Arvind Satyanarayan, Tim Kraska, Çagatay Demiralp, and César Hidalgo. Sherlock: A deep learning approach to semantic data type detection. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1500– 1508, 2019. doi:10.1145/3292500.3330993.
- 47 Naman Jain, Skanda Vaidyanath, Arun Iyer, Nagarajan Natarajan, Suresh Parthasarathy, Sriram Rajamani, and Rahul Sharma. Jigsaw: Large language models meet program synthesis. In Proceedings of the 44th International Conference on Software Engineering, pages 1219–1231, 2022. doi:10.1145/3510003.3510203.
- 48 Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12):1–38, 2023. doi:10. 1145/3571730.
- 49 Daniel Kahneman. Thinking, fast and slow. Macmillan, 2011.
- 50 Paschalia Kapli, Ziheng Yang, and Maximilian J Telford. Phylogenetic tree building in the genomic age. Nature Reviews Genetics, 21(7):428–444, 2020. doi:10.1038/s41576-020-0233-0.
- 51 Elisa F Kendall and Deborah L McGuinness. Ontology engineering. Morgan & Claypool Publishers, 2019. doi:10.1007/978-3-031-79486-5.
- 52 Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. Advances in neural information processing systems, 35:22199– 22213, 2022.
- 53 Keti Korini and Christian Bizer. Column type annotation using chatgpt. arXiv preprint arXiv:2306.00745, 2023. doi:10.48550/ARXIV. 2306.00745.
- 54 Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. Performance of chatgpt on usmle: Potential for aiassisted medical education using large language models. *PLoS digital health*, 2(2):e0000198, 2023. doi:10.1101/2022.12.19.22283643.
- 55 Himabindu Lakkaraju, Dylan Slack, Yuxin Chen, Chenhao Tan, and Sameer Singh. Rethinking explainability as a dialogue: A practitioner's perspective. CoRR, abs/2202.01875, 2022. doi: 10.48550/arXiv.2202.01875.
- 56 Wanhae Lee, Minki Chun, Hyeonhak Jeong, and Hyunggu Jung. Toward keyword generation

through large language models. In Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, IUI '23 Companion, pages 37–40, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/ 3581754.3584126.

- 57 Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9):1–35, 2023. doi:10.1145/3560815.
- 58 Michela Lorandi and Anya Belz. Data-to-text generation for severely under-resourced languages with gpt-3.5: A bit of help needed from google translate. arXiv preprint arXiv:2308.09957, 2023. doi:10.48550/ARXIV.2308.09957.
- 59 Arthur MacGregor. Naturalists in the field: collecting, recording and preserving the natural world from the fifteenth to the twenty-first century, volume 2. Brill, 2018.
- 60 Kyle Mahowald, Anna A Ivanova, Idan A Blank, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. Dissociating language and thought in large language models: a cognitive perspective. arXiv preprint arXiv:2301.06627, 2023. doi:10.48550/ARXIV.2301.06627.
- 61 Jose L. Martinez-Rodriguez, Aidan Hogan, and Ivan Lopez-Arevalo. Information extraction meets the semantic web: A survey. *Semantic Web*, 11(2):255–335, feb 2020. doi:10.3233/SW-180333.
- 62 Richard Menary. Writing as thinking. Language sciences, 29(5):621-632, 2007. doi:10.1016/j.langsci.2007.01.005.
- 63 Richard Menary. Dimensions of mind. Phenomenology and the Cognitive Sciences, 9:561–578, 2010. doi:10.1007/s11097-010-9186-7.
- 64 Hugo Mercier and Dan Sperber. The enigma of reason. Harvard University Press, 2017. doi: 10.4159/9780674977860.
- 65 Grégoire Mialon, Roberto Dessì, Maria Lomeli, Christoforos Nalmpantis, Ram Pasunuru, Roberta Raileanu, Baptiste Rozière, Timo Schick, Jane Dwivedi-Yu, Asli Celikyilmaz, et al. Augmented language models: a survey. arXiv preprint arXiv:2302.07842, 2023. doi:10.48550/ARXIV. 2302.07842.
- 66 Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Comput. Surv., jun 2023. Just Accepted. doi:10.1145/3605943.
- 67 Staffan Müller-Wille. Names and numbers:"data" in classical natural history, 1758–1859. Osiris, 32(1):109–128, 2017. doi:10.1086/693560.
- 68 Zara Nasar, Syed Waqar Jaffry, and Muhammad Kamran Malik. Named entity recognition and relation extraction: State-ofthe-art. ACM Computing Surveys (CSUR), 54(1):1–39, 2021. doi:10.1145/3445965.
- 69 Catarina Dutilh Novaes. Formal languages in logic: A philosophical and cognitive analysis. Cambridge University Press, 2012. doi:10.1017/ cbo9781139108010.

- 70 Catarina Dutilh Novaes. The dialogical roots of deduction: Historical, cognitive, and philosophical perspectives on reasoning. Cambridge University Press, 2020. doi:10.1017/9781108800792.
- 71 Alexandra Ortolja-Baird and Julianne Nyhan. Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive. *Digital Scholarship in the Humanities*, 37(3):844–867, 2022. doi:10.1093/LLC/FQAB065.
- 72 Roderic DM Page. Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Briefings in bioinformatics*, 9(5):345– 354, 2008. doi:10.59350/x3wmw-nws84.
- 73 Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. arXiv preprint arXiv:2306.08302, 2023. doi:10. 48550/ARXIV.2306.08302.
- 74 Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. arXiv preprint arXiv:2304.03442, 2023. doi:10.1145/3586183. 3606763.
- 75 Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? arXiv preprint arXiv:1909.01066, 2019. doi:10.18653/V1/D19-1250.
- 76 Valentina Presutti, Enrico Daga, Aldo Gangemi, and Eva Blomqvist. extreme design with content ontology design patterns. In Proc. Workshop on Ontology Patterns, pages 83–97, 2009. URL: https://ceur-ws.org/Vol-516/pap21.pdf.
- 77 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748– 8763. PMLR, 2021. URL: http://proceedings. mlr.press/v139/radford21a.html.
- 78 Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International Conference* on Machine Learning, pages 28492-28518. PMLR, 2023. URL: https://proceedings.mlr.press/ v202/radford23a.html.
- 79 Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical textconditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 1(2):3, 2022. doi:10.48550/ARXIV.2204.06125.
- 80 Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–7, 2021. doi:10.1145/3411763.3451760.
- 81 Tim Robertson, Markus Döring, Robert Guralnick, David Bloom, John Wieczorek, Kyle Braak, Javier Otegui, Laura Russell, and Peter Desmet. The gbif integrated publishing toolkit: facilitating the efficient publishing of biodiversity data

on the internet. *PloS one*, 9(8):e102623, 2014. doi:10.1371/journal.pone.0102623.

- 82 Víctor Rodríguez-Doncel and Elena Montiel-Ponsoda. Lynx: Towards a legal knowledge graph for multilingual europe. Law Context: A Socio-Legal J., 37:175, 2020. doi:10.26826/ law-in-context.v37i1.129.
- 83 Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. Highresolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684–10695, 2022. doi:10.1109/ CVPR52688.2022.01042.
- 84 Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools, 2023. doi:10.48550/ARXIV.2302.04761.
- 85 Guus Schreiber. Knowledge engineering. Foundations of Artificial Intelligence, 3:929–946, 2008. doi:10.1016/S1574-6526(07)03025-8.
- 86 Guus Schreiber, Hans Akkermans, Anjo Anjewierden, Nigel Shadbolt, Robert de Hoog, Walter Van de Velde, and Bob Wielinga. Knowledge engineering and management: the CommonKADS methodology. MIT press, 2000. doi: 10.7551/mitpress/4073.001.0001.
- 87 Guus Schreiber and Lora Aroyo. Principles for knowledge engineering on the web. In AAAI Spring Symposium: Symbiotic Relationships between Semantic Web and Knowledge Engineering, pages 78-82, 2008. URL: https://aaai.org/papers /0012-ss08-07-012-principles-for-knowledgeengineering-on-the-web/.
- 88 Nigel R Shadbolt, Paul R Smart, J Wilson, and S Sharples. Knowledge elicitation. Evaluation of human work, pages 163–200, 2015.
- 89 Murray Shanahan. Talking about large language models. arXiv preprint arXiv:2212.03551, 2022. doi:10.48550/ARXIV.2212.03551.
- 90 Steffen Staab and Rudi Studer. Handbook on ontologies. Springer Science & Business Media, 2010. doi:10.1007/978-3-540-24750-0.
- **91** Lise Stork. Knowledge extraction from archives of natural history collections. PhD thesis, Ph. D. Dissertation, Leiden University, 2021.
- 92 Lise Stork, Andreas Weber, Eulàlia Gassó Miracle, Fons Verbeek, Aske Plaat, Jaap van den Herik, and Katherine Wolstencroft. Semantic annotation of natural history collections. *Journal of Web Semantics*, 59:100462, 2019. doi:10.1016/J.WEBSEM. 2018.06.002.
- 93 Rudi Studer, V Richard Benjamins, and Dieter Fensel. Knowledge engineering: Principles and methods. Data & knowledge engineering, 25(1-2):161-197, 1998. doi:10.1016/S0169-023X(97) 00056-6.
- 94 Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. The neon methodology for ontology engineering. In Ontology engineering in a networked world, pages 9–34. Springer, 2011. doi:10.1007/ 978-3-642-24794-1_2.

3:18 Knowledge Engineering Using Large Language Models

- 95 Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. Introduction: Ontology engineering in a networked world. Springer, 2012. doi:10.1007/ 978-3-642-24794-1_1.
- 96 Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27, 2014. URL: https: //proceedings.neurips.cc/paper/2014/hash/ a14ac55a4f27472c5d894ec1c3c743d2-Abstract. html.
- 97 Anders Telenius. Biodiversity information goes public: Gbif at your service. Nordic Journal of Botany, 29(3):378–381, 2011. doi:10.1111/j. 1756-1051.2011.01167.x.
- 98 Ian Tenney, Dipanjan Das, and Ellie Pavlick. Bert rediscovers the classical nlp pipeline. arXiv preprint arXiv:1905.05950, 2019. doi:10.18653/V1/ P19-1452.
- 99 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. Artificial Intelligence, 302:103627, 2022. doi:10.1016/J.ARTINT.2021.103627.
- 100 Priyan Vaithilingam, Tianyi Zhang, and Elena L Glassman. Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *Chi conference on human factors in computing systems extended abstracts*, pages 1–7, 2022. doi:10.1145/3491101. 3519665.
- 101 Michael van Bekkum, Maaike de Boer, Frank van Harmelen, André Meyer-Vitali, and Annette ten Teije. Modular design patterns for hybrid learning and reasoning systems: a taxonomy, patterns and use cases. Applied Intelligence, 51(9):6528–6546, 2021. doi:10.1007/s10489-021-02394-3.
- 102 M.G.J. van Erp. Accessing natural history: Discoveries in data cleaning, structuring, and retrieval. PhD thesis, Tilburg University, 2010. Series: TiCC Ph.D. Series Volume: 13.
- 103 Frank Van Harmelen and Annette ten Teije. A boxology of design patterns for hybrid learning and reasoning systems. arXiv preprint arXiv:1905.12389, 2019. doi:10.48550/arXiv. 1905.12389.
- 104 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.
- 105 Chengyi Wang, Sanyuan Chen, Yu Wu, Ziqiang Zhang, Long Zhou, Shujie Liu, Zhuo Chen, Yanqing Liu, Huaming Wang, Jinyu Li, et al. Neural codec language models are zero-shot text to speech synthesizers. arXiv preprint arXiv:2301.02111, 2023. doi:10.48550/ARXIV.2301.02111.
- 106 Haohan Wang and Bhiksha Raj. On the origin of deep learning. arXiv preprint arXiv:1702.07800, 2017. doi:10.48550/arXiv.1702.07800.
- 107 Andreas Weber, Mahya Ameryan, Katherine Wolstencroft, Lise Stork, Maarten Heerlien, and Lambert Schomaker. Towards a digital infrastructure for illustrated handwritten archives. In Digital Cultural Heritage: Final Conference of

the Marie Sklodowska-Curie Initial Training Network for Digital Cultural Heritage, ITN-DCH 2017, Olimje, Slovenia, May 23–25, 2017, Revised Selected Papers, pages 155–166. Springer, 2018. doi:10.1007/978-3-319-75826-8_13.

- 108 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35:24824-24837, 2022. URL: http://papers.nips.cc/paper_files/paper/ 2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html.
- 109 Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. Toward general design principles for generative ai applications. arXiv preprint arXiv:2301.05578, 2023. doi:10.48550/ ARXIV.2301.05578.
- 110 Jules White, Sam Hays, Quchen Fu, Jesse Spencer-Smith, and Douglas C. Schmidt. Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design, 2023. doi:10.48550/ARXIV.2303.07839.
- 111 Bob J Wielinga, A Th Schreiber, and Jost A Breuker. Kads: A modelling approach to knowledge engineering. *Knowledge acquisition*, 4(1):5– 53, 1992. doi:10.1016/1042-8143(92)90013-q.
- 112 Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3(1), mar 2016. doi:10.1038/sdata.2016.18.
- 113 Lionel Wong, Gabriel Grand, Alexander K Lew, Noah D Goodman, Vikash K Mansinghka, Jacob Andreas, and Joshua B Tenenbaum. From word models to world models: Translating from natural language to the probabilistic language of thought. arXiv preprint arXiv:2306.12672, 2023. doi:10.48550/ARXIV.2306.12672.
- 114 Qianqian Xie, Jennifer Amy Bishop, Prayag Tiwari, and Sophia Ananiadou. Pre-trained language models with domain knowledge for biomedical extractive summarization. *Knowledge-Based Systems*, 252:109460, 2022. doi:10.1016/J. KNOSYS.2022.109460.
- 115 Zonglin Yang, Xinya Du, Rui Mao, Jinjie Ni, and Erik Cambria. Logical reasoning over natural lan-

guage as knowledge representation: A survey, 2023. doi:10.48550/ARXIV.2303.12023.

- 116 Kang Min Yoo, Dongju Park, Jaewook Kang, Sang-Woo Lee, and Woomyeong Park. Gpt3mix: Leveraging large-scale language models for text augmentation. arXiv preprint arXiv:2104.08826, 2021. doi:10.18653/v1/2021.findings-emnlp.192.
- 117 Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. arXiv preprint arXiv:2205.01917, 2022. doi:10.48550/ARXIV.2205.01917.
- 118 Wen Zhang, Yushan Zhu, Mingyang Chen, Yuxia Geng, Yufeng Huang, Yajing Xu, Wenting Song, and Huajun Chen. Structure pretraining and prompt tuning for knowledge graph transfer. In *Proceedings of the ACM Web Conference 2023*, WWW '23, pages 2581–2590, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3543507.3583301.
- 119 Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337-2348, 2022. doi: 10.48550/arXiv.2109.01134.

Knowledge Graph Embeddings: Open Challenges and Opportunities

Russa Biswas¹ \square \square Hasso-Plattner Institut, Potsdam, Germany

Michael Cochez 🖂 🏠 💿 Vrije Universiteit Amsterdam, The Netherlands Elsevier Discovery Lab, Netherlands

Theis E. Jendal ⊠© Aalborg University, Denmark

Vanessa Lopez 🖂 回 IBM Research Dublin, Ireland

Heiko Paulheim 🖂 🗅 Universität Mannheim, Germany

Edlira Kalemi Vakaj 🖂 🕼

Birmingham City University, UK

Lucie-Aimée Kaffee ⊠© Hasso-Plattner-Institut, Potsdam, Germany

StefaniaDumbrava 🖂 🏠 🔎 ENSIIE, France

Matteo Lissandrini 🖂 🕩 Aalborg University, Denmark

Eneldo Loza Mencía ⊠© research.eneldo.net, Frankfurt, Germany

Harald Sack 🖂 🗅 FIZ Karlsruhe, Germany Karlsruhe Institute of Technology, AIFB, Germany

Gerard de Melo ⊠© Hasso-Plattner Institut, Potsdam, Germany University of Potsdam, Germany

— Abstract –

While Knowledge Graphs (KGs) have long been used as valuable sources of structured knowledge, in recent years, KG embeddings have become a popular way of deriving numeric vector representations from them, for instance, to support knowledge graph completion and similarity search. This study surveys advances as well as open challenges and opportunities in this area. For instance, the most prominent embedding models focus primarily on structural information. However, there has been notable progress in incorporating further aspects, such as semantics, multi-modal, temporal, and mul-

tilingual features. Most embedding techniques are assessed using human-curated benchmark datasets for the task of link prediction, neglecting other important real-world KG applications. Many approaches assume a static knowledge graph and are unable to account for dynamic changes. Additionally, KG embeddings may encode data biases and lack interpretability. Overall, this study provides an overview of promising research avenues to learn improved KG embeddings that can address a more diverse range of use cases.

2012 ACM Subject Classification Computing methodologies \rightarrow Machine learning approaches; Computing methodologies \rightarrow Semantic networks

Keywords and phrases Knowledge Graphs, KG embeddings, Link prediction, KG applications

Digital Object Identifier 10.4230/TGDK.1.1.4

Category Survey

Funding Michael Cochez: Partially funded by the Graph-Massivizer project, funded by the Horizon Europe programme of the European Union (grant 101093202).

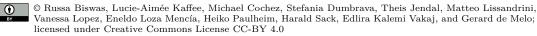
Theis E. Jendal: Supported by the Poul Due Jensen Foundation and the Independent Research Fund Denmark (DFF) under grant agreement no. DFF-8048- 00051B.

Received 2023-06-30 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

corresponding author



Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 4, pp. 4:1-4:32 Transactions on Graph Data and Knowledge



TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

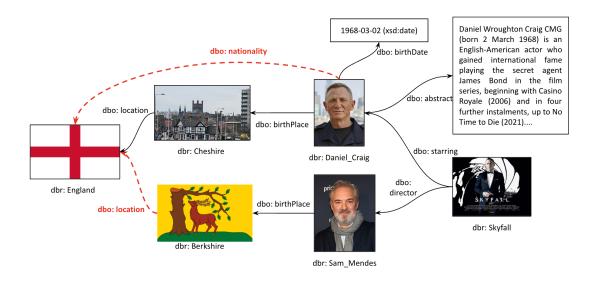


Figure 1 Excerpt from DBpedia, with red dashed lines representing possible inferred relations.

1 Introduction

A Knowledge Graph (KG) is a semantic network that organises knowledge in a graph using entities, relations, and attributes. It captures semantic relationships and connections between entities, allowing for rapid searching, reasoning, and analysis. KGs are directed labelled graphs that can represent a variety of structured knowledge across a wide range of domains including e-commerce [97, 130], media [137], and life science [24], to name a few. They enable the integration of structured knowledge from diverse sources, laying the groundwork for applications such as question-answering systems, recommender systems, semantic search, and information retrieval. Google [155], eBay [130], Amazon [97], and Uber [59] are examples of companies that have developed in-house enterprise KGs for commercial purposes, which are not publicly available. The term "Knowledge Graph" was first used in the literature in 1972 [149] and later revived by Google in 2012 with the introduction of the Google KG. Broad-coverage open KGs, such as DBpedia [11], Freebase [20], YAGO [158], and Wikidata [173], are either developed using heuristics, manually curated, or automatically or semi-automatically extracted from structured data.

While the structured knowledge in KGs can readily be used in many applications, KG embeddings open up new possibilities. A KG embedding encodes semantic information and structural relationships by representing entities and relations in a KG as dense, low-dimensional numeric vectors. This entails developing a mapping between entities and relations and vector representations that accurately capture their characteristics and relationships.

KG embeddings allow for effective computation, reasoning, and analysis while maintaining semantics and structural patterns. Link prediction and KG completion are perhaps the most well-known uses of KG embeddings. Although KGs store vast amounts of data, they are often incomplete. For instance, given the KG in Figure 1, which is an excerpt from DBpedia, it will not be possible to answer the following questions:

Q1: Where is Berkshire located?, and

Q2: What is the nationality of Daniel Craig?

Responding to Q1 requires the prediction of the missing entity in the triple <dbr:Berkshire¹, dbo:locatedIn, ?>. Similarly, for Q2, one would need to infer the nationality of Daniel Craig from the information available in the KG. The effectiveness of KG-based question-answering applications may therefore be enhanced by using embeddings to predict the missing links in a KG. This is referred to as KG completion.

Other applications of KG embeddings include similarity search, entity classification, recommender systems, semantic search, and question answering. Additionally, an embedding converts symbolic knowledge into numerical representations, making it possible to incorporate structured knowledge into machine learning and AI models, enabling reasoning across KGs.

Although prominent KG embedding models are widely used across diverse applications, there is potential to learn improved embeddings addressing an even broader range of input information and opening up new opportunities. For instance, one can account for additional signals in the KG beyond the structural information, such as multi-modal and hierarchical information, as well as external textual data, or information related to a certain domain or context. Some models struggle to adequately represent rare or long-tail entities, while others are unable to cope with little or no training data. Additionally, there is potential to design models that better account for dynamic and temporal information in the KG. Likewise, KGs are often multilingual, which may enable improved representations. Some models have trouble capturing asymmetric links as well as complex relationships such as hierarchical, compositional, or multi-hop relationships. The bias in KGs may also be reflected in the corresponding embeddings. Most models also lack explicit interpretability or explainability. This paper focuses on describing the relevant research addressing the aforementioned KG embedding models' inadequacies and then discussing the untapped areas for future research.

The rest of the paper is organised as: Section 2 gives an overview of the definitions and notations related to KGs, followed by Section 3 summarising mainstream KG embedding models. Next, Section 4 provides an overview of models that exploit additional kinds of information often neglected by traditional KG embedding models, along with a discussion of remaining open challenges. Section 5 sheds some light on important application areas of KG embeddings. Finally, Section 6 concludes the paper with a discussion and an outlook of future work.

2 Preliminaries

This section provides formal definitions and relevant notational conventions used in this paper.

▶ Definition 1 (Knowledge Graph). A KG \mathcal{G} is a labelled directed graph, which can be viewed as a set of knowledge triples $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times (\mathcal{E} \cup \mathcal{L})$, where \mathcal{E} is the set of nodes, corresponding to entities (or resources), \mathcal{R} is the set of relation types (or properties) of the entities, and \mathcal{L} is the set of literals. An entity represents a real-world object or an abstract concept. Often the labels of entities and relations are chosen to be URIs or IRIs (Internationalised Resource Identifiers).

▶ **Definition 2** (Triple). Given a KG \mathcal{G} , we call $(e_h, r, e_t) \in \mathcal{T}$ a triple, where $e_h \in \mathcal{E}$ is the subject, $r \in \mathcal{R}$ is the relation, and $e_t \in \mathcal{E} \cup \mathcal{L}$ is the object. The subject is also called the head entity, and an object $e_t \in \mathcal{E}$ may be referred to as the tail entity. Triples with literals as objects, i.e., $e_t \in \mathcal{L}$ are known as attributive triples. In this paper, we use the notation $\langle e_h, r, e_t \rangle$, with angle brackets, to indicate a triple.

¹ For example, we will often shorten the IRIs using prefixes. For example, in dbr:Berkshire, dbr: stands for http://dbpedia.org/resource/, and hence the identifier is a shorthand for http://dbpedia.org/resource/ Berkshire. Similarly, dbo: stands for http://dbpedia.org/ontology/.

4:4 Knowledge Graph Embeddings: Open Challenges and Opportunities

Relations (or Properties). Depending on the nature of the objects in a triple, one may distinguish two main kinds of relations:

- Object Relation (or Property), in which an entity is linked to another entity. For instance, in the triple <dbr:Daniel_Craig, dbo:birthPlace, dbr:Cheshire>, dbr:Daniel_Craig and dbr:Cheshire are head and tail entities, respectively, and dbo:birthPlace is an Object Relation (or Property).
- Data Type Relation (or Property), in which the entity is linked to a literal. For instance, we find the date "1868-03-02" in the triple < dbr:Daniel_Craig, dbo:birthDate, "1868-03-02">, and therefore the relation dbo:birthDate is a Data Type Relation (or Property).

Additionally, an entity e can also be linked to classes or semantic types of the entity. For example, DBpedia uses rdf:type as r, while Freebase uses isA. A triple of the form $\langle e, rdf:type, C_k \rangle$ hence implies that $e \in \mathcal{E}$ is an entity, $C_k \in \mathcal{C}$ is a class, \mathcal{C} is the set of semantic types or classes, and e is an instance of C_k . Often, the semantic types or the classes in a KG are organised in a hierarchical tree structure. An entity may belong to more than one class.

Literals. A KG can have many types of literal values and examples of common attribute types are as follows:

- *Text literals*: These store information in the form of free natural language text and are often used for labels, entity descriptions, comments, titles, and so on.
- Numeric literals: Dates, population sizes, and other data saved as integers, real numbers, etc. provide valuable information about an entity in a KG.
- Image literals: These literals can, for example, be used to store a visual representation of the entity, but can also contain the outcome of a medical scan, or a chart.

It is also possible that there is additional information (such as video or audio) stored external to the graph. The graph can then contain an IRI or other kind of identifier that references the external resource, its location, or both.

3 Knowledge Graph Embeddings

KG embedding models represent entities and relationships in a KG in a low-dimensional vector space for various downstream applications. A typical KG embedding model is characterised by the following aspects, as detailed by Ji et al. [83]: (1) The *Representation Space* may be a single standard Euclidean vector space, separate Euclidean vector spaces for entities and relations, or matrices, tensors, multivariate Gaussian distributions, or mixtures of Gaussians. Some methods also use complex vectors or hyperbolic space to better account for the properties of relationships. (2) A *scoring function* serves to represent relationships by quantifying the plausibility of triples in the KG, with higher scores for true triples and lower scores for false/negative/corrupted ones. (3) *Encoding models* are responsible for learning the representations by capturing relational interactions between entities. This is typically achieved by solving optimisation problems, often using factorisation approaches or neural networks. (4) *Auxiliary Information* in the KG may be incorporated, e.g., literals. This leads to enriched entity embeddings and relations, forming an ad-hoc scoring function integrated into the general scoring function.

An overview of different types of KG embedding models is given in Table 1. In the following, we explain each of these in more detail.

Translation-based models use distance-based scoring functions to measure the plausibility of a fact as the distance between two entities. There are numerous variants. TransE [22] represents entities and relations as vectors in the same space, while TransH [184] introduces relation-specific hyperplanes. TransR [114] uses relation-specific spaces but requires a projection matrix for each relation. TransD [81] simplifies TransR by using two vectors for each entity-relation pair.

Categories	Models
Translational Models	TransE [21] and its variants, RotatE [160], etc,
Gaussian Embeddings	KG2E [67], TransG [192]
Semantic Matching Models	RESCAL [124] and its extensions, DistMult [198], HoIE [123], SME [22]
Neural Network Models	NTN [156], HypER [14], ConvE [38], ConvKB [32]
Graph Neural Networks	GCN [93], R-GCN [148], GraphSAGE [61], GAT [172], KGAT [179], ComplEx-KG [170], SimlE [91]
Path-based Models	GAKE [44], PTransE [113], RSN, PConvKB [84], RDF2vec [141]

Table 1 Categorisation of classic Knowledge Graph Embedding Models.

TranSparse [82] employs two separate models, TranSparse(share) and TranSparse(separate), to modify projection vectors or matrices without considering other aspects. TransA [85] replaces the traditional Euclidean distance with the Mahalanobis distance, demonstrating better adaptability and flexibility as an indicator for performance improvement.

- **Gaussian Embeddings**: KG2E [67] and TransG [192] are probabilistic embedding models that incorporate uncertainty into their representation. KG2E uses multi-Gaussian distributions to embed entities and relations, representing the mean and covariance of each entity or relation in a semantic feature space. TransG, in contrast, uses a Gaussian mixture model to represent relations, addressing multiple relationship semantics and incorporating uncertainty. Both models offer unique approaches to representing entities and relations.
- **Semantic Matching** models rely on the notion of semantic similarity to define their scoring function. These include tensor decomposition models such as RESCAL, a tensor factorisation model that represents entities and relations as latent factors [124], capturing complex interactions between them. DistMult [199] simplifies the scoring function of RESCAL by using diagonal matrices, leading to more efficient computations. SimplIE [91] is a simpler model that uses a rule-based approach to extract relations from sentences. RotatE [161] introduces rotational transformations to model complex relationships in KGs. ComplEx [170] extends DistMult by introducing complex-valued embeddings, enabling it to capture both symmetric and antisymmetric relations. HolE [124] employs circular correlation to capture compositional patterns in KGs. TuckER [15] is a linear model based on Tucker decomposition of the binary tensor representation of triples.
- Neural network based models draw on the powerful representation learning abilities of modern deep learning. Neural Tensor Networks (NTN) [156] allow mediated interaction of entity vectors via a tensor. ConvE [38] uses 2D convolutions over embeddings to predict missing links in KGs. ConvKB [32] represents each triple as a 3-column matrix and applies convolution filters to generate multiple feature maps, which are concatenated into a single feature vector. This vector is multiplied with a weight vector to produce a score, used for predicting the validity of the triple. HypER [14] generates convolutional filter weights for each relation using a hyper-network approach.
- **Graph Neural Network** models are neural networks that operate directly on the graph structure, often with information propagation along edges. GCN [93] and GraphSAGE [61] are graph convolutional techniques that combine information from neighbouring nodes in a graph to enable efficient learning of node representations in large-scale graphs. R-GCN [148] extends GCN to handle different relationships between entities in graph-structured data using a CNN model to learn hidden layer representations that encode local network structure and node attributes, growing linearly with the number of graph edges. GAT [172] employs an

4:6 Knowledge Graph Embeddings: Open Challenges and Opportunities

attention mechanism to dynamically allocate weights to neighbouring nodes, focusing on salient neighbours and capturing expressive representations. KGAT [179] applies the concept of graph attention networks to KG embeddings, taking into account entity and relation information, as well as capturing complicated semantic linkages and structural patterns. ComplEx-KG [170] is a complex-valued embedding-based extension of ComplEx, a bilinear model for KG embeddings. SimplE [91] uses a scoring function for large KGs that is scalable and optimised for efficiency.

Path-based models such as PTransE [113] represents entities and relations in the KG as vectors and learn embeddings based on relation-specific translation operations along edge paths. RSN [204] models the KG as a recursive structure, aggregating embeddings of connected entities and capturing structural information through recursive path-based reasoning. PConvKB [84] extends the ConvKB model and uses an attention mechanism on the paths to measure the local importance in relation paths. GAKE [44] is a graph-aware embedding model that takes into consideration three forms of graph structure: neighbour context, path context, and edge context. RDF2Vec [141] uses random walks over the graph structure to generate node and edge sequences, which are then used as input for training word2vec skip-gram models, which yield entity and relation embeddings.

Traditional KG embedding methods primarily take into account the triple information but neglect other potentially valuable signals encountered in KGs, such as multimodality, temporality, multilinguality, and many more. Additionally, these models often assume KGs are static in nature and have cold-start problems when incorporating new entities and relations. Also, real-world KGs often exhibit sparsity, noisiness, and bias, which may adversely affect embedding models.

4 Opportunities and Challenges

KG embeddings are widely used to capture semantic meaning and enable improved comprehension, reasoning, and decision-making across a diverse range of applications. However, the traditional KG embedding models described earlier neglect a series of important opportunities and aspects. In the following, in Section 4.1, we consider auxiliary information that may be present in KGs but is often neglected in KG embeddings, e.g., multimodal, multilingual, and dynamic knowledge. Subsequently, in Section 4.2, we discuss further more general issues, such as bias and explainability. Recent research has made notable progress in addressing these issues. The remainder of the section summarises pertinent recent research along with a discussion of open research challenges.

4.1 Auxiliary Information

Prominent KG embedding models such as those enumerated in Section 3 focus primarily on the structure of the KG, i.e., on structural information pertaining to entities and their relationships. To improve the latent representations of entities and relations, new lines of research attempt to draw on additional forms of information present in the KG. This section offers an overview of existing research in this regard, along with discussions of relevant shortcomings and recommendations for further research.

4.1.1 Multimodal KG Embeddings

Many approaches for representation learning on entities and relations ignore the variety of data modalities in KGs. In a Multimodal KG (MKG), entities and attributes of these entities may have different modalities, each providing additional information about the entity. An effective learned representation captures correspondences between modalities for accurate predictions, as described by Gesese et al. [54]. The used modalities depend on the application area but can include text,

images, numerical, and categorical values. Inductive approaches are required for modelling MKGs that encompass a variety of data modalities, as assuming that all entities have been observed during training is impractical. Learning a distinct vector for each entity and using enumeration for all possible attribute multimodal values to predict links is usually infeasible.

- **Text:** One of the early approaches for text extends TransE by incorporating word2vec SkipGram and training a probabilistic version in the same embedding space, anchoring via Freebase entities and the word embedding model vocabulary [183]. This enables link prediction for previously unknown entities. Relations are treated without differentiation of types. A combination of DistMult and CNN [169] tackles this issue by modelling the textual relations via dependency paths extracted from the text. Other models such as DKRL [194] and Jointly (BOW) [196] use the word2vec Continuous Bag-Of-Words (CBOW) approach to encode keywords extracted from textual entity descriptions, while Text Literals in KGloVe [31] uses these in combination with the graph context to train a GloVe model. However, the alignment between KG and word model is achieved using string matching and therefore struggles with ambiguous entity names. Veira et al. [171] use Wikipedia articles to construct relation-specific weighted word vectors (WWV). Convolutional models, such as DKRL (CNN) [194] and RTKRL [66], use word order to represent relations, considering implicit relationships between entities. Multi-source Knowledge Representation Learning (MKRL) [164] uses position embedding and attention in CNNs to find the most important textual relations among entity pairs. STKRL [188] extracts reference sentences for each entity and treats the entity representation as a multi-instance learning model. Recurrent neural models such as Entity Descriptions-Guided Embedding (EDGE) [178] and Jointly (ALSTM) [196] use attention-based LSTMs with a gating mechanism to encode entity descriptions, capturing long-term relational dependencies. The LLM encoder BERT is used in Pretrain-KGE [213] to generate initial entity embeddings from entity descriptions and relations, and subsequently feed them into KG embedding models for final embeddings. Other research uses LLMs [17, 181, 120, 3] to produce representations at the word, sentence, and document levels, merging them with graph structure embeddings. KG-BERT [200] optimises the BERT model on KGs, followed by KG-GPT2 [18] fine-tuning the GPT-2 model. MTL-KGC [92] enhances the effectiveness of KG-BERT by combining prediction and relevance ranking tasks. Saxena et al. [147] similarly transform the link prediction task into a sequence-to-sequence problem by verbalising triplets into questions and answers, overcoming the scalability issues of KG-BERT. Masked Language Modelling (MLM) has been introduced to encode KG text, with MEMKGC [29] predicting masked entities using the MEM classification model. StAR [174] uses bi-encoder-style textual encoders for text along with a scoring module, while SimKGC leverages bi-encoding for the textual encoder. LP-BERT [105] is a hybrid method that combines MLM Encoding for pre-training with LLM and Separated Encoding for fine-tuning.
- **Numeric literals** are addressed by several prominent models. MT-KGNN [166] trains a relational network for triple classification and an attribute network for attribute value regression, focusing on data properties with non-discrete literal values. KBLRN [51] combines relational, latent, and numerical features using a probabilistic PoE method. LiteralE [98] incorporates literals into existing latent feature models for link prediction, modifying the scoring function and using a learnable transformation function. TransEA [190] has two component models: a new attribute embedding model and a translation-based structure embedding model, TransE. These embedding approaches, however, fail to fully comprehend the semantics behind literal and unit data types. Also, most models lack proper mechanisms to handle multi-valued literals.
- Image and Video models account for multimedia content. There is a large body of work on visual relationship detection, i.e., identifying triples portrayed in visual content, using datasets such as VisualGenome [96] and methods such as VTransE [208]. IKLR [193] enriches

4:8 Knowledge Graph Embeddings: Open Challenges and Opportunities

KG embeddings by retrieving images for each entity from ImageNet. The respective set of pre-trained image embeddings is subsequently combined by an attention-based multi-instance learning method into a joint representation space of entities and relations. This additionally enables identifying the most relevant images for each entity.

General multi-modal KG embedding models may be used both for better link prediction between existing entities and to impute missing values. One approach [128] combines different neural encoders to learn embeddings of entities and multimodal evidence types used to predict links. Then, DistMult or ConvE is employed to produce a score reflecting the probability that a triple is correct. In addition, neural decoders are applied over the learned embeddings to generate missing multimodal attributes, such as numerical values, text and images, from the information in the KG. Moreover, decoders can be invoked to generate entity names, descriptions, and images for previously unknown entities. A blueprint for multimodal learning from KGs is introduced by Ektefaie et al. [41]. Graph methods are employed to combine different datasets and modalities while leveraging cross-modal dependencies through geometric relationships. Graph Neural Networks (GNN) are used to capture interactions in multimodal graphs and learn a representation of the nodes, edges, subgraphs, or entity graph, based on message-passing strategies. Multimodal graphs find increasing application not only in computer vision and language modelling but also in natural sciences and biomedical networks [106], as well as in physics-informed GNNs that integrate multimodal data with mathematical models [154].

Limitations. Some of the key challenges reported in the literature that require further attention include: (1) Utilising multimodal information and multimodal fusion (from two or more modalities) to perform a prediction (e.g., classification, regression, or link prediction), even in the presence of missing modalities [128, 101, 41, 34]. (2) Modality collapse, that is when only a subset of the most helpful modalities dominates the training process. The model may overly rely on that subset of modalities and disregard information from the others that may be informative. This can be due to an imbalance in the learning process or insufficient data for one or more modalities and it can lead to sub-optimal representations [41]. (3) Generalisation across domains, modalities. and transfer learning of embeddings across different downstream tasks. In general, there is a high variance in the performance of multimodal methods [128, 110]. (4) Developing multimodal imputation models that are capable of generating missing multimodal values. While research in MKGs has predominantly focused on language (text) and vision (images) modalities, there is a need to explore multimodal research in other modalities and domains as well [128]. (5) Robustness to noise and controlling the flow of information within MKGs from more accurate predictions. While multimodal triples provide more information, not all parts of this additional data are necessarily informative for all prediction downstream tasks [101, 71, 128]. (6) Efficient and scalable frameworks that can handle the complexity during training and inference [34, 110]. Large KGs are challenging for all embedding-based link prediction techniques, and multimodal embeddings are not significantly worse because they can be viewed as having additional triples. However, multimodal encoders/decoders are more expensive to train [128] and techniques for batching and sampling are usually required for training. By addressing these challenges, we can unlock the full potential of MKGs and advance our understanding in various domains.

4.1.2 Schema/Ontology Insertion in KG Embeddings

While many real-world KGs come with schemas and ontologies, which may be rich and expressive, this does not hold for many of the benchmark datasets used in the evaluation of KG embeddings, in particular in the link prediction field. Therefore, the use of ontological knowledge for improving embeddings has drawn comparatively little attention.

In a very recent survey [209], the authors have reviewed approaches that combine ontological knowledge with KG embeddings. The authors distinguish between *pre* methods (methods applied before training the embedding), *joint* (during training of the embedding), and *post* (after training the embedding) methods. In their survey, joint methods are the most common approaches, usually incorporating the ontological knowledge in the loss function [10, 26, 40, 39, 52, 57, 99, 113, 143, 194, 206]. In such approaches, loss functions of existing KGE models are typically altered in a way such that ontologically non-compliant predictions are penalised. This is in line with a recent proposal of evaluation functions that not only take into account the ranking of correct triples but also the ontological compliance of predictions [75]. Some approaches also foresee the parallel training of class encoders [194] or class embeddings [65] to optimise the entity embeddings.

Pre methods observed in the literature come in two flavours. The first family of approaches exploit ontologies by inferring implicit knowledge in a preprocessing step and embedding the resulting graph enriched with inferred knowledge [76, 143]. The second family of approaches exploits ontologies in the process of sampling negative triples, implementing a sampling strategy that has a higher tendency to create ontologically compliant (and thus harder) negative examples [10, 58, 78, 99, 194], or builds upon adversarial training setups [116].

The *post* methods in the aforementioned survey are actually modifications of the downstream task, not the embedding method, and thus do not affect the embedding method per se.

The fact that most approaches fall into the *joint* category also limits them by being bound to one single embedding model, instead of being universally applicable. At the same time, most approaches have a very limited set of schema or ontology constraints they support (e.g., only domains and ranges of relations), while general approaches that are able to deal with the full spectrum of ontological definitions, or even more complex expressions such as SHACL constraints, remain very rare.

4.1.3 Relation Prediction Models

Relation prediction in KGs is a fundamental task that involves predicting missing or unobserved relations (properties) between entities in a KG. For instance, in Figure 1, relation prediction aims to predict the relation *dbo:starring* between entities *dbr:Daniel_Craig* and *dbr:Skyfall*.

Some of the classical KG embedding models such as translational models, and semantic matching models are often also used to predict missing relations. However, one of the pioneer models that focused on improving the relation prediction task is ProjE [153]. The model projected entity candidates onto a target vector representing input data, using a learnable combination operator to avoid transformation matrices followed by an optimised ranking loss of candidate entities. CNN-based models, in contrast, are argued to obtain richer and more expressive feature embeddings compared to traditional approaches. Attention-based embeddings enhance this approach further by capturing both entity and relation features in any given context or multihop neighbourhood [118]. Prior research on relation prediction, which was restricted to encyclopaedic KGs alone, disregarded the rich semantic information offered by lexical KGs, which resulted in the issue of shallow understanding and coarse-grained analysis for knowledge acquisition. HARP [182] extends earlier work by proposing a hierarchical attention module that integrates multiple semantic signals, combining structured semantics from encyclopaedic KGs and concept semantics from lexical KGs to improve relation prediction accuracy.

Self-supervised training objectives for multi-relational graph representation have also given promising results. This may be achieved using a simplistic approach by incorporating relation prediction into the commonly used 1-vs-All objective [28]. The previously mentioned path-based embedding models may also be used, but often overlook sequential information or limited-length entity paths, leading to the potential loss of crucial information. GGAE [107] is a novel global

4:10 Knowledge Graph Embeddings: Open Challenges and Opportunities

graph attention embedding network model that incorporates long-distance information from multi-hop paths and sequential path information for relation prediction. The effectiveness of KG embedding models for relation prediction is typically assessed using rank-based metrics, which evaluate the ability of models to give high scores to ground-truth entities.

Limitations. Although embedding-based models for relation prediction in KGs have advanced significantly, they have several shortcomings. (1) Most of the models struggle to capture transitivity, which is essential for understanding relations that change over time or apply in different contexts. (2) They also struggle to handle rare relations, which can result in biased predictions. (3) Although embedding techniques are intended to accommodate multi-relational data, capturing complex interactions between numerous relations remains challenging. (4) KGs can contain relations with different semantic heterogeneity. For example, imagine a KG with a relation called *hasPartner* that represents any type of close partnership, such as business partners or friends. This relationship is semantically different from *hasSpouse*. Relation prediction models are often unable to distinguish between such relations with related but different meanings. (5) Relation prediction models provide limited support for temporal and contextual information. Temporal information, however, is handled by the temporal KG embedding models presented in Section 4.1.5.

4.1.4 Hierarchical and *N*-to-*M* Modelling in KG Embeddings

Crucial to the success of using KG embeddings for link prediction is their ability to model relation connectivity patterns, such as symmetry, inversion, and composition. However, many existing models make deterministic predictions for a given entity and relation and hence struggle to adequately model N-to-M relationships, where a given entity can stand in the same relationship to many other entities, as for instance for the *hasFriend* relationship [121].

A particularly important case is that of hierarchical patterns, which, albeit ubiquitous, still pose significant challenges. Indeed, modelling them with knowledge embeddings often requires additional information regarding the hierarchical typing structure of the data [194] or custom techniques [212, 211], as discussed next.

Various approaches have been proposed for modelling hierarchical structures. Li et al. [108] proposes a joint embedding of entities and categories into a semantic space, by integrating structured knowledge and taxonomy hierarchies from large-scale knowledge bases, as well as a Hierarchical Category Embedding (HCE) model for hierarchical classification. This model additionally incorporates the ancestor categories of the target entity when predicting context entities, to capture the semantics of hierarchical concept category structures.

Another method used for hierarchical modelling centres around the usage of clustering algorithms [212]. The authors define a three-layer hierarchical relation structure (HRS) for KG relation clusters, relations, and subrelations. Based on this, they extend classic translational embedding models to learn better knowledge representations. Their model defines the embedding of a knowledge triple based on the sum of the embedding vectors for each of the HRS layers.

The Type-embodied Knowledge Representation Learning (TKRL) [194] model uses entity-type information in KG embeddings to model hierarchical relations. Following the TransE approach, relations are translated between head and tail KG entities in the embedding space. For each entity type, type-specific projection matrices are built using custom hierarchical type encoders, projecting the heads and tails of entities into their type spaces.

Limitations. Although they intend to better represent the structure of a KG, the limitations of such KG embeddings include: (1) It is challenging to model interactions that transcend numerous hierarchy levels, resulting in a limited ability to capture cross-hierarchy linkages. For instance,

4:11

Arnold Schwarzenegger is an actor, a film director as well as a politician, leading to the entity belonging to different branches of the class hierarchy in the KG. (2) The depth of the hierarchy or branching factor of an n-to-m relationship can affect how effective the embeddings are, e.g., in very fine-grained or coarse-grained hierarchies, performance may suffer. (3) Training and inference with hierarchical embeddings can be computationally intensive, particularly in ultrafine-grained hierarchies.

4.1.5 Temporal KG Embeddings

Most KG completion methods assume KGs to be static, which can lead to inaccurate prediction results due to the constant change of facts over time. For instance, neglecting the fact that $\langle Barack Obama, president Of, USA \rangle$ only holds from 2009 to 2017 can become crucial for KG completion. Emerging approaches for Temporal Knowledge Graph Completion (TKGC) incorporate timestamps into facts to improve the result prediction. These methods consider the dynamic evolution of KGs by adding timestamps to convert triples into quadruples using several strategies [23]:

- **Tensor Decomposition** based models in KG completion transform a KG into a 3-dimensional binary tensor, with three modes representing head, relation, and tail entities to learn their corresponding representations by tensor decomposition. The addition of timestamps as an additional mode of tensor (4-way tensor) for TKGC allows for low-dimensional representations of timestamps for scoring functions. For TKGC, Canonical Polyadic (CP) decomposition is used on quadruple facts [112]. The authors employ an imaginary timestamp for static facts, while complex-valued representation vectors may be used for asymmetric relations [100]. Temporal smoothness penalties are used to ensure that neighbouring timestamps obtain similar representations. Multivector representations [195] are learned using CP decomposition, allowing the model to adjust to both point timestamps and intervals. A temporal smoothness penalty for timestamps is created and expanded to a more generic autoregressive model. Tucker decomposition can be used for TKGC [151], treating KGs as 4-way tensors and scoring functions that consider interactions among entities, relations, and timestamps, relaxing the requirement for identical embedding dimensions of entities, relations, and timestamps.
- Timestamp-based Transformation models involve generating synthetic time-dependent relations by concatenating relations with timestamps (e.g., president Of: 2009-2017), converting <Barack Obama, presidentOf, USA> to <Barack Obama, presidentOf:2009-2017, USA> [102]. This however may lead to more synthetic relations than necessary. An improvement is to derive optimal timestamps for concatenating relations by splitting or merging existing time intervals [135]. The concatenation of relation and timestamp as a sequence of tokens is also provided as an input making the synthetic relation adaptive to different formats like points, intervals, or modifiers [50]. Others [177] argue that different relations rely on different time resolutions, such as a life span in years or a birth date in days. Multi-head self-attention is adopted on the timestamp-relation sequence to achieve adaptive time resolution. In the TKGC model, timestamps are often considered linear transformations that map entities/relations to corresponding representations. The timestamps are also treated as hyperplanes, dividing time into discrete time zones [33]. An additional relational matrix is included to map entities to be relation-specific to improve expressiveness for multi-relational facts [185]. To capture dynamics between hyperplanes, a GRU may be applied to the sequence of hyperplanes [163]. Another approach [103] encodes timestamps into a one-hot vector representing various time resolutions, such as centuries or days to achieve time precision.
- **KG Snapshots** can be considered as a series of snapshots/subgraphs taken from a KG, with each subgraph holding facts labelled with a timestamp. Therefore, a temporal subgraph evolves with changing relation connections. The link prediction problem can be solved

4:12 Knowledge Graph Embeddings: Open Challenges and Opportunities

by utilising Markov models [197] to infer the multi-relational interactions among entities and relations over time and can be trained using a recursive model. Probabilistic entity representations based on variational Bayesian inference can be adopted to model entity features and uncertainty jointly [111]. The dynamic evolution of facts can be modelled using an autoregressive approach [86], incorporating local multi-hop neighbouring information and a multi-relational graph aggregator. Alternatively, a multilayer GCN can capture dependencies between concurrent facts with gated components to learn long-term temporal patterns [109]. Continuous-time embeddings can encode temporal and structural data from historical KG snapshots [64].

Historical Context based models focus on the chronological order of facts in a KG, determined by the availability of timestamps, which enable predicting missing links by reasoning with the historical context of the query. An attention-based reasoning process has been proposed [63] as the expansion of a query-dependent inference subgraph, which iteratively expands by sampling neighbouring historical facts. Another approach uses path-based multi-hop reasoning by propagating attention using a two-stage GNN through the edges of the KG, using the inferred attention distribution [87]. The model captures displacements at two different granularities, i.e., past, present, and future and the magnitude of the displacement. Two heuristic-based tendency scores Goodness and Closeness [12] have been introduced to organise historical facts for link prediction. Historical facts are aggregated based on these scores, followed by a GRU for dynamic reasoning. It is observed that history often repeats itself in KGs [214], leading to the proposal of two modes of inference: Copy and Generation.

Limitations. Although recently many TKGC models have been proposed that resolve the issues of classical KG embedding models with timestamps, some intriguing possibilities for future studies on TKGC include: (1) External knowledge such as relational domain knowledge, entity types, and semantics of entities and relationships can be added to the limited structural/temporal information during model learning to enhance prediction accuracy. (2) Due to the time dimension and intricate relationships between facts and timestamps, time-aware negative sampling should be investigated in TKGC. (3) Most methods assume timestamps are available, while in some cases only relative time information is known. For example, we would know that a person lived in a city after they were born, but neither when the person was born, nor when they started living there. (4) With the constant evolution of the real-world KGs, TKGC should be regarded as an incremental or continual learning problem.

4.1.6 Dynamic KG Embeddings

As discussed in the previous section, incorporating timestamps is one way to handle changes; however, facts may be added, altered, or deleted over time, are not foreseen [95], and would typically require a complete re-computation of the embedding model. Such an approach might still be feasible for KGs like DBpedia, which have release cycles of weeks or months [70], but not for continuously updated KGs such as Wikidata, let alone examples of even more highly dynamic KGs, e.g., digital twins, which may continuously change every second. Moreover, naïvely recomputing embeddings for an only slightly changed KG may lead to drastic shifts in the embeddings of existing entities, e.g., due to stochastic training behaviour. This would require a recalibration of downstream models consuming those embeddings, as they would not be *stable* [187, 94].

While a few approaches for embedding dynamic graphs (not necessarily KGs) have been proposed [90], many of them focus on embedding a series of snapshots of KGs, rather than developing mechanisms for embedding a dynamic KG. Thus, they do not support *online learning*, i.e., continuously adjusting the KG embedding model whenever changes occur.

Approaches capable of online learning are much scarcer. One of the first was puTransE [165], which continuously learns new embedding spaces. Similarly, Wewer et al. [187] investigate updating the link prediction model by incorporating change-specific epochs forcing the model to update the embeddings related to added or removed entities and/or relations.

Embeddings based on random walks can be adapted to changes in the graph by extracting new walks around the changed areas [115], or by applying local changes to the corpus of random walks [146]. The latter approach also supports the deletion of nodes and edges. DKGE [189] learns embeddings using gated graph neural networks and requires retraining only vectors of affected entities in the online learning part. Similarly, OUKE first learns static embeddings and computes dynamic representations only locally using graph neural networks. The two representations are then combined into a dynamic embedding vector. The idea of only updating embeddings of affected entities is also pursued by RotatH [186]. A different strategy is considered by Navi [94], which learns a surrogate model to reconstruct the entity embeddings based on those of neighbouring existing entities. This surrogate model is then used to recompute the embedding vectors for new entities or entities with changed contexts.

Limitations. The main limitations in the existing approaches so far are threefold: (1) In most models, only addition to KGs is studied, while deletion is not the focus, an exception is the work by Wewer et al. [187].² (2) The stability of the resulting embeddings, which is crucial for downstream applications, has rarely been analysed systematically. (3) The applicability in a true real-time scenario, as it would be required, e.g., for digital twins, is unclear for most approaches, which are evaluated on snapshots.

4.1.7 Inductive KG Embedding

In the inductive setting, graph representation learning involves training and inference of partially or completely disjoint sets of nodes, edges, and possibly even relationships types. In practice, from the specific set of known structures, it tries to generalise knowledge that enables reasoning with unseen graph objects by exploiting information on the structures involving them and the data attached to them [47]. The case of link prediction involves being able to predict the existence of a link between two previously unseen nodes (head and tail) by reasoning about their connections to other known nodes (i.e., nodes observed during training) or by reasoning about their attributes (e.g., features similar to those of nodes seen during training).

Therefore, in the most common setting, relationship types do not change, but training involves a given KG and inference involves a completely or partially different graph. Overall, the crucial point is that there must be some form of shared information that allows for *inferring* a description of an unknown entity or edge from a small set of known attributes. For example, a common approach allows for predictions involving previously unseen, or out-of-sample, entities that attach to a known KG with a few edges adopting known relationship types [48]. In this case, a few nodes in the KG seen during training are used as anchors and called NodePieces. A full NodePiece vocabulary is then constructed from anchor nodes and relation types. Given a new node, an embedding representation is obtained using elements of the constructed NodePiece vocabulary extracting a *hash code* for it given by the sequence of k closest anchors, combined with discrete anchor distances, and a relational context connecting relations. Other approaches extract a local subgraph of one or more nodes and consider the structures within such a subgraph trying to learn

 $^{^2}$ Even for papers using different versions of public KGs e.g., DBpedia or YAGO, the majority of changes are additions, and most benchmarks used in the evaluation of the papers mentioned above, usually have much more additions than deletions.

4:14 Knowledge Graph Embeddings: Open Challenges and Opportunities

an inductive bias able to infer entity-independent relational semantics [167]. This approach is then also adopted to predict missing facts in KGs, i.e., to predict a missing relation between two entities. Similarly, NBFNet [215] instead encodes the representation of a pair of nodes using the generalised sum of all path representations between the two nodes and with each path representation as the generalised product of the edge representations in the path. In this case, the operation is modelled along the line of a generalised Bellman-Ford algorithm that computes the shortest paths from a single source vertex to all of the other vertices by taking into account edge weights. Here, operators to compute the length of the shortest path are learned for a specific downstream task.

The aforementioned methods are designed for the case where the only information available are triples connecting entities and do not take into account node or edge properties. Conversely, when properties are taken into account, e.g., textual data describing entities, this information can be exploited as node or edge features. A typical case is that of networks that adopt an auto-encoder architecture to encode node representations and decode edges as a function over the representation of node pairs. Among those, GraphSAGE [61] was the first inductive GNN able to efficiently generate embeddings for unseen nodes by leveraging node features, e.g., textual attributes. Later methods, including BLP [36] create embeddings for entities by encoding the description with a language model fine-tuned on a link prediction objective. This model can then be used inductively, as long as nodes have a description.

Limitations. All these approaches have only scratched the surface of the need for KG embeddings. In particular, challenges persist in terms of (1) scalability, e.g., the possibility of learning inductive biases from small representative samples of the graph: (2) exploiting well-known feature extraction from graphs and KGs, as existing methods tend to disregard the possibility of using structural features, e.g., betweenness, page rank, relational neighbourhood and characteristic sets [122]; (3) moreover, while GNNs seem the most promising and expressive architecture, their ability to produce inductive relation aware KG representations are limited in their treatment of rich vocabularies of relation types (typically limited to fewer than a hundred), their ability to exploit information at more than 3 hops of distance, and the possibility to generate a representation for very sparse feature sets. Finally, known challenges that apply to transductive methods, e.g., distribution shift and how to update the model or decide to train it from scratch, still apply. Finally, the ability to work in an inductive fashion might increase the risk of data leakages, which already exist in non-inductive settings [42]. The use of GNNs that learn how to aggregate information from node and edge attributes raises more concerns when the training data involves private data; how to ensure that private data is not leaked through the model, e.g., via differentially private KG embedding [62], is still an open question.

4.1.8 Multilingual KG Embeddings

Providing multilingual information in a KG is crucial to ensure wide adoption across different language communities [88]. Languages in KGs can have different representations; e.g., in Wikidata, each entity has a language-independent identifier, and labels in different languages are indicated with the rdfs:label property [89]. Therefore, in Wikidata, entities do not need alignment across languages. In DBpedia, there is one entity per language, derived from the respective language Wikipedia [104]. Therefore, different language entities on the same concept can have different facts stated about them. Here, an alignment using the owl:sameAs property is necessary to ensure the different entities are connected across languages and enable seamless access to information for all language communities. The different representations of languages in the different KGs can heavily influence which way the KG can be embedded. For example, if provided with a KG per language as in DBpedia, different language KGs might be embedded separately and then aligned or can be fused for usage in downstream applications [74].

One of the downstream tasks of multilingual KG embeddings is KG completion. Finding new facts given machine-readable data such as a KG is a tedious task for human annotators, even more so when the graph covers a wide range of languages. Addressing these challenges, recent work has employed KG embeddings across languages to predict new facts in a KG.

One of the large challenges of multilingual KG embeddings is the knowledge inconsistency across languages, i.e., the vastly different number of facts per language. Fusing different languages to overcome such knowledge inconsistencies for multilingual KG completion can improve performance across languages, especially for lower-resourced languages [74]. To fuse different languages, KGs need to be aligned across languages. Such alignment can be done jointly with the task of multilingual KG completion [25, 168, 27].

Another approach for multilingual KG completion is leveraging large language models' (LLM) knowledge about the world to add new facts to a KG. As LLMs are not trained towards KG completion and are biased towards English, Song et al. [157] introduce global and local knowledge constraints to constrain the reasoning of answer entities and to enhance the representation of query context. Hence, the LLMs are better adapted for the task of multilingual KG completion.

Limitations. Although most of the existing multilingual KG embedding models focus on having a unified embedding space across different language versions of the KGs, these embeddings have several shortcomings. (1) The potential of the model to learn and generalise relations between entities in different languages is often restricted by sparse cross-lingual links, resulting in less accurate cross-lingual representations of entities. (2) Polysemy, which occurs when a word has numerous meanings, can be difficult to address across languages, resulting in ambiguity in cross-lingual representations. (3) Entities and relations can have very context-dependent and language-specific meanings, which is a challenging task for multilingual embeddings to capture the nuances of the context. (4) Resource imbalances may result in low-resource languages having inadequate training data and linguistic resources, impacting the entity and relation embeddings.

4.2 General Challenges

In addition to the goal of accounting for a broader spectrum of available information, there are more general challenges and opportunities for KG embedding models: (1) KG embedding models can inherit biases from training data, thereby reinforcing societal preconceptions. (2) Scalable embedding approaches are required for large-scale KGs with millions or billions of elements and relations. (3) Improving the interpretability and explainability of embeddings remains a challenge.

4.2.1 Bias in KG Embeddings

KGs, which serve as the foundation for KG embeddings, are regarded as crucial tools for organising and presenting information, enabling us to comprehend the vast quantities of available data. Once constructed, KGs are commonly regarded as "gold standard" data sources that uphold the accuracy of other systems, thus making the objectivity and neutrality of the information they convey vital concerns. Biases inherent to KGs may become magnified and spread through KG-based systems [150]. Traditionally, bias can be defined as "a disproportionate weight in favour of or against an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair"³. Taking into account the bias networking effect for KGs, it is crucial that various types of bias are already acknowledged and addressed during KG construction [79].

³ Wikipedia article on bias. https://en.wikipedia.org/wiki/Bias, retrieved 2023-11-28.

4:16 Knowledge Graph Embeddings: Open Challenges and Opportunities

Biases within KGs, as well as the approaches to address them, differ from those found in linguistic models or image classification. KGs are sparse by nature, i.e., only a small number of triples are available per entity. In contrast, linguistic models acquire the meaning of a term through its contextual usage in extensive corpora, while image classification leverages millions of labelled images to learn classes. Biases in KGs can arise from various sources, including the design of the KG itself, the (semi-)automated generation of the source data, and the algorithms employed to sample, aggregate, and process the data. These source biases typically manifest in expressions, utterances, and textual sources, which can then permeate downstream representations and in particular KG embeddings. Additionally, we must also account for a wide range of human biases, such as reporting bias, selection bias, confirmation bias, overgeneralisation, and more.

Biases in KGs as the source of KG embeddings can arise from multiple sources. Data bias occurs already in the data collection process or simply from the available source data. Schema bias depends on the chosen ontology for the KG or simply is already embedded within the used ontologies [79]. Inferential bias might result from drawing inferences on the represented knowledge. Ontologies are typically defined by a group of knowledge engineers in collaboration with domain experts and consequently (implicitly) reflect the world views and biases of the development team. Ontologies are also prone to encoding bias depending on the chosen representation language and modelling framework. Moreover, biases in KG embeddings may in particular arise from the chosen embedding method as for instance induced by application-specific loss functions. Inferential biases, which may arise at the inferencing level, such as reasoning, querying, or rule learning, are mostly limited to KGs themselves and rarely propagate to KG embeddings. A simple example of inferencing bias might be the different SPARQL entailment regimes, which in consequence, might be responsible for different results that different SPARQL endpoints deliver despite containing the same KG [2, 55].

Collaboratively built KGs, such as DBpedia or GeoNames, also exhibit social bias, often arising from the western-centric world view of their main contributors [37]. In addition, some "truths" represented in such KGs may be considered controversial or opinionated, which underlines the importance of provenance information.

For KG embeddings that represent a vector space-based approximation of the structural and semantic information contained in a KG, one of the main sources of bias lies in the sparsity and incompleteness of most KGs. KG embeddings trained on incomplete KGs might favour entities for which more information is available [136]. Moreover, if the underlying KG is biased, then KG embeddings trained on this base data will as well be, and in fact, bias may even be amplified. De-biasing of KG embeddings requires methods for detecting as well as removing bias in KG embeddings. Depending on the underlying embedding model, this task might become complex and requires finetuning of embeddings with respect to certain sensitive relations [45, 46, 9].

4.2.2 Reliability and Scalability of KG Embeddings

KG embedding methods suffer from many issues in terms of scalability. For example, many studies experiment mainly on (poorly constructed) subsets of Freebase and Wordnet, the infamous FB15k and WN18 [1], which are known to suffer from information leakage. These datasets contain in the order of a few million triples and rarely go beyond 1,000 relationship types, usually focusing on subgraphs with 200 or fewer. Recently, more realistic datasets have been proposed in terms of the quality of the data involved and of the link prediction task adopted [145]. Nonetheless, even these are far from being representative of typical real-world KG applications. Consider that DBpedia contains 52M distinct triples involving 28M distinct literals and as many distinct entities, with 1.3K distinct relationship types. Indeed, a recent Wikidata snapshot contains 1.926 billion triples, involving more than 600M entities and 904M distinct literals across 9K relationship types [134].

The size of real-world KGs is far beyond the capabilities of current methods, and the current results on small controlled benchmarks cannot be seen as representative of their scalability and reliability on real-world deployment. This perhaps also suggests the need for methods designed end-to-end to consider cases where different models can be learned for different subgraphs and then combined in a modular fashion. Last but not least, as KG embedding methods are adopted for tasks that go beyond link prediction, e.g., KG alignment [159], we refer to the well-known issues of scale in terms of dataset size (number of triples) and in terms of heterogeneity (scale of the vocabulary of relationships and attributes), as well as to new important issues based on the number of KGs to align, i.e., scale in terms of the number of distinct KG sources [16].

4.2.3 Explainability of KG Embeddings

One of the persistent difficulties is the development of KG embedding methods to enhance interpretability and explainability. This includes comprehending the reasoning and decisionmaking processes of KG embedding models as well as providing explanations for their predictions. KG embeddings have several advantages over conventional representations produced by deep learning algorithms, including their absence of ambiguity and the ability to justify and explain decisions [125]. Additionally, they can offer a semantic layer to help applications such as questionanswering, which are normally handled by text-based brute force techniques. CRIAGE [129] is one such tool that can be used to understand the impact of adding and removing facts. GNNExplainer [203] is proposed for the explainability of the predictions done by GNNs. Deep Knowledge-Aware Networks [176] and Knowledge-aware Path Recurrent Networks [180] have witnessed a surge in attention to recommendation systems. They model sequential dependencies that link users and items. OpenDialKG [117] is a corpus that aligns KGs with dialogues and presents an attention-based model that learns pathways from dialogue contexts and predicts relevant novel entities. These models offer a semantic and explicable layer for conversational agents and recommendations, aiding in the completion and interpretation of the predictions.

Limitations. However, there are still a number of limitations: (1) The lack of standardised evaluation standards makes it difficult to compare different approaches and assess performance consistently. (2) Improving interpretability often comes at the expense of performance and striking a balance between interpretability and performance still remains a challenge. (3) User-centric evaluation is necessary to understand the practical utility of explainable KG embeddings. (4) Current research on KG embedding explainability often focuses on global or model-level explanations, ignoring the importance of contextual and domain-specific explanations.

4.2.4 Complex Logical Query Answering and Approximate Answering of Graph Queries

The link prediction task is often seen as a graph completion task. However, it can equivalently be cast as a query-answering task for a very simple query. For example, if we predict the tail of the triple $\langle h, r, ? \rangle$, the task is equivalent to answering the corresponding query as if the graph had all the missing information. Recently, researchers started investigating how we could answer such queries if they are more complex, a task known as complex logical query answering⁴. The goal is, given a graph with missing information and a graph query, to produce the answers to the query as if the graph were complete (or more commonly, produce a ranking of possible answers).

⁴ also sometimes approximate query answering, multi-hop reasoning, or query embedding

4:18 Knowledge Graph Embeddings: Open Challenges and Opportunities

One might naïvely assume that this can be solved by first completing the graph and then performing a traditional graph query on the completed graph. The issue is, however, that a very large KG can never be complete. This is because link prediction models do not yield a set of missing edges, but rather a ranking of possible completions for an incomplete triple.

We can distinguish three main lines of work in this area. The reader is referred to relevant surveys [138, 30] for more details. The first group of approaches are those that make use of a link predictor, like the ones introduced above. These methods *decompose* the query into triples and then use the link prediction model to make predictions for the triples. The first approach of this type was CQD [7], which uses fuzzy logic to combine the outputs of the link predictor. Further developments for this type of model include QTO [13], which materialises all intermediate scores for the link predictors and makes sure that edges existing in the graph are always regarded as more certain than those predicted by the link predictor. Another newer approach is Adaptive CQD [8], which improves CQD by calibrating the scores of the link predictor across different relation types.

A second group of approaches are referred to as projection approaches, and the earliest approaches in this domain are of this type. These methods are characterised by the restriction that they can only answer DAG-shaped graph queries. They are inspired by translation-based link predictors. Starting from the entities in the query (in this context called the anchors), they project them with a relation-specific model to a representation for the tail entity. This representation then replaces the other occurrences as a subject of the variable in the query. If a variable occurs in more than one object position, a model is invoked to combine the computed projections into a single representation (called the intersection). The first approach of this type was Graph Query Embedding (GQE) [60], which did the above using vectors as representations, simple linear projections, and an MLP with element-wise mean for the intersection. Later examples include Query2Box [139], which uses axis-aligned hyperplanes to represent the outcomes of projections and intersections, and BetaE [140], which instead uses the beta distribution.

A final group of approaches is message-passing-based. These are very flexible and can deal with more query shapes than the above. This method regards the query as a small graph and embeds that complete query into a single embedding. Then, answers to the query are found simply by retrieving the entities of which the embedding is close to that query in the embedded space. A notable example is MPQE [35], which uses a relational graph convolutional network (R-GCN) to embed the query. The flexibility of these models is illustrated by StarQE [4], which can even answer hyper-relational queries (very similar to RDF-star).

Limitations. As indicated in the survey by Ren et al. [138], there are still very many open questions in this domain. (1) One aspect is that current approaches only support small subsets of all possible graph queries. For example, hardly any work attempts to answer cyclic queries, queries with variables on the relation position, or only variables in the whole query. (2) Also, the graph formalism currently used is limited; only very few approaches can deal with literal data, and there is no word yet on temporal KGs or the use of background semantics.

5 Applications

Recent research on KG embeddings has shown broad potential across diverse application domains such as search engines [43], recommendation systems [49], question-answering systems [73], biomedical and healthcare informatics [5], e-commerce [210], social network analysis [152], education [201], and scientific research [119]. However, in this study, we highlight two such domains: recommendation and biomedical/therapeutic use cases.

5.1 KG Embedding for Recommendation

Recommender systems (RSs) are an integral part of many online services and applications to provide relevant content and products tailored to their users. Many RSs identify user preference patterns assuming that users with similar past behaviour have similar preferences, e.g., people that watch the same movies are likely to do so also in the future, an approach commonly referred to as collaborative filtering [69, 68]. Yet, many existing methods only work in a warm-start setting, where it is assumed that all users and items have been seen during training [61, 205]. Moreover, methods that try to deal with cold-start settings, where for some users or items only user-item interactions are known and only at inference time [202, 205], making them unable to handle situations where this type of data is sparse, e.g., long-tail users and items. Therefore, we can see this problem as a link prediction problem, and we can also distinguish between a transductive setting and an inductive setting. In the transductive setting, some approaches try to exploit other contextual information from KGs, e.g., semantic annotations, taxonomies, item descriptions, or categories, to overcome these problems. In particular, a large body of methods exploits both domain-specific and open-domain KGs integrated with user and item information. In practice, users and items are nodes connected by special domain-specific relation types, e.g., a rating or a purchase, and item nodes are represented with additional connections to other entities describing their categories, features, producers, and provenance. This information, in the form of a Collaborative KG, is adopted as additional side information in the recommendation process [179, 175, 126]. These methods can be grouped into three categories:

- 1. path-based methods, which capture information from distant nodes but tend to dismiss much of the structural information in KG and are very dependent on the paths selected during training [180, 191, 162];
- 2. embedding-based methods, which use existing transductive graph embedding approaches to capture the semantic relations of the graph structure, such as TransR [206] or Node2Vec [56], further applying them in recommendation scenarios [126, 207]; and
- **3.** structural-based methods, which use GNNs to aggregate structural information of each node's neighbourhood [175, 179].

Among these, GNNs have recently shown promising results thanks to their ability to model relations and capture high-order connectivity information by combining KGs and collaborative data (user-item interactions) [179]. Nonetheless, these approaches often rely on transductive methods, making them unable to handle frequent changes in the graph. Moreover, their user-item representation often is limited to a single relation type and still cannot fully exploit the contextual knowledge offered by open-domain KGs, due to only very few relation types being considered. Furthermore, these approaches need to be able to exploit both the structure of the graph and the attributes describing the items.

5.2 Multimodal KG Embeddings for Biomedical and Therapeutic Use

In the biomedical domain, KGs are a natural way to model and represent complex biomedical structured data, such as molecular interactions, signalling pathways and disease co-morbidities [106]. Information from a single source usually does not provide sufficient data, and various state-of-the-art studies have shown that incorporating multiple heterogeneous knowledge sources and modalities yields better predictions [101, 53, 71]. Learning an effective representation that leverages the topology of these multimodal and heterogeneous KGs to create optimised embedding representations is key to applying AI models. These optimised embeddings can then be fed into link prediction models, such as for interactions between proteins [80], drugs [53], drug-targets [53, 101], or drug indication/contraindications for diseases [71].

4:20 Knowledge Graph Embeddings: Open Challenges and Opportunities

For instance, Otter-Knowledge [101] uses MKGs built from diverse sources, where each node has a modality assigned, such as textual (e.g., protein function), numerical (e.g., molecule mass), categorical entities (e.g., protein family), and modalities for representing protein and molecules. For each modality in the graph, a model is assigned to compute initial embeddings, e.g., pre-trained language models such as ESM [142] and MolFormer [144] are used for protein sequences and molecules' SMILES, respectively. A GNN is then invoked to enrich the initial representations and train a model to produce knowledge-enhanced representations for drug molecules and protein entities. These representations can improve drug-target binding affinity prediction tasks [72], even in the presence of entities not encountered during training or having missing modalities.

During training, attribute modalities are treated as relational triples of structured knowledge instead of predetermined features, making them first-class citizens of the MKG [128, 101]. The advantage of this approach is that entity nodes are not required to carry all multimodal properties or project large property vectors with missing values. Instead, the projection is done per modality and only when such a modality exists for the entity.

6 Discussion and Conclusion

Currently, the vast majority of evaluations of knowledge graph embeddings are conducted on the task of link prediction. At the same time, embeddings created with such techniques are used across a wide range of diverse downstream tasks, such as recommender systems, text annotation and retrieval, fact validation, data interpretation and integration, to name just a few. This raises the question: How suitable is the effectiveness of a link prediction task as a predictor of the applicability of a particular KGE method for a particular downstream task?

While the evaluation of link prediction is quite standardised with respect to benchmark datasets and evaluation metrics, the field of downstream applications is much more diverse and less standardised. Some frameworks, such as GEval [127] and kgbench [19], offer a greater variety of tasks and evaluations, including evaluation metrics and dataset splits.

Some studies have looked into characterising the representation capabilities of different KGE methods. They, for instance, analyse whether different classes are separated in the embedding space [6, 77, 216]. More recently, the DLLC benchmark [132] has been proposed, which allows for analysing which types of classification problems embeddings produced by a particular method can address. Other studies analyse the distance function in the resulting embedding spaces, finding that while most approaches create embedding spaces that encode entity similarity, others focus on entity relatedness [131], and that some methods can actually be altered to focus more on similarity and relatedness [133].

In addition, link prediction, entity categorisation, KG completion, and KG embeddings are crucial for a number of downstream activities, such as entity recommendation, relation extraction, question-answering, recommender systems, semantic search, and information retrieval. Models that leverage user profiles, historical interactions, and KGs can deliver personalised recommendations, capture similarity and relevance, and increase accuracy and relevance. KG embeddings also improve the accuracy of relation extraction by adding structured knowledge. The majority of existing KG embedding models are generalised, that is, they are trained and evaluated on open KGs for KG completion. However, task-specific KG embeddings would be quite advantageous in various kinds of applications, which still remains an open research task. They can be optimised for creating representations for specific tasks, improving performance, focusing on relevant information extraction, resolving data scarcity, and thereby improving interpretability and explainability. With the use of domain-specific data or constraints, these embeddings can be trained to grasp and reason about the relationships and semantics unique to that domain.

Recent ongoing research also reveals that when KG embeddings and LLMs are combined, a symbiotic relationship results, maximising the benefits of each methodology. While LLMs help to integrate textual knowledge, improve entity and relation linking, promote cross-modal fusion, and increase the explainability of KG embeddings, KG embeddings provide structured knowledge representations that improve the contextual comprehension and reasoning of LLMs. Therefore, future research may focus on building more robust and comprehensive models for knowledge representation, reasoning, and language understanding as a result of these interrelated effects.

KG embeddings will continue to evolve and serve an important role in enabling effective knowledge representation, reasoning, and decision-making as KGs grow in scale and complexity. This study highlights the potential of KG embeddings to convert unstructured data into structured knowledge, reveal deeper insights, and enhance intelligent applications.

— References -

- 1 Farahnaz Akrami, Mohammed Samiul Saeef, Qingheng Zhang, Wei Hu, and Chengkai Li. Realistic re-evaluation of knowledge graph completion methods: An experimental study. In *Proceedings* of the 2020 ACM SIGMOD International Conference on Management of Data, pages 1995–2010. ACM, 2020. doi:10.1145/3318464.3380599.
- 2 Mehwish Alam, George Fletcher, Antonie Isaac, Aidan Hogan, Diana Maynard, Heiko Paulheim, Harald Sack, Elena Simperl, Lise Stork, Marieke van Erp, and Hideaki Takeda. Bias in knowledge graph systems. In Proc. of Dagstuhl Seminar 22372: Knowledge Graphs and their Role in the Knowledge Engineering of the 21st Century, 2022; Vol. 12(9), pages 106–109. Dagstuhl Reports, 2022. doi:10.4230/DagRep.12.9.60.
- 3 Mirza Mohtashim Alam, Md Rashad Al Hasan Rony, Mojtaba Nayyeri, Karishma Mohiuddin, MST Mahfuja Akter, Sahar Vahdati, and Jens Lehmann. Language model guided knowledge graph embeddings. *IEEE Access*, 10:76008–76020, 2022. doi:10.1109/ACCESS.2022.3191666.
- 4 Dimitrios Alivanistos, Max Berrendorf, Michael Cochez, and Mikhail Galkin. Query embedding on hyper-relational knowledge graphs. In 10th International Conference on Learning Representations, ICLR 2020, virtual, April 25-29, 2022. OpenReview.net, 2022. URL: https://openreview.net/ forum?id=4rLw09TgRw9.
- 5 Mona Alshahrani, Maha A Thafar, and Magbubah Essack. Application and evaluation of knowledge graph embeddings in biomedical data. *PeerJ Computer Science*, 7:e341, feb 2021. doi: 10.7717/PEERJ-CS.341.
- 6 Faisal Alshargi, Saeedeh Shekarpour, Tommaso Soru, and Amit Sheth. Concept2vec: Metrics for evaluating quality of embeddings for ontological concepts. arXiv preprint arXiv:1803.04488, abs/1803.04488, 2018. doi:10.48550/arXiv.1803. 04488.
- 7 Erik Arakelyan, Daniel Daza, Pasquale Minervini, and Michael Cochez. Complex query answering with neural link predictors. In *International Conference on Learning Representations*. OpenReview.net, 2021. URL: https://openreview.net/ forum?id=Mos9F9kDwkz.
- 8 Erik Arakelyan, Pasquale Minervini, Daniel Daza, Michael Cochez, and Isabelle Augenstein. Adapting neural link predictors for data-efficient com-

plex query answering. arXiv, jan 2023. arXiv: 2301.12313.

- 9 Mario Arduini, Lorenzo Noci, Federico Pirovano, Ce Zhang, Yash Raj Shrestha, and Bibek Paudel. Adversarial learning for debiasing knowledge graph embeddings. ArXiv, abs/2006.16309, jun 2020. doi:10.48550/arXiv.2006.16309.
- 10 Siddhant Arora, Srikanta Bedathur, Maya Ramanath, and Deepak Sharma. IterefinE: Iterative KG refinement embeddings using symbolic knowledge. In Automated Knowledge Base Construction, 2020. doi:10.24432/C5NP46.
- 11 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A nucleus for a web of open data. In *The semantic web*, volume 4825, pages 722–735. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-76298-0_52.
- 12 Luyi Bai, Xiangnan Ma, Mingcheng Zhang, and Wenting Yu. TPmod: A tendency-guided prediction model for temporal knowledge graph completion. ACM Transactions on Knowledge Discovery from Data, 15(3):1–17, jun 2021. doi: 10.1145/3443687.
- 13 Yushi Bai, Xin Lv, Juanzi Li, and Lei Hou. Answering complex logical queries on knowledge graphs via query computation tree optimization. *arXiv* preprint arXiv:2212.09567, abs/2212.09567, dec 2022. doi:10.48550/ARXIV.2212.09567.
- 14 Ivana Balažević, Carl Allen, and Timothy M Hospedales. Hypernetwork knowledge graph embeddings. In Proceedings of the International Conference on Artificial Neural Networks, volume 11731, pages 553–565. Springer International Publishing, 2019. doi:10.1007/978-3-030-30493-5_52.
- 15 Ivana Balažević, Carl Allen, and Timothy M. Hospedales. TuckER: Tensor factorization for knowledge graph completion. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, pages 5184–5193. Association for Computational Linguistics, 2019. doi:10.18653/V1/D19-1522.
- 16 Matthias Baumgartner, Daniele Dell'Aglio, Heiko Paulheim, and Abraham Bernstein. Towards the web of embeddings: Integrating multiple knowledge graph embedding spaces with FedCoder.

Journal of Web Semantics, 75:100741, 2023. doi: 10.1016/J.WEBSEM.2022.100741.

- 17 Russa Biswas, Harald Sack, and Mehwish Alam. MADLINK: Attentive multihop and entity descriptions for link prediction in knowledge graphs. *Semantic Web*, pages 1–24, 2022. doi:10.3233/ sw-222960.
- 18 Russa Biswas, Radina Sofronova, Mehwish Alam, and Harald Sack. Contextual language models for knowledge graph completion. In *MLS-MKG*, volume 2997. CEUR-WS.org, 2021. URL: https://ceur-ws.org/Vol-2997/paper3.pdf.
- 19 Peter Bloem, Xander Wilcke, Lucas van Berkel, and Victor de Boer. kgbench: A collection of knowledge graph datasets for evaluating relational and multimodal machine learning. In *European Semantic Web Conference*, volume 12731, pages 614–630. Springer International Publishing, 2021. doi:10.1007/978-3-030-77385-4_37.
- 20 Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. Freebase: A shared database of structured general human knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pages 1962–1963. AAAI Press, jul 2007. URL: http://www.aaai.org/Library/AAAI/2007/ aaai07-355.php.
- 21 Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Ok-Translating embeddings sana Yakhnenko. for modeling multi-relational data. Advances in neural information processing systems, 26:2787-2795, 2013. URL: https: //proceedings.neurips.cc/paper/2013/hash/ 1cecc7a77928ca8133fa24680a88d2f9-Abstract. html.
- 22 Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning structured embeddings of knowledge bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, volume 25, pages 301–306. Association for the Advancement of Artificial Intelligence (AAAI), 2011. doi:10.1609/AAAI.V25I1.7917.
- 23 Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. Temporal knowledge graph completion: A survey. arXiv preprint arXiv:2201.08236, abs/2201.08236:6545–6553, jan 2022. doi:10.24963/IJCAI.2023/734.
- 24 Alison Callahan, Jose Cruz-Toledo, Peter Ansell, and Michel Dumontier. Bio2RDF release 2: improved coverage, interoperability and provenance of life science linked data. In *Extended semantic web conference*, volume 7882, pages 200–212. Springer, Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-38288-8_14.
- 25 Soumen Chakrabarti, Harkanwar Singh, Shubham Lohiya, Prachi Jain, and Mausam. Joint completion and alignment of multilingual knowledge graphs. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 11922–11938. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.EMNLP-MAIN.817.
- 26 Kai-Wei Chang, Wen-tau Yih, Bishan Yang, and Christopher Meek. Typed tensor decomposition of knowledge bases for relation extraction. In

Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1568–1579. Association for Computational Linguistics, 2014. doi:10.3115/V1/D14-1165.

- 27 Xuelu Chen, Muhao Chen, Changjun Fan, Ankith Uppunda, Yizhou Sun, and Carlo Zaniolo. Multilingual knowledge graph completion via ensemble knowledge transfer. In *Findings of the Association* for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020, volume EMNLP 2020 of Findings of ACL, pages 3227–3238. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.FINDINGS-EMNLP.290.
- 28 Yihong Chen, Pasquale Minervini, Sebastian Riedel, and Pontus Stenetorp. Relation prediction as an auxiliary training objective for improving multi-relational graph representations. In 3rd Conference on Automated Knowledge Base Construction, AKBC 2021, Virtual, October 4-8, 2021, 2021. doi:10.24432/C54K5W.
- 29 Bonggeun Choi, Daesik Jang, and Youngjoong Ko. MEM-KGC: Masked entity model for knowledge graph completion with pre-trained language model. *IEEE Access*, 9:132025–132032, 2021. doi:10.1109/ACCESS.2021.3113329.
- 30 Michael Cochez, Dimitrios Alivanistos, Erik Arakelyan, Max Berrendorf, Daniel Daza, Mikhail Galkin, Pasquale Minervini, Mathias Niepert, and Hongyu Ren. Approximate answering of graph queries. In *Compendium of Neurosymbolic Artificial Intelligence*, volume 369, pages 373–386. IOS Press, aug 2023. doi:10.3233/FAIA230149.
- 31 Michael Cochez, Martina Garofalo, Jérôme Lenßen, and Maria Angela Pellegrino. A first experiment on including text literals in KGloVe. In Joint proceedings of the 4th Workshop on Semantic Deep Learning (SemDeep-4) and NLI-WoD4: Natural Language Interfaces for the Web of Data (NLIWOD-4) and 9th Question Answering over Linked Data challenge (QALD-9) colocated with 17th International Semantic Web Conference (ISWC 2018), Monterey, California, United States of America, October 8th - 9th, 2018, volume 2241 of CEUR Workshop Proceedings, pages 103-106. CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2241/paper-10.pdf.
- 32 Tu Dinh Nguyen Dai Quoc Nguyen, Dat Quoc Nguyen, and Dinh Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 327–333. Association for Computational Linguistics, 2018. doi:10.18653/V1/N18-2053.
- 33 Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In Proceedings of the 2018 conference on empirical methods in natural language processing, pages 2001– 2011. Association for Computational Linguistics, 2018. doi:10.18653/V1/D18-1225.
- 34 Daniel Daza, Dimitrios Alivanistos, Payal Mitra, Thom Pijnenburg, Michael Cochez, and Paul Groth. BioBLP: A modular framework for learning on multimodal biomedical knowledge graphs.

arXiv preprint arXiv:2306.03606, 2023. doi:10. 48550/ARXIV.2306.03606.

- 35 Daniel Daza and Michael Cochez. Message passing query embedding. In ICML Workshop - Graph Representation Learning and Beyond, 2020. doi: 10.48550/arXiv.2002.02406.
- 36 Daniel Daza, Michael Cochez, and Paul Groth. Inductive entity representations from text via link prediction. In *Proceedings of the Web Conference 2021*, pages 798–808. ACM, 2021. doi: 10.1145/3442381.3450141.
- 37 Gianluca Demartini. Implicit bias in crowdsourced knowledge graphs. In Companion Proceedings of The 2019 World Wide Web Conference, pages 624– 630. Association for Computing Machinery, 2019. doi:10.1145/3308560.3317307.
- 38 Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pages 1811–1818. Association for the Advancement of Artificial Intelligence (AAAI), 2018. doi:10.1609/AAAI.V32I1.11573.
- 39 Boyang Ding, Quan Wang, Bin Wang, and Li Guo. Improving knowledge graph embedding using simple constraints. arXiv preprint arXiv:1805.02408, pages 110–121, 2018. doi: 10.18653/V1/P18-1011.
- 40 Claudia d'Amato, Nicola Flavio Quatraro, and Nicola Fanizzi. Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In *European Semantic Web Conference*, volume 12731, pages 441–457. Springer, Springer International Publishing, 2021. doi: 10.1007/978-3-030-77385-4_26.
- 41 Yasha Ektefaie, George Dasoulas, Ayusa Noori, Maha Farhat, and Marinka Zitnik. Multimodal learning with graphs. Nature Machine Intelligence, 5:340–350, 2023. doi:10.1038/ S42256-023-00624-6.
- 42 Michael Ellers, Michael Cochez, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Privacy attacks on network embeddings. *CoRR*, abs/1912.10979, 2019. doi:10.48550/ arXiv.1912.10979.
- 43 Shahla Farzana, Qunzhi Zhou, and Petar Ristoski. Knowledge graph-enhanced neural query rewriting. In Companion Proceedings of the ACM Web Conference 2023, pages 911–919. ACM, apr 2023. doi:10.1145/3543873.3587678.
- 44 Jun Feng, Minlie Huang, Yang Yang, and Xiaoyan Zhu. GAKE: Graph aware knowledge embedding. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 641–651. ACL, 2016. URL: https://aclanthology.org/C16-1062/.
- 45 Joseph Fisher. Measuring social bias in knowledge graph embeddings. ArXiv, abs/1912.02761, dec 2019. doi:10.48550/arXiv.1912.02761.
- 46 Joseph Fisher, Arpit Mittal, Dave Palfrey, and Christos Christodoulopoulos. Debiasing knowledge graph embeddings. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 7332-7345. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.EMNLP-MAIN.595.

- 47 Mikhail Galkin, Max Berrendorf, and Charles Tapley Hoyt. An open challenge for inductive link prediction on knowledge graphs. arXiv preprint arXiv:2203.01520, abs/2203.01520, mar 2022. doi:10.48550/ARXIV.2203.01520.
- 48 Mikhail Galkin, Etienne Denis, Jiapeng Wu, and William L. Hamilton. NodePiece: Compositional and parameter-efficient representations of large knowledge graphs. In *International Conference* on Learning Representations. OpenReview.net, 2022. URL: https://openreview.net/forum?id= xMJWUKJnFSw.
- 49 Min Gao, Jian-Yu Li, Chun-Hua Chen, Yun Li, Jun Zhang, and Zhi-Hui Zhan. Enhanced multitask learning and knowledge graph-based recommender system. *IEEE Transactions on Knowledge and Data Engineering*, 35:10281–10294, 2023. doi:10.1109/TKDE.2023.3251897.
- 50 Alberto García-Durán, Sebastijan Dumancic, and Mathias Niepert. Learning sequence encoders for temporal knowledge graph completion. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 4816– 4821. Association for Computational Linguistics, Association for Computational Linguistics, 2018. doi:10.18653/v1/d18-1516.
- 51 Alberto García-Durán and Mathias Niepert. KBLRN: End-to-end learning of knowledge base representations with latent, relational, and numerical features. In Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018, pages 372-381. AUAI Press, 2018. URL: http://auai.org/uai2018/ proceedings/papers/149.pdf.
- 52 Dinesh Garg, Shajith Ikbal, Santosh K Srivastava, Harit Vishwakarma, Hima Karanam, and L Venkata Subramaniam. Quantum embedding of knowledge for reasoning. Advances in Neural Information Processing Systems, 32:5595-5605, 2019. URL: https: //proceedings.neurips.cc/paper/2019/hash/ cb12d7f933e7d102c52231bf62b8a678-Abstract. html.
- 53 Aryo Pradipta Gema, Dominik Grabarczyk, Wolf De Wulf, Piyush Borole, Javier Antonio Alfaro, Pasquale Minervini, Antonio Vergari, and Ajitha Rajan. Knowledge graph embeddings in the biomedical domain: Are they useful? a look at link prediction, rule learning, and downstream polypharmacy tasks. arXiv.org, abs/2305.19979, may 2023. doi:10.48550/ARXIV.2305.19979.
- 54 Genet Asefa Gesese, Russa Biswas, Mehwish Alam, and Harald Sack. A survey on knowledge graph embeddings with literals: Which model links better literal-ly? *Semantic Web*, 12:617–647, 2021. doi:10.3233/SW-200404.
- 55 Paul Groth, Elena Simperl, Marieke van Erp, and Denny Vrandečić. Knowledge graphs and their role in the knowledge engineering of the 21st century (Dagstuhl seminar 22372). Dagstuhl Reports, 12:60–120, 2023. doi:10.4230/DAGREP.12.9.60.
- 56 Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In 22nd ACM SIGKDD international conference on Knowledge

discovery and data mining, pages 855–864. ACM, 2016. doi:10.1145/2939672.2939754.

- 57 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 192–202. Association for Computational Linguistics, 2016. doi:10.18653/V1/ D16-1019.
- 58 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, pages 4816–4823. Association for the Advancement of Artificial Intelligence (AAAI), 2018. doi:10.1609/AAAI.V32I1.11918.
- 59 Ferras Hamad, Issac Liu, and Xian Xing Zhang. Food discovery with uber eats: Building a query understanding engine. uber engineering blog. https://eng.uber.com/uber-eats-queryunderstanding/., 2018. Accessed: 2022-07-03.
- 60 William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. In Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, Neur-IPS 2018, December 3-8, 2018, Montréal, Canada, pages 2030-2041, 2018. URL: https: //proceedings.neurips.cc/paper/2018/hash/ ef50c335cca9f340bde656363ebd02fd-Abstract. html.
- 61 William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 1025-1035, 2017. URL: https: //proceedings.neurips.cc/paper/2017/hash/ 5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract. html.
- Kiaolin Han, Daniele Dell'Aglio, Tobias Grubenmann, Reynold Cheng, and Abraham Bernstein. A framework for differentially-private knowledge graph embeddings. *Journal of Web Semantics*, 72:100696, 2022. doi:10.1016/J.WEBSEM.2021.100696.
- 63 Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In International Conference on Learning Representations, 2020. URL: https://openreview.net/forum?id= pGIHq1m7PU.
- 64 Zhen Han, Zifeng Ding, Yunpu Ma, Yujia Gu, and Volker Tresp. Learning neural ordinary equations for forecasting future links on temporal knowledge graphs. In Proceedings of the 2021 conference on empirical methods in natural language processing, pages 8352-8364. Association for Computational Linguistics, 2021. doi:10.18653/V1/ 2021.EMNLP-MAIN.658.
- 65 Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In Proceedings of the 25th ACM SIGKDD International

Conference on Knowledge Discovery & Data Mining, pages 1709–1719. ACM, 2019. doi:10.1145/ 3292500.3330838.

- 66 Ming He, Xiangkun Du, and Bo Wang. Representation learning of knowledge graphs via fine-grained relation description combinations. *IEEE Access*, 7:26466–26473, 2019. doi:10.1109/ACCESS.2019. 2901544.
- 67 Shizhu He, Kang Liu, Guoliang Ji, and Jun Zhao. Learning to represent knowledge graphs with gaussian embedding. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 623–632. ACM, 2015. doi:10.1145/2806416.2806502.
- 68 Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Light-GCN: Simplifying and powering graph convolution network for recommendation. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 639–648. ACM, 2020. doi: 10.1145/3397271.3401063.
- 69 Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017. doi:10.1145/3038912.3052569.
- 70 Nicolas Heist, Sven Hertling, Daniel Ringler, and Heiko Paulheim. Knowledge graphs on the weban overview. *Knowledge Graphs for eXplainable Artificial Intelligence*, pages 3–22, 2020. doi: 10.3233/SSW200009.
- 71 Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish Nadkarni, Benjamin Glicksberg, Nils Gehlenborg, and Marinka Zitnik. Zero-shot prediction of therapeutic use with geometric deep learning and clinician centered design. medRxiv, 2023. doi:10.1101/2023.03.19.23287458.
- 72 Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. Artificial intelligence foundation for therapeutic science. *Nature chemical biology*, 18:1033–1036, 2022. doi:10.1038/s41589-022-01131-2.
- 73 Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. Knowledge graph embedding based question answering. In Proceedings of the twelfth ACM international conference on web search and data mining, pages 105–113. ACM, 2019. doi: 10.1145/3289600.3290956.
- 74 Zijie Huang, Zheng Li, Haoming Jiang, Tianyu Cao, Hanqing Lu, Bing Yin, Karthik Subbian, Yizhou Sun, and Wei Wang. Multilingual knowledge graph completion with self-supervised adaptive graph alignment. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, volume abs/2203.14987, pages 474–485. Association for Computational Linguistics, 2022. doi: 10.18653/V1/2022.ACL-LONG.36.
- 75 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Knowledge graph embed-

dings for link prediction: Beware of semantics! In DL4KG@ ISWC 2022: Workshop on Deep Learning for Knowledge Graphs, held as part of ISWC 2022: the 21st International Semantic Web Conference, volume 3342. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3342/paper-4.pdf.

- 76 Andreea Iana and Heiko Paulheim. More is not always better: The negative impact of A-box materialization on RDF2vec knowledge graph embeddings. In Workshop on Combining Symbolic and Sub-symbolic Methods and their Applications (CSSA), volume abs/2009.00318. CEUR-WS.org, sep 2020. URL: https://ceur-ws.org/Vol-2699/ paper05.pdf.
- 77 Nitisha Jain, Jan-Christoph Kalo, Wolf-Tilo Balke, and Ralf Krestel. Do embeddings actually capture knowledge graph semantics? In *The Semantic* Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18, volume 12731, pages 143–159. Springer, Springer International Publishing, 2021. doi:10.1007/ 978-3-030-77385-4_9.
- 78 Nitisha Jain, Trung-Kien Tran, Mohamed H Gad-Elrab, and Daria Stepanova. Improving knowledge graph embeddings with ontological reasoning. In *International Semantic Web Conference*, volume 12922, pages 410–426. Springer, Springer International Publishing, 2021. doi: 10.1007/978-3-030-88361-4_24.
- 79 Krzysztof Janowicz, Bo Yan 0003, Blake Regalia, Rui Zhu, and Gengchen Mai. Debiasing knowledge graphs: Why female presidents are not like female popes. In Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018, volume 2180 of CEUR Workshop Proceedings. CEUR-WS.org, 2018. URL: http://ceur-ws.org/Vol-2180/ ISWC_2018_Outrageous_Ideas_paper_17.pdf.
- 80 Kanchan Jha, Sriparna Saha, and Snehanshu Saha. Prediction of protein-protein interactions using deep multi-modal representations. In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE, jul 2021. doi: 10.1109/IJCNN52387.2021.9533478.
- 81 Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing, pages 687-696. Association for Computational Linguistics, 2015. doi:10.3115/V1/P15-1067.
- 82 Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. Knowledge graph completion with adaptive sparse transfer matrix. In *Proceedings of the Thirtieth AAAI conference on artificial intelligence*, volume 30, pages 985–991. Association for the Advancement of Artificial Intelligence (AAAI), 2016. doi:10.1609/AAAI.V30I1.10089.
- 83 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Net*-

works and Learning Systems, 33(2):494-514, 2021. doi:10.1109/TNNLS.2021.3070843.

- 84 Ningning Jia, Xiang Cheng, and Sen Su. Improving knowledge graph embedding using locally and globally attentive relation paths. In European Conference on Information Retrieval, volume 12035, pages 17–32. Springer International Publishing, 2020. doi:10.1007/978-3-030-45439-5_2.
- 85 Yantao Jia, Yuanzhuo Wang, Hailun Lin, Xiaolong Jin, and Xueqi Cheng. Locally adaptive translation for knowledge graph embedding. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, volume 30, pages 992–998. Association for the Advancement of Artificial Intelligence (AAAI), 2016. doi:10.1609/AAAI.V30I1.10091.
- 86 Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 6669–6683. Association for Computational Linguistics, 2020. doi:10.18653/V1/ 2020.EMNLP-MAIN.541.
- 87 Jaehun Jung, Jinhong Jung, and U Kang. Learning to walk across time for interpretable temporal knowledge graph completion. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pages 786–795. ACM, 2021. doi:10.1145/3447548.3467292.
- 88 Lucie-Aimée Kaffee. Multilinguality in knowledge graphs. PhD thesis, University of Southampton, 2021. URL: https://eprints.soton.ac.uk/ 456783/.
- 89 Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. A glimpse into babel: An analysis of multilinguality in wikidata. In Proceedings of the 13th International Symposium on Open Collaboration, OpenSym 2017, Galway, Ireland, August 23-25, 2017, pages 14:1–14:5. ACM, 2017. doi:10.1145/3125433.3125465.
- 90 Seyed Mehran Kazemi, Rishab Goel, Kshitij Jain, Ivan Kobyzev, Akshay Sethi, Peter Forsyth, and Pascal Poupart. Representation learning for dynamic graphs: A survey. The Journal of Machine Learning Research, 21:2648-2720, 2020. URL: http://jmlr.org/papers/v21/19-447.html.
- 91 Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. Advances in neural information processing systems, 31:4289-4300, 2018. URL: https: //proceedings.neurips.cc/paper/2018/hash/ b2ab001909a8a6f04b51920306046ce5-Abstract. html.
- 92 Bosung Kim, Taesuk Hong, Youngjoong Ko, and Jungyun Seo. Multi-task learning for knowledge graph completion with pre-trained language models. In *COLING*, pages 1737–1743. International Committee on Computational Linguistics, 2020. doi:10.18653/V1/2020.COLING-MAIN.153.
- 93 Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, abs/1609.02907, 2016. doi:10.48550/arXiv.1609.02907.
- 94 Franz Krause. Dynamic knowledge graph embeddings via local embedding reconstructions.

In The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece, May 29-June 2, 2022, Proceedings, volume 13384, pages 215-223. Springer, 2022. doi:10.1007/ 978-3-031-11609-4_36.

- 95 Franz Krause, Tobias Weller, and Heiko Paulheim. On a generalized framework for time-aware knowledge graphs. In Towards a Knowledge-Aware AI: SEMANTICS 2022—Proceedings of the 18th International Conference on Semantic Systems, 13-15 September 2022, Vienna, Austria, volume 55, page 69. IOS Press, IOS Press, 2022. doi:10.3233/ssw220010.
- 96 Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123:32–73, 2017. doi:10.1007/ S11263-016-0981-7.
- 97 Arun Krishnan. Making search easier: How amazon's product graph is helping customers find products more easily. amazon blog. https://blog.aboutamazon.com/innovation/ making-search-easier, 2018. Accessed: 2022-07-03.
- 98 Agustinus Kristiadi, Mohammad Asif Khan, Denis Lukovnikov, Jens Lehmann, and Asja Fischer. Incorporating literals into knowledge graph embeddings. In *Proceedings of the International* Semantic Web Conference, volume 11778, pages 347–363. Springer International Publishing, 2019. doi:10.1007/978-3-030-30793-6_20.
- 99 Denis Krompaß, Stephan Baier, and Volker Tresp. Type-constrained representation learning in knowledge graphs. In *International semantic web conference*, volume 9366, pages 640–655. Springer, Springer International Publishing, 2015. doi: 10.1007/978-3-319-25007-6_37.
- 100 Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. Tensor decompositions for temporal knowledge base completion. In *International Conference on Learning Representations*, 2019. URL: https://openreview.net/forum?id=rke2P1BFwS.
- 101 Hoang Thanh Lam, Marco Luca Sbodio, Marcos Martinez Gallindo, Mykhaylo Zayats, Raul Fernandez-Diaz, Victor Valls, Gabriele Picco, Cesar Berrospi Ramis, and Vanessa Lopez. Otter-knowledge: benchmarks of multimodal knowledge graph representation learning from different sources for drug discovery. arXiv.org, abs/2306.12802, jun 2023. doi:10.48550/ARXIV. 2306.12802.
- 102 Julien Leblay and Melisachew Wudage Chekol. Deriving validity time in knowledge graph. In Companion Proceedings of the The Web Conference 2018, pages 1771–1776. ACM Press, 2018. doi:10.1145/3184558.3191639.
- 103 Julien Leblay, Melisachew Wudage Chekol, and Xin Liu. Towards temporal knowledge graph embeddings with arbitrary time precision. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pages 685– 694. ACM, 2020. doi:10.1145/3340531.3412028.

- 104 Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. Dbpedia–a largescale, multilingual knowledge base extracted from Wikipedia. Semantic web, 6(2):167–195, 2015. doi:10.3233/SW-140134.
- 105 Da Li, Ming Yi, and Yukai He. LP-BERT: Multi-task pre-training knowledge graph bert for link prediction. arXiv preprint arXiv:2201.04843, abs/2201.04843, 2022. doi:10.48550/arXiv.2201. 04843.
- 106 Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6:1–17, 2022. doi:10.1038/s41551-022-00942-x.
- 107 Qian Li, Daling Wang, Shi Feng, Cheng Niu, and Yifei Zhang. Global graph attention embedding network for relation prediction in knowledge graphs. *IEEE Transactions on Neural Networks and Learning Systems*, 33:6712–6725, 2021. doi:10.1109/TNNLS.2021.3083259.
- 108 Yuezhang Li, Ronghuo Zheng, Tian Tian, Zhiting Hu, Rahul Iyer, and Katia P. Sycara. Joint embedding of hierarchical categories and entities for concept categorization and dataless classification. In *COLING*, volume abs/1607.07956, pages 2678– 2688. ACL, 2016. URL: https://aclanthology. org/C16-1252/.
- 109 Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. Temporal knowledge graph reasoning based on evolutional representation learning. In Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pages 408–417. ACM, 2021. doi:10.1145/3404835.3462963.
- 110 Paul Pu Liang, Yiwei Lyu, Xiang Fan, Zetian Wu, Yun Cheng, Jason Wu, Leslie Chen, Peter Wu, Michelle A. Lee, Yuke Zhu, Ruslan Salakhutdinov, and Louis-Philippe Morency. MultiBench: Multiscale benchmarks for multimodal representation learning. *CoRR*, abs/2107.07502, 2021. doi:10.48550/arXiv.2107.07502.
- 111 Siyuan Liao, Shangsong Liang, Zaiqiao Meng, and Qiang Zhang. Learning dynamic embeddings for temporal knowledge graphs. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 535–543. ACM, 2021. doi:10.1145/3437963.3441741.
- 112 Lifan Lin and Kun She. Tensor decompositionbased temporal knowledge graph embedding. In 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), pages 969–975. IEEE, IEEE, nov 2020. doi:10.1109/ ICTAI50040.2020.00151.
- 113 Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, volume abs/1506.00379, pages 705–714. Association for Computational Linguistics, 2015. doi:10.18653/V1/D15-1082.
- 114 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation

embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, pages 2181–2187. Association for the Advancement of Artificial Intelligence (AAAI), 2015. doi:10.1609/AAAI.V29I1.9491.

- 115 Sedigheh Mahdavi, Shima Khoshraftar, and Aijun An. dynnode2vec: Scalable dynamic network embedding. In 2018 IEEE International Conference on Big Data (Big Data), pages 3762–3765. IEEE, IEEE, 2018. doi:10.1109/BIGDATA.2018. 8621910.
- 116 Pasquale Minervini, Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel. Adversarial sets for regularising neural link predictors. In UAI2017, the 33rd Conference on Uncertainty in Artificial Intelligence, volume abs/1707.07596, pages 1–10. AUAI Press, 2017. URL: http://auai. org/uai2017/proceedings/papers/306.pdf.
- 117 Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In Proceedings of the 57th annual meeting of the association for computational linguistics, pages 845–854. Association for Computational Linguistics, 2019. doi:10.18653/ V1/P19-1081.
- 118 Deepak Nathani, Jatin Chauhan, Charu Sharma, and Manohar Kaul. Learning attention-based embeddings for relation prediction in knowledge graphs. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pages 4710–4723. Association for Computational Linguistics, 2019. doi:10.18653/V1/P19-1466.
- 119 Mojtaba Nayyeri, Gokce Muge Cil, Sahar Vahdati, Francesco Osborne, Mahfuzur Rahman, Simone Angioni, Angelo Salatino, Diego Reforgiato Recupero, Nadezhda Vassilyeva, Enrico Motta, et al. Trans4E: Link prediction on scholarly knowledge graphs. *Neurocomputing*, 461:530–542, oct 2021. doi:10.1016/J.NEUCOM.2021.02.100.
- 120 Mojtaba Nayyeri, Zihao Wang, Mst Akter, Mirza Mohtashim Alam, Md Rashad Al Hasan Rony, Jens Lehmann, Steffen Staab, et al. Integrating knowledge graph embedding and pretrained language models in hypercomplex spaces. arXiv preprint arXiv:2208.02743, abs/2208.02743, 2022. doi:10.48550/ARXIV.2208.02743.
- 121 Mojtaba Nayyeri, Bo Xiong, Majid Mohammadi, Mst. Mahfuja Akter, Mirza Mohtashim Alam, Jens Lehmann, and Steffen Staab. Knowledge graph embeddings using neural Itô process: From multiple walks to stochastic trajectories. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7165–7179, 2023. doi: 10.18653/V1/2023.FINDINGS-ACL.448.
- 122 Thomas Neumann and Guido Moerkotte. Characteristic sets: Accurate cardinality estimation for RDF queries with multiple joins. In 2011 IEEE 27th International Conference on Data Engineering, pages 984–994. IEEE, IEEE, 2011. doi:10.1109/ICDE.2011.5767868.
- 123 Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on*

Artificial Intelligence, volume 30, pages 1955–1961. Association for the Advancement of Artificial Intelligence (AAAI), 2016. doi:10.1609/AAAI.V30I1. 10314.

- 124 Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the International Conference on Machine Learning*, volume 11, pages 3104482–3104584. Omnipress, 2011. URL: https://icml.cc/2011/papers/438_ icmlpaper.pdf.
- 125 Matteo Palmonari and Pasquale Minervini. Knowledge graph embeddings and explainable AI. Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges, 47:49, 2020. doi:10.3233/SSW200011.
- 126 Enrico Palumbo, Diego Monti, Giuseppe Rizzo, Raphaël Troncy, and Elena Baralis. entity2rec: Property-specific knowledge graph embeddings for item recommendation. Expert Systems with Applications, 151:113235, aug 2020. doi:10.1016/J. ESWA.2020.113235.
- 127 Maria Angela Pellegrino, Abdulrahman Altabba, Martina Garofalo, Petar Ristoski, and Michael Cochez. GEval: a modular and extensible evaluation framework for graph embedding techniques. In *European Semantic Web Conference*, volume 12123, pages 565–582. Springer, Springer International Publishing, 2020. doi:10.1007/ 978-3-030-49461-2_33.
- 128 Pouya Pezeshkpour, Liyan Chen, and Sameer Singh. Embedding multimodal relational data for knowledge base completion. In *Proceedings of the* 2018 Conference on Empirical Methods in Natural Language Processing, volume abs/1809.01341, pages 3208–3218, Brussels, Belgium, 2018. Association for Computational Linguistics. doi: 10.18653/V1/D18-1359.
- 129 Pouya Pezeshkpour, CA Irvine, Yifan Tian, and Sameer Singh. Investigating robustness and interpretability of link prediction via adversarial modifications. In *Proceedings of NAACL-HLT*, pages 3336–3347. Association for Computational Linguistics, 2019. doi:10.18653/V1/N19-1337.
- 130 R.J. Pittman, Srivastava. Amit, Sanjika Hewavitharana, Ajinkya Kale, and Saab Mansour. Cracking the code on conversational commerce. ebay blog. https://www.ebayinc.com/stories/ news/cracking-the-code-on-conversational commerce/, 2017. Accessed: 2022-07-03.
- 131 Jan Portisch, Nicolas Heist, and Heiko Paulheim. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction two sides of the same coin? *Semantic Web*, 13:399– 422, 2022. doi:10.3233/SW-212892.
- 132 Jan Portisch and Heiko Paulheim. The DLCC node classification benchmark for analyzing knowledge graph embeddings. In *The Semantic Web–ISWC* 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings, volume abs/2207.06014, pages 592–609. Springer, Springer International Publishing, 2022. doi:10.1007/978-3-031-19433-7_34.
- 133 Jan Portisch and Heiko Paulheim. Walk this way! entity walks and property walks for RDF2vec.

CoRR, abs/2204.02777:133-137, 2022. doi:10.48550/ARXIV.2204.02777.

- 134 Kashif Rabbani, Matteo Lissandrini, and Katja Hose. Extraction of validating shapes from very large knowledge graphs. *Proceedings of the VLDB Endowment*, 16(5):1023–1032, jan 2023. doi: 10.14778/3579075.3579078.
- 135 Wessel Radstok, Mel Chekol, and Yannis Velegrakis. Leveraging static models for link prediction in temporal knowledge graphs. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), pages 1034–1041. IEEE, IEEE, nov 2021. doi:10.1109/ICTAI52525.2021.00165.
- 136 Wessel Radstok, Melisachew Wudage Chekol, and Mirko Tobias Schäfer. Are knowledge graph embedding models biased, or is it the data that they are trained on? In Wikidata@ISWC, volume 2982. CEUR-WS.org, 2021. URL: https://ceur-ws. org/Vol-2982/paper-5.pdf.
- 137 Yves Raimond, Tristan Ferne, Michael Smethurst, and Gareth Adams. The BBC world service archive prototype. *Journal of web semantics*, 27:2–9, aug 2014. doi:10.2139/ssrn.3199103.
- 138 Hongyu Ren, Mikhail Galkin, Michael Cochez, Zhaocheng Zhu, and Jure Leskovec. Neural graph reasoning: Complex logical query answering meets graph databases. arXiv.org, abs/2303.14617, mar 2023. doi:10.48550/ARXIV.2303.14617.
- 139 Hongyu Ren, Weihua Hu, and Jure Leskovec. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, volume abs/2002.05969. OpenReview.net, feb 2020. URL: https://openreview.net/forum? id=BJgr4kSFDS.
- 140 Hongyu Ren and Jure Leskovec. Beta embeddings for multi-hop logical reasoning in knowledge graphs. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, volume abs/2010.11465, 2020. URL: https: //proceedings.neurips.cc/paper/2020/hash/ e43739bba7cdb577e9e3e4e42447f5a5-Abstract. html.
- 141 Petar Ristoski and Heiko Paulheim. Rdf2Vec: RDF graph embeddings for data mining. In Proceedings of the International Semantic Web Conference, volume 9981, pages 498–514. Springer, Springer International Publishing, 2016. doi: 10.1007/978-3-319-46523-4_30.
- 142 Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15):e2016239118, apr 2021. doi:10.1073/PNAS.2016239118.
- 143 Tim Rocktäschel, Sameer Singh, and Sebastian Riedel. Injecting logical background knowledge into embeddings for relation extraction. In *Proceedings of the 2015 conference of the north American*

Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1119–1129. Association for Computational Linguistics, 2015. doi:10.3115/V1/N15-1118.

- 144 Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Molformer: Large scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 2022. doi:10.21203/rs.3.rs-1570270/v1.
- 145 Tara Safavi and Danai Koutra. CoDEx: A comprehensive knowledge graph completion benchmark. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8328–8350. Association for Computational Linguistics, 2020. doi:10.18653/V1/ 2020.EMNLP-MAIN.669.
- 146 Hooman Peiro Sajjad, Andrew Docherty, and Yuriy Tyshetskiy. Efficient representation learning using random walks for dynamic graphs. arXiv preprint arXiv:1901.01346, abs/1901.01346, jan 2019. doi:10.48550/arXiv.1901.01346.
- 147 Apoorv Saxena, Adrian Kochsiek, and Rainer Gemulla. Sequence-to-sequence knowledge graph completion and question answering. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume abs/2203.10321, pages 2814– 2828, Dublin, Ireland, 2022. Association for Computational Linguistics. doi:10.18653/V1/2022. ACL-LONG.201.
- 148 Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *Proceedings of the European semantic web conference*, volume 10843, pages 593–607. Springer International Publishing, 2018. doi:10.1007/978-3-319-93417-4_38.
- 149 Edward W Schneider. Course modularization applied: The interface system and its implications for sequence control and data analysis. In *PsycEX-TRA Dataset.* ERIC, nov 1973.
- 150 Lena Schwertmann, Manoj Prabhakar Kannan Ravi, and Gerard de Melo. Model-agnostic bias measurement in link prediction. In *Findings* of the Association for Computational Linguistics: EACL 2023, Dubrovnik, Croatia, May 2-6, 2023, pages 1587–1603. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023. FINDINGS-EACL.121.
- 151 Pengpeng Shao, Dawei Zhang, Guohua Yang, Jianhua Tao, Feihu Che, and Tong Liu. Tucker decomposition-based temporal knowledge graph completion. *Knowledge-Based Systems*, 238:107841, feb 2022. doi:10.1016/J.KNOSYS. 2021.107841.
- 152 Yinghan Shen, Xuhui Jiang, Zijian Li, Yuanzhuo Wang, Chengjin Xu, Huawei Shen, and Xueqi Cheng. UniSKGRep: A unified representation learning framework of social network and knowledge graph. *Neural Networks*, 158:142–153, jan 2023. doi:10.1016/J.NEUNET.2022.11.010.
- 153 Baoxu Shi and Tim Weninger. ProjE: Embedding projection for knowledge graph completion. In Proceedings of the AAAI Conference on Artificial

Intelligence, volume 31, pages 1236–1242. AAAI Press, 2017. doi:10.1609/AAAI.V31I1.10677.

- 154 Khemraj Shukla, Mengjia Xu, Nathaniel Trask, and George E. Karniadakis. Scalable algorithms for physics-informed neural and graph networks. *Data-Centric Engineering*, 3:e24, 2022. doi:10. 1017/dce.2022.24.
- 155 Amit Singhal. Introducing the knowledge graph: things, not strings. Google Blog. https://www. blog.google/products/search/introducingknowledge-graph-things-not/, 2012. Accessed: 2022-07-03.
- 156 Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In Proceedings of the Advances in neural information processing systems, pages 926-934, 2013. URL: https: //proceedings.neurips.cc/paper/2013/hash/ b337e84de8752b27eda3a12363109e80-Abstract. html.
- 157 Ran Song, Shizhu He, Shengxiang Gao, Li Cai, Kang Liu, Zhengtao Yu, and Jun Zhao. Multilingual knowledge graph completion from pretrained language models with knowledge constraints. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7709–7721. Association for Computational Linguistics, 2023. doi: 10.18653/V1/2023.FINDINGS-ACL.488.
- 158 Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: a core of semantic knowledge. In Proceedings of the 16th international conference on World Wide Web, pages 697–706. ACM, 2007. doi:10.1145/1242572.1242667.
- 159 Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. A benchmarking study of embeddingbased entity alignment for knowledge graphs. Proc. VLDB Endow., 13(12):2326–2340, jul 2020. doi: 10.14778/3407790.3407828.
- 160 Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In Proceedings of the 7th International Conference on Learning Representations. OpenReview.net, feb 2019. URL: https://openreview.net/forum?id= HkgEQnRqYQ.
- 161 Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge graph embedding by relational rotation in complex space. In 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, 2019. URL: https://openreview.net/ forum?id=HkgEQnRqYQ.
- 162 Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM Conference* on Recommender Systems, pages 297–305. ACM, 2018. doi:10.1145/3240323.3240361.
- 163 Xiaoli Tang, Rui Yuan, Qianyu Li, Tengyun Wang, Haizhi Yang, Yundong Cai, and Hengjie Song. Timespan-aware dynamic knowledge graph embedding by incorporating temporal evolution. *IEEE Access*, 8:6849–6860, jan 2020. doi:10.1109/ ACCESS.2020.2964028.

- 164 Xing Tang, Ling Chen, Jun Cui, and Baogang Wei. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations. Information Processing & Management, 56(3):809– 822, may 2019. doi:10.1016/J.IPM.2019.01.005.
- 165 Yi Tay, Anh Luu, and Siu Cheung Hui. Nonparametric estimation of multiple embeddings for link prediction on dynamic knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 1243–1249. Association for the Advancement of Artificial Intelligence (AAAI), 2017. doi:10.1609/AAAI.V31I1.10685.
- 166 Yi Tay, Luu Anh Tuan, Minh C Phan, and Siu Cheung Hui. Multi-task neural network for non-discrete attribute prediction in knowledge graphs. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 1029–1038. ACM, 2017. doi:10.1145/3132847.3132937.
- 167 Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In International Conference on Machine Learning, volume 119, pages 9448–9457. PMLR, PMLR, 2020. URL: http://proceedings.mlr. press/v119/teru20a.html.
- 168 Vinh Tong, Dat Quoc Nguyen, Trung Thanh Huynh, Tam Thanh Nguyen, Quoc Viet Hung Nguyen, and Mathias Niepert. Joint multilingual knowledge graph completion and alignment. In Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 4646–4658. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.FINDINGS-EMNLP.341.
- 169 Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. Representing text for joint embedding of text and knowledge bases. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1499–1509. Association for Computational Linguistics, 2015. doi:10.18653/V1/D15-1174.
- 170 Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In International conference on machine learning, volume 48, pages 2071–2080. PMLR, JMLR.org, 2016. URL: http://proceedings.mlr.press/v48/ trouillon16.html.
- 171 Neil Veira, Brian Keng, Kanchana Padmanabhan, and Andreas G Veneris. Unsupervised embedding enhancements of knowledge graphs using textual associations. In *IJCAI*, pages 5218–5225. International Joint Conferences on Artificial Intelligence Organization, 2019. doi:10.24963/IJCAI.2019/ 725.
- 172 Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In International Conference on Learning Representations. OpenReview.net, 2018. URL: https:// openreview.net/forum?id=rJXMpikCZ.
- 173 Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57:78–85, 2014. doi:10.1145/ 2629489.

4:30 Knowledge Graph Embeddings: Open Challenges and Opportunities

- 174 Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, Ying Wang, and Yi Chang. Structure-augmented text representation learning for efficient knowledge graph completion. In *Proceedings of the Web Conference 2021*, pages 1737–1748. ACM, 2021. doi:10.1145/3442381.3450043.
- 175 Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. RippleNet: Propagating user preferences on the knowledge graph for recommender systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pages 417–426. ACM, mar 2018. doi:10.1145/3269206.3271739.
- 176 Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the* 2018 world wide web conference, pages 1835–1844. ACM, 2018. doi:10.1145/3178876.3186175.
- 177 Jingbin Wang, Wang Zhang, Xinyuan Chen, Jing Lei, and Xiaolian Lai. 3DRTE: 3d rotation embedding in temporal knowledge graph. *IEEE Ac*cess, 8:207515–207523, 2020. doi:10.1109/ACCESS. 2020.3036897.
- 178 Rui Wang, Bicheng Li, Shengwei Hu, Wenqian Du, and Min Zhang. Knowledge graph embedding via graph attenuated attention networks. *IEEE* Access, 8:5212–5224, 2019. doi:10.1109/ACCESS. 2019.2963367.
- 179 Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. KGAT: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 950–958. ACM, 2019. doi: 10.1145/3292500.3330989.
- 180 Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5329–5336. Association for the Advancement of Artificial Intelligence (AAAI), 2019. doi: 10.1609/AAAI.V33I01.33015329.
- 181 Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. KEPLER: A unified model for knowledge embedding and pre-trained language representation. Transactions of the Association for Computational Linguistics, 9:176–194, 2021. doi: 10.1162/TACL_A_00360.
- 182 Yashen Wang and Huanhuan Zhang. HARP: A novel hierarchical attention model for relation prediction. ACM Transactions on Knowledge Discovery from Data (TKDD), 15:1–22, 2021. doi: 10.1145/3424673.
- 183 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph and text jointly embedding. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1591–1601, Doha, Qatar, 2014. Association for Computational Linguistics. doi:10.3115/V1/D14-1167.
- 184 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of*

the AAAI conference on artificial intelligence, volume 28, pages 1112–1119. Association for the Advancement of Artificial Intelligence (AAAI), 2014. doi:10.1609/AAAI.V28I1.8870.

- 185 Zhihao Wang and Xin Li. Hybrid-TE: Hybrid translation-based temporal knowledge graph embedding. In 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pages 1446–1451. IEEE, IEEE, 2019. doi:10. 1109/ICTAI.2019.00205.
- 186 Yuyang Wei, Wei Chen, Zhixu Li, and Lei Zhao. Incremental update of knowledge graph embedding by rotating on hyperplanes. In 2021 IEEE International Conference on Web Services (ICWS), pages 516–524. IEEE, 2021. doi:10.1109/ICWS53863. 2021.00072.
- 187 Christopher Wewer, Florian Lemmerich, and Michael Cochez. Updating embeddings for dynamic knowledge graphs. arXiv.org, abs/2109.10896, sep 2021. doi:10.48550/arXiv.2109.10896.
- 188 Jiawei Wu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Knowledge representation via joint learning of sequential text and knowledge graphs. arXiv preprint arXiv:1609.07075, abs/1609.07075, sep 2016. doi:10.48550/arXiv.1609.07075.
- 189 Tianxing Wu, Arijit Khan, Melvin Yong, Guilin Qi, and Meng Wang. Efficiently embedding dynamic knowledge graphs. *Knowledge-Based Systems*, 250:109124, aug 2022. doi:10.1016/J.KNOSYS. 2022.109124.
- 190 Yanrong Wu and Zhichun Wang. Knowledge graph embedding with numeric attributes of entities. In Proceedings of the Rep4NLP@ACL, pages 132–136. Association for Computational Linguistics, 2018. doi:10.18653/V1/W18-3017.
- 191 Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of SIGIR 2019*, pages 285–294, 2019. doi:10.1145/3331184.3331203.
- 192 Han Xiao, Minlie Huang, Yu Hao, and Xiaoyan Zhu. TransG: A generative mixture model for knowledge graph embedding. arXiv preprint arXiv:1509.05488, abs/1509.05488, 2015. doi: 10.48550/arXiv.1509.05488.
- 193 Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Image-embodied knowledge representation learning. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pages 3140–3146. International Joint Conferences on Artificial Intelligence Organization, 2017. doi:10.24963/IJCAI.2017/438.
- 194 Ruobing Xie, Zhiyuan Liu, Maosong Sun, et al. Representation learning of knowledge graphs with hierarchical types. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2965–2971. IJCAI/AAAI Press, 2016. URL: http://www.ijcai.org/Abstract/16/421.
- 195 Chengjin Xu, Yung-Yu Chen, Mojtaba Nayyeri, and Jens Lehmann. Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2569–2578.

Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.NAACL-MAIN.202.

- 196 Jiacheng Xu, Xipeng Qiu, Kan Chen, and Xuanjing Huang. Knowledge graph representation with jointly structural and textual encoding. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pages 1318–1324. International Joint Conferences on Artificial Intelligence Organization, aug 2017. doi: 10.24963/IJCAI.2017/183.
- 197 Youri Xu, E Haihong, Meina Song, Wenyu Song, Xiaodong Lv, Wang Haotian, and Yang Jinrui. RTFE: A recursive temporal fact embedding framework for temporal knowledge graph completion. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5671–5681. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021. NAACL-MAIN.451.
- 198 Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In Proceedings of the 3rd International Conference on Learning Representations, 2015. URL: http://arxiv.org/abs/1412.6575.
- 199 Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575, dec 2014. doi:10.48550/arXiv.1412.6575.
- 200 Liang Yao, Chengsheng Mao, and Yuan Luo. KG-BERT: BERT for knowledge graph completion. CoRR, abs/1909.03193, sep 2019. doi:10.48550/ arXiv.1909.03193.
- 201 Siyu Yao, Ruijie Wang, Shen Sun, Derui Bu, and Jun Liu. Joint embedding learning of educational knowledge graphs. Artificial Intelligence Supported Educational Technologies, pages 209–224, 2020. doi:10.1007/978-3-030-41099-5_12.
- 202 Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for webscale recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 974– 983. ACM, 2018. doi:10.1145/3219819.3219890.
- 203 Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GNNExplainer: Generating explanations for graph neural networks. Advances in neural information processing systems, 32:9240–9251, 2019. URL: https: //proceedings.neurips.cc/paper/2019/hash/ d80b7040b773199015de6d3b4293c8ff-Abstract. html.
- 204 Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, Jun Zhao, Zheng Liu, Jing Li, and Maosong Sun. Recursive neural networks for complex relation extraction. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 945–951, 2014.
- 205 Chuxu Zhang, Huaxiu Yao, Lu Yu, Chao Huang, Dongjin Song, Haifeng Chen, Meng Jiang, and Nitesh V Chawla. Inductive contextual relation learning for personalization. ACM Transactions on Information Systems (TOIS), 39(3):1–22, jul 2021. doi:10.1145/3450353.

- 206 Fuxiang Zhang, Xin Wang, Zhao Li, and Jianxin Li. TransRHS: a representation learning method for knowledge graphs with relation hierarchical structure. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, pages 2987–2993, 2021. doi:10.24963/IJCAI.2020/413.
- 207 Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 353–362. ACM, 2016. doi:10.1145/2939672.2939673.
- 208 Hanwang Zhang, Zawlin Kyaw, Shih-Fu Chang, and Tat-Seng Chua. Visual translation embedding network for visual relation detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5532–5540. IEEE, 2017. doi:10.1109/CVPR.2017.331.
- 209 Wen Zhang, Jiaoyan Chen, Juan Li, Zezhong Xu, Jeff Z Pan, and Huajun Chen. Knowledge graph reasoning with logics and embeddings: Survey and perspective. arXiv preprint arXiv:2202.07412, abs/2202.07412, 2022. doi:10.48550/arXiv.2202. 07412.
- 210 Wen Zhang, Shumin Deng, Mingyang Chen, Liang Wang, Qiang Chen, Feiyu Xiong, Xiangwen Liu, and Huajun Chen. Knowledge graph embedding in e-commerce applications: Attentive reasoning, explanations, and transferable rules. In Proceedings of the 10th International Joint Conference on Knowledge Graphs, pages 71–79. ACM, 2021. doi:10.1145/3502223.3502232.
- 211 Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. Learning hierarchy-aware knowledge graph embeddings for link prediction. In AAAI, volume 34, pages 3065–3072. AAAI Press, apr 2020. doi:10.1609/AAAI.V34I03.5701.
- 212 Zhao Zhang, Fuzhen Zhuang, Meng Qu, Fen Lin, and Qing He. Knowledge graph embedding with hierarchical relation structure. In *EMNLP*, pages 3198–3207. Association for Computational Linguistics, 2018. doi:10.18653/V1/D18-1358.
- 213 Zhiyuan Zhang, Xiaoqian Liu, Yi Zhang, Qi Su, Xu Sun, and Bin He. Pretrain-KGE: learning knowledge representation from pretrained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 259–266. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.FINDINGS-EMNLP.25.
- 214 Cunchao Zhu, Muhao Chen, Changjun Fan, Guangquan Cheng, and Yan Zhang. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In Proceedings of the AAAI conference on artificial intelligence, volume 35, pages 4732–4740. Association for the Advancement of Artificial Intelligence (AAAI), 2021. doi:10.1609/AAAI.V35I5.16604.
- 215 Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural bellmanford networks: A general graph neural network framework for link prediction. Advances in Neural Information Processing Systems, 34:29476-29490, 2021. URL: https: //proceedings.neurips.cc/paper/2021/hash/ f6a673f09493afcd8b129a0bcf1cd5bc-Abstract. html.

4:32 Knowledge Graph Embeddings: Open Challenges and Opportunities

216 Amal Zouaq and Felix Martel. What is the schema of your knowledge graph? leveraging knowledge graph embeddings and clustering for expressive

taxonomy learning. In Proceedings of the international workshop on semantic big data, pages 1–6. ACM, 2020. doi:10.1145/3391274.3393637.

Knowledge Graphs for the Life Sciences: Recent Developments, Challenges and Opportunities^{*}

Jiaoyan Chen ⊠©

Department of Computer Science, University of Manchester, UK Department of Computer Science, University of Oxford, UK

Hang Dong ⊠©

Department of Computer Science, University of Oxford, UK

Janna Hastings 🖂 💿

Institute for Implementation Science in Health Care, University of Zurich, Switzerland School of Medicine, University of St. Gallen, Switzerland

Ernesto Jiménez-Ruiz 🖂 🏠 💿

City, University of London, UK SIRIUS, University of Oslo, Norway

Vanessa López ⊠ IBM Research Europe, Dublin, Ireland

Pierre Monnin ⊠ ☆ [©] Université Côte d'Azur, Inria, CNRS, I3S, France

Catia Pesquita ⊠[©] LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Petr Škoda 🖂 💿

Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague, Czechia

Valentina Tamma 🖂 🗈

Department of Computer Science, University of Liverpool, UK

— Abstract -

The term *life sciences* refers to the disciplines that study living organisms and life processes, and include chemistry, biology, medicine, and a range of other related disciplines. Research efforts in life sciences are heavily data-driven, as they produce and consume vast amounts of scientific data, much of which is intrinsically relational and graphstructured.

The volume of data and the complexity of scientific concepts and relations referred to therein promote the application of advanced knowledgedriven technologies for managing and interpreting data, with the ultimate aim to advance scientific discovery.

In this survey and position paper, we discuss

recent developments and advances in the use of graph-based technologies in life sciences and set out a vision for how these technologies will impact these fields into the future. We focus on three broad topics: the construction and management of Knowledge Graphs (KGs), the use of KGs and associated technologies in the discovery of new knowledge, and the use of KGs in artificial intelligence applications to support explanations (explainable AI). We select a few exemplary use cases for each topic, discuss the challenges and open research questions within these topics, and conclude with a perspective and outlook that summarizes the overarching challenges and their potential solutions as a guide for future research.

^{*} Authors are listed in alphabetic order with authors' contributions presented in the metadata block. Jiaoyan Chen, Hang Dong, Janna Hastings, and Valentina Tamma are corresponding authors.



© Jiaoyan Chen, Hang Dong, Janna Hastings, Ernesto Jiménez-Ruiz, Vanessa López, Pierre Monnin, Catia Pesquita, Petr Škoda, and Valentina Tamma; licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 5, pp. 5:1-5:33



Transactions on Graph Data and Knowledge **TGDK** Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

5:2 Knowledge Graphs for Life Sciences

2012 ACM Subject Classification Information systems \rightarrow Graph-based database models; Computing methodologies \rightarrow Knowledge representation and reasoning; Applied computing \rightarrow Life and medical sciences

Keywords and phrases Knowledge graphs, Life science, Knowledge discovery, Explainable AI Digital Object Identifier 10.4230/TGDK.1.1.5

Category Position

Funding Jiaoyan Chen: supported by the EPSRC project ConCur (EP/V050869/1).

Hang Dong: supported by the EPSRC project ConCur (EP/V050869/1).

Janna Hastings: supported by the School of Medicine of the University of St. Gallen.

Ernesto Jiménez-Ruiz: supported by the SIRIUS Centre for Scalable Data Access (Research Council of Norway, project 237889).

Catia Pesquita: funded by the FCT through LASIGE Research Unit (ref. UIDB/00408/2020 and ref. UIDP/00408/2020), and also partially supported project 41, HfPT: Health from Portugal, funded by the Portuguese Plano de Recuperação e Resiliência.

Acknowledgements We would like to thank Uli Sattler (University of Manchester) for proposing the topic of this paper and Terry Payne (University of Liverpool) for the useful comments on a previous draft. We would also like to thank the TGDK editors in chief for organizing this inaugural issue.

Received 2023-06-30 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

Authors' Contributions All authors participated in the planning and discussions of this work. JH and HD finished the abstract and "Introduction". VT, JC and EJR contributed to "Knowledge Graphs in the Life Sciences". VT contributed to the main part of "Knowledge Graph Construction and Management", with contributions of use cases from JC, HD, PM, EJR, and JH. VL and JC contributed to "Life Science Knowledge Discovery". PM, PS, HD, and CP contributed to "Knowledge Graphs for Explainable AI". HD, JC, and CP contributed to "Discussion and Conclusion" based on discussions with other team members. All authors contributed to the final revision of this paper.

1 Introduction

The term *life sciences* refers to those disciplines that study living organisms and life processes, and include chemistry, biology, medicine, and a range of other related areas. Research efforts in life sciences are increasingly data-driven, as they produce and consume vast amounts of scientific data, much of which is intrinsically relational and graph-structured.

Much of this data is large-scale, complex, and presents many interrelationships and dependencies, thus being well suited to be represented in graph structures. For this reason, graph-based technologies are frequently used in the life sciences, and these disciplines have been drivers and early adopters of innovative methods and associated technologies.

In this brief survey and position paper we discuss recent developments and advances in the use of graph-based technologies in life sciences, and set out a vision for how these technologies will impact these fields in future. We illustrate the contribution in this paper in Figure 1.

We consider Knowledge Graphs (KGs) and their associated technologies to broadly include

- (i) different forms of graph-based representations,
- (ii) the logical languages that assign explicit semantics to such representations, and their associated automated reasoning technologies, and
- (iii) machine learning approaches that ingest data in graph-based representations and that process these graph-based representations to perform some task, e.g., data analytics.

J. Chen et al.

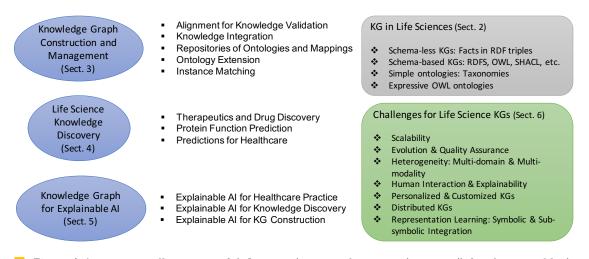


Figure 1 An overview illustration of definitions (upper right, in gray), topics (left column, in blue), use cases (middle), and challenges (bottom right, in green) for the research of KGs in the life sciences.

These different forms of graph-based representations can be further categorized based on the type of content represented. We therefore distinguish schema-less and schema-based Knowledge Graphs. More specifically, a typical KG contains either or both a schema part (terminologies or $TBox^1$) and a data part (facts, assertions, or ABox). The formal semantics of KGs can be expressed with the OWL ontology language².

In the remainder of this paper we will focus on three broad topic areas in which graph-based technologies have been used extensively, and we illustrate each area with some specific projects or use cases that guide our discussion and summary of the challenges that have been encountered.

- The construction and management of KGs to represent life science knowledge;
- The use of KGs and associated technologies in the discovery of new knowledge;
- The use of KGs in artificial intelligence applications to support explanations (eXplainable AI or XAI).

We then provide a summary of the general challenges across the topics, that include intrinsic characteristics of KGs (e.g., scalability, evolution, heterogeneity) and their operational aspects in the real world (e.g., human interaction, personalization, distributed setting, and representation learning). We present the challenges by means of use cases and the current research efforts that address them. It is worth mentioning that while we aim to focus on the life sciences, many of the topics and challenges discussed in this work, especially those of KG construction and management in Section 3, are general and applicable to KGs in other domains such as finance, e-commerce, material, and urban management [111, 31], etc. The KG-based problem modeling and solving approaches in life science knowledge discovery could be applicable for addressing many other use cases and problems in a broader domain of AI for scientific discovery [175, 60].

In the next section, we introduce several different categories of KGs as they have been used in life sciences. Thereafter in Sections 3-5, each of the above topics is described in a dedicated section together with a survey of recent advances. Finally, in Section 6 we synthesize the overarching challenges and trends into a perspective on the outlook for the future.

¹ We introduce a list of key terms relevant to Knowledge Graphs and Life Sciences in Appendix A.

 $^{^2}$ Web Ontology Language: https://www.w3.org/OWL/

2 Knowledge Graphs in the Life Sciences

KGs represent semantically-described real-world entities, typically through ontologies (vocabularies or schemas) [68, 61] and the data instantiating them, and thus provide descriptions of the entities of interest and their interrelations, by means of links to ontology classes describing them, organized in a graph [158]. KGs have been widely adopted in the life sciences, as can be seen in the composition of the Linked Open Data Cloud³, where life sciences represent one of the largest subdomains. A prominent example is the KG representing annotations regarding proteins by means of terms in the Gene Ontology describing different protein functions [4].

Whilst KGs are becoming increasingly popular in different domains including the life sciences, there is no single accepted definition of KG [43]. A KG can be formally described as a directed, edge-labeled graph $\mathcal{G} = (V, E)$, where V refers to the vertices or nodes, representing real-world entities of interest (e.g., proteins, genes, compounds, cellular components, but also pathways, biological processes and molecular functions, to name a few) while E refers to the edges in the graph, representing relationships or links between the entities in V (e.g., binds, associates, etc.). These may be represented as statements about entities in the form of RDF⁴ triples: (subject, predicate, object).

However, this formal definition only focuses on the components of KGs, but does not pose any constraint on what a KG should model or represent, and how. This is particularly true in life sciences, where the term *Knowledge Graph* has been used to refer to diverse graph data structures, typically interconnected, but often isolated.

Many of the everyday tasks faced by researchers in this domain require the systematic processing and integration of data and knowledge from data sources that are characterized by heterogeneous syntaxes and structures, formats, entity notation, schemas and scope, e.g., ranging from molecular mechanisms to phenotypes. Researchers in this area have been early adopters of Semantic Web and linked data approaches as a means to facilitate knowledge integration and processing to support tasks including semantic search, clinical decision support, enrichment analysis, data annotation and integration. However, a recent analysis of life science open data has identified several stand-alone data sources that exist in isolation, are not interlinked with other sources, and are schema-less (or use unpublished schemas), with limited reuse or mappings to other data sources [88]. Therefore, we can define a life sciences KG, following [130], as a data resource integrating one or more possibly curated sources of information into a graph whose nodes represent entities and edges represent relationships between two entities. This definition is consistent with other definitions found in the literature, e.g., [135].

These considerations underlie the reasons why KGs in life sciences can be of different types, and can be categorized across different dimensions. One of the most critical dimensions (in terms of support for complex queries and integration) is the categorization of KGs into schema-based and schema-less knowledge bases. In turn, the expressivity of the schema provides a further categorization criterion, depending on whether schemas are modeled as simple taxonomies (e.g., the NCBI taxonomy [154] included in the UMLS Metathesaurus [9]), RDFS⁵ vocabularies or (fully axiomatized) OWL ontologies. In particular, this paper refers to this broad definition of KGs, which we then divide into:

Schema-less KGs composed of only relational facts in the form of RDF triples. Examples include the PharmaGKB dataset, an integrated online knowledge resource capturing how genetic variation contributes to variation in drug response [180]. Note that many semantic networks (defined in Appendix A) could be assigned to this category as their triples form a multi-relational graph.

³ http://cas.lod-cloud.net

⁴ Resource Description Framework: https://www.w3.org/RDF/

⁵ RDF Schema: https://www.w3.org/TR/rdf-schema/

- Schema-based KGs composed of relational facts and their schema (meta information) in e.g., RDFS, OWL, and constraint languages such as SHACL⁶. Examples include Wikidata with its property constraints, and DBpedia with its DBpedia ontology. Whilst Wikidata and DBpedia are general-purpose KGs, they also include large-scale life science knowledge.
- Simple ontologies representing taxonomies. Notable examples include the tree structure of the UMLS Semantic Network⁷ and the International Classification of Diseases, version 10 (ICD-10) [182].
- Expressive OWL ontologies, with complex axioms beyond simple taxonomies. OWL ontologies may be composed of a TBox and an ABox. Depending on the expressivity of the axioms modeled in the ontology, i.e., the basic statements that an OWL ontology expresses, OWL ontologies can fall into one of the previous categories: for instance, an OWL ontology with just an ABox can be seen as the case above of a KG composed of relational facts alone. In this final category we include fully axiomatized OWL ontologies, e.g., with complex classes and property restrictions. Notable examples of these ontologies include SNOMED CT [38], the Gene Ontology [4, 28], and the Food Ontology (FoodOn)⁸.

3 Knowledge Graph Construction and Management

The adoption of KGs in the life sciences is motivated by the need for standardization of taxonomies and vocabularies to support the integration, exchange and analysis of data. More recently, richly annotated data is also being used in combination with machine learning methods for many applications, including helping to overcome issues related to the sparsity of data and helping to select promising candidates for reducing expensive and time-consuming physical experiments [64]. Graph-based machine learning approaches such as Graph Neural Networks have been applied to a number of life science tasks [49], including drug repurposing [120] and predicting polypharmacy side effects [196].

Given the diverse nature of the knowledge and tasks supported by KGs, the focus of stateof-the-art approaches has been the description of how individual KGs are developed within the specific domain [190], typically in terms of the specific approaches used for the development of the KG (e.g., data extraction process, relation extraction and entity discovery), rather than on the overall development process. More recently, some efforts have focused on providing an overview of development approaches and pipelines for the construction of KGs in the life sciences, and beyond [130, 164]. The process of constructing a KG depends heavily on:

- The type of data sources integrated and annotated by the KG, e.g., CSV files, public and proprietary data sources, structured databases, full-text publications, etc.
- The granularity of the KG to be constructed, e.g., schema-less KG, simple or expressive ontology.
- The usability expectations in downstream applications, e.g., the ability to customize and manipulate the graph to support different use cases, or the ease of consumption as input to machine learning methods [51].

A recent systematic review [164] surveyed different KG development approaches to determine a general development framework. The review identified six main phases that are common across different KG development approaches:

⁶ https://www.w3.org/TR/shacl/

⁷ https://uts.nlm.nih.gov/uts/umls/semantic-network/root

⁸ http://foodon.org

5:6 Knowledge Graphs for Life Sciences

- 1) Data source selection.
- 2) Ontology construction.
- 3) Knowledge extraction.
- 4) Knowledge ingestion and validation.
- 5) KG storage and inspection.
- 6) KG maintenance and evolution.

In the remainder of this section we will present the individual phases and the role they play in a KG development process by means of two use cases, where we illustrate the construction of KGs and discuss how these support knowledge integration and validation (Section 3.2). We then present some recent technical developments in Section 3.3, while Section 3.4 discusses open challenges for the construction and management of KGs.

3.1 Knowledge Graph Construction Phases

This section provides more details on the phases involved in the KG construction process, with the aim of identifying recent trends, rather than providing an exhaustive literature survey. These phases are discussed in order of execution, however the *ontology construction* phase can occur either together with the data source selection (if an ontology covering the domain of interest already exists or can be constructed through a set of given requirements) or as part of the *knowledge ingestion and validation* phase, where an ontology is built semi-automatically from the available data or through modularization and alignment of existing ontologies.

3.1.1 Data source selection

This phase identifies the data sources that are to be integrated by the KG, which in turn affects the choice of knowledge extraction techniques. Generally, life science KGs ingest knowledge from structured, semi-structured and unstructured data sources. By *structured* we refer to data modeled according to an existing structure, e.g., data in tables or public or proprietary reference (relational) databases such as UniProt [29] or ChEMBL [50]. Semi-structured data refer to, e.g., XML documents [116], whereas unstructured data refer to data that do not conform to a given structure, i.e., free-text sources, such as scientific publications from PubMed⁹. Data ingested from manually curated databases [130] and semi-structured sources constitute the foundation of a KG [51], generally defining the entities and some of the relations in the KG. This data is then further enriched by performing text mining on large-scale free text sources, in order to extract relationships, which is the objective of the *knowledge extraction* phase.

3.1.2 Ontology construction

The aim of this phase is to define a common, consensus-based, controlled vocabulary to describe the data in an *ontology* [146]. The existence of a common structure, or schema, supports querying, integration and reasoning tasks over the KG.

Traditional ontology engineering approaches are divided into top-down or bottom-up. Topdown approaches are based on more or less formal ontology engineering methodologies [45, 95, 131] or common practices [3] to build ontologies from a description of the domain elicited from domain experts [129], and/or by reusing or extending existing ontologies [82]. Ontology engineering methodologies define the ontology development process in terms of requirement analysis, entity

⁹ https://pubmed.ncbi.nlm.nih.gov

and property definitions, ontology reuse, validation and population. In contrast, bottom-up approaches utilize semi-automatic data driven techniques, e.g., ontology learning from text [110], and can be used to refine and validate an ontology. These approaches are discussed in more detail when presenting the *knowledge ingestion and validation* phase.

Whilst general purpose ontology engineering methodologies have evolved to be used in the development of KGs [139], a considerable number of ontologies in the life science domain have been built as part of the Open Biological and Biomedical Ontologies (OBO) Foundry effort,¹⁰ which defines a set of development principles for biological and biomedical ontologies and provides a suite of high-quality, interoperable, free and open source tools that support ontology development [115].

3.1.3 Knowledge extraction

Knowledge extraction refers to the identification of entities and their relations from the data sources, which is a crucial step in the development of a KG [164]. *Entity extraction* identifies entities from the various data sources selected using Natural Language Processing (NLP) approaches and text mining techniques to analyze and extract relevant information from large text corpora [178, 103, 71]. Named entity recognition (NER) supports the identification of named entities in text, such as drug names, diseases, or chemical compounds, and their classification according to pre-defined entity types [127]. NER approaches in the life sciences are typically based on labor intensive tasks such as the definition of generic (e.g., orthographic, morphological, or dictionary-based) and specific rules that are typically defined by experts, and are not easily applicable to other corpora [195]. There are a number of issues hindering these approaches:

- a) the pace of scientific discovery and the identification of new entities;
- b) the large number of synonyms and term variations associated with an entity; and
- c) entity identifiers that are composed of a mixture of letters, symbols and punctuation, often in large sentences [101].

More recent approaches have proposed the use of supervised machine learning methods (e.g., conditional random fields, or Support Vector Machines, SVMs, neural networks, and neural language models in particular) [112, 86, 35] either in isolation, or combined in hybrid approaches to improve accuracy [149].

Entity recognition generates entities that are isolated and not linked [164]. The goal of *Relation* extraction is to discover relationships of interest between a pair of entities, thus describing their interaction. Relation extraction is a necessary step for entities defined in semi-structured or unstructured sources, whereas structured data sources are characterized by explicitly identifiable relationships. Typical approaches for relation extraction include rule-based [75, 145, 144], supervised [106, 48] and unsupervised approaches [98, 130]. Rule-based relation extraction identifies keywords (based on existing ontologies or expert defined dictionaries) and grammatical patterns to discover relations between entities. Supervised relationship extraction methods utilize publicly available pre-labelled datasets (e.g., BioInfer [141] or BioCreative II [97]) to construct generalized patterns that separate positive examples (sentences implying the existence of a relationship) from negative ones. Supervised approaches include SVMs, Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) [6, 130]. Unsupervised relation extraction methods [113] have emerged to address the lack of scalability of supervised relation extraction methods, due to the high cost of human annotation. Unsupervised methods involve some form of clustering or statistical computation to detect the co-occurrence of two entities in the same text [130].

 $^{^{10} {\}tt https://obofoundry.org}$

5:8 Knowledge Graphs for Life Sciences

More recently, end-to-end approaches (End-to-End Relation Extraction – RE) have been used to tackle both tasks simultaneously. In this scenario, a model is trained simultaneously on both the NER and Relation Extraction objectives [74]. Furthermore, rule-based approaches can be combined with relation classification using specialized pre-trained language models adapted for life science domains, e.g., BioBERT [102], SapBERT [108], and RoBERTa-PM [104], to name a few. There is also a recent trend to probe and prompt pre-trained language models to extract relations (e.g., disease-to-disease, disease-to-symptoms) [187, 163].

3.1.4 Knowledge ingestion and validation

The aim of this phase is to ingest the entities and relationships extracted in a previous phase, which models knowledge from different sources. These entities and relations can be incomplete, ambiguous or redundant, and need to be appropriately aligned and integrated, and finally annotated according to the ontology constructed in phase 2.

Knowledge integration or fusion can critically improve the quality of data by performing *entity resolution*, i.e., the detection of different descriptions of the same real-world entity (also called entity matching, deduplication, entity linkage or entity canonicalization), prior to ingesting them in the KG. This reconciliation step is particularly crucial in the life sciences, where duplication can be caused by data modeled using different vocabularies or ontologies, or when data is extracted from literature sources that are rapidly changing. The severity of the ambiguity depends on the number of ontologies available for the domain. For instance, the number of gene vocabularies is far smaller than the number of disease vocabularies that could be present in the ingested datasets. Linking these entities requires costly alignment processing; in particular the alignment of disease entities is especially problematic given the number of different coding systems, whose conversion is often not trivial [51]. We further explore this issue in two of the use cases presented in Section 3.3, where we explore the problem of aligning vocabularies and ontologies through the use of mapping repositories and instance matching in automated clinical coding.

Entities are assigned unique identifiers (URI or IRI) that support the definition of bespoke namespaces, and support integration by reusing identifiers in related namespaces. Entity resolution is based on clustering similar entities together in a *block*, where similarity measures are used to detect duplicates [164]. Typical methods include sorted neighborhoods and traditional blocking; and machine learning methods are commonly used for similarity computation, e.g., feature vector computation [93].

This phase may also include the bottom-up construction of the ontology for those applications where a top-down approach is not feasible. Bottom-up approaches extract the relevant knowledge first, and then they construct the data schema / ontology based on the extracted data, typically using (semi-)automated methods, based on machine learning. Ontologies define the structure of the knowledge graph, which supports querying and data analytics. In bottom-up ontology development the structure of the knowledge graph is determined based on the extracted knowledge, thus providing a structure for this knowledge [69].

Often the construction of ontologies (either bottom-up or top-down) relies on the ability to correctly align and reuse entities defined across different domains and KGs. Furthermore, reuse of (or conformance to) existing upper level ontologies, e.g., BFO (Basic Formal Ontology) [3] provides the basis for the consistent and unambiguous formal definition of entities and relations that prevents errors in coding and annotation. The alignment of ontologies in life sciences and other domains is an active area of research, and we provide an overview of recent technical developments and challenges in Section 3.3.

Whilst bottom-up approaches, especially those based on alignment, are becoming more viable, especially given the support of language models, such as BERT [63], their performance is not always adequate for the task, as discussed in the second challenge in Section 3.4.

Knowledge enrichment and completion improve the KG quality by performing reasoning (KG materialization), inference [56] and optimization. Reasoning and inference support the assertion of new relations based either on logical reasoning (e.g., [128, 170]) or machine learning techniques (e.g., statistical relational learning or through embedding based link predictors for new concepts [34, 35, 66, 76] and node classifiers, also called KG refinement [135]). The extent and type of logical inferences depends on the expressivity of the ontology built in phase 2, or in a bottom-up fashion in this phase, together with any associated mappings. Description Logic formalisms, such as OWL, use logic-based reasoning for detecting and correcting incorrect assertions and ontology alignments [24].

3.1.5 KG storage and inspection

KGs need to be accessible to support a variety of different tasks, beyond the mere integration of different knowledge sources, and thus KG storage management [164, 142, 177] is an active area of research. Current KG storage mechanisms are divided into relation based stores (e.g., [1]) and native graph stores (e.g., [197]). Relational KG stores, either based on relational databases or through NOSQL databases and / or triple stores such as Jena TDB¹¹, have reached a considerable level of maturity and have been optimized in order to avoid common problems, e.g., a large number of null values in columns or optimized query performance [142]. Graph databases store nodes, edges and properties of graphs natively, and support query and graph mining tasks. Examples of state of the art implementations include Neo4J¹², GraphDB¹³, and RDFox¹⁴. The evolution of the performance of these systems has been the object of systematic studies [8], whereas [168] explicitly focuses on biomedical use cases.

Storage management has implications on the ways KGs support expressive queries for nodes and edges and visualization, to support data analysis, navigation and discovery of related knowledge [93, 162]. Graph databases often provide built-in tools for visualization, e.g., Neo4J, whereas different Javascript libraries (e.g., SigmaJS¹⁵) are available for developing visualization front ends. Support for complex queries is also either built in a graph database or a triple store by supporting the SPARQL query language [140, 197], or proprietary query languages such as Cypher [47], supported by Neo4J.

3.1.6 Knowledge maintenance and evolution

Given the rapid scientific development in the life sciences, and the consequent continuous update of ontologies for this domain, artifacts annotated with these ontologies can become outdated very quickly, and require some form of update (also called ontology extension). These update mechanisms need to be automated to ensure that they scale to the size of KGs. Automatic update approaches are based on the periodical detection and extraction of new knowledge that is then mapped to existing entities and relations in the KG [183].

Update mechanisms are typically based on the detection of *changes* [121] that can affect an ontology, e.g., addition, removal or modification of meta-entities (i.e., entities, relations and their definitions). These changes include renaming concepts and properties, setting domain and range restrictions, or setting a subsumption relation. To date, the most comprehensive account of

 $^{^{11} \}tt https://jena.apache.org/documentation/tdb/index.html$

¹² https://neo4j.com

¹³ https://graphdb.ontotext.com

¹⁴ https://www.oxfordsemantic.tech/product

¹⁵ https://github.com/jacomyal/sigma.js

5:10 Knowledge Graphs for Life Sciences

ontology change is given in [46], where change is described for different sub-fields, e.g., ontology alignment, matching and mapping, morphisms, articulation, translation, evolution, debugging, versioning, integration and merging; each with different requirements and implications. The study [137] further investigates the impact of biomedical ontology evolution on materialization.

Currently available tools and methodologies use (semi)-automated methods to perform many of the operations that trigger a change in an ontology and the consequent creation of a new version [54, 63]. Different ontology management platforms and portals mandate different principles and frameworks for handling ontology versioning (e.g., OBO foundry¹⁶ or BioPortal¹⁷), but these are typically implemented by ontology developers with limited tool support. Section 3.3 presents an example of automated ontology extension that relies on machine learning to cope with the scale of data.

3.2 Examples of Life Science KG Construction

In this section we provide two examples of life science KGs that illustrate in practice the phases composing the generic KG construction process discussed in Section 3; namely a KG for Pharmacogenomics, PGxLOD [118], and one for Ecotoxicological Analysis, TERA [124, 125].

Alignment for Knowledge Validation: An Example of Pharmacogenomics. As mentioned in Section 3, the task of aligning knowledge in KGs supports several downstream applications and domains. For instance, pharmacogenomics studies the influence of genetic factors on drug response phenotypes (e.g., expected effect, side effect). Hence, pharmacogenomics is of interest for personalized medicine. The atomic knowledge unit in pharmacogenomics is a ternary relationship between a drug, a genetic factor, and a phenotype. Such a relationship states that a patient being treated with the specified drug while having the specified genetic factor may experience the described phenotype. Semantic Web and KG technologies have been employed in this application domain, for example by building ontologies in which patients and pharmacogenomic knowledge are represented, and then using deductive reasoning mechanisms to conditionally recommend genetic testing before drug prescription [153]. However, the knowledge relevant to pharmacogenomics is scattered across several sources including reference databases such as PharmGKB, and the biomedical literature. Additionally, this knowledge may lack sufficient validation to be implemented in clinical practice. For example, some relationships may have only been observed in smaller cohorts of patients or in non-replicated studies. Hence, there is a need to align different sources of pharmacogenomic knowledge to detect additional evidence validating (or moderating) a knowledge unit. To this aim, the PGxLOD KG was proposed [118]. Automatic knowledge extraction approaches were applied on semi-structured and unstructured data from PharmGKB and the biomedical literature to represent their knowledge in the KG. Then, matching approaches were developed to align knowledge units from various sources [117, 119]. The resulting alignments outlined some agreements between PharmGKB and the biomedical literature, which was expected since PharmGKB is manually completed by experts after reviewing the literature. Interestingly, this automatic knowledge extraction pipeline could guide the manual review process by automatically pointing out studies confirming or mentioning a pharmacogenomic knowledge unit.

Knowledge Integration: An Example of Ecotoxicological Analysis. In ecotoxicological analysis, data and knowledge from different domains such as chemistry and biology are often needed. These are usually located in different sources such as spreadsheets or CSV files for local experimental results, open databases for public research results, and ontologies for domain knowledge. Thus knowledge integration becomes a critical and fundamental challenge before real analysis can be conducted. In the study by Myklebust et al. [124, 125], which aims to predict adverse biological

¹⁶ http://www.obofoundry.org/principles/fp-004-versioning.html

¹⁷ https://bioportal.bioontology.org

effects of chemicals on species, a toxicological effect and risk assessment KG named TERA was constructed for knowledge integration. TERA includes three sub-KGs: (i) the Chemical sub-KG, which is constructed by integrating the vocabulary MeSH (Medical Subject Headings) with selective knowledge from two chemical databases PubChem and ChEMBL utilizing the chemical mappings in Wikidata; (ii) the Taxonomy sub-KG, which is constructed by integrating EOL (Environment Ontology for Livestock) and the NCBITaxon ontology utilizing NIBI-EOL mappings in Wikidata; and (iii) the ECOTOX sub-KG, which is composed of RDF triples transformed from experimental risk results and is aligned with the other two sub-KGs by the ontology alignment system LogMap [80] and the chemical mappings in Wikidata. Another example of knowledge integration is for drug repurposing, where the KG Hetionet¹⁸ is created by integrating 29 public resources, including biomedical KGs and other types of data [67].

3.3 What has been done: recent technical developments

Given the many existing ontologies in life sciences, e.g., ontologies available in the OBO Foundry collection or in BioPortal [132], KG construction usually involves the reuse, alignment, and enrichment of state-of-the-art ontologies. The existing ontologies in life sciences need to be updated given the new discoveries in the field. This is broadly a key issue in the management, maintenance, and evolution of ontologies. We select a few promising use cases below to highlight some recent developments that support the KG construction in the life sciences.

Repositories of Ontologies and Mappings. Ontologies and their mappings play a central role in semantically enabled products and services consumed by life science companies, academic institutions and universities, as highlighted by the Pistoia Alliance ontology mapping project [58].¹⁹ Ontology mappings are essential in knowledge graph construction tasks to bridge the knowledge provided by different ontologies and expand their coverage. Ontology mappings can also play a key role when identifying the right ontologies to be reused as they will enable the retrieval of the relevant (overlapping) ontologies for the domain of interest. For this reason, a number of notable efforts in life sciences have created large repositories of ontologies and mappings to serve the research within the community. Prominent examples include the UMLS Metathesaurus [9], BioPortal [132, 152], MONDO [172], and the EBI services: OLS [174], OxO [84] and the RDF platform [85]. The UMLS Metathesaurus is a comprehensive effort for integrating biomedical ontologies through mappings. In its 2023AA version, it integrates more than two hundred vocabularies, with more than 3 million unique concepts and more than 15 million concept names. BioPortal is a repository containing more than 1,000 biomedical ontologies and more than 79 million lexically computed mappings among them (as of July 13, 2023). The Mondo Disease Ontology (MONDO) is a manually curated effort to harmonize and integrate disease conceptualizations and definitions across state-of-the-art ontologies (e.g., HPO [96], DO [155], ICD, SNOMED CT, etc.). The services provided by the European Bioinformatics Institute (EBI) also deserve a special mention. The Ontology Lookup Service (OLS) has become a reference to explore the latest versions of more than two hundred ontologies via its graphical interface or programmatically via its API. OxO is a repository of ontology mappings and cross-references extracted from the OLS and UMLS. OxO allows users to visually traverse the graph of mappings to identify additional potential mappings beyond direct ones (i.e., multi-hop mappings). Finally, the EBI RDF platform provides a unified KG with all the RDF resources at the EBI. Complementary to the efforts from the life sciences, the Semantic Web

¹⁸ https://github.com/hetio/hetionet

¹⁹ https://www.pistoiaalliance.org/projects/current-projects/ontologies-mapping/

5:12 Knowledge Graphs for Life Sciences

has also contributed to the systematic evaluation of mappings in public repositories (e.g., [81, 44]) and mappings produced by automated ontology mapping systems (e.g., the Ontology Alignment Evaluation Initiative (OAEI) [138]). Automatically generated mappings of high quality have the potential to be integrated within the aforementioned repositories and hence, the OAEI has always had a special focus on life science test cases with evaluation tracks like Anatomy [39], LargeBio [83], Phenotype [59] and the newly created track BioML [64]. The Simple Standard for Sharing Ontological Mappings (SSSOM) [114] represents a joint effort between the life sciences and Semantic Web communities to facilitate the exchange of mappings across different parties and repositories, while keeping the provenance and other relevant characteristics of the mappings.

Ontology Extension. Ontology extension in life sciences aims to connect new concepts and their relations to an ontology from updated sources, e.g., scientific papers in PubMed and chemical information in PubChem²⁰. Manual ontology extension, while essential for the development of gold standard resources, is not scalable to the full scope of large domains due to its high cost and low efficiency, and sometimes is even unfeasible as human beings may not be able to review the quantities of new information at the rate they become available. Thus machinelearning-based, automated methods are needed. One recent example is the use of deep learning, specifically a Transformer-based model, to categorize new chemical entities within the ChEBI ontology²¹ [53]. In addition, recent studies have explored enriching SNOMED CT by mining new concepts from texts [35] and placing them into the ontology [109, 34]. A new concept can be identified by NIL entity linking, i.e., exploring unlinkable mentions, usually through setting a "linkable" score threshold or through classification [35]. Resolution and disambiguation of NIL mentions with clustering can help to represent NIL entities [66, 91]. For concept placement, similar to the aforementioned CHEBI ontology extension [53], machine learning, especially in the form of Transformer-based deep learning, has been applied to predict subsumption relations between a new concept and the existing concepts. Complex concepts in OWL ontologies that contain logical operators (e.g., existential quantifier and conjunction in SNOMED CT) can be supported in subsumption prediction [23] and new concept placement [34]. Another group of studies use post-coordination or formalising a new term with existing concepts and attributes [16, 92], which is similar to composing subsumption axioms with complex concepts. The methods include using lexical features [92], word embeddings and KG embeddings [16]. Pre-trained and Large Language Models, through fine-tuning, zero-shot and few-shot prompting have the potential to support the mining [35] and placement of new concepts (e.g., by subsumption prediction [23, 65]).

Instance Matching: Automated Clinical Coding. A main source for patients' KG construction is Electronic Health Records (EHR). Using medical ontologies as backbones, it is possible to add a layer of data by instance matching (or patient matching) through *Clinical Coding*. Clinical coding is the task of transforming medical information in EHR into structured codes described in medical ontologies [36], e.g., ICD and SNOMED CT. Recent approaches mainly formulate the problem as a multi-label classification problem. Various neural network architectures have been proposed and knowledge plays a key role to enhance the neural architectures [36, 79]. Pre-trained language models, e.g., BERT [32], have been applied to clinical coding and gradually achieved better results with adapted modeling methods and more advanced language models, e.g., PLM-ICD [70] with RoBERTa-PM [104], according to studies [36, 42, 78]. Other studies formulate the task as a Named Entity Recognition and Linking (NER+L) problem, by extraction of concepts and linking them with the ontologies [36]. Overall, the recent progress in clinical coding, along with

²⁰https://pubchem.ncbi.nlm.nih.gov/

²¹ https://www.ebi.ac.uk/chebi/

the advent of Large Language Models (LLMs) suggests a trend in this area for patients' KG construction from EHR. However, there is still room for improvement in knowledge integration to better address explainability (see Section 5 for more details) and in zero-shot learning problems, i.e., for classifying into rare codes or concepts [36, 42, 79]. There are also further recent examples of instance matching with EHR data, including the works [15, 166].

3.4 What are the challenges?

KG construction and management often play a fundamental role in supporting life sciences with computation. There are still quite a few technical challenges, and many of the current tools and algorithms can be improved by modern machine learning and AI techniques. Here we present some critical and fundamental technical challenges.

- How to construct a customized KG? For a specific application, we often need to extract relevant data and knowledge from multiple sources, and at the same time integrate extracted knowledge from different sources. Considering a case study of personal health assistance, a customized KG with knowledge of at least exercise (sports), food, disease and medicine are required, while fine-grained knowledge of these aspects will lie in different domain KGs. The key challenge for integrating different ontology modules lies in estimating the semantic similarity and discovering the equivalence of two knowledge elements with their contexts considered, as well as the subsequent refinement like KG completion and knowledge representation canonicalization. Adequate tool support to minimize manual curation but enabling the user involvement when required is also paramount (e.g., [105]).
- How to ensure adequate performance using machine learning based approaches for automated KG construction? At the TBox level, the state-of-the-art alignment between classes (especially for subsumption relations) seems to not yet be achieving good enough performance, as reflected in recent biomedical ontology alignment benchmarking [64]. At the ABox level, predicting missing facts for practical KG construction expects high precision (e.g., beyond 90% or 95%) but only a few relations can be populated with a precision above 80% using prompt learning with BERT as evaluated in [173]. This is also the case to associate patients' EHR (as a part of ABox) with clinical codes or concepts in medical ontologies, where a micro F_1 score is below 60% [36]. Learning subsymbolic representations (see defined in Appendix A) of KG and data sources may help address the challenge. Transformer-based language models have achieved great performance in recent years. Among them, pre-trained language models such as BERT have been applied for KG construction with a promising performance achieved (see e.g., the package DeepOnto [63]), while the more recent and more powerful generative language models like GPT series [13] have not been well applied at the time of writing, especially in the life science domain.
- **How to ensure reliable semi-automated deep learning-based KG construction with human interaction?** Many tasks in the KG life cycle unavoidably rely on human experts to achieve consensus on reliable knowledge; on the other hand, as the automated KG construction process is growing opaque with deep learning methods, it is important to ensure trustworthiness and reliability [191]. Apart from enhancing performance metrics with novel methods, results with certain explainability are needed, for example, highlighting key parts in the data input when they are used as sources for KG construction. We discuss other aspects of explainability with KG, on life science knowledge discovery and healthcare decision making, in Section 5. Human-in-the-loop learning design for explainable KG construction may ensure the use of experts' knowledge for the task across the KG life cycle, which still remains a challenge for future research [191].

4 Life Science Knowledge Discovery

Research into AI technologies – including machine learning and KG-based reasoning – to accelerate the pace of scientific discovery is an emerging and rapidly developing field. The challenge lies in assisting scientists to uncover new knowledge and solutions, such as discovering novel therapeutic opportunities, identifying candidate molecular drugs to treat complex diseases or alternatively new uses for existing drugs, and supporting more personalized predictions.

Knowledge Graphs are powerful tools for representing complex biomedical knowledge, including molecular interactions, signaling pathways, disease co-morbidities, and more. Overviews of graph representation learning in biomedicine for healthcare applications and polypharmacy tasks are presented in [107] and [52] respectively. In graph representation learning, the graph's topology is leveraged to create compact vector embeddings. Through nonlinear transformations, highdimensional information about a node's graph neighborhood is distilled into low-dimensional vectors, where similar nodes are embedded close together in the vectorial space. Embeddings have been shown to be valuable for handling numerous relations in a KG while efficiently exploiting relation sparsity using vector computations. These optimized representations are subsequently used to train downstream models for various tasks, such as predicting property values of specific nodes (e.g., protein function), predicting links between nodes (e.g., binding affinity between molecules and protein targets), or performing classification tasks (e.g., predicting the toxicity profile of a candidate drug, or risk of readmission for a patient).

It is worth mentioning that among the existing works for life science knowledge discovery, different kinds of KGs have been exploited. The schema-less KG can be used to model different kinds of interaction between instances such as proteins and drugs; the taxonomy alike simple ontology is often used to represent concepts and their hierarchy such as protein functions defined in the gene ontology, chemical compounds, species, and diseases; expressive OWL ontologies and schema-based KGs can be used to model complex logical relationships between concepts, besides simple interaction between instances. Such diverse knowledge representation capabilities make KGs more flexible in modeling the input data and prediction targets of different knowledge discovery tasks, than graphs and tabular data that are widely used in previous pure machine learning-based methods.

In the following, we present some typical use cases, where machine learning techniques (including graph representation learning and language models) are applied over KGs built from diverse sources and domain ontologies, to facilitate life science discovery.

4.1 What has been done: use cases and their recent developments

Therapeutics and Drug Discovery: Learning a representation using multi-modal and heterogeneous knowledge. Drug discovery entails exploring an extremely large space of potential drug candidates. AI can help to accelerate this process by narrowing down the most promising candidates before expensive experimentation. The key to leveraging predictive and generative models for candidate solution generation lies in learning an effective multi-modal representation of protein targets, molecules and diseases among others. Recent research has focused on applying language models over large databases of proteins or molecules for self-supervised representation learning, such as ESM [148] and ProteinBERT [10] for protein sequences, or Molformer for the molecule simplified molecular-input line-entry system (SMILES) [151]. These models have exhibited remarkable success in tasks such as predicting protein interactions, binding affinity between drugs and targets, and protein functions and structures. However, these existing pre-trained sequence-based models often neglect to incorporate background knowledge from diverse sources, for example, biological structural knowledge.

Nonetheless, recent research indicates that incorporating existing expressive factual knowledge can improve results in downstream machine learning tasks. To enhance Protein Language Models (PLM), approaches such as OntoProtein [192] and KeAP [194] use a KG of protein sequences augmented with textual annotations from the Gene Ontology (GO). OntoProtein was the first to inject gene ontology descriptions into a PLM for sequences to predict protein interactions, function and contact prediction. OntoProtein proposes to reconstruct masked amino acids while minimizing the embedding distance between the contextual representation of proteins and associated knowledge terms. Similarly, ProtST [186] uses a dataset of protein sequences augmented with textual property descriptions from biomedical texts and jointly trains a PLM with a biomedical language model.

Knowledge Graphs are suitable data models for expressing heterogeneous knowledge and facilitating end-to-end learning [181]. An entity in a KG can have multiple attributes with different modalities - where each modality provides extra information about the entity - as well as relations to and from entities in other sources. Graph Neural Networks (GNN) have been used to capture inter-dependencies and diverse types of interactions between heterogeneous entity types and multimodal attributes in KGs [100]. They achieve this by iteratively aggregating information from neighboring nodes (through a process called message passing) and employing scoring functions to optimize the learned embeddings for downstream tasks. Otter-Knowledge [100] incorporates a heterogeneous KG (schema-based, containing concepts and their attributes) from diverse sources and modalities, i.e., each node has a particular mode that qualifies its type (text, image, protein sequence, molecule, etc.) and initial embeddings for each node are computed based on their modality. A GNN is then used to enrich protein and molecule representations and train a model to produce final node embeddings. The model is able to produce representations for entities that were not seen during training and achieve state-of-the-art results in the Therapeutic Data Commons (TDC) benchmarks [73] for drug-target binding affinity prediction. TxGNN [72] uses a GNN pre-trained on a large heterogeneous, multi-relational KG of diseases and therapeutic candidates constructed from various knowledge bases. TxGNN obtains a signature vector for each disease based on its neighboring proteins, exposure and other biomedical entities to compute a disease similarity and predict drug indication/contraindication for poorly characterized diseases.

Protein Function Prediction with the Gene Ontology. Conducting physical experiments for identifying protein functions is time and resource consuming. With the development of machine learning, protein function prediction (which is the task of predicting a given protein with multiple and potentially hierarchical classes – functions – defined in GO) has been widely investigated in recent years [193, 171]. A large part of these works such as GOLabler [189] focus on exploring feature extraction, feature ensemble, and automatic feature learning of the proteins. For example, GO-Labler [189] utilizes five kinds of different protein sequence information while DeepGraphGO [188] builds a network of proteins and learns protein features via a Graph Neural Network. Recent methods attempt to further exploit inter-function (class) relationships that are defined in GO for better performance. For example, DeepGOZero [99] and HMI [185] use formal semantics including the class hierarchy, class disjointness axioms and complex class restrictions in OWL as additional constraints for training the multi-label classifier for protein function prediction. Protein function prediction is a representative multi-label classification problem where complex relationships of the labels are defined in a KG and can be used for performance augmentation. It is quite common in machine learning applications in the life sciences, such as the above mentioned automated clinical coding where the codes' semantics are modeled by the ICD ontology, and ecotoxicological effect prediction where the multiple affected species of a chemical to predict form a taxonomy.

Predictions for Healthcare using Ontologies with Clinical Data. Digital Healthcare involves predictions using clinical data and ontologies, including diagnosis (e.g., rare diseases) and procedure

5:16 Knowledge Graphs for Life Sciences

predictions (e.g., ICU readmissions). A related concept is personalized medicine, which is achieved through the matching and fusion of knowledge from diverse sources, and plays a significant role in the prediction tasks. This often involves matching multiple ontologies [156], integrating curated databases (e.g., pharmacogenomics, molecules and proteins knowledge bases), mining knowledge from scientific literature [184] and person-centered clinical knowledge extracted from EHR or claim data, with distinguishing risk factors or cohorts' demographics (e.g., age and gender), which could enhance predictions related to adverse effects [123] or rare diseases for which there are not enough labeled datasets [2]. For example, SHEPHERD [2] incorporates a multi-relational KG (extracted from PrimeKG [19]) of diseases, phenotypes and genes, and leverages patient simulated data to discover novel connections between patients' clinical, phenotype and gene information to accelerate the diagnoses of rare diseases. Knowledge-guided learning is achieved by training a GNN to represent each patient's subgraphs of phenotypes in relation to other gene, phenotype, and disease associations within the KG, such that embeddings are informed by all of the existing biomedical knowledge captured in the network topology.

The approach in [15] constructs a KG (using expressive OWL ontologies) to predict ICU (intensive care units) readmission risk by enriching EHR data with semantic annotations from various biomedical ontologies in BioPortal. These predictions are based on KG embedding, such as RDF2vec, OPA2vec, and TransE, and classical machine learning methods, such as Logistic Regression, Random Forest, Naive Bayes and Support Vector Machines. Drawing from the Health & Social Person-centric Ontology (HSPO) [165], which focuses on multiple clinical, social and demographic facets for a patient or cohort, the approach presented in [166] builds a person-centric KG (expressive OWL ontology with TBox and ABox) from structured and unstructured data in EHR). Subsequently, a representation learning approach using GNNs is used to predict readmissions to the ICU.

4.2 What are the challenges?

We present four of the open challenges to unlock the full potential of methods to advance knowledge discovery for the life sciences using KGs, based on the use cases above.

- How to incorporate the semantics from a KG in machine learning? Many life science knowledge discovery tasks are modeled as a machine learning classification problem, whose input and output labels have additional valuable information in one or multiple external KGs. The challenge lies in extracting this information, optionally encoding it into vector representations, and injecting that knowledge into machine learning and pre-trained language models. Doing this effectively remains an important open challenge especially for protein-related pre-trained language models [192, 186, 194]. Besides improving the accuracy in knowledge discovery, injecting semantics from KGs can also contribute to making the model more explainable (see Section 5), but to this end, much research is still required.
- How to deal with the long-tail phenomenon in machine learning with KGs? In machine learning classification for real-world life science knowledge discovery, the candidate labels often exhibit a long-tailed distribution, i.e., a small ratio of them are common with a large number of training samples available, while most of them are infrequent or even have never appeared before. For example, imbalance in training data may occur for rare diseases or adverse drug effects that affect only a small portion of the population [2, 72, 37]. KGs sometimes have encoded the relationships of the labels, and could be used to help train the model for predicting those long-tailed labels or enable the inference of such labels.
- How to create an efficient multi-modal representation of knowledge to enable discovery? Most current state-of-the-art methods build learned graph representations based on isolated modalities. Multimodal KGs can explicitly capture labeled nodes and edges, each with

well-defined meanings, across heterogeneous node types, relations and modalities (such as text, images, protein sequences, molecules fingerprints, diseases and more) [19, 100]. Incorporating KGs with multiple modalities for representation learning requires computationally scalable methods to compute the initial embeddings for each modality, as a preliminary step to learn computable representations of large knowledge. Furthermore, robust learning techniques are needed for generalizing the learned representations to nodes with unseen or missing modalities, thereby enabling the discovery of new knowledge. An example would be inferring properties of proteins for which only the sequence is known.

How to efficiently utilize and fuse heterogeneous datasets, such as human-curated domain knowledge bases, scientific literature and person-centered health records, for knowledge discovery? State of the art shows that representations can be enhanced by incorporating richer information available across different sources [72, 100, 156]. Bringing in more data during training is needed to learn representations that can be applied to a broader range of downstream prediction tasks. However, learning from large and diverse KGs requires addressing challenges such as alignment, noise handling, balancing rich expressive knowledge with scalability and dealing with knowledge inconsistency. Moreover, more robust learning methods are needed for generalizing the learned representation to multiple downstream tasks (e.g., knowledge-aware transfer, zero-shot and few-shot learning [22]). An important aspect in this regard is addressing the disparity between all of the knowledge accessible during pre-training and the knowledge accessible or relevant for downstream fine-tuning [72, 100].

5 Knowledge Graphs for Explainable AI

Machine Learning (ML) and Artificial Intelligence (AI) methods are widely employed to tackle complex problems in many domains, including life sciences such as chemistry or biomedicine. Yet many of those methods operate as a "black-box", not enabling domain experts to understand the reasoning behind their predictions [90]. This is a major concern, especially for applications in areas with a potential impact on human lives, or areas with legally enforced accountability or transparency [143]. Moreover, understanding the workings of AI methods is also crucial in the context of scientific applications, such as those described in Section 4, where explaining the prediction process can help elucidate natural phenomena [40].

One way to address this issue is to employ the methods of eXplainable Artificial Intelligence (XAI). Although this is a topic long explored in the AI research community, there is still no widely-accepted definition of explainability, with many terms being used interchangeably, such as interpretability, comprehensibility, understandability and transparency [7]. Barredo et al. define explainability as the ability of a model to make its functioning clearer to an audience [7]. A slightly different definition is given in the previous survey [55]: "an interface between humans and a decision maker that is at the same time both an accurate proxy of the decision maker and comprehensible to humans". Both definitions focus on the audience, for *whom* is the model explainable, but the second suggests an explanation is another artifact produced by a model or alongside the model.

There are two distinguishable audiences in the context of the life sciences: scientists (researchers) and healthcare practitioners [167]. For the first group, the explanation is used as a guide to understanding within life sciences research for scientific discovery. As a result, the explanation may exist in a well-bounded context of a hypothesis or research project. On the other hand, practitioners are involved directly in decisions with impact on healthcare. They need to consider the output of the model in an open context, and sometimes also to explain the output to a patient who is not a domain expert.

5:18 Knowledge Graphs for Life Sciences

A number of approaches for XAI emerge from the literature and broadly contain two parts: (1) transparent box design, which includes algorithms such as decision trees, where models can be directly interpreted by users and therefore an explanation of an output results in simply following the decision paths that relate input to output; (2) post hoc interpretability, which provides an explanation to a black-box model using additional methods such as probing, perturbing, or by constructing surrogate models for general ML or AI methods [90, 167].

Utilization of KGs can greatly enhance XAI qualities as KGs are ideal for improving the model's interpretability, explainability, and understandability. Some methods are directly built around KGs and thus take full advantage of them. Examples of those methods may include methods that are using paths [161], predicting links, or performing reasoning [33]. Other methods can be enhanced using the KG (e.g., [126]). Yet the enhancement effect greatly depends on the place where KGs are employed and iteratively applied: *pre-model* (e.g., KG construction, potentially multi-modal), *in-model* (e.g., integrating KG with machine learning models), and *post-model* (e.g., reviewing and updating KG by domain experts to be applied in the next iteration to enhance machine learning models and their explanability) [143]. For example in in-model use, a model can be pre-trained using a KG, and an example of a pre-trained language model is SapBERT [108], which utilizes synonyms in the UMLS Metathesaurus to further pre-train a BERT language model. This can not only be beneficial for performance [192], but can also potentially enhance post-model explanation since the trained features are aligned with the KG [143].

5.1 What has been done: use cases and recent developments

Explainable AI for Healthcare Practice. The utilization of AI in healthcare practice raises the concern of leaving life-critical decisions to black-box models [143, 167]. For example, in the field of precision medicine which aims at tailoring drug treatments and dosages to each patient, clinicians require more information from a model than a simple binary decision [7]. The interpretability and explainability of AI models is thus an essential characteristic to make outputs understandable and transparent. This would enforce both clinicians' and patients' trust in models by complementing (and not substituting) clinicians' explanations [20, 143, 167].

To illustrate, this direction has been envisioned for several healthcare scenarios. Explainable AI models could support the experts in finding clinical trials that are appropriate based on patient history [167]. Counterintuitive or unreliable predictions that could have serious consequences could be explained, and thus prevented [167, 14, 90]. Some also envision such models to be used to explain and debunk healthcare-related misinformation [143]. As aforementioned, it is noteworthy that different kinds of explanations should be employed depending on the target audience, e.g., scientific explanations for evidence or trace-based explanations for treatment [20].

Explainable AI for Knowledge Discovery. As introduced in Section 4, KGs can support knowledge discovery in life science, including the explainability of the process and the discovered units. In this view, Ritoski and Paulheim [147] explain that ontologies, linked data, and KGs are used in the interpretation step of a data mining process, e.g., for interpreting sequential patterns in patient data [77], or to describe subgroups in a semantic subgroup discovery process [169]. KGs can also serve both as the basis for knowledge discovery processes and the interpretation process. For example, Linked Open Data connecting drugs and adverse reactions can be analyzed with Hidden Conditional Random Fields to predict adverse drug reactions, where the paths from selected drugs to outcomes visually explain the prediction [87]. Similarly, Bresso et al. [12] leverage features extracted from KGs (interpretable features such as paths, neighbors, path patterns) and white box models (e.g., decision trees) to reproduce expert classifications of drugs causing or not specific adverse drug reactions. The rules extracted from the decision trees contain

features that provide explanations for the molecular mechanisms behind these adverse reactions according to experts. Sousa et al. [159] employ KGs to explain both protein-protein interaction predictions and gene-disease association predictions based on shared semantic aspects.

Explainable AI for KG Construction. The final use case considers the situation that XAI is applied to KGs themselves. We discussed the challenge to support human intervention in KG construction in Section 3.4. Recent KG construction gradually relies on data-driven, deep learning based methods to automatically induce knowledge from data. The deep learning models are opaque, and thus the process requires explainability. The resulting KG may not be accountable to be used for downstream applications. *Trustworthy KG engineering* is proposed in [191] to highlight the importance of embedding explainable AI and human intervention in the KG life cycle. XAI methods have been applied in many NLP related tasks (entity and relation extraction, entity resolution, link prediction, etc.) in KG construction from texts. The XAI methods rely either on feature-based explanations or knowledge-based explanations. While feature-based explanations from the data or the models' interpretation of the data, knowledge-based explanations aim to interpret the process with rules, reasoning paths, and structured contextual information. Rules and paths have mainly been used for explanation, especially for link prediction, a task comprehensively surveyed in [191].

5.2 What are the challenges?

- How to integrate KGs for better XAI, especially with recent deep learning and language model based methods? KG may provide better data provenance for the model output. This can ensure explainability for communicating the model to domain experts in data science applications [7]. In terms of recent generative LLMs, life science KGs, with careful curation based on scientific publications, may help to provide provenance data to the answers generated by LLMs. Studies need to understand to what extent, and how, LLMs can be applied to induce knowledge (e.g., by probing LLMs with biomedical ontologies [65]), which then may provide a foundation to create better approaches to integrate KGs with LLMs. Another area is neuro-symbolic methods which may provide models that are inherently more interpretable (see further discussions in Section 6.1). Also, regarding language models (especially LLMs), they are capable of generating fluent texts, which can potentially serve as textual explanation generators from symbolic knowledge for XAI. Meanwhile, a key issue is the hallucination of LLMs, and KGs may support better prompting, fine-tuning and interpretable inference of LLMs for higher decisiveness and trustworthiness [134].
- How to evaluate XAI methods that involve KG? How to measure the quality of explanations, to ensure they are corresponding to users? The majority (around 70%) of XAI studies for KG construction do not evaluate the quality of the explanations or only informally visualize or comment on a limited number of cases to show the intuitive outcome [191]. Also, an XAI method needs to consider the target audience, as the explainability is to be finally received by a group of humans [7]. For instance, only a small number of current approaches to XAI for KG construction involve a user study, human evaluation or task-specific metrics [191]. Evaluating the quality of explanations requires some expert evaluation performed as ex-post evaluation, and well-defined metrics are needed for this task. An example is in [57] to use a combination of users' scores for each predicted explanation in a KG link prediction task, where there are multiple possible explanations. More expert validated and automated evaluation methods and associated metrics are required for KG-related XAI.

5:20 Knowledge Graphs for Life Sciences

6 Discussion and Conclusion

In this work, we have summarized the recent developments of KG research in life science on three important topics – KG Construction and Management, Life Science Knowledge Discovery, and KG for XAI. While each topic has its specific challenges, there are some common challenges and trends for the life science KG research in general.

6.1 Overall challenges and trends

Meanwhile, more scalable and efficient knowledge retrieval, query and reasoning systems, including life science KGs and mapping repositories, are still worthy of investigation and development.

Evolution and Quality Assurance of KGs. KGs need to be updated as new data and knowledge are emerging, and the schema and facts can easily become outdated or less useful for existing applications in life sciences. In terms of KG construction, we discussed ontology extension as a use case to address the evolution issue or emergence of new concepts and relations, and also instance matching to extend new instances for the KG. Updating KGs is also a prerequisite for life science knowledge discovery and knowledge discovery methods should be able to support the evolution of KGs with e.g., the capabilities of continuous learning and zero-shot learning. Quality assurance is another issue for KGs, including the tasks of knowledge error detection and correction, knowledge completion, knowledge canonicalization, etc. On the one hand, more effective KG quality assurance methods and systems should be developed, including schema and constraint languages for quality verification and learning-based models for prediction (e.g., [24] combines both for fact correction); on the other hand, knowledge discovery methods should be robust to noisy KGs by investigating e.g., robust KG embeddings and multi-modal representation learning.

Heterogeneity in KGs: Multi-domain and Multi-modality. KGs contain heterogeneous information, which brings challenges to their construction, representation, and reasoning. Different schema and data in KGs can have different focuses in their scopes and domains. Integrating data of different domains for building *multi-domain* KGs is difficult with challenges in e.g., ontology and data matching. Besides, recent studies have explored integrating different modalities to construct *Multi-modal* KGs [26, 122, 176], for instance text [133], images [179], etc. One challenge to address is how to learn effective machine learning models over multi-modal KGs fused from different sources (patients' records, curated knowledge bases, and scientific literature) to support scientific discovery as well as KG construction and management. Another challenge is developing accurate and efficient knowledge representation approaches for texts and images in multi-modal KG construction. For example, careful consideration should be given to when to simply use an annotation property to associate an image with an entity, and when to use a property with specific semantics to connect an image and an entity.

Human Interaction and Explainability with KGs. In KG construction, human experts are required for many sub-tasks of KG construction and provide oversight [191]. In life science knowledge discovery, human experts are necessary to finally validate the predicted new knowledge. The whole process of interacting with KG in life sciences requires explainability, especially when sub-symbolic models (e.g., pre-trained language models) are used. How to generate clear explanations for human interaction and how to evaluate the quality of explanations remains a challenge, as well as how to achieve consensus regarding scientific understanding with automatically discovered knowledge when organizing knowledge in life science [129]. The recent growth of *Neuro-Symbolic methods* suggests their support for explainability [89, 90, 150]. A recent survey [90] summarizes

XAI in bioinformatics with a chapter on knowledge-based explanations, whereas Karim [89, Chapter 8] provides a neuro-symbolic framework for KG construction and utilization for medical experts' decision making in the cancer domain. The approach presented in [150] is another recent example of neuro-symbolic integration for image classification with KG-based XAI in the cultural heritage domain.

Personalized and Customized KGs. A key challenge for KG construction is customization, as we discussed in Section 3, to construct application-oriented KGs, where relevant sub-KGs have to be extracted for large-scale KGs (a.k.a. modularization) and integrated with other knowledge and data from different sources. Besides, many life science KGs are about individuals, e.g., patients in healthcare applications, where Personal Health KG enables the integration of instance-level (or patient-level) information and their computation is required [122]. An example is the Personal Health KG in [21] that supports the dietary recommendation for users, where the construction and population of the KG requires reusing and integrating existing ontologies, dietary guidelines, and time-series patient data. The other examples of KGs integrating patients' EHR data [166, 15] are presented in Section 4.1. In personal KG construction, personal data should be protected. KG scalability should also be considered in order to be used on small devices such as cellphones. This is still a big challenge that has been rarely considered in using KGs in the life sciences.

Distributed KGs. The value of healthcare data for improving clinical knowledge and standard of care and the potential of semantic technologies to further enhance it are well recognized. However, a responsible use of healthcare data at the global level (beyond each healthcare provider and even each country) must take into account both legal and ethical issues in data sharing, privacy and security. Distributed knowledge graphs can mitigate these issues, by allowing for access control and privacy protection. Furthermore, distributed knowledge graphs can also address the challenges of scientific data ownership and stewardship by enabling the decentralized publishing of high quality data. Several approaches for federated querying and embedding of knowledge graphs have been proposed in recent years [25, 136, 157], however a wide adoption of semantic technologies in healthcare is still lacking, with a proliferation of terminological standards and a disconnection between data and meaning.

Representation Learning with KGs: Symbolic and Sub-symbolic Integration. Across the topics and use cases, we see the importance of transforming symbolic knowledge into sub-symbolic representations or combining both representations. The combination of both the neural and the traditional symbolic representation methods leads to a trend in neural-symbolic approaches in the field [11]. Recently, Pre-trained and Large Language Models provide new methods to transfer self-supervised learning from a vast amount of corpora to support KG construction, e.g., OntoGPT [17] and OntoLAMA [65]. LLMs are especially good at representing texts of life science publications in sub-symbolic spaces for semantic understanding. KGs may also provide a layer of explainability by validating the output of LLMs. A recent survey [134] proposes a roadmap for integrating LLMs and KGs. OntoProtein [192] is a recent example of how to integrate KGs into the process of pre-training LLMs in the bioinformatic domain, thus achieving improved results on protein-related knowledge discovery tasks. Also, geometry-informed representations of more formal KGs, especially in hyperbolic spaces or using complex geometric structures, e.g., [18, 99], can usually represent the structure of the KG with low dimensional vectors. Graph Neural Networks may also support the encoding of KG structures in a more explainable way with logical rules [30].

5:22 Knowledge Graphs for Life Sciences

6.2 Conclusion

Knowledge Graphs have become a popular and effective method to represent heterogeneous concepts, relations, and data in life sciences. They require scalable solutions to represent and reason with heterogeneous data and require constant updates. Throughout this work, we covered the main topics and their corresponding use cases of KGs in multiple life science domains such as protein analysis, drug discovery, ecotoxicology, and healthcare, and summarized the corresponding challenges. As new methods in knowledge representation appear, for instance the recent trends of human-in-the-loop, sub-symbolic knowledge representations, pre-trained and large language models, and neuro-symbolic integration, we envisage deeper applications of KGs to life science processes, that support the construction of more applicable KGs and the discovery of more reliable scientific knowledge, with explainability and human interaction better supported. KGs in combination with other modern machine learning and natural language processing techniques will become a foundation for AI for the life sciences.

— References ·

- 1 Daniel J Abadi, Adam Marcus, Samuel R Madden, and Kate Hollenbach. SW-Store: a vertically partitioned DBMS for Semantic Web data management. *The VLDB Journal*, 18:385–406, 2009. doi:10.1007/S00778-008-0125-Y.
- 2 Emily Alsentzer, Michelle M. Li, Shilpa N. Kobren, Undiagnosed Diseases Network, Isaac S. Kohane, and Marinka Zitnik. Deep learning for diagnosing patients with rare genetic diseases. *medRxiv*, 2022. doi:10.1101/2022.12.07.22283238.
- 3 Robert Arp, Barry Smith, and Andrew D. Spear. Building Ontologies With Basic Formal Ontology. The MIT Press, aug 2015. doi:10.7551/mitpress/ 9780262527811.001.0001.
- 4 Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000. doi:10.1038/75556.
- 5 Franz Baader, Ian Horrocks, Carsten Lutz, and Uli Sattler. An Introduction to Description Logic. Cambridge University Press, Cambridge, 2017. doi:10.1017/9781139025355.
- 6 Nguyen Bach and Sameer Badaskar. A review of relation extraction. *Literature review for Language* and Statistics II, 2:1–15, 2007.
- 7 Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion, 58:82–115, 2020. doi:10.1016/J.INFFUS.2019.12.012.
- 8 Maciej Besta, Robert Gerstenberger, Emanuel Peter, Marc Fischer, Michał Podstawski, Claude Barthels, Gustavo Alonso, and Torsten Hoefler. Demystifying graph databases: Analysis and taxonomy of data organization, system designs, and graph queries. ACM Comput. Surv., 2023. doi: 10.1145/3604932.
- 9 Olivier Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical

terminology. Nucleic Acids Res., 32(Database-Issue):267-270, 2004. doi:10.1093/NAR/GKH061.

- 10 Nadav Brandes, Dan Ofer, Yam Peleg, Nadav Rappoport, and Michal Linial. ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics*, 38(8):2102–2110, feb 2022. doi:10.1093/BIOINFORMATICS/BTAC020.
- 11 Anna Breit, Laura Waltersdorfer, Fajar J. Ekaputra, Marta Sabou, Andreas Ekelhart, Andreea Iana, Heiko Paulheim, Jan Portisch, Artem Revenko, Frank van Harmelen, and Annette ten Teije. Combining machine learning and semantic web: A systematic mapping study. ACM Computing Surveys, 2023. doi:10.1145/3586163.
- 12 Emmanuel Bresso, Pierre Monnin, Cédric Bousquet, François-Élie Calvier, Ndeye Coumba Ndiaye, Nadine Petitpain, Malika Smaïl-Tabbone, and Adrien Coulet. Investigating ADR mechanisms with explainable AI: a feasibility study with knowledge graph mining. BMC Medical Informatics Decis. Mak., 21(1):171, 2021. doi: 10.1186/S12911-021-01518-6.
- 13 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditva Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc. URL: https: //proceedings.neurips.cc/paper/2020/hash/ 1457c0d6bfcb4967418bfb8ac142f64a-Abstract. html.
- 14 Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Con-

ference on Knowledge Discovery and Data Mining, pages 1721–1730, 2015. doi:10.1145/2783258. 2788613.

- 15 Ricardo MS Carvalho, Daniela Oliveira, and Catia Pesquita. Knowledge Graph Embeddings for ICU readmission prediction. BMC Medical Informatics and Decision Making, 23(1):12, 2023. doi:10.1186/S12911-022-02070-7.
- 16 Javier Castell-Díaz, Jose Antonio Miñarro-Giménez, and Catalina Martínez-Costa. Supporting SNOMED CT postcoordination with knowledge graph embeddings. *Journal of Biomedical Informatics*, 139:104297, 2023. doi:10.1016/J. JBI.2023.104297.
- 17 J Harry Caufield, Harshad Hegde, Vincent Emonet, Nomi L Harris, Marcin P Joachimiak, Nicolas Matentzoglu, HyeongSik Kim, Sierra AT Moxon, Justin T Reese, Melissa A Haendel, et al. Structured prompt interrogation and recursive extraction of semantics (spires): A method for populating knowledge bases using zero-shot learning. arXiv preprint arXiv:2304.02711, 2023. doi: 10.48550/ARXIV.2304.02711.
- 18 Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. arXiv preprint arXiv:2005.00545, 2020. doi:10.18653/v1/2020.acl-main.617.
- 19 Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, 2023. doi:10.1038/s41597-023-01960-3.
- 20 Shruthi Chari, Oshani Seneviratne, Daniel M. Gruen, Morgan A. Foreman, Amar K. Das, and Deborah L. McGuinness. Explanation ontology: A model of explanations for user-centered AI. In Jeff Z. Pan, Valentina A. M. Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal, editors, The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, volume 12507 of Lecture Notes in Computer Science, pages 228–243, 2020. doi: 10.1007/978-3-030-62466-8_15.
- 21 Ching-Hua Chen, Daniel Gruen, Jonathan Harris, James Hendler, Deborah L McGuinness, Marco Monti, Nidhi Rastogi, Oshani Seneviratne, and Mohammed J Zaki. Semantic technologies for clinically relevant personal health applications. In Personal Health Informatics: Patient Participation in Precision Health, pages 199–220. Cham, 2022. doi:10.1007/978-3-031-07696-1_10.
- 22 Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Jeff Z Pan, Yuan He, Wen Zhang, Ian Horrocks, and Huajun Chen. Zero-Shot and Few-Shot Learning With Knowledge Graphs: A Comprehensive Survey. *Proceedings of the IEEE*, 2023. doi: 10.1109/JPR0C.2023.3279374.
- 23 Jiaoyan Chen, Yuan He, Yuxia Geng, Ernesto Jiménez-Ruiz, Hang Dong, and Ian Horrocks. Contextual semantic embeddings for ontology subsumption prediction. World Wide Web, pages 1–23, 2023. doi:10.1007/S11280-023-01169-9.
- 24 Jiaoyan Chen, Ernesto Jiménez-Ruiz, Ian Horrocks, Xi Chen, and Erik Bryhn Myklebust. An asser-

tion and alignment correction framework for large scale knowledge bases. *Semantic Web*, 14(1):29–53, 2023. doi:10.3233/SW-210448.

- 25 Mingyang Chen, Wen Zhang, Zonggang Yuan, Yantao Jia, and Huajun Chen. Federated knowledge graph completion via embedding-contrastive learning. *Knowledge-Based Systems*, 252:109459, 2022. doi:10.1016/J.KNOSYS.2022.109459.
- 26 Yong Chen, Xinkai Ge, Shengli Yang, Linmei Hu, Jie Li, and Jinwen Zhang. A survey on multimodal knowledge graphs: Construction, completion and applications. *Mathematics*, 11(8):1815, 2023. doi:10.3390/math11081815.
- 27 E. Coiera. Guide to Health Informatics, chapter Chapter 23 Healthcare terminologies and classification systems, pages 381–399. CRC Press, Taylor & Francis Group, Boca Raton, 2015. doi: 10.1201/b13617.
- 28 The Gene Ontology Consortium, Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, David P Hill, Raymond Lee, Huaiyu Mi, Sierra Moxon, Christopher J Mungall, Anushya Muruganugan, Tremayne Mushavahama, Paul W Sternberg, Paul D Thomas, Kimberly Van Auken, Jolene Ramsey, Deborah A Siegele, Rex L Chisholm, Petra Fey, Maria Cristina Aspromonte, Maria Victoria Nugnes, Federica Quaglia, Silvio Tosatto, Michelle Giglio, Suvarna Nadendla, Giulia Antonazzo, Helen Attrill, Gil dos Santos, Steven Marygold, Victor Strelets, Christopher J Tabone, Jim Thurmond, Pinglei Zhou, Saadullah H Ahmed, Praoparn Asanitthong, Diana Luna Buitrago, Meltem N Erdol, Matthew C Gage, Mohamed Ali Kadhum, Kan Yan Chloe Li, Miao Long, Aleksandra Michalak, Angeline Pesala, Armalya Pritazahra, Shirin C C Saverimuttu, Renzhi Su, Kate E Thurlow, Ruth C Lovering, Colin Logie, Snezhana Oliferenko, Judith Blake, Karen Christie, Lori Corbani, Mary E Dolan, Harold J Drabkin, David P Hill, Li Ni, Dmitry Sitnikov, Cynthia Smith, Alayne Cuzick, James Seager, Laurel Cooper, Justin Elser, Pankaj Jaiswal, Parul Gupta, Pankaj Jaiswal, Sushma Naithani, Manuel Lera-Ramirez, Kim Rutherford, Valerie Wood, Jeffrey L De Pons, Melinda R Dwinell, G Thomas Hayman, Mary L Kaldunski, Anne E Kwitek, Stanley J F Laulederkind, Marek A Tutaj, Mahima Vedi, Shur-Jen Wang, Peter D'Eustachio, Lucila Aimo, Kristian Axelsen, Alan Bridge, Nevila Hyka-Nouspikel, Anne Morgat, Suzi A Aleksander, J Michael Cherry, Stacia R Engel, Kalpana Karra, Stuart R Miyasato, Robert S Nash, Marek S Skrzypek, Shuai Weng, Edith D Wong, Erika Bakker, Tanya Z Berardini, Leonore Reiser, Andrea Auchincloss, Kristian Axelsen, Ghislaine Argoud-Puy, Marie-Claude Blatter, Emmanuel Boutet, Lionel Breuza, Alan Bridge, Cristina Casals-Casas, Elisabeth Coudert, Anne Estreicher, Maria Livia Famiglietti, Marc Feuermann, Arnaud Gos, Nadine Gruaz-Gumowski, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Philippe Le Mercier, Damien Lieberherr, Patrick Masson, Anne Morgat, Ivo Pedruzzi, Lucille Pourcel, Sylvain Poux, Catherine Rivoire, Shyamala Sundaram,

Alex Bateman, Emily Bowler-Barnett, Hema Bye-A-Jee, Paul Denny, Alexandr Ignatchenko, Rizwan Ishtiaq, Antonia Lock, Yvonne Lussi, Michele Magrane, Maria J Martin, Sandra Orchard, Pedro Raposo, Elena Speretta, Nidhi Tyagi, Kate Warner, Rossana Zaru, Alexander D Diehl, Raymond Lee, Juancarlos Chan, Stavros Diamantakis, Daniela Raciti, Magdalena Zarowiecki, Malcolm Fisher, Christina James-Zorn, Virgilio Ponferrada, Aaron Zorn, Sridhar Ramachandran, Leyla Ruzicka, and Monte Westerfield. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, mar 2023. doi:10.1093/genetics/iyad031.

- 29 The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1):D506-D515, nov 2018. doi: 10.1093/NAR/GKY1049.
- 30 David Jaime Tena Cucala, Bernardo Cuenca Grau, Egor V Kostylev, and Boris Motik. Explainable GNN-Based Models over Knowledge Graphs. In International Conference on Learning Representations, 2021. URL: https://openreview.net/ forum?id=CrCvGNHAIrz.
- 31 Shumin Deng, Chengming Wang, Zhoubo Li, Ningyu Zhang, Zelin Dai, Hehong Chen, Feiyu Xiong, Ming Yan, Qiang Chen, Mosha Chen, et al. Construction and applications of billion-scale pretrained multimodal business knowledge graph. In 2023 IEEE 39th International Conference on Data Engineering (ICDE), pages 2988–3002. IEEE, 2023. doi:10.1109/ICDE55515.2023.00229.
- 32 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019. doi:10.18653/V1/N19-1423.
- 33 Ivan Donadello, Mauro Dragoni, and Claudio Eccher. Persuasive explanation of reasoning inferences on dietary data. In Joint Proceedings of PROFILES 2019 and SEMEX 2019, 1st Workshop on Semantic Explainability (SemEx 2019), co-located with the 18th International Semantic Web Conference (ISWC '19), PROFILES-SEMEX 2019, pages 46-61. CEUR-WS, 2019. URL: http://ceur-ws.org/Vol-2465/semex_paper2.pdf.
- 34 Hang Dong, Jiaoyan Chen, Yuan He, and Ian Horrocks. Ontology enrichment from texts: A biomedical dataset for concept discovery and placement. In Proceedings of the 32nd ACM International Conference on Information & Knowledge Management, 2023. doi:10.1145/3583780.3615126.
- 35 Hang Dong, Jiaoyan Chen, Yuan He, Yinan Liu, and Ian Horrocks. Reveal the unknown: Out-ofknowledge-base mention discovery with entity linking. In Proceedings of the 32nd ACM International Conference on Information & Knowledge Management, 2023. doi:10.1145/3583780.3615036.
- 36 Hang Dong, Matúš Falis, William Whiteley, Beatrice Alex, Joshua Matterson, Shaoxiong Ji, Jiaoyan Chen, and Honghan Wu. Automated clinical coding: what, why, and where we are? NPJ digital medicine, 5(1):159, 2022. doi:10.1038/S41746-022-00705-7.

- 37 Hang Dong, Víctor Suárez-Paniagua, Huayu Zhang, Minhong Wang, Arlene Casey, Emma Davidson, Jiaoyan Chen, Beatrice Alex, William Whiteley, and Honghan Wu. Ontology-driven and weakly supervised rare disease identification from clinical notes. BMC Medical Informatics and Decision Making, 23(1):1–17, 2023. doi: 10.1186/S12911-023-02181-9.
- 38 Kevin Donnelly et al. SNOMED-CT: The advanced terminology and coding system for ehealth. In Medical and Care Computeries 3, volume 121 of Studies in health technology and informatics, pages 279–290. IOS Press, 2006.
- 39 Zlatan Dragisic, Valentina Ivanova, Huanyu Li, and Patrick Lambrix. Experiences from the anatomy track in the ontology alignment evaluation initiative. J. Biomed. Semant., 8(1):56:1– 56:28, 2017. doi:10.1186/S13326-017-0166-5.
- 40 Juan M Durán. Dissecting scientific explanation in ai (sxai): A case for medicine and healthcare. Artificial Intelligence, 297:103498, 2021. doi:10.1016/J.ARTINT.2021.103498.
- 41 Artur S d'Avila Garcez, Luís C Lamb, and Dov M Gabbay. Neural-symbolic learning systems, pages 35–54. Springer, 2009. doi:10.1007/ 978-1-4471-0211-3.
- 42 Joakim Edin, Alexander Junge, Jakob D. Havtorn, Lasse Borgholt, Maria Maistro, Tuukka Ruotsalo, and Lars Maaløe. Automated medical coding on MIMIC-III and MIMIC-IV: A critical review and replicability study. *CoRR*, abs/2304.10909, 2023. doi:10.48550/ARXIV.2304.10909.
- 43 Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. In Michael Martin, Martí Cuquet, and Erwin Folmer, editors, Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16) co-located with the 12th International Conference on Semantic Systems (SEMANTiCS 2016), Leipzig, Germany, September 12-15, 2016, volume 1695 of CEUR Workshop Proceedings. CEUR-WS.org, 2016. URL: https://ceur-ws.org/Vol-1695/paper4.pdf.
- 44 Daniel Faria, Ernesto Jiménez-Ruiz, Catia Pesquita, Emanuel Santos, and Francisco M. Couto. Towards Annotating Potential Incoherences in BioPortal Mappings. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig A. Knoblock, Denny Vrandecic, Paul Groth, Natasha F. Noy, Krzysztof Janowicz, and Carole A. Goble, editors, The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference. Proceedings, Part II, volume 8797 of Lecture Notes in Computer Science, pages 17–32, 2014. doi:10.1007/978-3-319-11915-1_2.
- 45 Mariano Fernandez, Asuncion Gomez-Perez, and Natalia Juristo. Methontology: from ontological art towards ontological engineering. In Proc. of the AAA197 Spring Symposium Series on Ontological Engineering, pages 33–40. Stanford, USA, 1997.
- 46 Giorgos Flouris, Dimitris Manakanatas, Haridimos Kondylakis, Dimitris Plexousakis, and Grigoris Antoniou. Ontology change: classification and

survey. *Knowl. Eng. Rev.*, 23(2):117–152, 2008. doi:10.1017/S0269888908001367.

- 47 Nadime Francis, Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Mats Rydberg, Petra Selmer, and Andrés Taylor. Cypher: An Evolving Query Language for Property Graphs. In Proceedings of the 2018 International Conference on Management of Data, SIGMOD '18, pages 1433–1445, 2018. doi:10.1145/3183713.3190657.
- 48 Katrin Fundel, Robert Küffner, and Ralf Zimmer. Relex—relation extraction using dependency parse trees. *Bioinformatics*, 23(3):365–371, 2007. doi:10.1093/BI0INFORMATICS/BTL616.
- 49 Thomas Gaudelet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy B R Hayter, Richard Vickers, Charles Roberts, Jian Tang, David Roblin, Tom L Blundell, Michael M Bronstein, and Jake P Taylor-King. Utilizing graph machine learning within drug discovery and development. Briefings in Bioinformatics, 22(6):bbab159, may 2021. doi: 10.1093/BIB/BBAB159.
- 50 Anna Gaulton, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, and John P. Overington. Chembl: A large-scale bioactivity database for drug discovery. Nucleic acids research, 40(D1):D1100–D1107, 2012. doi:10.1093/NAR/GKR777.
- 51 David Geleta, Andriy Nikolov, Gavin Edwards, Anna Gogleva, Richard Jackson, Erik Jansson, Andrej Lamov, Sebastian Nilsson, Marina Pettersson, Vladimir Poroshin, et al. Biological insights knowledge graph: an integrated knowledge graph to support drug development. *BioRxiv*, pages 2021– 10, 2021. doi:10.1101/2021.10.28.466262.
- 52 Aryo Pradipta Gema, Dominik Grabarczyk, Wolf De Wulf, Piyush Borole, Javier Antonio Alfaro, Pasquale Minervini, Antonio Vergari, and Ajitha Rajan. Knowledge graph embeddings in the biomedical domain: Are they useful? a look at link prediction, rule learning, and downstream polypharmacy tasks. arXiv preprint arXiv:2305.19979, 2023. doi:10.48550/ARXIV. 2305.19979.
- 53 Martin Glauer, Adel Memariani, Fabian Neuhaus, Till Mossakowski, and Janna Hastings. Interpretable Ontology Extension in Chemistry. Semantic Web Journal, 2022. doi:10.3233/SW-233183.
- 54 Anika Groß, Cédric Pruski, and Erhard Rahm. Evolution of biomedical ontologies and mappings: Overview of recent approaches. *Computational* and Structural Biotechnology Journal, 14:333–340, 2016. doi:10.1016/j.csbj.2016.08.002.
- 55 Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. ACM Comput. Surv., 51(5), aug 2018. doi:10.1145/3236009.
- 56 Ricardo Guimarães and Ana Ozaki. Reasoning in Knowledge Graphs. In International Research School in Artificial Intelligence in Bergen (AIB 2022), volume 99 of Open Access Series in Informatics (OASIcs), pages 2:1–2:31, 2022. doi:10.4230/OASICS.AIB.2022.2.

- 57 Nicholas Halliwell, Fabien Gandon, and Freddy Lécué. User scored evaluation of non-unique explanations for relational graph convolutional network link prediction on knowledge graphs. In Anna Lisa Gentile and Rafael Gonçalves, editors, K-CAP '21: Knowledge Capture Conference, Virtual Event, USA, December 2-3, 2021, pages 57–64, 2021. doi:10.1145/3460210.3493557.
- 58 I. Harrow, R. Balakrishnan, E. Jimenez-Ruiz, S. Jupp, J. Lomax, J. Reed, M. Romacker, C. Senger, A. Splendiani, J. Wilson, and P. Woollard. Ontology mapping for semantically enabled applications. *Drug Discovery Today*, may 2019. doi:10.1016/j.drudis.2019.05.020.
- 59 Ian Harrow, Ernesto Jiménez-Ruiz, Andrea Splendiani, Martin Romacker, Peter Woollard, Scott Markel, Yasmin Alam-Faruque, Martin Koch, James Malone, and Arild Waaler. Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. J. Biomed. Semant., 8(1):55:1–55:13, 2017. doi:10.1186/ S13326-017-0162-9.
- 60 J. Hastings. AI for Scientific Discovery. AI for everything series. CRC Press, Milton, 2023. doi:10.1201/9781003226642.
- 61 Janna Hastings. Primer on Ontologies. In Christophe Dessimoz and Nives Škunca, editors, *The Gene Ontology Handbook*, volume 1446, pages 3–13. Humana Press, SpringerOpen, New York, New York, NY, 2017. doi:10.1007/ 978-1-4939-3743-1_1.
- 62 Janna Hastings, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, and Christoph Steinbeck. Chebi in 2016: Improved services and an expanding collection of metabolites. Nucleic acids research, 44(D1):D1214– D1219, 2016. doi:10.1093/NAR/GKV1031.
- 63 Yuan He, Jiaoyan Chen, Hang Dong, Ian Horrocks, Carlo Allocca, Taehun Kim, and Brahmananda Sapkota. DeepOnto: A Python package for ontology engineering with deep learning. arXiv preprint arXiv:2307.03067, 2023. doi: 10.48550/ARXIV.2307.03067.
- 64 Yuan He, Jiaoyan Chen, Hang Dong, Ernesto Jiménez-Ruiz, Ali Hadian, and Ian Horrocks. Machine learning-friendly biomedical datasets for equivalence and subsumption ontology matching. In Ulrike Sattler, Aidan Hogan, C. Maria Keet, Valentina Presutti, João Paulo A. Almeida, Hideaki Takeda, Pierre Monnin, Giuseppe Pirrò, and Claudia d'Amato, editors, The Semantic Web ISWC 2022 21st International Semantic Web Conference, Proceedings, volume 13489 of Lecture Notes in Computer Science, pages 575–591, 2022. doi:10.1007/978-3-031-19433-7_33.
- 65 Yuan He, Jiaoyan Chen, Ernesto Jiménez-Ruiz, Hang Dong, and Ian Horrocks. Language model analysis for ontology subsumption inference. Findings of the Association for Computational Linguistics: ACL 2023, pages 3439–3453, 2023. doi: 10.18653/V1/2023.FINDINGS-ACL.213.
- 66 Nicolas Heist and Heiko Paulheim. NASTyLinker: NIL-Aware Scalable Transformer-Based Entity Linker. In Catia Pesquita, Ernesto Jiménez-Ruiz,

Jamie P. McCusker, Daniel Faria, Mauro Dragoni, Anastasia Dimou, Raphaël Troncy, and Sven Hertling, editors, *The Semantic Web - 20th International Conference, ESWC 2023, Hersonissos, Crete, Greece, May 28 - June 1, 2023, Proceedings, volume 13870 of Lecture Notes in Computer Science, pages 174–191. Springer, 2023.* doi:10.1007/978-3-031-33455-9_11.

- 67 Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, 6:e26726, 2017. doi:10.7554/eLife. 26726.
- 68 Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. The role of ontologies in biological and biomedical research: a functional perspective. *Briefings in Bioinformatics*, 16(6):1069–1080, apr 2015. doi:10.1093/BIB/BBV011.
- 69 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. Number 22 in Synthesis Lectures on Data, Semantics, and Knowledge. Springer, 2021. doi:10.2200/S01125ED1V01Y202109DSK022.
- 70 Chao-Wei Huang, Shang-Chi Tsai, and Yun-Nung Chen. PLM-ICD: Automatic ICD coding with pretrained language models. In *Proceedings of the* 4th Clinical Natural Language Processing Workshop, pages 10–20, 2022. doi:10.18653/v1/2022. clinicalnlp-1.2.
- 71 Chung-Chi Huang and Zhiyong Lu. Community challenges in biomedical text mining over 10 years: success, failure and the future. *Briefings in bioinformatics*, 17(1):132–144, 2016. doi: 10.1093/BIB/BBV024.
- 72 Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish Nadkarni, Benjamin Glicksberg, Nils Gehlenborg, and Marinka Zitnik. Zero-shot prediction of therapeutic use with geometric deep learning and clinician centered design. medRxiv, 2023. doi:10.1101/2023.03.19.23287458.
- 73 Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. Artificial intelligence foundation for therapeutic science. *Nature chemical biology*, 2022. doi:10.1038/s41589-022-01131-2.
- 74 Pere-Lluís Huguet Cabot and Roberto Navigli. REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381, Punta Cana, Dominican Republic, nov 2021. Association for Computational Linguistics. doi:10.18653/V1/2021.FINDINGS-EMNLP. 204.
- 75 Lawrence Hunter, Zhiyong Lu, James Firby, William A Baumgartner, Helen L Johnson, Philip V Ogren, and K Bretonnel Cohen.

Opendmap: an open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC bioinformatics*, 9(1):1–11, 2008. doi:10.1186/1471-2105-9-78.

- 76 Anastasiia Iurshina, Jiaxin Pan, Rafika Boutalbi, and Steffen Staab. NILK: Entity Linking Dataset Targeting NIL-Linking Cases. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22, pages 4069–4073, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/ 3511808.3557659.
- 77 Nicolas Jay and Mathieu d'Aquin. Linked data and online classifications to organise mined patterns in patient data. In AMIA 2013, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 16-20, 2013, 2013. URL: https: //knowledge.amia.org/amia-55142-a2013e-1. 580047/t-05-1.583941/f-005-1.583942/ a-224-1.584151/a-226-1.584146.
- 78 Shaoxiong Ji, Matti Hölttä, and Pekka Marttinen. Does the magic of BERT apply to medical code assignment? A quantitative study. *Computers in biology and medicine*, 139:104998, 2021. doi:10.1016/J.COMPBIOMED.2021.104998.
- 79 Shaoxiong Ji, Wei Sun, Hang Dong, Honghan Wu, and Pekka Marttinen. A unified review of deep learning for automated medical coding. arXiv preprint arXiv:2201.02797, 2022. doi:10.48550/ arXiv.2201.02797.
- 80 Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. LogMap: Logic-based and scalable ontology matching. In The Semantic Web-ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10, pages 273-288, 2011. doi:10.1007/ 978-3-642-25073-6_18.
- 81 Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Ian Horrocks, and Rafael Berlanga Llavori. Logicbased assessment of the compatibility of UMLS ontology sources. J. Biomed. Semant., 2(S-1):S2, 2011. doi:10.1186/2041-1480-2-S1-S2.
- 82 Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Ulrike Sattler, Thomas Schneider, and Rafael Berlanga Llavori. Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support. In *The Semantic Web: Research and Applications, 5th European Semantic Web Conference, Proceedings*, pages 185–199, 2008. doi:10.1007/978-3-540-68234-9_16.
- 83 Ernesto Jiménez-Ruiz, Christian Meilicke, Bernardo Cuenca Grau, and Ian Horrocks. Evaluating mapping repair systems with large biomedical ontologies. In Thomas Eiter, Birte Glimm, Yevgeny Kazakov, and Markus Krötzsch, editors, Informal Proceedings of the 26th International Workshop on Description Logics, Ulm, Germany, July 23 - 26, 2013, volume 1014 of CEUR Workshop Proceedings, pages 246-257. CEUR-WS.org, 2013. URL: https://ceur-ws.org/Vol-1014/paper_63.pdf.
- 84 Simon Jupp, Thomas Liener, Sirarat Sarntivijai, Olga Vrousgou, Tony Burdett, and Helen E.

Parkinson. OxO - A Gravy of Ontology Mapping Extracts. In Matthew Horridge, Phillip Lord, and Jennifer D. Warrender, editors, *Proceedings of the 8th International Conference on Biomedical Ontology (ICBO 2017)*, volume 2137 of *CEUR Workshop Proceedings*, 2017. URL: https://ceur-ws.org/Vol-2137/paper_27.pdf.

- 85 Simon Jupp, James Malone, Jerven T. Bolleman, Marco Brandizi, Mark Davies, Leyla J. García, Anna Gaulton, Sebastien Gehant, Camille Laibe, Nicole Redaschi, Sarala M. Wimalaratne, Maria Jesus Martin, Nicolas Le Novère, Helen E. Parkinson, Ewan Birney, and Andrew M. Jenkinson. The EBI RDF platform: linked open data for the life sciences. *Bioinform.*, 30(9):1338–1339, 2014. doi:10.1093/BIOINFORMATICS/BTT765.
- 86 Daniel Jurafsky and James H. Martin. Speech and Language Processing (3rd Edition). Online, 2023. URL: https://web.stanford.edu/ ~jurafsky/slp3/ed3book_jan72023.pdf.
- 87 Maulik R. Kamdar and Mark A. Musen. PhLeGrA: Graph Analytics in Pharmacology over the Web of Life Sciences Linked Open Data. In Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *Proceedings of the 26th International Conference on World Wide Web*, pages 321–329. ACM, 2017. doi:10.1145/3038912. 3052692.
- 88 Maulik R. Kamdar and Mark A. Musen. An empirical meta-analysis of the life sciences linked open data on the web. *Scientific Data*, 8, 2020. doi:10.1038/s41597-021-00797-y.
- 89 Md. Rezaul Karim. Interpreting black-box machine learning models with decision rules and knowledge graph reasoning. Dissertation, RWTH Aachen University, Aachen, 2022. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2022. doi:10.18154/RWTH-2022-07610.
- 90 Md Rezaul Karim, Tanhim Islam, Md Shajalal, Oya Beyan, Christoph Lange, Michael Cochez, Dietrich Rebholz-Schuhmann, and Stefan Decker. Explainable ai for bioinformatics: Methods, tools and applications. *Briefings in bioinformatics*, 24(5):bbad236, 2023. doi:10.1093/BIB/BBAD236.
- 91 Nora Kassner, Fabio Petroni, Mikhail Plekhanov, Sebastian Riedel, and Nicola Cancedda. EDIN: An end-to-end benchmark and pipeline for unknown entity discovery and indexing. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 8659–8673, Abu Dhabi, United Arab Emirates, dec 2022. Association for Computational Linguistics. doi:10.18653/V1/2022.EMNLP-MAIN.593.
- 92 Rohit J Kate. Automatic full conversion of clinical terms into SNOMED CT concepts. Journal of Biomedical Informatics, 111:103585, 2020. doi:10.1016/J.JBI.2020.103585.
- 93 Mayank Kejriwal. Domain-Specific Knowledge Graph Construction. Springer Publishing Company, Incorporated, 1st edition, 2019. doi:10. 1007/978-3-030-12375-8.
- 94 Troy D. Kelley. Symbolic and sub-symbolic representations in computational models of human cognition: What can be learned from biology?

Theory & Psychology, 13(6):847-860, 2003. doi: 10.1177/0959354303136005.

- 95 Elisa F. Kendall and Deborah L. McGuinness. Ontology Engineering. Synthesis Lectures on the Semantic Web: Theory and Technology. Springer, Cham, Switzerland, 2019. doi:10.1007/ 978-3-031-79486-5.
- 96 Sebastian Köhler, Michael Gargano, Nicolas Matentzoglu, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, et al. The human phenotype ontology in 2021. Nucleic acids research, 49(D1):D1207–D1217, 2021. doi:10.1093/NAR/GKAA1043.
- 97 Martin Krallinger, Florian Leitner, Carlos Rodriguez-Penagos, and Alfonso Valencia. Overview of the protein-protein interaction annotation extraction task of BioCreative II. Genome biology, 9:1–19, 2008. doi:10.1186/gb-2008-9-s2-s4.
- 98 Michael Kuhn, Damian Milosz Szklarczyk, Sune Pletscher-Frankild, Thomas H Blicher, Christian von Mering, Lars J Jensen, and Peer Bork. Stitch 4: integration of protein-chemical interactions with user data. *Nucleic Acids Research*, 42(D1):D401– D407, 2013. doi:10.1093/NAR/GKT1207.
- 99 Maxat Kulmanov and Robert Hoehndorf. Deep-GOZero: improving protein function prediction from sequence and zero-shot learning based on ontology axioms. *Bioinformatics*, 38(Supplement_1):i238-i245, 2022. doi:10.1093/ bioinformatics/btac256.
- 100 Hoang Thanh Lam, Marco Luca Sbodio, Marcos Martinez Gallindo, Mykhaylo Zayats, Raul Fernandez-Diaz, Victor Valls, Gabriele Picco, Cesar Berrospi Ramis, and Vanessa Lopez. Otter-Knowledge: benchmarks of multimodal knowledge graph representation learning from different sources for drug discovery. CoRR, abs/2306.12802, 2023. doi:10.48550/ARXIV.2306.12802.
- 101 Robert Leaman, Ritu Khare, and Zhiyong Lu. Challenges in clinical natural language processing for automated disorder normalization. *Journal of biomedical informatics*, 57:28–37, 2015. doi:10.1016/J.JBI.2015.07.010.
- 102 Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, sep 2019. doi:10.1093/BI0INFORMATICS/BTZ682.
- 103 Ulf Leser and Jörg Hakenberg. What makes a gene name? Named entity recognition in the biomedical literature. Briefings in bioinformatics, 6(4):357– 369, 2005. doi:10.1093/BIB/6.4.357.
- 104 Patrick Lewis, Myle Ott, Jingfei Du, and Veselin Stoyanov. Pretrained language models for biomedical and clinical tasks: Understanding and extending the state-of-the-art. In Proceedings of the 3rd Clinical Natural Language Processing Workshop, pages 146–157, 2020. doi:10.18653/V1/ 2020.CLINICALNLP-1.17.
- 105 Huanyu Li, Zlatan Dragisic, Daniel Faria, Valentina Ivanova, Ernesto Jiménez-Ruiz, Patrick Lambrix, and Catia Pesquita. User validation in ontology alignment: functional assessment and

impact. *Knowl. Eng. Rev.*, 34:e15, 2019. doi: 10.1017/S0269888919000080.

- 106 Jiao Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wiegers, and Zhiyong Lu. Biocreative V CDR task corpus: a resource for chemical disease relation extraction. *Database J. Biol. Databases Curation*, 2016, 2016. doi:10.1093/DATABASE/BAW068.
- 107 Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, pages 1–17, 2022. doi:10.1038/s41551-022-00942-x.
- 108 Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. Self-alignment pretraining for biomedical entity representations. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4228–4238, 2021. doi:10.18653/V1/2021. NAACL-MAIN.334.
- 109 Hao Liu, Yehoshua Perl, and James Geller. Concept Placement Using BERT Trained by Transforming and Summarizing Biomedical Ontology Structure. J. of Biomedical Informatics, 112(C), 2020. doi:10.1016/J.JBI.2020.103607.
- 110 Kaihong Liu, William R Hogan, and Rebecca S Crowley. Natural language processing methods and systems for biomedical ontology learning. *Jour*nal of biomedical informatics, 44(1):163–179, 2011. doi:10.1016/J.JBI.2010.07.006.
- 111 Yu Liu, Jingtao Ding, Yanjie Fu, and Yong Li. Urbankg: An urban knowledge graph system. ACM Transactions on Intelligent Systems and Technology, 14(4):1–25, 2023. doi:10.1145/3588577.
- 112 Takaki Makino, Yoshihiro Ohta, Jun'ichi Tsujii, et al. Tuning support vector machines for biomedical named entity recognition. In Proceedings of the ACL-02 workshop on Natural language processing in the biomedical domain, pages 1–8, 2002. doi:10.3115/1118149.1118150.
- 113 Diego Marcheggiani and Ivan Titov. Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations. Transactions of the Association for Computational Linguistics, 4:231– 244, jun 2016. doi:10.1162/TACL_A_00095.
- 114 Nicolas Matentzoglu, James P. Balhoff, Susan M. Bello, Chris Bizon, Matthew H. Brush, Tiffany J. Callahan, Christopher G. Chute, William D. Duncan, Chris T. A. Evelo, Davera Gabriel, John Graybeal, Alasdair J. G. Gray, Benjamin M. Gyori, Melissa A. Haendel, Henriette Harmse, Nomi L. Harris, Ian Harrow, Harshad Hegde, Amelia L. Hoyt, Charles Tapley Hoyt, Dazhi Jiao, Ernesto Jiménez-Ruiz, Simon Jupp, Hyeongsik Kim, Sebastian Köhler, Thomas Liener, Qinqin Long, James Malone, James A. McLaughlin, Julie A. McMurry, Sierra A. T. Moxon, Monica C. Munoz-Torres, David Osumi-Sutherland, James A. Overton, Bjoern Peters, Tim E. Putman, Núria Queralt-Rosinach, Kent A. Shefchek, Harold Solbrig, Anne E. Thessen, Tania Tudorache, Nicole A. Vasilevsky, Alex H. Wagner, and Christopher J. Mungall. A simple standard for sharing ontological mappings (SSSOM). Database

J. Biol. Databases Curation, 2022(2022), 2022. doi:10.1093/DATABASE/BAAC035.

- 115 Nicolas Matentzoglu, Damien Goutte-Gattat, Shawn Zheng Kai Tan, James P Balhoff, Seth Carbon, Anita R Caron, William D Duncan, Joe E Flack, Melissa Haendel, Nomi L Harris, William R Hogan, Charles Tapley Hoyt, Rebecca C Jackson, HyeongSik Kim, Huseyin Kir, Martin Larralde, Julie A McMurry, James A Overton, Bjoern Peters, Clare Pilgrim, Ray Stefancsik, Sofia MC Robb, Sabrina Toro, Nicole A Vasilevsky, Ramona Walls, Christopher J Mungall, and David Osumi-Sutherland. Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies. Database, 2022:baac087, oct 2022. doi:10.1093/DATABASE/BAAC087.
- 116 Jamie P. McCusker, Neha Keshan, Sabbir Rashid, Michael Deagen, Cate Brinson, and Deborah L. McGuinness. NanoMine: A Knowledge Graph for Nanocomposite Materials Science. In The Semantic Web – ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II, pages 144–159, 2020. doi:10.1007/978-3-030-62466-8_10.
- 117 Pierre Monnin, Miguel Couceiro, Amedeo Napoli, and Adrien Coulet. Knowledge-based matching of n-ary tuples. In Mehwish Alam, Tanya Braun, and Bruno Yun, editors, Ontologies and Concepts in Mind and Machine - 25th International Conference on Conceptual Structures, Proceedings, volume 12277 of Lecture Notes in Computer Science, pages 48–56, 2020. doi:10.1007/ 978-3-030-57855-8_4.
- 118 Pierre Monnin, Joël Legrand, Graziella Husson, Patrice Ringot, Andon Tchechmedjiev, Clément Jonquet, Amedeo Napoli, and Adrien Coulet. PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison. BMC Bioinform., 20-S(4):139:1–139:16, 2019. doi:10.1186/S12859-019-2693-9.
- 119 Pierre Monnin, Chedy Raïssi, Amedeo Napoli, and Adrien Coulet. Discovering alignment relations with graph convolutional networks: A biomedical case study. *Semantic Web*, 13(3):379–398, 2022. doi:10.3233/SW-210452.
- 120 Deisy Morselli Gysi, Ítalo Do Valle, Marinka Zitnik, Asher Ameli, Xiao Gan, Onur Varol, Susan Dina Ghiassian, JJ Patten, Robert A Davey, Joseph Loscalzo, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proceedings of the National Academy of Sciences, 118(19):e2025581118, 2021. doi:10.1073/pnas.2025581118.
- 121 Boris Motik and Ljiljana Stojanovic. Ontology evolution within ontology editors. In OntoWeb-SIG3 Workshop at the 13th International Conference on Knowledge Engineering and Knowledge Management EKAW 2002; Siguenza (Spain), 30th September 2002, 2002. URL: https://ceur-ws. org/Vol-62/EON2002_Stojanovic.pdf.
- 122 Lino Murali, G. Gopakumar, Daleesha M. Viswanathan, and Prema Nedungadi. Towards electronic health record-based medical knowledge graph construction, completion, and applications:

A literature study. Journal of Biomedical Informatics, 143:104403, 2023. doi:10.1016/J.JBI. 2023.104403.

- 123 Emir Muñoz, Vít Nováček, and Pierre-Yves Vandenbussche. Facilitating prediction of adverse drug reactions by using knowledge graphs and multilabel learning models. *Briefings in Bioinformatics*, 20(1):190–202, aug 2017. doi:10.1093/BIB/ BBX099.
- 124 Erik B Myklebust, Ernesto Jimenez-Ruiz, Jiaoyan Chen, Raoul Wolf, and Knut Erik Tollefsen. Knowledge graph embedding for ecotoxicological effect prediction. In The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18, pages 490–506, 2019. doi:10.1007/978-3-030-30796-7_30.
- 125 Erik B Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, Raoul Wolf, and Knut Erik Tollefsen. Prediction of adverse biological effects of chemicals using knowledge graph embeddings. *Semantic Web*, 13(3):299–338, 2022. doi:10.3233/SW-222804.
- 126 Erik Bryhn Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, Raoul Wolf, and Knut Erik Tollefsen. Understanding Adverse Biological Effect Predictions Using Knowledge Graphs. CoRR, abs/2210.15985, 2022. doi:10.48550/ARXIV.2210. 15985.
- 127 David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26, 2007. doi: 10.1075/bct.19.03nad.
- 128 Yavor Nenov, Robert Piro, Boris Motik, Ian Horrocks, Zhe Wu, and Jay Banerjee. RDFox: A Highly-Scalable RDF Store. In 14th International Semantic Web Conference, volume 9367 of Lecture Notes in Computer Science, pages 3–20. Springer, 2015. doi:10.1007/978-3-319-25010-6_1.
- 129 Fabian Neuhaus and Janna Hastings. Ontology development is consensus creation, not (merely) representation. Applied Ontology, 17(4):495–513, 2022. doi:10.3233/A0-220273.
- 130 David N. Nicholson and Casey S. Greene. Constructing knowledge graphs and their biomedical applications. *Computational and Structural Biotechnology Journal*, 18:1414–1428, 2020. doi: 10.1016/j.csbj.2020.05.017.
- 131 N. Noy and D.L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05 and SMI-2001-0880, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, 2001.
- 132 Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, et al. Bio-Portal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, 37(suppl_2):W170–W173, 2009. doi:10.1093/ NAR/GKP440.
- 133 Vardaan Pahuja, Yu Gu, Wenhu Chen, Mehdi Bahrami, Lei Liu, Wei-Peng Chen, and Yu Su. A Systematic Investigation of KB-Text Embedding Alignment at Scale. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International

Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1764–1774, 2021. doi:10.18653/V1/2021.ACL-LONG.139.

- 134 Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. CoRR, abs/2306.08302, 2023. doi:10.48550/ ARXIV.2306.08302.
- 135 Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. Semant. Web, 8(3):489–508, jan 2017. doi: 10.3233/SW-160218.
- 136 Hao Peng, Haoran Li, Yangqiu Song, Vincent Zheng, and Jianxin Li. Differentially private federated knowledge graphs embedding. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pages 1416–1425, 2021. doi:10.1145/3459637.3482252.
- 137 Romana Pernisch, Daniele Dell'Aglio, and Abraham Bernstein. Beware of the hierarchy—an analysis of ontology evolution and the materialisation impact for biomedical ontologies. Journal of Web Semantics, 70:100658, 2021. doi: 10.1016/J.WEBSEM.2021.100658.
- 138 Mina Abd Nikooie Pour, Alsayed Algergawy, Patrice Buche, Leyla Jael Castro, Jiaoyan Chen, Hang Dong, Omaima Fallatah, Daniel Faria, Irini Fundulaki, Sven Hertling, Yuan He, Ian Horrocks, Martin Huschka, Liliana Ibanescu, Ernesto Jiménez-Ruiz, Naouel Karam, Amir Laadhar, Patrick Lambrix, Huanyu Li, Ying Li, Franck Michel, Engy Nasr, Heiko Paulheim, Catia Pesquita, Tzanina Saveta, Pavel Shvaiko, Cássia Trojahn, Chantelle Verhey, Mingfang Wu, Beyza Yaman, Ondrej Zamazal, and Lu Zhou. Results of the Ontology Alignment Evaluation Initiative 2022. In Proceedings of the 17th International Workshop on Ontology Matching (OM 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), volume 3324 of CEUR Workshop Proceedings, pages 84-128, 2022. URL: https: //ceur-ws.org/Vol-3324/oaei22_paper0.pdf.
- 139 María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. LOT: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence, 111:104755, 2022. doi: 10.1016/J.ENGAPPAI.2022.104755.
- 140 Eric Prud'hommeaux, Steve Harris, and Andy Seaborne. SPARQL 1.1 Query Language. Technical report, W3C, 2013. URL: https://www.w3. org/TR/sparql11-query/.
- 141 Sampo Pyysalo, Filip Ginter, Juho Heimonen, Jari Björne, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. Bioinfer: a corpus for information extraction in the biomedical domain. *BMC bioinformatics*, 8:1–24, 2007. doi:10.1186/ 1471-2105-8-50.
- 142 Zhixin Qi, Hongzhi Wang, Ziming Shen, and Donghua Yang. PreKar: A learned performance predictor for knowledge graph stores. World Wide Web, 26(1):321–341, 2023. doi:10.1007/ S11280-022-01033-2.
- 143 Enayat Rajabi and Somayeh Kafaie. Knowledge graphs and explainable AI in healthcare. Inf., 13(10):459, 2022. doi:10.3390/INF013100459.

5:30 Knowledge Graphs for Life Sciences

- 144 K. E. Ravikumar, Majid Rastegar-Mojarad, Majid Rastegar-Mojarad, and Hongfang Liu. Belminer: adapting a rule-based relation extraction system to extract biological expression language statements from bio-medical literature evidence sentences. Database: The Journal of Biological Databases and Curation, 2017, 2017. doi:10.1093/DATABASE/ BAW156.
- 145 KE Ravikumar, Kavishwar B Wagholikar, and Hongfang Liu. Towards pathway curation through literature mining-a case study using pharmgkb. In *Biocomputing* 2014, pages 352-363. World Scientific, 2014. URL: http://psb.stanford.edu/psb-online/ proceedings/psb14/ravikumarke.pdf, doi:10.1142/9789814583220_0034.
- 146 Alan Rector, Stefan Schulz, Jean Marie Rodrigues, Christopher G Chute, and Harold Solbrig. On beyond Gruber: "Ontologies" in today's biomedical information systems and the limits of OWL. Journal of Biomedical Informatics, 100:100002, 2019. doi:10.1016/J.YJBINX.2019.100002.
- 147 Petar Ristoski and Heiko Paulheim. Semantic web in data mining and knowledge discovery: A comprehensive survey. J. Web Semant., 36:1–22, 2016. doi:10.2139/ssrn.3199217.
- 148 Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences of the United States of America, 118(15), 2021. doi:10.1073/PNAS. 2016239118.
- 149 Tim Rocktäschel, Michael Weidlich, and Ulf Leser. ChemSpot: a hybrid system for chemical named entity recognition. *Bioinformatics*, 28(12):1633–1640, 2012. doi:10.1093/BI0INFORMATICS/BTS183.
- 150 Natalia Díaz Rodríguez, Alberto Lamas, Jules Sanchez, Gianni Franchi, Ivan Donadello, Siham Tabik, David Filliat, Policarpo Cruz, Rosana Montes, and Francisco Herrera. Explainable neural-symbolic learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The monumai cultural heritage use case. Inf. Fusion, 79:58–83, 2022. doi:10.1016/J.INFFUS.2021.09.022.
- 151 Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large scale chemical language representations capture molecular structure and properties. Nature Machine Intelligence, 4(12):1256–1264, 2022. doi:10.1038/S42256-022-00580-7.
- 152 Manuel Salvadores, Paul R. Alexander, Mark A. Musen, and Natalya Fridman Noy. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. Semantic Web, 4(3):277–284, 2013. doi:10.3233/SW-2012-0086.
- 153 Matthias Samwald, José Antonio Miñarro-Giménez, Richard D. Boyce, Robert R. Freimuth, Klaus-Peter Adlassnig, and Michel Dumontier. Pharmacogenomic knowledge representation, reasoning and genome-based clinical decision support based on OWL 2 DL ontologies. BMC

Medical Informatics Decis. Mak., 15:12, 2015. doi:10.1186/S12911-015-0130-1.

- 154 Conrad L Schoch, Stacy Ciufo, Mikhail Domrachev, Carol L Hotton, Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe, Richard Mcveigh, Kathleen O'Neill, Barbara Robbertse, Shobha Sharma, Vladimir Soussov, John P Sullivan, Lu Sun, Seán Turner, and Ilene Karsch-Mizrachi. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020:baaa062, aug 2020. doi:10.1093/DATABASE/ BAAA062.
- 155 Lynn M Schriml, James B Munro, Mike Schor, Dustin Olley, Carrie McCracken, Victor Felix, J Allen Baron, Rebecca Jackson, Susan M Bello, Cynthia Bearer, et al. The human disease ontology 2022 update. Nucleic acids research, 50(D1):D1255-D1261, 2022. doi:10.1093/NAR/ GKAB1063.
- 156 Marta Contreiras Silva, Daniel Faria, and Catia Pesquita. Matching multiple ontologies to build a knowledge graph for personalized medicine. In *The Semantic Web: 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29–June 2, 2022, Proceedings*, pages 461–477. Springer, 2022. doi:10.1007/ 978-3-031-06981-9_27.
- 157 Ana Claudia Sima, Tarcisio Mendes de Farias, Erich Zbinden, Maria Anisimova, Manuel Gil, Heinz Stockinger, Kurt Stockinger, Marc Robinson-Rechavi, and Christophe Dessimoz. Enabling semantic queries across federated bioinformatics databases. *Database*, 2019:baz106, 2019. doi: 10.1093/DATABASE/BAZ106.
- 158 Rita T Sousa, Sara Silva, and Catia Pesquita. Evolving knowledge graph similarity for supervised learning in complex biomedical domains. BMC bioinformatics, 21:1–19, 2020. doi:10.1186/ S12859-019-3296-1.
- **159** Rita T Sousa, Sara Silva, and Catia Pesquita. Explainable representations for relation prediction in knowledge graphs. *arXiv e-prints*, pages arXiv–2306, 2023. doi:10.24963/KR.2023/62.
- 160 John F Sowa et al. Semantic networks. Encyclopedia of artificial intelligence, 2:1493–1511, 1992. doi:10.1002/0470018860.s00065.
- 161 Lise Stork, Ilaria Tiddi, René Spijker, and Annette ten Teije. Explainable drug repurposing in context via deep reinforcement learning. In Catia Pesquita, Ernesto Jimenez-Ruiz, Jamie McCusker, Daniel Faria, Mauro Dragoni, Anastasia Dimou, Raphael Troncy, and Sven Hertling, editors, *The Semantic Web*, pages 3–20, 2023. doi:10.1007/978-3-031-33455-9_1.
- 162 Kai Sun, Yuhua Liu, Zongchao Guo, and Changbo Wang. Visualization for knowledge graph based on education data. Int. J. Softw. Informatics, 10, 2016. URL: http://www.ijsi.org/ijsi/article/ abstract/i227.
- 163 Mujeen Sung, Jinhyuk Lee, Sean Yi, Minji Jeon, Sungdong Kim, and Jaewoo Kang. Can language models be biomedical knowledge bases? In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 4723-4734, 2021. doi:10.18653/V1/2021. EMNLP-MAIN.388.

- 164 Gyte Tamasauskaite and Paul Groth. Defining a knowledge graph development process through a systematic review. ACM Trans. Softw. Eng. Methodol., 32(1), 2023. doi:10.1145/3522586.
- 165 HSPO Team. Health and Social Person-centric Ontology, sep 2022. URL: https://github.com/ IBM/hspo-ontology.
- 166 Christos Theodoropoulos, Natasha Mulligan, Thaddeus Stappenbeck, and Joao Bettencourt-Silva. Representation learning for person or entitycentric knowledge graphs: An application in healthcare. CoRR, abs/2305.05640, 2023. doi:10.48550/ ARXIV.2305.05640.
- 167 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. Artif. Intell., 302:103627, 2022. doi:10.1016/J.ARTINT.2021.103627.
- 168 Santiago Timón-Reina, Mariano Rincón, and Rafael Martínez-Tomás. An overview of graph databases and their applications in the biomedical domain. *Database*, 2021:baab026, may 2021. doi:10.1093/DATABASE/BAAB026.
- 169 Igor Trajkovski, Nada Lavrac, and Jakub Tolar. SEGS: search for enriched gene sets in microarray data. J. Biomed. Informatics, 41(4):588–601, 2008. doi:10.1016/J.JBI.2007.12.001.
- 170 Efthymia Tsamoura, David Carral, Enrico Malizia, and Jacopo Urbani. Materializing knowledge bases via trigger graphs. *Proc. VLDB Endow.*, 14(6):943– 956, 2021. doi:10.14778/3447689.3447699.
- 171 Serbulent Unsal, Heval Atas, Muammer Albayrak, Kemal Turhan, Aybar C Acar, and Tunca Doğan. Learning functional properties of proteins with language models. *Nature Machine Intelligence*, 4(3):227–245, 2022. doi:10.1038/ S42256-022-00457-9.
- 172 Nicole A. Vasilevsky, Shahim Essaid, Nicolas Matentzoglu, Nomi L. Harris, Melissa A. Haendel, Peter N. Robinson, and Christopher J. Mungall. Mondo Disease Ontology: Harmonizing Disease Concepts Across the World (short paper). In Janna Hastings and Frank Loebe, editors, Proceedings of the 11th International Conference on Biomedical Ontologies (ICBO) joint with the 10th Workshop on Ontologies and Data in Life Sciences (ODLS) and part of the Bolzano Summer of Knowledge (BoSK 2020), volume 2807 of CEUR Workshop Proceedings, pages 1–2. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2807/abstractY.pdf.
- 173 Blerta Veseli, Sneha Singhania, Simon Razniewski, and Gerhard Weikum. Evaluating language models for knowledge base completion. In European Semantic Web Conference, pages 227–243, 2023. doi:10.1007/978-3-031-33455-9_14.
- 174 Olga Vrousgou, Tony Burdett, Helen E. Parkinson, and Simon Jupp. Biomedical Ontology Evolution in the EMBL-EBI Ontology Lookup Service. In Themis Palpanas and Kostas Stefanidis, editors, Proceedings of the Workshops of the EDBT/ICDT 2016 Joint Conference, EDBT/ICDT Workshops 2016, volume 1558 of CEUR Workshop Proceedings, 2016. URL: https://ceur-ws.org/ Vol-1558/paper12.pdf.
- 175 Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea

Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023. doi:10.1038/S41586-023-06221-2.

- 176 Meng Wang and Ningyu Zhang. Cross-modal knowledge discovery, inference, and challenges. In Reasoning Web. Causality, Explanations and Declarative Knowledge: 18th International Summer School 2022, Tutorial Lectures, pages 199–209. Springer Nature Switzerland Springer, Cham, 2023. doi:10.1007/978-3-031-31414-8_6.
- 177 Xin Wang and Weixue Chen. Knowledge graph data management: Models, methods, and systems. In International Conference on Web Information Systems Engineering, pages 3–12. Springer, 2020. doi:10.1007/978-981-15-3281-8_1.
- 178 Xu Wang, Chen Yang, and Renchu Guan. A comparative study for biomedical named entity recognition. International Journal of Machine Learning and Cybernetics, 9:373–382, 2018. doi: 10.1007/S13042-015-0426-6.
- 179 Xuwu Wang, Junfeng Tian, Min Gui, Zhixu Li, Rui Wang, Ming Yan, Lihan Chen, and Yanghua Xiao. WikiDiverse: A Multimodal Entity Linking Dataset with Diversified Contextual Topics and Entity Types. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 4785-4797, 2022. doi:10.18653/V1/2022. ACL-LONG.328.
- 180 M. Whirl-Carrillo, R. Huddart, L. Gong, K. Sangkuhl, C. F. Thorn, R. Whaley, and T. E. Klein. An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics*, 110(3):563-572, 2021. doi:10.1002/cpt.2350.
- 181 Xander Wilcke, Peter Bloem, and Victor De Boer. The knowledge graph as the default data model for learning on heterogeneous knowledge. *Data Sci*ence, 1(1-2):39–57, 2017. doi:10.3233/DS-170007.
- 182 World Health Organization. International statistical classification of diseases and related health problems. ICD-10. World Health Organization, Geneva, Switzerland, fifth edition, 2016.
- 183 Tianxing Wu, Guilin Qi, Cheng Li, and Meng Wang. A Survey of Techniques for Constructing Chinese Knowledge Graphs and Their Applications. Sustainability, 10(9):1–26, 2018. doi: 10.3390/su10093245.
- 184 Eryu Xia, Wen Sun, Jing Mei, Enliang Xu, Ke Wang, and Yong Qin. Mining diseasesymptom relation from massive biomedical literature and its application in severe disease diagnosis. AMIA Annual Symposium Proceedings, 2018:1118–1126, 2018. URL: https://knowledge.amia.org/67852-amia-1. 4259402/t004-1.4263758/t004-1.4263759/ 2976467-1.4263778/2976026-1.4263775.
- 185 Bo Xiong, Michael Cochez, Mojtaba Nayyeri, and Steffen Staab. Hyperbolic embedding inference for structured multi-label prediction. Advances in Neural Information Processing Systems, 35:33016-33028, 2022. URL: http:// papers.nips.cc/paper_files/paper/2022/hash/ d51ab0fc62fe2d777c7569952f518f56-Abstract-Conference.html.

5:32 Knowledge Graphs for Life Sciences

- 186 Minghao Xu, Xinyu Yuan, Santiago Miret, and Jian Tang. Protst: Multi-modality learning of protein sequences and biomedical texts. In Proceedings of the 40th International Conference on Machine Learning, ICML'23. JMLR.org, 2023. URL: https: //proceedings.mlr.press/v202/xu23t.html.
- 187 Zonghai Yao, Yi Cao, Zhichao Yang, Vijeta Deshpande, and Hong Yu. Extracting biomedical factual knowledge using pretrained language model and electronic health record context. In AMIA Annual Symposium Proceedings, volume 2022, page 1188, 2022. URL: https://knowledge.amia.org/76677-amia-1.4637602/f006-1.4642154/f006-1.4642155/82-1.4642180/1077-1.4642177.
- 188 Ronghui You, Shuwei Yao, Hiroshi Mamitsuka, and Shanfeng Zhu. DeepGraphGO: graph neural network for large-scale, multispecies protein function prediction. *Bioinformatics*, 37(Supplement_1):i262-i271, 2021. doi:10.1093/ BIOINFORMATICS/BTAB270.
- 189 Ronghui You, Zihan Zhang, Yi Xiong, Fengzhu Sun, Hiroshi Mamitsuka, and Shanfeng Zhu. GO-Labeler: improving sequence-based large-scale protein function prediction by learning to rank. *Bioinformatics*, 34(14):2465–2473, 2018. doi:10.1093/ BIOINFORMATICS/BTY130.
- 190 Jianbo Yuan, Zhiwei Jin, Han Guo, Hongxia Jin, Xianchao Zhang, Tristram Smith, and Jiebo Luo. Constructing biomedical domain-specific knowledge graph with minimum supervision. *Knowl*edge and Information Systems, 62(1), 2020. doi: 10.1007/S10115-019-01351-4.
- 191 Bohui Zhang, Albert Meroño Peñuela, and Elena Simperl. Towards explainable automatic knowledge graph construction with human-in-the-loop. In *HHAI 2023: Augmenting Human Intellect*,

pages 274–289. IOS Press, Amsterdam, 2023. doi: 10.3233/FAIA230091.

- 192 Ningyu Zhang, Zhen Bi, Xiaozhuan Liang, Siyuan Cheng, Haosen Hong, Shumin Deng, Qiang Zhang, Jiazhang Lian, and Huajun Chen. OntoProtein: Protein Pretraining With Gene Ontology Embedding. In International Conference on Learning Representations, 2022. URL: https://openreview. net/forum?id=yfe1VMYAXa4.
- 193 Yingwen Zhao, Jun Wang, Jian Chen, Xiangliang Zhang, Maozu Guo, and Guoxian Yu. A literature review of gene function prediction by modeling gene ontology. *Frontiers in genetics*, 11:400, 2020. doi:10.3389/fgene.2020.00400.
- 194 Hong-Yu Zhou, Yunxiang Fu, Zhicheng Zhang, Bian Cheng, and Yizhou Yu. Protein representation learning via knowledge enhanced primary structure reasoning. In *International Conference on Learning Representations*, 2023. URL: https://openreview.net/pdf?id=VbCMhg7MRmj.
- 195 Qile Zhu, Xiaolin Li, Ana Conesa, and Cécile Pereira. GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics*, 34(9):1547–1554, 2018. doi:10.1093/BI0INFORMATICS/BTX815.
- 196 Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018. doi:10.1093/ BIOINFORMATICS/BTY294.
- 197 Lei Zou, M Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gStore: a graph-based SPARQL query engine. The VLDB journal, 23:565–590, 2014. doi:10.1007/ S00778-013-0337-7.

A Terms in Knowledge Graphs and Life Sciences

Below we provide a list of key terms used in this paper, as well as their definitions and explanations. Note we mainly use the original sentences in the sources that are referenced as the definitions.

- **Description Logics:** a family of knowledge representation languages that can be used to represent knowledge of an application domain. DLs differ from their predecessors, such as semantic networks and frames, in that they are equipped with logic-based semantics, the same semantics as that of classical first-order logic. Most ontologies are implemented in OWL, whose semantics are given by the Description Logic SROIQ. [5]
- **TBox and ABox:** the two components of domain knowledge in Description Logics, i.e., a terminological part called the TBox and an assertional part called the ABox, with the combination of a TBox and an ABox being called a knowledge base (KB). The TBox represents knowledge about the structure of the domain (similar to a database schema), while the ABox represents knowledge about a concrete situation (similar to a database instance). [5]
- **Semantic Networks:** a graph structure for representing knowledge in patterns of interconnected nodes and arcs [160]. We use the term to denote a graph of concepts and relations without formal semantics.
- **Gene Ontology:** The Gene Ontology (GO) knowledgebase provides a comprehensive, structured, computer-accessible representation of gene function, for genes from any cellular organism or virus [4, 28].

- **SNOMED-CT:** Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a structured clinical vocabulary. It has a general and comprehensive coverage of clinical terms to support electronic healthcare systems and clinical applications. [38, 27]
- UMLS (UMLS Metahesaurus and UMLS Semantic Networks): Unified Medical Language System (UMLS) is a repository of biomedical vocabularies developed by the US National Library of Medicine. The UMLS is composed of three "knowledge sources", a Metathesaurus, a semantic network, and a lexicon. The UMLS Metathesaurus is a comprehensive effort for integrating biomedical ontologies through mappings. The UMLS Semantic Networks define the types or categories, or Semantic Types, of all Metathesaurus concepts and their relationships, or Semantic Relations. [9, 27]
- **ChEBI:** Chemical Entities of Biological Interest (ChEBI) is a database and ontology containing information about chemical entities of biological interest. [62]
- Symbolic vs. subsymbolic representations: Rooted in cognitive science, symbolic systems of human cognition are related to the representation and manipulation of symbols; sub-symbolic or connectionist systems are most generally associated with the metaphor of a neuron, e.g., perceptrons as an early system [94]. In terms of AI, symbolic systems contain logic-based and knowledge representations, while subsymbolic systems typically contain neural networks and deep learning based methods [41]. Neural language models and pre-trained language models [86] are also classified under subsymbolic systems.
- **Pre-trained and Large Language Models:** Neural language modeling is the task of using neural network approaches to predict words from prior their contexts in a sequence. Pre-training is the process of learning some sort of representation (usually neural embedding based) of meaning for words or sentences by processing very large amounts of text (or other data in a sequence form, e.g., proteins and KG facts). This results in pre-trained language models. The dominating architecture for neural language modeling is Transformer-based models, including BERT, its domain specific versions, and later large variants, like the GPT series. The pre-trained language models of very large sizes are recently coined Large Language Models (LLMs). [86]
- **Neuro-symbolic representations:** refers to the integration of neural networks and symbolic representations to design AI models that base their prediction on both data and knowledge. [41]

Towards Ordinal Data Science

Gerd Stumme¹ $\square \land \bigcirc$

Knowledge & Data Engineering Group, Research Center for Information System Design & Department of Electrical Engineering and Computer Science, University of Kassel, Germany

Dominik Dürrschnabel 🖂 🏠 💿

Knowledge & Data Engineering Group, Research Center for Information System Design & Department of Electrical Engineering and Computer Science, University of Kassel, Germany

Tom Hanika 🖂 🏠 💿

Institute of Computer Science, University of Hildesheim, Germany Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Germany

— Abstract -

Order is one of the main instruments to measure the relationship between objects in (empirical) data. However, compared to methods that use numerical properties of objects, the amount of ordinal methods developed is rather small. One reason for this is the limited availability of computational resources in the last century that would have been required for ordinal computations. Another reason – particularly important for this line of research – is that order-based methods are often seen as too mathematically rigorous for applying them to realworld data. In this paper, we will therefore discuss different means for measuring and 'calculating' with ordinal structures – a specific class of directed graphs – and show how to infer knowledge from them. Our aim is to establish Ordinal Data Science as a fundamentally new research agenda. Besides cross-fertilization with other cornerstone machine learning and knowledge representation methods, a broad range of disciplines will benefit from this endeavor, including, psychology, sociology, economics, web science, knowledge engineering, scientometrics.

2012 ACM Subject Classification Computing methodologies \rightarrow Ontology engineering; Computing methodologies \rightarrow Nonmonotonic, default reasoning and belief revision; Computing methodologies \rightarrow Semantic networks; Computing methodologies \rightarrow Algebraic algorithms; Computing methodologies \rightarrow Boolean algebra algorithms; Computing methodologies \rightarrow Unsupervised learning; Computing methodologies \rightarrow Rule learning

Keywords and phrases Order relation, data science, relational theory of measurement, metric learning, general algebra, lattices, factorization, approximations and heuristics, factor analysis, visualization, browsing, explainability

Digital Object Identifier 10.4230/TGDK.1.1.6

Category Vision

Related Version arXiv Version: https://arxiv.org/abs/2307.09477

Received 2023-06-30 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1 Introduction

Order is a predominant concept for perceiving and organizing our physical and social environment, to infer meaning and explanation from observation, and to search and rectify decisions. For instance, we admire the highest mountain on earth, observe pecking order among animals, schedule events in time, and structure our organizations, libraries, and diseases hierarchically. The notion

¹ corresponding author



© Gerd Stumme, Dominik Dürrschnabel, and Tom Hanika;

licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 6, pp. 6:1–6:39 Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

6:2 Towards Ordinal Data Science

of order is deeply embedded in our language, as every adjective gives rise to a comparative (e.g., better, more expensive, more beautiful). Specific technical and social processes have been established for dealing with ordinal structures, e.g., scheduling routines for aircraft take-offs and production planning, queuing at bus stops, deriving the succession order as depth-first linear extension of the royal family tree, or discussing only the borderline cases in scientific program committees. These processes, however, are rather task-specific – there exist only few generic data analysis and machine learning tasks that are particularly tailored for ordinal data in general.

It comes thus as little surprise that order is one of the two main types of relations between objects in empirical data, together with proximity/distance [39]. While there exists a wide range of scientific work on analysis methods for ordinal data, their amount and coverage is far less than for proximity data. As a consequence, many data analysts resort to methods designed for other data types. We therefore call interested researchers to join us in the aim to establish the new field of Ordinal Data Science – both by organizing existing work in a unifying framework and by extending these approaches: Contributions are sought in two specific areas. Theoretical foundations for analyzing ordinal structures, in particular with respect to knowledge extraction and representation, on the one hand; and algorithmic methods that provide the means to measure and "calculate" with ordinal structures and closely related algebraic notions on the other hand.

A wide range of applied disciplines will benefit from the new field of Ordinal Data Science, a. o. psychology, web science, knowledge engineering, scientometrics. Besides these disciplines, ordinal data are related to the large family of relational data which have received high interest of the computer science community in the last years. This is due to developments in fields such as sociology ("relational turn") [141] or genetics [84] or epidemiology [36], and in particular sociotechnical developments such as the rise of online social networks or knowledge graphs. This means that, for the analysis of ordinal data, one can benefit from all kinds of measures and methods for relational data, as for instance centrality measures and clustering algorithms for (social) network data, or inductive logic programming. The specific structure of ordinal data, however, allows additionally to tap on the rich – but up to date mostly unexploited for data science – tool-set of mathematical order theory [34] and lattice² theory [42].

We expect Ordinal Data Science to become a scientific field with many ramifications. In this paper, we will discuss – in an exemplary fashion and to kindle the research field – five out of many potential subfields as programmatic research tasks:

1) Ordinal Measurement Theory. Representational Measurement Theory, the most influential theory of measurement to date, discusses how empirically observed phenomena can be measured. It considers a measurement as a mapping from some empirical relational structure to a numerical relational structure. While there seems to be agreement that the latter could be any mathematical structure that is appropriate for the purpose at hand, almost all theory is centered around the real numbers (and variations) as numerical relational structure. As order is the second main type of relation in empirical data beside proximity/distance, we will discuss, in Section 5, the vision of an order-theoretic version of measurement theory.

2) Ordered Metric Spaces. Real-world datasets are typically heterogeneous and contain both ordinal and numerical dimensions. The latter naturally give rise to a metric. In Section 6, we will motivate the need for a theory about the compatibility (or consistency) of an order relation and a metric. To this end, we introduce ordered metric spaces (om-spaces), in the spirit of metric measure spaces. Such a theory should be able to answer a range of foundational questions about

 $^{^{2}}$ In Section 4.1, we will recall that lattices are specific order relations with additional structure.

G. Stumme, D. Dürrschnabel, and T. Hanika

the relationship between order relations and metrics, including: a) If a metric has been learned based on information about the order, how consistent is the metric with that order? This will provide a fundamental new quality criterion for order-related machine learning tasks. b) How strong do order and metrics "agree" with their view on a dataset? An answer might be highly beneficial to an investigation whether order and metrics reflect the same hidden structure. c) How justified is it to represent an order relation by means of the metric? If, for a given dataset, the answer is satisfactory, this would justify in this case the use of the rich toolkit of numerical data analysis and machine learning methods for analyzing the ordinal structure.

3) Algebraic constructions/decompositions for complexity reduction. Algebraic operations provide a large variety of methods for reducing the complexity of ordered sets, in particular lattices. However, they are very sensitive to small perturbations in the data; and approximations and heuristics in the style of data mining are not (yet) considered in algebraic research. In Section 7, we will discuss how they could be made less sensitive in order to make them applicable to data science tasks.

4) Ordinal Factor Analysis. Observed and measured data is often highly correlated and interlinked, being caused by a small amount of factors. A significant task in data analysis is the identification of these factors. In Section 8, we discuss – in analogy to the classical factor analysis which is based on linear algebra – potential ways for extracting these factors in ordinal data based on the relation product.

5) Visualization, Exploration, and Browsing Ordinal Data. The typical means of presenting ordered sets to humans is via line diagrams. Open problems include in particular the specification of HCI-founded, formal optimization criteria for graph drawing (and the development of corresponding layout algorithms), as well as the (semi-)automatic break-down of large ordered sets into smaller, visualizable parts together with suitable means for their interactive exploration. In Section 9, we will call for the development of new paradigms for exploring and browsing ordinal datasets.

Before diving in the challenges of these subfields in more detail, we will first discuss the nature of ordinal data and the role of hierarchies in data science in Sections 2 and 3, resp., and will provide some order-theoretical foundations in Section 4.

2 Order as Foundational Principle for Organizing Data

Quantification with real numbers has been boosted by different factors, including *i*) the development of scientific measuring instruments since the scientific revolution, *ii*) the claim that the social sciences (starting with psychology) should use the same numerical methods which had been successful in natural sciences [63], and *iii*) nowadays by the instant availability of an enormous range of datasets to almost all aspects of science and everyday life. Indeed, in many cases, entities can be ordered through real-valued valuation functions like price or size. As the real numbers constitute an ordered field, the analysis of such data benefits from the existence of the operators $+, -, \cdot, /, 0, 1$ together with total³ comparability \leq . Moreover, this combination allows for measures of tendency (such as mean, variance, and skewness) and transformations. If more than one real-valued dimension is present, this yields to a real vector space \mathbb{R}^n , which results in additional descriptive measures and metric properties, such as volumes, angles, correlation, covariance. This is the standard setting for the majority of data analysis and machine learning

³ Total refers to the property that for any two elements a, b either $a \le b$ or $b \le a$ is true. A total order relation is often also called *linear order*.

6:4 Towards Ordinal Data Science

models, and algorithms (e.g., density-based clustering, logistic regression, SVMs, to name just a few).

However, organizing hierarchical relationships by means of numerical values is not always adequate, as this kind of organization presupposes two important conditions:

- 1. every pair of entities has to be comparable, and
- 2. the magnitudes of the differences between numerical values are meaningful and thus comparable themselves.

In many situations, however, this is not the case: (i) does not hold, e.g., in concept hierarchies ("mankind" is neither a subconcept nor a superconcept of "ocean") nor in organizations (a member of parliament is neither above nor below a Secretary of State), and also teaching curricula are far from being linearly ordered; (ii) does not hold, e.g., in school grades (In the German system is the difference between "1 – very good" and "2 – good" equal to the difference between "4 – sufficient" and "5 – insufficient/fail"?) nor in organizations (In the European Commission, is an advisor closer to a deputy director general than a head of group to a director?).

To address such variations of data types, S. S. Stevens has distinguished in [174] four *levels* of measurement: nominal, ordinal, interval, and ratio. For data on the ratio level (e.g., height), all above-mentioned operations are allowed (division, for instance, provides ratios). Data on the interval level (e.g., temperature measured in Celsius or Fahrenheit) do not have a meaningful zero as point of reference and thus do not allow for ratios, while the comparison of differences is still meaningful. Data on the ordinal level (e.g., the parent relation) can be compared hierarchically only, and on nominal level (e.g., eye color) only up to equality. Over time, this classification was discussed, refined and expanded, but never discarded [144, 35].

The mathematical field of Order and Lattice Theory [97, 34, 163, 21, 88, 42] is a canonical theory for modeling ordinal data. Its development started – independently of the above-mentioned quest for a suitable theory for measurements – in the second half of the 19th century as algebraic logic with the aim of George Boole [23], Charles S. Peirce [150] and Ernst Schröder [164] to formalize a calculus for concept hierarchies as cornerstone for a mathematical logic. In symbolic knowledge representation, ordered sets are extensively used for this purpose. For instance, in RDFS, the set of all rdfs:Classes with the rdfs:subClassOf relation is an ordered set, and in the field of conceptual graphs, one frequently assumes that the type hierarchy is even a lattice [168].

3 Hierarchies in Data Science

Data Science is the scientific field of analyzing data and extracting knowledge from data. This understanding of data science as an interdisciplinary field grew over the past decades [208, 152, 58, 41, 46] as a confluence of methods from statistics, data mining/knowledge discovery, machine learning, data management and big data. Here, we briefly discuss the relationship between some of these areas to ordinal data.

In statistics, several correlation measures exist for rankings (e.g., Kendall's τ [111], Spearman's ρ [170], Goodman and Kruskal's γ [83]) and tests for the significance of correlation (e.g., Mann–Whitney U test and Wilcoxon signed-rank test [213]). However, these measures and tests are designed only for *linear orders* (i. e., multi-inheritance and pairs of incomparable elements are not considered). In *data mining and machine learning*, most methods are designed for numerical (and some for categorical) data. When ordered sets occur, then mostly as part of the model, e.g., in decision trees [27], as topology of neural networks [114], as inclusion hierarchy for frequent item sets [3] (cf. next subsection), or as structure for coarser and finer set systems in hierarchical agglomerative clustering [212]. Often data points are organized in (linear) rankings, as a result from some numerical score, e.g., the predicted accuracy in supervised learning, the cosine similarity to a given query in information retrieval, or the result of a learning-to-rank approach. Here again,

G. Stumme, D. Dürrschnabel, and T. Hanika

these methods produce linear orders. A setting where the resulting order is non-linear is *ontology learning* (see below).

Ordinal Data Science will profit from and contribute to the field of machine-learned ranking [129, 85], particularly preference learning [106], which is a subfield of AI concerned with label ranking and instance ranking. Other areas in AI do also turn to ordinal data, e.g., "Deep Ordinal Reinforcement Learning" [230] attempts to adapt reinforcement learning to ordinal data. Machine learning methods for linear ordered sets are already somewhat further developed, e.g., treatment of linear ordered classes in supervised learning [32]. The same is true for the important machine learning method metric embedding, which was recently adapted to linear ordered data [165] and ordinal constraints inferred from metrics [115, 199]. An even more extensive study on the use of linear ordered data [109] concluded that even in the presence of metric features the study of purely ordinal features might be fruitful. Early methods of how this might be done are already being explored in a rather practical way [231]. Finally, Ordinal Data Science is strongly related to the discipline statistical relational learning, a research area that itself has a lot in common with many other AI areas, such as reasoning and knowledge representation. Foremost their advances in treating graph data [159] is of high relevance to Ordinal Data Science and vice versa.

4 Order-Theoretical Foundations

4.1 Ordered Sets

The key concept for studying ordinal data is an ordered set. An ordered set (P, \leq) is a directed graph (i. e., it consists of a set P and a binary relation \leq on P) such that \leq is reflexive $(p \leq p)$, transitive $(p \leq q, q \leq r \implies p \leq r)$, and antisymmetric $(p \leq q, q \leq p \implies p = q)$ [21]. This is equivalent to a transitive, directed acyclic graph with self-loops. Ordered sets could thus be seen as special cases of directed graphs or networks, and indeed some results of graph theory and network analysis are used; but order theory extends this with its own, specific flavor: For instance, the acyclicity of an ordered set simplifies its decomposition (e.g., , by filters and ideals), results in more specific structural measures such as height, width and dimension, and allows for more readable visualisations.

A typical ordered set is the set of all human beings together with the is_ancester_of relation. Another example are the dependencies of tasks in production planning. Note that, in general, we do not require comparability of all elements. The latter would mean that the order is *total* (also called *linear*), i. e., that, for all $p, q \in P$, one of p < q, p = q or p > q holds. To emphasize that comparability is not required, ordered sets are also called *partially ordered sets* (or *posets* for short) in the literature, especially in computer science.

While one cannot exploit numerical operations for describing and analyzing ordinal data, they come with other constructions, such as *order filters, order intervals, Pareto optima* etc.

For finite ordered sets – as studied in data science – we can reduce \leq without loss of information to the *neighboring relation* \prec by removing all pairs that can be deduced by reflexivity and transitivity. This is the relation predominantly used for visualization. Note that in many cases, the order is not *linear* (i. e., there may be incomparable elements and multi-inheritance).

A particularly useful kind of order relations are *lattices*, as they provide more structure and operations: In a *lattice*, any two elements p and q always have a *unique least common upper bound* $p \lor q$ and a *unique greatest common lower bound* $p \land q$. These two operations are often called *join* and *meet* respectively. One could say that a lattice is a hierarchy in which one can navigate up and down with these two operations. A *complete lattice* is an ordered set where every subset of P (including the empty set and infinite subsets) has meet and join. This particular property is only relevant if P is not finite, since each finite lattice is complete. Although the datasets considered in

6:6 Towards Ordinal Data Science

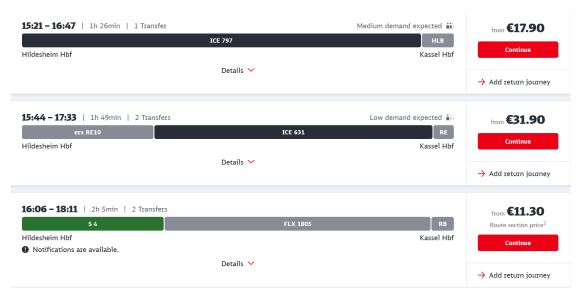


Figure 1 Pareto optima of train connections between the cities Hildesheim and Kassel as inferred from the German railway corporation Deutsche Bahn via www.bahn.de.

data science are always finite, a deeper understanding of the infinite case, and hence of complete lattices, can provide new insights.

A complexity measure for an ordered set or lattice (P, \leq) is its order dimension, i.e., the number *n* of linear orders \leq_1, \ldots, \leq_n on *P* with $\bigcap_{i=1}^n \leq_i = \leq$. In other terms, it is the smallest $n \in \mathbb{N}$ such that the ordered set is embeddable in the Cartesian product of *n* linear orders [206]. In a previous work [50] we explored how to reduce the complexity of a dataset by identifying parts that contribute to it having a high order dimension. However, computing (i.e., deciding) the order dimension is \mathcal{NP} -complete if the dimension is higher than two [228]. Even approximations are proven to be \mathcal{NP} -complete [99].

4.2 Ordinal Structures

As data may be ordered by different criteria, we define, more generally, an ordinal structure $\mathbf{P} = \langle P, (\leq_i)_{i \in I} \rangle$ as a set P of objects that is equipped with quasi-orders $\leq_i [176]$.⁴ Quasi-orders are transitive, reflexive relations. Not insisting on anti-symmetry for \leq_i means that there may exist objects that are indistinguishable in this relation (e.g., there may be different objects with identical prize).

▶ Example 1 (Ordinal Structure). Train connections between two cities are an example for an ordinal structure. They can be ordered by different criteria, e.g., by departure time (closer to time given in the query is better), travel duration (shorter is better), number of transfers (fewer is better), expected demand (lower is better), and price (lower is better). Figure 1 shows the *pareto optima* for connections between Hildesheim and Kassel for a requested departure at 3pm, i.e., those connections for which it is impossible to improve a criterion without becoming worse in any of the other criteria.

In our vision, we extend the notion of ordinal scale to allow for the analysis of all kinds of ordinal data. With terms like "order", "ordinal", we always refer to arbitrary orders – be they

⁴ The entities may – and in most cases will – have additional attributes on other levels of measurement. The development of hybrid analysis methods is of particular interest here.

linear or not.⁵ We thus go beyond Stevens' original understanding of "ordinal scale" which only referred to linear orders. Even though this seems to be only a minor modification at first glance, it has far-reaching consequences for the question of what "measurement" actually is, as well as for the complexity of computational problems.

Ordered sets in general do not provide much structure. However, every ordered set is embeddable into a complete lattice, its *Dedekind-MacNeille completion* [135] (which generalizes Dedekind's construction for embedding \mathbb{Q} in \mathbb{R}). This allows for tapping into the rich algebraic theory of lattices and of universal algebra [20, 87, 30]. The class of lattices forms a so-called *variety*, which implies that every homomorphic image, every factor lattice and every subset of a lattice that is closed under join and meet, as well as all direct products and direct sub-products of lattices are lattices again. Hence, these operators can be exploited for composing and decomposing lattices – and thus eventually of the ordered sets that generated them. In the worst case, the size of the Dedekind-MacNeille completion is exponential in the size of the ordered set. However, this growth is expected to be tame for real-world data, due to low amounts of multi-inheritance in it. Ganter and Kuznetsov provided an algorithm for its computation [73] in $O(c \cdot w \cdot n^2)$ with c being the size of the completion, w the width of the ordered set and n its size. The new nodes bear (yet unexplored) potential for identifying substructures in ordered sets.

4.3 Formal Concept Analysis and knowledge spaces

A direct application of lattice theory to data analysis has been established as *Formal Concept Analysis* (FCA) by R. Wille [216]. It follows the spirit of Boole, Peirce and Schröder by computing concept hierarchies from datasets. In its most basic form, FCA derives a *concept lattice* from a so-called *formal context*, a dataset (G, M, I) consisting of a set of objects G, a set of binary attributes M, and $I \subseteq G \times M$ indicating which objects have which attributes. *Concepts* are all maximal pairs (A, B) with $A \subseteq G$ and $B \subseteq M$ s. t. all objects in the *extent* A share all attributes in the *intent* B, and vice versa. The set of formal concepts is denoted by $\mathfrak{B}(G, M, I)$ and the relation $\leq \subseteq \mathfrak{B}(G, M, I) \times \mathfrak{B}(G, M, I)$ with $(A, B) \leq (C, D) :\iff A \subseteq C$ constitutes an order relation on $\mathfrak{B}(G, M, I)$. Thereby (A, B) is designated as a subconcept of (C, D). More importantly, \leq is a lattice order. This fact follows from the *basic theorem on Concept Lattices* [78, Theorem 3]. This theorem states that for any subset $\{(A_i, B_i)\}_{i\in I}$ of $\mathfrak{B}(G, M, I)$ there exists an infimum (meet) $\bigwedge_{i\in I}(A_i, B_i)$ and a supremum (join) $\bigvee_{i\in I}(A_i, B_i)$ in \leq . The set of all concept extents forms a closure system, i. e., it is a set system which is closed under intersection, and the same holds for the set of all concept intents.

Example 2. The following table depicts a real world data set about paintings by Rembrandt (the object set G) and their properties (M), taken from "Concept lattices and conceptual knowledge systems" [215]. The formal context $\mathbb{K} = (G, M, I)$ and its incidence I is represented by means of a *cross table*:

K	Family Portrait	Group Portrait	Oak	Canvas	≥ 1660
Nightwatch		×		×	
Anatomical lessons		×		×	×
Portrait Titus	×			×	×
Staalmeesters				×	×
Mother	×		×		

⁵ When addressing linear orders, we will state this explicitly. For sake of linguistic simplicity, we will also summaries all levels of measurement above ordinal – essentially interval and ratio – as "numerical".

6:8 Towards Ordinal Data Science

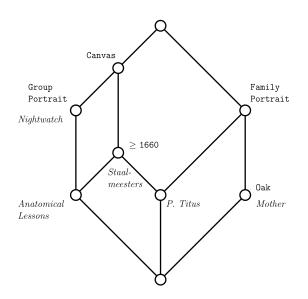


Figure 2 Line diagram of the concept lattice from Example 2. The drawing uses a short-hand notation for the labels. The nodes represent the set of formal concepts. Attributes are drawn on top of nodes and objects below. Only the first occurrence of an attribute and the last occurrence of an object is annotated (in the reading direction from top to bottom).

The line diagram of the concept lattice corresponding to \mathbb{K} is shown in Figure 2. Such a line diagram displays the covering relation, i.e., the subset of \leq which contains only the comparable elements that are immediate neighbors. From the lattice we can infer various types of information. For example, we find that all paintings that were painted in 1660 or later were painted on canvas. More intriguingly, family portraits that were painted on canvas occurred in 1660 or later. For a comprehensive list of how-to infer information from lattices, and in particular lattice diagrams, we refer the reader to the literature [78]. For a more in-depth discussion of questions concerning the diagrammatic representation of order structures in the specific case of the present work, please refer to Section 9.

Via *conceptual scaling* [77], FCA can be applied to any kind of data. Since its foundation in 1982, FCA has grown to an extensive theory [79, 33, 74, 64] and has seen several extensions, for instance, to fuzzy concepts [11]. It has been applied in a large variety of domains (see, e.g., [195, 68] for surveys). FCA has been connected to other disciplines, such as Software Engineering [167], Psychology [169], Scientometrics [44], and Social Choice Theory [108].

Similar to FCA is the theory of knowledge spaces by J.–P. Doignon and J.–C. Falmagne [45]. They define a knowledge structure as a pair (Q, \mathcal{K}) where Q is a non-empty set and \mathcal{K} is a family of subsets of Q. The set Q is called *domain* and its elements are called *questions* or *items*. The subsets in \mathcal{K} are called *(knowledge) states*. The intuition for this definition is that Q is a collection of questions of some school subject, and that the sets in \mathcal{K} describe which combinations of questions could be solved by individual students. A *knowledge space* is then a knowledge structure that is closed under set union. This is inspired by the assumption that if two students have knowledge states K_1 and K_2 , resp., then it should be possible for them to learn what the other student knows, resulting in $K_1 \cup K_2$.

A knowledge space is thus a kernel system, i.e., a set system closed under union. If \mathcal{K} is a kernel system, then $\{Q \setminus K \mid K \in \mathcal{K}\}$ is a closure system, and vice versa. This means in particular that all theoretical and algorithmic results of FCA can directly be transferred to knowledge spaces, and vice versa (see in particular [72]). From a mathematical viewpoint, the structures of closure

systems, kernel systems, complete lattices and concept lattices are equivalent [216], as there exist natural 1-to-1 correspondences. These structures are also in a natural 1-to-1 correspondence to sets of implications (of propositional Horn logic) [214] and functional dependencies (of database theory). These natural correspondences mean that for the computation, analysis and (de)compositions of such structures one can switch to the representation that is most suitable for the task at hand. Formal contexts play an important role when it comes to computations, as their sizes are logarithmic in comparison to the sizes of the corresponding (concept) lattices.

4.4 Ordered Sets and Lattices in Computer Science

Specific ordered sets and lattices are frequently found in computer science: For instance, queues and lists are linear orders, (B-)trees are used for indexing data, tries for Information Retrieval, CPOs for modeling the semantics of recursion, and Boolean algebras for propositional logic implemented in logical gates. Inheritance (including multiple inheritance) in object-oriented programming induces a partial order on the set of classes. An introduction to the mathematical foundations of order theory relevant to *Data Mining* is provided in [166]. Ordered sets are also the key ingredient for modeling dependencies between tasks in scheduling problems.

FCA is used extensively in the field of Ontology Engineering / Semantic Web, in particular for conceptual information systems [192], knowledge acquisition [186, 187, 189], for knowledge base completion in Description Logics [6, 7], for computing probably approximately correct implication bases [24], for ontology management [185, 76, 37], [160, 14], and text corpus analysis [104]. An extensive survey on FCA based models and techniques in knowledge processing is given in [153].

4.5 Algorithms and Software for Ordered Sets and Lattices

The number of algorithms and software tools in the field of order theory and (concept) lattices is extensive. Listing all relevant works would go far beyond the scope of this article; we only mention the most important and recent tools. Ordinal Data Science may benefit from algorithms for the uniform sampling of linear extensions [29] and the uniform sampling of ordered sets [119], and the even more complicated problem for sampling binary relations in general, which was tackled in [61]. The research field of Formal Concept Analysis can contribute algorithms for the fast generation of concept lattices, e.g. In-Close4 [5] or Parallel Close-by-One [120]. Also, the computation of (inter-feature-) dependencies will be an essential asset. The computation of such dependencies can be done with proven parallel canonical base algorithms [122]. One may also resort to approximate methods, since the related computational problems become computationally intractable with increasing input size. For example, the dependencies can be computed with probably approximately correct learning methods [25, 24], a combination of order theory and classical machine learning model. Many of the mentioned algorithms are already available, or are currently being implemented, in the analysis software conexp-clj [91], which is a reimplementation of conexp by means of the functional programming language *Clojure*. It is currently maintained and further developed in the group of the authors. There are many further FCA software tools under active developed (and even more historic ones), most with a particular specialization. Just to name a few: $Graph-FCA^6$, an extension of FCA to knowledge graphs; *lattice* miner [142]; conexp- ng^7 , an educational variant of conexp; $fcaR^8$, a package for the statistical software suite R.

⁶ https://bitbucket.org/sebferre/graph-fca/src/master/

⁷ https://github.com/fcatools/conexp-ng

⁸ https://cran.r-project.org/web/packages/fcaR/fcaR.pdf

6:10 Towards Ordinal Data Science

In the next five sections, we will now discuss exemplarily five out of many possible lines of research in the new field of Ordinal Data Science in more detail.

5 Ordinal Measurement Theory

While measurements in physics were accepted without objections, other disciplines, in particular psychology, asked in the 20th century – and still do so today – if and under which conditions non-physical sensations such as loudness, fear, or intelligence can be measured. The most widely acknowledged answer to this is the *Representational Theory of Measurement (RTM)* [121, 198, 131], which aims at identifying criteria that justify a measurement on a given level. These criteria can (at least in principle) be verified empirically without actually performing the measurement. RTM defines *measurement* as a homomorphism from an empirical relational structure $\mathbf{A} = \langle A, (R_i)_{i \in I} \rangle$ – i.e., a set A of empirical objects (such as a collection of rods) equipped with qualitative relations R_i (such as "rod a is shorter than rod b" and "when laid end to end, rods a and b together have the same extension as rod c") – to a *numerical relational structure* $\mathbf{B} = \langle B, (S_i)_{i \in I} \rangle$ [151, 121]. In the latter, B is usually the set \mathbb{R} of real numbers (or \mathbb{R}^n and other variations) together with relations such as x < y and x + y = z, even though several researchers [158, p. 253] observe that in some cases it might be beneficial to allow for other algebraic – "non-numerical" – structures.

RTM puts a strong emphasis on measurements into the real numbers, and less emphasis on ordinal phenomena. As it allows for linear ordinal scales only, RTM does not allow for studying phenomena involving incomparability or multi-inheritance. Therefore, inspired by RTM, R. Wille et al. [75] developed a theory of *conceptual measurement* where **B** takes the form of specific lattices, called *conceptual scales* – a first step towards an ordinal version of RTM. In this context several studies followed [77, 224, 154, 188, 182]. Conceptual measurement provides structural statements about homomorphisms to standard scales, but it does not yet provide a full theory about the conditions under which ordinal phenomena can be measured with the different types of ordinal scales. In order to provide a *theoretical foundation to Ordinal Data Science*, we propose to extend RTM by non-numerical scales. This will allow for empirically testing on a given dataset which scales are appropriate for its analysis.

To this end, in this section, we suggest to analyze under which conditions measurements into (linear and in particular in non-linear) ordinal scales exist, and how unique they are (i. e., under which permissible transformations they remain invariant). We further suggest investigating how the – potentially high – order dimension of the original data relates to the order dimensions of a set of scales which together fully measure the dataset. At last, we consider non-linear versions of Guttman scales [89], to allow for measuring simultaneously objects and their features.

In the long run, this endeavor may open the avenue to a "Grand Unified Theory of Measurement" which will allow, for instance, in (the philosophy of) physics to discuss the analysis of symmetries of elementary particles (by letting \mathbf{B} be a (non-ordered) symmetry group) in the same terms as the numerical measurement of their mass and charge.

5.1 State of the Art and Open Questions

Based on the successful role of measurement in physics and the demand of other disciplines – in particular psychology – , in the late 19th and early 20th century a universal definition of measurement was sought [63]. S. S. Stevens [174] provided an operational approach by distinguishing four *levels of measurement*: nominal, ordinal, interval, and ratio; and characterized them by families of automorphisms of the set of real numbers that respect the scale type ("admissible functions"). Stevens' levels of measurement have been (and still are [162, 202]) under dispute. A particular controversial question among practitioners that is discussed since Stevens' paper for over 70 years is whether computing the mean of ordinal data is an allowed operation or not (see, e.g., the

football number dispute [130, 162] as illustrative example). In practice, this is frequently done (e.g., aggregating the jury votes in gymnastics or figure skating), while in other cases this is considered bad practice (e.g., aggregating reviewer judgments in scientific peer review, where often the average is considered as bearing little meaning, and therefore consensus is sought). Several extensions were proposed for Stevens' levels (e.g., [144, 35]), see [201] for a systematic survey. All those extensions, however, consider ordinal data as a separate category; so our endeavor is widely agnostic to this discussion.

Stevens did not answer the question of how to decide, in a given setting, which level is appropriate. *Representational Theory of Measurement (RTM)* [121, 198, 131], [151, 158] provides an answer with the notion of conjoint measurement. RTM defines a *measurement* as a mapping of an empirical relational structure (ERS), i. e., a set of objects equipped with qualitative relations such as "is larger than", to a numerical relational structure (NRS). Luce and Tukey [132] proved that whether some empirical sensation is a continuous variable depends on characteristics of the ERS and is thus in principle empirically testable. How feasible this testing is in practice is one of the questions of the still ongoing discussion on RTM, in particular in psychology [204, 28]. The envisaged substantially new methods in Ordinal Data Science for measuring ordinal data may turn RTM more acceptable to practitioners, as it is easier to show that ordinal methods are legible in a given situation than it is for methods requiring interval or even ratio level.

Several researchers working on RTM (e. g., [158, p. 253]) stated that it might be beneficial not to focus solely on the ordered field of real numbers as NRS, but to allow instead for other mathematical structures. However, this approach has been followed rarely. For instance, R. Wille and his group took up this line of thought, as described above: On one hand the group studied the representation of ordinal structures in the real numbers [223, 224, 226] and weaker algebras such as ordered quasi-groups [225]. On the other hand, work on a theory of *conceptual measurement* [75, 77, 188, 182, 154] was started, where \mathbb{R} is replaced by complete lattices which can be understood as *partial ontologies*.

As Dedekind-MacNeille completions preserve important properties of ordered sets (e.g., order dimension) and are isomorphic to a concept lattice, conceptual measurement is just a special – but conceptually and computationally convenient – way of formalizing a general theory of ordinal measurement. Conceptual measurement has frequently been used in applications. However, a theoretical investigation in terms of RTM (e.g., what are the equivalences of cancellation and solvability when \mathbb{R} is replaced by a lattice?) is still lacking.

5.2 Promising Research Questions

The main objective in this field of research will be to extend Relational Measurement Theory to non-numerical scales, as it would allow to empirically test on given data which of these scales are appropriate. We will address three key aspects: The first is to establish theorems about the existence and uniqueness of measurements into non-linear ordinal scales and to explore how well the theory can be extended to fundamentally different scale types (e.g., (non-ordered) symmetry groups), as a first step towards a "Grand Unified Theory of Measurement". As the order dimension is a strong indicator for the complexity of an ordered set, we will then discuss the relationship between the order dimensions of an ordinal structure and its scales, in order to provide guidance for suitable decompositions and factorizations in Sections 7 and 8; and extend the discussion to scales that measure simultaneously the objects and their attributes.

6:12 Towards Ordinal Data Science

5.2.1 Measuring into ordinal scales

Research in Representational Measurement Theory is centered around the real numbers as "numerical relational structure" (including derived structures, such as semiorders and interval orders for modeling error and variation) and real vector spaces. Measurements with other types of ordered sets have not been studied systematically, except for the theory of conceptual measurement as described above. So a general research question will be: how does measurement theory look like if we replace $(\mathbb{R}, \leq, +, -, \cdot, 0, 1)$ by a specific ordered set (P, \leq) or (complete) lattice (L, \leq, \wedge, \vee) ? Are there equivalences to the notions of cancellation, solvability and conjointness, and how are they constituted?

Following the line of traditional RTM, we suggest to study in particular the representation problem and the uniqueness problem for specific scale types. The *representation problem* deals with the question of which conditions a dataset has to fulfil in order to permit a measurement in a given scale. For binary attributes on the empirical side and many standard ordinal scales (nominal, linear, interval, multi-ordinal, contra-nominal, contra-ordinal, and convex-ordinal scales) on the "numerical" side, this problem was addressed in [79, Chapt. 7]. In this case, the task is reduced to translating the results to the terminology of RTM. For other types of empirical data, and for other types of scales (e. g., trees) and ordinal scales in general⁹ on the "numerical" side, the necessary theorems still have to be established. Once a measure has been identified, one should also consider the *uniqueness problem* by analyzing invariance under permissible transformations: How unique is the resulting measure or scale?

Our results constitute the first steps towards a "Grand Unified Theory of Measurement": Once a theory for ordinal measurement as been established, one may start to explore – as a side-trip and preparation of more extensive research – how well this theory can be extended to fundamentally different scales, as for instance to non-ordered groups (such as symmetry groups).

5.2.2 Dimensionality

The complexity of an ordered set is not only determined by its cardinality, but also to a large extent by its order dimension. In contrast to numerical data, where most frequently each feature spans one dimension, the identification of the different dimensions of an ordered set is not straightforward – even determining their number is \mathcal{NP} -complete [228]. The question of dimensionality thus takes a completely different stance as in standard RTM (where dimensional analysis means that physical dimensions such as length, mass, time duration, speed, and their relationships are studied [121, Chapt. 10]). Breaking down a complex ordered set into smaller, less complex parts (i. e., parts with lower dimensionality) is an important – but challenging – task for data science, which we will discuss further in Sections 7 and 8. The following example shows that the development of a theory about the dependencies between the ordinal dimension of an ordered set and the dimensions of the scales it can be measured to might be of high interest.

▶ Example 3 (Order Dimension). Table 1 depicts the final standings of the German Soccer Bundesliga season in 2022/23. In this dataset, the strength of the soccer clubs is measured using four different scales. In the columns "Won" and "Goals for", a high value implies a strong club, while the converse is true for "Lost" and "Goals against". The direct product of these four linear orders describes an order relation in which two clubs are comparable, if one dominates the other in all four categories. Two clubs may be incomparable in this ordering, as for instance Borussia Dortmund has more wins than FC Bayern München, but also more losses.

⁹ Reminder: In this project, ordinal scales may be any type of ordered set, either linear or non-linear. What is usually called ordinal scale (such as Likert scale or Beaufort) is a special, linear case thereof.

Pos	Team	W	D	\mathbf{L}	\mathbf{GF}	GA	GD	\mathbf{Pts}
1	FC Bayern München	21	8	5	92	38	+54	71
2	Borussia Dortmund	22	5	7	83	44	+39	71
3	RB Leipzig	20	6	8	64	41	+23	66
4	1. FC Union Berlin	18	8	8	51	38	+13	62
5	SC Freiburg	17	8	9	51	44	+7	59
6	Bayer 04 Leverkusen	14	8	12	57	49	+8	50
7	Eintracht Frankfurt	13	11	10	58	52	+6	50
8	VfL Wolfsburg	13	10	11	57	48	+9	49
9	1. FSV Mainz 05	12	10	12	54	55	-1	46
10	Borussia Mönchengladbach	11	10	13	52	55	-3	43
11	1. FC Köln	10	12	12	49	54	-5	42
12	TSG 1899 Hoffenheim	10	6	18	48	57	-9	36
13	Werder Bremen	10	6	18	51	64	-13	36
14	VfL Bochum	10	5	19	40	72	-32	35
15	FC Augsburg	9	7	18	42	63	-21	34
16	VfB Stuttgart	7	12	15	45	57	-12	33
17	FC Schalke 04	7	10	17	35	71	-36	31
18	Hertha BSC	7	8	19	42	69	-27	29

Table 1 The final league table of the 2022–23 Bundesliga, Germany's premier soccer league. (Pos: Position, W: Won, D: Drawn, L: Lost, GF: Goals for, GA: Goals against, GD: Goal difference, Pts: Points)

The order dimension of the domination order of the 2022/23 season is three. This means that only three linear orders are necessary to represent the combination of the four scales. There is hence some redundancy in the data – but not enough to produce a total ranking. The latter would hold if the order dimension of the domination order would be one. It would then, for every pair of clubs, provide an answer to the natural question of which of the two is better than the other. As this is not the case, further rules are used to compute the final position of each club in the league. This results in a linear extension of this domination order. The choice of these rules captures the understanding which quality criteria are considered most important.

The current rules of Bundesliga [127] are one possible way to reach such a linear extension: First, a "Points" scale is derived, where each team is awarded three Points for every win and one point for every draw. This results in a linear extension of the direct product of the "Won" and "Lost" scales, which are considered to be the most important criteria. However, this may not yet result in a total ranking as multiple teams can have the same number of points (for instance Borussia Dortmund and Bayern München). In this case, the tie is broken using the "Goal difference", which itself is a linear extension of the direct product of "Goals for" and "Goals against". Thus, from the original four scales, the two linear extensions "Points" and "Goal difference" are computed as derived scales. Then, these two give the final ranking by taking once again a linear extension of the two derived scales.¹⁰

Note that the rules are arbitrarily chosen (e.g., , one could assign only two points for a win, as is done in other sports (and was done earlier in Bundesliga), or consider the direct comparison of clubs in a tie as second criterion) and for every linear extension of the domination order there is a set of rules that gives rise to the ranking described by this extension. Thus, one could consider each linear extension a valid final ranking of the participating teams.

 $^{^{10}}$ [127] provides further rules for cases where this still does not resolve all ties.

6:14 Towards Ordinal Data Science

A next research question could be a dimensional analysis of combinations of the standard ordinal scales – for each of them studied separately, the answer is straightforward, but dimensional properties of full measures of an ordered set in a (heterogeneous) collection of scales have not yet been studied systematically. As deciding the order dimension of an ordered set is \mathcal{NP} -complete, one has to keep an eye on the computational tractability: Are exact algorithms fast enough for typically sized datasets? Alternatively one may resort to computing upper bounds, together with the *realizers* (i. e., the spanning linear orders) [229]. Also, the development of efficient algorithms for computing lower bounds (and eventually also the critical pairs) will be of interest.

5.2.3 Joint scales on objects and attributes

Guttman scales have been introduced as joint scales for objects and attributes. Consider, for instance, a set G of students, a set M of exercises of (assumingly) increasing difficulty, and a relation I stating which student solved which exercise. If the corresponding concept lattice is linearly ordered then it provides a linear ranking both for the difficulty of the exercises and for the ability of the students. This can be interpreted such that there is no exception to the rule: a student who is capable of solving a specific exercise can also solve all easier exercises. This is equivalent to the original definition [89] which requires the existence of maps $s: G \to \mathbb{R}$ and $e: M \to \mathbb{R}$ s.t. $(g,m) \in I \iff s(g) \le e(m)$. In practice, Guttman scales can be observed frequently, but often only up to some "noise" or "impurity". Standard approaches are then either to assume that the assumptions hold in principle and to statistically find best-fitting functions [38]. or to analyze whether the data result from more than one Guttman scale. We propose to follow two new order-theoretic approaches for analyzing these "impurities" instead. The first approach is to relax the condition of linear order and to explore if a more general scale – e.g., a tree – may be suitable as a joint scale. One can then identify necessary and sufficient conditions for the existence of such generalized Guttman scales as well as criteria for their minimality (e.g., the lowest cardinality, the lowest dimensionality, least multi-inheritance). The second approach is to analyze more closely the disturbance of the linear scale by identifying which parts of the scale increase its *local dimension* [207]. Such a tool would allow analysts to investigate whether a linear feature is disturbed by "noisy data" or whether it really shows some unexpected non-linear characteristics.

6 Ordered Metric Spaces

Real-world data is feature-heterogeneous [31], i.e., it is (apart from others) composed of ordinal and numerical dimensions. The former can be viewed as a partially ordered set, and the latter often leads naturally to a metric. In principle, both can be analyzed separately, but this approach has two disadvantages. First, the results calculated in this way can be contradictory to each other. Second, the knowledge about the connection of the two feature dimensions is not exploited.

We therefore propose to develop theoretical and practical methods that honor the compatibility (or consistency) of order relations and metrics on a set. To this end, we envision the unified representation of such data by means of *ordered metric spaces* (*om-spaces*) (P, R, d), where Pdenotes a set, R is an order relation on P, and a d a metric $d: P \times P \to \mathbb{R}^+$. This modeling is inspired by similar investigations in *metric measure spaces*¹¹ [146, 196].

¹¹ In metric measure spaces, *measures* are those of mathematical analysis, being defined over σ -algebras [90]. They model a different aspect of the concept of measurement as the homomorphisms between relational structures in RTM do, and should not be confused with them.

The corresponding highly ambitious task, due to its novelty, is to develop a structure theory for om-spaces. Particularly interesting in this context is the study of om-space embeddings into Euclidean space. Since the latter is a very commonly used space for machine learning, the question of how to measure *ordinal distortions* is imperative. Vice versa, it is important to understand to which extent the extraction of linear orders from data via valuation functions is meaningful with respect to om-spaces. With a new class of distortion measures – and computationally tractable approximation algorithms – which allow to quantify the compatibility of R and d, one might be able to answer a range of fundamental questions about the relationship between orders and metrics, including: 1) If a metric has been (machine) learned based on information about the order, how consistent is the metric with the order? This will provide a fundamental new quality criterion for order-related machine learning tasks. 2) How strong do order and metrics "agree" with their view on the dataset? An answer might be highly beneficial to an investigation whether order and metrics reflect the same hidden structure. 3) How justified is it to represent the order by means of a metric? If the answer is satisfactory for a given dataset, this would justify the use of the rich toolkit of numerical data analysis and machine learning methods for analyzing the ordinal structure. If not, this would indicate that alternative (ordinal) methods are mandatory in this particular case.

6.1 State of the Art and Open Questions

Order relations entail many geometric properties. Likewise, metric spaces, which play a dominant role in data science, have a rich geometric structure. Examples for datasets exhibiting both a metric and an order relation are numerous. For example: Humans have (distances between their) birthplaces and an ancestry order relation; scientists have an ancestry relation (doctoral advisor) and different proximities (topical, geographic, social); mountains are ranked by height and prominence [161] and have a geographic distance. A more general example for an om-space is the set of Airports together with the connecting plane routes relation and the geodesic distances.

The analysis of ordinal data is often based on a (numerical) metrification. Imposing metric structures on ordered sets¹² is a widely studied [143], yet still very active, field of research. Almost all mathematical disciplines contribute to this, e.g., algebra [60], combinatorics and geometry [128], and analysis [110]. Furthermore, the task of (distance) metric learning in the realm of (weakly) supervised machine learning addresses a similar problem, e.g., based on the generalized *Mahalanobis distance*. Contemporary developments [211] do often use deep learning settings. However, recent works claim that the reported achievements are rather marginal [145]. We contributed to the task of finding meaningful metrics in ordinal data in two ways. First, we adapted the notions of dominance and prominence from the research field of orometrics to network science [161]. This process revealed a whole class of new metrics to study. Second, we have further developed the popular *word2vec* method, which is based on a neural network model, for the special requirements of (concept) lattices [48].

A popular embedding structure for ordinal and metric data in the realm of machine learning is the Euclidean space, especially represented as \mathbb{R}^n . Depending on the properties of the relation (symmetric, transitive, reflexive, etc.) and the kind of embedding, e.g., *distance*, *similarity*, or *transitional*, different bounds for n [18] are required for distortion free, i.e., isometric, embeddings. Often n equals |P|, which may be computationally intractable. Moreover, it is a well known fact [136] that many finite metric spaces are not distortion-free embeddable into \mathbb{R}^n , for any $n \in \mathbb{N}$. A simple example for this is a 4-element ring with the graph distance as a metric. Therefore, low

¹² In the realm of machine learning a natural consequence when learning an "embedding".

6:16 Towards Ordinal Data Science

distortion embeddings [1] and different means of measuring distortion [210] are widely studied. Hence, the investigation of distortions for embedding finite om-spaces into \mathbb{R}^n , which takes into account metric and order relational characteristics, is at hand, highly ambitious, and up to our knowledge, not conducted. We have already found initial evidence in a special case that the distortion can be extensive [92].

Understanding different metric embeddings of the same ordinal data is possible by comparing the resulting metric spaces. Amongst others, the Gromov-Hausdorff (GH) distance (by D. Edwards [55, 209], M. Gromov [86]) enables such a comparison naturally. However, since GH is computationally intractable [138] one might resort to computable lower bound methods, e.g., the modified GH distance [147], which is time polynomial.

To further mitigate the computational demands in the last paragraphs, we can refer the reader to various works we carried out, e.g., the identification of important features in relational data [94] based on entropy maximization and structure preservation, dataset size reduction using cores in formal contexts [93], and the identification of clones [43], i. e., concept lattice preserving permutations. All works are applicable to om-spaces, as every order relation gives rise to a (unique) concept lattice; and every metric on P can be lifted to the power set.

To the best of our knowledge, studies purely concerned with the interplay of order relations and metrics with respect to data analysis are rare. We can identify in the literature few results about fix points [126, 17] and contractions [2]. Related topological results for ordered metric spaces are of less interest for our envisioned task, as the considered om-spaces are almost always finite, which in turn implies that the induced topological space is discrete.

Coming full circle with our motivation from mm-spaces, we propose to consider different results on valuation functions in ordered sets. This topic is already studied for order relations and lattices [143, 124]. We already contributed to this by transferring ideas from orometrics to bounded metric spaces endowed with binary relations [179, 180]. Even more related to the realm of mm-spaces is our work on an intrinsic dimension of geometric datasets [95, 178]. Although the proposed dimension function relies on mm-spaces, it is fully capable for the dimension analysis of om-spaces.

Finally, comparing real-world data with randomly generated data often reveals new structural insights and provides benchmarks. Yet, uniform sampling of ordered sets is a difficult problem, at least for reasonable set sizes [119]. Our previous work allows to some extent for randomly generating binary relations [61] and null-model generation for (order) relations [62]. To which extent these methods can be useful for the study of om-spaces is an open question.

6.2 Promising Research Questions

Central to this research is the theoretical and experimental study of the triple (P, R, d), as introduced in the last section as ordered metric space (om-space). One should distinguish between originally metric data (and therefore spaces) and those where the metric was obtained through an external (e. g., machine learning) method. Special interpretations of this structure are, for example, directed acyclic graph metric spaces (if the metric values are additive along paths), or weighted directed acyclic graphs (when considering metric values for connected pairs, i.e., relational pairs, only). One should approach the goal for a structure theory of om-spaces for ordinal data science from different angles, most importantly graph theory, order theory, and the analysis of metric spaces.

Any new result will surely build up on modelings concerned with R being a tree or forest relation and d a graph metric on it. The guiding methodology should be to examine the extensive apparatus of order theory with respect to the om-space. In particular, it is necessary to develop new concepts and methods to assess algorithmically the compatibility of R and d. A special focus

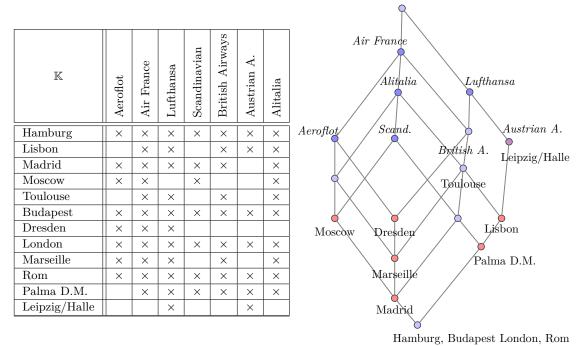


Figure 3 Left: Simplified formal context from real-world data about cities (objects) that are serviced

by airlines (attributes). **Right:** Corresponding concept lattice.

should be set on understanding if and how different kinds of dimensionality (e.g., order dimension, intrinsic dimension) do inflict the compatibility of R and d as well as a possible later embedding of (P, R, d) into Euclidean space.

6.2.1 Structure Theory of om-Space Datasets

To explore the distortion of order relations on sets with respect to indigenous and externally imposed metric functions is the focus of this research task. A first attempt to assess the distortion of an order relation in this sense can be based on the distortion of a map φ between two metric spaces (X, d_X) and (Y, d_Y) . In the realm of machine learning numerous distortion functions are studied [210]. A promising start may be with $dis(\varphi) := \sup_{x,\hat{x}\in X} |d_X(x,\hat{x}) - d_Y(\varphi(x),\varphi(\hat{x}))|$. To study the distortion introduced by relation R in om-space (P, R, d), one could then study the canonical (by means of the Galois connection) map $\varphi : P \to 2^P$ defined by $x \mapsto \{y \in P \mid (x, y) \in R\}$ from metric space (P, d) to $(2^P, d_H)$ where d_H is the Hausdorff metric derived from d. This setting allows for analytical tools to assess the relational distortion (as well as expansion, extraction, etc.) of ordered metric spaces for real world datasets. If R is a purely reflexive relation, i.e., the simplest order relation on P, there will be no distortion, i.e., $dis(\varphi) = 0$. Yet, for arbitrary relations in real-world datasets the change of distortion is unknown.

Example 4. The data shown in Figure 3 and Table 2 can be comprehended as an om-space. The elements of P are the cities, the map d assigns any two cities their geodesic distance, and any two cities are in relation R iff there is an airline servicing both cities.

Applying this new measure on real-world data would require non-incremental research to develop efficient algorithms to compute them, as computational demands increase drastically with increasing |P|. One might face this highly ambitious challenge by drawing from results on closed sets, which arise naturally from the closure operator generated by R. Hence, one may employ

	Hamburg	Lisbon	Madrid	Moscow	Toulouse	Budapest	Dresden	London	Marseille	Rom	Palma D.M.	Leipzig/Halle
Hamburg	0	2198	1781	1765	1278	950	378	745	1186	1326	1658	289
Lisbon	2198	0	513	3892	1029	2480	2250	1564	1307	1839	1027	2174
Madrid	1781	513	0	3420	537	1975	1784	1246	797	1330	546	1718
Moscow	1765	3892	3420	0	2883	1563	1643	2508	2666	2399	3119	1722
Toulouse	1278	1029	537	2883	0	1453	1249	883	312	911	468	1188
Budapest	950	2480	1975	1563	1453	0	573	1490	1180	836	1590	674
Dresden	378	2250	1784	1643	1249	573	0	988	1070	1044	1546	111
London	745	1564	1246	2508	883	1490	988	0	989	1444	1348	879
Marseille	1186	1307	797	2666	312	1180	1070	989	0	602	479	1032
Rom	1326	1839	1330	2399	911	836	1044	1444	602	0	839	1070
Palma D.M.	1658	1027	546	3119	468	1590	1546	1348	479	839	0	1511
Leipzig/Halle	289	2174	1718	1722	1188	674	111	879	1032	1070	1511	0

Table 2 Geodesic distances between cities (in nautical miles [nmi]) that are the objects of the formal context in Figure 3.

fast algorithms for the generation of closed sets (or their generators), which will facilitate the computation of the Hausdorff distances for large sets P. Starting from this promising modeling, the next step would then be to test, examine and develop further distortion measures in the realm of om-spaces.

6.2.2 Formal Context Mediated Metrics and Orders by Valuation

In a more general setting one might also want to study arbitrary binary relations, as often investigated in the form of a formal context (G, M, I). In this regard, the following question is directly obvious. How does the binary incidence relation $I \subseteq G \times M$, and therefore the inherent lattice structure, mediate a metric from one set to another? In detail, given the incidence Ibetween sets G and M, where G is equipped with a metric d, one can induce a metric on Musing the afore mentioned Hausdorff metric on (G, d) and the map $m \mapsto \{g \in G \mid (g, m) \in I\}$. What properties are expected of the resulting metric space (M, d_M) ? What alterations of I, i.e., operations based on adding or removing relational pairs, do change the properties of (M, d_M) ? Based on this one might also want to investigate the special case in which both sets G and M are equipped with an indigenous metric. Then, the natural question arises: how compatible are these indigenous metrics with the mutually induced ones mediated by the incidence I? For this, the computation of distance between metric spaces is essential, for which it might be necessary to employ the modified GH distance [147].

▶ **Example 5.** It is clear that in Example 4 only the object set G is equipped with a distance function (actually a metric) d_G . In order to obtain a distance function between the airlines, i.e. the attributes, the most obvious way is to calculate a Hamming distance between them. This distance, however, would not consider the existing data for G. Based on the proposition above, we can infer a distance function that incorporates d_G . For example, d_M (Scandinavian, Austrian) would be mapped to the Hausdorff distance $d_H(M_1, M_2)$ where

 $M_1 = \{$ Hamburg, Madrid, Moscow, Budapest, London, Rom, Palma D.M. $\}$

 $M_2 = \{\text{Hamburg, Lisbon, Budapest, London, Rom, Palma D.M., Leipzig/Halle}\}.$

Hence, one has to compute

$$\max\left\{\sup_{x\in M_1} d(x, M_2), \sup_{y\in M_2} d(M_1, y)\right\},\$$

where $d(u, V) \coloneqq \inf_{v \in V} d(u, v)$, i.e., the distance from u to V. In our example the computation of the Hausdorff distance results in max{836, 1563} = 1563 nmi.

Closely related to the setting above is the application of valuation functions to generate (linear) order relations. An example is to rank objects from G based on sizes of their shared related subsets from M, i.e., every $g \in G$ is valued by $|\{m \in M \mid (g,m) \in I\}|$. Since valuation functions almost always map into a numerical scale, e.g., the real line, this results in a metric and in turn into a linear order. Dual to the question in the first paragraph one can investigate how a valuation map distorts an indigenous ordinal property of the set, for example extracted through the notion of concept lattices. As a start, one may count how many wrongly ordered relation tuples are present. Furthermore, it might be fruitful to investigate to which extent the order created by the valuation differs from an arbitrarily generated linear extension [29]. Finally, one might re-evaluate our preliminary work in this area [161, 179, 180]. This can shed light on the question whether the valuation functions we transferred from orometry are superior when dealing with indigenous ordinal properties.

6.2.3 Intrinsic Dimension of om-Spaces and Randomness

There are numerous notions for measuring the complexity of a dataset. For ordinal data the order dimension is a natural candidate, as described in Section 4.1. Another approach is the notion for *intrinsic dimension*. Put simply, this value reflects the number of dimensions required to accurately represent the data. For incidence structures, such as formal contexts, we proposed in our previous work "Intrinsic Dimension of Geometric Data Sets" [95] a novel intrinsic dimension function that incorporates measure- and metric properties. This function is also applicable to ordinal data. This is in particular true, if the ordinal data is scaled via the *general ordinal scale*.

Yet, it is unknown if the intrinsic dimension does relate to the order dimension in any way. Still, it is to be expected that the intrinsic dimension can be used to capture ordinal complexity to some extent. To substantiate the hypothesis, an extensive experimental evaluation is needed and should be conducted, using real-world as well as randomly generated datasets. For the latter we might refer the reader to our work on random relational data generation [61]. Moreover, one can compare our results to recent notions of estimating dimensionality in the absence of distance functions [116]. Building up on this evaluation, and even more progressive, is the further development of an extension of the intrinsic dimension to om-spaces. This extension can open up a new research avenue for dataset dimensionality – however, it is dependent on highly ambitious theoretical and algorithmic results from all paragraphs above.

7 Algebraic constructions/decompositions for complexity reduction

For algebraic structures such as lattices, there exist many ways of reducing or decomposing them to smaller parts, for instance factorizations or sub-direct decompositions. However, these constructions are very sensitive to small perturbations in the data; and approximations and heuristics in the style of data mining are not (yet) considered in algebraic research.¹³ The challenge is thus to equip the algebraic constructions with approximations and heuristics s.t. they become suitable for data analysis.

We propose to develop methods for clustering, aggregating and factorizing ordered sets and lattices. For ordered sets this may include order-based modifications of clustering algorithms such as k-modes, hierarchical agglomerative clustering and density-based clustering; and one might explore which signals for clustering can be obtained from their Dedekind-MacNeille completions.

¹³ There are other approximations in universal algebra, but of a different nature, which is not relevant here: they deal with approximating infinite algebras [137] and serve, e.g., for defining the semantics of programming languages.

6:20 Towards Ordinal Data Science

For lattices, we suggest exploiting constructions of universal algebra and lattice theory (such as factorizations, tolerance relations, atlas decompositions etc.) [42] [190]. These methods, however, have not been developed with real-world data with noise or disturbances in mind – up to now there are no notions such as "the lattice is decomposable to 95%", and no algorithms dealing with them.

7.1 State of the Art and Open Questions

In universal algebra [20, 87, 30] it is a well-known fact that every homomorphic image and every subalgebra of a lattice or of a finite complete lattice and every direct product thereof is a lattice again [19].¹⁴ This means that universal algebra provides us already with a rich toolkit for decomposing and aggregating lattices. In particular, every congruence relation of the lattice (i. e., an equivalence relation that is compatible with meet and join) provides a clustering of the lattice that additionally respects the lattice operations. The interpretation as clustering works much better for lattices than for other types of algebras (e.g., groups), as each congruence class is a convex set ("cluster") and is therefore not shattered throughout the lattice. The corresponding factor lattice can be understood as an aggregate of the original lattice. Several approaches exist that weaken the requirement of a congruence relation. Tolerance relations [217], for example, allow for overlapping clusters, while our newly introduced interval relations [118] produce smaller clusters (on the cost of not respecting all meets and joins any more). Another way of reducing the complexity of a lattice is by exploiting the fact that every lattice can be decomposed into smaller, irreducible factors.

Many of these algebraic operations were linked to datasets (in the form of a formal context) within the field of FCA [69]; in particular for subdirect decompositions [155, 156], for direct products of convex-ordinal scales [175], for substitution decomposition [133, 172] and substitution product [220, 173], as well as for tensorial decomposition [219] and tensor products [222].

▶ **Example 6.** Figure 4 shows the subdirect product of two conceptual scales for a conceptual information system about flight movements at Frankfurt Airport. With this composition of scales, one can analyze the distribution of the flight movements over different dimensions. In the figure, the outer scale indicates the location of the baggage conveyor assigned to an aircraft. For instance, we can see at the left of the diagram that 2503 times a baggage conveyor in Halle A was assigned to an aircraft. The inner scale shows the position of the aircraft. Of the 2503 aircraft mentioned before, 78 were positioned on the apron (V = Vorfeld) and 2425 at Terminal 1 (T1). Both assignments are reasonable because Halle A is part of Terminal 1. However, the system also helps us to discover apparent mismatches. For instance, we find that there are 180 aircraft that had a baggage conveyor assigned at Terminal 2 but which were positioned at Terminal 1 (T1). Focussing on these 180 flight movements with further scales would support us to discover reasons for this apparent mismatch.

Complete congruence relations of concept lattices [157, 117] are a good starting point for clustering and aggregating. Complete tolerance relations and atlas decompositions [219] have been developed explicitly for allowing large lattices to be visualized in smaller parts just like several map pages in a road atlas. Decompositions and factorizations are more straightforward the more structure the lattice shows. A particular strong structure is distributivity ($x \land (y \lor z) = (x \land y) \lor (x \land z)$ and $x \lor (y \land z) = (x \lor y) \land (x \lor z)$) [19, 218, 42]. In this case, the lattice can be subdirectly decomposed into linear orders (which is beneficial for ordinal factor analysis as is discussed in Section 8).

¹⁴ The same holds for all classes of algebras that are defined solely by equations, e.g., for groups and rings.

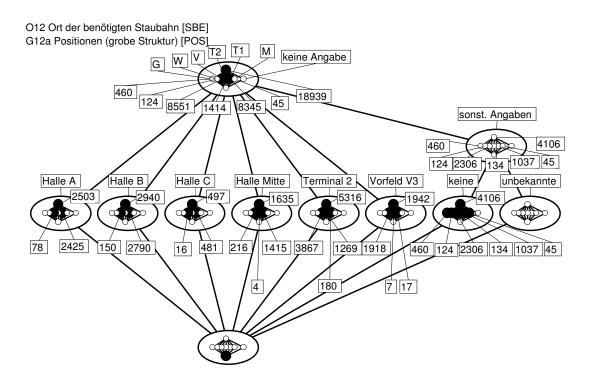


Figure 4 Subdirect product of the scales *position of baggage conveyor* and *position of aircraft* at Frankfurt Airport.

However, many of these constructions have never been used for larger data analysis tasks. This is due to the fact that lattices resulting from real-world data often do not precisely fulfil the structural conditions required for applying a decomposition or a factorization. In such a case, a lattice is considered as irreducible; there is no such notion as "95% distributive" or "reducible with an error of x". Techniques for dealing with partly imperfect data (as they exist in numerical settings, as for instance soft margins for support vector machines) are not in the spirit of universal algebra, and have thus not been developed so far. The same holds for distributivity, which is rarely observed in the whole concept lattice of real-world data. Nevertheless, empirical evidence shows that many lattices contain "a large distributive part" – a still to-be-defined concept.

The main scientific challenge in this field is thus to *develop methods for applying algebraic constructions to lattices even if the necessary structural conditions are not satisfied.* This will include the definition of measures for the degree of satisfaction of these conditions, the provision of preprocessing methods for "cleaning" the data appropriately, the establishment of a theory for compositions, decompositions and aggregations with structural impurity, and the development of algorithms and heuristics for their computation on large datasets.

7.2 Promising Research Questions

We propose three specific research questions in this area. The first two consider clustering in ordered sets and lattices, resp., while the third one focuses on decompositions of lattices. Considering the required research approaches, it is the latter two questions that resemble more, as both can draw on the richer structure of lattices. In those two, the challenge is how to either "structurally clean up" the data before applying the algebraic operations, or how to turn the operations more "fault-tolerant".

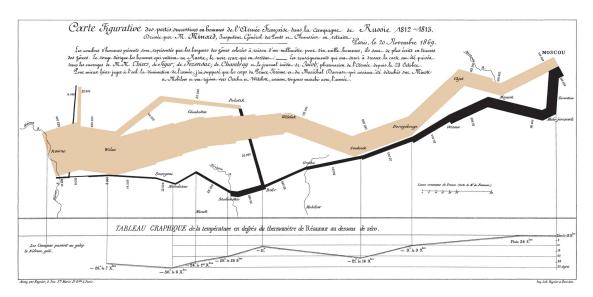


Figure 5 Minard's map of Napoleon's campaign in Russia as inspiration for a possible density-based clustering algorithm of an ordered set.

7.2.1 Clustering in Ordered Sets

There are (at least) two straightforward approaches for clustering algorithms for ordered sets, the first being the adaptation of classical clustering techniques to ordered sets, and the second being the direct exploitation of order-theoretic constructs. For the adaption approach, we suggest to start with exploring modifications of standard clustering algorithms that rely on the comparability of pairs of objects, as opposed to numerical distances. These include for instance k-modes [105] and hierarchical agglomerative clustering [212] (where "hierarchical" does refer to the set inclusion hierarchy of the clusters). We also consider it extremely promising to develop ordinal versions of modularity-based clustering of networks [22], and – inspired by Ch. Minard's seminal map of Napoleon's campaign in Russia (Figure 5, [140]) – ordinal versions of density-based clustering [59].

In the second approach, one might investigate how to exploit order-theoretic concepts (such as common upper and lower bounds, order filters and ideals, and intervals) for new kinds of ordinal clustering. Particularly promising is to consider the new nodes in a Dedekind-MacNeille completion as cluster nuclei. This is due to the observation that, in the **ancestor_of** hierarchy of a human genealogy, a couple with two or more children always generates a new node in the Dedekind-MacNeille completion, which may be interpreted as a node representing the family.

7.2.2 Clustering and Factorizations in Lattices

If an ordered set is a lattice (including the case when we enforce this by the Dedekind-MacNeille completion), one can benefit from more structure. As stated above, in theory congruence relations and tolerance relations already provide clusterings that are even compatible with the lattice operations. However, they may not always be applicable. Again, there are (at least) two promising approaches to remediate this: The first is to assess the proportion of a lattice that has to be modified such that a congruence relation or a tolerance relation with pre-defined requirements exists; and derive methods to efficiently identify such modifications. The second is to weaken the notions of congruence relations and tolerance relations, so that they may respect only most of the operations. The questions here are again how to define them, how to compute (or approximate) them efficiently, and how to interpret the results. This may result in a new notion of "operator-

aware modularity". An expected problem case in modularity based graph clustering is a cluster that is a fully connected subgraph. In lattices, this corresponds to (convex) Boolean sublattices. This correspondence could give rise to a theory and algorithms for their compression in a cluster.

Another line of research one may draw from is employing other representation modes of the lattice: One could for instance study if there are meaningful concepts for clustering sets of implications, and how they translate back again to the clustering problem for lattices. In particular in the case of a large dataset one could also resort to probabilistic methods to derive the propositional logic and dependencies [181]. One might also employ results from preliminary work on probably approximately correct learning [24, 25] of those dependencies.

7.2.3 Decompositions of lattices

Concerning the complexity reduction of lattices by decompositions, one can follow the same research methodology as for clusterings and factorizations: There exists a large body of algebraic decomposition constructions for lattices, such as direct and sub-direct decompositions, or tensor decompositions [79], but these do not admit for minor structural disturbances. Hence, a natural task would be to adapt them to more robust versions, either by suitable preprocessing or by modifying the decomposition mechanism directly.

A promising observation with respect to preprocessing is that many real-world lattices are "to a large extent" distributive at their top [177], and that distributive lattices allow for particularly simple decompositions (e.g., subdirect decomposition into linear orders). Hence, the question for how to "repair" a lattice s.t. it becomes distributive is meaningful. In the case of "almost distributive" lattices, one could resort to the study of how (potentially non-distributive) ordered sets can be freely completed (in the sense of universal algebra) to lattices [190]. A potential means to identify the non-distributive part of a lattice might then be to establish a way to reverse this construction and develop efficient computation methods.

Additionally, one might want to explore – as described above – how other equivalent representations (in particular sets of implications) can be decomposed, and how this can be translated into meaningful decompositions of lattices.

8 Ordinal Factor Analysis

Factor analysis (in its traditional sense) is a method of multi-variate statistics to reduce data consisting of observations in different manifest variables to a lower-dimensional space spanned by "latent variables" with as little loss as possible. As the variables are considered to range in the real numbers, we will call this task numerical factor analysis in the sequel. A variety of approaches exists, the most prominent being principal component analysis. However, all these approaches require the data to be at least on interval scale level. Ordinal factor analysis, on the other hand, has been developed with the same intention as its numerical counterpart, but focussing on data of ordinal scale type.

Based on this foundation, we are looking forward to a *comprehensive theory and algorithms* for an ordinal version of factor analysis, that will keep track of all operations and thus avoid the problem of numerical factor analysis where distances may be distorted. For ordinal data the factors are known to be ordered sets. One might continue research in this direction by focussing on linear ordered sets as factors, as they provide an intuitive representation, in particular when their number is low. This constitutes an instance of dimension reduction and is computationally hard for ordinal data. Hence, one may have to resort to approximate methods for real-world data applications. In a second step, the theoretical foundation of ordinal factor analysis might then be adapted to non-linear factors. Our ambition is to establish ordinal factor analysis as a tool for ordinal data analysis, that will provide a more intuitive representation of complex ordinal data.

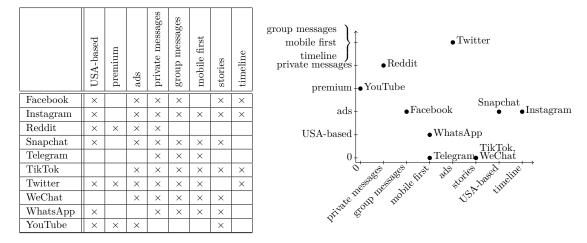


Figure 6 A dataset on social networks together with a visualization of a factorization into its two largest ordinal factors. All incidences can be deduced from the projection except for (TikTok, timeline), (WhatsApp, stories), (Facebook, timeline), (YouTube, stories), (Facebook, stories). The ordinal projection does not contain false data.

8.1 State of the Art and Open Questions

In multivariate statistics empirical observations are used to draw conclusions about underlying independent random variables. This was first developed by Spearman [170] in 1904 for evaluating intelligence tests, tracing them back to a single "general factor", and was generalized to multiple variables in 1919 by Garnett [80]. Building on this notion, nowadays two variants of factor analysis are employed. On one hand, confirmatory factor analysis [96] is used in social research to test whether the data fits to hypothesized measurement models. On the other hand, explorative factor analysis [149] is used to discover hidden structures in underlying data without the numbers of factors being known beforehand. This explorative factor analysis is closely related and often confused with principal component analysis as noted in [107], even though both techniques differ in execution as well as in their aim. In the area of recommender systems, factor analysis came to prominence through the "Netflix Prize" that was held between 2006 and 2009 where Funk [67] showed how to use factor analysis techniques to predict users preferences of movies. This approach was improved and refined in [123, 98]. A similar but not equal way of doing factor analysis is called non-negative matrix factorization and applied in cluster analysis of documents [171] or in astronomy [15]. One problem of these approaches is that the resulting factors are difficult (or even impossible) to interpret for a data analyst.

To avoid this problem, steps towards an explainable factor analysis have been made in the field of FCA. This goes back to Kerpt and Snásel [112, 113] and was further developed by Belohlavek and Vychodil [12, 13] under the name *Boolean factor analysis*. They compute a factorization of a Boolean matrix into two binary matrices, such that their Boolean matrix product results in the original matrix.

For such a factorization there is a corresponding factorizing family in the lattice, consisting of k formal concepts, called Boolean factors. Even though it is known to be \mathcal{NP} -complete to decide whether a Boolean matrix has a factorization into k Boolean factors, they introduce a set of algorithms for Boolean factor analysis which optimize k. Building on these notions, *ordinal factor analysis* was introduced by Ganter and Glodeanu in [71, 70], as a technique to meaningfully group Boolean factors, together with a visualization technique to depict ordinal factors in a biplot.

▶ **Example 7.** Figure 6 shows, on the left, a small dataset about prominent social networking platforms. On the right, a factorization of the dataset into its two largest ordinal data is shown. In the plot, an object contains all attributes that are listed below on the horizontal axis and all attributes that are listed left of it on the horizontal axis. The plot contains all information of the original dataset, except the four incidences listed in the caption of the figure. In order to cover these relations as well, the third factor would have to be included as well, leading to a 3D visualisation.

The authors restrict their work to ordinal factors that form a chain in the concept lattice. This *linear factor analysis* is closely related to dimension theory of ordered sets [206], as in both cases a minimal number of chains is computed that covers an ordered set. In [82] the authors show that linear factor analysis can be applied to data sets from real-world applications and in [81] the theory of linear factor is transferred to triadic concept analysis. A modification of the visualization technique for the linear factors and the idea to extract two maximal linear factors that cover most of the data is proposed in [70]. A way that tries to cover most of a concept lattice with two chains and can thus be seen as a variant of linear factor analysis is described in our work [47].

While linear factor analysis seems to be a promising tool for a data analyst, it is still not yet fleshed out enough to be applied in practice. This is mainly due to the fact of being computationally expensive, and thus algorithms have not yet been developed (except for the Boolean case). Furthermore, the theoretical backbone of the theory is not yet broad enough to provide a data analyst with a robust toolkit of methods and variants.

8.2 Promising Research Questions

The overarching research question here is how to develop ordinal factor analysis to a mature (visual) data analysis tool, that can be applied using little manual effort. In particular this calls for enriching the theoretical backbone and the development of algorithms to make ordinal factor analysis applicable in practice. The algorithms should be able to extract structure that is already encapsulated in data and to deal with incremental changes in the data. Finally, a variant that also deals with non-linear factors (e.g., trees) might be of interest.

8.2.1 Computing Linear Factorizations in Two and More Dimensions

A first step in this direction is our greedy algorithm [52] that iteratively extracts the largest remaining ordinal factor from a formal context. Furthermore, our second work [53] builds on an idea proposed by Ganter [70] to cover a maximum subset of the dataset with a small set of linear factors. This proposed method allows for the incorporation of heuristics to approximate a factorization into two dimensions, and examines the computational boundaries.

We assume two-dimensional linear factor analysis to perform better on smaller datasets, as large datasets tend to comprise more complex dependencies. Thus, a second step may leap significantly beyond the state of the art by exploring how complex data can be represented with factor analysis in higher dimensions. As the two-dimensional case is already known to be of high computational complexity, we do expect this to be an even more complicated problem, which will definitely require the development of heuristics.

As the aim of these factorizations is to improve the understanding of correlations in data sets for a data analyst, novel visualization techniques are required. The practicability of resulting breakthrough approaches and algorithms of ordinal factor analysis in higher dimensions will have to be demonstrated. As it hardly seems possible to visualize factors far beyond dimension three on a static two-dimensional document, interactive navigation methods will have to be developed as well.

8.2.2 Background Orders and Evolving Data

In real-world settings, parts of the data may already be (linearly) ordered because of additional background information. Such data often appears in practice, as especially data containing numerical values (including time) are already ordered. An ordinal factor analysis should respect such *background orders* as it is unnatural for an analyst to have concepts such as time divided and distributed over different factors. One could follow (at least) two possible routes to avoid this problem. The first one is to disregard the ordered part of the data from for the computation of an ordinal factorization. Then in a second step this order would be added back as an additional factor. While this guarantees the purity of this additional factor, it might not depict relationships to a sufficient degree. Thus, one might want to follow a more ambitious path by developing algorithms which do not split such an ordinal structure into different factors but are completely contained in a single factor. This factor is however still allowed to contain additional information and is contrary to the first approach not restricted to the background order.

Another typical requirement in practice is to update the factorization when data is evolving. Computing an ordinal factorization of a large dataset will not be (computationally) cheap because of the high computational complexity. Thus, if a small detail changes in the data it should be possible to integrate it in an existing ordinal factorization without having to repeat the whole computational effort. Furthermore, the structure of the factorization should be stable to a sufficient degree, as a human data analyst looking a second time at slightly modified data should be presented with familiar ordinal factors. This proposes two research questions, the first being if one already has to follow additional restrictions when computing the first factorization to allow a later change in the data. The second is how one can insert the changes into an existing factorization.

8.2.3 Non-Linear Factor Analysis

As often complex data is not correlated linearly, further research on non-linear factorizations seems to be promising. The definition of ordinal factors stemming from a linear order of Boolean factors allows the definition to be extended to factorization into arbitrarily ordered sets. However, these non-linear factors have to be further restricted as otherwise every lattice is its own, trivial factorization. We propose to investigate several restrictions for non-ordinal factors that seem reasonable, such as planarity, two-dimensionality, distributivity, and trees. Another possible approach to gain a reasonable factorization is to fix the number of factors beforehand and to allow arbitrary non-ordinal factors that should approximately be of similar cardinality. For all those restrictions no preliminary work is available, making this endeavor highly challenging because of anticipated potential complexity as well as visualization problems.

9 Visualizing, Exploring and Explaining Ordinal Data

Real-world ordinal data are usually too large to be analyzed by simply browsing the set along the edges of the order relation. The typical means of presenting ordered sets to humans is via line diagrams. A surprisingly hard conceptual problem that we will discuss in this section is the specification of HCI-founded, formal optimization criteria for graph drawing, for which one can then develop *efficient layout algorithms* [203, 47]. A second conceptually demanding problem – still addressed manually today – is the automatic break-down of large ordered sets into smaller, visualizable parts together with suitable means for their *interactive exploration* [182]. We will motivate the development of new interaction paradigms based on the decomposition and factorization methods discussed in Sections 7 and 8. These are particularly challenging as they also have to transmit the information about the effects caused by approximations.

9.1 State of the Art and Open Questions

Graph Drawing has a long research history, represented a.o. by the series of Graph Drawing Symposia¹⁵ since 1992. Partial orders and lattices are specific types of graphs, and their drawings can benefit from cycle freeness. Surprisingly, though, there exist only a few criteria for formalizing readability (e.g., maximizing distances between nodes and lines, maximizing angles of crossing lines, minimizing the number of different edge directions, organizing nodes in layers). There is agreement in the community that further criteria are needed, and that they have to be empirically evaluated from the HCI perspective.

In order theory, a common tool for visualizing and investigating ordered set is the order diagram, that is sometimes also called line diagram or Hasse diagram. The order diagram of an ordered set (X, \leq) is a directed graph, where the edge $(a, b) \in X \times X$ exists if and only if a < b and there is no $c \in X$ such that a < c < b. In a drawing of this graph the direction of an edge is not signaled by an arrow but by the y-coordinates, with the dot marking the position of b being above the dot marking the position of a. The relative horizontal positions are thus used as visual variable [16]. While readable order diagrams can be drawn from an experienced expert by hand, this is not viable in practice, as it is a time-consuming task. The automatic generation of well interpretable graph drawings is a surprisingly hard task. The problem starts with the fact that the readability criteria described above are partly conflicting and that their relative importance varies in different settings. In the standard work about graph drawing [203, Sect. 3.1] several divide-and-conquer algorithms for drawing trees, which are closely related to order diagrams, are described.

In the case of a planar graph (which can be checked in linear time [103]), a drawing only consisting of straight lines without bends or curves always exists [148, Sect. 4.2 & 4.3] and should be preferred. For an order diagram with a unique maximum and minimum it can be checked whether it is planar and in this case an upward planar drawing can be computed in linear time [203, Sect. 6]. However, while lattices by definition contain a maximum and minimum element, they are usually not planar if they are derived from real world data [4]. A drawing algorithm [203, Sect. 3.2] for "serial parallel graphs", a special family of planar, acyclic graphs straight-line drawings can be equipped with a modification to produce symmetries based on the automorphism group [102]. Sugiyama et al. [197] published in 1981 an algorithm framework to compute upwards layered drawings of acyclic graphs. Force directed algorithms were introduced in [54] and further refined in [66]. In [65, 51] this idea is applied to order diagrams. Our work [194] proposes a geometrical heuristic to support a human to draw order diagrams manually. By trying to emphasize the intrinsic structure of ordered sets our work [47] proposes a drawing algorithm based on the computation of two linear extensions. The algorithm is included as a tool in the software-suite conexp-clj¹⁶ under the name DimDraw [49].

▶ **Example 8.** The example in Figure 7 shows that it is possible to achieve competitive order diagrams using this approach. The concept lattice was created for structuring an educational film about living beings and water [200, 79]. The left diagram was hand-drawn by an FCA expert. The diagram in the middle was generated with the algorithm of Sugiyama et al. [197] while the right one was drawn with the DimDraw-algorithm [47] which leverages the dimensional structure of the order. Experiments have shown that the DimDraw results are often rather close to hand-drawn diagrams.

6:27

¹⁵https://dblp.uni-trier.de/db/conf/gd/

¹⁶ https://github.com/tomhanika/conexp-clj

6:28 Towards Ordinal Data Science

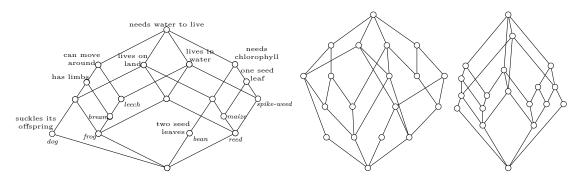


Figure 7 Three order diagrams for the visualization of the same ordered set: hand-drawn by an expert, by the Sugiyama algorithm and by the DimDraw algorithm.

As ordered sets and lattices even of modest size (from, say, 30-50 elements upwards) are hard to visualize such that they are still beneficial for a data analyst, further means of interaction are necessary for larger ordered sets. One way is to compute a diagram in 3D, which can be done using [65] or a modification of [47], and then allow to rotate it on the screen, but this does not increase the size of presentable nodes significantly. Another way is to display only a single node together with its upper and lower neighbors at a time, and to shift the focus when the user selects one of these neighbors. This has for instance been used by Carpineto and Romano for Information Retrieval [33]. A third approach follows the divide & conquer paradigm. Its first realization for concept lattices decomposed the lattice in two or more factors of a sub-direct semi-product, and combined their smaller diagrams in a nested line diagram [221, 188, 182]. This is also the archetype for the kind of algebraic constructions that we suggested studying in Section 7. This approach became part of a model for *Conceptual Information Systems* [8] and has been implemented in the ToscanaJ tool suite¹⁷ [9, 10, 40]. ToscanaJ also allows, by means of conceptual scaling, to visualize concept lattices derived from data types of any kind, and to navigate intuitively through the data by switching between the conceptual scales and zooming into concepts. However, the initialization of the system requires conceptual and manual work, in particular a meaningful selection of the factors and a manual layout of the order diagram for each factor. The model of Conceptual Information Systems has been generalized in various ways: as ordinal version of Online-Analytical Processing [191, 183, 56, 100, 193], for accessing databases and conceptual graphs [125] and knowledge bases [64]. A good survey over navigation paradigms and layout algorithms for concept lattices with hybrid data is provided by [57].

9.2 Promising Research Questions

The main objective of the visualization of ordinal structures is to support human analysts in analyzing and exploring large ordinal data. In particular, we emphasize on the urgent need of new layout algorithms that allow for a completely automatic, well interpretable visualization of medium-sized ordered sets and lattices. These might include new interaction paradigms based on the decomposition and factorization methods as discussed in Sections 7 and 8. This approach is particularly challenging as it also has to transmit the information about which aspects of the data are not represented in the visualization, due to an approximation.

These (static) visualizations may pave the way for (dynamic) interactive browsing and exploring of complete datasets. Again, the opportunities for new browsing paradigms may benefit from advances in the field of algebraic decompositions and ordinal factor analysis as discussed in Sections 7 and 8.

¹⁷ http://toscanaj.sourceforge.net/

9.2.1 Visualization of Ordered Sets and Lattices

Ordinal factor analysis, as discussed in Section 8, will provide a set of chains that cover all points represented in a dataset. Using ideas from [47], this could be applied for a novel drawing algorithm, which, however, would not reduce the – typically too large – order dimension. One could hence exploit means for order dimension reduction (towards dimension 2) as discussed in Section 7, or employ force directed drawing algorithms in three or more dimensions. The challenge is to incorporate the order constraints such that they are all reflected properly in the resulting diagram. Another potential approach might be machine-learning-based dimension reduction. This is challenging since the requirement that the *y*-coordinates of data points have to obey the order relation is not encoded in non-linear methods such as t-SNE [134], in contrast to linear methods such as PCA [227]. This is also true for deep learning procedures, which nevertheless might be worth being adapted to ordinal data.

For some special families of lattices (such as distributive lattices, two-dimensional lattices or Boolean lattices), methods exist for generating readable drawings. Thus, in a third line of research one could explore modifications (removing or adding vertices and/or edges) such that the lattice "partly fulfills" these properties – which first needs to be defined. In a second step, the drawing of this structurally simpler part could then be used to derive a readable drawing for the whole lattice. A supporting study of this endeavor would be to empirically investigate how large the two-dimensional, distributive or Boolean parts of real-world lattices have to be to justify the applicability of such an algorithm.

9.2.2 Browsing and Exploring Large Orders

There may be many ways for future approaches to browsing and exploring ordered sets. We assume that some of them will be based on the outcomes of research along the lines described in Sections 7 and 8. From today's perspective we anticipate these approaches as follows. The *slice* \mathscr{C} *dice* paradigm of On-Line Analytical Processing (OLAP) provides a good scheme for analyzing multidimensional numerical data, and first applications to ordinal data exist [191, 183, 56, 100, 193]. A similar approach might be followed for the clusterings, decompositions and factorizations developed as described in Sections 7 and 8. To date it is not obvious, though, whether the same visualization and interaction paradigms can also be applied to the new constructions. What is definitely missing in the preliminary work is the handling of disturbances. In a second step, we might therefore address the visualization of and interaction with algebraic clusterings, decompositions and factorizations that have been approximated. As there does not exist any preliminary work on this, this part of the research will be of non-incremental nature. Our assumption is that the analyst will be informed about the existence of any non-covered and eventually wrongly assigned data, at least on request. Following the paradigm of Explainable AI, the analyst should be enabled to drill down to the original data if necessary, and to obtain explanations in case the construction could not fully respect the data.

10 Challenges and Long-Term Perspective

While a wide range of methods for Data Analysis and Knowledge Discovery has been developed for data that are on ratio level (i. e., can be adequately modeled with real numbers), there are fewer methods for data on ordinal level. In this paper, we call for a joint activity to bring these methods together and to extend them with new ones, and to establish thus the new research field of Ordinal Data Analysis. The amount of ordinal data in the real world is large, and so is the requirement for adequate methods and algorithms. There are many research opportunities in this

6:30 Towards Ordinal Data Science

field, and we can provide only some teasers in this paper. We are looking forward to all kinds of contributions to the new field of Ordinal Data Science, both within and outside the five subfields discussed in the previous sections.

We see Ordinal Data Science in a long tradition of Conceptual Knowledge Processing, where we aim at a highly human-centered process of asking, exploring, analyzing, interpreting, and learning about the data in interaction with the underlying database [101, 184]. In this line, we follow R. S. Brachman et al. [26], who introduced the notion of Data Archaeology for knowledge discovery tasks in which a precise specification of the discovery strategy, the crucial questions, and the basic goals of the task have to be elaborated during such an unpredictable interactive exploration of the data.

We expect that all future contributions to this new field will share a range of challenges that are specific for research on the intersection of computer science and mathematics, which both come with different philosophical foundations, research aims and working paradigms. One challenge will be to develop sound theoretical theories, another one to find computationally feasible solutions and/or approximations. Last but not least the algorithms and methods have to be made accessible to their users, e.g., by suitable visualizations, interaction paradigms and human-computer interfaces.

We will conclude this paper by illustrating, for the five fields that we discussed in the previous sections, the challenges and long-term perspectives that we expect to face:

We expect to meet the largest theoretical challenges in the endeavor of developing an ordinal version of the Representational Theory of Measurement. The replacement of the numerical relational structure in RTM by ordinal relational structures (and in the long run by other, non-ordinal relational structures such as symmetry groups) implies the loss of focus on additivity. A generalization of RTM thus requires a much more flexible framework that allows for a large variety of (numerical and) non-numerical scales, and it is not obvious a priori how such a framework might look like. A major criticism to RTM is that in practice the pre-conditions for applicability of measures on interval or ratio scale are hard to verify, and that nominal measurements (whose pre-conditions are easier to verify) do not provide enough insight. When we succeed, we expect a new ordinal theory of measurement to fill the gap between these extremes and to become a significant contribution to the still intense discussion about the applicability of RTM for psychological studies ([139, 204, 205], a good survey is [201]).

Generalized ordinal scales are just a special case of allowing arbitrary algebraic structures as ranges of measurements. In the long run we have the ambition to establish a "Grand Unified Theory of Measurement", which is likely to be of high interest to (the philosophical foundations of) physics (where traditionally the adaptation of measurement to psychological phenomena has been eyed with some suspicion), as it might allow for describing the study of, e.g., the symmetries of elementary particles in the same way as their numerical properties.

In Section 6, we proposed to develop a joint theory for the interplay of metric functions and ordinal relations, based on the archetype of metric measure spaces. While, in the latter, both metrics and measures operate on the set of real numbers, our proposed approach tries to marry two structurally different domains. We might thus face the task of developing a new paradigm of interaction which has no precedents that can be followed.

In Sections 7 to 9, we argued for the exploitation of algebraic structures for data analysis tasks, which unfortunately go along with large size and computationally hard problems. E.g., the size of concept lattices grows in the worst case exponentially in the size of the data, which in the Information Age are already large by themselves; and determining the order dimension of an

ordered set is \mathcal{NP} -hard. In Data Science, the typical approach is to use projections, decompositions, approximations, and heuristics. However, these kinds of approximations and heuristics are no subject of study in classical universal algebra, and still have to be established. In particular, we suggest to make use of algebraic composition and decomposition techniques for a divide & conquer approach for browsing and exploring the data. The transformation of the (approximated) algebraic constructions to human-computer interaction paradigms that support users without knowledge of universal algebra and order theory is challenging, as there is little precedence. Nonetheless, if this works out, it might provide a whole new family of analysis techniques for ordinal data.

— References -

- 1 Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. Advances in Mathematics, 228(6):3026-3126, 2011. doi: 10.1016/j.aim.2011.08.003.
- 2 Ravi P Agarwal, MA El-Gebeily, and Donal O'Regan. Generalized contractions in partially ordered metric spaces. *Applicable Analysis*, 87(1):109–116, 2008. doi:10.1080/ 00036810701556151.
- 3 Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207– 216, jun 1993. doi:10.1145/170035.170072.
- 4 Alexandre Albano and Bogdan Chornomaz. Why concept lattices are large: extremal theory for generators, concepts, and vc-dimension. *International Journal of General Systems*, 46(5):440–457, 2017. doi:10.1080/03081079.2017.1354798.
- 5 Simon Andrews. Making use of empty intersections to improve the performance of cbo-type algorithms. In *International Conference on Formal Concept Analysis*, pages 56–71. Springer, 2017. doi:10.1007/978-3-319-59271-8_4.
- 6 Franz Baader. Computing a minimal representation of the subsumption lattice of all conjunctions of concepts defined in a terminology. In Proceedings of the International Symposium on Knowledge Retrieval, Use, and Storage for Efficiency, KRUSE 95, pages 168–178, Santa Cruz, USA, 1995.
- 7 Franz Baader, Bernhard Ganter, Baris Sertkaya, and Ulrike Sattler. Completing description logic knowledge bases using formal concept analysis. In Manuela M. Veloso, editor, *IJCAI*, pages 230–235, 2007. doi:10.25368/2022.155.
- 8 K. Becker, G. Stumme, R. Wille, U. Wille, and M. Zickwolff. Conceptual information systems discussed through an it-security tool. In R. Dieng and O. Corby, editors, *Knowledge Engineering and Knowledge Management. Methods, Models, and Tools.*, volume 1937 of *LNAI*, pages 352– 365, Heidelberg, 2000. Springer. doi:10.1007/ 3-540-39967-4_27.
- 9 P. Becker, J. Hereth, and G. Stumme. ToscanaJ: An open source tool for qualitative data analysis. In V. Duquenne, B. Ganter, M. Liquiere, E. M. Nguifo, and G. Stumme, editors, Advances in Formal Concept Analysis for Knowledge Discovery in Databases., pages 1–2, Lyon, France, 7 23 2002.
- 10 Peter Becker and Joachim Hereth Correia. The toscanaj suite for implementing conceptual information systems. In *Formal Concept Analy*-

sis, pages 324–348. Springer, 2005. doi:10.1007/11528784_17.

- 11 Radim Belohlavek. Fuzzy relational systems: foundations and principles, volume 20. Springer Science & Business Media, 2012. doi:10.1007/ 978-1-4615-0633-1.
- 12 Radim Belohlávek and Vilém Vychodil. Formal concepts as optimal factors in boolean factor analysis: Implications and experiments. In Peter W. Eklund, Jean Diatta, and Michel Liquiere, editors, Proceedings of the Fifth International Conference on Concept Lattices and Their Applications, CLA 2007, Montpellier, France, October 24-26, 2007, volume 331 of CEUR Workshop Proceedings. CEUR-WS.org, 2007. URL: https: //ceur-ws.org/Vol-331/Belohlavek1.pdf.
- 13 Radim Belohlavek and Vilem Vychodil. Discovery of optimal factors in binary data via a novel method of matrix decomposition. *Journal of Computer and System Sciences*, 76(1):3–20, 2010. doi:10.1016/J.JCSS.2009.05.002.
- 14 R. Bendaoud, A. Napoli, and Y. Toussaint. Formal concept analysis: A unified framework for building and refining ontologies. In *EKAW*, 2008. doi:10.1007/978-3-540-87696-0_16.
- 15 Olivier Berne, C. Joblin, Y. Deville, J. D. Smith, M. Rapacioli, J. P. Bernard, J. Thomas, W. Reach, and A. Abergel. Analysis of the emission of very small dust particles from spitzer spectro-imagery data using blind signal separation methods. Astronomy & Astrophysics, 469(2):575–586, 2007. doi:10.1051/0004-6361:20066282.
- 16 J. Bertin and M. Barbut. Sémiologie graphique: les diagrammes, les réseaux, les cartes. Gauthier Villars, 1967.
- 17 T Gnana Bhaskar and V Lakshmikantham. Fixed point theorems in partially ordered metric spaces and applications. Nonlinear Analysis: Theory, Methods & Applications, 65(7):1379-1393, 2006. doi:10.1016/j.na.2005.10.017.
- 18 Robi Bhattacharjee and Sanjoy Dasgupta. What relations are reliably embeddable in euclidean space? In Aryeh Kontorovich and Gergely Neu, editors, Proceedings of the 31st International Conference on Algorithmic Learning Theory, volume 117 of Proceedings of Machine Learning Research, pages 174–195, San Diego, California, USA, 08 February-11 February 2020. PMLR. URL: http://proceedings.mlr.press/ v117/bhattacharjee20a.html.
- 19 G. Birkhoff. On the structure of abstract algebras. Proceedings of the Cambridge Philosoph-

ical Society, 31(4):433–454, 1935. doi:10.1017/ s0305004100013463.

- 20 Garrett Birkhoff. Universal algebra. In Comptes Rendus du Premier Congrès Canadien de Mathématiques, pages 310–326, Toronto, 1946. University of Toronto Press.
- 21 Garrett Birkhoff. Lattice theory. Rev. ed, volume 25 of Collog. Publ., Am. Math. Soc. American Mathematical Society (AMS), 1948.
- 22 Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal* of Statistical Mechanics: Theory and Experiment, 2008:10008, 2008. doi:10.1088/1742-5468/2008/ 10/p10008.
- 23 George Boole. Investigation of The Laws of Thought On Which Are Founded the Mathematical Theories of Logic and Probabilities. Walton and Maberly, 1853. Also available from Dover, New York 1958, ISBN 0-486-60028-9.
- 24 Daniel Borchmann, Tom Hanika, and Sergei Obiedkov. On the usability of probably approximately correct implication bases. In Karell Bertet, Daniel Borchmann, Peggy Cellier, and Sébastien Ferré, editors, *ICFCA*, volume 10308 of *LNCS*, pages 72–88. Springer, 2017. doi: 10.1007/978-3-319-59271-8_5.
- 25 Daniel Borchmann, Tom Hanika, and Sergei Obiedkov. Probably approximately correct learning of horn envelopes from queries. *Discrete Applied Mathematics*, 273:30–42, 2020. Advances in Formal Concept Analysis: Traces of CLA 2016. doi:10.1016/J.DAM.2019.02.036.
- 26 Ronald J. Brachman, Peter G. Selfridge, Loren G. Terveen, Boris Altman, Alex Borgida, Fern Halper, Thomas Kirk, Alan Lazar, Deborah L. McGuinness, and Lori Alperin Resnick. Integrated support for data archaeology. In *Proceedings of the 2nd International Conference on Knowledge Discovery in Databases*, AAAIWS'93, pages 197–211. AAAI Press, 1993. doi:10.1142/S0218215793000083.
- 27 L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984.
- 28 Laura F. Bringmann and Markus I. Eronen. Heating up the measurement debate: What psychologists can learn from the history of physics. *Theory & Psychology*, 26(1):27–43, 2016. doi:10.1177/0959354315617253.
- 29 Russ Bubley and Martin Dyer. Faster random generation of linear extensions. Discrete mathematics, 201(1-3):81–88, 1999. doi:10.1016/ S0012-365X(98)00333-1.
- 30 S. Burris and H. P. Sankappanavar. A Course in Universal Algebra. Springer, New York, 1981.
- 31 Xiao Cai, Feiping Nie, and Heng Huang. Multiview k-means clustering on big data. In Twenty-Third International Joint conference on artificial intelligence. Citeseer, 2013. URL: http://ijcai. org/Abstract/13/383.
- 32 Jaime S Cardoso, Joaquim F Pinto da Costa, and Maria J Cardoso. Modelling ordinal relations with svms: An application to objective aesthetic

evaluation of breast cancer conservative treatment. *Neural Networks*, 18(5-6):808-817, 2005. doi:10.1016/J.NEUNET.2005.06.023.

- 33 Claudio Carpineto and Giovanni Romano. Concept Data Analysis: Theory and Applications. John Wiley & Sons, 2004.
- 34 N. Caspard, B. Leclerc, and B. Monjardet. Finite Ordered Sets: Concepts, Results and Uses. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2012. doi: 10.1017/cbo9781139005135.
- 35 Nicholas R. Chrisman. Rethinking levels of measurement for cartography. Cartography and Geographic Information Systems, 25(4):231–242, 1998. doi:10.1559/152304098782383043.
- 36 Constanze Ciavarella, Laura Fumanelli, Stefano Merler, Ciro Cattuto, and Marco Ajelli. School closure policies at municipality level for mitigating influenza spread: a model-based evaluation. BMC Infectious Diseases, 16(1):576, 2016. doi:10.1186/s12879-016-1918-z.
- 37 Philipp Cimiano, Andreas Hotho, Gerd Stumme, and Julien Tane. Conceptual knowledge processing with formal concept analysis and ontologies. In Peter Eklund, editor, *Concept Lattices*, volume 2961 of *LNAI*, pages 189–207, Heidelberg, 2004. Second International Conference on Formal Concept Analysis, ICFCA 2004, Springer. doi:10.1007/978-3-540-24651-0_18.
- 38 C. H. Coombs, L. C. Coombs, and J. C. Lingoes. Stochastic cumulative scales. In S. Shye, editor, *Theory construction and data analysis in the behavioral sciences*. Jossey-Bass., San Francisco, 1978.
- **39** Clyde H. Coombs. A Theory of Data. Wiley, New York, 1964.
- 40 Frithjof Dau and Baris Sertkaya. An extension of ToscanaJ for FCA-based data analysis over triple stores. In CUBIST Workshop, 2011. URL: https://ceur-ws.org/Vol-753/ paper02_cubistws11.pdf.
- 41 Thomas H. Davenport and Jeanne G. Harris. Competing on Analytics: The New Science of Winning. Harvard Business School Press, 2007. doi:10.5860/choice.44-6322.
- 42 B. A. Davey and H. A. Priestley. Introduction to Lattices and Order. Cambridge mathematical textbooks. Cambridge University Press, 2002. doi:10.1017/CB09780511809088.
- 43 Stephan Doerfel, Tom Hanika, and Gerd Stumme. Clones in graphs. In Michelangelo Ceci, Nathalie Japkowicz, Jiming Liu, George A. Papadopoulos, and Zbigniew W. Ras, editors, *ISMIS*, volume 11177 of *LNCS*, pages 56–66. Springer, 2018. doi:10.1007/978-3-030-01851-1_6.
- 44 Stephan Doerfel, Robert Jäschke, and Gerd Stumme. Publication analysis of the formal concept analysis community. In F. Domenach, D.I. Ignatov, and J. Poelmans, editors, *ICFCA 2012*, volume 7278 of *Lecture Notes in Artificial Intelligence*, pages 77–95, Berlin/Heidelberg, 2012. Springer. doi:10.1007/978-3-642-29892-9_12.
- 45 Jean-Paul Doignon and Jean-Claude Falmagne. Knowledge Spaces. Imprint: Springer, Berlin, Heidelberg, 1999. doi:10.1007/978-3-642-58625-5.

- 46 Michel Dumontier and Tobias Kuhn. Data science – methods, infrastructure, and applications. *Data Science*, 1:1–5, 2017. doi:10.3233/DS-170013.
- 47 Dominik Dürrschnabel, Tom Hanika, and Gerd Stumme. Drawing order diagrams through twodimension extension, 2019. arXiv:1906.06208.
- 48 Dominik Dürrschnabel, Tom Hanika, and Maximilian Stubbemann. FCA2VEC: Embedding techniques for formal concept analysis. In Rokia Missaoui, Léonard Kwuida, and Talel Abdessalem, editors, Complex Data Analytics with Formal Concept Analysis, pages 47–74. Springer International Publishing, 2022. doi:10.1007/978-3-030-93278-7_3.
- 49 Dominik Dürrschnabel, Tom Hanika, and Gerd Stumme. Dimdraw – a novel tool for drawing concept lattices. In Diana Cristea, Florence Le Ber, Rokia Missaoui, Léonard Kwuida, and Baris Sertkaya, editors, *ICFCA (Supplements)*, volume 2378 of *CEUR Workshop Proceedings*, pages 60– 64. CEUR-WS.org, 2019. URL: https://ceur-ws. org/Vol-2378/shortAT6.pdf.
- 50 Dominik Dürrschnabel, Maren Koyda, and Gerd Stumme. Attribute selection using contranominal scales. In Tanya Braun, Marcel Gehrke, Tom Hanika, and Nathalie Hernandez, editors, Graph-Based Representation and Reasoning - 26th International Conference on Conceptual Structures, ICCS 2021, Virtual Event, September 20-22, 2021, Proceedings, volume 12879 of Lecture Notes in Computer Science, pages 127–141. Springer, 2021. doi:10.1007/978-3-030-86982-3_10.
- 51 Dominik Dürrschnabel and Gerd Stumme. Forcedirected layout of order diagrams using dimensional reduction. In Agnès Braud, Aleksey Buzmakov, Tom Hanika, and Florence Le Ber, editors, Formal Concept Analysis - 16th International Conference, ICFCA 2021, Strasbourg, France, June 29 - July 2, 2021, Proceedings, volume 12733 of Lecture Notes in Computer Science, pages 224–240. Springer, 2021. doi:10.1007/ 978-3-030-77867-5_14.
- 52 Dominik Dürrschnabel and Gerd Stumme. Greedy discovery of ordinal factors. CoRR, abs/2302.11554, 2023. doi:10.48550/ARXIV.2302. 11554.
- 53 Dominik Dürrschnabel and Gerd Stumme. Maximal ordinal two-factorizations. In Proceedings of the 27th International Conference on Conceptual Structures, ICCS 2022, 2023. doi:10.1007/ 978-3-031-40960-8_5.
- 54 Peter Eades. A heuristic for graph drawing. Congressus numerantium, 42:149–160, 1984.
- 55 David A Edwards. The structure of superspace. In *Studies in topology*, pages 121–133. Elsevier, 1975. doi:10.1016/b978-0-12-663450-1.50017-7.
- 56 P. Eklund, B. Groh, G. Stumme, and R. Wille. Contextual-logic extension of toscana. In B. Ganter and G. W. Mineau, editors, *Conceptual Structures: Logical, Linguistic, and Computational*, volume 1867 of *LNAI*, pages 453–467, Heidelberg, 2000. Springer. doi:10.1007/10722280_31.
- 57 Peter W. Eklund and Jean Villerd. A survey of hybrid representations of concept lattices in conceptual knowledge processing. In Léonard

Kwuida and Baris Sertkaya, editors, *ICFCA*, volume 5986 of *LNCS*, pages 296–311. Springer, 2010. doi:10.1007/978-3-642-11928-6_21.

- 58 Yves Escoufier, Bernard Fichet, Edwin Diday, Ludovic Lebart, Chikio Hayashi, Noburu Ohsumi, and Yasumasa Baba, editors. *Data Science and its Applications*. Acadamic Press/Harcourt Brace, Tokyo, 1995.
- 59 Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han, and Usama M. Fayyad, editors, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, pages 226-231. AAAI Press, 1996. URL: http://www.aaai.org/Library/KDD/ 1996/kdd96-037.php.
- 60 Ernesto Estrada. The communicability distance in graphs. Linear Algebra and its Applications, 436(11):4317-4328, 2012. doi:10.1016/j.laa. 2012.01.017.
- 61 Maximilian Felde and Tom Hanika. Formal context generation using dirichlet distributions. In Dominik Endres, Mehwish Alam, and Diana Sotropa, editors, *ICCS*, volume 11530 of *LNCS*, pages 57–71. Springer, 2019. doi:10.1007/ 978-3-030-23182-8_5.
- 62 Maximilian Felde, Tom Hanika, and Gerd Stumme. Null models for formal contexts. J. Information, 11(3, 135), 2020. doi:10.3390/INF011030135.
- 63 A. Ferguson, C. S. Myers, R. J. Bartlett, H. Banister, F. C. Bartlett, W. Brown, N. R. Campbell, K. J. W. Craik, J. Drever, J. Guild, R. A. Houstoun, J. O. Irwin, G. W. C. Kaye, S. J. F. Philpott, L. F. Richardson, J. H. Shaxby, T. Smith, R. H. Thouless, and W. S. Tucker. Quantitative estimates of sensory events. Advancement of Science, 2:331–349, 1940. (Final report of a committee appointed by the British Association for the Advancement of Science in 1932 to consider the possibility of measuring intensities of human sensation.). doi:10.1038/130334a0.
- 64 Sébastien Ferré, Marianne Huchard, Mehdi Kaytoue, Sergei O Kuznetsov, and Amedeo Napoli. Formal concept analysis: from knowledge discovery to knowledge processing. A Guided Tour of Artificial Intelligence Research: Volume II: AI Algorithms, pages 411–445, 2020. doi:10.1007/ 978-3-030-06167-8_13.
- 65 Ralph Freese. Automated lattice drawing. In International Conference on Formal Concept Analysis, pages 112–127. Springer, 2004. doi:10.1007/ 978-3-540-24651-0_12.
- 66 Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. Software: Practice and experience, 21(11):1129–1164, 1991. doi:10.1002/SPE.4380211102.
- 67 Simon Funk. Netflix update: Try this at home, 2006. [Online; accessed 11-August-2020].
- 68 B. Ganter, G. Stumme, and R. Wille, editors. Formal Concept Analysis: Foundations and Applications, volume 3626 of LNAI, Heidelberg, 2005. Springer. doi:10.1007/978-3-540-31881-1.
- **69** Bernhard Ganter. Composition and decomposition in formal concept analysis. In Hans-Herrmann

Bock, editor, *Classification and related methods of data analysis*, pages 561–566, Amsterdam, 1988. North–Holland.

- 70 Bernhard Ganter. Diskrete mathematik: Geordnete mengen. Springer, 2013. doi:10.1007/ 978-3-642-37500-2.
- 71 Bernhard Ganter and Cynthia Vera Glodeanu. Ordinal factor analysis. In International Conference on Formal Concept Analysis, pages 128–139. Springer, 2012. doi:10.1007/978-3-642-29892-9_15.
- 72 Bernhard Ganter and Cynthia Vera Glodeanu. Factors and skills. In Cynthia Vera Glodeanu, Mehdi Kaytoue, and Christian Sacarea, editors, Formal Concept Analysis 12th International Conference, ICFCA 2014, Cluj-Napoca, Romania, June 10-13, 2014. Proceedings, volume 8478 of Lecture Notes in Computer Science, pages 173–187. Springer, 2014. doi:10.1007/978-3-319-07248-7_13.
- 73 Bernhard Ganter and Sergei O. Kuznetsov. Stepwise construction of the dedekind-macneille completion. In Marie-Laure Mugnier and Michel Chein, editors, *Conceptual Structures: Theory, Tools and Applications*, pages 295–302, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg. doi: 10.1007/BFB0054922.
- 74 Bernhard Ganter and Sergei A. Obiedkov. Conceptual Exploration. Springer, 2016. doi:10.1007/ 978-3-662-49291-8.
- 75 Bernhard Ganter, Jürgen Stahl, and Rudolf Wille. Conceptual measurement and many-valued contexts. In W. Gaul and Martin Schader, editors, *Classification as a tool of research*, pages 169–176, Amsterdam, 1986. North–Holland.
- 76 Bernhard Ganter and Gerd Stumme. Creation and merging of ontology top-levels. In Aldo de Moor, Wilfried Lex, and Bernhard Ganter, editors, Conceptual Structures for Knowledge Creation and Communication., volume 2746 of LNAI, pages 131–145, Heidelberg, 2003. Springer. doi: 10.1007/978-3-540-45091-7_9.
- 77 Bernhard Ganter and Rudolf Wille. Conceptual scaling. In Frank Roberts, editor, Applications of combinatorics and graph theory to the biological and social sciences, pages 139–167. Springer–Verlag, New York, 1989. doi:10.1007/978-1-4684-6381-1_6.
- 78 Bernhard Ganter and Rudolf Wille. Formale Begriffsanalyse: Mathematische Grundlagen. Springer, Heidelberg, 1996. doi:10.1007/ 978-3-642-61450-7.
- 79 Bernhard Ganter and Rudolf Wille. Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg, 1999. doi:10.1007/ 978-3-642-59830-2.
- 80 J. C. Maxwell Garnett. General ability, cleverness and purpose. British Journal of Psychology, 9(3):345, 1919.
- 81 Cynthia Vera Glodeanu. Tri-ordinal factor analysis. In International Conference on Formal Concept Analysis, pages 125–140. Springer, 2013. doi:10.1007/978-3-642-38317-5_8.
- 82 Cynthia Vera Glodeanu and Bernhard Ganter. Applications of ordinal factor analysis. In International Conference on Formal Concept Analy-

sis, pages 109-124. Springer, 2013. doi:10.1007/978-3-642-38317-5_7.

- 83 Leo A. Goodman and William H. Kruskal. Measures of Association for Cross Classifications. Springer Verlag, New York, 1979. doi:10.1007/ 978-1-4612-9995-0.
- 84 S. Goto, H. Bono, H. Ogata, W. Fujibuchi, T. Nishioka, K. Sato, and M. Kanehisa. Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput*, pages 175– 186, 1997.
- 85 Olivier Gouvert, Thomas Oberlin, and Cédric Févotte. Ordinal non-negative matrix factorization for recommendation. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, volume 119 of Proceedings of Machine Learning Research, pages 3680-3689. PMLR, 2020. URL: http://proceedings.mlr.press/ v119/gouvert20a.html.
- 86 Mikhael Gromov. Groups of polynomial growth and expanding maps. Publications Mathématiques de l'Institut des Hautes Études Scientifiques, 53(1):53-78, 1981. doi:10.1007/bf02698687.
- 87 George Grätzer. Universal Algebra. Springer, 1968. doi:10.1007/978-0-387-77487-9.
- 88 George Grätzer. General Lattice Theory. Birkhäuser Verlag, 2. edition, 1998. doi:10.1016/ s0079-8169(08)x6085-4.
- 89 L. Guttman. A basis for scaling qualitative data. American Sociological Review, 9:139–150, 1944. doi:10.2307/2086306.
- 91 Tom Hanika and Johannes Hirth. Conexp-clj a research tool for fca. In Diana Cristea, Florence Le Ber, Rokia Missaoui, Léonard Kwuida, and Baris Sertkaya, editors, *ICFCA (Supplements)*, volume 2378 of *CEUR Workshop Proceedings*, pages 70–75. CEUR-WS.org, 2019. URL: https://ceur-ws.org/Vol-2378/shortAT8.pdf.
- 92 Tom Hanika and Johannes Hirth. Quantifying the conceptual error in dimensionality reduction. In Tanya Braun, Marcel Gehrke, Tom Hanika, and Nathalie Hernandez, editors, Graph-Based Representation and Reasoning 26th International Conference on Conceptual Structures, ICCS 2021, Virtual Event, September 20-22, 2021, Proceedings, volume 12879 of Lecture Notes in Computer Science, pages 105–118. Springer, 2021. doi:10.1007/978-3-030-86982-3_8.
- 93 Tom Hanika and Johannes Hirth. Knowledge cores in large formal contexts. Ann. Math. Artif. Intell., 90(6):537–567, 2022. doi:10.1007/ S10472-022-09790-6.
- 94 Tom Hanika, Maren Koyda, and Gerd Stumme. Relevant attributes in formal contexts. In Dominik Endres, Mehwish Alam, and Diana Sotropa, editors, *ICCS*, volume 11530 of *LNCS*, pages 102–116. Springer, 2019. doi:10.1007/ 978-3-030-23182-8_8.
- 95 Tom Hanika, Friedrich Martin Schneider, and Gerd Stumme. Intrinsic dimension of geometric data sets. *Tohoku Mathematical Journal*, 74(1):23–52, 2022. doi:10.2748/tmj.20201015a.

- **96** D. Harrington. *Confirmatory Factor Analysis*. Oxford scholarship online: Social Work module. Oxford University Press, USA, 2009.
- 97 Egbert Harzheim. Ordered Sets. Springer, 2005. doi:10.1007/b104891.
- 98 Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017. doi:10.1145/3038912.3052569.
- 99 Rajneesh Hegde and Kamal Jain. The hardness of approximating poset dimension. *Electronic Notes* in Discrete Mathematics, 29:435–443, 2007. European Conference on Combinatorics, Graph Theory and Applications. doi:10.1016/J.ENDM.2007.07. 084.
- 100 J. Hereth and G. Stumme. Reverse pivoting in conceptual information systems. In H. Delugach and G. Stumme, editors, *Conceptual Structures: Broadening the Base.*, volume 2120 of *LNAI*, pages 202–215, Heidelberg, 2001. Springer. doi:10.1007/3-540-44583-8_15.
- 101 Joachim Hereth, Gerd Stumme, Rudolf Wille, and Uta Wille. Conceptual knowledge discovery – a human-centered approach. Journal of Applied Artificial Intelligence (AAI), 17(3):281–301, 2003. doi:10.1080/713827122.
- 102 Seok-Hee Hong, Peter Eades, and Sang-Ho Lee. Drawing series parallel digraphs symmetrically. *Computational Geometry*, 17(3-4):165–188, 2000. doi:10.1016/S0925-7721(00)00020-1.
- 103 John Hopcroft and Robert Tarjan. Efficient planarity testing. Journal of the ACM (JACM), 21(4):549–568, 1974. doi:10.1145/ 321850.321852.
- 104 Andreas Hotho, Steffen Staab, and Gerd Stumme. Explaining text clustering results using semantic structures. In Nada Lavrač, Dragan Gamberger, and Hendrik BlockeelLjupco Todorovski, editors, Knowledge Discovery in Databases: PKDD 2003, 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, volume 2838 of LNAI, pages 217–228, Heidelberg, 2003. Springer. doi:10.1007/978-3-540-39804-2_21.
- 105 Zhexue Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min. Knowl. Discov., 2(3):283–304, 1998. doi:10.1023/A:1009769707641.
- 106 Eyke Hüllermeier, Johannes Fürnkranz, Weiwei Cheng, and Klaus Brinker. Label ranking by learning pairwise preferences. Artificial Intelligence, 172(16-17):1897–1916, 2008. doi:10.1016/ J.ARTINT.2008.08.002.
- 107 IT Joliffe and BJT Morgan. Principal component analysis and exploratory factor analysis. *Statisti*cal methods in medical research, 1(1):69–95, 1992. doi:10.1177/096228029200100105.
- 108 Toshihiro Kamishima and Jun Fujiki. Clustering orders. In Gunter Grieser, Yuzuru Tanaka, and Akihiro Yamamoto, editors, *Discovery Science*, pages 194–207, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. doi:10.1007/978-3-540-39644-4_ 17.
- 109 K Keller and E Petrov. Ordinal spaces. Acta Mathematica Hungarica, 160(1):119–152, 2020. doi:10.1007/s10474-019-00972-z.

- 110 Matthias Keller. Intrinsic metrics on graphs: A survey. In Delio Mugnolo, editor, Mathematical Technology of Networks, pages 81–119, Cham, 2015. Springer International Publishing. doi:10.1007/978-3-319-16619-3_7.
- 111 Maurice G. Kendall. Rank correlation methods. C. Griffin, 1948.
- 112 Ales Keprt and Václav Snásel. Binary factor analysis with help of formal concepts. In CLA, volume 110, pages 90–101, 2004. URL: https: //ceur-ws.org/Vol-110/paper10.pdf.
- 113 Aleš Keprt. Algorithms for Binary Factor Analysis. PhD thesis, PhD thesis, 2006.
- 114 S.C. Kleene. Representation of events in nerve nets and finite automata. Annals of Mathematics Studies, 34:3-41, 1956. doi:10.1515/ 9781400882618-002.
- 115 Matthäus Kleindessner and Ulrike Luxburg. Uniqueness of ordinal embedding. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, Proceedings of The 27th Conference on Learning Theory, volume 35 of Proceedings of Machine Learning Research, pages 40-67, Barcelona, Spain, 13-15 June 2014. PMLR. URL: http://proceedings.mlr.press/ v35/kleindessner14.html.
- 116 Matthäus Kleindessner and Ulrike Luxburg. Dimensionality estimation without distances. In Artificial Intelligence and Statistics, pages 471–479, 2015. URL: http://proceedings.mlr.press/v38/ kleindessner15.html.
- 117 Sigrid Knecht and Rudolf Wille. Congruence lattices of finite lattices as concept lattices. In J. Almeida, G. H. Bordalo, and P. Dwinger, editors, *Lattices, semigroups, and universal algebra*, pages 323–325. Plenum Press, New York–London, 1990. doi:10.1007/978-1-4899-2608-1_32.
- 118 Maren Koyda and Gerd Stumme. Factorizing lattices by interval relations. International Journal of Approximate Reasoning, 157:70-87, 2023. doi:10.1016/J.IJAR.2023.03.003.
- 119 Patryk Kozieł and Małgorzata Sulkowska. Uniform random posets. *Information Sciences*, 515:294–301, 2020. doi:10.1016/J.INS.2019.12.018.
- 120 Petr Krajca, Jan Outrata, and Vilem Vychodil. Parallel algorithm for computing fixpoints of galois connections. Annals of Mathematics and Artificial Intelligence, 59(2):257–272, 2010. doi: 10.1007/S10472-010-9199-5.
- 121 David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. Foundations of Measurement Additive and Polynomial Representations, volume 1. Academic Press, 1971. doi: 10.1016/C2013-0-11004-5.
- 122 Francesco Kriegel and Daniel Borchmann. Nextclosures: parallel computation of the canonical base with background knowledge. *International Journal of General Systems*, 46(5):490–510, 2017. doi: 10.1080/03081079.2017.1349570.
- 123 Rajeev Kumar, BK Verma, and Shyam Sunder Rastogi. Social popularity based svd++ recommender system. International Journal of Computer Applications, 87(14), 2014. doi:10.5120/15279-4033.
- 124 Léonard Kwuida and Stefan E. Schmidt. Valuations and closure operators on finite lattices.

Discrete Applied Mathematics, 159(10):990–1001, 2011. doi:10.1016/J.DAM.2010.11.022.

- 125 Jens Kötters and Peter W. Eklund. Conjunctive query pattern structures: A relational database model for formal concept analysis. *Discrete Applied Mathematics*, 273:144–171, 2020. Advances in Formal Concept Analysis: Traces of CLA 2016. doi:10.1016/J.DAM.2019.08.019.
- 126 Abdul Latif, Antonio-Francisco Roldán-López-de Hierro, and Wutiphol Sintunaravat. On common α -fuzzy fixed points with applications. *Fixed Point Theory Appl.*, 2014:22, 2014. Id/No 234. doi:10.1186/1687-1812-2014-234.
- 127 Deutsche Fußball Liga. Spielordnung, 2019. URL: https://media.dfl.de/sites/2/2019/09/ Spielordnung-SpOL-2019-08-22-Stand.pdf.
- 128 Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. doi:10.1007/BF01200757.
- 129 Tie-Yan Liu. Learning to rank for information retrieval. Foundations and Trends[®] in Information Retrieval, 3(3):225–331, 2009. doi:10.1561/ 1500000016.
- 130 F. M. Lord. On the statistical treatment of football numbers. American Psychologist, 8(12):750–751, 1953. doi:10.1037/h0063675.
- 131 R. Duncan Luce, David H. Krantz, Patrick Suppes, and Amos Tversky. Foundations of Measurement Representation, Axiomatization, and Invariance, volume 3. Academic Press, 1990. doi:10.1016/C2009-0-21635-7.
- 132 R. Duncan Luce and John W. Tukey. Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychol*ogy, page 27, 1964. doi:10.1016/0022-2496(64) 90015-x.
- 133 Peter Luksch and Rudolf Wille. Substitution decomposition of concept lattices. In G. Eigenthaler, H. K. Kaiser, W. B. Müller, and W. Nöbauer, editors, *Contributions to general algebra*, volume 5, pages 213–220, Wien, 1987. Hölder–Pichler– Tempsky.
- 134 Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. Journal of machine learning research, 9(Nov):2579-2605, 2008. URL: http://jmlr.org/papers/v9/ vandermaaten08a.html.
- 135 H. M. MacNeille. Partially ordered sets. Trans. Am. Math. Soc., 42:416–460, 1937. doi:10.1090/ s0002-9947-1937-1501929-x.
- 136 Hiroshi Maehara. Euclidean embeddings of finite metric spaces. Discrete Mathematics, 313(23):2848– 2856, 2013. doi:10.1016/J.DISC.2013.08.029.
- 137 K. Meinke and J. V. Tucker. Universal Algebra, pages 189–368. Oxford University Press, Inc., USA, 1993. URL: https://dl.acm.org/doi/10.5555/ 162573.162534.
- 138 Facundo Memoli. On the use of gromov-hausdorff distances for shape comparison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi: 10.2312/SPBG/SPBG07/081-090.
- 139 Joel Michell. Quantitative science and the definition of measurement in psychology. British

Journal of Psychology, 88(3):355-383, 1997. doi: 10.1111/j.2044-8295.1997.tb02641.x.

- 140 Charles Minard. Carte figurative des pertes successives en hommes de l'armée française dans la campagne de russie 1812-1813, 1869. URL: https:// commons.wikimedia.org/wiki/File:Minard.png.
- 141 Ann Mische. Relational sociology, culture, and agency. The SAGE handbook of social network analysis, pages 80–97, 2011. doi:10.4135/ 9781446294413.n7.
- 142 Rokia Missaoui and Kevin Emamirad. Lattice miner-a formal concept analysis tool. In 14th International Conference on Formal Concept Analysis, Supplementary Proceedings, page 91, 2017.
- 143 B. Monjardet. Metrics on partially ordered sets—a survey. Discrete Mathematics, 35(1):173–184, 1981. Special Volume on Ordered Sets. doi: 10.1016/0012-365X(81)90206-5.
- 144 Frederick Mosteller and John W Tukey. Data analysis and regression. a second course in statistics. Addison-Wesley series in behavioral science: quantitative methods, 1977.
- 145 Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check, 2020. doi: 10.1007/978-3-030-58595-2_41.
- 146 Facundo Mémoli. Metric structures on datasets: stability and classification of algorithms. In International Conference on Computer Analysis of Images and Patterns, pages 1–33. Springer, 2011. doi:10.1007/978-3-642-23678-5_1.
- 147 Facundo Mémoli. Some properties of gromovhausdorff distances. Discrete & Computational Geometry, 48(2):416–440, 2012. doi:10.1007/ S00454-012-9406-8.
- 148 Takao Nishizeki and Md Saidur Rahman. Planar graph drawing, volume 12. World Scientific Publishing Company, 2004. doi:10.1142/5648.
- 149 Megan Norris and Luc Lecavalier. Evaluating the use of exploratory factor analysis in developmental disability psychological research. Journal of autism and developmental disorders, 40(1):8–20, 2010. doi:10.1007/s10803-009-0816-2.
- 150 Charles S. Peirce. Collected Papers. Harvard Universit Press, Cambridge, 1931–1935. doi: 10.5860/choice.37-1485.
- 151 J. Pfanzagl. Theory of Measurement. Physica, Heidelberg, 1971. doi:10.1007/978-3-662-41488-0.
- 152 Gregory Piatetsky-Shapiro. Knowledge discovery in real databases: A report on the ijcai-89 workshop. AI Magazine, 11(5):68–70, 1991. doi:10.1609/AIMAG.V11I4.873.
- 153 Jonas Poelmans, Sergei O. Kuznetsov, Dmitry I. Ignatov, and Guido Dedene. Formal concept analysis in knowledge processing: A survey on models and techniques. *Expert Syst. Appl.*, 40(16):6601– 6623, 2013. doi:10.1016/J.ESWA.2013.05.007.
- 154 Silke Pollandt and Rudolf Wille. Functorial scaling of ordinal data. Discret. Appl. Math., 147(1):101– 111, 2005. doi:10.1016/J.DAM.2004.06.024.
- 155 Klaus Reuter. A note on the number of irreducibles of subdirect products. Algebra Universalis, 22:304– 305, 1986. doi:10.1007/bf01224035.
- 156 Klaus Reuter. Matchings of linearly indecomposable modular lattices. Discrete Mathematics, 63:245–247, 1987. doi:10.1016/0012-365X(87) 90013-6.

- 157 Klaus Reuter and Rudolf Wille. Complete congruence relations of concept lattices. Acta Sci. Math., 51:319-327, 1987. URL: http://acta.bibl. u-szeged.hu/id/eprint/15144.
- 158 Fred S. Roberts. Measurement Theory. Cambridge University Press, 1984. doi:10.1017/ CB09780511759871.
- 159 Ryan A Rossi, Luke K McDowell, David William Aha, and Jennifer Neville. Transforming graph data for statistical relational learning. *Journal of Artificial Intelligence Research*, 45:363–441, 2012. doi:10.1613/JAIR.3659.
- 160 Camille Roth and Paul Bourgine. Lattice-based dynamic and overlapping taxonomies: The case of epistemic communities. *Scientometrics*, 69(2):429– 447, 2006. doi:10.1007/S11192-006-0161-6.
- 161 Andreas Schmidt and Gerd Stumme. Prominence and dominance in networks. In Catherine Faron Zucker, Chiara Ghidini, Amedeo Napoli, and Toussaint Yannick, editors, Proceedings of the 21th International Conference on Knowledge Engineering and Knowledge Management (EKAW), LNCS, pages 370–385. Springer, 2018. doi:10.1007/978-3-030-03667-6_24.
- 162 Annemarie Zand Scholten and Denny Borsboom. A reanalysis of lord's statistical treatment of football numbers. Journal of Mathematical Psychology, 53(2):69-75, 2009. doi:10.1016/j.jmp.2009.01. 002.
- 163 Bernd Schröder. Ordered sets. An introduction with connections from combinatorics to topology. Basel: Birkhäuser/Springer, 2nd edition edition, 2016. doi:10.1007/978-3-319-29788-0.
- 164 Ernst Schröder. Algebra der Logik I, II, III. 1890, 1891, 1895. Thoemmes Press, Bristol, 2001.
- 165 Yuan Shi, Wenzhe Li, and Fei Sha. Metric learning for ordinal data. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pages 2030–2036, 2016. doi:10.1609/AAAI.V30I1. 10280.
- 166 Dan A. Simovici and Chabane Djeraba. Mathematical Tools for Data Mining - Set Theory, Partial Orders, Combinatorics. Second Edition. Advanced Information and Knowledge Processing. Springer, 2014. doi:10.1007/978-1-4471-6407-4.
- 167 Gregor Snelting and Frank Tip. Understanding class hierarchies using concept analysis. ACM Trans. Program. Lang. Syst., 22(3):540–582, 2000. doi:10.1145/353926.353940.
- 168 John Sowa. Conceptual Structures: Information Processing in Mind and Machine. The Systems Programming Series. Addison-Wesley, 1984.
- 169 N Spangenberg and KE Wolff. Conceptual grid evaluation, classification and related methods of data analysis, 1988.
- 170 C. Spearman. The proof and measurement of association between two things. *The Ameri*can Journal of Psychology, 15(1):72–101, 1904. doi:10.2307/1422689.
- 171 Suvrit Sra and Inderjit S. Dhillon. Generalized nonnegative matrix approximations with bregman divergences. In Advances in neural information processing systems, pages 283-290, 2006. URL: https: //proceedings.neurips.cc/paper/2005/hash/ d58e2f077670f4de9cd7963c857f2534-Abstract. html.

- 172 Jörg Stephan. Substitution decomposition of lattices. PhD thesis, TH Darmstadt, 1991.
- 173 Jörg Stephan. Substitution products of lattices. In D. Dorninger, G. Eigenthaler, H. K. Kaiser, and W. B. Müller, editors, *Contributions to general algebra*, volume 7, pages 321–336, Wien, 1991. Hölder–Pichler–Tempsky.
- 174 S. S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946. doi: 10.1126/science.103.2684.677.
- 175 Selma Strahringer. Direct products of convexordinal scales. Order, 11:361–383, 1994. doi: 10.1007/bf01108768.
- 176 Selma Strahringer and Rudolf Wille. Towards a structure theory for ordinal data. In M. Schader, editor, Analysing and modelling data and knowledge, pages 129–139, Heidelberg, 1992. Springer-Verlag. doi:10.1007/978-3-642-46757-8_14.
- 177 Selma Strahringer, Rudolf Wille, and Uta Wille. Mathematical support for empirical theory building. In Harry S. Delugach and Gerd Stumme, editors, Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings, volume 2120 of LNCS, pages 169–186. Springer, 2001. doi: 10.1007/3-540-44583-8_13.
- 178 Maximilian Stubbemann, Tom Hanika, and Friedrich Martin Schneider. Intrinsic dimension for large-scale geometric learning. *Transactions on Machine Learning Research*, 2023. URL: https: //openreview.net/forum?id=85BfDdYMBY.
- 179 Maximilian Stubbemann, Tom Hanika, and Gerd Stumme. Orometric methods in bounded metric data. In Michael R. Berthold, Ad Feelders, and Georg Krempl, editors, Advances in Intelligent Data Analysis XVIII 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27-29, 2020, Proceedings, volume 12080 of LNCS, pages 496–508. Springer, 2020. doi:10.1007/978-3-030-44584-3_38.
- 180 Maximilian Stubbemann and Gerd Stumme. The mont blanc of twitter: Identifying hierarchies of outstanding peaks in social networks. In Proc. 34th European Conference on Machine Learning / 27th Intl. Conference on Prinicples of Knowledge Discovery in Databases, 2023. doi:10.1007/978-3-031-43418-1_11.
- 181 Milan Studeny. Probabilistic Conditional Independence Structures. Springer Publishing Company, Incorporated, 1st edition, 2010. doi:10. 1007/b138557.
- 182 G. Stumme. Hierarchies of conceptual scales. In T. B. Gaines, R. Kremer, and M. Musen, editors, *Proc. Workshop on Knowledge Acquisition, Model*ing and Management (KAW'99), volume 2, pages 78–95. Banff, October 16-22 1999.
- 183 G. Stumme. Conceptual on-line analytical processing. In K. Tanaka, S. Ghandeharizadeh, and Y. Kambayashi, editors, *Information Organization and Databases*, chapter 14, pages 191–203. Kluwer, Boston-Dordrecht-London, 2000. doi: 10.1007/978-1-4615-1379-7_14.
- 184 G. Stumme. Off to new shores conceptual knowledge discovery and processing. Intl. J. Human-

Comuter Studies (IJHCS), 59(3):287–325, sep 2003. doi:10.1016/S1071-5819(03)00044-2.

- 185 G. Stumme and A. Maedche. Fca-merge: Bottomup merging of ontologies. In B. Nebel, editor, Proc. 17th Intl. Conf. on Artificial Intelligence (IJCAI '01), pages 225-230, Seattle, WA, USA, 2001. URL: https://www.ijcai.org/ Proceedings/01/IJCAI-2001-b.pdf.
- 186 Gerd Stumme. Knowledge acquisition by distributive concept exploration. In G. Ellis, R. Levinson, W. Rich, and J. F. Sowa, editors, Conceptual structures: applications, implementation and theory, number 954 in Lecture Notes in Artificial Intelligence, Berlin-Heidelberg-New York, 1995. Springer-Verlag. (supplementary proceedings). doi:10.1007/3-540-60161-9.
- 187 Gerd Stumme. Attribute exploration with background implications and exceptions. In H.-H. Bock and W. Polasek, editors, *Data analysis* and information systems, pages 457–469, Berlin-Heidelberg-New York, 1996. Springer-Verlag. doi: 10.1007/978-3-642-80098-6_39.
- 188 Gerd Stumme. Local scaling in conceptual data systems. In P. W. Eklund, G. Ellis, and G. Mann, editors, *Conceptual Structures: Knowledge Repre*sentation as Interlingua. Proc. ICCS'96, volume 1115 of LNAI, pages 308–320, Heidelberg, 1996. Springer. doi:10.1007/3-540-61534-2_20.
- 189 Gerd Stumme. Concept exploration a tool for creating and exploring > conceptual hierarchies. In D. Lukose, H. Delugach, M. Keeler, L. Searle, and J. F. Sowa, editors, *Conceptual* structures: Fulfilling Peirce's dream, volume 1257 of Lecture Notes in Artificial Intelligence, Berlin-Heidelberg-New York, 1997. Springer-Verlag. doi: 10.1007/BFB0027880.
- 190 Gerd Stumme. Free distributive completions of partial complete lattices. In: Order, 14:179–189, 1998. doi:10.1023/A:1006002507294.
- 191 Gerd Stumme. On-line analytical processing with conceptual information systems. In K. Tanaka and S. Ghandeharizadeh, editors, Proc. 5th Intl. Conf. on Foundations of Data Organization (FODO'98), pages 117–126, November 12-13 1998.
- 192 Gerd Stumme. Acquiring expert knowledge for the design of conceptual information systems. In D. Fensel and R. Studer, editors, *Knowledge* Acquisition, Modeling, and Management. Proc. 11th European, volume 1621 of LNAI, pages 275– 290, Heidelberg, 1999. Springer. doi:10.1007/ 3-540-48775-1_17.
- 193 Gerd Stumme. A finite state model for on-line analytical processing in triadic contexts. In Bernhard Ganter and Robert Godin, editors, *Proc. 3rd Intl. Conf. on Formal Concept Analysis*, volume 3403 of *LNCS*, pages 315–328, Heidelberg, 2005. Springer. doi:10.1007/978-3-540-32262-7_22.
- 194 Gerd Stumme and Rudolf Wille. A geometrical heuristic for drawing concept lattices. In R. Tamassia and I.G. Tollis, editors, *Graph Drawing*, volume 894 of *LNCS*, pages 452–459, Heidelberg, 1995. Springer. doi:10.1007/3-540-58950-3_399.
- 195 Gerd Stumme and Rudolf Wille, editors. Begriffliche Wissensverarbeitung – Methoden und Anwendungen, Heidelberg, 2000. Springer. doi: 10.1007/978-3-642-57217-3.

- 196 Karl-Theodor Sturm et al. On the geometry of metric measure spaces. Acta mathematica, 196(1):65– 131, 2006. doi:10.1007/s11511-006-0002-8.
- 197 Kozo Sugiyama, Shojiro Tagawa, and Mitsuhiko Toda. Methods for visual understanding of hierarchical system structures. *IEEE Transactions* on Systems, Man, and Cybernetics, 11(2):109–125, 1981. doi:10.1109/TSMC.1981.4308636.
- 198 Patrick Suppes, David H. Krantz, R. Duncan Luce, and Amos Tversky. Foundations of Measurement – Geometrical, Threshold, and Probabilistic Representations, volume 2. Academic Press, 1989. doi:10.1016/C2009-0-21665-5.
- 199 Atsushi Suzuki, Atsushi Nitanda, Jing Wang, Linchuan Xu, Kenji Yamanishi, and Marc Cavazza. Generalization error bound for hyperbolic ordinal embedding. In Marina Meila and Tong Zhang, editors, Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 10011– 10021. PMLR, 18–24 July 2021. URL: http:// proceedings.mlr.press/v139/suzuki21a.html.
- **200** Viola Takàcs. Two applications of Galois graphs in pedagogical research. Manuscript of a lecture given at TH Darmstadt, 1984.
- 201 Eran Tal. Measurement in science. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philos*ophy. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017. URL: https://plato. stanford.edu/Entries/measurement-science/.
- 202 M. Thomas. Mathematization, not measurement: A critique of stevens' scales of measurement. Journal of Methods and Measurement in the Social Sciences, 10(2):76-94, 2020. doi:10.2458/v10i2. 23785.
- 203 Ioannis Tollis, Giuseppe Di Battista, Peter Eades, and Roberto Tamassia. Graph drawing. Algorithms for the visualization of graphs. Upper Saddle River, NJ: Prentice Hall, 1999.
- 204 Günter Trendler. Measurement theory, psychology and the revolution that cannot happen. Theory & Psychology, 19:579–599, oct 2009. doi: 10.1177/0959354309341926.
- 205 Günter Trendler. Conjoint measurement undone. Theory & Psychology, 29(1):100–128, 2019. doi: 10.1177/0959354318788729.
- **206** William T. Trotter. Combinatorics and Partially Ordered Sets: Dimension Theory. Johns Hopkins University Press, 2001.
- 207 William T. Trotter and Bartosz Walczak. Boolean dimension and local dimension. *Electron. Notes Discret. Math.*, 61:1047–1053, 2017. doi:10.1016/ J.ENDM.2017.07.071.
- 208 John W. Tukey. The future of data analysis. Annals of Mathematical Statistics, 33:1-67, 1962. doi:10.1214/aoms/1177704711.
- 209 Alexey A. Tuzhilin. Who invented the gromov-hausdorff distance?, 2016. cite arxiv:1612.00728Comment: 7 pages, 10 references. doi:10.48550/arXiv.1612.00728.
- 210 Leena Chennuru Vankadara and Ulrike von Luxburg. Measures of distortion for machine learning. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, Advances in Neural Information Processing

Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada, pages 4891-4900, 2018. URL: https: //proceedings.neurips.cc/paper/2018/hash/ 4c5bcfec8584af0d967f1ab10179ca4b-Abstract. html.

- 211 Fei Wang and Jimeng Sun. Survey on distance metric learning and dimensionality reduction in data mining. *Data Min. Knowl. Discov.*, 29(2):534–564, 2015. doi:10.1007/S10618-014-0356-Z.
- 212 Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the Ameri*can Statistical Association, 58(301):236–244, 1963. doi:10.1080/01621459.1963.10500845.
- 213 Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. doi:10.1007/978-1-4612-4380-9_16.
- 214 Marcel Wild. A theory of finite closure spaces based on implications. Advances in Mathematics, 108(1):118-139, 1994. doi:10.1006/aima.1994. 1069.
- 215 R. Wille. Concept lattices and conceptual knowledge systems. Computers and Mathematics with Applications, 23:493–515, 1992. doi:10.1016/ 0898-1221(92)90120-7.
- 216 Rudolf Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In Ivan Rival, editor, *Ordered sets*, pages 445–470. Reidel, 1982. doi:10.1007/978-94-009-7798-3_15.
- 217 Rudolf Wille. Complete tolerance relations of concept lattices. In G. Eigenthaler, H. K. Kaiser, W. B. Müller, and W. Nöbauer, editors, *Contributions to general algebra*, volume 3, pages 397–415. Hölder–Pichler–Tempsky, Wien, 1985.
- 218 Rudolf Wille. Finite distributive lattices as concept lattices. Atti Inc. Logica Mathematica, 2:635–648, 1985.
- 219 Rudolf Wille. Tensorial decomposition of concept lattices. Order, 2:81–95, 1985. doi:10.1007/ bf00337926.
- 220 Rudolf Wille. Subdirect product construction of concept lattices. Discrete Mathematics, 63:305– 313, 1987. doi:10.1016/0012-365X(87)90019-7.
- 221 Rudolf Wille. Lattices in data analysis: how to draw them with a computer. In Ivan Rival, editor, Algorithms and order, pages 33–58, Dordrecht-Boston, 1989. Kluwer. doi:10.1007/ 978-94-009-2639-4_2.

- 222 Rudolf Wille. Tensor products of complete lattices as closure systems. In D. Dorninger, G. Eigenthaler, H. K. Kaiser, and W. B. Müller, editors, *Contributions to general algebra*, volume 7, pages 381–385. Hölder–Pichler–Tempsky, Wien, 1991.
- 223 Rudolf Wille and Uta Wille. Uniqueness of coordinatizations of ordinal structures. In *Contributions* to general algebra, volume 9, pages 321–324, Wien, 1995. Hölder–Pichler–Tempsky.
- 224 Rudolf Wille and Uta Wille. Coordinatization of ordinal structures. Order, 13, 1996. doi: 10.1007/bf00338747.
- 225 Uta Wille. Representation of ordinal contexts by ordered n-quasigroups. Eur. J. Comb., 17(2-3):317– 333, 1996. doi:10.1006/EUJC.1996.0027.
- 226 Uta Wille. Linear measurement models axiomatizations and axiomatizability. Journal of Mathematical Psychology, 44(4):617-650, 2000. doi:10.1006/jmps.2000.1326.
- 227 Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. doi:10.1016/0169-7439(87)80084-9.
- 228 Mihalis Yannakakis. The complexity of the partial order dimension problem. SIAM Journal on Algebraic Discrete Methods, 3(3):351-358, 1982. doi:10.1137/0603036.
- 229 Javier Yáñez and Javier Montero. A poset dimension algorithm. J. Algorithms, 30(1):185–208, 1999. doi:10.1006/JAGM.1998.0974.
- 230 Alexander Zap, Tobias Joppen, and Johannes Fürnkranz. Deep ordinal reinforcement learning. In Ulf Brefeld, Élisa Fromont, Andreas Hotho, Arno J. Knobbe, Marloes H. Maathuis, and Céline Robardet, editors, Machine Learning and Knowledge Discovery in Databases – European Conference, ECML PKDD 2019, Würzburg, Germany, September 16-20, 2019, Proceedings, Part III, volume 11908 of LNCS, pages 3–18. Springer, 2019. doi:10.1007/978-3-030-46133-1_1.
- 231 Yiqun Zhang and Yiu ming Cheung. An ordinal data clustering algorithm with automated distance learning. In The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, pages 6869–6876. AAAI Press, 2020. doi:10.1609/AAAI.V34I04.6168.

Rule Learning over Knowledge Graphs: A Review

Hong Wu ⊠©

School of Information and Communication Technology, Griffith University, Australia

Zhe Wang ⊠©

School of Information and Communication Technology, Griffith University, Australia

Kewen Wang¹ \square \square

School of Information and Communication Technology, Griffith University, Australia

Pouya Ghiasnezhad Omran 🖂 🖻

School of Computing, The Australian National University, Australia

Jiangmeng Li 🖂 🗈

Institute of Software, Chinese Academy of Sciences, China

— Abstract -

Compared to black-box neural networks, logic rules express explicit knowledge, can provide humanunderstandable explanations for reasoning processes, and have found their wide application in knowledge graphs and other downstream tasks. As extracting rules manually from large knowledge graphs is labour-intensive and often infeasible, automated rule learning has recently attracted significant interest, and a number of approaches to rule learning for knowledge graphs have been proposed. This survey aims to provide a review of approaches and a classification of state-of-the-art

systems for learning first-order logic rules over knowledge graphs. A comparative analysis of various approaches to rule learning is conducted based on rule language biases, underlying methods, and evaluation metrics. The approaches we consider include inductive logic programming (ILP)-based, statistical path generalisation, and neuro-symbolic methods. Moreover, we highlight important and promising application scenarios of rule learning, such as rule-based knowledge graph completion, fact checking, and applications in other research areas.

2012 ACM Subject Classification Computing methodologies \rightarrow Knowledge representation and reasoning; Information systems \rightarrow Data mining

Keywords and phrases Rule learning, Knowledge graphs, Link prediction

Digital Object Identifier 10.4230/TGDK.1.1.7

Category Survey

Acknowledgements The authors would like to thank the editors and the anonymous referees for their constructive comments that have helped improve the quality of this paper.

Received 2023-07-02 Accepted 2023-08-31 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1 Introduction

Knowledge graphs are a popular form of knowledge bases that describe facts about real-world entities and their relations. They serve as powerful tools for organising and modelling information in a way that allows for efficient storage, retrieval, and reasoning. Recently, they have garnered significant attention in both academia and industry. Many public knowledge graphs have been developed, such as Freebase [8], WordNet [46], YAGO [69], DBPedia [3], and WikiData [76]. They

¹ The corresponding author.



© Hong Wu, Zhe Wang, Kewen Wang, Pouya G. Omran and Jiangmeng Li; licensed under Creative Commons License CC-BY 4.0 Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 7, pp. 7:1-7:23

Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

7:2 Rule Learning over Knowledge Graphs: A Review

highlight the pivotal role these graphs play as expansive and valuable resources, supporting a wide array of applications in artificial intelligence, data analysis, and knowledge representation. Additionally, numerous commercial knowledge graphs have been created, such as Google KG [18], Microsoft Satori [34], and Facebook Graph Search [19]. These knowledge graphs have demonstrated their capability to provide more efficient services for other products in their companies. They enable efficient querying and reasoning, allowing users and applications to gain valuable insights and make informed decisions based on interconnected knowledge. Many knowledge graphs are large-scale with millions of entities and facts. For instance, in the case of the DBpedia Core Release, the 2016-04 edition of the DBpedia dataset, it contains 6.0 million entities and 9.5 billion triples. These numbers underscore the substantial growth in both the size and complexity of knowledge graphs. Some medium-sized knowledge graphs encompass tens of thousands of entities and several hundreds of thousands of facts.

The rise of knowledge graphs is intricately linked to the advancement of the Semantic Web [7]. Abiding by the triple-based definition of the Semantic Web, a knowledge graph is a set of RDF triples such as (Allen, livesIn, NewYork), which means Allen lives in the city of New York. As highlighted by some researchers [49, 6], a knowledge graph goes beyond being a simple graph database. Data management and reasoning in knowledge graphs can be empowered by a layer of conceptual knowledge, known as the ontology layer, and such conceptual knowledge can be represented as logic rules. For instance, consider a rule 0.9: hasBirthPlace $(x, y) \land isRegionOf(y, z) \rightarrow$ hasNationality(x, z). This rule implies that if x was born in the place y of country z, then x may have the nationality of z with a confidence degree of 0.9. Such rules can be used for reasoning over knowledge graphs, enabling the prediction of new facts based on existing ones. Unlike black-box deep neural networks, rules offer explicit high-level knowledge and can provide humanunderstandable explanations for the reasoning processes. However, crafting rules manually for large knowledge graphs is challenging, if not practically impossible. Therefore, the automatic extraction of high-level rules becomes both useful and important for knowledge graphs. The learned rules can be directly applied to reasoning in knowledge graph completion tasks. Additionally, they can be indirectly utilised to enhance the interpretability of neural network models for knowledge graph reasoning.

Learning Horn clauses has been studied extensively in the *inductive logic programming (ILP)* literature [47, 47, 15]. In the context of ILP-based approaches, first-order Horn rules are explored by systematically exploring the rule space through various refinement operators. Classical ILP systems [92, 68] cannot be used directly to handle knowledge graphs due to the lack of negative examples and the large data sizes. Recently, ILP-based rule learners such as AMIE [24] and its extensions [23, 37] have been developed with the aim of handling knowledge graphs. Thus, predicates in rules learned by AMIE+ are binary. This language bias helps significantly reduce the search space in rule learning. They use plausibility metrics adapted from association rule mining to address the lack of negative examples. ILP-based rule learners for knowledge graphs usually assign a confidence score for each learned rule. Such approaches are also referred to as *Probabilistic* Inductive Logic Programming (PILP) in the literature [65]. Another group of approaches generate candidate rules by directly exploring frequent patterns or paths of different granularity. A typical system in this group is AnyBURL [45, 44], which samples paths within knowledge graphs and generalises them to form rules. It generalises path instances by substituting entities from sampled paths with variables, thus forming rule patterns. The effectiveness of such an approach relies on the ability to sample representative paths and apply suitable statistical metrics. Apart from instance-level paths, some other rule learners explore ontological-level paths within knowledge graphs [13, 12, 56], which significantly reduces the search space of paths. Recently, there has been an emerging interest in integrating neural networks into the realm of rule learning. Neural

H. Wu, Z. Wang, K. Wang, P. G. Omran and J. Li

networks have shown remarkable success in various machine learning tasks and can automatically learn feature representations from raw data, including knowledge graphs. These neuro-symbolic approaches have the advantage of simultaneously learning both rule structures and parameters. Neural LP [89] is the first attempt to propose a framework combining both of the learning in an end-to-end differentiable model. Another group combines neural network models with other rulelearning strategies through knowledge graph embeddings, like EMBEDRULE [88] and RLvLR [51]. The incorporation of embedding can improve the scalability of rule learning over large knowledge graphs.

In this paper, we survey major approaches to learning first-order Horn rules over knowledge graphs, aiming to serve as a resource for researchers and practitioners in rule discovery over knowledge graphs. We are unaware of any similar survey paper on automated rule learning over knowledge graphs. More specifically, this paper provides a comprehensive review of state-of-the-art rule learners. It presents a comparative analysis of various approaches to rule learning, considering factors such as language bias, evaluation metrics, and underlying methods. Secondly, a categorisation of rule learning methods and techniques is provided. The surveyed approaches encompass three main categories: ILP-based, statistical path generalisation, and neuro-symbolic methods. Thirdly, this survey investigates the important and promising application scenario of logic rules, offering valuable insights into the current and future directions of this important field. By providing a comprehensive overview of the state-of-the-art approaches and highlighting the challenges and opportunities in rule learning, we hope to inspire further research and innovations in this area. We believe that this survey will facilitate knowledge exchange and collaboration among scholars and industry professionals, ultimately leading to significant contributions to the field of rule learning and beyond.

The rest of this paper is organised as follows. Section 2 provides an overview of our survey paper, introducing the definition of knowledge graphs, first-order Horn rules, a classification of learning methods, and the rule evaluation metrics. Sections 3, 4 and 5 focus on three categories of rule learning methods, respectively. Section 6 introduces the applications of first-order Horn rules automatically learned by rule learners. Finally, we discuss the future directions of rule learning and conclude the paper in Section 7.

2 Overview

In this section, we will first fix some definitions and notations in knowledge graphs and rule languages that will be used in the paper. Then, we formulate the problem of rule learning in knowledge graphs, and then propose a classification of rule learning methods. A method of rule learning is essentially a process of ranking candidate rules. So, we will also introduce three confidence measures of rules before we discuss specific methods of rule learning later.

2.1 Knowledge Graphs and Rules

In this subsection, we introduce the basics of knowledge graphs and rules.

Knowledge graphs represent real-world entities, such as persons and places, and binary relations among them. A knowledge graph (KG) is often expressed as a set of triples of the form (s, p, o), where entities s and o are called the subject and object of the triple, respectively, and p is the relation. A KG is essentially a directed multi-relational graph by viewing the entity (the subjects and objects) as the vertices and a triple (s, p, o) as an edge from s to o with the label p. For instance, a triple (airline-NY, hasBase, airport-JFK) describes that the two entities airline-NY and airport-JFK are connected by the relation hasBase. Following the convention in knowledge representation, a triple is also denoted as a fact p(s, o).

7:4 Rule Learning over Knowledge Graphs: A Review

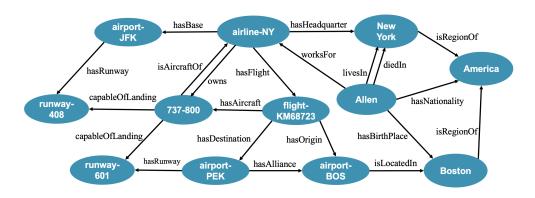


Figure 1 An example knowledge graph.

Formally, let \mathcal{E} and \mathcal{P} be respectively the sets of entities and relations in a KG \mathcal{G} . For a relation p, p^- denotes its *inverse*, i.e., triple (runway-408, hasRunway⁻¹, airport-JFK) is equivalent to (airport-JFK, hasRunway, runway-408). And \mathcal{P}^* denotes the set of the relations and their inverse relations, i.e., $\mathcal{P} \cup \{p^- \mid p \in \mathcal{P}\}$. In this paper, we consider the class of *first-order Horn rules*, which is sufficiently expressive for many practical applications in the Semantic Web and AI, and allows efficient reasoning algorithms. Moreover, in KGs, only binary and unary relations are considered. While a binary relation connects two entities, a unary relation represents a type (or class) of entities. In first-order logic, a relation is expressed as a predicate. Whenever no confusion is caused, we use these two terms alternatively.

A term is either an entity or a variable. If p is a binary relation, $p(t_1, t_2)$ is an atom, where t_1 and t_2 are terms. Similarly, if p is a unary relation, p(t) is an atom, where t is a term.

A first-order Horn rule r is of the form

$$b_1 \wedge \ldots \wedge b_n \to h \tag{1}$$

where h, b_1, \ldots, b_n are atoms. The atom h is the head of r, denoted head(r), and the conjunction of atoms b_1, \ldots, b_n is the body of r, denoted body(r). Intuitively, the rule r reads that if b_1, \ldots , and b_n hold, then h holds too. The *length* of the above rule body is n (the number of body atoms in the rule).

Due to the enormous search space of first-order Horn rules over large KGs, existing rule learning approaches often adopt certain language biases to restrict the forms of rules to learn, such as constraining the maximum length of rules, to effectively reduce the search space. This enables the rule-learning algorithm to be more efficient and applicable in practical scenarios. These constraints strike a balance between the size of the search space and the expressiveness of rules.

A most common language bias is to learn rules that represent path patterns in KGs, that is, the class of *closed-path* (*CP*) rules [81, 51]. Intuitively, in a CP rule, the body atoms form a path from the subject to the object of the head atom (involving only variables not entities). Formally, a *closed-path* rule is of the form

$$p_1(x_0, x_1) \land p_2(x_1, x_2) \land \dots \land p_n(x_{n-1}, x_n) \to p(x_0, x_n),$$
 (2)

where $p \in \mathcal{P}$, $p_i \in \mathcal{P}^*$ $(1 \le i \le n)$ and x_j 's $(0 \le j \le n)$ are variables. Note that CP rules allow recursion, i.e., the head predicate can occur in the body. The advantage of closed-path rules lies in their ability to capture specific and meaningful patterns in the data. These rules can reveal intricate dependencies, cyclic patterns, and sequential behaviours present in the data, providing

H. Wu, Z. Wang, K. Wang, P. G. Omran and J. Li

deeper insights into the underlying associations. Focusing on CP rules reduces the search for candidate rules to the problem of path finding and ranking in KGs. This type of rules have been widely adopted by major rule learners [88, 25, 13, 51, 56].

As the class of CP rules is too limited for some applications, one way to expand the class of CP rules is to allow rules that are *closed* and *connected*. Two atoms in a rule are connected if they share a variable or an entity. A rule is connected if every atom in the rule is connected to another atom. A variable in a rule is closed if it appears at least twice in the rule. A rule is closed if all of its variables are closed.

To avoid learning rules with unrelated atoms, some methods, such as AMIE and its variants [24, 23, 37], require that the graphs of learned rules to be *connected*. In addition, to avoid having variables with existential quantifier in the rule head, a learned rule must be *closed*. For example, the rule $p_1(x_0, x_1) \wedge p_2(x_1, x_2) \rightarrow p(x_0, x_1)$ is connected but not closed, while the rule $p_1(x_0, x_1) \wedge p_2(x_1, x_2) \rightarrow p(x_0, x_1)$ is connected. By the definition, it is clear that a CP rule is both connected and closed, but not vice versa. For example, $p_1(x_0, x_1) \wedge p_2(x_1, x_2) \wedge p_3(x_0, x_1) \rightarrow p(x_0, x_2)$ is both connected and closed, but not a CP rule.

Another natural extension of CP rules is to allow unary predicates, i.e., classes (or types), in the rules [82]. Such a rule also describes a path pattern in the KG, but allows to specify the classes (or types) of the nodes on the paths. The rule $person(x_0) \wedge hasBirthPlace(x_0, x_1) \wedge city(x_1) \wedge isRegionOf(x_1, x_2) \wedge country(x_2) \rightarrow hasNationality(x_0, x_2)$ is a typed rule. It specifies the classes (types) of x_0, x_1, x_2 to be Person, City, Country.

Some rule learners can learn more expressive rules beyond first-order Horn rules, allowing negations [21, 31, 75], numeric values [54, 77], temporal values [52, 42], etc. Such extensions are useful for practical applications. Specifically, some approaches can learn rules that involve comparisons among numeric values, for example, hasBirthYear(x, v_0) \land hasBirthYear(y, v_1) $\land v_0 > v_1 \rightarrow$ younger(x, y). This rule says if a person x was born after another person y, then x is younger than y. Here, the relation hasBirthYear takes literal numbers as its object datatype. Some approaches focus on learning nonmonotonic rules (or negated rules) with negated atoms in the rule body, such as bornln(x, y) \land not immigrate(x, z) \rightarrow livesln(x, y) says that a person x who was born in a place y and is not known to have migrated to z lives in y. Moreover, temporal rules can be learned over temporal KGs where every atom has a timestamp. For example, the rule bornln(x, y, t) \rightarrow diedln(x, y, t + 80) indicates that if a person x born in city y at timestamp t usually die in the same place at time t + 80.

2.2 A Classification of Rule Learning Methods

The task of *rule learning* is to automatically extract a set of first-order logic Horn rules over a given KG. Formally, given a KG \mathcal{G} , a rule learning system, a.k.a. *rule learner*, learns a set of rules r of the form $b_1 \wedge \ldots \wedge b_n \to h$ with a confidence degree $0 \leq \alpha_r \leq 1$ associated with each rule r. The relations and (possibly) entities in r are from \mathcal{G} , and r is considered plausible if there are many instances of r obtained by substituting the variables in r with entities in \mathcal{G} , such that the atoms in these instances are facts occurring in the KG \mathcal{G} . The more such instances exist the more plausible r is. The confidence degree α_r is calculated to reflect the plausibility of r, i.e., the more plausible r is the higher α_r should be.

Rule learning involves both learning the rule structures and estimating their plausibility. For example, in traditional Inductive Logic Programming, rule structure learning is achieved by systematically exploring the rule space by adding, updating, or deleting atoms at a time in the rule bodies. Rule plausibility is measured via example coverage by the set of learned rules. That is, given two sets of positive examples (true facts in the data) and negative examples (false or absent facts), the rule learner aims to induce a set of rules that cover as many positive examples

7:6 Rule Learning over Knowledge Graphs: A Review

and as few negative examples as possible. Yet, due to the large sizes and the lack of negative examples in KGs, the traditional ILP methods of traversing rule space for structure learning and measuring example coverage for plausibility estimation are not directly applicable to KGs.

To handle the large sizes, complexity, incompleteness and dynamics of KGs, many novel and efficient rule learners have been developed. We classify them by two dimensions as shown in Figure 2, according to their structure learning and confidence measure methods.

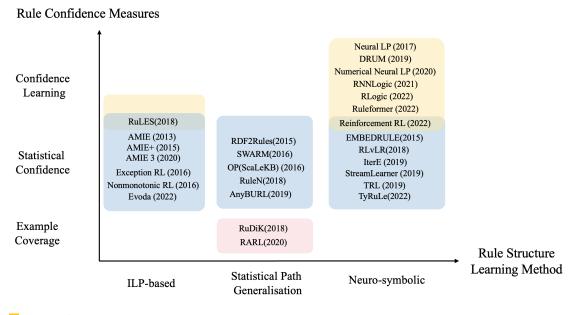


Figure 2 The categorisation of rule learners.

According to the rule structure learning methods, we can broadly categorise existing methods into three groups: the *inductive logic programming (ILP)-based* approaches, which use refinement operators to guide the search in the rule space, the *statistical path generalisation* methods, which extract frequent patterns from sampled paths or sub-graphs of the KGs, and the *neuro-symbolic* approaches, which directly or indirectly utilise neural networks to learn rules. We will introduce these three groups of methods with their representative works in Sections 3, 4, and 5.

According to the confidence measures, we classify the existing methods into three groups: the *example coverage* measures, which are in the same spirit as ILP approaches, by generating negative examples from KGs via a form of closed world assumption; the *statistical confidence* measures, which adapt statistical measures from association rule mining to address the lack of negative examples; and the *confidence learning* approaches, which learn confidence degrees as parameters of some neural networks. We will discuss these confidence measures in further detail in the remainder of this section.

Generally speaking, earlier works on KG rule learning are mostly ILP-based or statistical path generalisation methods, while most recent ones are largely neuro-symbolic methods. Statistic confidence measures are the most widely used, as they do not rely on the existence of negative examples and have better interpretability, i.e., statistical meanings. While confidence learning is only adopted by neuro-symbolic methods, example coverage measures have also been employed in statistical path generalisation methods. Interestingly, we have not found an ILP-based rule learner for KGs that adopts the example coverage measure which originates from ILP.

2.3 Rule Confidence Measures

Rule confidence measures play a crucial role in rule learning, as they indicate the plausibility of the learned rules, and the accuracy of such measures is critical for downstream tasks based on the learned rules. There are three main approaches for rule confidence measures, example coverage, statistical, and confidence learning measures.

2.3.1 Example Coverage

Traditional rule learning approaches, such as inductive logic programming (ILP), typically measure the confidence of rules via their coverage of observed positive and negative examples. It views a collection of rules as a classification model for the observed examples, and the goal of rule learning is to find a set of rules that cover as many positive examples as possible and as few or zero negative examples as possible. Specifically, for a first-order Horn rule r, intuitively a *positive example* is a tuple of entities **a** that satisfy the head of the rule head(r), i.e., when the variables **x** in h are substituted with **a**, the resulting atom is a fact in the KG. Similarly, a *negative example* is a tuple of entities **a'** that do not satisfy the head of the rule head(r), i.e., the substituted atom is not a fact in the KG. The body of r covers (or simply say r covers) a tuple of entities **a** if when **x** is substituted with **a**, there is a way to substitute the other variables in r to make all the facts obtained from the body of r facts in the KG.

Given two sets \mathcal{A}^+ and \mathcal{A}^- of respectively positive and negative examples, let $ex_{\mathcal{A}^+}(r)$ and $ex_{\mathcal{A}^-}(r)$ consist of respectively positive and negative examples covered by r. Formally, for a set of rules \mathcal{R} with the same head h, the *coverage* of \mathcal{R} w.r.t. \mathcal{A}^+ and \mathcal{A}^- is defined as

$$cover(\mathcal{R}, \mathcal{A}^+, \mathcal{A}^-) = \beta \cdot \frac{|\bigcup_{r \in \mathcal{R}} ex_{\mathcal{A}^+}(r)|}{|\mathcal{A}^+|} - (1 - \beta) \cdot \frac{|\bigcup_{r \in \mathcal{R}} ex_{\mathcal{A}^-}(r)|}{|\mathcal{A}^-|}.$$
(3)

Some rule learning approaches measure the confidence of individual rules based on variants of the coverage [54, 56], by applying it to individual rules r instead of a rule set \mathcal{R} . However, a major challenge is that KGs adopt the Open World Assumption (OWA), that is, missing triples are not necessarily false but just unknown, which makes them inherently lack negative examples. To address this issue, certain negative examples generation strategies are employed [54, 56]. For example, negative examples are generated in [54] by considering (s, p, o) a negative example if it does not occur in the KG and at least one triple of the form (s, p, o'), (s, p', o) or (s', p, o) occurs in the KG, essentially following the PCA assumption (will be discussed in Section 2.3.2). Although several safeguards are employed in the negative example generation processes, they may still introduce noisy examples.

2.3.2 Statistical Confidence

Inspired by association rule mining and viewing logic rules as frequent patterns in KGs, some statistical measures such as *support* and *confidence* have been adapted for rule learning over KGs [24, 23, 37]. One advantage of such measures is that negative examples are not required.

For a first-order Horn rule r of the form (1), $ex_H(r)$ consists of all the tuples of entities **a** that satisfy the head of the rule h. Similarly, $ex_B(r)$ consists of all the tuples of entities **a** covered by the body or r, i.e., when the variables **x** in the head h is substituted with **a**, there is a way to substitute the other variables in r to make all the facts obtained from the body of r facts in the KG. Then, the *support* of r is defined as $supp(r) = |ex_H(r) \cap ex_B(r)|$. That is, the support of ris defined as the number of entities that satisfy both the head and the body of r. The *standard confidence* (*SC*) and *head coverage* (*HC*) of r are defined as follows

7:8 Rule Learning over Knowledge Graphs: A Review

$$sc(r) = \frac{|ex_H(r) \cap ex_B(r)|}{|ex_B(r)|}$$
 and $hc(r) = \frac{|ex_H(r) \cap ex_B(r)|}{|ex_H(r)|}$ (4)

SC is the normalisation of support through the number of entity pairs that satisfy the body, while HC is the normalisation of support through the number of entity pairs that satisfy the head. The higher the values are for these measures, the more plausible the rule is.

These statistical measures have been widely applied to estimate the plausibility of rules learned from KGs [23, 13, 51, 31, 44, 94, 71]. And several variants of these measures have been proposed for better estimations.

It is argued that the standard confidence (SC) is not suitable for KGs that are highly incomplete. Some approaches refine it by introducing *partial completeness assumption* (*PCA*) [24, 23, 37]. PCA assumes the KG has complete information about an entity w.r.t. a relation if the KG contains at least one fact about the entity and relation. For example, if it is mentioned in KG that (Allen, worksFor, airline-NY), then we assume that he only has one job in New York airline. Some other variants along this line include *completeness confidence* [72, 73], which proposes to rely on explicitly incompleteness information to determine an instance as a counterexample, and *soft confidence* [81] which refines the SC with the entity type information. These statistical measures have also been extended to other forms of rules, such as negated rules [21, 75, 31], numerical rules [54, 77], temporal rules [52], and typed rules [82].

2.3.3 Confidence Learning

Traditionally, rule learning can be divided into two main steps: rule structure learning and rule confidence estimation. Recently, some approaches based on neural networks have been developed to perform rule structure and confidence learning simultaneously [89, 77, 62, 14, 87]. These approaches learn the confidence degree of a rule r as a learnable parameter $\alpha_r \in [0, 1]$, which will be updated during the training of the neural networks. This allows the trained neural networks to score any potential rules in the language bias and tightly couples the rule structure learning and confidence learning.

These approaches typically model the learning process through rule-based reasoning tasks, e.g., using rules to infer a triple. Hence, the objective function of the neural network is to assess the plausibility of each possible triple so that those existing in the KG have the highest plausibility. The input of the neural networks includes some latent representations (called embeddings) of the entities and relations in the KG, denoted \mathbf{e} for $e \in \mathcal{E}$ and \mathbf{p} for $p \in \mathcal{P}$. For a triple (s, p, o) and a rule r with its confidence α_r , let $score(\mathbf{s}, \mathbf{p}, \mathbf{o}, r)$ be a scoring function defined on the embeddings and the rule to score the plausibility of the triple. The neural networks are trained to update the parameters α_r that maximize scores for all the triples in the KG \mathcal{G} . This allows the neural networks to simultaneously learn both rule structure r and the rule confidence degrees α_r . By learning the confidence of rules as a parameter, the model can capture the uncertainty and ambiguity in the data, allowing it to assign appropriate confidence to different rules. This adaptability makes the rule-learning process more data-driven and helps the model to make more accurate predictions and better generalisations on unseen data. We will defer the detailed discussions of neural-based structure and parameter learning to Section 5.

In the following sections, we will talk about the three major groups of rule learning methods: ILP-based, statistical path generalisation, and neuro-symbolic methods.

3 ILP-based Methods

The classical ILP methods learn rule structures by systematically exploring the rule space (either in a top-down or a bottom-up manner) through refinement operators, such as adding, updating, or deleting one atom at a time in the rule bodies. However, existing ILP rule learners face at least two challenges when they are applied to KGs. First, the traditional rule search methods have difficulties in scaling to large KGs due to their high computational complexity. A KG can be much larger in size than datasets typically considered in the ILP literature. As a result, classical ILP systems like ALEPH [68] and QuickFOIL [92] cannot be used directly to perform rule learning over medium-sized or large KGs. Moreover, in traditional ILP settings, both positive and negative examples are normally provided for evaluating the rule coverage. However, KGs only contain positive facts, and obtaining negative examples becomes more challenging due to the *Open World Assumption (OWA)*, which assumes that facts missing from the KG are not necessarily false but just unknown.

In recent years, many ILP-based rule learning methods have been proposed in response to these challenges. The AMIE series are notable representatives in this line of research, which introduces novel rule confidence measures to avoid the step of generating negative examples. Another line of research within ILP addresses the lack of negative examples by generating them. They focus on learning rules that can express exceptions or negations, known as nonmonotonic logic programs.

3.1 Without Negative Examples

As traditional ILP methods cannot handle rule learning due to the scale of search space and the lack of negative examples, recent ILP-based rule learners such as AMIE [24, 23, 37] tackle these challenges in rule learning over KGs by employing certain language biases and new confidence measures adapted form association rule mining. AMIE, especially its extension AMIE+ [23] is one of the earliest and most widely referenced KG rule learners.

To reduce the search space, AMIE learns *connected* and *closed* rules. The rule structure learning process of AMIE is a standard ILP top-down search to explore the rule search space. Top-down rule search starts with a general rule and then refines it by progressively adding more atoms. Based on defined refinement operators, such as adding atoms to make the rule closed and connected, it iteratively expands rules. Instead of using standard confidence (SC), AMIE calculates rule confidence under partial completeness assumption (PCA). If the rule satisfies the confidence measure thresholds, the rule is selected as a candidate rule.

AMIE + and AMIE 3 [37] are the extensions of AMIE with a series of improvements and optimisations that allow the system to run over large-scale KGs. Specifically, AMIE + speeds up the rule refinement phase for specific kinds of rules, simplifies the query of support, and approximates the PCA computations by an upper bound; AMIE 3 utilises an in-memory database and parallel computation to store and process large-scale KGs.

Some other rule learning methods use enhanced refinement operators for rule search. For example, Evoda [84] uses a *Genetic Logic Programming* algorithm that is combined with *Evolutionary* Algorithms (EA) to define refinement operators. Hence, three rule transformation operators are proposed, *mutation*, crossover and selection. This allows Evoda to learn rules that are not necessarily closed, e.g., it can learn a rule like $\mathsf{owns}(x_0, x_1) \land \mathsf{capableOfLanding}(x_1, x_2) \rightarrow \mathsf{isAircraftOf}(x_1, x_0)$, which says if x_0 owns x_1 that can land on x_2 , then x_1 may be the aircraft of the airline x_0 . Evoda also adopts the PCA measure for rule confidence estimation.

3.2 Generate Negative Examples

Some other approaches address the lack of negative examples by generating them, this is particularly necessary for learning rules with negations, traditionally known as *nonmonotonic logic programs* [33, 64]. Such rules can express a form of exceptions and support nonmonotonic reasoning.

Exception-Enriched Rule Learning [21] is an ILP-based method that refines learned Horn rules by adding negated atoms (i.e., exceptions) into their rule bodies. They primarily focus on mining rules over unary predicates, by converting binary predicates into multiple unary ones. For example, the binary predicate hasOrigin(\cdot , \cdot) can be translated into several unary ones like hasOriginPEK(\cdot) and hasOriginBOS(\cdot). This is because unary predicates are easier to search for negated atoms. But this makes KG a flattened representation containing just unary facts. To overcome this problem, Nonmonotonic Relational Learning [75] extends [21] to learn exception rules with binary predicates in KGs.

The lack of negative examples is essentially related to the incompleteness and noisiness of KGs, and approaches like RuLES [31] address this through KG embeddings. Compared to AMIE+, RuLES has two more refinement operators to allow negated atoms in a rule body. From Figure 2, we can see the rule evaluation for RuLES is a hybrid combination of statistical and embeddingguided confidence. A weight is used to allow one to choose whether to rely more on the classical measure μ_1 (like standard or PCA confidence), or on the embedding-based measure $\mu_2(\mathcal{G}_r, \phi)$. \mathcal{G}_r extends \mathcal{G} with facts derived from \mathcal{G} by applying rule r. So, $\mu_2(\mathcal{G}_r, \phi)$ capture the information about facts missing in \mathcal{G} that are relevant for r by loss function $\phi(\cdot)$ pre-trained by KG embedding and text corpus models.

4 Statistical Path Generalisation

Statistical Relational Learning (SRL) [26] is a subfield of machine learning that focuses on modelling and learning complex relational data. It combines principles from statistical learning and relational databases to address learning tasks that involve structured data with rich inter-dependencies and uncertainty. There are some SRL approaches to learning first-order rules by using heuristic searching algorithms, for instance, both ProPPR [80, 79] and CoR-PRA [38] are based on the Path Ranking Algorithm (PRA). Several rule learners have emerged to utilise various path-searching strategies based on statistics to directly generate rules. In this section, we focus on sampling-based approaches to mining frequency patterns or path patterns for rule generation in KGs.

4.1 Heuristic Path Sampling

This group of approaches generate candidate rules by directly exploring paths in KGs using heuristic strategies, and these approaches typically focus on learning CP rules generalised from KG paths. As many KGs contain scheme-level (a.k.a., ontological) knowledge such as entity types as well as relation domains and ranges, which is different from instance-level knowledge about relations between individual entities. So, a KG containing ontological knowledge can be split into the ontology graph and the instance (sub)graph.

Given a KG $\mathcal{G} = \mathcal{G}_O \cup \mathcal{G}_I$, \mathcal{G}_O is the ontology graph and \mathcal{G}_I is the instance graph. The instance graph \mathcal{G}_I has entities as vertices and describes instance-level knowledge about entities and their relations; for instance, a triple (airline-NY, hasBase, airport-JFK) describes that the two entities airline-NY and airport-JFK are associated by the relation hasBase. \mathcal{G}_I also describes the classes of entities, such as a triple (airline-NY, rdf:type, Airline) expressing that airline-NY is a member of the class Airline. The ontology graph \mathcal{G}_O , on the other hand, has classes as vertices and describes schema-level (or ontological) knowledge about the relations between classes. For example, triple

H. Wu, Z. Wang, K. Wang, P. G. Omran and J. Li

(Airline, rdfs:subClassOf, Organisation) says that class Airline is a subclass of Organisation. Also, the two triples (hasBase, rdfs:domain, Airline) and (hasBase, rdfs:range, Airport) state that relation hasBase has a domain type Airline and a range type Airport. This can be expressed as an edge (Airline, hasBase, Airport) in the ontology graph.

A path is a sequence of triples (s_1, p_1, o_1) , $(s_2, p_2, o_2), \ldots, (s_n, p_n, o_n)$ in \mathcal{G} where $o_i = s_{i+1}$ $(1 \leq i < n)$. There are two kinds of paths in KGs. A *instance path* (resp., *ontological path*) is a path where s_i, o_i are entities (resp., classes). Existing path exploration methods can thus be classified by whether the paths are from the ontology graph or the instance graph. Generalising instance paths to form rule patterns is also called a *bottom-up* approach, while generating rules from ontological paths is called a *top-down* approach.

4.1.1 Bottom-up Approaches

The AnyBURL series [45, 44, 43] are typical bottom-up approaches, which learn rules by sampling and generalising instance-level paths. To generalise path instances to form rule patterns, AnyBURL substitutes entities from sampled paths in the KG with variables. For example, the paths worksFor(Allen, airline-NY), hasHeadquater(airline-NY, NewYork) and livesIn(Allen, NewYork) can be generalised to a rule worksFor $(x, y) \wedge \mathsf{hasHeadquater}(y, z) \to \mathsf{livesIn}(x, z)$. Unlike the bottom-up approaches in ILP, the generalisation does not add or delete atoms but only substitutes their parameters. This bottom-up approach allows AnyBURL to learn CP rules with constants (i.e., non-substituted entities) in predefined places, e.g., worksFor $(x, y) \land$ hasHeadquater $(y, NewYork) \rightarrow$ lives $\ln(x, \text{NewYork})$. Due to the huge space of semi-grounded CP rules, it is infeasible for AnyBURL to systematically explore the whole search space. Instead, it introduces an anytime algorithm where users can specify the learning time. AnyBURL is much more efficient and effective for rule learning in the sense that it is able to learn more rules in a shorter time than most existing rule learners. For the task of link prediction in KGs, AnyBURL also outperforms many embedding-based methods. In addition, Reinforced AnyBURL[43] is an extension of AnyBURL that introduces reinforcement learning to find more reasonable rules earlier. To achieve this, it incorporates three different reward strategies based on statistical confidence measures to guide the sampling process.

Another example of bottom-up approaches is RuDiK [54]. Different from AnyBURL, it starts from some positive examples and searches the paths connecting the subjects and objects in the positive examples using A^* search. In this approach, the rule generation is guided by assigning a so-called *marginal weight* to each path. In each iteration, the method picks the most promising paths (that is, paths with minimum marginal weight) in the queue of invalid paths. Such paths are expanded and evaluated. If valid, it is added to the output and not further expanded. And the resulting invalid paths will be put back in the queue. This process continues until the queue is empty. RuDiK extends the language bias to allow \perp in the rule heads and literal comparison in the rule bodies. Thus, it can learn constraints as well as rules. To learn rules with literal comparison, it adds comparison edges between pairs of literals of the same type and treats the comparison operators as normal predicates. Unlike AnyBURL using SC for confidence estimation, RuDiK adopts ILP-style example coverage metrics.

4.1.2 Top-down Approaches

Rule learners exploring ontological paths in KGs include Ontological Pathfinding [12], ScaLeKB [13], and RARL [56]. They construct path patterns directly from ontology graphs, and the space of ontological paths is much smaller than that of instance paths. For example, the ontology path hasBirthPlace(Person, City), isRegionOf(City, Country) and hasNationality(Person, Country) can be generalised to a rule hasBirthPlace(x_0, x_1) \land isRegionOf($(x_1, x_2) \rightarrow$ hasNationality(x_0, x_2).

7:12 Rule Learning over Knowledge Graphs: A Review

ScaLeKB extends Ontological Pathfinding and is a typical example of top-down approaches. It achieves its efficiency by storing candidate rules in relational tables according to *structural equivalence*. Two first-order rules are defined to be structurally equivalent if the number of body predicates of the rule and the position of the variable are the same. After that, the rule mining algorithm can process the join queries and partition tables into smaller inputs in parallel, so as to break the mining tasks into smaller independent sub-tasks with Spark.

However, ScaLeKB typically uses type information as hard constraints on the candidate rules, i.e., each entity going through the path must belong to some type. Yet in practice, type information in KGs is often highly incomplete, which would make it too restrictive.

Instead of directly searching ontological paths as hard constraints, RARL [56] samples paths step by step according to the domain and range information of the relations. It reduces the path search space by computing the relatedness of the relations on the paths based on *term frequencyinverse document frequency* (*TF-IDF*) weighting factor, an information retrieval technique adapted to KGs. RARL adopts ILP-style example coverage metrics similar to RuDiK. It also samples a reduced instance graph for fast confidence computation. Moreover, in TyRuLe [82], type information is present in the learned rules and is encoded as latent representations (known as embeddings) to guide the rule search.

4.2 Frequent Pattern Sampling

Some techniques in association rule mining have been adapted to first-order rule learning [81, 4, 5]. Such techniques are originally developed to discover meaningful relationships or associations among items in a dataset and thus, they are modified to generate first-order rules by identifying statistically frequent patterns occurring in KGs.

RDF2Rules [81] samples the so-called frequent predicate cycles (FPCs). A predicate cycle is a sequence of variables and predicates of the form $(x_0, p_1, x_1, ..., p_n, x_0)$, which are essentially generalised instance paths. If a predicate cycle has a sufficient number of instance paths in the KG as its instantiations, it is called a frequent predicate cycle. Rules can be generated from FPCs. For instance, a FPC $(x_0, hasOrigin^{-1}, x_1, hasDestination, x_2, hasAlliance, x_0)$ can generate three rules hasDestination $(x_1, x_2) \land hasAlliance(x_2, x_0) \rightarrow hasOrigin(x_1, x_0), hasOrigin(x_1, x_0) \land$ hasAlliance⁻¹ $(x_0, x_2) \rightarrow hasDestination(x_1, x_2), and hasDestination^{-1}<math>(x_2, x_1) \land hasOrigin(x_1, x_0) \rightarrow$ hasAlliance (x_2, x_0) . RDF2Rules uses a greedy algorithm to iteratively mine FPCs.

SWARM [4, 5] converts triples in KGs into transaction data to apply association rule mining. Association rules capture frequent items in transaction data. As for KG, they convert a triple (s, p, o) into a 2-tuple (s, (p, o)) or $(o, (p^{-1}, s))$. Here both of the s or (p, o) can be seen as items, also the fact tuple (s, (p, o)) can be one transaction where these two items appear at the same time. The frequent transaction items can be generalised to rule patterns. For example, we found two frequent transaction items having a common or mostly overlapping item set like ({Allen, Ally}, (livesln, NewYork)) and ({Allen, Ally}, (diedln, NewYork)). They can be generalised to the association rule {Allen, Ally} : (livesln, NewYork) \rightarrow (diedln, NewYork). Using the type information like rdf:type and rdfs:subClassOf in the ontology graph, they could generate the rule {Person} : (livesln, City) \rightarrow (diedln, City).

5 Neuro-symbolic Methods

In this section, we review rule learning methods that utilise deep neural networks (DNNs). While a rich body of DNN-based methods have been proposed for KG completion, including Graph Neural Networks and embedding-based methods, there is an increasing interest in developing or applying DNN-based methods for rule learning over KGs.

H. Wu, Z. Wang, K. Wang, P. G. Omran and J. Li

There are two major streams of research in this group. Those that learn both rule structures and parameters (i.e., rule plausibility degrees) through DNN models [89, 77, 62, 57, 87, 14, 11], and those that combine DNN models with other rule learning strategies through KG embeddings [88, 51, 94, 82].

5.1 End-to-end Models

In this subsection, we discuss about approaches that use DNNs to learn rules directly by optimising objective functions that roughly correspond to plausible path patterns.

5.1.1 Neural Logic Programming

Neural LP [89] was among the first attempts to combine rule structure learning and confidence learning in an end-to-end differentiable model. It is based on a differentiable probabilistic logic called *TensorLog*, which models CP rule inferences with sparse matrix multiplications. TensorLog maps each entity $e_i \in \mathcal{E}$ to a one-hot vector $\mathbf{v}_i \in \{0, 1\}^{|\mathcal{E}|}$ where only the *i*-th entry is 1, and each relation $p_k \in \mathcal{P}^*$ to a matrix $\mathbf{M}_k \in \{0, 1\}^{|\mathcal{E}| \times |\mathcal{E}|}$ such that its (i, j) entry is 1 if $p(e_i, e_j)$ is a fact in the KG. Then, the application of a rule $p_1(x, z) \wedge p_2(z, y) \to p(x, y)$ on an entity $x = e_i$ can be captured by matrix multiplications $\mathbf{M}_1 \cdot \mathbf{M}_2 \cdot \mathbf{v}_i = \mathbf{s}$. The non-zero entries of the score vector \mathbf{s} represent the instances of y (as entities in the KG) when the rule is applied. Hence, the rule-based inference is captured as:

$$\max_{\{\alpha_r, L_r\}} \sum_{\{x, y\}} score(y|x) = \max_{\{\alpha_r, L_r\}} \sum_{\{x, y\}} \mathbf{v}_y^T (\sum_r (\alpha_r(\prod_{k \in L_r} \mathbf{M}_k \mathbf{v}_x))),$$
(5)

where r indexes over all possible rules, α_r is the confidence associated with rule r and L_r is an ordered list of all predicates in this particular rule. The rule structure along with its confidence α_r are learned by maximising the score. Neural LP thus uses gradient-based programming and optimisation algorithms for the rule learning task. A challenge is a large number of learnable parameters, and Neural LP reduces the learnable parameters by approximating the optimisation objective functions. Another limitation of the above approach is that it is bound to learn rules with a fixed length, and Neural LP addresses this by using an LSTM and attention mechanisms to learn rules of variable lengths.

Neural-Num-LP [77] extends Neural-LP to learn rules with negations and numeric values. It also improves Neural-LP by representing some necessary matrix operations implicitly, including using dynamic programming, cumulative sums operation for numerical comparison features, and low-rank factorisations for negated atoms. Yet it is found that Neural-LP may inevitably learn meaningless rules with high confidence that share atoms with quality rules, and DRUM [62] addresses this issue by utilising bidirectional RNNs to prune the potential incorrect rules.

5.1.2 Decoupling Models

Neural logic programming approaches may still face challenges of exponentially large rule search space, as well as the computational cost of large matrix multiplications. Also, the complexity of simultaneously learning rule structures and confidences makes the optimisation of the computation nontrivial.

To overcome this challenge, new models have been proposed by decoupling the rule structure learning and confidence learning. RNNLogic [57] uses a separate rule generator module for structure learning and a reasoning predictor module for confidence learning. Such a separation allows for more efficient optimisation, by adapting an *Expectation-Maximisation* (*EM*) algorithm [48], which

7:14 Rule Learning over Knowledge Graphs: A Review

enables RNNLogic to handle large KGs effectively. In the same spirit, Ruleformer [87] adopts an encoder-decoder model based on the Transformer architecture. The logic rules are indirectly parsed from the reasoning paths by trained parameters, rather than learned directly. Since Transformer is a sequence-to-sequence model, a converter is used to convert context sub-graph structure into a sequence. A relational attention mechanism is utilised for encoding multi-relational KGs in Transformer. Moreover, RLogic [14] proposes a sequential rule learning algorithm into small atomic models in a recursive way. Given a relation path of rule body $[p_1, \ldots, p_n]$, the relation path encoder first reduces it into a single head p_h by recursively merging relation pairs by a greedy algorithm. Then, the close ratio predictor bridges the gap between "ideal prediction" p_h following logical rules and "real observation" p_t given in KGs. A multi-layer perception (MLP) is used to learn the probability of replacing a relation pair with a single relation and the ratio that a path will close. Finally, Reinforcement Rule Learning [11] formalises the rule generation problem as a sequential decision problem, by adopting reinforcement learning (RL) to generate rules step by step. As shown in Figure 2, the plausibility of generated candidate rules is estimated with a hybrid measure combining explicit statistical confidence and latent embedding measures.

It should be noted that many DNN-based models aim to learn instance paths in KGs to predict missing triples in link prediction. Logical rules are extracted from these instance paths as the side product [86, 74, 16, 85].

5.2 Embedding-based Methods

Representation learning for KGs has attracted intensive interest, which maps entities, relations, and types to low-dimensional vector or matrix spaces, called *embedding* [61, 35], to capture semantic associations between them. Another stream of rule learning approaches use existing or new KG embeddings to learn rule structures and combine them with other rule search strategies and/or rule confidence measures. Utilising embeddings enhances the efficiency of rule learning to allow the handling of large KGs and improves the interpretability of DNN-based rule learning methods.

The authors of [88] were among the first to suggest using KG embeddings extracted from DNNs for rule structure learning, called EMBEDRULE [88]. In EMBEDRULE, the plausibility of rules will be first estimated via an embedding-based *scoring function* before the more expensive computation of PCA confidence. It first embeds entities $e \in \mathcal{E}$ and predicates $p \in \mathcal{P}^*$ as respectively vectors $\mathbf{e} \in \mathbb{R}^d$ and diagonal matrices $\mathbf{P} \in \mathbb{R}^{d \times d}$. Similar to the bilinear transformation of Neural LP, the embeddings satisfy $\mathbf{e}_1^T \cdot \mathbf{P} \cdot \mathbf{e}_2 \approx 1$ for each fact $p(e_1, e_2)$ in the KG; that is, $\mathbf{e}_1^T \cdot \mathbf{P} \approx \mathbf{e}_2^T$. Consider a CP rule r of the form $p_1(x_0, x_1) \wedge p_2(x_1, x_2) \wedge \cdots \wedge p_n(x_{n-1}, x_n) \rightarrow p(x_0, x_n)$, there should be many instance paths that support it, i.e., $p_1(e_0, e_1), p_2(e_1, e_2), \ldots, p_n(e_{n-1}, e_n)$ and $p(e_0, e_n)$ in the KG. Hence, the embeddings satisfy $\mathbf{e}_0^T \cdot \mathbf{P}_1 \approx \mathbf{e}_1^T$, $\mathbf{e}_1^T \cdot \mathbf{P}_2 \approx \mathbf{e}_2^T$, $\ldots, \mathbf{e}_{n-1}^T \cdot \mathbf{P}_n \approx \mathbf{e}_n^T$, and $\mathbf{e}_0^T \cdot \mathbf{P} \approx \mathbf{e}_n^T$; that is, $\mathbf{e}_0^T \cdot \mathbf{P}_1 \cdot \mathbf{P}_2 \cdots \mathbf{P}_n \approx \mathbf{e}_n^T \approx \mathbf{e}_0^T \cdot \mathbf{P}$. Since rule r must hold for many such entities e_0 , the rule can be captured by $\mathbf{P}_1 \cdot \mathbf{P}_2 \cdots \mathbf{P}_n \approx \mathbf{P}$. The scoring function for r is defined via embeddings as follows:

$$score_1(r) = sim(\mathbf{P}_1 \cdot \mathbf{P}_2 \cdots \mathbf{P}_n, \mathbf{P}),$$
(6)

and the similarity between two matrices $sim(\mathbf{M}_1, \mathbf{M}_2)$ can be defined in various ways such as the Frobenius norm, i.e., $sim(\mathbf{M}_1, \mathbf{M}_2) = exp(-||\mathbf{M}_1 - \mathbf{M}_2||_F)$. Based on similar intuitions, different scoring functions via embeddings for rule learning.

RLvLR [51] uses a KG sampling strategy and path embedding methods, which enables it to handle large-scale KGs like DBpedia or Wikidata. The step of sampling based on *n*-hop paths (for rules with maximum length n) can effectively reduce the search space and can handle massive benchmarks efficiently. The sampled smaller KG contains only those entities and facts that are relevant to the target predicate p. Then, RLvLR uses the proposed *co-occurrence scoring function*

H. Wu, Z. Wang, K. Wang, P. G. Omran and J. Li

to guide and prune the search for plausible rules. The selected candidates are kept for the final evaluation by standard confidence and head coverage. The co-occurrence means that for a rule of the form $p_1(x_0, x_1) \wedge p_2(x_1, x_2) \wedge \cdots \wedge p_n(x_{n-1}, x_n) \rightarrow p(x_0, x_n)$, the objects of p_i share many common entities with the subjects of p_{i+1} , where embeddings of the subjects and objects of p_i , denoted \mathbf{p}_i^s and \mathbf{p}_i^s , are defined as the averages of the embeddings of the entities occurring in the corresponding positions. Their co-occurrence scoring function is defined as:

$$score_2(r) = sim(\mathbf{p}_1^s, \mathbf{p}^s) + sim(\mathbf{p}_1^o, \mathbf{p}_2^s) + \dots + sim(\mathbf{p}_{n-1}^o, \mathbf{p}_n^s) + sim(\mathbf{p}_n^o, \mathbf{p}^o).$$
(7)

Later approaches focus on improving the embedding methods to further enhance rule learning performance [83, 94, 53] or learn more expressive forms of rules [52, 82]. In particular, R-Linker [83] improves RLvLR with a hierarchical sampling and lightweight embedding method, and IterE [94] improves the KG embeddings of (especially sparse) KGs through an iterative enhancement process. Rules are learned from embedding with traverse and select strategies, while embedding is refined according to new triples inferred by rules. StreamLearner [52] extends RLvLR to learn temporal rules, and TyRuLe [82] extends the embedding to learn rules with entity type information. Finally, embeddings have been used to transfer rules from one KG to another KG [53].

6 Applications of Rule Learning

Logical rules have been applied in a wide range of scenarios and play a significant role in Explainable AI. In this section, we focus on the applications of first-order Horn rules automatically learned by rule learners.

6.1 Applications in KG Completion

Many existing KGs are large-scale and subject to regular updates. Yet the knowledge contained in them is still far from complete and contains noise. Manual maintenance of large-scale KGs is costly, if not impossible. Hence, automated reasoning for KG completion and verification is essential, including common tasks such as *link prediction* and *fact checking*. While link prediction aims to discover missing links between entities, fact-checking focuses on validating existing triples. Both tasks are important for building KGs and enhancing the quality of existing KGs.

Rule-based Link Prediction. Link prediction is the task of extracting missing triples in a KG and thus it is a subtask of KG completion. Formally, given an entity $s \in \mathcal{E}$ (resp., $o \in \mathcal{E}$) and a relation $p \in \mathcal{P}$ in a KG, the task is to predict entities o (resp., s) such that the triple (s, p, o) is plausibly in the KG.

In recent years, a large number of link prediction models have been proposed. Most major models for link prediction are based on DNNs, especially, embedding models and thus, lack transparency and explainability. Some researchers proposed to develop explainable link prediction models that are based on logic rules [23, 88, 45]. Rules express explicit knowledge and are easy to understand for human beings. For example, given a query of (Allen, hasNationality,?) in Figure 1, the rule hasBirthPlace $(x, y) \land$ isRegionOf $(y, z) \rightarrow$ hasNationality(x, z) can predict hasNationality(Allen, America) and provide the reason that hasBirthPlace(Allen, Boston) \land isRegionOf (Boston, America). Thus, rules provide a promising approach to explainable link prediction. This has become feasible when scalable and effective rule learners are available. On the other hand, link prediction offers an important benchmark for validating and evaluating rule learners for KGs. Experiments show that rule-based link prediction methods possess competitive accuracy and scalability compared to embedding-based ones [44, 56].

7:16 Rule Learning over Knowledge Graphs: A Review

Unlike embedding-based methods that rank all the entities via scoring functions, a rule-based method derives plausible new facts through logical reasoning, and the confidence of the new facts is determined by the rules deriving them. One research direction is how to calculate the confidence of the derived facts from those of the rules. Noisy-OR [23] and Max-Aggregation [44] are two popular methods for obtaining the confidence degree of a plausible triple that is derived from learned rules. Noisy-OR is defined by aggregating the SC of all the rules deriving the triple. The intuition is that facts inferred by more rules should have a higher confidence degree. Max-Aggregation [44] ranks a triple based on the maximum SC of all rules deriving it. It only considers the most confident rule that derives the triple. These two methods for obtaining confidence degrees have their advantages and disadvantages. When aggregating using Noisy-OR, the redundant rules can lead to overestimating the confidence of predicted entities. In order to mitigate this disadvantage, a Non-redundant Noisy-OR [55] is proposed to cluster rules based on their redundancy degree prior to Noisy-OR. Predictions of rules in a cluster are aggregated using the Max-Aggregation, while predictions of the different clusters are then further aggregated using the Noisy-OR.

Rule-guided Embeddings. Rules can also be used to guide the training of existing embedding models so that logic relations are incorporated into the embeddings. In this way, inferences by the generated embeddings are expected to satisfy the rules and are more interpretable, which essentially combines symbolic rule reasoning with neural networks.

In the early literature of rule-guided embeddings, only a small number of manually created hard rules are used in the models [78, 28]. Given the availability of scalable and effective rule learners for KGs, recent efforts focus on combining embedding models and rules learned automatically. RUGE [29] and SoLE [93] use *t-norm fuzzy logics* [30] to incorporate grounded rules, as the t-norm fuzzy logics define how to calculate the probabilities of compositions (e.g., conjunctions and disjunctions) of clauses from the probabilities of the individual clauses. Some other works, such as pLogicNet [58], use *Markov Logic Network (MLN)* [60] to combine the grounded rules with probabilistic graphical models. These approaches need to first compute all the groundings of the rules, which is computationally expensive, especially for long rules. To avoid this, some other approaches incorporate rules as a sequence of relations [17, 27, 50]. Some other models incorporate rules via different learning frameworks, such as RuleGuider [39] and RARL [32], which train reinforcement learning agents guided by rules, and AR-KGAN [95], which incorporates rules via Graph Attention Networks (GAT).

Rule-based Fact Checking. *Fact checking* is the task of verifying the facts in KGs, by predicting the plausibility of the facts. Traditionally, fact checking is done by manual verification which is extremely time-consuming. While black-box models can be used to estimate the plausibility of facts, rule-based fact checking offers explainability. Moreover, rules contain domain-specific knowledge, which is particularly useful for verifying ambiguous facts.

Some rule-based fact checking approaches can generate evidence or explanations for the facts under examination (called target facts) to assess their plausibility [20, 67, 22]. CHEEP [20] generates evidence as paths in the ontology graph according to rules, while ExFaKT [22] uses rules to rewrite each target fact into a set of other easier-to-spot facts as explanations from both text and KGs.

There are also some approaches dedicated to designing new forms of rules for fact checking. Lin et al. [40] propose a kind of graph-fact-checking (GFC) rules to discover a discriminantdirected graph associated with the target facts. OGFC [41] extends GFC with more topological and ontological information to group similar triples. Rules in disjunctive normal form (DNF), i.e., a disjunction of multiple conjunctions, are used in CHAI [9] to filter facts. Some other methods transform rule-based fact-checking problems into answer set program formulations, like EXPCLAIM [1]. It uses first-order rule discovery and Web text mining to gather the evidence to assess target facts, and the fact-checking task is modelled as an inference problem in the answer set programs.

6.2 Other Applications of Rule Learning

There has been a heightened interest in the interpretability of AI models, as understanding how and why a model arrives at a particular decision is pivotal for trust and transparency in decisionmaking. Logic rules, with their inherent ability to support human-comprehensible reasoning processes, have emerged as a valuable tool in diverse fields.

Several domains are actively exploring the use of automatically learned rules to complement and enhance existing AI models. Rules can naturally be applied to other tasks related to KGs, such as entity alignment [10, 36] and knowledge base question answering (KBQA) [70]. In the entity alignment task, RTEA-RA [36] enhances the embeddings of individual entities by injecting the grounded rules into the model to produce hybrid embeddings. MuGNN [10] reconciles the structural differences of two KGs before entity alignment by employing rules induced by AMIE+ for KG completion and pruning. Also, these rules are transferred between KGs based on the knowledge invariant assumption.

As for KBQA, RuKBC-QA [70] uses rule-based knowledge base completion (KBC) in general question answering (QA) systems. Both the origin knowledge base and inferred missing facts by selected rules are used as input of RuKBC-QA for predicting the answers.

Beyond the tasks with KGs as the primary forms of data, several attempts involving rules have been made in the context of natural language processing (NLP), computer vision (CV), and biomedical applications. For NLP, RuleBERT [63] tries to teach the pre-trained language models (PLMs) with the common-sense knowledge provided by Horn rules. KoRC [91] uses rules learned by background KGs to construct the reasoning chain for Reading Comprehension.

Rule learning has also found applications in computer vision, like reasoning in the sub-graph extracted from the images. LOGICDEF [90] constructs a defence model that uses first-order logic rules mined from the extracted scene graph, to explain the object classification and detect the attacks of the adversarial vision model. The document image model [59] uses inductive rules by extracting textual subgraphs corresponding to the text entities in the documents for information extraction.

In the biomedical domain, rule learning has also shown promising applications. In drug discovery [66], generating the explanation paths by Horn rules for drug-disease (entity) pairs. As for drug-gene interaction prediction [2], rules can predict missing links between drug and gene nodes in a graph that contains relevant biomedical knowledge.

These studies collectively showcase the versatility and potential of rule learning across diverse domains, addressing various challenges and improving outcomes in different application areas.

7 Discussion and Conclusion

Rule learning in KGs is a fast-developing research area with promising applications in various fields. In this section, we discuss some future research directions for rule learning and conclude the paper.

7:18 Rule Learning over Knowledge Graphs: A Review

7.1 Future Directions

Despite the rapid advancements in rule learning techniques over KGs, this is still a relatively new research area, with a lot of potential for further development in several aspects, including the complexity and quality of learned rules, evaluation metrics, and rule learning and reasoning strategies.

First of all, the complexity of learned rules can be enhanced by expanding the form of rules (i.e., the language bias) to be learned. Most existing rule learning approaches focus on rules that represent path patterns in KGs. While path patterns capture important structural information in KGs, more complex structural patterns such as sub-graph patterns can reveal more refined knowledge and first-order Horn rules can express complex sub-graph patterns. Scalability of rule learning and identification of the most useful sub-graph patterns remain major challenges in learning more complex rules. Besides, learning rules that can capture more useful information such as attributes or more complex logical connections in KGs is another aspect of rule complexity.

Also, the form of rules and their learning strategies can be tailored according to the applications. For instance, to learn domain-specific rules that are designed to express prior knowledge in specific domains, can potentially lead to more scalable, accurate, and robust performance in the concerned applications. This is especially relevant for applications in knowledge-intensive domains such as biomedical science, healthcare, finance, or legal applications, which involve domain-specific knowledge and specialised applications.

Moreover, to enhance the quality of learned rules, it is crucial to establish effective and robust rule quality evaluation measures. Existing rule confidence measures often lack sufficient granularity or interpretability. Yet designing suitable rule quality metrics has received less attention in the literature. Meanwhile, there has been insufficient emphasis on evaluating the semantic validity of learned rules, that is, to measure how meaningful the learned rules are to human beings. This is particularly important for learning domain-specific rules, and it is desirable to develop quality measures that take into consideration both the data semantics and data distributions.

Finally, new rule learning and reasoning strategies that tightly integrate symbolic AI and deep learning techniques hold great promise. This line of research may involve developing hybrid models that leverage the strengths of both rule-based reasoning and data-driven neural networks. Yet a tighter integration of them has always been a pursuit of the academic communities and the industry. With the recent development of neuro-symbolic approaches, where neural networks are used for symbolic rule learning and reasoning, the other direction is gaining increasing interest, that is, to apply symbolic knowledge, logical constraints, and rule guidance in neural network predictions. This research direction also foresees a tighter coupling of rule learning and prediction models that utilise KGs.

7.2 Conclusion

In this paper, we have provided a systematic review of state-of-the-art rule learning for knowledge graphs. We studied major categories of logic rule learning approaches over knowledge graphs, including the ILP-based, statistical path generalisation, and neuro-symbolic approaches, with discussions on their developments and limitations. Besides, we also discussed the rule confidence measures, which play a crucial role in rule learning. As for the applications of rule learning, we introduced applications of rule-based knowledge graph inferences, as well as wider applications in natural language processing, computer vision, and biomedical science. Finally, we pointed out several promising future directions for rule learning research.

— References -

- Naser Ahmadi, Joohyung Lee, Paolo Papotti, and Mohammed Saeed. Explainable fact checking with probabilistic answer set programming. In Proceedings of the 2019 Truth and Trust Online Conference (TTO-19), London, UK, October 4-5, 2019, 2019. URL: https://truthandtrustonline.com/ wp-content/uploads/2019/09/paper_15.pdf.
- 2 Fotis Aisopos and Georgios Paliouras. Comparing methods for drug-gene interaction prediction on the biomedical literature knowledge graph: performance versus explainability. BMC Bioinform., 24(1):272, 2023. doi:10.1186/ S12859-023-05373-2.
- 3 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. Dbpedia: A nucleus for a web of open data. In Proceedings of the 6th International Semantic Web Conference (ISWC-07), Busan, Korea, November 11-15, 2007, volume 4825, pages 722–735. Springer, 2007. doi:10.1007/978-3-540-76298-0_52.
- 4 Molood Barati, Quan Bai, and Qing Liu. SWARM: an approach for mining semantic association rules from semantic web data. In Proceedings of the 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI-16), Phuket, Thailand, August 22-26, 2016, volume 9810, pages 30–43. Springer, 2016. doi:10.1007/ 978-3-319-42911-3_3.
- 5 Molood Barati, Quan Bai, and Qing Liu. Mining semantic association rules from RDF data. *Knowl. Based Syst.*, 133:183–196, 2017. doi: 10.1016/J.KNOSYS.2017.07.009.
- 6 Luigi Bellomarini, Davide Benedetto, Georg Gottlob, and Emanuel Sallinger. Vadalog: A modern architecture for automated reasoning with large knowledge graphs. *Inf. Syst.*, 105:101528, 2022. doi:10.1016/J.IS.2020.101528.
- 7 Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. Scientific American, 284(5):34-43, 2001. doi:10.1038/ scientificamerican0501-3.
- 8 Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD-08), Vancouver, BC, Canada, June 10-12, 2008, pages 1247–1250. ACM, 2008. doi:10.1145/1376616.1376746.
- 9 Agustín Borrego, Daniel Ayala, Inma Hernández, Carlos R. Rivero, and David Ruiz. Generating rules to filter candidate triples for their correctness checking by knowledge graph completion techniques. In Proceedings of the 10th International Conference on Knowledge Capture (K-CAP-19), Marina Del Rey, CA, USA, November 19-21, 2019, pages 115–122. ACM, 2019. doi: 10.1145/3360901.3364418.
- 10 Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. Multi-channel graph neural network for entity alignment. In Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL-19),

Florence, Italy, July 28- August 2, pages 1452– 1461. Association for Computational Linguistics, 2019. doi:10.18653/V1/P19-1140.

- 11 Lihan Chen, Sihang Jiang, Jingping Liu, Chao Wang, Sheng Zhang, Chenhao Xie, Jiaqing Liang, Yanghua Xiao, and Rui Song. Rule mining over knowledge graphs via reinforcement learning. *Knowl. Based Syst.*, 242:108371, 2022. doi: 10.1016/J.KNOSYS.2022.108371.
- 12 Yang Chen, Sean Goldberg, Daisy Zhe Wang, and Soumitra Siddharth Johri. Ontological pathfinding. In Proceedings of the 2016 International Conference on Management of Data Conference (SIGMOD-16), San Francisco, CA, USA, June 26 -July 01, 2016, pages 835–846. ACM, 2016. doi:10.1145/2882903.2882954.
- 13 Yang Chen, Daisy Zhe Wang, and Sean Goldberg. Scalekb: scalable learning and inference over large knowledge bases. VLDB J., 25(6):893–918, 2016. doi:10.1007/S00778-016-0444-3.
- 14 Kewei Cheng, Jiahao Liu, Wei Wang, and Yizhou Sun. Rlogic: Recursive logical rule learning from knowledge graphs. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-22), Washington, DC, USA, August 14 18, 2022, pages 179–189. ACM, 2022. doi:10.1145/3534678.3539421.
- 15 Andrew Cropper, Sebastijan Dumancic, Richard Evans, and Stephen H. Muggleton. Inductive logic programming at 30. Mach. Learn., 111(1):147–172, 2022. doi:10.1007/S10994-021-06089-1.
- 16 David Jaime Tena Cucala, Bernardo Cuenca Grau, Egor V. Kostylev, and Boris Motik. Explainable gnn-based models over knowledge graphs. In Proceedings of the 10th International Conference on Learning Representations (ICLR-22), Virtual Event, April 25-29, 2022. OpenReview.net, 2022. URL: https://openreview.net/forum?id= CrCvGNHAIrz.
- 17 Boyang Ding, Quan Wang, Bin Wang, and Li Guo. Improving knowledge graph embedding using simple constraints. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL-18), Melbourne, Australia, July 15-20, 2018, pages 110–121. Association for Computational Linguistics, 2018. doi: 10.18653/V1/P18-1011.
- 18 Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-14), New York, USA, August 24-27, 2014, pages 601–610. ACM, 2014. doi:10.1145/2623330.2623623.
- 19 Facebook. Graph api facebook for developers. https://developers.facebook.com/ docs/graph-api/, 2022.
- 20 Valeria Fionda and Giuseppe Pirrò. Fact checking via evidence patterns. In *Proceedings of the* 27th International Joint Conference on Artificial Intelligence (IJCAI-18), July 13-19, 2018, Stock-

holm, Sweden, pages 3755–3761. ijcai.org, 2018. doi:10.24963/IJCAI.2018/522.

- 21 Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. Exceptionenriched rule learning from knowledge graphs. In Proceedings of the 15th International Semantic Web Conference (ISWC-16), Kobe, Japan, October 17-21, 2016, pages 234–251, 2016. doi: 10.1007/978-3-319-46523-4_15.
- 22 Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. Exfakt: A framework for explaining facts over knowledge graphs and text. In Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM-19), Melbourne, VIC, Australia, February 11-15, 2019, pages 87–95. ACM, 2019. doi:10.1145/3289600.3290996.
- 23 Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. VLDB J., 24(6):707–730, 2015. doi:10.1007/ S00778-015-0394-1.
- 24 Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In Proceedings of the 22nd International World Wide Web Conference (WWW-13), Rio de Janeiro, Brazil, May 13-17, 2013, pages 413–422. International World Wide Web Conferences Steering Committee / ACM, 2013. doi:10.1145/2488388.2488425.
- 25 Matt Gardner, Partha P. Talukdar, and Tom Michael Mitchell. Combining vector space embeddings with symbolic logical inference over open-domain text. In Proceedings of the 2015 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 22-25, 2015. AAAI Press, 2015. URL: http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10230.
- 26 Lise Getoor and Ben Taskar. Introduction to statistical relational learning. MIT press, 2007.
- 27 Shu Guo, Lin Li, Zhen Hui, Lingshuai Meng, Bingnan Ma, Wei Liu, Lihong Wang, Haibin Zhai, and Hong Zhang. Knowledge graph embedding preserving soft logical regularity. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM-20), Virtual Event, Ireland, October 19-23, 2020, pages 425–434. ACM, 2020. doi:10.1145/3340531.3412055.
- 28 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP-16), Austin, Texas, USA, November 1-4, 2016, pages 192–202. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1019.
- 29 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 4816–4823. AAAI Press, 2018. doi:10.1609/AAAI.V32I1.11918.

- 30 Petr Hájek. Metamathematics of Fuzzy Logic, volume 4. Kluwer, 1998. doi:10.1007/ 978-94-011-5300-3.
- 31 Vinh Thinh Ho, Daria Stepanova, Mohamed H. Gad-Elrab, Evgeny Kharlamov, and Gerhard Weikum. Rule learning from knowledge graphs guided by embedding models. In *Proceedings of the 17th International Semantic Web Conference (ISWC-18), Monterey, CA, USA, October 8-12, 2018*, volume 11136, pages 72–90. Springer, 2018. doi:10.1007/978-3-030-00671-6_5.
- 32 Zhongni Hou, Xiaolong Jin, Zixuan Li, and Long Bai. Rule-aware reinforcement learning for knowledge graph reasoning. In Findings of the Association for Computational Linguistics (ACL/IJCNLP-21), Online Event, August 1-6, 2021, volume ACL/IJCNLP 2021, pages 4687– 4692. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.FINDINGS-ACL.412.
- 33 Katsumi Inoue and Yoshimitsu Kudoh. Learning extended logic programs. In Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97), Nagoya, Japan, August 23-29, 1997, 2 Volumes, pages 176–181. Morgan Kaufmann, 1997. URL: http://ijcai. org/Proceedings/97-1/Papers/029.pdf.
- 34 Lei Ji. Knowledge mining api microsoft research. https://www.microsoft.com/en-us/research/ project/knowledge-mining-api/, 2016.
- 35 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Networks Learn. Syst.*, 33(2):494–514, 2022. doi:10.1109/TNNLS. 2021.3070843.
- 36 Tingting Jiang, Chenyang Bu, Yi Zhu, and Xindong Wu. Two-stage entity alignment: Combining hybrid knowledge graph embedding with similarity-based relation alignment. In Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI-19), Cuvu, Yanuca Island, Fiji, August 26-30, 2019, volume 11670, pages 162–175. Springer, 2019. doi:10.1007/978-3-030-29908-8_13.
- 37 Jonathan Lajus, Luis Galárraga, and Fabian M. Suchanek. Fast and exact rule mining with AMIE 3. In Proceedings of the 17th Extended Semantic Web Conference (ESWC-20), Heraklion, Crete, Greece, May 31-June 4, 2020, volume 12123, pages 36–52. Springer, 2020. doi:10.1007/ 978-3-030-49461-2_3.
- 38 Ni Lao, Einat Minkov, and William W. Cohen. Learning relational features with backward random walks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL-15), July 26-31, 2015, Beijing, China, pages 666–675. The Association for Computer Linguistics, 2015. doi:10.3115/V1/P15-1065.
- 39 Deren Lei, Gangrong Jiang, Xiaotao Gu, Kexuan Sun, Yuning Mao, and Xiang Ren. Learning collaborative agents with rule guidance for knowledge graph reasoning. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP-20), Online, November 16-20, 2020, pages 8541–8547. Associ-

ation for Computational Linguistics, 2020. doi: 10.18653/V1/2020.EMNLP-MAIN.688.

- 40 Peng Lin, Qi Song, Jialiang Shen, and Yinghui Wu. Discovering graph patterns for fact checking in knowledge graphs. In Proceedings of the 23rd International Conference on Database Systems for Advanced Applications (DASFAA-18), Gold Coast, QLD, Australia, May 21-24, 2018, volume 10827, pages 783-801. Springer, 2018. doi:10.1007/978-3-319-91452-7_50.
- 41 Peng Lin, Qi Song, and Yinghui Wu. Fact checking in knowledge graphs with ontological subgraph patterns. *Data Sci. Eng.*, 3(4):341–358, 2018. doi:10.1007/S41019-018-0082-4.
- 42 Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22), Virtual Event, February 22
 March 1, 2022, pages 4120–4127. AAAI Press, 2022. doi:10.1609/AAAI.V36I4.20330.
- 43 Christian Meilicke, Melisachew Wudage Chekol, Manuel Fink, and Heiner Stuckenschmidt. Reinforced anytime bottom up rule learning for knowledge graph completion. *CoRR*, abs/2004.04412, 2020. doi:10.48550/arXiv.2004.04412.
- 44 Christian Meilicke, Melisachew Wudage Chekol, Daniel Ruffinelli, and Heiner Stuckenschmidt. Anytime bottom-up rule learning for knowledge graph completion. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, China, August 10-16, 2019, pages 3137–3143. ijcai.org, 2019. doi: 10.24963/IJCAI.2019/435.
- 45 Christian Meilicke, Manuel Fink, Yanjie Wang, Daniel Ruffinelli, Rainer Gemulla, and Heiner Stuckenschmidt. Fine-grained evaluation of ruleand embedding-based systems for knowledge graph completion. In Proceedings of the 17th International Semantic Web Conference (ISWC-18), Monterey, CA, USA, October 8-12, 2018, volume 11136, pages 3–20. Springer, 2018. doi:10.1007/ 978-3-030-00671-6_1.
- 46 George A. Miller. Wordnet: A lexical database for english. Commun. ACM, 38(11):39–41, 1995. doi:10.1145/219717.219748.
- 47 Stephen H. Muggleton. Inductive logic programming. In Proceedings of the 1st International Workshop on Algorithmic Learning Theory (ALT-90), Tokyo, Japan, October 8-10, 1990, pages 42–62. Springer/Ohmsha, 1990.
- 48 Radford M. Neal and Geoffrey E. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, volume 89, pages 355–368. Springer Netherlands, 1998. doi:10.1007/978-94-011-5014-9_ 12.
- 49 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *IEEE*, 104(no. 1):11–33, 2016. doi:10.1109/JPR0C.2015. 2483592.
- 50 Guanglin Niu, Yongfei Zhang, Bo Li, Peng Cui, Si Liu, Jingyang Li, and Xiaowei Zhang.

Rule-guided compositional representation learning on knowledge graphs. In *Proceedings of the* 34th AAAI Conference on Artificial Intelligence (AAAI-20), New York, NY, USA, February 7-12, 2020, pages 2950–2958. AAAI Press, 2020. doi:10.1609/AAAI.V34I03.5687.

- 51 Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. Scalable rule learning via learning representation. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18), July 13-19, 2018, Stockholm, Sweden, pages 2149–2155. ijcai.org, 2018. doi: 10.24963/IJCAI.2018/297.
- 52 Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. Learning temporal rules from knowledge graph streams. In Proceedings of the AAAI 2019 Spring Symposium on Combining Machine Learning with Knowledge Engineering (AAAI-MAKE-19), Stanford University, Palo Alto, California, USA, March 25-27, 2019, volume 2350. CEUR-WS.org, 2019. URL: https://ceur-ws. org/Vol-2350/paper15.pdf.
- 53 Pouya Ghiasnezhad Omran, Zhe Wang, and Kewen Wang. Knowledge graph rule mining via transfer learning. In Proceedings of the 23th Pacific-Asia Knowledge Discovery and Data Mining Conference (PAKDD-19), Macau, China, April 14-17, 2019, volume 11441, pages 489–500. Springer, 2019. doi:10.1007/978-3-030-16142-2_38.
- 54 Stefano Ortona, Venkata Vamsikrishna Meduri, and Paolo Papotti. Rudik: Rule discovery in knowledge bases. *PVLDB*, 11(12):1946–1949, 2018. doi:10.14778/3229863.3236231.
- 55 Simon Ott, Christian Meilicke, and Matthias Samwald. SAFRAN: an interpretable, rule-based link prediction method outperforming embedding models. In Proceedings of the 3rd Conference on Automated Knowledge Base Construction (AKBC-21), Virtual, October 4-8, 2021, 2021. doi:10.24432/ C5MK57.
- 56 Giuseppe Pirrò. Relatedness and tbox-driven rule learning in large knowledge bases. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, (AAAI-20), New York, NY, USA, February 7-12, 2020, pages 2975–2982. AAAI Press, 2020. doi:10.1609/AAAI.V34I03.5690.
- 57 Meng Qu, Junkun Chen, Louis-Pascal A. C. Xhonneux, Yoshua Bengio, and Jian Tang. Rnnlogic: Learning logic rules for reasoning on knowledge graphs. In *Proceedings of the 9th International Conference on Learning Representations (ICLR-21), Virtual Event, Austria, May 3-7, 2021.* Open-Review.net, 2021. URL: https://openreview. net/forum?id=tGZu6DlbreV.
- 58 Meng Qu and Jian Tang. Probabilistic logic neural networks for reasoning. In Proceedings of the 33rd Annual Conference on Neural Information Processing Systems 2019 (NeurIPS-19), December 8-14, 2019, Vancouver, BC, Canada, pages 7710– 7720, 2019. doi:10.48550/arXiv.1906.08495.
- 59 Mouli Rastogi, Syed Afshan Ali, Mrinal Rawat, Lovekesh Vig, Puneet Agarwal, Gautam Shroff, and Ashwin Srinivasan. Information extraction from document images via FCA based template detection and knowledge graph rule induction.

In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW-20), Seattle, WA, USA, June 14-19, 2020, pages 2377-2385. Computer Vision Foundation / IEEE, 2020. doi:10.1109/ CVPRW50498.2020.00287.

- 60 Matthew Richardson and Pedro M. Domingos. Markov logic networks. Mach. Learn., 62(1-2):107– 136, 2006. doi:10.1007/S10994-006-5833-1.
- 61 Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. Knowledge graph embedding for link prediction: A comparative analysis. ACM Trans. Knowl. Discov. Data, 15(2):14:1–14:49, 2021. doi:10.1145/ 3424672.
- 62 Ali Sadeghian, Mohammadreza Armandpour, Patrick Ding, and Daisy Zhe Wang. DRUM: end-to-end differentiable rule mining on knowledge graphs. In Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS-19), December 8-14, 2019, Vancouver, BC, Canada, pages 15321-15331, 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/0c72cb7ee1512f800abe27823a792d03-Abstract.html.
- 63 Mohammed Saeed, Naser Ahmadi, Preslav Nakov, and Paolo Papotti. Rulebert: Teaching soft rules to pre-trained language models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP-21), Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021, pages 1460–1476. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.EMNLP-MAIN.110.
- 64 Chiaki Sakama. Induction from answer sets in nonmonotonic logic programs. ACM Trans. Comput. Log., 6(2):203–231, 2005. doi:10.1145/1055686. 1055687.
- 65 Abdus Salam, Rolf Schwitter, and Mehmet A. Orgun. Probabilistic rule learning systems: A survey. ACM Comput. Surv., 54(4):79:1–79:16, 2022. doi:10.1145/3447581.
- 66 Kara Schatz, Cleber C. Melo-Filho, Alexander Tropsha, and Rada Chirkova. Explaining drugdiscovery hypotheses using knowledge-graph patterns. In Proceedings of the IEEE International Conference on Big Data, Orlando, FL, USA, December 15-18, 2021, pages 3709–3716. IEEE, 2021. doi:10.1109/BIGDATA52589.2021.9672006.
- 67 Baoxu Shi and Tim Weninger. Discriminative predicate path mining for fact checking in knowledge graphs. *Knowl. Based Syst.*, 104:123–133, 2016. doi:10.1016/J.KNOSYS.2016.04.015.
- 68 Ashwin Srinivasan. The aleph manual, 2007. URL: https://www.cs.ox.ac.uk/activities/ programinduction/Aleph/aleph_toc.html.
- 69 Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web (WWW-07), Banff, Alberta, Canada, May 8-12, 2007, pages 697–706. ACM, 2007. doi:10.1145/1242572.1242667.
- 70 Qilin Sun and Weizhuo Li. Rukbc-qa: A framework for question answering over incomplete kbs enhanced with rules injection. In

Proceedings of the 9th Natural Language Processing and Chinese Computing (NLPCC-20), Zhengzhou, China, October 14-18, 2020, volume 12431, pages 82–94. Springer, 2020. doi:10.1007/ 978-3-030-60457-8_7.

- 71 Martin Svatos, Steven Schockaert, Jesse Davis, and Ondrej Kuzelka. Strike: Rule-driven relational learning using stratified k-entailment. In Proceedings of the 24th European Conference on Artificial Intelligence (ECAI-20), 29 August-8 September 2020, Santiago de Compostela, Spain, volume 325, pages 1515–1522. IOS Press, 2020. doi:10.3233/FAIA200259.
- 72 Thomas Pellissier Tanon, Daria Stepanova, Simon Razniewski, Paramita Mirza, and Gerhard Weikum. Completeness-aware rule learning from knowledge graphs. In Proceedings of the 16th International Semantic Web Conference (ISWC-17), Vienna, Austria, October 21-25, 2017, volume 10587, pages 507–525. Springer, 2017. doi:10. 1007/978-3-319-68288-4_30.
- 73 Thomas Pellissier Tanon, Daria Stepanova, Simon Razniewski, Paramita Mirza, and Gerhard Weikum. Completeness-aware rule learning from knowledge graphs. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18), July 13-19, 2018, Stockholm, Sweden, pages 5339–5343. ijcai.org, 2018. doi:10.24963/IJCAI.2018/749.
- 74 Komal K. Teru, Etienne G. Denis, and William L. Hamilton. Inductive relation prediction by subgraph reasoning. In *Proceedings of the 37th International Conference on Machine Learning (ICML-*20), 13-18 July 2020, Virtual Event, volume 119, pages 9448–9457. PMLR, 2020. URL: http: //proceedings.mlr.press/v119/teru20a.html.
- 75 Hai Dang Tran, Daria Stepanova, Mohamed H. Gad-Elrab, Francesca A. Lisi, and Gerhard Weikum. Towards nonmonotonic relational learning from knowledge graphs. In *Proceedings of the* 26th International Conference on Inductive Logic Programming (ILP-16), London, UK, September 4-6, 2016, volume 10326, pages 94–107. Springer, 2016. doi:10.1007/978-3-319-63342-8_8.
- 76 Denny Vrandecic and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. Commun. ACM, 57(10):78-85, 2014. doi:10.1145/2629489.
- 77 Po-Wei Wang, Daria Stepanova, Csaba Domokos, and J. Zico Kolter. Differentiable learning of numerical rules in knowledge graphs. In Proceedings of the 8th International Conference on Learning Representations (ICLR-20), Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020. URL: https://openreview.net/forum?id= rJleKgrKwS.
- 78 Quan Wang, Bin Wang, and Li Guo. Knowledge base completion using embeddings and rules. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI-15), Buenos Aires, Argentina, July 25-31, 2015, pages 1859– 1866. AAAI Press, 2015. URL: http://ijcai.org/ Abstract/15/264.
- 79 William Yang Wang and William W. Cohen. Joint information extraction and reasoning: A scalable statistical relational learning approach. In Proceedings of the 53rd Annual Meeting of the Asso-

ciation for Computational Linguistics (ACL-15), July 26-31, 2015, Beijing, China, pages 355–364. The Association for Computer Linguistics, 2015. doi:10.3115/V1/P15-1035.

- 80 William Yang Wang, Kathryn Mazaitis, and William W. Cohen. Structure learning via parameter learning. In Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM-14), Shanghai, China, November 3-7, 2014, pages 1199–1208. ACM, 2014. doi:10.1145/2661829.2662022.
- 81 Zhichun Wang and Juan-Zi Li. Rdf2rules: Learning rules from RDF knowledge bases by mining frequent predicate cycles. *CoRR*, abs/1512.07734, 2015. doi:10.48550/arXiv.1512.07734.
- 82 Hong Wu, Zhe Wang, Kewen Wang, and Yi-Dong Shen. Learning typed rules over knowledge graphs. In Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR-22), Haifa, Israel, July 31-August 5, 2022, 2022. doi:10.24963/kr.2022/51.
- 83 Hong Wu, Zhe Wang, Xiaowang Zhang, Pouya Ghiasnezhad Omran, Zhiyong Feng, and Kewen Wang. A system for reasoning-based link prediction in large knowledge graphs. In Proceedings of the 18th International Semantic Web Conference Satellite Tracks (Posters & Demonstrations, Industry, and Outrageous Ideas) (ISWC-19), Auckland, New Zealand, October 26-30, 2019, volume 2456, pages 121-124. CEUR-WS.org, 2019. URL: https:// ceur-ws.org/Vol-2456/paper32.pdf.
- 84 Lianlong Wu, Emanuel Sallinger, Evgeny Sherkhonov, Sahar Vahdati, and Georg Gottlob. Rule learning over knowledge graphs with genetic logic programming. In Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE-22), Kuala Lumpur, Malaysia, May 9-12, 2022, pages 3373–3385. IEEE, 2022. doi:10.1109/ICDE53745.2022.00318.
- 85 Yi Xia, Junyong Luo, Mingjing Lan, Gang Zhou, Zhibo Li, and Shuo Liu. Reason more like human: Incorporating meta information into hierarchical reinforcement learning for knowledge graph reasoning. *Appl. Intell.*, 53(11):13293–13308, 2023. doi:10.1007/S10489-022-04147-2.
- 86 Wenhan Xiong, Thien Hoang, and William Yang Wang. Deeppath: A reinforcement learning method for knowledge graph reasoning. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP-17), Copenhagen, Denmark, September 9-11, 2017, pages 564-573. Association for Computational Linguistics, 2017. doi:10.18653/V1/D17-1060.
- 87 Zezhong Xu, Peng Ye, Hui Chen, Meng Zhao, Huajun Chen, and Wen Zhang. Ruleformer: Context-aware rule mining over knowledge graph. In Proceedings of the 29th International Conference on Computational Linguistics (COLING-

22), Gyeongju, Republic of Korea, October 12-17, 2022, pages 2551-2560. International Committee on Computational Linguistics, 2022. URL: https://aclanthology.org/2022.coling-1.225.

- 88 Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In Proceedings of the 3rd International Conference on Learning Representations (ICLR-2015), San Diego, CA, USA, May 7-9, 2015, 2015. URL: http://arxiv.org/abs/1412.6575.
- 89 Fan Yang, Zhilin Yang, and William W. Cohen. Differentiable learning of logical rules for knowledge base reasoning. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NeurIPS-17), December 4-9, 2017, Long Beach, CA, USA, pages 2319-2328, 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/0e55666a4ad822e0e34299df3591d979-Abstract.html.
- 90 Yuan Yang, James Clayton Kerce, and Faramarz Fekri. LOGICDEF: an interpretable defense framework against adversarial examples via inductive scene graph reasoning. In Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22), Virtual Event, February 22 - March 1, 2022, pages 8840–8848. AAAI Press, 2022. doi:10.1609/AAAI.V36I8.20865.
- 91 Zijun Yao, Yantao Liu, Xin Lv, Shulin Cao, Jifan Yu, Juanzi Li, and Lei Hou. Korc: Knowledge oriented reading comprehension benchmark for deep text understanding. In *Findings of the Association for Computational Linguistics (ACL-23), Toronto, Canada, July 9-14, 2023*, pages 11689–11707. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.FINDINGS-ACL.743.
- 92 Qiang Zeng, Jignesh M. Patel, and David Page. Quickfoil: Scalable inductive logic programming. Proc. VLDB Endow., 8(3):197–208, 2014. doi: 10.14778/2735508.2735510.
- 93 Jindou Zhang and Jing Li. Enhanced knowledge graph embedding by jointly learning soft rules and facts. *Algorithms*, 12(12):265, 2019. doi:10.3390/A12120265.
- 94 Wen Zhang, Bibek Paudel, Liang Wang, Jiaoyan Chen, Hai Zhu, Wei Zhang, Abraham Bernstein, and Huajun Chen. Iteratively learning embeddings and rules for knowledge graph reasoning. In Proceedings of the World Wide Web Conference (WWW-19), San Francisco, CA, USA, May 13-17, 2019, pages 2366-2377. ACM, 2019. doi:10.1145/3308558.3313612.
- 95 Zhenghao Zhang, Jianbin Huang, and Qinglin Tan. Association rules enhanced knowledge graph attention network. *Knowl. Based Syst.*, 239:108038, 2022. doi:10.1016/J.KNOSYS.2021.108038.

Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

Claudia d'Amato 🖂 🏠 💿 University of Bari, Italy

Louis Mahon 🖂 💿 Oxford University, United Kingdom

Pierre Monnin 🖂 🏠 💿 Université Côte d'Azur, Inria, CNRS, I3S, France

Giorgos Stamou 🖂 🗈 National Technical University of Athens, Greece

— Abstract -

The graph model is nowadays largely adopted to model a wide range of knowledge and data, spanning from social networks to knowledge graphs (KGs), representing a successful paradigm of how symbolic and transparent AI can scale on the World Wide Web. However, due to their unprecedented volume, they are generally tackled by Machine Learning (ML) and mostly numeric based methods such as graph embedding models (KGE) and deep neural networks (DNNs). The latter methods have been proved lately very efficient, leading the current AI spring. In this vision paper, we introduce some of the main existing methods for combining KGs

and ML, divided into two categories: those using ML to improve KGs, and those using KGs to improve results on ML tasks. From this introduction, we highlight research gaps and perspectives that we deem promising and currently under-explored for the involved research communities, spanning from KG support for LLM prompting, integration of KG semantics in ML models to symbol-based methods, interpretability of ML models, and the need for improved benchmark datasets. In our opinion, such perspectives are stepping stones in an ultimate view of KGs as central assets for neuro-symbolic and explainable AI.

2012 ACM Subject Classification Information systems \rightarrow World Wide Web; Computing methodologies \rightarrow Artificial intelligence

Keywords and phrases Graph-based Learning, Knowledge Graph Embeddings, Large Language Models, Explainable AI, Knowledge Graph Completion & Curation

Digital Object Identifier 10.4230/TGDK.1.1.8

Category Vision

Funding Claudia d'Amato: Partially supported by the project FAIR – Future AI Research (PE00000013), spoke 6 - Symbiotic AI (https://future-ai-research.it/), under the NRRP MUR program funded by the NextGenerationEU and by the project HypeKG – Hybrid Prediction and Explanation with Knowledge Graphs (H53D23003700006), under PRIN 2022 program funded by MUR.

Pierre Monnin: Supported by the AT2TA project (https://at2ta.loria.fr), funded by the French National Research Agency ("Agence Nationale de la Recherche" – ANR) under grant ANR-22-CE23-0023.

Received 2023-06-30 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge



© Claudia d'Amato, Louis Mahon, Pierre Monnin, and Giorgos Stamou;

licensed under Creative Commons License CC-BY 4.0 Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 8, pp. 8:1-8:35

Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

8:2 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

1 Introduction

Graph data refers to data that lends itself naturally to being represented as a graph-based data model. Examples of graph data are social networks, computer networks, entailment graphs [93], concept graphs [26]. Several standards have been proposed to represent graph data, including the W3C devised standards OWL, RDF, and RDFS. These enable easy sharing and combining of graph data from different sources, and so further facilitate the adoption of the graph formalism.

Among the several types of graph data in widespread use, one prominent example is the Knowledge Graph (KG). A KG aims to convey knowledge of the real world and represent it conforming to a graph-based data model, where nodes represent concepts of interest, such as human or lion, and edges represent possibly different relations between these entities, such as isTypeOf or isPredatorOf [68]. A closely related concept that we do not discuss any further is Property Graph, where both nodes and edges can have multiple properties which are represented as key-value pairs (the interested reader may refer to [68] for further details). Graphs data may be stored in native graph databases or relational databases [68].

When referring to the representation of information, the term 'knowledge', as opposed to 'data', is usually what is predicated of humans. It suggests the information is stored in a more structured and actionable manner, e.g. that it enables reasoning. This distinction from 'data' was first made in relation to the concept of a knowledge base (KB) [104], in the context of expert systems [64], in order to distinguish them from databases using, e.g., lookup tables or hash tables. A KB is a representation of information as a set of facts or sentences [161].

A KG can be formalized as a triple of sets $\langle E, R, T \rangle$, where E is a set of entities, R a set of relations, and T is of the form $\{(s, p, o) \mid s, o \in E, p \in R\}$ [29], by which it is immediately equivalent to a KB, considered as a set of facts. Moreover, a graph $G = (\mathcal{N}, \mathcal{E})$, can be written equivalently as a set of facts, by equating \mathcal{N} with the set of all entities appearing as arguments to facts, and equating each fact $\langle s, p, o \rangle$ to a directed edge from s (*subject*) to o (*object*) labeled p (*predicate*). On a higher level, one difference between a KG and KB as a set of facts, is that the former has a greater emphasis on the connection to the graph-based data model, and is more directly associated with the agreed formatting standards for graph data. Our discussion here does not require precise disambiguation of the term and in the remainder of this paper, we use the two terms interchangeably.

A closely related concept to a KG is an ontology. Intuitively, an ontology defines a set of object types, and how these types relate to each other. For example, if the domain is living things, then an ontology would specify that human and lion are two distinct types of a mammal, mammal and reptile are two distinct types of vertebrates etc. Formally, an ontology has been defined as comprising two components, the TBox, which introduces the vocabulary of an application domain, and the ABox, which contains assertions about named individuals in terms of this vocabulary [11]. Often the set of concepts in a KG forms an ontology, and their ontological relations can be incorporated into the structure of the KG. In the remainder of this paper, we will treat the term "ontology" as interchangeable with "knowledge base", as defined above.

Many important applications, such as e-Commerce [208], financial trading [29], semantic search [205], fact-checking [165], recommendation [195], (medical) decision support systems [202], question answering [73] and even machine translation [221, 135] benefit from access to real-world knowledge in a form that is both machine-readable and human-interpretable (i.e. entities, properties, relations and types). There has thus been a general convergence on KGs as the means to represent and store such knowledge. This interest from academia and especially from industry, has led to several large-scale efforts at constructing KGs. Some are freely available and accessible,

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

such as DBpedia [9]¹, Freebase [18]², YAGO [172]³, Wikidata [187]⁴. Others are private, developed for commercial use by companies such as Google, Amazon, IKEA, Uber, Microsoft, Facebook and LinkedIn. The interested reader could refer to [69] for a comprehensive overview of the history and current use of KGs.

The amount of data that may be of interest to KG applications is very large, e.g., Englishlanguage Wikipedia contains close to seven million articles at the time of writing⁵. Developing KGs of this size is a difficult, expensive process, requiring the integration of multiple sources of information, along with input from human experts and crowdsourcing. Despite significant efforts for making KGs as comprehensive and reliable as possible, they tend to suffer from incompleteness and noise, due to the complex building process [69, 193]. This has prompted a search for automatic construction and enrichment [83, 190], often through the use of machine learning (ML). Indeed, the ML world has advanced considerably in the past decade, particularly with the rise of deep learning. From the victory of AlexNet in the ILSVRC in 2012 [96], to the release of ChatGPT in 2022, deep learning has come to dominate ML research and powers many industry applications.

One method of combining the world of knowledge and KGs with ML, and especially deep learning, is to form a vector representation of each node and edge in the KG, by optimizing some loss function based on the graph structure. The resulting set of vector representations is known as a knowledge graph embedding (KGE) and it enables several important use cases. In one direction, KGEs allow the use of predictive machine learning techniques to improve the KG, for example, by KG completion, where sparse KGs, such as those automatically constructed from text [89], are augmented with missing triples. Also, by using the deep neural network (DNN) feature vector extracted from a video, KGEs have been used to represent the content of a video as a graph [120]. Other uses of KGEs include triple (fact) classification, for assessing if a fact within the KG is correct or not, KG question answering and node clustering. Node clustering indeed can reveal similarities and differences between groups of nodes in the KG [59] and this can, for example, help uncover certain types of users in a social network, or article subjects, in a citation network. KG question answering uses the information in a KG to answer natural language questions [73]. In the other direction, KGEs allow KGs to be used to improve ML performance: for example, knowledge-aware visual question-answering [107], or reasoning of large language models (LLMs) [212].

In this paper, we introduce some of the main existing methods for combining KGs and ML, divided into two categories: those using ML to improve KGs, and those using KGs to improve results on ML tasks. From this introduction, we draw research gaps and perspectives that we consider urgent as well as promising. These gaps and perspectives are summarized in Table 1 (and analyzed and developed in section 3) and are concerned with the topics: LLM prompting, KG semantics and KGE models, symbol-based methods, ML model interpretability, and benchmark datasets. For each topic, we provide a description of some unsolved problems (gaps) that we consider to be of particular importance for future research work, and provide our views, claims, and proposals to overcome them. In particular, we support the use of KGs to formalize LLM prompting (e.g., providing concept, defining sequencing). We claim that KGE could benefit from the injection of KG semantics and usage of various reasoning capabilities, e.g., in terms of performance or negative generation. Informative negatives could also be generated by exploiting symbol-based method learning disjointness axioms (that are often missing). With respect to interpretability

¹ https://www.dbpedia.org/

² https://developers.google.com/freebase

³ https://yago-knowledge.org/

⁴ https://www.wikidata.org/

⁵ https://en.wikipedia.org/wiki/Wikipedia:Statistics

8:4 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

using KG, we argue that little progress has been made, and that in-model KG-based approaches that demonstrably produce reliable explanations are needed to validate ML results. Assessing these improvements in KGE performance or interpretability also calls for extensive empirical evaluations. Such evaluations require benchmark datasets that feature various schema constructs or levels of semantics that are currently lacking, unnoticed, or uncommon in the state of the art. That is why, we call for a systematic characterization and collection of available datasets as well as the creation of synthetic KG generators to produce tailored datasets to support experiments.

The remainder of this paper is organized as follows. Section 2 provides an overview of existing work linking KG and ML, under the framework of KGs for ML (Section 2.1) and ML for KGs (Section 2.2). Section 3 describes some gaps in the literature that we deem important, and outlines our vision of future research directions for filling these gaps. We particularly focus on: the use of KGs for prompting LLMs (Section 3.1), the integration of KG semantics and associated reasoning capabilities in KGE models for improved performance and handling of negatives (Section 3.2), the potential of symbol-based ML for KGs (Section 3.3), the attempts to use KGs for explainable AI (Section 3.4), and the need for further benchmark datasets and metrics to assess improvements brought by aforementioned directions (Section 3.5). Finally, Section 4 concludes and summarizes this work.

2 Machine Learning and Knowledge Graphs

In this section we focus on the interlink between ML and KGs. As sketched in [19], two main perspectives can be drawn: a) KGs as input to ML, whose main goal is to improve the performance in many learning tasks, e.g. question answering, image classification, instance disambiguation, text summarization, etc.; b) ML as input to KG, whose main goal is to improve the KG itself, e.g. in terms of coverage, quality, and adding new facts. In the following, we analyze the most impactful approaches in the literature, along these two perspectives.

2.1 Knowledge Graphs as Input to Machine Learning

KGs, as representations of background and contextual knowledge in a structured form, have gained significant interest from both academia and industry in the area of machine learning, enabling models to tackle complicated tasks that need prior knowledge [44]. ML models are knowledge-aware and thus can benefit from the incorporation of information that effectively captures the semantic meanings [83]. From traditional ML to modern DNNs, KGs can offer advantages, enhancing the functionality of ML systems by addressing various challenges and solving problems. In the following, we will briefly describe key applications of KGs in ML. Specifically, in Section 2.1.1, we elaborate on the key methodologies for incorporating KGs in ML, with a particular emphasis on the shortcomings they seek to mitigate. In Section 2.1.2, our focus shifts to recent advancements in describing large language models (LLMs) enhancement using KGs, a domain we believe will be increasingly significant in the future, given the widespread adoption of LLMs.

2.1.1 Addressing Machine Learning Challenges with Knowledge Graphs

KGs represent semantic descriptions of entity types and properties with a well-defined meaning. Hence, they can be employed when attempting to automatically extract features (that are difficult to measure or quantify directly) from data points [92, 133]. A feature extractor is a transformation function that maps data from a higher-dimensional space to a lower-dimensional vector space, encompassing a wide range of dimensionality reduction techniques. Early approaches map the output of feature extractors to hierarchies [100, 41] or use hierarchies as input to feature extraction [162], or use large-scale real world labels and their inter-relations [140, 39]. Many recent **Table 1** Overview of the research topics considered, the identified gaps, and our claims and proposals to address them.

Topics	Gaps	Claims & Proposals
LLM prompting	LLM hallucinationsNo formalized process to inter- act with LLMs	 Use KG at inference time to formalize the dialogue process between humans and LLM Ground prompts in knowledge (e.g., adding context, analyzing response, defining prompt sequence)
KG semantics & KGE models	 Semantics not (fully) considered Deductive capabilities not (fully) considered 	 Investigate the full exploitation of KG semantics (e.g., to improve model performance, to generate informative negatives) Possibly with different reasoning types (deductive, analogical) Empirical full assessment of the role of semantics
Symbol-based methods	Largely disregardedScalability issues	 Leverage mining of disjointness ax- ioms to generate informative negatives needed in ML models training Alleviate scalability issues
Interpretability of ML models	 Pre-/post-model approaches do not fulfill necessary require- ments In-model KG-based explainable approaches not proved to im- prove interpretability 	 Infuse KG in ML training Demonstrate that this improves ML interpretability
Benchmark datasets	 Lack of needed characteristics (e.g., schemas) Some datasets under-used or un- noticed 	 Develop a unified repository of datasets Automatically crawl in the wild and qualify datasets w.r.t. needed characteristics Create synthetic KG generators that generate both tailored schemas and KGs

8:6 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

approaches rely on image annotation that is linked to KGs, such as WordNet [127], like the image databases that have been established based on these concepts (see for example [40, 94]). On the other hand, knowledge graph embedding methods can be also seen as methods to build semantic feature extractors. This involves the mapping of entities and relations into low-dimensional vectors, effectively capturing their semantic meanings in a form that is more compatible to deep learning models [192, 138]. This field of research offers significant opportunities for exploration and advancement [145, 112] and will be analyzed in more detail in the next sections.

DNNs require a substantial amount of data for training. Sometimes, the data can either be unavailable or come with a high cost of collection. This issue, commonly referred to as the sample shortage, comes with different settings. Among them, the zero-shot learning (ZSL) [141] [49, 194] and the *few-shot learning* (FSL) [196] has recently gained significant research attention and call for the use of structured knowledge [71]. ZSL is formally defined as the task of predicting labels for new classes that have never been encountered during training, while FSL involves the task of predicting labels for new classes for which only a small number of labeled samples are provided. In both cases, the proposed solutions try to somehow transfer knowledge from seen classes to unseen classes (see [228] for recent advances on transfer learning, specifically describing knowledge transfer). Here, KGs play an important role, since they can represent background knowledge such as class hierarchies, instances of classes (samples), features, properties, relations as well as meta information like model parameters, providing the necessary auxiliary information. The interested reader can check [194] and [196] for a systematic review on ZSL and FSL, [71] and [27] for ZSL and FSL based on external knowledge (covering some works that use KGs as the background knowledge), [133] covers the use of knowledge graphs specifically for visual transfer learning and [28] that is a recent thorough survey paper that specifically classifies and analyzes methods utilizing KGs for ZSL and FSL.

The capabilities of DNNs have enabled the development of numerous models and techniques to address challenging problems, particularly those involving multimodal data. In this context, multimodal machine learning [14, 61, 132] has emerged as one of the rapidly advancing fields within artificial intelligence, addressing various challenging problems, including visual question answering, visual reasoning, image captioning, image-text retrieval, visual storytelling, visual dialoguing and others [3, 66, 217, 204, 169, 45, 45, 95]. Not surprisingly, the proposed DNNs models (mainly based on transformers) often struggle with generalization to various concepts and scenarios that demand commonsense knowledge, or understanding of abstract entities, facts. and real-world events, due to the lack of formal representation of background, contextual and commonsense knowledge [150, 74, 90]. Hence, integrating external knowledge at different stages of multimodal learning, especially in pre-training or fine-tuning, augments the capabilities of models, enabling them to better address a broader range of real-world scenarios. Several proposed DNNs models are based on external knowledge that is represented using semantic descriptions stored in KGs. In particular, there have been proposed datasets that leverage external knowledge [122, 177, 149, 200] linked to web resources and KGs [106] to learn the alignment between visual and textual information [30] in order to solve multimodal learning tasks. The interested reader can find information in several survey papers classifying and analyzing methods in the area of multimodal learning (see for example [14, 61, 132], specifically presenting works that make use of KGs [119]).

The adoption of symbolic knowledge representation and reasoning as a means to address the opacity of machine learning classifiers is a research domain that has recently garnered significant attention from researchers [58]. The need to provide explanations grounded in domain knowledge with formal semantics has driven the utilization of KGs in explainable AI [32, 111, 42, 25, 180]. As this field holds considerable interest and offers numerous prospects for future research, we discuss it in more detail in Section 3.4.

2.1.2 Knowledge Graphs for Large Language Models

The current ML literature is dominated by Deep Learning solutions that have been proved very effective in multiple domains and for multiple tasks. Particularly, nowadays LLMs and related systems are catalyzing the attention of the scientific and industrial community for their impressive ability to provide highly accurate results in a very limited amount of time, as for the case of ChatGPT⁶ and similar solutions. LLMs behind these systems (like the GPT models [22] that currently guide ChatGPT) are usually deep learning models that have been trained on huge amounts of text data and are capable of understanding and generating human-like text. Typically, they get a text in their input and provide a text as a response. Lately, they can be also directly connected to other generative models like Midjourney⁷ and DALLE-3⁸ that get text as input and give image or videos in the output, advancing the user experience and extending the scope of application domains.

There are many ways of using KGs to improve or understand the operation of LLMs. There are works that aim to enhance the text generation (see for example the survey [216]) or more generally to enhance visiolinguistic learning with knowledge (see for example the survey [119]). In [142] several methods are discussed that try to unify LLMs and KGs, combining their advantages. Among others, methods that use KGs to improve the operation of LLMs are analyzed. An interesting approach is to incorporate knowledge graph information into LLMs in order to enhance their performance, by advancing the factual knowledge understanding. This is a way to improve the LLM performance on knowledge-intensive tasks, and to generate more informed and contextually grounded text. In particular, there are works that try to enhance word representations with knowledge graph embeddings providing context, improving the model's performance [146], or to learn contextualized representations that capture both linguistic and factual knowledge [118], or to use KGs in pre-training to enhance the model's understanding of factual knowledge [174, 109]. Other works in the area try to decompose knowledge into separate modules to improve its natural language understanding capabilities [219], or to integrate KG and language understanding in a joint pre-training framework [215].

Moreover, there are other approaches for graph-to-text generation integrating knowledge from a knowledge graph into the text generation process, trying to produce more informative and coherent outputs [214]. In this framework, combining language representations with knowledge graph embeddings can be used to enhance the representation of contextualized knowledge [173, 171, 65]. Sentiment knowledge can be also incorporated with the use of KGs, thus enhancing the performance of language models with respect to sentiment analysis accuracy [178].

Finally, KGs can be used to prob and possibly understand different aspects of the operation of LLMs. In particular, KGs can be used to elicit knowledge from language models using automatically generated prompts, enabling targeted information retrieval from the model's knowledge base [164], or for querying language models effectively, through a query generation technique that leverages explicit context [2], or to contrastively probing LLMs to investigate the domain knowledge of pretrained language models by comparing their performance to specially designed contrast models [125]. Prompting can be also used for understanding the limitations LLMs, revealing scenarios where language models may produce unreliable or incorrect responses [121], or to enable the exploration and understanding of the underlying knowledge captured by LLMs [176], or to understand how LLMs capture factual knowledge and identify the key factors that contribute to their acquisition of factual information [108].

TGDK

⁶ https://openai.com/blog/chatgpt

⁷ https://www.midjourney.com/home/

⁸ https://openai.com/dall-e-3

8:8 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

Of particular significance in this context is the utilization of KGs to validate LLMs, mitigating the issue of hallucination, that causes the generation of factually incorrect content [84]. Hallucination of LLMs poses a substantial challenge to their reliability [15]. Although some LLMs are equipped with the ability to explain their predictions, their explanatory capabilities also suffer from hallucination and this has been particularly connected to the criticism that LLMs have limited ability to encode factual knowledge [229, 188, 56]. Hence, it becomes crucial to examine and authenticate the knowledge embedded within LLMs to prevent hallucination. Recently, there is some work in the area of utilizing KGs for hallucination detection. Specifically, KGs are used as an external source to validate LLMs reliability [85], or to develop fact-checking models, identifying and mitigating hallucinations [48]. This is a very interesting area for future work.

In Section 3.1, we outline our viewpoint on the most important research areas that require attention in order to address the challenges discussed here.

2.2 Machine Learning as Input to Knowledge Graphs

From the perspective of ML as input to KGs, the main objective is to improve the quality of existing KGs overall. Particularly, given the well-known issues concerning noise and incompleteness of KGs, most solutions have focused on KG refinement which actually encompasses several tasks. Among the others, triple classification (aiming at assessing the correctness of a statement in a KG and generally regarded as a binary classification problem) and mostly link/type prediction (aiming at predicting missing links/types between entities and generally regarded as a learning to rank problem) gained most of the attention, aiming at improving/limiting KG incompleteness.

Different approaches have been developed over the years, with the goal of improving effectiveness (mostly targeting the link prediction problems) while scaling to very large KGs. Mostly, numericbased methods have been investigated. Among the very first proposals, *probabilistic latent variable models* from the Statistical Relational Learning (SRL) [54] field (having as main goal the creation of statistical models for relational/graph-based data) have been formalized. Successive and very efficient solutions have been represented by *Knowledge Graph Embedding* (KGE) models. Other approaches focusing on propositionalization techniques, recently also exploiting Graph Neural Networks (GNN) [201]) have been also pursued. Complementary to these numeric-based solutions, research directions targeting symbol-based models have been also proposed, particularly focusing on rule-based methods for predicting triples in KGs.

In the following we summarize the most representative methods for each of the aforementioned categories. We dedicate particular attention to KGE methods that represent the main subject of study for our successive proposals, illustrated in Section 3.2.

2.2.1 Probabilistic Latent Variable Models

Probabilistic Latent Variable Models explain relations between entities by associating each resource to a set of intrinsic latent attributes (i.e. attributes not directly observable in the data) and conditions the probability distribution of the relations between two resources on their latent attributes. All relations are considered conditionally independent given the latent attributes. This allows the information to propagate through the network of interconnected latent variables.

One of the first solutions belonging to this category is the *Infinite Hidden Semantic Model* (IHSM) [153]. It formalizes a probabilistic latent variable that associates a latent class variable with each node and makes use of constraints expressed in First Order Logic during the learning process. IHSM showed promising results but was found to have limited scaling on large data collections, because of the complexity of the probabilistic inference and learning, which is intractable in general [91].

2.2.2 Knowledge Graph Embedding Models

KGE models have received considerable attention because of their impressive ability to scale on very large KGs. KGE are numeric-based approaches that convert the data graph into an optimal low-dimensional space in which graph structural information and graph properties are preserved as much as possible [23, 83]. The embedding procedure consists of learning embeddings such that the score of a valid (positive) triple is lower than the score of an invalid triple, i.e. the invalid triples function as negative examples. Graph embedding methods may differ in their main building blocks: the representation space (e.g. point-wise, complex, discrete, Gaussian, manifold), the encoding model (e.g. linear, factorization, neural models) and the scoring function (that can be based on distance, energy, semantic matching or other criteria) [83]. Over the years, several models have been developed. Some are presented below. It should also be noted that several libraries or frameworks such as Deep Graph Library⁹ [191], PyKEEN¹⁰ [6], or PyTorch-BigGraph¹¹ [105] have been developed and provide unified implementations of wide ranges of models.

One of the first solutions that has been proposed is RESCAL [139], which performs graph embedding by computing a three-way factorization of an adjacency tensor that represents the multi-graph structure of the data collection. It resulted in a powerful model that was also able to capture complex relational patterns over multiple hops in a graph, however it was not able to scale on very large graph-based data collections (e.g. the whole YAGO or DBpedia). The main limitation was represented by the parameter learning phase, which may take rather long for converging to optimal solutions.

The very first highly scalable embedding model is TRANSE [20]. It introduces a simple but effective and efficient model: each entity is represented by an embedding vector and each predicate is represented by a (vector) *translation operation*. The score of a triple is given by the similarity of the translated subject embedding to the object embedding. The optimal embedding and translation vectors for predicates are learned jointly. The method relies on a *stochastic optimization process*, that iteratively updates the distributed representations by increasing the score of the positive triples i.e. the observed triples, while lowering the score of unobserved triples standing as negative examples. The embedding of all entities and predicates in the KG is learned by minimizing a *margin-based ranking loss*.

Despite its scalability and effectiveness, TRANSE remained limited in properly representing various types of properties such as *reflexivity*, and 1-to-N, N-to-1 and N-to-N relations that can be easily found in KGs (e.g. typeOf as an example of N-to-N relationship). To tackle this limitation while keeping the ability to scale to very large KGs, a large family of models has been developed that build on TRANSE, such as TRANSH [197] and TRANSR [113].

Specifically, TRANSR adopts a score function that projects entities into a different vectorial space for each relation through a suitable projection matrix. TRANSR associates to typeOf, and to all other properties, a specific vector space in which entity vectors are projected. This leads to training specific projection matrices for typeOf (and any other relation) so that the projected entities can be located more suitably to be linked by the vector translation associated to the (typeOf) relation. This differs from TRANSE, which models typeOf as simple vector translation. The considered individuals and classes may be quite different in terms of the properties and attributes they are involved in, thus determining strong semantic differences (according to [210]) taking place at small reciprocal distances in the underlying vector space, hence revealing the weakness of employing the mere translation.

With the goal of capturing additional properties in the data, such as inverse relationship, symmetry, anti-symmetry and composition, more complex embedding models have been formalized,

⁹ https://www.dgl.ai/

¹⁰ https://github.com/pykeen/pykeen

¹¹ https://github.com/facebookresearch/PyTorch-BigGraph

8:10 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

either targeting more complex vector representation spaces, such as the complex representation, as for the case of COMPLEX [184] and (PATH-)ROTATE [224], Gaussian representation, as for the case of KG2E [67] and TRANSG [203], and manifold representation, as for the case of MURP [13] and DIHEDRAL [206], or targeting more complex encoding models such as neural models, as for the case of CONVKB [137] and COMPGCN [185]. Nevertheless, these additional models became rather computationally expensive, which limits their usefulness.

Nevertheless, several additional semantic aspects that are generally available within KGs, such as hierarchies of concepts and roles, type constraints and transitivity of relationships, are currently almost disregarded by existing KGE models. The need for *semantic embedding methods* has been argued [33, 144, 82]. In [60] a KG embedding method considering logical rules has been proposed, where triples in the KG and rules are represented in a unified framework. Specifically, triples are represented as atomic formulae while rules are represented as more complex formulae modeled by t-norm fuzzy logics. A common loss function over both representations is defined, which is minimized to learn the embeddings. This proposal resulted in a novel solution but the specific form of prior knowledge that has to be available constitutes its main drawback. A similar drawback also applies to [129], where a solution based on adversarial training is formalized, exploiting Datalog clauses to encode assumptions which are used to regularize neural link predictors.

Complementary solutions, directly targeting rich representation languages as RDFS and OWL and not requiring additional formalism for representing prior knowledge have been proposed. Particularly, [128] has proven the effectiveness of combining embedding methods and strategies relying on reasoning services for the injection of prior Background Knowledge (BK) to enhance the performance of a specific predictive model. Following this line, TRANSOWL, aiming at injecting schema level information, particularly during the learning process, and its upgraded version TRANSROWL, have been formalized [36, 35]. The main focus is on the application of this idea to enhance well-known basic scalable models, namely TRANSE [20] and TRANSR [113], even if, in principle, the proposed approach could be applied to more complex embedding methods. with an additional formalization. In TRANSOWL the original TRANSE setting is maintained while resorting to reasoning with schema axioms to derive further triples to be considered for training and that are generated consistently with the semantics of the properties. Particularly, for each considered axiom, TRANSOWL defines, on the score function, specific constraints that guide the way embedding vectors are learned. A set of different axioms, specifically equivalentClass, equivalentProperty, inverseOf and subClassOf, are employed for the definition of constraints on the score function so that the resulting vectors, related to such axioms, reflect their specific properties. As a consequence, new triples are added to the training set on the grounds of the specified axioms. TRANSROWL further develops TRANSOWL by adopting TRANSR as the base model in order to handle non 1-to-1 properties in a more proper way. TRANSOWL and TRANSROWL have been proven to improve their effectiveness on link prediction and triple classification tasks when compared to the baseline models (TRANSE and TRANSR) that focus on structural graph properties. Some additional efforts in the formalization of KGE and Deep Learning solutions taking into account limited semantics can be found in the literature [57, 12, 72, 62, 99]. Nevertheless, none of the existing KGE model is able to exploit the full expressiveness that a KG may have in principle.

Independently of the specific model, another important issue needs to be highlighted: most of the existing KGs only contain positive (training) examples, since usually false facts are generally not encoded. However, training a learning model in all-positive examples could be tricky, because the model might easily overgeneralize. As such, in order to obtain the negative examples that are needed to train KGE models, two different approaches are generally adopted: either *corrupting* true/observed triples randomly, with the goal of generating plausible negative examples or

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

adopting a *local-closed world assumption* (LCWA) in which the data collection is assumed as *locally* complete [138]. In both cases, wrong negative triples may be generated and thus used when training and learning the embedding models.

In Section 3.2, we present our perspective on the research directions that need to be tackled to cope with the problems illustrated particularly in this section.

2.2.3 Neural Methods for Vector Space Embeddings

Another research direction focused on the exploitation of vector space embeddings for obtaining a propositional feature vector representation of a KG. One of the first solutions targeting this research direction is RDF2Vec [154], which adapts the well-known Word2Vec technique, devised for natural language processing, to graph representations. A two-step approach is adopted. First the data graph is converted into a set of sequences of entities (two different approaches can be used for this purpose: graph walks and Weisfeiler-Lehman Subtree RDF graph kernels). In the second step, the obtained sequences are used to train a neural language model to estimate the likelihood of a sequence of entities appearing in a graph. The result is that each entity in the graph is represented as a vector of latent numerical features. In order to show that the obtained vector representation is independent of the downstream task and the specific algorithm, an experimental evaluation involving a number of classification and regression tasks has been performed.

An upgrade of RDF2Vec has been presented in [31], where global patterns are considered (differently from the initial RDF2Vec proposition grounded on local patterns). These solutions cannot cope with literals.

Another way to better capture global information is to use a more powerful model, such as a graph neural network (GNN). These are a class of methods for allowing artificial neural networks to operate on graph data. Given that graphs are a very general data structure, GNNs can take a wide variety of forms. It has also been shown that many popular deep learning architectures, such as convolutional neural networks, recurrent neural networks, and transformers, can be seen as a GNN for a suitably defined graph [21]. In a GNN, as for RDF2Vec and KGE models, nodes are represented as vectors. These vectors are fed through a sequence of message-passing layers, where nodes update their values based on their neighbors' values, and local pooling layers, where groups of neighboring nodes are combined into a single vector representation. The final layer aggregates the entire input into a single vector representation for the entire graph. Because of this iterative process, GNNs are better able to capture multi-hop relations and global graph structure, compared to RDF2Vec [154]. They are also able to reduce an entire graph to a single embedding vector, as well as computing embedding vectors for each node. See [223] or [226] for an overview of GNN design and applications.

Several works have applied GNNs to construct or enhance KGs. [227] integrates Bellman-Ford into the GNNs training procedure, and then uses the resulting model for link prediction on KGs. [143] show that GNNs can be trained, in a supervised setting, to accurately estimate node importance in a KG. GNNs have also been used for entity alignment, which seeks to discover when the same entity appears in two different knowledge graphs. [198] embeds entities in both KGs and then uses the distance between the embeddings to identify when nodes in different KGs correspond to the same entity. More recent works have built on this method, for example by capturing time-sensitive information [207] or multi-modal inputs [170]. Another common use of GNNs for KG is to improve the use of KGs in recommender systems [52], and inference [136]. For an overview of the use for GNNs for KGs, see [213].

8:12 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

2.2.4 Rule Learning Solutions

With the goal of finding new facts (namely new triples) that are missing in a KG, AMIE $[51, 50]^{12}$ has been proposed. AMIE represents one of the most well-known and efficient solutions grounded on a symbol-based approach. Inspired by *association rule mining* [4] and the Inductive Logic Programming (ILP) literature, AMIE is able to learn logic rules from KGs, that are ultimately used for predicting new unseen triples. Interestingly AMIE is tailored to support the Open World Assumption (OWA) characterizing KGs, differently from all numeric-based solutions that are grounded on the Closed World Assumption (CWA). Nevertheless, AMIE mines rules inspecting the triples that are directly observable in the KG and it does not exploit the additional semantics that is available in the KG as well as any form of deductive reasoning.

A related rule mining system, based on a level-wise generate and test strategy has been further proposed [37], with the goal of learning SWRL rules [70] while exploiting schema level information and deductive reasoning capabilities during learning. As for AMIE, the goal was to exploit the discovered rules for predicting new facts. This system actually outperformed AMIE in terms of new predicted triples, and this was due to the exploitation of the schema level information and reasoning capabilities. Nevertheless, they have been also the main cause of the reduced ability of the system to scale on large KGs, when compared to AMIE.

More recently AnyBURL [123] has been proposed. It is a scalable bottom-up rule learning system for KG completion that works by sampling random paths that are generalized into Horn rules. Reinforcement learning is exploited to guide path sampling and make efficient use of computational resources. AnyBURL also showed improved scalability and competitive performance in comparison to numeric-based approaches. Even more so, it has been also shown that AnyBURL can be used to explain predictions made by a latent model when restricting the types of learned rules. Nevertheless, as for AMIE, no exploitation of the KG semantics and reasoning capabilities can be found.

3 Gaps in Machine Learning and Knowledge Graphs and Next Challenges

In this section we analyze existing gaps of the class of methods illustrated in Section 2 that we identify as important. Hence, for each of them, we provide our perspective on the research directions that need to be pursued in order to fill these gaps. Specifically, the following Section 3.1 primarily focuses on the need of having a clear methodology for interleaving LLMs with KGs and drafts a preliminary proposal. Section 3.2 primarily focuses and provides preliminary proposals for the need of taking into account reasoning capabilities and schema level information of KGs, to be used for setting up a more informative way for generating negative training examples as well as for injecting schema level information in KGE. Beyond the gaps, Section 3.3 presents our view supporting that symbolic ML methods may still have a role in KG, particularly for KG refinement and more specifically for mining disjointness axioms, that are quite often missing in KGs and related ontologies. Section 3.4 presents our position on the need for an approach that demonstrably produces reliable explanations to validate ML results when applied to KGs. Hence, Section 3.5 shows the need for diverse, high-quality benchmark datasets when combining ML and KGs as well as new metrics for capturing new behaviors.

¹²AMIE system is currently at its third version. For more details see https://github.com/dig-team/amie.

3.1 Knowledge Graphs for Prompting Large Language Models

From what has been described in section 2.1, we understand that the use of KGs, as an additional tool, during the (pre-)training phase or during the inference phase of LLMs are important fields of research, attracting the interest of many researchers, and could potentially improve the operation of the LLM and the results of LLMs, respectively. Although the operation of modern LLMs and respective systems (like chatGPT) is impressive and traditional machine learning gaps (like reasoning capabilities) have started to close, **there is still room for improvement**, and the use of KGs as an additional tool during the training and fine-tuning phases can play an important role, here. Specifically, KGs can provide background knowledge (encyclopedic, commonsense, domain-specific, multimodal etc), represent human-oriented processes, and explain opaque machine operation. On the other hand, the practical use of LLMs increases dramatically and **there is a great need for advancing the use of LLMs inference, making the process of dialoguing LLMs more formal and systematic**. Therefore, the use of KGs during the phase of the design of the input to be given to LLMs and during the phase of the analysis of the LLM response seems to have a great potential.

Following the above, interesting open problems and challenges is the use of KGs in LLM *prompt engineering* or simply LLM *prompting* [142, 116]. Prompting is the process of providing a sequence of instructions or queries to a LLM in order to get the desired output or to check the LLM's operation and characteristics. It is actually a dialogue between a user (human or agent) and a LLM, that reflects the user's intent and finally results in the desired task or information that the user wants to get from the model. Although the field is new, there are some attempts to formalize the process (see for example the *Automatic Prompt Engineer (APE)* approach [225]). The formalization of the dialogue process should be grounded on some type of background knowledge, so there is a need for representing and using this knowledge. Here, we describe the great potential of using KGs in LLM prompting, based on the nature of prompts, their types and effectiveness, the tasks and the methodology to provide adequate prompts during the prompting process, focusing on the potential use of KGs.

There are many ways to modify the prompt that is given to LLM, using KGs. First, the instruction or question can be more explicit and specific, capturing the user requirements, since it is well-understood that the more specific the prompt the better the chance of guiding the LLM to the desired response. For example, the instruction "Summarize text A" can be specified as "Summarize the text A in 200 words", using the knowledge that an abstract should be between 200 and 300 words. Or the question "Is there any recent paper in the area of prompting machine learning systems?" can be specified as "Is there any recent paper in the area of prompting LLMs?". On the other hand, sometimes it may be helpful, depending on the instruction or the question, to generalize it, for example, the question "Is there any recent paper in the area of prompting machine learning systems?" can be generalized as "Is there any recent work in the area of prompting machine learning systems?". Also, may be useful to contextualize or style the prompt, by providing examples ("Suggest romantic musicals, like "La La Land"), or conditions ("Suggest papers for prompting LLM, published in top conferences"), or style ("Paraphrase text A, using more formal language). It is not difficult to see that KGs can be very helpful in constructing knowledge-enhanced prompts like the above (and not restricted to them), guiding prompt changes, as they capture formal domain knowledge descriptions. Interesting ideas can be found in [225] that the instruction generation is framed as natural language program synthesis, in [166] that simple and effective prompts are constructed to improve GPT-3's reliability, in [189] that multi-step reasoning tasks are tackled by constructing planning and solving prompts, in [222] that LLMs are asked to provide explanations for their choices (in this case for a specific task that is model selection) and in [116] that prompting with generated knowledge rectifies model prediction.

8:14 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

Response analysis. Another interesting issue that could be considered is to use KGs to characterize the prompt, for example to measure its *effectiveness* or *reliability*, by analyzing and evaluating the response. The effectiveness of prompts depends on the response of the LLM, i.e. the answer to a specific prompt in comparison with the desired output, given the task. Depending on the prompt and response languages, it is important to formalize effectiveness or reliability evaluation measures that guide a process of iterative refinement of the results, by using formal knowledge represented in KGs. Interesting ideas can be found in [142, 118, 166].

Prompt sequencing. Designing and controlling prompting, i.e. producing a sequence of prompts to elicit a desired output, can be a challenging task that requires a systematic strategy, evaluation and experimentation. Although LLMs are powerful, their operation is complex and unpredictable and thus a dialogue for producing a sequence of prompts may be helpful to understand LLM characteristics, like complex reasoning capabilities. There is lately some work in the area, for example: Chain-of-Thought (CoT) prompts [199] decompose complex reasoning capabilities into a set of simpler reasoning steps; In [116], the usefulness of using knowledge in common sense reasoning is discovered, extracting knowledge from an LLM and then using this knowledge as additional input to refine the prompt result. The APE methodology proposed in [225] uses ideas from program synthesis in order to optimize the prompt selection process, based on efficient score estimations. Future steps would benefit from the use of KGs as formal knowledge representations, because there is a clear requirement formalizing the prompting extraction methodology.

3.2 Handling Semantics, Reasoning and Negative Information in Knowledge Graph Embedding Methods

One of the key features of KGs is that they can be enriched with schema-level information. For the purpose ontologies are generally adopted, which coupled with deductive reasoners, allow to make explicit knowledge which is implicitly coded in a KG¹³. For example, given a KG containing the triple <c typeOf Woman> (or equivalently Woman(c), by adopting a Description Logic formalism) and referring to the following simple ontology formalizing a hierarchy of concepts Man \sqsubseteq Human and Woman \sqsubseteq Human, the fact Human(c) can be derived by the use of a deductive reasoner. Similarly, new knowledge can be derived when additional axioms are available, such as equivalence axioms, disjointness axioms, as well as restrictions on domain and ranges¹⁴. However, due to the limited ability of reasoners to scale on very large KGs, deductive reasoning is currently almost disregarded.

Indeed, when talking about ML methods coupled with KGs, as for the case of KGE methods, generally only facts that can be directly observed are taken into account e.g. when projecting the data graph into a lower vectorial representation space. This is clearly a limitation, since knowledge that is somehow already available within the KGs (as for the fact Human(c) in the example above) and that may play a role when considering KGE is ignored. For instance, by considering the fact Human(c), a more appropriate vectorial representation for the entity c could be provided thus limiting errors also when solving downstream tasks. By only considering observable facts, schema level information, that is a seminal element of knowledge, and all additional knowledge that can be derived are actually fully dismissed.

¹³ Several reasoners exist and may be used for the purpose. Some examples are RDFox (https://www.oxfordsemantic.tech/rdfox), HermiT (http://www.hermit-reasoner.com/), FaCT++ (http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/ for the full list of reasoners.

¹⁴See https://www.w3.org/TR/owl2-overview/ for details on the representation language.

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

Abboud et al. [1] analyzed the shortcomings of the existing embedding models. These shortcomings can be summarized in: theoretical inexpressiveness, lack of support for inference patterns and higher-arity relations, need for logical rule incorporation.

Here, we specifically claim that KGE methods need to be equipped with the full usage of KGs semantics which comprises the exploitation of all axioms that can be found in the ontologies that are used for supplying (rich) schema level information to KGs, as well as the exploitation of deductive reasoning services that allow to obtain additional knowledge both at schema and assertion level. Indeed, whilst the need for semantic embedding methods has been advocated [33, 144, 82], only a few proposals can be found in the literature that actually address this problem (see section 2.2.2 for details) and mainly focusing on equivalentClass, equivalentProperty, inverseOf and subClassOf axioms. To the best of our knowledge, none of the existing methods is able to exploit all kinds of axioms that in principle can be found in expressive ontologies. Even more so, a complementary research direction would be needed, calling for a solid and extensive experimental evaluation aiming at providing a clear position on the need (or not) to fully exploit the KG semantics as well as reasoning capabilities. Specifically, we claim that a comprehensive experimental evaluation, involving most of the KGE methods currently available, is needed. Two main scenarios should be considered: the first one (currently adopted) where only observable facts are considered; the second one where the full knowledge available within KG is made explicit by considering schema-level information (e.g. transitivity, equivalence axioms, same as axioms etc.) and reasoning capabilities. Hence performances on the very same downstream tasks, adopting the two settings, should be compared, in order to experimentally prove the value added, if any, of exploiting the KGs entirely. Importantly the second scenario could be possibly divided into two intermediate steps, one where knowledge is partially completed by considering the schema level information but no exploitation of deductive reasoners and a second step where the actual full knowledge is gained by adopting available deductive reasoners. This is on one hand, for assessing the impact of the usage of the full knowledge and on the other hand, for assessing if some complexity, due to reasoning, can be saved whilst still trying to make knowledge explicit as much as possible.

Another issue with KGE models is given by the need of negative examples (for training KGE models) that anyhow are generally missing in KGs, where generally only positive information is coded. As illustrated in section 2.2.2, this problem is usually addressed either by *corrupting* true/observed triples randomly, that is by replacing either the subject or the object of the observed triple with an entity picked randomly from the KG, or by adopting a *local-closed world assumption* (LCWA), in which the data collection is assumed as *locally* complete [138]. In both cases, wrong negative triples may be generated and thus used when training and learning the embedding models. In order to mitigate this issue, preliminary proposals tried to take under control the number of negatives that are randomly generated [43]. Clearly this solution does not solve the problem of generating false negatives, but rather simply tries to somehow control the effect of the false negatives. One of the first proposals trying to generate and materialize actual negative triples has been formalized in [8]. Nevertheless, the proposed solution is grounded on the exploitation of additional and external sources of information besides KGs. Specifically, the proposed solution is grounded on two complementary approaches: a statistical ranking for statements obtained based on related entities, and a pattern-based text extraction, applied to search engine query logs.

On the contrary, here we claim that KGs semantics should be fully and solely exploited for making explicit correct negative statements. For instance, given a restriction on domain and/or range of predicate appearing in a true observed triple, the restriction can be exploited for generating negative triples where e.g. the object entity of the negative triple can be deductively proved to be out of the declared range restriction. Similarly, given an observed true triple with a

8:16 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

predicate having a functional restriction, negative statements may be generated by constructing triples having objects that are different from the object in the true statement. More generally, the approach for generating correct negative statements that is envisioned, is deeply grounded on the semantics of the schema axioms. The approach should basically construct triples that are in the complement of the set of triples representing the semantics of a given schema axiom.

An initial proposal in this direction can be found in [36, 35, 117], where only domain, range, disjointWith and functionalProperty constraints are considered. Whilst we consider this proposal a valuable way to go as in agreement with the envisioned solution, it needs to be extended for comprising all axioms and constraints that can be possibly found in a KGs, e.g. transitivity, same-as, equivalence axioms, for citing a few. Even more so, we consider it worthwhile to conduct an extensive experimental study comparing the different settings for generating negative examples in order to prove experimentally the actual role of semantics, if any.

Up to now, when referring to reasoning we basically meant deductive reasoning applied to ontologies/KGs [11]. Nevertheless, besides deductive reasoning, other forms of reasoning could be investigated. These different reasoning forms could be useful in KG-related tasks, and conversely, knowledge contained in KGs could be leveraged in their reasoning process. Here we specifically focus on analogical reasoning that is a remarkable capability of the human mind [131] relying on analogical proportions. They are statements of the form "A is to B as C is to D" that can be formalized as quadruples A:B::C:D [126]. An example of such a quadruple is "leg: human :: paw : dog". Analogical reasoning relies on similarity and dissimilarity to extrapolate knowledge between objects of potentially different domains. To illustrate, the given example quadruple leverages the similarity between body parts and whole, and the relation linking them to constitute a valid analogy. Analogical reasoning is mainly concerned with two tasks: analogy detection that aims to determine whether a quadruple A:B::C:D is a valid analogy, and analogy solving that aims to predict a missing element X, given three elements A, B, and C such that A: B:: C: D constitutes a valid analogy. When elements are represented as vectors, analogies can be thought of as parallelograms, *i.e.*, $e_B - e_A = e_D - e_C$. Such a view can thus be adopted with embeddings, which attracted works on ML-based analogy for various Natural Language Processing tasks, e.g., word morphology [7] or machine translation [101]. In the realm of KGs, to the best of our knowledge, only a few works consider analogical reasoning. However, KG embeddings are suited for analogical formalization. For instance, by using translations to model relations, TRANSE inherently validates the parallelogram rule. This motivated Portisch et al. [147] to investigate whether some KG embedding models are well-suited for the task of analogy detection with standard analogical datasets. But analogical reasoning could also be directly applied to KGs. In the link prediction task, it is natural to extrapolate edges from one (part of a) KG to another (part), which motivated the ANALOGY model [115]. Interestingly, ANALOGY is based on the parallelogram rule and the authors showed that it subsumes some other models such as DISTMULT, COMPLEX, and HOLE. Analogical reasoning can also be considered as an enhancer of existing KGE models by using triples, relations or entities in analogies to enrich the training process [211]. In fact, the integration of analogical reasoning into KG-related tasks and KGE models is not limited to one formalization or one task. Jarnac et al. re-used a convolutional model for analogy detection and applied it on pre-trained graph embeddings to select subgraphs of interest from Wikidata to bootstrap a domain-specific KG [81]. Analogies also inherently appear in several other tasks, e.g., Semantic Table Interpretation, matching, or recommendation [134]. It remains to explore both theoretically and empirically the best formalizations, models, improvement in performance, and interactions with other forms of reasoning, especially deductive reasoning that is inherently permitted by ontologies.

3.3 Symbol-based Methods for Knowledge Graphs

Given KGs volumes, the need for scalable ML solutions has obfuscated the attention to symbol-based ML solutions. Nevertheless, the important gain, in terms of scalability, that numeric-based methods (such as KGEs) are obtaining is penalizing: a) the possibility to have interpretable models as a result of a learning process (see Section 3.4 for more details); b) the ability to exploit deductive (and complementary forms of) reasoning (see Section 3.2 for more details); c) the expressiveness of the representations to be considered and related assumptions (such as the Open World Assumption (OWA)).

Indeed, suitable symbol-based methods, often inspired by the *Inductive Logic Programming* (ILP) [151] field (aiming at inducing a hypothesized logic program from background knowledge and a collection of examples), have been proposed [34, 86, 103, 51, 179]. Most of them are able to cope with expressive representation languages such as Description Logics (DLs) [11], theoretical foundation for OWL, and the *Open World Assumption* (OWA) typically adopted, differently from the *Closed World Assumption* (CWA) that is usually assumed in the traditional ML settings. Also, problems such as ontology refinement and enrichment at terminology/schema level have been proposed [46, 47, 102, 186, 157].

Particularly, with the purpose of enriching ontologies at the terminological level, methods for learning concept descriptions for a concept name have been formalized. The problem has been regarded as a supervised concept learning problem aiming at approximating an intensional DLs definition, given a set of individuals of an ontological KB acting as positive/negative training examples. Various solutions, e.g. DL-FOIL¹⁵ [46] and CELOE [102] (part of the DL-LEARNER suite¹⁶), have been formalized. They are mostly grounded on a separate-and-conquer (sequential covering) strategy: a new concept description is built by specializing, via suitable refinement operators, a partial solution to correctly cover (i.e. decide a consistent classification for) as many training instances as possible. Whilst DL-FOIL works under OWA, CELOE works under CWA. Both of them may yield sub-optimal solutions. In order to overcome such issues, DL-FOCL¹⁷ [159, 158], PARCEL [182] and SPACEL [183] have been proposed. DL-FOCL is an optimized version of DL-FOIL, implementing a base greedy covering strategy. PARCEL combines top-down and bottom-up refinements in the search space. Specifically, the learning problem is split into various sub-problems, according to a divide-and-conquer strategy, that are solved by running CELOE as a subroutine. Once the partial solutions are obtained, they are combined in a bottom-up fashion. SPACEL extends PARCEL by performing a symmetrical specialization of a concept description. All these solutions proved to be able to learn approximated concept descriptions for a target concept name to be used for possibly introducing new (inclusion or equality) axioms in the KB. Nevertheless, quite often, relatively small ontological KBs have been considered for the experiments, revealing that, currently, they have **limited ability to scale** on very large KGs.

A few scalable exceptions are represented by rule learning systems for KG completion such as AMIE and most of all AnyBURL (see section 2.2.4 for more details). Nevertheless, most of the existing symbol-based methods cannot scale to very large KGs [158].

Here we want to highlight particularly the **role that symbolic ML solutions may play in assessing disjointness axioms** within ontologies. Indeed, disjointness axioms are essential for making explicit the negative knowledge about a domain, yet they are often overlooked during the modeling process [193]. Furthermore, disjointness axioms would be absolutely beneficial for setting up an informed generation of negative examples in KGE models (see section 3.2 for details), thus limiting false negatives that random corruption may inject.

TGDK

¹⁵System publicly available at: https://bitbucket.org/grizzo001/dl-foil/src/master/

¹⁶Suite publicly available at: https://dl-learner.org/

¹⁷System publicly available at: https://bitbucket.org/grizzo001/dlfocl/src/master/

8:18 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

To tackle this problem, automated methods for discovering disjointness axioms from the data distribution have been devised. A solution grounded on association rule mining [4] has been proposed in [186]. It is based on studying the correlation between classes comparatively, namely by considering association rules, negative association rules and correlation coefficient. Background knowledge and reasoning capabilities are used to a limited extent. A different solution has been proposed in [157, 156], where, moving from the assumption that two or more concepts may be mutually disjoint when the sets of their (known) instances do not overlap, the problem has been regarded as a clustering problem, aiming at finding partitions of similar individuals of the knowledge base, according to a *cohesion* criterion quantifying the degree of homogeneity of the individuals in an element of the partition. Specifically, the problem has been cast as a conceptual clustering problem, where the goal is both to find the best possible partitioning of the individuals and also to induce intensional definitions of the corresponding classes expressed in the standard representation languages. Emerging disjointness axioms are captured by the employment of terminological cluster trees (TCTs) and by minimizing the risk of mutual overlap between concepts. Once the TCT is grown, groups of (disjoint) clusters located at sibling nodes identify concepts involved in candidate disjointness axioms to be derived¹⁸. Unlike [186], that is based on the statistical correlation between instances, the empirical evaluation of [157, 156] showed the system ability to discover disjointness axioms also involving complex concept descriptions, thanks to the exploitation of the underlying ontology as background knowledge.

Here, we claim that, when tackling the problem of learning disjointness axioms, a two-level analysis needs to be conducted. One level relates to the expressiveness of the axioms that can be learned. The other level is related to the usage of the learned axioms from a user/knowledge engineering perspective. The goal of this two-level analysis should be finding a trade-off between expressiveness and utility from a user modeling perspective. Whilst the former analysis, concerning the expressiveness of the discovered axioms, has been conducted (as reported just above) the latter, requiring an actual user study is currently missing, whilst we consider it necessary for coming up with the aforementioned trade-off between expressiveness and utility of the discovered disjointness axioms. Furthermore, additional efforts should be devoted to the scalability of the developed methods that, even if not very limited, still do not appear to be able to scale on the existing KGs.

3.4 Knowledge Graphs for Interpretable Machine Learning

When considering the relation of KGs to deep learning, via KGEs for example, a popular research objective is to use KGs for interpretability. The internal dynamics of DNNs are typically opaque, and there is hope that KGs can be used to help provide (satisfying) explanations of their behavior. The general goal of producing explanations for behavior of machine learning models is sometimes referred to as *explainable AI* (XAI).

As argued in [55], the concepts of explainability and interpretability are intertwined in the context of XAI, because what we really seek is an interpretable explanation. One could, for example, detail exactly the activations of each hidden layer in a neural network to explain why it produced the output from the corresponding input, but this is not a human-interpretable explanation, so is unhelpful for XAI. Despite a strong incentive for interpretable machine learning [114], especially in the area of healthcare [130, 5], and despite significant research attention, how to make complex machine learning model interpretable and explainable remains an open problem [87, 110].

In this section, we give an overview of existing work, and needed future work, on using KGs for interpretable machine learning. We follow our above framework and divide the discussion into two parts: ML for KG and KG for ML.

¹⁸System publicly available at: https://github.com/Giuseppe-Rizzo/TCTnew

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

The former uses ML techniques to augment or construct a KG. With respect to interpretability, the idea is that a KG is a human-readable representation of information. Once it is constructed, it can be used to produce an answer that is highly interpretable, because we can identify the facts and inference rules from which the answer was derived. The problem is that the construction itself, which is often a complex process, remains uninterpretable. The same also applies to work that uses LLMs for KG construction, such as [63, 98], which use BERT-based models to build a clinical KG for medical and financial applications, respectively. Once constructed, the KG can perhaps be used in an interpretable way, but the LLM that constructs it is not interpretable. Methods which then use the KG as input to another stage, may see interpretability gains at those other stages. For example, [16] iteratively use a KG to augment the training data, and then use predictions from augmented training data to extend the KG. However, the initial creation of the KG remains uninterpretable.

In the other direction, there are several works which aim to use KGs to enhance the performance of ML models. There, the possible approaches to using KG for interpretable ML models can, following [152] be divided into three types, pre-model, post-model and in-model.

Pre-model, refers to using the KG as input to a DNN often referred to as "conditioning on the KG", [99]. The idea is that the KG contains higher-quality structured information than images or free-form text, which can then be used by the DNN to solve the given task. This could potentially help interpretability if the network uses an attention mechanism that can be inspected to see which parts of the KG are attended to, as shown by [209] (although, interestingly, the authors were not motivated by explainability in the design of their model). A similar method was later also used by [218]. Similarly, [220] proposed a question-answering model that attends to paths in a KG from a question to the answer, and claims the attention map over these paths constitutes an explanation of the model output. However, these provide at best, only partial interpretability, because it is unclear how/why the model's attention mechanism focuses on the information from the KG that it does.

Post-model, refers to obtaining the output of a ML model, and then invoking a KG to try to produce an explanation for where that output came from. For example, [53] proposes a visual classifier that matches the predicted classes to KG entities, and then uses the KG structure to give an explanation. Similarly, [167] claims to propose an explainable textual entailment model that, after predicting whether one text entails another, finds evidence for this entailment in a KG. The problem with generating post-hoc explanations is that they depend only on the model output and not on the processes internal to the model which produced that output, even though it is precisely the latter that explanations are supposed to shed light on. Two different ML models that produced the same output by very different means would, by methods such as [167] and [53]. automatically receive the same 'explanation'. For example, consider two visual classifiers which both assign the same label to an input image. Suppose one of these classifiers has been trained on and memorized the test set, while the other has actually learned relevant visual features and used these to infer the label. We would surely want the explanation for the outputs of these two classifiers to be different, but if we use only the assigned label to produce an explanation, then they will automatically be the same. Thus, post-model XAI methods that invoke a KG after prediction are precluded from the outset from producing satisfactory explanations, because the explanation is independent of internal model behavior (given the output), which is exactly the thing we want to explain.

In-model, the third manner of enhancing ML models with a KG, involves the KG during the training of the model itself. In the case of DNNs, this faces the difficulty of connecting discrete data from the KG, to a continuous loss function. Beyond some exploratory work, [99, 163], few methods have attempted this approach. Additionally, even if one successfully improved predictive performance, it is not immediately obvious that it would improve interpretability. It is possible

8:20 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

that such an in-model method, were it to be designed, would involve a complex interactive passing of information between a KG and a DNN, which is highly uninterpretable. One such method did explicitly target explanations [160], however this was a bespoke system that requires the KG to consist of part-whole relationships only, as well as additional annotation of the images of object-part classes.

The use of KGs for interpretable ML remains an open problem, either to devise a generalizable method of infusing KG in ML training that demonstrably improves interpretability, or to determine that such a method is not feasible. At the moment, there is interest in the use of KGs for interpretable ML, but we do not have a KG-based method that demonstrably improves interpretability in ML. This gap in the research was also noted by [38]. Moreover, in order for KGs to be of significant help for explainability, we contend that they must be involved internally in the model itself. Using machine learning to generate KGs means that this generation process itself is not interpretable, and invoking the KG after the operation of the machine learning model means that it cannot distinguish between models that produced the same output, even if by very different means.

3.5 Benchmark datasets, and metrics

The ever-expanding number of available methods targeting KG construction, refinement, or usage in ML approaches entails a need for appropriate benchmark datasets and metrics to evaluate their capabilities. Some datasets are considered as *de facto* standards to evaluate approaches developed for KG-related tasks such as FB15k-237 and WN18RR for link prediction, or Citeseer for node classification. However, we claim that current datasets do not suffice for a sound and complete evaluation of the capabilities of developed approaches. Indeed, they present several issues such as:

- unwanted leakages between train and test sets;
- absence of shared patterns between train and test sets;
- lack of necessary characteristics to support the use of background knowledge in ML models (e.g., presence of inverse axioms, hierarchy of classes or properties).
- scattering of datasets across several repositories hindering their discovery and re-usage

In the following, we briefly illustrate and discuss each of these issues and propose possible ways to overcome them.

Several datasets have been made available to the community over the past few years, e.g., by using (fragments of) open KGs [17, 148, 155]. At first, the presence of patterns in train and test sets was regarded with a concern for unwanted leakages. For example, the two datasets FB15k and WN18 were previously widely adopted to evaluate link prediction approaches. It was later discovered that both datasets present data leakage between train and test sets due to inverse relations [43, 181]. A link prediction approach can then easily learn to predict a test triple (t, r^{-1}, h) if triple (h, r, t) is in the train set, where r^{-1} denotes the inverse relation of r. Two filtered versions named FB15k-237 [181] and WN18RR [43] were thus created by filtering such triples, to avoid spurious performance measures. Nevertheless, patterns such as inversion, symmetry, hierarchy or composition and their capture by KGE models are now argued to be of interest, especially if adequately considered in the experimental and evaluation setting [117, 24]. In particular, some authors claim that test triples should be inferable from patterns learned and premises existing in the train set. This imposes additional constraints when constituting datasets but enables to evaluate the ability of KGE models to efficiently model, capture, and implement those patterns [117, 24]. In this view, train sets should contain samples of premises and conclusions of the considered patterns to learn. Test sets should contain conclusions that can be inferred from patterns learned and premises in train sets. This empirical evaluation is of interest to substantiate some theoretical guarantees of model design or, conversely, to outline some unexpected

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

abilities. For instance, some KGE models such as ROTATE [175] are theoretically designed to capture patterns such as symmetry, antisymmetry, inversion, and composition and should be evaluated accordingly. It follows that detecting (and potentially removing) some patterns is an important step of dataset preprocessing. For now, detection (and removal) of inverses is performed statistically, as featured in the AYNEC/AYNEXT system [10, 168]. They detect whether two relations r_1 and r_2 are inverses of each other if some proportions of triples involving r_1 have their counterpart involving r_2 . The identification of other patterns also relies on statistical approaches such as rule mining for their detection [117]. It is noteworthy that ontologies provide definitions of inverses, symmetric predicates and hierarchies of properties and classes. Hence, besides statistical approaches, ontological axioms should be taken into account to detect or implement patterns. Indeed, train sets could be completely based on ontological axioms and deductive reasoning to feature the needed patterns to learn or remove some unwanted ones.

Also, we previously outlined the interest in studying the role and usage of background knowledge in ML models. For now, datasets are often regarded as simple graph data without consideration for (or association with) additional knowledge potentially provided by ontologies. Beside improving datasets by adding triples respecting patterns or removing unwanted ones, the association of ontological axioms with datasets could support the development of learning techniques, settings, and models that consider them, following our claim for further consideration of knowledge in KGE models. To illustrate, instead of enriching datasets with triples respecting patterns, models could be evaluated on their ability to consider patterns stated by ontological axioms to predict missing triples in the test set. It is noteworthy that knowledge is already leveraged to enrich the training process in some proposals. For instance, Type-Constrained Negative Sampling [97] replaces the head or the tail of a triple with an entity of the same type when generating negative triples. d'Amato et al. [36] use a reasoner to deduce additional triples from axioms defining equivalent classes, equivalent properties, inverses, or subclasses. Similarly, Iana and Paulheim [79] test whether materializing all triples induced by transitive properties, symmetric properties, and sub-properties leads to improved embeddings. Ontological information is also needed to evaluate the semantics captured by KGE models. In this view, Jain et al. [80] relies on the existence of types of entities. They learn embeddings on the YAGO3-10 and FB15k-237 datasets with various KGE models and then use these embeddings to predict entity types with classification or clustering approaches. Their analysis shows that semantic representation in the embedding space is not universal across models. In a similar fashion, the DLCC node classification benchmark was introduced to evaluate the capability of classification approaches to reproduce classes defined by Description Logic Constructors [148]. For example, the constructor $\exists r. \top$ is used to group nodes having a particular outgoing relation. Interestingly, they propose two gold standards: one based on the real graph DBpedia and another synthetic standard that is generated by a gold standard generator publicly available. The analysis of ontological information captured by KGE models also motivates new metrics besides traditional metrics such as precision, recall, Hits@K, or Mean Reciprocal Rank. For example, Hubert *et al.* introduced the Sem@K metric [76, 75, 77] to measure the number of predicted triples that respect domain and range of relations among the top-K predicted triples. This metric can thus be seen as measuring the ability of KGE models to capture the semantic profiles of relations. The aforementioned work highlights an interest in using ontological information in KGE model design, learning process, or evaluation. Consequently, we advocate for the further development of benchmark datasets that include various ontological axioms, separately or combined. The availability of such datasets would in turn encourage and support the development of neuro-symbolic methods leveraging such axioms. However, it is noteworthy that not all current benchmarks offer the ontological information that is needed by particular approaches. That is why some authors resort to synthetic KG generators [124, 148],

8:22 Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

sometimes with a fixed ontology. To further this research direction, synthetic KG generators should be enriched with the synthetic generation of schemas with different levels of expressiveness and constructs. This would allow an on-demand generation of specific ontologies and knowledge graphs featuring the needed ontological axioms.

To further support the research community, we also call for a more systematic approach in the development, characterization, and collection of benchmark datasets. For now, benchmark datasets (or versions of) are scattered across several repositories such as GitHub or Zenodo. This leads to some of them being widely adopted (e.g., FB15k-237) and some others to be only re-used in a few papers. A unified repository, similar to the UCI Machine Learning repository¹⁹, is needed to encourage their reuse and adoption by the community. Constituting such a repository first requires to crawl (semi-)automatically several sources, including GitHub or Zenodo, and links in papers available in digital libraries, arXiv, or PapersWithCode. Additionally, given that different approaches may leverage different characteristics of datasets (e.g., DL constructors [148], sub-properties [36, 79], domain and range of predicates [78], patterns in train and test sets [117]), datasets should be qualified w.r.t. the presence or absence of these characteristics. This would help researchers and developers to select suitable datasets to evaluate their approaches. To this aim. scalable automatic methods need to be developed to crawl and analyze KG-based datasets in the wild and detect a broad range of characteristics including those aforementioned. This qualification process will produce metadata that enrich usual dataset metadata such as providers, or license. To represent these new dataset metadata, an additional perspective thus lies in extending existing ontologies describing datasets (e.g., VoID, DCAT). Ontologies introduced to describe mining processes and their features such as DMOP [88] could offer sources of inspiration in this matter.

4 Conclusion

The interrelation between knowledge graphs and machine learning has been supporting advances in both fields. Machine learning methods have indeed allowed efficient construction and refinement of large knowledge graphs. Conversely, knowledge graphs have been leveraged in various machine learning tasks to improve performance, e.g., in question answering, or image classification.

However, this interrelation still does not consider parts of knowledge graphs and ML methods summarized in Table 1 that we deem to be important and to offer promising research directions. In particular, we believe KGs constitute a major structure for prompting Large Languages Models and could allow researchers to formalize interactions (e.g., providing contexts in prompts, or deciding prompt sequencing). Additionally, rich semantics of KGs and knowledge actionable by various forms of reasoning capabilities could benefit KGE models through a deeper integration. This could lead to improved performance, or a better handling or generation of informative negatives which are essential in model learning. Regarding informative negatives, we also believe that symbol-based ML, which is particularly adapted to the symbolic structure of KGs, could provide an interesting perspective, especially with the mining of disjointness axioms. KGs are human- and machine-interpretable, and thus are a promising structure on which to construct in-model interpretable ML models. Nevertheless, the infusion of KGs directly within ML models and an actual demonstration of the production of more interpretable and reliable explanations are open challenges. To assess improved performance or interpretability of ML models thanks to KGs. extensive experimental evaluations are needed, which require datasets showcasing different levels of semantics, or schema constructs to assess their individual impacts. That is why, we also call for a more systematic collection and characterization of datasets, as well as the creation of synthetic KG generators to enrich the collection of available benchmarks.

¹⁹ https://archive.ics.uci.edu/

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

In our view, such integrations and interactions open promising challenges to foster both fields of research. We believe these directions to be stepping stones to place KGs as central assets towards neuro-symbolic and explainable AI.

— References -

- 1 Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. BoxE: A Box Embedding Model for Knowledge Base Completion. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020., 2020. URL: https: //proceedings.neurips.cc/paper/2020/hash/ 6dbbe6abe5f14af882ff977fc3f35501-Abstract. html.
- 2 Leonard Adolphs, Shehzaad Dhuliawala, and Thomas Hofmann. How to Query Language Models?, 2021. doi:10.48550/arXiv.2108.01928.
- 3 Aishwarya Agrawal, Jiasen Lu, Stanislaw Antol, Margaret Mitchell, C. Lawrence Zitnick, Devi Parikh, and Dhruv Batra. VQA: Visual Question Answering - www.visualqa.org. International Journal of Computer Vision, 123(1):4–31, 2017. doi:10.1007/S11263-016-0966-6.
- 4 Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26-28, 1993, pages 207–216. ACM Press, 1993. doi:10.1145/170035.170072.
- 5 Muhammad Aurangzeb Ahmad, Carly Eckert, and Ankur Teredesai. Interpretable Machine Learning in Healthcare. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2018, Washington, DC, USA, August 29 -September 01, 2018, pages 559–560. ACM, 2018. doi:10.1145/3233547.3233667.
- 6 Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. Journal of Machine Learning Research, 22:82:1–82:6, 2021. URL: http: //jmlr.org/papers/v22/20-825.html.
- 7 Safa Alsaidi, Amandine Decker, Puthineath Lay, Esteban Marquer, Pierre-Alexandre Murena, and Miguel Couceiro. A Neural Approach for Detecting Morphological Analogies. In 8th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2021, Porto, Portugal, October 6-9, 2021, pages 1–10. IEEE, 2021. doi:10.1109/DSAA53316.2021.9564186.
- 8 Hiba Arnaout, Simon Razniewski, and Gerhard Weikum. Enriching Knowledge Bases with Interesting Negative Statements. In Conference on Automated Knowledge Base Construction, AKBC 2020, Virtual, June 22-24, 2020, 2020. doi: 10.24432/C5101K.
- 9 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. DBpedia: A Nucleus for a Web of Open

Data. In The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007, volume 4825 of Lecture Notes in Computer Science, pages 722-735. Springer, 2007. doi:10.1007/ 978-3-540-76298-0_52.

- 10 Daniel Ayala, Agustín Borrego, Inma Hernández, Carlos R. Rivero, and David Ruiz. AYNEC: All You Need for Evaluating Completion Techniques in Knowledge Graphs. In The Semantic Web - 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2-6, 2019, Proceedings, volume 11503 of Lecture Notes in Computer Science, pages 397–411. Springer, 2019. doi:10.1007/978-3-030-21348-0_26.
- 11 Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *Description Logic Handbook*, 2nd edition. Cambridge University Press, 2010. doi:10.1017/CB09780511711787.
- 12 Samy Badreddine, Artur S. d'Avila Garcez, Luciano Serafini, and Michael Spranger. Logic Tensor Networks. Artificial Intelligence, 303:103649, 2022. doi:10.1016/J.ARTINT.2021.103649.
- 13 Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. Multi-relational Poincaré Graph Embeddings. In Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019., pages 4465-4475, 2019. URL: https: //proceedings.neurips.cc/paper/2019/hash/ f8b932c70d0b2e6bf071729a4fa68dfc-Abstract. html.
- 14 Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 41(2):423–443, 2019. doi:10.1109/TPAMI. 2018.2798607.
- 15 Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity, 2023. doi:10.48550/ARXIV.2302.04023.
- 16 Shreyansh P. Bhatt, Amit P. Sheth, Valerie L. Shalin, and Jinjin Zhao. Knowledge Graph Semantic Enhancement of Input Data for Improving AI. *IEEE Internet Computing*, 24(2):66–72, 2020. doi:10.1109/MIC.2020.2979620.
- 17 Peter Bloem, Xander Wilcke, Lucas van Berkel, and Victor de Boer. kgbench: A Collection of Knowledge Graph Datasets for Evaluating Relational and Multimodal Machine Learning. In The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021,

Proceedings, volume 12731 of *Lecture Notes in Computer Science*, pages 614–630. Springer, 2021. doi:10.1007/978-3-030-77385-4_37.

- 18 Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. Freebase: A Shared Database of Structured General Human Knowledge. In Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada, pages 1962–1963. AAAI Press, 2007. URL: http://www.aaai.org/Library/AAAI/ 2007/aaai07-355.php.
- 19 Piero Andrea Bonatti, Stefan Decker, Axel Polleres, and Valentina Presutti. Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Dagstuhl Reports, 8(9):29–111, 2018. doi:10.4230/DAGREP.8.9.29.
- 20 Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States, pages 2787-2795, 2013. URL: https: //proceedings.neurips.cc/paper/2013/hash/ 1cecc7a77928ca8133fa24680a88d2f9-Abstract. html.
- 21 Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021. doi:10.48550/arXiv.2104.13478.
- 22 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020. URL: https: //proceedings.neurips.cc/paper/2020/hash/ 1457c0d6bfcb4967418bfb8ac142f64a-Abstract. html.
- 23 Hongyun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang. A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018. doi: 10.1109/TKDE.2018.2807452.
- 24 Yixin Cao, Xiang Ji, Xin Lv, Juanzi Li, Yonggang Wen, and Hanwang Zhang. Are Missing Links Predictable? An Inferential Benchmark for Knowledge Graph Completion. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers),

Virtual Event, August 1-6, 2021, pages 6855–6865. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.ACL-LONG.534.

- 25 Shruthi Chari, Oshani Seneviratne, Mohamed Ghalwash, Sola Shirai, Daniel M. Gruen, Pablo Meyer, Prithwish Chakraborty, and Deborah L McGuinness. Explanation Ontology: A General-Purpose, Semantic Representation for Supporting User-Centered Explanations. *Semantic Web*, (to appear), 2023. doi:10.3233/SW-233282.
- 26 Michel Chein and Marie-Laure Mugnier. Graphbased Knowledge Representation - Computational Foundations of Conceptual Graphs. Advanced Information and Knowledge Processing. Springer, 2009. doi:10.1007/978-1-84800-286-9.
- 27 Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Ian Horrocks, Jeff Z. Pan, and Huajun Chen. Knowledge-aware Zero-Shot Learning: Survey and Perspective. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJ-CAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021, pages 4366–4373. ijcai.org, 2021. doi:10.24963/IJCAI.2021/597.
- 28 Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Jeff Z. Pan, Yuan He, Wen Zhang, Ian Horrocks, and Huajun Chen. Zero-Shot and Few-Shot Learning With Knowledge Graphs: A Comprehensive Survey. *Proceedings of the IEEE*, 111(6):653–685, 2023. doi:10.1109/JPR0C.2023.3279374.
- 29 Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. Expert Systems with Applications, 141, 2020. doi:10.1016/J.ESWA.2019.112948.
- 30 Zhuo Chen, Yufeng Huang, Jiaoyan Chen, Yuxia Geng, Yin Fang, Jeff Z. Pan, Ningyu Zhang, and Wen Zhang. LaKo: Knowledge-driven Visual Question Answering via Late Knowledge-to-Text Injection. In Proceedings of the 11th International Joint Conference on Knowledge Graphs, IJCKG 2022, Hangzhou, China, October 27-28, 2022, pages 20–29. ACM, 2022. doi:10.1145/3579051.3579053.
- 31 Michael Cochez, Petar Ristoski, Simone Paolo Ponzetto, and Heiko Paulheim. Global RDF Vector Space Embeddings. In The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part I, volume 10587 of Lecture Notes in Computer Science, pages 190–207. Springer, 2017. doi:10.1007/978-3-319-68288-4_12.
- 32 Roberto Confalonieri and Giancarlo Guizzardi. On the Multiple Roles of Ontologies in Explainable AI, 2023. doi:10.48550/ARXIV.2311.04778.
- 33 Claudia d'Amato. Machine Learning for the Semantic Web: Lessons learnt and next research directions. Semantic Web, 11(1):195-203, 2020. doi:10.3233/SW-200388.
- 34 Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito. Query Answering and Ontology Population: An Inductive Approach. In The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008, Proceedings, volume 5021 of Lecture Notes in Computer Science, pages 288–302. Springer, 2008. doi:10.1007/978-3-540-68234-9_23.

C. d'Amato, L. Mahon, P. Monnin, and G. Stamou

- 35 Claudia d'Amato, Nicola Flavio Quatraro, and Nicola Fanizzi. Embedding Models for Knowledge Graphs Induced by Clusters of Relations and Background Knowledge. In Inductive Logic Programming - 30th International Conference, ILP 2021, Virtual Event, October 25-27, 2021, Proceedings, volume 13191 of Lecture Notes in Computer Science, pages 1–16. Springer, 2021. doi:10.1007/978-3-030-97454-1_1.
- 36 Claudia d'Amato, Nicola Flavio Quatraro, and Nicola Fanizzi. Injecting Background Knowledge into Embedding Models for Predictive Tasks on Knowledge Graphs. In The Semantic Web -18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings, volume 12731 of Lecture Notes in Computer Science, pages 441–457. Springer, 2021. doi:10.1007/ 978-3-030-77385-4_26.
- 37 Claudia d'Amato, Andrea G. B. Tettamanzi, and Duc Minh Tran. Evolutionary Discovery of Multi-relational Association Rules from Ontological Knowledge Bases. In Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, volume 10024 of Lecture Notes in Computer Science, pages 113–128, 2016. doi:10.1007/978-3-319-49004-5_8.
- 38 Tirtharaj Dash, Sharad Chitlangia, Aditya Ahuja, and Ashwin Srinivasan. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports*, 12(1):1040, 2022. doi:10.1038/s41598-021-04590-0.
- 39 Jia Deng, Nan Ding, Yangqing Jia, Andrea Frome, Kevin Murphy, Samy Bengio, Yuan Li, Hartmut Neven, and Hartwig Adam. Large-Scale Object Classification Using Label Relation Graphs. In Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I, volume 8689 of Lecture Notes in Computer Science, pages 48–64. Springer, 2014. doi:10.1007/978-3-319-10590-1_4.
- 40 Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA, pages 248–255. IEEE Computer Society, 2009. doi:10.1109/CVPR.2009. 5206848.
- 41 Jia Deng, Jonathan Krause, Alexander C. Berg, and Li Fei-Fei. Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012, pages 3450–3457. IEEE Computer Society, 2012. doi: 10.1109/CVPR.2012.6248086.
- 42 Edmund Dervakos, Konstantinos Thomas, Giorgos Filandrianos, and Giorgos Stamou. Choose your data wisely: A framework for semantic counterfactuals. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China, pages 382–390. ijcai.org, 2023. doi:10.24963/IJCAI.2023/43.

- 43 Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D Knowledge Graph Embeddings. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 1811–1818. AAAI Press, 2018. doi:10.1609/AAAI.V32I1.11573.
- 44 Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery* and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014, pages 601–610. ACM, 2014. doi:10.1145/2623330.2623623.
- 45 Alaaeldin El-Nouby, Shikhar Sharma, Hannes Schulz, R. Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Yoshua Bengio, and Graham W. Taylor. Tell, Draw, and Repeat: Generating and Modifying Images Based on Continual Linguistic Instruction. In 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019, pages 10303-10311. IEEE, 2019. doi:10.1109/ICCV.2019.01040.
- 46 Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. DL-FOIL Concept Learning in Description Logics. In Inductive Logic Programming, 18th International Conference, ILP 2008, Prague, Czech Republic, September 10-12, 2008, Proceedings, volume 5194 of Lecture Notes in Computer Science, pages 107–121. Springer, 2008. doi:10.1007/978-3-540-85928-4_12.
- 47 Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. Metric-based stochastic conceptual clustering for ontologies. *Information Systems*, 34(8):792–806, 2009. doi:10.1016/J.IS.2009.03.008.
- 48 Shangbin Feng, Vidhisha Balachandran, Yuyang Bai, and Yulia Tsvetkov. FactKB: Generalizable Factuality Evaluation using Language Models Enhanced with Factual Knowledge, 2023. doi: 10.48550/ARXIV.2305.08281.
- 49 Yanwei Fu, Tao Xiang, Yu-Gang Jiang, Xiangyang Xue, Leonid Sigal, and Shaogang Gong. Recent Advances in Zero-Shot Recognition: Toward Data-Efficient Understanding of Visual Content. *IEEE Signal Processing Magazine*, 35(1):112–125, 2018. doi:10.1109/MSP.2017.2763441.
- 50 Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. *The VLDB Journal*, 24(6):707–730, 2015. doi:10.1007/ S00778-015-0394-1.
- 51 Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, pages 413-422. International World Wide Web

- 52 Yang Gao, Yi-Fan Li, Yu Lin, Hang Gao, and Latifur Khan. Deep Learning on Knowledge Graph for Recommender System: A Survey, 2020. doi:10.48550/arXiv.2004.00387.
- 53 Yuxia Geng, Jiaoyan Chen, Zhiquan Ye, Zonggang Yuan, Wei Zhang, and Huajun Chen. Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs. Semantic Web, 12(5):741–765, 2021. doi:10.3233/ SW-210435.
- 54 Lise Getoor and Ben Taskar, editors. Introduction to Statistical Relational Learning. MIT Press, 2007.
- 55 Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael A. Specter, and Lalana Kagal. Explaining Explanations: An Overview of Interpretability of Machine Learning. In 5th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2018, Turin, Italy, October 1-3, 2018, pages 80–89. IEEE, 2018. doi: 10.1109/DSAA.2018.00018.
- 56 Olga Golovneva, Moya Chen, Spencer Poff, Martin Corredor, Luke Zettlemoyer, Maryam Fazel-Zarandi, and Asli Celikyilmaz. ROSCOE: A Suite of Metrics for Scoring Step-by-Step Reasoning. In The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. URL: https://openreview.net/pdf?id=xYlJRpzZtsY.
- 57 Ramanathan V. Guha. Towards A Model Theory for Distributed Representations. In 2015 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 22-25, 2015. AAAI Press, 2015. URL: http://www.aaai.org/ocs/ index.php/SSS/SSS15/paper/view/10220.
- 58 Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys, 51(5):93:1– 93:42, 2019. doi:10.1145/3236009.
- 59 Lin Guo and Qun Dai. Graph Clustering via Variational Graph Embedding. Pattern Recognition, 122:108334, 2022. doi:10.1016/J.PATCOG.2021. 108334.
- 60 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly Embedding Knowledge Graphs and Logical Rules. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, pages 192–202. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1019.
- 61 Wenzhong Guo, Jianwen Wang, and Shiping Wang. Deep Multimodal Representation Learning: A Survey. *IEEE Access*, 7:63373–63394, 2019. doi: 10.1109/ACCESS.2019.2916887.
- 62 Víctor Gutiérrez-Basulto and Steven Schockaert. From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules. In Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018, pages 379–388. AAAI

Press, 2018. URL: https://aaai.org/ocs/index.php/KR/KR18/paper/view/18013.

- 63 Ayoub Harnoune, Maryem Rhanoui, Mounia Mikram, Siham Yousfi, Zineb Elkaimbillah, and Bouchra El Asri. BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis. *Computer Methods and Programs in Biomedicine Update*, 1:100042, 2021. doi:10.1016/j.cmpbup.2021.100042.
- 64 Frederick Hayes-Roth, Donald A Waterman, and Douglas B Lenat. Building expert systems. Addison-Wesley Longman Publishing Co., Inc., 1983.
- 65 Bin He, Daoyuan Zhou, Jingjing Xiao, Xiangyang Jiang, Qun Liu, Nianwen J. Yuan, and Tao Xu. BERT-MK: Integrating Graph Contextualized Knowledge into Pre-trained Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2281–2290, Online, nov 2020. Association for Computational Linguistics. doi:10.18653/v1/2020.findings-emnlp. 207.
- 66 Feijuan He, Yaxian Wang, Xianglin Miao, and Xia Sun. Interpretable visual reasoning: A survey. Image and Vision Computing, 112:104194, 2021. doi:10.1016/J.IMAVIS.2021.104194.
- 67 Shizhu He, Kang Liu, Guoliang Ji, and Jun Zhao. Learning to Represent Knowledge Graphs with Gaussian Embedding. In Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015, pages 623–632. ACM, 2015. doi:10.1145/2806416. 2806502.
- 68 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021. doi:10.2200/ S01125ED1V01Y202109DSK022.
- 69 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. ACM Computing Surveys, 54(4):1–37, 2021. doi:10.1145/3447772.
- 70 Ian Horrocks, Peter. F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. SWRL: A semantic web rule language combining OWL and RuleML, 2004. URL: http: //www.daml.org/2004/04/swrl/rules-all.html.
- 71 Yang Hu, Adriane Chapman, Guihua Wen, and Wendy Hall. What Can Knowledge Bring to Machine Learning? - A Survey of Low-shot Learning for Structured Data. ACM Transactions on Intelligent Systems and Technology, 13(3):48:1–48:45, 2022. doi:10.1145/3510030.
- 72 Zhiwei Hu, Víctor Gutiérrez-Basulto, Zhiliang Xiang, Xiaoli Li, Ru Li, and Jeff Z. Pan. Typeaware Embeddings for Multi-Hop Reasoning over

Knowledge Graphs. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022, pages 3078–3084. ijcai.org, 2022. doi:10.24963/IJCAI.2022/427.

- 73 Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. Knowledge graph embedding based question answering. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019, pages 105–113. ACM, 2019. doi:10.1145/3289600.3290956.
- 74 Zhicheng Huang, Zhaoyang Zeng, Yupan Huang, Bei Liu, Dongmei Fu, and Jianlong Fu. Seeing Out of the Box: End-to-End Pre-Training for Vision-Language Representation Learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25,* 2021, pages 12976–12985. Computer Vision Foundation / IEEE, 2021. doi:10.1109/CVPR46437. 2021.01278.
- 75 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Knowledge Graph Embeddings for Link Prediction: Beware of Semantics! In Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Virtual Conference, online, October 24, 2022, volume 3342 of CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3342/paper-4.pdf.
- 76 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. New Strategies for Learning Knowledge Graph Embeddings: The Recommendation Case. In Knowledge Engineering and Knowledge Management - 23rd International Conference, EKAW 2022, Bolzano, Italy, September 26-29, 2022, Proceedings, volume 13514 of Lecture Notes in Computer Science, pages 66–80. Springer, 2022. doi:10.1007/978-3-031-17105-5_5.
- 77 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Sem@K: Is my knowledge graph embedding model semantic-aware? Semantic Web, (to appear), 2023. doi:10.48550/arXiv. 2301.05601.
- 78 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Treat Different Negatives Differently: Enriching Loss Functions with Domain and Range Constraints for Link Prediction, 2023. doi:10.48550/arXiv.2303.00286.
- 79 Andreea Iana and Heiko Paulheim. More is not Always Better: The Negative Impact of A-box Materialization on RDF2vec Knowledge Graph Embeddings. In Proceedings of the CIKM 2020 Workshops co-located with 29th ACM International Conference on Information and Knowledge Management (CIKM 2020), Galway, Ireland, October 19-23, 2020, volume 2699 of CEUR Workshop Proceedings. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2699/paper05.pdf.
- 80 Nitisha Jain, Jan-Christoph Kalo, Wolf-Tilo Balke, and Ralf Krestel. Do Embeddings Actually Capture Knowledge Graph Semantics? In The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings, volume 12731 of Lecture Notes in Com-

puter Science, pages 143–159. Springer, 2021. doi:10.1007/978-3-030-77385-4_9.

- 81 Lucas Jarnac, Miguel Couceiro, and Pierre Monnin. Relevant Entity Selection: Knowledge Graph Bootstrapping via Zero-Shot Analogical Pruning. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023, pages 934–944. ACM, 2023. doi:10.1145/3583780.3615030.
- 82 Mirantha Jayathilaka, Tingting Mu, and Uli Sattler. Visual-Semantic Embedding Model Informed by Structured Knowledge. In Proceedings of the 9th European Starting AI Researchers' Symposium 2020 co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), Santiago Compostela, Spain, August, 2020, volume 2655 of CEUR Workshop Proceedings. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2655/ paper23.pdf.
- 83 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Net*works and Learning Systems, 33(2):494–514, 2022. doi:10.1109/TNNLS.2021.3070843.
- 84 Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12):248:1–248:38, 2023. doi:10.1145/3571730.
- 85 Ziwei Ji, Zihan Liu, Nayeon Lee, Tiezheng Yu, Bryan Wilie, Min Zeng, and Pascale Fung. RHO: Reducing Hallucination in Open-domain Dialogues with Knowledge Grounding. In *Findings of the* Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023, pages 4504–4522. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.FINDINGS-ACL. 275.
- 86 J. Józefowska, A Lawrynowicz, and T. Lukaszewski. The role of semantics in mining frequent patterns from knowledge bases in description logics with rules. *Theory and Practice of Logic Programming*, 10(3):251–289, 2010. doi:10.1017/S1471068410000098.
- 87 Serhiy Kandul, Vincent Micheli, Juliane Beck, Markus Kneer, Thomas Burri, François Fleuret, and Markus Christen. Explainable AI: A Review of the Empirical Literature. SSRN 4325219, 2023. doi:10.2139/ssrn.4325219.
- 88 C. Maria Keet, Agnieszka Lawrynowicz, Claudia d'Amato, Alexandros Kalousis, Phong Nguyen, Raúl Palma, Robert Stevens, and Melanie Hilario. The Data Mining OPtimization Ontology. Journal of Web Semantics, 32:43–53, 2015. doi:10.1016/J.WEBSEM.2015.01.001.
- 89 Mayank Kejriwal. Domain-Specific Knowledge Graph Construction. Springer Briefs in Computer Science. Springer, 2019. doi:10.1007/ 978-3-030-12375-8.
- **90** Wonjae Kim, Bokyung Son, and Ildoo Kim. ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision. In *Proceedings of the*

38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, volume 139 of Proceedings of Machine Learning Research, pages 5583-5594. PMLR, 2021. URL: http: //proceedings.mlr.press/v139/kim21k.html.

- 91 Daphne Koller and Nir Friedman, editors. Probabilistic graphical models: principles and techniques. MIT Press, 2009. URL: http://mitpress.mit.edu/catalog/item/ default.asp?ttype=2&tid=11886.
- 92 Wei Kun Kong, Xin Liu, Teeradaj Racharak, Guanqun Sun, Jianan Chen, Qiang Ma, and Le-Minh Nguyen. WeExt: A Framework of Extending Deterministic Knowledge Graph Embedding Models for Embedding Weighted Knowledge Graphs. *IEEE Access*, 11:48901–48911, 2023. doi:10.1109/ACCESS.2023.3276319.
- 93 Lili Kotlerman, Ido Dagan, Bernardo Magnini, and Luisa Bentivogli. Textual entailment graphs. Natural Language Engineering, 21(5):699–724, 2015. doi:10.1017/S1351324915000108.
- 94 Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. International Journal of Computer Vision, 123(1):32– 73, 2017. doi:10.1007/S11263-016-0981-7.
- 95 Anastasia Kritharoula, Maria Lymperaiou, and Giorgos Stamou. Large Language Models and Multimodal Retrieval for Visual Word Sense Disambiguation, 2023. doi:10.48550/ARXIV.2310. 14025.
- 96 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the* ACM, 60(6):84–90, 2017. doi:10.1145/3065386.
- 97 Denis Krompaß, Stephan Baier, and Volker Tresp. Type-Constrained Representation Learning in Knowledge Graphs. In *The Semantic Web - ISWC* 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I, volume 9366 of Lecture Notes in Computer Science, pages 640–655. Springer, 2015. doi:10.1007/978-3-319-25007-6_37.
- 98 Abhijeet Kumar, Abhishek Pandey, Rohit Gadia, and Mridul Mishra. Building knowledge graph using pre-trained language model for learning entity-aware relationships. In 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), pages 310–315. IEEE, 2020. doi:10.1109/GUCON48875. 2020.9231227.
- 99 Ugur Kursuncu, Manas Gaur, and Amit P. Sheth. Knowledge Infused Learning (K-IL): Towards Deep Incorporation of Knowledge in Deep Learning. In Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice, AAAI-MAKE 2020, Palo Alto, CA, USA, March 23-25, 2020, Volume I, volume 2600 of CEUR Workshop Proceedings. CEUR-WS.org, 2020. URL: https:// ceur-ws.org/Vol-2600/paper14.pdf.
- 100 Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen ob-

ject classes by between-class attribute transfer. In 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA, pages 951–958. IEEE Computer Society, 2009. doi:10.1109/CVPR.2009.5206594.

- 101 Philippe Langlais, François Yvon, and Pierre Zweigenbaum. Improvements in Analogical Learning: Application to Translating Multi-Terms of the Medical Domain. In EACL 2009, 12th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Athens, Greece, March 30 - April 3, 2009, pages 487–495. The Association for Computer Linguistics, 2009. URL: https://aclanthology.org/E09-1056/.
- 102 Jens Lehmann, Sören Auer, Lorenz Bühmann, and Sebastian Tramp. Class expression learning for ontology engineering. *Journal of Web Semantics*, 9(1):71-81, 2011. doi:10.1016/J.WEBSEM.2011. 01.001.
- 103 Jens Lehmann and Lorenz Bühmann. ORE - A Tool for Repairing and Enriching Knowledge Bases. In The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part II, volume 6497 of Lecture Notes in Computer Science, pages 177–193. Springer, 2010. doi:10.1007/ 978-3-642-17749-1_12.
- 104 Douglas B. Lenat, Alan Borning, David W. Mc-Donald, Craig Taylor, and Steven Weyer. Knoesphere: Building Expert Systems With Encyclopedic Knowledge. In Proceedings of the 8th International Joint Conference on Artificial Intelligence. Karlsruhe, FRG, August 1983, pages 167–169. William Kaufmann, 1983. URL: http://ijcai. org/Proceedings/83-1/Papers/034.pdf.
- 105 Adam Lerer, Ledell Wu, Jiajun Shen, Timothée Lacroix, Luca Wehrstedt, Abhijit Bose, and Alex Peysakhovich. Pytorch-BigGraph: A Large Scale Graph Embedding System. In Proceedings of Machine Learning and Systems 2019, MLSys 2019, Stanford, CA, USA, March 31 - April 2, 2019. mlsys.org, 2019. URL: https://proceedings.mlsys. org/book/282.pdf.
- 106 Paul Lerner, Olivier Ferret, and Camille Guinaudeau. Multimodal Inverse Cloze Task for Knowledge-Based Visual Question Answering. In Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part I, volume 13980 of Lecture Notes in Computer Science, pages 569–587. Springer, 2023. doi:10.1007/978-3-031-28244-7_36.
- 107 Guohao Li, Xin Wang, and Wenwu Zhu. Boosting Visual Question Answering with Contextaware Knowledge Aggregation. In MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020, pages 1227–1235. ACM, 2020. doi:10.1145/3394171.3413943.
- 108 Shaobo Li, Xiaoguang Li, Lifeng Shang, Zhenhua Dong, Chengjie Sun, Bingquan Liu, Zhenzhou Ji, Xin Jiang, and Qun Liu. How Pre-trained Language Models Capture Factual Knowledge? A

Causal-Inspired Analysis. In Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 1720– 1732. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.FINDINGS-ACL.136.

- 109 Shaobo Li, Xiaoguang Li, Lifeng Shang, Chengjie Sun, Bingquan Liu, Zhenzhou Ji, Xin Jiang, and Qun Liu. Pre-training Language Models with Deterministic Factual Knowledge. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 11118-11131. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. EMNLP-MAIN.764.
- 110 Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12):3197–3234, 2022. doi:10.1007/S10115-022-01756-8.
- 111 Jason Liartis, Edmund Dervakos, Orfeas Menis-Mastromichalakis, Alexandros Chortaras, and Giorgos Stamou. Searching for explanations of black-box classifiers in the space of semantic queries. Semantic Web, (to appear), 2023. doi: 10.3233/SW-233469.
- 112 Yankai Lin, Xu Han, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Knowledge Representation Learning: A Quantitative Review, 2018. doi: 10.48550/arXiv.1812.10901.
- 113 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA, pages 2181–2187. AAAI Press, 2015. doi:10.1609/AAAI.V29I1.9491.
- 114 Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1):18, 2021. doi:10.3390/E23010018.
- 115 Hanxiao Liu, Yuexin Wu, and Yiming Yang. Analogical Inference for Multi-relational Embeddings. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, volume 70 of Proceedings of Machine Learning Research, pages 2168–2178. PMLR, 2017. URL: http://proceedings.mlr.press/v70/liu17d.html.
- 116 Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. Generated Knowledge Prompting for Commonsense Reasoning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 3154-3169. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. ACL-LONG.225.
- 117 Shuwen Liu, Bernardo Cuenca Grau, Ian Horrocks, and Egor V. Kostylev. Revisiting Inferential Benchmarks for Knowledge Graph Completion. In Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR 2023, Rhodes, Greece,

September 2-8, 2023, pages 461-471, 2023. doi: 10.24963/KR.2023/45.

- 118 Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-BERT: Enabling Language Representation with Knowledge Graph. In The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, pages 2901–2908. AAAI Press, 2020. doi:10.1609/AAAI.V34I03.5681.
- 119 Maria Lymperaiou and Giorgos Stamou. The Contribution of Knowledge in Visiolinguistic Learning: A Survey on Tasks and Challenges. In Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023, volume 3433 of CEUR Workshop Proceedings. CEUR-WS.org, 2023. URL: https:// ceur-ws.org/Vol-3433/paper18.pdf.
- 120 Louis Mahon, Eleonora Giunchiglia, Bowen Li, and Thomas Lukasiewicz. Knowledge Graph Extraction from Videos. In 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020, Miami, FL, USA, December 14-17, 2020, pages 25–32. IEEE, 2020. doi:10.1109/ICMLA51294.2020.00014.
- 121 Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, pages 9802–9822. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.546.
- 122 Kenneth Marino, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. OK-VQA: A Visual Question Answering Benchmark Requiring External Knowledge. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR* 2019, Long Beach, CA, USA, June 16-20, 2019, pages 3195–3204. Computer Vision Foundation / IEEE, 2019. doi:10.1109/CVPR.2019.00331.
- 123 Christian Meilicke, Melisachew Wudage Chekol, Manuel Fink Patrick Betz, and Heiner Stuckeschmidt. Anytime bottom-up rule learning for large-scale knowledge graph completion. The VLDB Journal, 2023. doi:10.1007/ s00778-023-00800-5.
- 124 André Melo and Heiko Paulheim. Synthesizing Knowledge Graphs for Link and Type Prediction Benchmarking. In The Semantic Web -14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I, volume 10249 of Lecture Notes in Computer Science, pages 136–151, 2017. doi: 10.1007/978-3-319-58068-5_9.
- 125 Zaiqiao Meng, Fangyu Liu, Ehsan Shareghi, Yixuan Su, Charlotte Collins, and Nigel Collier. Rewire-then-Probe: A Contrastive Recipe for

Probing Biomedical Knowledge of Pre-trained Language Models. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 4798– 4810. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.ACL-LONG.329.

- 126 Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning. Journal of Artificial Intelligence Research, 32:793–824, 2008. doi:10.1613/JAIR.2519.
- 127 George A. Miller. WordNet: A Lexical Database for English. Communications of the ACM, 38(11):39-41, 1995. doi:10.1145/219717.219748.
- 128 Pasquale Minervini, Luca Costabello, Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Regularizing Knowledge Graph Embeddings via Equivalence and Inversion Axioms. In Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part I, volume 10534 of Lecture Notes in Computer Science, pages 668–683. Springer, 2017. doi:10.1007/978-3-319-71249-9_40.
- 129 Pasquale Minervini, Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel. Adversarial Sets for Regularising Neural Link Predictors. In Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017. AUAI Press, 2017. URL: http://auai.org/uai2017/ proceedings/papers/306.pdf.
- 130 Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018. doi:10.1093/BIB/BBX044.
- 131 Melanie Mitchell. Abstraction and Analogy-Making in Artificial Intelligence. Annals of the New York Academy of Sciences, 1505(1):79–101, 2021. doi:10.1111/nyas.14619.
- 132 Aditya Mogadala, Marimuthu Kalimuthu, and Dietrich Klakow. Trends in Integration of Vision and Language Research: A Survey of Tasks, Datasets, and Methods. Journal of Artificial Intelligence Research, 71:1183–1317, 2021. doi: 10.1613/JAIR.1.11688.
- 133 Sebastian Monka, Lavdim Halilaj, and Achim Rettinger. A survey on visual transfer learning using knowledge graphs. *Semantic Web*, 13(3):477–510, 2022. doi:10.3233/SW-212959.
- 134 Pierre Monnin and Miguel Couceiro. Interactions Between Knowledge Graph-Related Tasks and Analogical Reasoning: A Discussion. In Workshop Proceedings of the 30th International Conferece on Case-Based Reasoning co-located with the 30th International Conference on Case-Based Reasoning (ICCBR 2022), Nancy (France), September 12-15th, 2022, volume 3389 of CEUR Workshop Proceedings, pages 57-67. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3389/ ICCBR_2022_Workshop_paper_75.pdf.
- 135 Diego Moussallem, Mihael Arcan, Axel-Cyrille Ngonga Ngomo, and Paul Buitelaar. Augmenting Neural Machine Translation with

Knowledge Graphs, 2019. arXiv:1902.08816, doi:10.48550/arXiv.1902.08816.

- 136 Daniel Neil, Joss Briody, Alix Lacoste, Aaron Sim, Páidí Creed, and Amir Saffari. Interpretable Graph Convolutional Neural Networks for Inference on Noisy Knowledge Graphs, 2018. doi: 10.48550/arXiv.1812.00279.
- 137 Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Q. Phung. A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers), pages 327–333. Association for Computational Linguistics, 2018. doi:10.18653/V1/N18-2053.
- 138 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A Review of Relational Machine Learning for Knowledge Graphs. Proceedings of the IEEE, 104(1):11–33, 2016. doi: 10.1109/JPR0C.2015.2483592.
- 139 Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. In Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011, pages 809–816. Omnipress, 2011. URL: https://icml.cc/2011/papers/438_ icmlpaper.pdf.
- 140 Vicente Ordonez, Jia Deng, Yejin Choi, Alexander C. Berg, and Tamara L. Berg. From Large Scale Image Categorization to Entry-Level Categories. In *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, pages 2768–2775. IEEE Computer Society, 2013. doi:10.1109/ICCV.2013.344.
- 141 Mark Palatucci, Dean Pomerleau, Geoffrey E. Hinton, and Tom M. Mitchell. Zero-shot Learning with Semantic Output Codes. In Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009., pages 1410-1418. Curran Associates, Inc., 2009. URL: https: //proceedings.neurips.cc/paper/2009/hash/ 1543843a4723ed2ab08e18053ae6dc5b-Abstract. html.
- 142 Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying Large Language Models and Knowledge Graphs: A Roadmap, 2023. doi:10.48550/ARXIV.2306. 08302.
- 143 Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. Estimating Node Importance in Knowledge Graphs Using Graph Neural Networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019, pages 596– 606. ACM, 2019. doi:10.1145/3292500.3330855.
- 144 Heiko Paulheim. Make Embeddings Semantic Again! In Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey,

USA, October 8th - to - 12th, 2018, volume 2180 of CEUR Workshop Proceedings. CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2180/ ISWC_2018_Outrageous_Ideas_paper_4.pdf.

- 145 Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. Knowledge Graphs: Opportunities and Challenges. Artificial Intelligence Review, 56(11):13071–13102, 2023. doi:10.1007/ S10462-023-10465-9.
- 146 Matthew E. Peters, Mark Neumann, Robert L. Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. Knowledge Enhanced Contextual Word Representations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, pages 43-54. Association for Computational Linguistics, 2019. doi:10.18653/V1/D19-1005.
- 147 Jan Portisch, Nicolas Heist, and Heiko Paulheim. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction - two sides of the same coin? Semantic Web, 13(3):399-422, 2022. doi:10.3233/SW-212892.
- 148 Jan Portisch and Heiko Paulheim. The DLCC Node Classification Benchmark for Analyzing Knowledge Graph Embeddings. In *The Semantic* Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings, volume 13489 of Lecture Notes in Computer Science, pages 592–609. Springer, 2022. doi:10.1007/978-3-031-19433-7_34.
- 149 Chen Qu, Hamed Zamani, Liu Yang, W. Bruce Croft, and Erik G. Learned-Miller. Passage Retrieval for Outside-Knowledge Visual Question Answering. In SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021, pages 1753–1757. ACM, 2021. doi:10.1145/3404835.3462987.
- 150 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, volume 139 of Proceedings of Machine Learning Research, pages 8748–8763. PMLR, 2021. URL: http://proceedings.mlr.press/ v139/radford21a.html.
- 151 Luc De Raedt, editor. Logical and Relational Learning: From ILP to MRDM (Cognitive Technologies). Springer-Verlag, 2008. doi:10.1007/ 978-3-540-68856-3.
- 152 Enayat Rajabi and Kobra Etminani. Knowledgegraph-based explainable AI: A systematic review. *Journal of Information Science*, 2022. doi:10. 1177/016555152211128.
- 153 Achim Rettinger, Matthias Nickles, and Volker Tresp. Statistical Relational Learning with Formal Ontologies. In Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, Septem-

ber 7-11, 2009, Proceedings, Part II, volume 5782 of Lecture Notes in Computer Science, pages 286-301. Springer, 2009. doi:10.1007/978-3-642-04174-7_19.

- 154 P. Ristoski and H. Paulheim. RDF2Vec: RDF Graph Embeddings for Data Mining. In The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Proceedings, Part I, volume 9981 of LNCS, pages 498–514. Springer, 2016. doi:10.1007/978-3-319-46523-4_30.
- 155 Petar Ristoski, Gerben Klaas Dirk de Vries, and Heiko Paulheim. A Collection of Benchmark Datasets for Systematic Evaluations of Machine Learning on the Semantic Web. In The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II, volume 9982 of Lecture Notes in Computer Science, pages 186–194, 2016. doi:10.1007/978-3-319-46547-0_20.
- 156 Giuseppe Rizzo, Claudia d'Amato, and Nicola Fanizzi. An unsupervised approach to disjointness learning based on terminological cluster trees. *Semantic Web*, 12(3):423–447, 2021. doi:10.3233/ SW-200391.
- 157 Giuseppe Rizzo, Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito. Terminological Cluster Trees for Disjointness Axiom Discovery. In The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I, volume 10249 of Lecture Notes in Computer Science, pages 184–201, 2017. doi:10.1007/978-3-319-58068-5_12.
- 158 Giuseppe Rizzo, Nicola Fanizzi, and Claudia d'Amato. Class expression induction as concept space exploration: From DL-Foil to DL-Focl. Future Generation Computing Systems, 108:256–272, 2020. doi:10.1016/J.FUTURE.2020.02.071.
- 159 Giuseppe Rizzo, Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. A Framework for Tackling Myopia in Concept Learning on the Web of Data. In Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings, volume 11313 of Lecture Notes in Computer Science, pages 338–354. Springer, 2018. doi:10.1007/978-3-030-03667-6_22.
- 160 Natalia Díaz Rodríguez, Alberto Lamas, Jules Sanchez, Gianni Franchi, Ivan Donadello, Siham Tabik, David Filliat, Policarpo Cruz, Rosana Montes, and Francisco Herrera. EXplainable Neural-Symbolic Learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. Information Fusion, 79:58–83, 2022. doi:10.1016/J.INFFUS.2021.09.022.
- 161 Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach (4th Edition). Pearson, 2020. URL: http://aima.cs.berkeley.edu/.
- 162 Babak Shahbaba and Radford M. Neal. Improving Classification When a Class Hierarchy is Available Using a Hierarchy-Based Prior, 2005. doi:10.48550/arXiv.math/0510449.
- 163 Amit P. Sheth, Manas Gaur, Ugur Kursuncu, and Ruwan Wickramarachchi. Shades of Knowledge-Infused Learning for Enhancing Deep Learning.

IEEE Internet Computing, 23(6):54–63, 2019. doi: 10.1109/MIC.2019.2960071.

- 164 Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pages 4222–4235. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.EMNLP-MAIN.346.
- 165 Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. Finding Streams in Knowledge Graphs to Support Fact Checking. In 2017 IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017, pages 859–864. IEEE Computer Society, 2017. doi:10.1109/ICDM.2017.105.
- 166 Chenglei Si, Zhe Gan, Zhengyuan Yang, Shuohang Wang, Jianfeng Wang, Jordan L. Boyd-Graber, and Lijuan Wang. Prompting GPT-3 To Be Reliable. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023.* OpenReview.net, 2023. URL: https://openreview.net/pdf?id=98p5x51L5af.
- 167 Vivian Dos Santos Silva, André Freitas, and Siegfried Handschuh. Exploring Knowledge Graphs in an Interpretable Composite Approach for Text Entailment. In The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, pages 7023–7030. AAAI Press, 2019. doi:10.1609/ AAAI.V33I01.33017023.
- 168 Fernando Sola, Daniel Ayala, Rafael Ayala, Inma Hernández, Carlos R. Rivero, and David Ruiz. AYNEXT - tools for streamlining the evaluation of link prediction techniques. *SoftwareX*, 23:101474, 2023. doi:10.1016/J.SOFTX.2023.101474.
- 169 Matteo Stefanini, Marcella Cornia, Lorenzo Baraldi, Silvia Cascianelli, Giuseppe Fiameni, and Rita Cucchiara. From Show to Tell: A Survey on Deep Learning-Based Image Captioning. *IEEE Transactions on Patterns Analysis* and Machine Intelligence, 45(1):539–559, 2023. doi:10.1109/TPAMI.2022.3148210.
- 170 Fenglong Su, Chengjin Xu, Han Yang, Zhongwu Chen, and Ning Jing. Neural entity alignment with cross-modal supervision. *Information Pro*cessing and Management, 60(2):103174, 2023. doi: 10.1016/J.IPM.2022.103174.
- 171 Yan Su, Xu Han, Zhiyuan Zhang, Yankai Lin, Peng Li, Zhiyuan Liu, Jie Zhou, and Maosong Sun. CokeBERT: Contextual Knowledge Selection and Embedding towards Enhanced Pre-trained Language Models. AI Open, 2:127–134, 2021. doi:10.1016/J.AIOPEN.2021.06.004.
- 172 Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta,

Canada, May 8-12, 2007, pages 697-706. ACM, 2007. doi:10.1145/1242572.1242667.

- 173 Tianxiang Sun, Yunfan Shao, Xipeng Qiu, Qipeng Guo, Yaru Hu, Xuanjing Huang, and Zheng Zhang. CoLAKE: Contextualized Language and Knowledge Embedding. In Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020, pages 3660–3670. International Committee on Computational Linguistics, 2020. doi:10.18653/V1/2020.COLING-MAIN.327.
- 174 Yu Sun, Shuohuan Wang, Yu-Kun Li, Shikun Feng, Hao Tian, Hua Wu, and Haifeng Wang. ERNIE 2.0: A Continual Pre-Training Framework for Language Understanding. In The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, pages 8968–8975. AAAI Press, 2020. doi:10.1609/AAAI.V34I05.6428.
- 175 Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net, 2019. URL: https://openreview.net/forum?id=HkgEQnRqYQ.
- 176 Vinitra Swamy, Angelika Romanou, and Martin Jaggi. Interpreting Language Models Through Knowledge Graph Extraction, 2021. doi:10. 48550/arXiv.2111.08546.
- 177 Niket Tandon, Gerard de Melo, and Gerhard Weikum. Acquiring Comparative Commonsense Knowledge from the Web. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada, pages 166–172. AAAI Press, 2014. doi:10.1609/AAAI.V28I1.8735.
- 178 Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, pages 4067–4076. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020. ACL-MAIN.374.
- 179 Ilaria Tiddi, Mathieu d'Aquin, and Enrico Motta. Dedalo: Looking for Clusters Explanations in a Labyrinth of Linked Data. In The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings, volume 8465 of Lecture Notes in Computer Science, pages 333–348. Springer, 2014. doi:10.1007/ 978-3-319-07443-6_23.
- 180 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. Artificial Intelligence, 302:103627, 2022. doi:10.1016/J.ARTINT.2021.103627.
- 181 Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on*

Continuous Vector Space Models and their Compositionality, CVSC 2015, Beijing, China, July 26-31, 2015, pages 57-66. Association for Computational Linguistics, 2015. doi:10.18653/V1/ W15-4007.

- 182 An C. Tran, Jens Dietrich, Hans W. Guesgen, and Stephen Marsland. An Approach to Parallel Class Expression Learning. In Rules on the Web: Research and Applications - 6th International Symposium, RuleML 2012, Montpellier, France, August 27-29, 2012. Proceedings, volume 7438 of Lecture Notes in Computer Science, pages 302–316. Springer, 2012. doi:10.1007/ 978-3-642-32689-9_25.
- 183 An C. Tran, Jens Dietrich, Hans W. Guesgen, and Stephen Marsland. Parallel Symmetric Class Expression Learning. *Journal of Machine Learning Research*, 18:64:1-64:34, 2017. URL: http: //jmlr.org/papers/v18/14-317.html.
- 184 Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex Embeddings for Simple Link Prediction. In Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, volume 48 of JMLR Workshop and Conference Proceedings, pages 2071– 2080. JMLR.org, 2016. URL: http://proceedings. mlr.press/v48/trouillon16.html.
- 185 Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. Composition-based Multi-Relational Graph Convolutional Networks. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020. URL: https://openreview.net/forum?id=By1A_C4tPr.
- 186 Johanna Völker, Daniel Fleischhacker, and Heiner Stuckenschmidt. Automatic acquisition of class disjointness. Journal of Web Semantics, 35:124–139, 2015. doi:10.1016/J.WEBSEM.2015.07.001.
- 187 Denny Vrandecic and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10):78-85, 2014. doi: 10.1145/2629489.
- 188 Jindong Wang, Xixu Hu, Wenxin Hou, Hao Chen, Runkai Zheng, Yidong Wang, Linyi Yang, Haojun Huang, Wei Ye, Xiubo Geng, Binxing Jiao, Yue Zhang, and Xing Xie. On the Robustness of Chat-GPT: An Adversarial and Out-of-distribution Perspective, 2023. doi:10.48550/ARXIV.2302.12095.
- 189 Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, pages 2609– 2634. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.147.
- 190 Meihong Wang, Linling Qiu, and Xiaoli Wang. A Survey on Knowledge Graph Embeddings for Link Prediction. Symmetry, 13(3):485, 2021. doi: 10.3390/SYM13030485.
- 191 Minjie Wang, Lingfan Yu, Da Zheng, Quan Gan, Yu Gai, Zihao Ye, Mufei Li, Jinjing Zhou, Qi Huang, Chao Ma, Ziyue Huang, Qipeng Guo,

Hao Zhang, Haibin Lin, Junbo Zhao, Jinyang Li, Alexander J. Smola, and Zheng Zhang. Deep Graph Library: Towards Efficient and Scalable Deep Learning on Graphs, 2019. doi:10.48550/ arXiv.1909.01315.

- 192 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017. doi:10.1109/TKDE.2017. 2754499.
- 193 Taowei David Wang, Bijan Parsia, and James A. Hendler. A Survey of the Web Ontology Landscape. In The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings, volume 4273 of Lecture Notes in Computer Science, pages 682–694. Springer, 2006. doi:10.1007/11926078_49.
- 194 Wei Wang, Vincent W. Zheng, Han Yu, and Chunyan Miao. A Survey of Zero-Shot Learning: Settings, Methods, and Applications. ACM Transactions on Intelligent Systems and Technology, 10(2):13:1–13:37, 2019. doi:10.1145/3293318.
- 195 Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. KGAT: Knowledge Graph Attention Network for Recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019, pages 950–958. ACM, 2019. doi:10.1145/3292500.3330989.
- 196 Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a Few Examples: A Survey on Few-shot Learning. ACM Computing Surveys, 53(3):63:1–63:34, 2021. doi: 10.1145/3386252.
- 197 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge Graph Embedding by Translating on Hyperplanes. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada, pages 1112–1119. AAAI Press, 2014. doi:10.1609/AAAI.V28I1.8870.
- 198 Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, pages 349–357. Association for Computational Linguistics, 2018. doi:10.18653/V1/D18-1032.
- 199 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, 2022. URL: http://papers.nips.cc/paper_files/paper/ 2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html.
- 200 Jialin Wu and Raymond J. Mooney. Entity-Focused Dense Passage Retrieval for Outside-Knowledge Visual Question Answering. In Proceedings of the 2022 Conference on Empirical Methods

in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 8061-8072. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. EMNLP-MAIN.551.

- 201 Lingfei Wu, Peng Cui, Jian Pei, and Liang Zhao, editors. Graph Neural Networks: Foundations, Frontiers, and Applications. Springer, 2022. doi: 10.1007/978-981-16-6054-2.
- 202 Xiayu Xiang, Zhongru Wang, Yan Jia, and Binxing Fang. Knowledge Graph-Based Clinical Decision Support System Reasoning: A Survey. In Fourth IEEE International Conference on Data Science in Cyberspace, DSC 2019, Hangzhou, China, June 23-25, 2019, pages 373–380. IEEE, 2019. doi:10.1109/DSC.2019.00063.
- 203 Han Xiao, Minlie Huang, and Xiaoyan Zhu. TransG : A Generative Model for Knowledge Graph Embedding. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers. The Association for Computer Linguistics, 2016. doi: 10.18653/V1/P16-1219.
- 204 Ning Xie, Farley Lai, Derek Doran, and Asim Kadav. Visual Entailment Task for Visually-Grounded Language Learning, 2018. doi:10. 48550/arXiv.1811.10582.
- 205 Chenyan Xiong, Russell Power, and Jamie Callan. Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding. In Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017, pages 1271–1279. ACM, 2017. doi: 10.1145/3038912.3052558.
- 206 Canran Xu and Ruijiang Li. Relation Embedding with Dihedral Group in Knowledge Graph. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pages 263-272. Association for Computational Linguistics, 2019. doi: 10.18653/V1/P19-1026.
- 207 Chengjin Xu, Fenglong Su, and Jens Lehmann. Time-aware Graph Neural Networks for Entity Alignment between Temporal Knowledge Graphs, 2022. doi:10.48550/ARXIV.2203.02150.
- 208 Da Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. Product Knowledge Graph Embedding for E-commerce. In WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020, pages 672–680. ACM, 2020. doi:10.1145/3336191.3371778.
- 209 Bishan Yang and Tom M. Mitchell. Leveraging Knowledge Bases in LSTMs for Improving Machine Reading. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30
 August 4, Volume 1: Long Papers, pages 1436– 1446. Association for Computational Linguistics, 2017. doi:10.18653/V1/P17-1132.
- 210 Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding Entities and Relations for Learning and Inference in Knowledge

Bases. In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015. URL: http://arxiv.org/abs/1412. 6575.

- 211 Zhen Yao, Wen Zhang, Mingyang Chen, Yufeng Huang, Yi Yang, and Huajun Chen. Analogical Inference Enhanced Knowledge Graph Embedding. In Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023, pages 4801–4808. AAAI Press, 2023. doi:10.1609/AAAI.V37I4.25605.
- 212 Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021, pages 535–546. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.NAACL-MAIN.45.
- 213 Zi Ye, Yogan Jaya Kumar, Goh Ong Sing, Fengyan Song, and Junsong Wang. A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs. *IEEE Access*, 10:75729–75741, 2022. doi:10.1109/ACCESS.2022.3191784.
- 214 Jason Youn and Ilias Tagkopoulos. KGLM: Integrating Knowledge Graph Structure in Language Models for Link Prediction. In Proceedings of the The 12th Joint Conference on Lexical and Computational Semantics, *SEM@ACL 2023, Toronto, Canada, July 13-14, 2023, pages 217–224. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.STARSEM-1.20.
- 215 Donghan Yu, Chenguang Zhu, Yiming Yang, and Michael Zeng. JAKET: Joint Pre-training of Knowledge Graph and Language Understanding. In Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022, pages 11630–11638. AAAI Press, 2022. doi:10.1609/AAAI.V36I10.21417.
- 216 Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. A Survey of Knowledge-enhanced Text Generation. ACM Computing Surveys, 54(11s):227:1–227:38, 2022. doi:10.1145/3512467.
- 217 Rowan Zellers, Yonatan Bisk, Ali Farhadi, and Yejin Choi. From Recognition to Cognition: Visual Commonsense Reasoning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR* 2019, Long Beach, CA, USA, June 16-20, 2019, pages 6720–6731. Computer Vision Foundation / IEEE, 2019. doi:10.1109/CVPR.2019.00688.
- 218 Kunli Zhang, Linkun Cai, Yu Song, Tao Liu, and Yueshu Zhao. Combining external medical knowledge for improving obstetric intelligent diagnosis:

model development and validation. *JMIR medical informatics*, 9(5):e25304, 2021. doi:10.2196/ 25304.

- 219 Tong Zhang, Cheng Wang, Ning Hu, Minlie Qiu, Chen Tang, Xiaodong He, and Jian Huang. DKPLM: Decomposable Knowledge-Enhanced Pretrained Language Model for Natural Language Understanding. In Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, pages 11 703–11 711. AAAI Press, February 22 March 1 2022. doi:10.1609/aaai.v36i10.21425.
- 220 Yingying Zhang, Shengsheng Qian, Quan Fang, and Changsheng Xu. Multi-modal Knowledgeaware Hierarchical Attention Network for Explainable Medical Question Answering. In Proceedings of the 27th ACM International Conference on Multimedia, MM 2019, Nice, France, October 21-25, 2019, pages 1089–1097. ACM, 2019. doi:10.1145/3343031.3351033.
- 221 Yang Zhao, Lu Xiang, Junnan Zhu, Jiajun Zhang, Yu Zhou, and Chengqing Zong. Knowledge Graph Enhanced Neural Machine Translation via Multi-task Learning on Sub-entity Granularity. In Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020, pages 4495–4505. International Committee on Computational Linguistics, 2020. doi: 10.18653/V1/2020.COLING-MAIN.397.
- 222 Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc V. Le, and Ed H. Chi. Least-to-Most Prompting Enables Complex Reasoning in Large Language Models. In The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. URL: https://openreview.net/pdf?id=WZH7099tgfM.
- 223 Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural

networks: A review of methods and applications. *AI Open*, 1:57–81, 2020. doi:10.1016/J.AIOPEN. 2021.01.001.

- 224 Xiaohan Zhou, Yunhui Yi, and Geng Jia. Path-RotatE: Knowledge Graph Embedding by Relational Rotation of Path in Complex Space. In 10th IEEE/CIC International Conference on Communications in China, ICCC 2021, Xiamen, China, July 28-30, 2021, pages 905–910. IEEE, 2021. doi:10.1109/ICCC52777.2021.9580273.
- 225 Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large Language Models are Human-Level Prompt Engineers. In The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. URL: https: //openreview.net/pdf?id=92gvk82DE-.
- 226 Yu Zhou, Haixia Zheng, Xin Huang, Shufeng Hao, Dengao Li, and Jumin Zhao. Graph Neural Networks: Taxonomy, Advances, and Trends. ACM Transactions on Intelligent Systems and Technology, 13(1):15:1–15:54, 2022. doi:10.1145/ 3495161.
- 227 Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction. In Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual, pages 29476-29490, 2021. URL: https: //proceedings.neurips.cc/paper/2021/hash/ f6a673f09493afcd8b129a0bcf1cd5bc-Abstract. html.
- 228 Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1):43–76, 2021. doi:10.1109/JPR0C.2020.3004555.
- 229 Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. Exploring AI Ethics of ChatGPT: A Diagnostic Analysis, 2023. doi: 10.48550/ARXIV.2301.12867.

Trust, Accountability, and Autonomy in Knowledge Graph-Based AI for Self-Determination

Luis-Daniel Ibáñez ⊠©

Department of Electronics and Computer Science, University of Southampton, UK

John Domingue 🖂 🗅

Knowledge Media Institute, The Open University, Milton Keynes, UK

Sabrina Kirrane 🖂 🗅

Institute for Information Systems & New Media, Vienna University of Economics and Business, Austria

Oshani Seneviratne ⊠©

Department of Computer Science, Rensselaer Polytechnic Institute, USA

Aisling Third ⊠© Knowledge Media Institute, The Open University, Milton Keynes, UK

Maria-Esther Vidal ⊠©

Leibniz University of Hannover, Germany TIB-Leibniz Information Centre of Science and Technology, Hannover, Germany L3S Research Centre, Hannover, Germany

— Abstract -

Knowledge Graphs (KGs) have emerged as fundamental platforms for powering intelligent decisionmaking and a wide range of Artificial Intelligence (AI) services across major corporations such as Google, Walmart, and AirBnb. KGs complement Machine Learning (ML) algorithms by providing data context and semantics, thereby enabling further inference and question-answering capabilities. The integration of KGs with neuronal learning (e.g., Large Language Models (LLMs)) is currently a topic of active research, commonly named neuro-symbolic AI. Despite the numerous benefits that can be accomplished with KG-based AI, its growing ubiquity within online services may result in the loss of selfdetermination for citizens as a fundamental societal issue. The more we rely on these technologies, which are often centralised, the less citizens will be able

to determine their own destinies. To counter this threat, AI regulation, such as the European Union (EU) AI Act, is being proposed in certain regions. The regulation sets what technologists need to do, leading to questions concerning How the output of AI systems can be trusted? What is needed to ensure that the data fuelling and the inner workings of these artefacts are transparent? How can AI be made accountable for its decision-making? This paper conceptualises the foundational topics and research pillars to support KG-based AI for self-determination. Drawing upon this conceptual framework, challenges and opportunities for citizen self-determination are illustrated and analysed in a real-world scenario. As a result, we propose a research agenda aimed at accomplishing the recommended objectives.

2012 ACM Subject Classification Social and professional topics \rightarrow Computing / technology policy; Computing methodologies \rightarrow Knowledge representation and reasoning; Human-centered computing \rightarrow Collaborative and social computing theory, concepts and paradigms; Security and privacy \rightarrow Human and societal aspects of security and privacy; Computing methodologies \rightarrow Distributed artificial intelligence Keywords and phrases Trust, Accountability, Autonomy, AI, Knowledge Graphs

Digital Object Identifier 10.4230/TGDK.1.1.9

Category Vision

Related Version Previous Version: https://arxiv.org/abs/2310.19503

 $\label{eq:Funding Luis-Daniel Ibáñez: Partially funded by the European Union's Horizon Research and Innovation Actions under Grant Agreement nº 101093216 (UPCAST)$

Sabrina Kirrane: Partially funded by the FWF Austrian Science Fund and the Internet Foundation



© Luis-Daniel Ibáñez, John Domingue, Sabrina Kirrane, Oshani Seneviratne, Aisling Third, and Maria-Esther Vidal;

licensed under Creative Commons License CC-BY 4.0 Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 9, pp. 9:1–9:32

Transactions on Graph Data and Knowledge



TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

9:2 Trust, Accountability, and Autonomy in KG-Based AI

Austria under the FWF Elise Richter and netidee SCIENCE programmes as project number V 759-N. *Oshani Seneviratne*: Partially funded by NSF IUCRC CRAFT center research grant (CRAFT Grant #22008) and the Algorand Centres of Excellence programme managed by Algorand Foundation. The opinions expressed in this publication do not necessarily represent the views of NSF IUCRC CRAFT or the Algorand Foundation.

Maria-Esther Vidal: Partially funded by Leibniz Association, program "Leibniz Best Minds: Programme for Women Professors", project TrustKG-Transforming Data in Trustable Insights; Grant P99/2020 Received 2023-06-30 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1 Introduction

Modern Artificial Intelligence (AI) can be traced back to a workshop held at Dartmouth College in the summer of 1956 [66] and is most commonly defined as the use of computers to simulate human intelligence, in particular human reasoning, learning, and problem-solving. Since 1956, AI has lived through times of increased interest and funding, and also "AI Winters", such as after the 1974 Lighthill report [63], when overall funding was reduced. Over the last few years, however, funding and interest in AI have been high and exploded in November 2022, when ChatGPT, a type of Generative AI, was announced by OpenAI, exposing the power of Large Language Models (LLMs) to the general public. Since its release, ChatGPT has become the fastest-growing app in history, reaching 100M users in just two months, and is now estimated to have 200M users. Generative AI will continue to grow following a significant investment by Microsoft into OpenAI and announcements by Microsoft and Google on how Generative AI will be embedded in future products [36]. Data-centric AI [114] recognises the immense value of data as crucial resources for training, optimising, and evaluating AI systems. Databricks, a prominent AI company, has defined data-centric AI as the challenge of designing processes for data collection, labelling, and quality monitoring in machine learning (ML) datasets [85] highlighting the need for continuous re-running and re-training, actionable monitoring, and the difficulties of incorporating data inaccessible to human annotators due to privacy concerns as primary research directions. Knowledge Graphs have been used as a resource and structure to support data-centric AI processes.

The term *Knowledge Graph* (KG) was first introduced by Google in 2012 and is usually defined as a type of knowledge structure that uses a graph data model to integrate data. KGs are strongly linked to the work of the Semantic Web community, which first began in around 2001 and was introduced in a seminal paper by Tim Berners-Lee [13]. The Semantic Web initiative produced a stack of web standards on which KGs are based. These include the Resource Description Framework (RDF), where data is encoded as subject-predicate-object triples, and the Web Ontology Language (OWL), a set of web-based languages mostly based on description logic. The common theme of these semantic representations is that they facilitate the publishing, use, and re-use of data at the web-scale. In particular, they allow disparate heterogeneous data sources to be integrated continuously at scale. Over the past decade, KGs have become a mainstay for several key large-scale applications found online. For example, KGs underpin Google Search, which saw 5,900,000 searches in just one minute in April 2022. Similarly, the same minute saw 1,700,000 pieces of content shared on Facebook, 1,000,000 hours streamed, and 347,200 tweets shared on Twitter. All of this content and data are linked to a plethora of AI services that have increasingly been based on KGs, as mentioned above, founded upon machine-readable data and schema representations based on a web stack of standards. AI services cover a wide number of areas, including content recommendation, user input prediction, and large-scale search and discovery and form the basis for the business models of companies like Google, Netflix, Spotify, and Facebook. Given the above, we define KG-based AI as an AI system (replicating some aspect of human intelligence) based on a KG that possibly uses the web standards produced by the Semantic Web community.

In addition to privacy concerns, there has been a growing worry about how personal data can be abused and, thus, how AI services impinge on citizen rights. For example, the over-centralisation of data and its misuse led Sir Tim Berners-Lee to call the Web "anti-human" in an interview in 2018 [18]. Since 2016, hundreds of United States (US) Immigration and Customs Enforcement employees have faced investigations into abuse of confidential law enforcement databases, from stalking and harassment, to passing data to criminals [69]. The subject of much of the proposed legislation today is ensuring that digital platforms, including AI platforms, provide real societal benefit. Within Europe, the proposed European Union (EU) AI Act¹ aims to support safe AI that respects fundamental human rights. The regulation sets what technologists need to do. The concept of data *self-determination*, which is often used in a legal context, implies that individuals are not only aware of who knows what about them but can also influence data processing that concerns them [60]. Given that nowadays, data processing is conducted by opaque AI algorithms behind corporate firewalls, sometimes even without our knowledge, data self-determination is harder than ever before. When it comes to trust in web data and services, Berners-Lee and Fischetti [12] envisaged an "Oh yeah?" button embedded into Web browsers that would provide justifications as to why a page or a service should be trusted. Alas, their vision was never realised in popular web browsers². Instead, we have dedicated websites, e.g., the Ecommerce Europe $Trustmark^3$ that are used to perform company reputation checks and fact-checking websites, such as Snopes⁴, that can be used to check the validity of information posted online. Although some automated fact-checking techniques have been proposed [87], they are used solely for developing trust in information resources and cannot provide any guarantees with respect to AI-based data processing. As we move beyond trust towards accountability, policies have already been used to specify legal data processing requirements that serve as the basis for automated compliance checking, for example, [82]. But what happens when service providers or AI algorithms do not comply? How far can technology go in terms of helping us determine non-compliance and make service providers accountable for their actions?

In this paper, we propose a research agenda for ensuring that KG-based AI approaches contribute to user self-determination instead of hindering it. Our vision, which is depicted in Figure 1, is structured around three pillar research topics - trust, accountability, and autonomy - that represent the desired goals for how AI can benefit society and facilitate self-determination. The pillars combine fundamental principles of the proposed EU AI Act and self-determination theory. Both trust and accountability are imperative for safeguarding against adverse impacts caused by AI systems, while autonomy is critical for ensuring individuals are able to determine their own destinies. The pillars are supported via four foundational research topics - machine-readable norms and policies are needed for humans to declare regulatory frameworks, privacy and usage constraints that can be interpreted by the machines that process their data; decentralised KG management and decentralised infrastructure to provide alternatives to approaches where a central entity controls a whole process, that are prone to abuse of power; and explainable neuro-symbolic AI to clearly communicate and prove the decisions AI systems make. We posit the following research questions:

¹ https://artificialintelligenceact.eu/

² However, a linked browser prototype, the Tabulator, incorporated this feature in an *Justification UI* (http://dig.csail.mit.edu/TAMI/2008/JustificationUI/howto.html#useTab).

³ https://ecommercetrustmark.eu/

⁴ https://www.snopes.com/

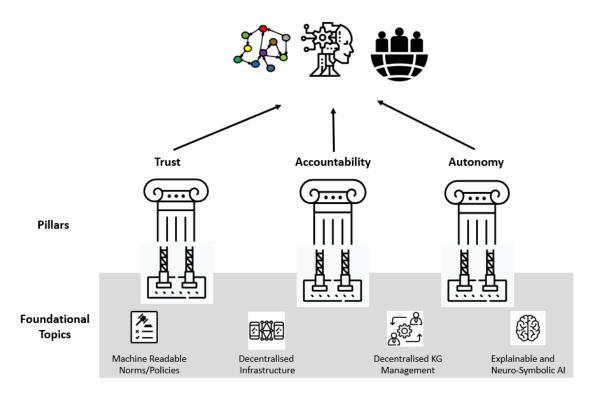


Figure 1 KG-based AI for Self-determination Conceptualisation. KG-based AI for self-determination is supported by the pillars of trust, accountability, and autonomy, built upon the foundational topics of machine-readable norms and policies, decentralised infrastructure, decentralised KG management, and explainable neuro-symbolic AI.

- Q1 What are the key requirements for an AI system to produce trustable results?
- Q2 How can AI be made accountable for its decision-making?
- Q3 How can citizens maintain autonomy as users or subjects of KG-based AI systems?

In order to facilitate exposition, we ground our discussion in a healthcare scenario inspired by the recently proposed regulation on European Health Data Space^5 that aims to ensure that "natural persons in the EU have increased control in practice over their electronic health data" and to facilitate access to health data by various stakeholders in order to "promote better diagnosis, treatment and well-being of natural persons, and lead to better and well-informed policies". The proposed healthcare scenario, which is illustrated in Figure 2, is composed of the following actors and interactions:

- Individuals manage their Personal Knowledge Graphs (PKGs) (aligned with the original Semantic Web vision and modern interpretations [7, 47]). They collect knowledge about their medical conditions, symptoms, treatments, reactions to treatments, etc. Individuals get services from KG-based AI applications that utilise their PKGs, e.g., therapy bots or health assistants.
- **Experts** in healthcare also have PKGs where they collect their knowledge about diseases, results of the treatments they have suggested in the past, links to general medical knowledge graphs, etc. Experts may also be assisted by KG-based AI models.

⁵ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197

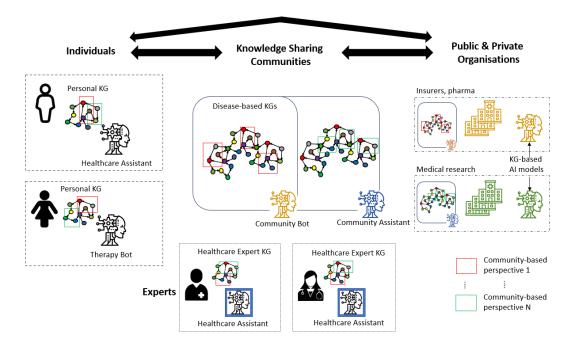


Figure 2 Illustrative Scenario for KG-Based AIs in the healthcare domain. Individuals use AI assistants to make sense of data collected in their PKGs. They may also share perspectives of their PKGs with other individuals and healthcare experts in knowledge-sharing communities that aggregate and curate data to power AI services for the benefit of all members. Public and private organisations can negotiate access to data from communities and individuals to train KG-based AI models, which in turn are used to build services for them.

- Knowledge sharing communities are spaces where individuals and healthcare experts may share subsets of their PKGs in the context of specific knowledge, e.g., diseases. We call these *community-based perspectives*. Perspectives from different contributors are aggregated into community KGs (e.g., disease-based). AI applications use these KGs for community benefit, e.g., assessing if a treatment that worked for an individual may work for a different individual.
- **Public and private organisations** may negotiate access to data and knowledge from communities to train large KG-based AI models to either improve internal processes or power products sold to communities, experts, or individuals, completing the cycle.

The remainder of the paper is structured as follows: Section 2 introduces the necessary background in terms of KG-based AI. Section 3 highlights the importance of trust, accountability, and autonomy when it comes to ensuring that AI benefits society. Section 4 presents several KG-based tools and techniques that can be used to facilitate trust, accountability, and selfdetermination. In Section 5, we propose a research roadmap that includes several challenges and opportunities for KG-based AI that benefits individuals and society. Finally, we conclude and outline important first steps in Section 6.

2 Knowledge Graph-based AI

In his seminal publication, "*Thinking, Fast and Slow*", Daniel Kahneman [49] presents a comprehensive theory of human intelligence, offering profound insights into the workings of the human mind. This groundbreaking work separates intuition from rationality when approaching problem-solving tasks, defining them as two sets of abilities or *systems*. System 1 operates at an

9:6 Trust, Accountability, and Autonomy in KG-Based AI

unconscious level, generating responses effortlessly and swiftly. In contrast, System 2 requires conscious attention and concentration, enabling the generation of responses needing complex computations. Kahneman's characterisation of mental cognition aligns with statistical and symbolic learning models that seek to simulate human thinking processes [16]. These systems are known as neuro-symbolic systems [17], and there is a growing interest in emerging hybrid approaches that aim to integrate cognitive capabilities. Specifically, they strive to combine the power of neural networks, such as LLMs, with the interpretability offered by symbolic processing, particularly semantic reasoning over KGs.

2.1 Knowledge Graphs

Google first introduced KGs in 2012 when they enabled "Knowledge Panels" to contain descriptions, including pictures, for search items. For example, if one types "London" in Google Search, the Knowledge Panel displays pictures, the current weather, a map, directions, elevation, and related entities (e.g., Paris). The seed for the Google KG was Freebase – a community knowledge base initially launched in 2007 with an add-on RDF service launched at the International Semantic Web Conference in 2008. In 2010, Google bought Metaweb, the company that owned Freebase and extended the knowledge base into the Google KG⁶.

In 2011, Bing, Google and Yahoo! launched Schema.org, a reference website for common data schemas related to web search engines. The proposal was that website owners would use the published schemas alongside Semantic Web standards such as RDFa and JSON-LD. A number of the schemas, such as **Organisation**, influence the results returned by Google KG search. Schema.org is an example of a shared vocabulary for semantic representation; the use of such vocabularies or ontologies in KGs, along with the ability to map between equivalent schemas in them, enables the integration of heterogeneous data at scale.

Today, KGs are used in a wide range of areas and products outside of search. For example, Netflix, Amazon, and Facebook all use KGs as the foundation for their recommendation engines for television programmes and films, consumer products and posts⁷, whereas in the healthcare sector, KGs are used to integrate medical knowledge and support drug discovery.⁸

2.2 Large Language Models

A Large Language Model (LLM) is a specialised machine learning model constructed using a transformer architecture, a category of deep neural networks [120]. LLMs are primarily designed for predicting the next word in a sequence, making them flexible tools for various text-processing tasks, such as text generation, summarisation, translation, and text completion. Examples of existing LLMs include OpenAI's ChatGPT [93] and Google's PALM [24]. These models have demonstrated high performance in Natural Language Processing (NLP) tasks like code generation, text generation, tool manipulation, and comprehension across diverse domains, often achieving high-quality results in zero-shot and few-shot settings. This success has stimulated advancements in LLM architectures, training techniques, prompt engineering, and question answering [72].

Despite their unquestionable capabilities in emulating human-like conversations, there is an ongoing debate regarding the intelligence exhibited by LLMs, particularly, since their fluency in language does not necessarily imply a cognitive understanding of real-world problems [72]. Additionally, LLMs can only learn knowledge when it appears in the training data and may

⁶ https://en.wikipedia.org/wiki/Schema.org

⁷ https://builtin.com/data-science/knowledge-graph

⁸ https://www.wisecube.ai/blog/20-real-world-industrial-applications-of-knowledge-graphs/

perform badly when answering questions involving long-tailed facts [30]. Moreover, they may struggle to absorb new knowledge and are not easy to audit [73], suggesting potential risks of discrimination and information hazards.

2.3 Neurosymbolic AI

LLMs and machine learning models, in general, are trained on extensive datasets, resulting in high-quality outcomes whenever applied to specific prediction tasks. However, LLMs, like OpenAI's ChatGPT [93], lack causal understanding and may hallucinate in cases which are not statistical in nature (e.g. memories or explanations) [40]. On the other hand, symbolic AI systems are capable of emulating human-like conscious processes required for causality, logic, and counterfactual reasoning, as well as maintaining long-term memory. As a result, symbolic systems can empower LLMs by modelling human learning and combining knowledge extracted (e.g., from KGs) to formulate prompts that allow for more fluent communication with users.

Neuro-symbolic AI provides the basis for integrating the discrete approaches implemented by Symbolic AI with high-dimensional vector spaces managed by LLMs. They must decide when and how to combine both systems, e.g. following a principled integration (combining neural and symbolic while maintaining a clear separation between their roles and representations) or integrated (e.g., a symbolic reasoner integrated into the tuning process of an LLM). Recently, van Bekkum et al. [109] propose 17 fundamental design patterns to model neuro-symbolic systems. These patterns encompass many scenarios where the symbiotic relationship between symbolic reasoning and ML models becomes apparent. Since these combinations may enable symbolic reasoning and enhance contextual knowledge, neuro-symbolic systems may empower explainability and, as a result, also improve transparency by showing how a system works based on the symbolic explanations deduced by the hybrid system.

3 KG-based AI that Benefits Individuals and Society

Considering our vision that KG-based AI can facilitate self-determination, we start by discussing the pertinent role played by trust, accountability, and autonomy when it comes to ensuring that AI benefits society. In each case, we highlight existing challenges and present arguments in favour of a KG-based AI system.

3.1 Trust and KG-based AI

One of the primary objectives of the proposed EU AI Act is the "development of an ecosystem of trust by proposing a legal framework for trustworthy AI". The Merriam-Webster dictionary definition of trust includes a "firm belief in the reliability, truth, or ability of someone or something" [70]. Questions we address in this paper include understanding how KG-based AI systems can demonstrate reliability, truth, and ability through mechanisms which add transparency to all elements involved in KG-reasoning. These include comprehensive provenance tracking of data sources and data elements used for any output; understanding repeatability for all KG-based AI reasoning (e.g. if datasets are altered or disappear altogether or if other reasoning methods, such as LLMs, are involved); and alleviation mechanisms when KG-based AI system responses are untruthful.

The proliferation of misinformation on the internet has risen significantly in recent years, coinciding with the advancements in generative AI technologies. As AI becomes more sophisticated, it has inadvertently provided tools and techniques for the creation and dissemination of false

9:8 Trust, Accountability, and Autonomy in KG-Based AI

information, leading to widespread confusion and societal harm [25, 121]. For instance, AIgenerated deepfake videos have become a concerning source of misinformation. Deepfakes use AI algorithms to manipulate and superimpose faces onto existing videos, making it difficult to discern real from fabricated content [113]. This technology has been used to create fake videos of public figures saying or doing things they never actually did, leading to potential defamation and manipulation of public opinion. AI-powered chatbots and automated accounts on social media platforms have been employed to spread false information and manipulate public sentiment. These bots can mimic human-like conversations and flood social media platforms with fake news, propaganda, and divisive narratives, influencing public opinion and sowing discord, and have even contributed to misinformation in medical literature [65]. AI-powered recommendation algorithms used by platforms like social media and video-sharing websites can inadvertently contribute to the spread of misinformation. These algorithms aim to maximise user engagement by suggesting content based on user preferences and behaviour. They can create filter bubbles, reinforcing users' existing beliefs and exposing them to a limited range of perspectives, potentially amplifying false information and preventing users from accessing accurate and diverse sources of information [84].

Amidst these challenges, KG technologies have emerged as a potential solution to curb misinformation and enhance trust. Leveraging the power of crowd-supplied and verified knowledge sources, such as Wikidata [111], KGs enable comprehensive fact-checking capabilities. By integrating diverse and reliable information from various trusted sources, these graphs can potentially identify and flag misleading or inaccurate content more effectively. By utilising the collective intelligence of a crowd, KG technologies empower users to contribute to the verification process, enhancing the accuracy and credibility of the information presented. Through collaborative efforts and the utilisation of KG technologies, it is possible to combat the rising tide of misinformation, safeguard the integrity of online information and foster a more informed digital society. Coupled with distributed ledgers KG-based AI may combat misinformation on the web [95]. There is already a growing body of work in this space, which shows some promise. For example, Mayank et al. [68] and Koloski et al. [57] describe systems that leverage KGs to detect fake news; Kou et al. [59] and Shang et al. [96] describe how crowd-sourced KGs can be used to mitigate COVID-19 misinformation; and Kazenoff et al. [50] use semantic graph analysis to detect cryptocurrency scams propagating in social media.

3.2 Accountability and KG-based AI

According to the proposed EU AI Act, when it comes to high-risk AI, "accuracy, reliability and transparency are particularly important to avoid adverse impacts, retain public trust and ensure accountability and effective redress". Accountability in a KG-based AI context assumes that data scientists, computer scientists, and software engineers will follow best practices and ensure compliance with relevant legislation. In the purely symbolic world, such properties can be achieved via consistency and compliance checking based on formal requirements specified in policy languages such as LegalRuleML [4] and ODRL [46]. When it comes to the sub-symbolic world, these principles are particularly challenging, as ML algorithms are often opaque and could potentially infer confidential information during the training process. In recent years, various Explainable AI (XAI) techniques have been used to build or to be applied to the output of models such that they can be interpreted and understood by various stakeholders [56]. In the context of KG-based AI, this will require the intersection between two strains of explainability: the explanation of why a statement is in the KG that supports the AI and the explanation of how the model used the statements from the KG to reach a particular decision. KGs can also be used to support the modelling, capturing, and auditing of records useful for accountability throughout the system life cycle [75]

When it comes to AI and accountability, technical research should go hand in hand with the interdisciplinary research conducted in communities like $FaccT^9$. A recent paper [26] revisited the four barriers of accountability that were developed in the 1990s for accountability of computerised systems in the light of the rise of AI, finding that they are even more important than before. The main barrier is the problem of many hands – the large amount of actors involved in the construction of an AI service creates difficulties in the assignment of responsibilities in case of harm. Advancing efficient provenance collection and verifiability will be the key technical intervention to overcome this barrier. Fields such as data science require strong guarantees for provenance to build context-aware KGs [94]. Similar to explainability, we consider two different approaches that need to be combined: the provenance of statements in the KG and the provenance of the pipeline that was followed to construct the ML model.

3.3 Autonomy and KG-based AI

Alongside accountability and trust, the third pillar needed to support self-determination is *autonomy*, defined from a self-determination theory¹⁰ perspective as "the belief that one can choose their own behaviours and actions". In the current context, we take this to mean that individuals should be able to make their own decisions about their uses of KG-based AI and about its uses of their data (and have their wishes respected). Assuming that AI systems can be made to be trustable and accountable, how can we best support autonomy in this way? That is to say, if we can know that an AI will behave in a desired and known way and that its decisions and processes are transparent and traceable, how can we express and enable control over what it does in regard to an individual? A number of approaches have emerged in recent years which facilitate individuals' data sovereignty and how they represent and express their identity online.

The concept of a PKG, introduced in our illustrative scenario (Figure 2), is one means of facilitating autonomy; Solid pods [91, 67] are secure decentralised data stores accessible through standard semantic interfaces for applications that generate and consume linked data. Currently, the default model on the Web is for service providers to host and control access to user data by means of a user account. This denies autonomy to the individuals concerned since all access is mediated via applications and interfaces designed and controlled by service providers. The PKG model is that personal data is independent of any application; PKGs are the primary source of data under the control of individuals, and they mediate service access via standard interfaces. On top of shifting control away from service providers, this approach makes it technically simpler to implement data usage policies, as they can be stored with the data and evaluated at the PKG level.

One prominent way of achieving the second goal is through the notion of Self-Sovereign Identity (SSI) [27]. Traditional digital identity (e.g. as in OpenID Authentication [41]) has been modelled in terms of Identity Providers (IdPs). An individual and an IdP establish a relationship, and the IdP generates a digital identity for them. If the individual wants to authenticate with a third party, the IdP confirms the relationship to them and then asserts that identity to the relying party. Crucially, sovereignty over that identity and decisions about who can see it, the data associated with it, or whether it continues to exist are taken by the IdP. With SSI, an individual generates their *own* digital identity (e.g. a cryptographic key pair), makes their own identity assertions, and therefore has full control over that identity, with correlations between two identities (digital or physical)

⁹ https://facctconference.org/index.html

¹⁰ https://en.wikipedia.org/wiki/Self-determination_theory

9:10 Trust, Accountability, and Autonomy in KG-Based AI

relying explicitly on attestation by others, and trust relationships with them¹¹. The autonomy enabled by SSI makes *selective disclosure* possible, meaning that what identity information gets shared with whom can be made contextually and on a case-by-base basis – much like presenting different aspects of one's personal identity in daily life (e.g., work and home personas).

Considerations of identity pervade technical ones for safeguarding self-determination. It seems uncontroversial that there will be scenarios in which an individual's identity is relevant to what they wish to do with a KG-based AI, whether in training, KG contents, or inference, and indeed, even where anonymity is desired, identity must be considered in order to avoid revealing it. Identity is also fundamental to the concept of trust; trust in a person, organisation, system, AI model, KG, etc., is useful only insofar as it is possible to identify relevant entities as needed, and accountability cannot be tracked or apportioned without it. We consider autonomy in terms of the identity, data, and sovereignty afforded to an individual or organisation in terms of what they or others communicate to a KG-based AI ecosystem or elements thereof, what they or others receive from those, and what happens to those (including respect of choices) as data is processed in the ecosystem, with each of these evaluated through the lenses of selective disclosure, relevant identities, and utility.

4 A KG Toolbox for Trust, Accountability, and Autonomy

In order to ground our pillars, we motivate and introduce relevant literature and highlight open research challenges and opportunities concerning our foundational topics: machine-readable norms and policies, decentralised infrastructure, decentralised KG management, and explainable neurosymbolic AI, each of which plays a pivotal role in facilitating trust, accountability, and autonomy in KG-based AI.

4.1 Machine-readable Norms and Policies

When it comes to KG-based AI, norms and policies could potentially be used to inform data processing based on legal requirements, social norms, privacy preferences, and licensing. Legal documents are designed in natural language for human consumption; thus in order to enable machines and automated agents to evaluate and enforce the agreements embodied in documents, we need to translate them to formats they can read and process efficiently.

4.1.1 Norm and Policy Encoding

Languages to express policies, including but not limited to data access, can be categorised as either general or specific. In the former, the syntax caters to a diverse range of functional requirements (e.g. access control, query answering, service discovery, negotiation), whereas the latter focuses on just one functional requirement. In the early days of the Semantic Web, research into general policy languages that leverage semantic technologies (e.g. KAoS [108], Rei [48], AIR [51], and Protune [15]) was an active area of research. However, despite the huge potential offered by these general-purpose languages to date, none of them achieved mainstream adoption [55]. More recently, researchers have proposed ontologies that can be used to represent licences, privacy preferences, and regulatory obligations [53]. When it comes to the legal domain specifically, Semantic Web researchers have proposed cross-domain ontologies that can be used to encode legal text in a

¹¹ As it ultimately does in traditional digital identity, where trust in a small number of well-known IdPs serves as a simplified proxy for more detailed or fine-grained considerations of trust networks.

machine-readable format using LegalRuleML¹² and adaptations thereof (e.g. [3, 80]). Others focus on facilitating legal document indexing and search using the European Law Identifier $(ELI)^{13}$ and the European Case Law Identifier $(ECLI)^{14}$ (e.g. [78, 22]), or bridging the gap between the EU and member state legal terminology (e.g. [1, 14]). Besides these cross-domain activities, there have also been various domain-specific initiatives. For instance, the ELI ontology has to be extended to facilitate the encoding of the text of the General Data Protection Regulation $(\text{GDPR})^{15}$ (e.g. [83]). At the same time, others have focused specifically on modelling privacy policies (e.g. [79, 82]). The Open Digital Rights Language (ODRL)¹⁶, which is a W3C recommendation, has gained a lot of traction in recent years in terms of intellectual property rights management (e.g. [42, 74]). Additionally, the ODRL model and vocabularies have been extended in order to model contracts [39], personal data processing consent [32], and data protection regulatory requirements [110]. There has also been some work on automatically extracting rights and conditions from textual documents (e.g. [21, 20]) or extracting important information from legal cases (e.g., [116, 77]). Although many of the proposed approaches are based on existing standards, there is a lot of overhead involved for systems that need to consider different types of policies that are encoded using different languages. General-purpose policy languages are particularly attractive in such scenarios as they lessen the administrative burden. However, considering the potential complexity of such a language, there is a need for policy profiles with well-defined semantics and complexity classes.

4.1.2 Norm and Policy Encoding

From a policy governance perspective, LegalRuleML researchers have proposed automated compliance approaches based on auditing (e.g. [28, 82]) and business processes (e.g. [81, 9]). While [37] shows how LegalRuleML, together with semantic technologies, is used for business process regulatory compliance checking based on a rule-based logic combining defeasible and deontic logic. One of the advantages of description logic-based approaches, when it comes to consistency and compliance checking, is that they can leverage generic reasoners, such as Pellet¹⁷ (e.g. [33]). Although there are presently no ODRL-specific reasoning engines, researchers have demonstrated how ODRL can be translated into rules that can be processed by Answer Set Programming (ASP) [8] solvers such as Clingo [35] (e.g. [42, 110]). Additionally, there have been several custom applications that are designed to support ODRL enforcement or compliance checking, such as a licence-based search engine [74]; generalised contract schema and role-based access control enforcement [39]; and access request matching and authorisation [32]. Despite existing efforts, challenges arise when it comes to ensuring that AI and processing algorithms adhere to the policies. This could potentially be achieved either before or during processing using Trusted Execution Environments (TEEs) [10] or after execution by detecting data misuse via automated compliance checking using system logs [54]. The combination of exante and ex-post compliance checking is particularly appealing for supporting risk-based conformance checking such as that envisaged in the proposed EU AI Act. Nevertheless, the practicality, performance and scalability of these proposals remain to be determined. In order to further support self-determination, data owners and processors should

¹⁶ https://www.w3.org/TR/odrl-model/

¹² https://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/os/legalruleml-core-specv1.0-os.html

¹³https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01)

¹⁴ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011XG0429(01)

¹⁵ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1681238509224

¹⁷ https://github.com/stardog-union/pellet

9:12 Trust, Accountability, and Autonomy in KG-Based AI

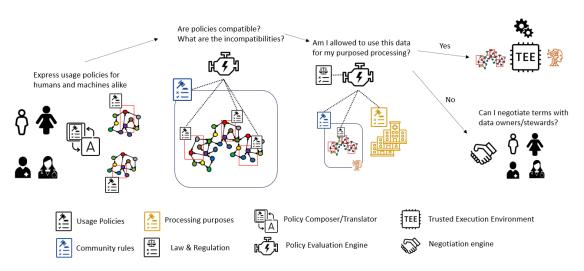


Figure 3 Machine-readable norms and policies to support self-determination. A Policy Composer/Translator assists individuals in writing data usage policies, communities in defining their rules, and organisations in declaring their data processing purposes in both human- and machine-readable formats. Policy Evaluation Engines assess the acceptability of perspectives in a community by evaluating policies and rules. Engines assess organisations' data usage compliance with regulations. If permitted, processing can occur in a Trusted Execution Environment, ensuring compliance. If not allowed, a Negotiation engine may be utilised to seek agreement with data owners/stewards under relevant regulations.

be able to engage in on-demand negotiation over policies, assisted by technology that ensures a safe and fair space and helps assess the compliance of negotiated terms with existing regulations. Negotiation between automated agents has been a topic of interest since the early 2000s, but in the context of self-determination, we must pay attention to the right balance between artificial representation and human involvement [5, 6].

4.1.3 Grounding based on our Illustrative Scenario

Figure 3 illustrates how machine-readable policies and norms can be used to support selfdetermination. Considering our illustrated scenario (Figure 2), individuals may want to establish policies to precisely define the subset of their PKGs to be shared with communities and what forwarding they allow. For example, share with the diabetes community my blood in sugar values measured by my connected device and the output of my AI healthcare assistant, or only share and forward anonymised aggregates to medical research institutions, or contact me for negotiation if the pharma company is interested in using my data for clinical studies. Communities may do the same, e.g. requiring specific data to be shared to join the community, but also requiring agreements in order to ensure that participants will abide by social and behavioural norms needed for self-regulation. Public and private organisations may need to adhere not only to privacy preferences and licences but also to various general regulations, e.g., the GDPR, the proposed AI Act in the EU or the Health Insurance Portability and Accountability Act (HIPAA)¹⁸ in the US, as well as domain-specific regulations (e.g. advanced therapy medicinal products¹⁹ and rare diseases²⁰).

¹⁸ https://www.hhs.gov/hipaa/index.html

¹⁹ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32007R1394

²⁰ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009H0703%2802%29

4.2 Decentralised Infrastructure

Over the last 15-20 years, a number of communities have come to accept that centralised computing systems, despite many benefits, can lead to issues such as the over-centralisation of power, the risk of single points of failure, potential abuse of personal data and creation of data silos which can inhibit innovation. A boon from this realisation is that we now have a number of technologies, standards, and approaches to decentralisation which offer benefits in terms of scalability, diversity, and privacy, as well as individually-centred flexibility and control, and is an appealing basis for maintaining and increasing trust, accountability, and autonomy with KG-based AI.

4.2.1 Personal Knowledge Graphs

The concept of a Personal Knowledge Graph (PKG) is that an individual can keep their personal or private data in a space belonging to them, rather than with siloed centralised service providers with limited access and control [7]. A Solid pod²¹ is an example of a PKG platform, and the key to the vision of Solid is that there should be standard interfaces and authorisation models to grant or deny access to the contents of a PKG at a granular level. This is argued in particular²² to enable a highly decentralised architecture for Web applications. Rather than a provider aggregating data from all users into a single location controlled by the provider and application code accessing such data there, an individual permits (or does not permit) Web applications of their choice to access whatever subsets of their data they decide from their PKG. As well as autonomy, this enables greater accountability since access to the PKG can be filtered via personal machine-readable policies at source, and activities can be tracked directly (e.g. [29]). Although PKGs offer great potential, they also come with challenges in terms of performance and scalability, as applications will need to interact with multiple distributed data sources as opposed to a single backend server. These challenges, however, may also simultaneously be opportunities for scalability trade-offs, querying over multiple low-powered data sources rather than a high-powered central one.

4.2.2 Distributed Ledger Technology

Distributed Ledger Technology (DLT) [104] promotes trust and empowerment through the replication of data across contributing nodes, which are geographically distributed across many sites, and the use of consensus algorithms which enable collective fair decision-making with no central control. Blockchains are a type of distributed ledger where an ever-growing list of records in blocks is tied together with cryptographic hashes, often, although not necessarily, associated with a securely exchangeable token system or "cryptocurrency". This technology rose to prominence following the release of Bitcoin [76] in 2008 - a blockchain-based currency that has now been adopted by El Salvador as their legal tender. Ethereum [115], a blockchain platform released in 2015, contains the notion of a "Smart Contract" [19] (originally coined in the 1990s by Nick Szabo [105]), which is a collection of code that executes in a fully decentralised way. Smart Contracts have been used to implement a range of decentralised applications, including Decentralised Autonomous Organisations (DAOs) [64], which are organisations where decisions are made through blockchain consensus mechanisms. The best-known example of a DAO was "The DAO", which at one point was worth more than \$70M; they have been applied to a number of different activities, including scholarly publishing [43]. Despite the fact that immutability and transparency guarantees offered by DLT are very attractive when dealing with personal data, both the ledgers and the smart

²¹ https://solidproject.org/

 $^{^{22}}$ https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/

9:14 Trust, Accountability, and Autonomy in KG-Based AI

contracts themselves will need to be protected against unauthorised access and usage. Personal data itself is neither stored in, or derivable from, immutable DLT records. Smart contracts may also introduce scalability issues: the default Ethereum model involves every contributing node executing every run of a smart contract and thus has inherent scale limitations. Relaxing this model may, however, affect trust.

4.2.3 Self Sovereign Identity

In the Web space, Self-Sovereign Identity (SSI) is being developed through a combination of Decentralised Identifiers (DIDs) [101] and Verifiable Credentials (VCs) [102], W3C standards for identity and verifiable attestation claims, respectively. DLT is one of the ways in which DIDs can be grounded, although, by design, the DID standard is open in terms of method. A DID is a URL (did:<method>:<...>) which can be resolved in a method-specific manner (e.g. HTTP(S) dereferencing, reading from a smart contract, etc.) to obtain a DID document, a Linked Data set containing information about digital identity in a standard form - for example, how to verify it (e.g. a public key), methods for communicating with the entity controlling it, and so on. DIDs enable SSI; the creation and use of DIDs are open and decentralised, and by using different DIDs with different audiences, individuals can minimise how easily their information can be tracked or correlated across services and can contextually and selectively disclose personal information as desired. This grants individuals significantly greater autonomy than current practices. There is a potential trade-off with trust and accountability of an individual when it comes to information that others need to rely on, which is that effective anonymity of a unique DID can be used to misrepresent oneself (e.g. fake a qualification or entitlement) or pretend to be someone else. VCs are a proposed solution to this. The VC data model is for sharing data alongside information that a recipient can use to verify its integrity or origin, such as a digital signature or DLT record. If a DID is presented to a service that is restricted to legal adults, for example, the DID owner may also present a VC issued by a government body confirming their adulthood; methods for selective disclosure supported by both DID and VC standards allow this to be done verifiably without requiring disclosure of real-world identity. These technologies are relatively new in comparison with standard digital identity models, and while intended and designed to address issues in those models, they may also introduce new difficulties or enable different vulnerabilities to, e.g. identity fraud, than current standards.

4.2.4 Federated Learning

In the context of data-driven AI and decentralised infrastructure, there are also techniques for decentralised machine learning. Federated Learning (FL) [118] is the idea that rather than aggregating training data in one location controlled by a model developer (thereby compromising subject privacy), data holders can run learning algorithms to generate model weights for their own data locally and privately, and then send only the weights to the developer to be incorporated into the larger model. An example might be a smartphone text prediction personalisation algorithm, where a user's own writing is used to generate predictive weights on the device, where periodic selections of these can be aggregated to improve general text prediction models. Refinements of FL approaches include sending not the actual learned model weights but a set of weights with statistically similar properties [112] to further reduce the risk of privacy breaches without affecting model performance. A related approach takes this concept even further, with the idea of embeddings in a larger model, e.g. "Textual Inversion" [34] to personalise large generative image diffusion models. The intuition here is that if someone wants certain personalised types of output from a generative AI, then if a model is sufficiently large, there is a good chance that the desired

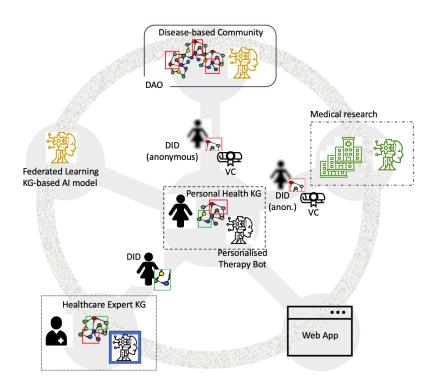


Figure 4 Decentralised Infrastructure supporting self-determination, shown from the perspective of one individual with a PKG²³. According to individual wishes, portions of the PKG can be shared either directly with a healthcare provider, with web applications for health, or indirectly with peer or research communities. Identity is via DIDs (anonymous in the latter cases), with VCs used for trustable selective disclosure. KG-based AI models can be trained and personalised in federated and private ways on knowledge from diverse sources.

concept already exists within it. More recently, the idea of federating for preserving privacy has been applied specifically to deep learning, in particular in the context of the Internet of Things. [119] proposes an architecture with a control layer including a distributed ledger, while [117] propose advanced cryptographic mechanisms to reduce the risk of privacy leaks, following more general approaches that apply either differential privacy, homomorphic encryption or secure multi-party computation. Federation also has the positive side effect of potentially speeding up model training when the privacy constraints allow for a helpful distribution of the process [11]. However, when opening the process to multiple parties, there are a number of attack vectors that do not exist in a centralised approach for which we need protection and pay a communication and computation overhead [44].

4.2.5 Grounding based on our Illustrative Scenario

A decentralised infrastructure supporting self-determination for our illustrative scenario (Figure 2) is depicted in Figure 4. Health data is highly sensitive and private, and individuals may want or need to interact with multiple services where it is relevant, including KG-based AI systems. It thus makes sense to create a personal health knowledge graph (PKG) to be a comprehensive

²³ The full picture would have knowledge exchange between multiple parties; to avoid an unreadable cluttered figure, this is left implied by the background network.

9:16 Trust, Accountability, and Autonomy in KG-Based AI

and interconnected representation of an individual's health information, including their medical history, lifestyle choices, genetic data, and real-time health monitoring data from IoT devices. Data from various sources, such as wearable devices, mobile applications, electronic health records, and even genomic sequencing, can be linked together to form a holistic view of an individual's health in such a personal health knowledge graph. An early example of a PKG was in [106], where medical, lifestyle, and IoT health monitoring data in a PKG was integrated into a (patient-focused) decision support system built around a public medically-curated KG representing cardiovascular risk factors, giving individuals the autonomy to gain deeper insights into their own health patterns and risks, identify correlations, and make more informed decisions.

More recently, BlockIoT [97, 98] aims to integrate health data seamlessly in a decentralised PKG using blockchain and KG technologies, addressing this trust aspect and using PKG-driven smart contracts to trigger the personalised recommendations for lifestyle modifications, medication adjustments, or even timely interventions by the healthcare providers. Furthermore, the PKG can serve as a powerful tool for healthcare beyond the individual. Communities of patients, providers, researchers, etc., or combinations thereof, can share knowledge about various aspects of, e.g., particular conditions, whether that is clinical evidence and best practice, peer advice and support on living with a condition, or data on novel or rare symptoms and side effects, with this knowledge used for support, care, or medical research across populations. De-identified and aggregated data from multiple individual KGs can be collected in community KGs, with trust securely established using DIDs and VCs, and accessed by community, practitioner, researcher, and service provider stakeholders, allowing for decentralised large-scale analysis and identification of broader health trends from multiple perspectives and intersecting factors. This can lead to advancements in disease prevention, treatment protocols, and the development of personalised medicine in a collaborative manner [99]. KG-based AI systems can be both trained and used across this ecosystem, with FL being applied to train larger models (e.g., the organisation models in Figure 2) and personalised embeddings used by individuals to get the best experience from their therapy bots and healthcare assistants while maintaining privacy and autonomy.

4.3 Decentralised KG Management

As the amount of data and knowledge grows exponentially, managing and harnessing this vast information becomes increasingly complex. Traditional centralised approaches to KG management face challenges in terms of scalability, privacy, and control over data, and to address these issues, decentralised KG management emerges as a promising solution. This section explores the key aspects and open challenges in decentralised KG management to enable trust, accountability, and self-determination for individuals in a rapidly evolving AI ecosystem.

4.3.1 Decentralised KG Access and Management

Efficient query processing infrastructures are fundamental for traversing decentralised KGs. There have been notable efforts such as Fedbench [92] in the past. However, these infrastructures should be capable of executing queries against the available KGs while respecting privacy and adhering to norms and policies. With the increasing emphasis on privacy protection with regulations such as GDPR, it is crucial to develop mechanisms that allow users to access and extract knowledge from KGs without compromising sensitive information or violating privacy regulations. Several research directions are worth considering to address the open challenges in decentralised KG management. Firstly, developing the formalisms to describe KG management semantically can provide a common ground for understanding and interoperability across different decentralised KG systems. Such formalisms can enable standardised representations of KGs in the form of

ontologies and facilitate seamless integration and collaboration among diverse knowledge sources. Architectures supporting new protocols and standards specific to decentralised KGs are essential for establishing interoperability and seamless communication between knowledge sources and systems. By defining and adopting common protocols and standards, decentralised KGs can collaborate more effectively, share insights, and facilitate cross-domain knowledge discovery.

Note that if we add LLMs to the picture, their current training and execution processes are currently centralised. Decentralised KG management is useful for providing transparency in data used for their training. For approaches involving the interaction between LLM and KGs, the transparency of the LLM itself still depends on the owner.

4.3.2 Provenance and Explanations

Furthermore, explainable methods for data integration and curation, as well as KG validation and distribution, such as the *Explanation Ontology* for user-centric AI, are necessary to ensure the reliability and accuracy of decentralised KGs [23]. By providing transparent and interpretable approaches, users can have better insights into knowledge integration and validation, enhancing trust and accountability of the knowledge contained in the KG and the insights derived. This is especially critical because, in decentralised KGs, data may come from various sources and be represented in different ways. The standardised framework provided in the Explanation Ontology for representing domain-specific explanations of KG entities and relationships helps users and applications understand the meaning and context of the data in the KG. Provenance and traceability also play a vital role in decentralised KG management. Establishing mechanisms to track and validate the origin, history, and lineage of knowledge within KGs is crucial for accountability and the ability to trace back the sources and transformations that contribute to the resulting knowledge. The W3C Provenance Data Management standards [71] provides the basis for encoding provenance attributes in KGs, and subsequent nanopublications specification [38] has gained a lot of traction in the biomedical domains. While these solutions exist, there needs to be a cohesive framework that ties together explanation provenance data management in a decentralised KG context and ensures that users can trace the origins, transformations, and sources of the data, which is crucial for trust, accountability, and data quality assurance. The W3C provenance data management suite of recommendations provides normative interoperable guidance on recording information about data sources, contributors, and how data is collected or transformed, making integrating heterogeneous data into a coherent KG easier. When data quality issues arise, users can trace back to the source of the problem and take corrective actions, ensuring the KG remains accurate and reliable. The W3C recommendations for decentralised provenance management provide a mechanism for attributing data to its sources or contributors. This attribution is essential for accountability, especially when multiple parties contribute to a KG.

4.3.3 Blockchain Technologies and Tokenomics

In recent years, the integration of blockchain technologies and tokenomics has gained attention in the context of decentralised KG management. Projects such as OriginTrail²⁴ have contributed to the development of ownable DKGs, which leverage blockchain's inherent properties to enhance trust, provenance, and accountability. By utilising blockchain, KG management systems can ensure the integrity and traceability of data and metadata across various nodes in the network. The OriginTrail protocol aims to create a trustless environment where data providers, consumers,

²⁴https://origintrail.io

9:18 Trust, Accountability, and Autonomy in KG-Based AI

and verifiers can interact and validate the authenticity and reliability of data stored within the knowledge graph. Their protocol issues tokens as incentives for data contributors, validators, and curators within the KG ecosystem. The integration of blockchain technologies and tokenomics in decentralised KG management addresses several critical aspects. Firstly, blockchain's immutability and transparency enable the traceability and provenance of data and metadata, ensuring accountability throughout the KG management pipeline. Secondly, the decentralised nature of blockchain mitigates single points of failure and promotes the distribution of knowledge and decision-making power among participants. This decentralised approach aligns with the principles of self-determination, empowering individuals to have control over their data and knowledge. By rewarding contributors, validators, and curators with tokens, these systems encourage continuous improvement, data quality assurance, and community engagement. Token-based economies can drive the development of sustainable KG management pipelines, enabling the growth and evolution of DKGs over time. However, the tokenomics have to be carefully designed and monitored to avoid the possibility contributors have a motivation (possibly extrinsic) to misbehave. There is also the risk that a sudden churn in blockchain participants impacts performance and availability. There is also the question of the performance of the consensus algorithm of a specific blockchain itself.

4.3.4 Grounding based on our Illustrative Scenario

An approach to decentralised knowledge graph management in the context of healthcare in our illustrative scenario (Figure 2), where users retain control over their personal information while benefiting from enhanced privacy measures and seamless collaboration in a community, is illustrated in Figure 5. At the heart of this framework lies the concept of PKGs, such as Solid, which empowers individuals to store and manage their personal health data securely. Central to the architecture are specific components aimed at safeguarding user privacy and ensuring data transparency. The process begins with knowledge sanitisation, which anonymises sensitive information and filters the data according to the user's preferences and data policies. These policies encompass not only globally recognised regulations like GDPR and HIPAA but also individual data policies, enabling users to set granular restrictions on how their data is used, such as opting out of genetic data usage for medical research. To ensure interoperability and standardisation, the creation of knowledge graphs leverages community-defined ontologies and vocabularies. These shared frameworks facilitate seamless integration and alignment of personal knowledge graphs within the broader ecosystem, promoting data exchange and collaboration. Users are incentivised to aggregate their knowledge graphs, contributing to the construction of community-based knowledge graphs focused on specific diseases. Through community-based verification, validation, and knowledge aggregation processes, these disease-based knowledge graphs are created, providing valuable insights and fostering collaborative efforts among healthcare professionals, researchers, and the wider community. Blockchain-based incentives drive user participation, rewarding both community users and healthcare experts for their verification, validation, and aggregation activities. The utilisation of an immutable ledger and verifiable credentials ensures the integrity and trustworthiness of the verification process. The validation process, powered by RDF SHACL and Shape descriptions, further enhances data quality and consistency, instilling confidence in the aggregated knowledge. The integrated knowledge graphs. encompassing personal, community-based, and healthcare expert knowledge, can be queried using federated querying mechanisms powered by SPARQL. This allows various institutions, including insurers, pharmaceutical companies, and medical research organisations, to access and leverage the rich insights stored within the knowledge graphs, enabling evidence-based decision-making and advancing medical research and healthcare practices. By combining decentralised knowledge graph management, user-centric privacy controls, and collaborative data sharing, this innovative

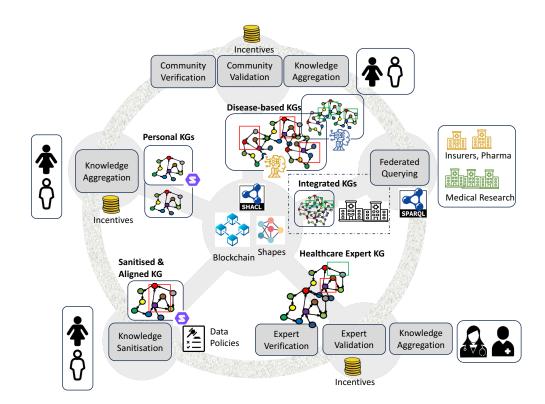


Figure 5 Decentralised KG Management Process in Healthcare. Emphasising user empowerment, privacy, and seamless collaboration, users maintain control over their personal health data through personal data stores like Solid, and community and healthcare experts enhance different facets of the KGs in the ecosystem. Decentralised KG management involves anonymisation, filtering based on data policies (including GDPR and HIPAA), and alignment with community-defined ontologies. Incentives, driven by blockchain technology, encourage user participation in aggregating KGs and incentivise healthcare professionals for verification, validation, and aggregation activities. SHACL shapes ensure KG validation and federated querying mechanisms enable access to the KGs to stakeholders, e.g. insurers, pharma, and medical research organisations. Integrated KGs are iteratively generated; they comprise a federation of KGs that may be autonomous, distributed, and heterogeneous. A federation query engine enables the traversal of these integrated and connected KGs to provide useful insights to the stakeholders involved.

framework represents a significant step forward in transforming decentralised KG management, fostering a secure, privacy-enhanced environment that empowers users, facilitates collaboration, and drives advancements in domains such as medical knowledge and patient care.

4.4 Explainable Neuro-symbolic AI

Neuro-symbolic systems go beyond generating explanations solely based on the trained model or the individual results derived from applying the model to specific data. They can produce symbolic explanations capturing the essence of an AI model itself. These explanations can be classified as either *instance-level* explanations generated for each specific result of the model or *model-level* explanations of the structure of a learned model. Previous work on the role of KGs in AI has focused on explainability. [61] frames explainability as a dimension of *trustable* AI and presents challenges, existing approaches, limitations and opportunities for KGs to bring explainable AI to the right level of semantics and interpretability. [107] and [88] conducted independent systematic reviews of existing explainable AI systems to characterise KGs' impact. These results put into perspective the role of KGs in providing symbolic reasoning and learning capabilities with the potential to be precise, as shown by Akrami et al. [2], in addition to being explainable.

9:20 Trust, Accountability, and Autonomy in KG-Based AI

4.4.1 Reasoning and AI

Despite the unquestionable reasoning features of symbolic systems and the studies reporting limitations of LLMs in human-like tasks (e.g., explanations, memories, and reasoning over factual statements) [40], there is an ongoing debate about LLM's reasoning their causal inference capabilities [52]. Although LLMs excel at certain reasoning tasks, they do poorly in others, raising the question of whether they genuinely engage in causal reasoning or merely function as unreliable mimics, generating memorised responses (e.g. [45]). Methods to reason can be roughly divided into methods using only the LLM itself (e.g. with prompt engineering) and methods combining the LLM with an external reasoner and/or external source of knowledge (e.g. a Knowledge Graph) [86]. Our vision posits that external help will always be needed, especially for concrete use cases. There are discussions about the need for knowledge graphs in the era of LLMs. Sun et al. [103] and Dong [31] report on an empirical assessment of ChatGPT [93] with respect to DBpedia, illustrating the need of symbolic systems that over-fit for the truth whenever factual statements are collected from KGs. In addition, symbolic approaches can support sanity checking and be easily auditable and traceable. These features position the combination of both approaches in neuro-symbolic AI as a feasible option to provide KG-based AI. Neuro-symbolic AI delivers the basis to integrate the discrete methods implemented by symbolic AI with high-dimensional vector spaces managed by LLMs. They must decide when and how to combine both systems, e.g., following a principled integration (combining neural and symbolic while maintaining a clear separation between their roles and representations) or integrated (e.g. a symbolic reasoner integrated into the tuning process of an LLM).

4.4.2 Trust and AI

Trust in AI systems stems from various factors, including transparency, reproducibility, predictability, and explainability. Neuro-symbolic systems play a vital role in enhancing trustworthiness by enabling communication between modules and facilitating tracing. Modularity enables the specification, verification, and validation of each component and its interactions. As a result, a system's behaviour can be traced and validated. Specifically, within the domain of KG-based AI for self-determination, the seamless integration of KGs and symbolic semantic reasoning offers a comprehensive and unified perspective on curated knowledge. This integration holds immense value in addressing critical tasks such as validating, refuting, and explaining incorrect, biased, or misleading information that may potentially be generated by LLMs. By combining symbolic reasoning over KGs with LLMs, the propagation of misinformation can be mitigated while simultaneously enhancing the transparency and trustworthiness of AI-generated outputs. Consequently, KG-based AI systems can effectively emulate human behaviour by subjecting mistakes arising from false or incomplete information to a process of validation and enrichment using curated and potentially peer-reviewed sources of knowledge [109].

4.4.3 Quality and AI

A notable application of KGs in neuro-symbolic AI is as a source of informative prior knowledge to increase the quality of machine learning models. An example is the work by Rivas et al. [89], where a deductive database, expressed in Datalog, establishes an axiomatic system of the pharmacokinetic behaviour of a treatment's drugs and enables the deduction of new drug-drug interactions in cancer treatments. This prior knowledge plays a crucial role in elucidating the characteristics of a therapy and justifying its efficacy by considering all the interactions and the dynamic movement of drugs within the body. It encompasses factors such as the absorption, bioavailability, metabolism, and excretion of drugs over time. A KG embedding model improves its prediction

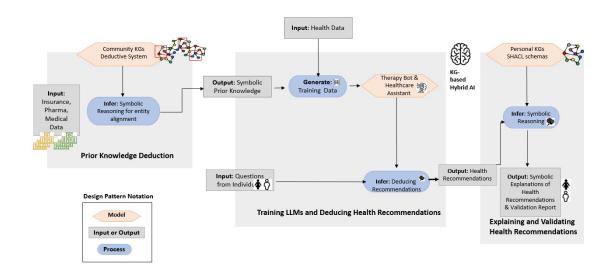


Figure 6 Design Patterns for Hybrid AI. Extension of patterns by van Bekkum et al. [109] for running example in Figure 2. The patterns represent an explainable system with prior knowledge created by the alignments of data from health-related data sources (e.g. insurance, pharma, and medical data).

of the effectiveness of a treatment based on prior knowledge, which encodes statements about a treatment's characteristics; these statements are inferred by a deductive system which comprises the symbolic component of the hybrid approach. An approach for explaining link prediction (e.g. [90]) allows the justification of why this added prior knowledge affects the model's decisions, potentially improving trust in the model's results.

4.4.4 Grounding Based on Our Illustrative Scenario

Grounding on the example presented in Figure 2, when individuals and professionals engage in communities with bots and assistants powered by AI models, it is critical to ensure the transparency of their decision-making process. However, despite the increasing focus on LLMs in healthcare and their continual improvement in terms of precision and accuracy [100], their outcomes can still be susceptible to hidden biases and a lack of traceability [62]. To tackle these challenges, the utilisation of a neuro-symbolic system can enhance LLMs by incorporating reasoning capabilities. This system operates as a deductive system on a user's Knowledge Graph (KG). These hybrid AI systems can be effectively modelled using patterns proposed by [109]. Figure 6 depicts a pattern describing a hybrid AI system that enhances the explainability of the LLMs described in our running example. At the community level, symbolic reasoning applied to the ontology of shared PKGs can generate prior knowledge, enabling precise and concrete questioning of an LLM and providing additional contextual information. Moreover, a symbolic system facilitates the linking of shared PKGs with corresponding entities in KGs related to insurance, pharmaceuticals, and medical research. By incorporating this prior knowledge, the LLM's answers are improved and validated with the assistance of the symbolic system. The systems operating at the community level and involving heterogeneous sources can be described using the explainable system with prior knowledge pattern; data alignments comprising prior knowledge enhance contextual knowledge provided to the therapy bot, facilitating thoughtful health recommendations.

5 Proposed KG-based AI for Self-determination Research Agenda

In this section, we derive a set of requirements concerning KG-based AI for self-determination and map them to the concrete research goals introduced at the start of this vision paper.

5.1 Trust, Accountability, and Autonomy Foundational Goals

In the following, we highlight five open research challenges and opportunities in each of our proposed foundational topics (machine-readable norms and policies, decentralised infrastructure, decentralised KG management, and explainable neuro-symbolic AI). Considering the complex nature of each of these requirements, an assessment of the maturity of existing technologies with respect to the various requirements is beyond the scope of a vision paper.

- **MRP1:** Seamless policy translation. There is a need for humans to express policies in machinereadable format and for machines to express them in natural language or via appropriate visualisations. A major challenge involves checking that machine-readable policies faithfully represent their human-readable counterpart.
- **MRP2:** Multi-level policy evaluation. Several policy languages exist. However, many of them do not have corresponding enforcement mechanisms. Given that usage constraints, community rules, and regulations operate at different yet interconnected levels, there is a need to devise effective and efficient enforcement and/or compliance-checking strategies.
- **MRP3:** Negotiation. Facilitate autonomy via fair and safe negotiation between individuals, communities, and organisations. Here, there is a need to study the benefits and trade-offs between merely assisting humans in making decisions and developing automated approaches that alleviate individuals from constant affirmations (e.g. the cookie problem).
- MRP4: Compliance verification. Provide support for both ex-ante and ex-post compliance checking mechanisms. Despite their potential, it remains to be seen which machine-readable agreements can actually be enforced by TEEs. Additionally, in scenarios where it does not pay data processors to cheat, game theoretic approaches could be used to underpin honours-based compliance-checking.
- **MRP5:** Data misuse detection. Instil trust and ensure accountability in KG-based AI by developing mechanisms that can detect if any party violated policies and norms. In this context, causal reasoning and explanations could potentially be used to both detect misuse and to better understand the root cause.

5.1.1 Decentralised Infrastructure

- **DI1: Comprehensive recording.** A DLT can provide an immutable ledger, but work remains on how best to connect KG-based AI activities, e.g. to a possible federated query engine.
- **DI2:** Personalised tracing. Providing individual and community owners of PKGs with personalised traces of how acquired data was processed and used will involve dis-aggregating KG-processing and inferencing according to different user data and ensuring that privacy is not violated when individual results are returned.
- **D13:** "Decency" check. There is a need for easy-to-use services that allow users and communities to check if an organisation has behaved in a "decent" way when it processes acquired data. Research here will examine how "decency" can be defined and validated by comparing PKG declarations of use (e.g. policies) with generated traces of use.

- **DI4:** Interoperability. Develop mechanisms that facilitate comprehensive, interoperable identification of human and machine participants in KG-based AI processes. For example, users and communities will wish to know and be able to validate claims that a data request comes from a particular organisation, unit and even individual KG processor. This will provide a foundation for accountability at all levels of granularity.
- **DI5:** Self-sovereignty. True self-sovereign KG-based AI needs to be: (i) based upon easy-to-use self-sovereign identities and data management; and (ii) capable of supporting the continuous monitoring of organisational behaviours in a transparent fashion.

5.1.2 Decentralised KG Management

- **DKG1:** Knowledge Sanitisation. Develop robust techniques for knowledge sanitisation that ensure user privacy by anonymising and filtering sensitive information based on data policies. These policies can be regulations such as GDPR and HIPAA, as well as individual-level data policies enforced at their personal data store, empowering users to specify their sharing preferences and control the aspects of data they disclose.
- **DKG2:** Knowledge Graph Aggregation. Design and implement mechanisms to encourage users to contribute their PKGs towards aggregated knowledge graphs, such as a concerted effort towards developing specific disease KGs. Blockchain-based incentive models that reward users for contributing to constructing such knowledge graphs, fostering collaborative efforts, and enriching the overall quality of shared knowledge are components of this goal.
- **DKG3:** Knowledge Verification. Develop community-based and expert processes to verify the knowledge available in the global KGs. On the community front, it is critical to ensure that a knowledge item that was previously contributed through an individual has not been altered (either through error or with malicious intent), for instance, via blockchain primitives, as explained in the previous section.
- **DKG4:** Knowledge Validation. Validation of knowledge is paramount to ensure KG interoperability and the consumption of knowledge in target applications. By employing RDF and SHACL technologies, we ensure that the DKGs across different data stores conform to a specific template, thus enabling their integration with community-supported KGs.
- **DKG5: Federated Querying.** Explore and implement federated querying mechanisms, specifically utilising SPARQL, to enable efficient querying across integrated KGs. This process includes developing techniques to support various institutions, such as insurers, pharmaceutical companies, and medical research organisations, accessing and extracting insights from the knowledge graphs to enhance decision-making and advance their respective domains.

5.1.3 Explainable Neuro-Symbolic AI

- XNS1: User-dependent Recommendations. Neuro-symbolic systems need to be empowered to present results transparently to the users according to their interests. For example, in our illustrative scenario (Figure 2), an individual may not expect the same level of detail in a health recommendation as a medical doctor or a community representative.
- XNS2: Adaptive Hybrid AI. Define models that can adaptively combine predictive models with logical reasoning, encompassing abilities such as generalisation and causal inference. For accountability, the neuro-symbolic system should explain when the combination of logical reasoning with a therapy bot or healthcare assistant will be beneficial. For autonomy, the neuro-symbolic system should include the user in the loop and consider their opinion in this decision. Finally, trust requires verifying and validating these decisions.

9:24 Trust, Accountability, and Autonomy in KG-Based AI

Table 1 Mapping of foundational requirements to pillars. A checkmark signifies that the corresponding requirement is necessary for answering a research question related to a pillar. (Q1) is related to Trust, (Q2) to Accountability, and (Q3) to Autonomy.

	Machine-readable norms and policies				
	Trust	Accountability	Autonomy		
MRP1	\checkmark		\checkmark		
MRP2	\checkmark		\checkmark		
MRP3			\checkmark		
MRP4	\checkmark		\checkmark		
MRP5	\checkmark	\checkmark	\checkmark		
	Decentralised Infrastructure				
	Trust	Accountability	Autonomy		
DI1	\checkmark		✓		
DI2	\checkmark		✓		
DI3	\checkmark		\checkmark		
DI4		✓			
DI5			\checkmark		
		nagement.			
	Trust	Accountability	Autonomy		
DKG1			✓		
DKG2	,	√	✓		
DKG3	√				
DKG4	\checkmark				
DKG5		√ 	~		
	Explainable Neuro-Symbolic AI				
	Trust	Accountability	Autonomy		
XNS1		✓			
XNS2	✓	✓			
XNS3	\checkmark	✓			
XNS4		✓			
XNS5		\checkmark			

- **XNS3: Contextual-based Hybrid AI.** Equip neuro-symbolic systems with contextual knowledge, reasoning capabilities, and causal inference to effectively evaluate the strengths and limitations of machine learning components. This goal empowers the system to identify optimal combinations of statistical and symbolic AI methods, requiring the definition of causal models on top of KGs capable of combining reasoning over KGs with causal inference.
- XNS4: Symbolic Reasoning. Employ inference processes, both inductive and deductive, on knowledge graphs to enable ML models, and LLMs in particular, to adjust hyper-parameters and a model's configuration to new environments (i.e., Personal, community-based, and integrated healthcare KGs) and provide explanations for their decisions. Despite the advances of Automated Machine Learning (AutoML) systems (e.g., AutoML²⁵ and AutoWeka [58], to the best of our knowledge, there are no developments for AutoML over KGs or for neuro-symbolic systems, which will enhance accountability, autonomy, and trust.
- **XNS5: Learning Transparency.** Investigate if existing XAI mechanisms can be tailored for learning transparency, such that it is possible to explain what action was taken, how the decision making was performed, and why this was perceived as the outcome offering the greatest expected satisfaction.

²⁵https://www.automl.org/

5.2 AI for Self-determination

The identified foundational research topic challenges and opportunities can be used to better contextualise concrete goals in relation to trust, accountability, and autonomy from a KG-based AI for self-determination perspective. An overview of this mapping, which is depicted in Table 1, is provided by attempting to answer the overarching questions that guide our vision paper.

- (Q1) What are the key requirements for an AI system to produce trustable results? From a trust perspective, it is important that machine-readable policies faithfully represent the human-readable policies (MRP1) in a manner that can be verified automatically (MRP2). Regardless of whether systems are automated or semi-automated, we need to be able to verify that processes behave as expected (MRP4) and any misuse can be detected and rectified (MRP5). Trust could potentially be facilitated via auditing (DI1) and tracing (DI2), as well as certification mechanisms that support decency checks (DI3) and (semi-)automated knowledge verification (DKG3) and validation (DKG4) techniques. While human involvement is paramount to establishing trust in adaptive (XNS2) and contextualised (XNS3) hybrid AI.
- (Q2) How can Al be made accountable for its decision-making? The first step to achieving accountability is to ensure it is possible to detect if any party violated policies and norms (MRP5) and that the recommendations given and decisions taken using both induction and deduction (XNS4) are comprehensible from a user perspective, for instance via user focuses recommendations (XNS1), providing explanations for recommendations and decisions (XNS2), facilitating learning transparency (XNS5), and contextualisation based on causal inference (XNS3). Considering that machines can only work with the knowledge that it has at hand, it is important that systems are able to integrate knowledge from disparate sources (DI4) and are capable of querying (DKG5) and aggregating (DKG2) relevant sources.
- (Q3) How can citizens maintain autonomy as users or subjects of KG-based AI systems? Citizens' autonomy in a KG-based AI context is necessary to ensure that humans are able to control not only who has access to their personal data, but also that its usage is in line with existing regulatory requirements. This could be achieved with automated compliance checking (MRP4) and misuse detection (MRP5) built on top of machine-readable policies (MRP1) and evaluation mechanisms (MRP2). Negotiation could potentially enable organisations to gain access to better quality data (MRP3) or to foster collaboration via aggregation (DKG2) and strong privacy guarantees via anonymisation (DKG1). Meanwhile, self-sovereign identities (DI5), auditing (DI1), tracing (DI2), and decency certification (DI3) have a major role to play when it comes to continuous monitoring.

6 Conclusion

This paper presents a compelling argument for integrating KG-based AI to empower individuals' selfdetermination and benefit society. This overarching goal is supported by three fundamental pillars: trust, accountability, and autonomy. We advocate that the foundations of these pillars require focused research in four areas: machine-readable norms and policies, decentralised infrastructure, decentralised KG management, and explainable neuro-symbolic AI. By drawing on a concrete scenario within the healthcare domain, we demonstrate the relevance of each foundational topic and outline a comprehensive research agenda for each of them.

We aspire for the insights presented in this paper to catalyse the creation of AI services that genuinely support citizens while upholding their rights. Responsible advancement of the foundational topics is crucial to ensure that future KG-based AI solutions are comprehensive and possess the qualities of being traceable, verifiable, and interpretable. It is essential that relevant

9:26 Trust, Accountability, and Autonomy in KG-Based AI

legislation, such as the EU AI Act, provides clear guidance to steer the development of these forthcoming applications, emphasising the need for accurate, reliable, and transparent AI systems. Within this context, we recognise the Semantic Web community as uniquely positioned to drive transformative change and contribute solutions that illuminate the workings of opaque AI models. Through this concerted effort, we envision a paradigm shift in KG management and analytics that establishes KG-based AI to empower individuals in their pursuit of self-determination.

— References

- 1 Gianmaria Ajani, Guido Boella, Luigi Di Caro, Livio Robaldo, Llio Humphreys, Sabrina Praduroux, Piercarlo Rossi, and Andrea Violato. The European Taxonomy Syllabus: A multi-lingual, multi-level ontology framework to untangle the web of European legal terminology. *Applied Ontology*, 11(4):325–375, 2016. doi:10.3233/A0-170174.
- 2 Farahnaz Akrami, Mohammed Samiul Saeef, Qingheng Zhang, Wei Hu, and Chengkai Li. Realistic re-evaluation of knowledge graph completion methods: An experimental study. In Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020, pages 1995–2010, 2020. doi: 10.1145/3318464.3380599.
- 3 Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Oasis legalruleml. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, ICAIL '13, pages 3–12, 2013. doi:10.1145/2514601.2514603.
- 4 Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. LegalRuleML: Design Principles and Foundations, pages 151–188. Springer International Publishing, 2015. doi:10.1007/978-3-319-21768-0_6.
- 5 Reyhan Aydogan and Catholijn M. Jonker. A survey of decision support mechanisms for negotiation. In Rafik Hadfi, Reyhan Aydogan, Takayuki Ito, and Ryuta Arisaka, editors, *Recent Advances in Agent-Based Negotiation: Applications and Competition Challenges, ACAN@IJCAI 2022, Vienna, Austria, July 24, 2022*, volume 1092 of *Studies in Computational Intelligence*, pages 30–51. Springer, 2022. doi:10.1007/978-981-99-0561-4_3.
- 6 Tim Baarslag, Michael Kaisers, Enrico H. Gerding, Catholijn M. Jonker, and Jonathan Gratch. Selfsufficient, Self-directed, and Interdependent Negotiation Systems: A Roadmap Toward Autonomous Negotiation Agents, pages 387–406. Springer International Publishing, 2022. doi:10.1007/978-3-030-76666-5_18.
- 7 Krisztian Balog and Tom Kenter. Personal Knowledge Graphs: A Research Agenda. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, pages 217–220, sep 2019. doi:10.1145/3341981.3344241.
- 8 Chitta Baral. Knowledge Representation, Reasoning and Declarative Problem Solving. Cambridge University Press, 2003. doi:10.1017/ CB09780511543357.
- 9 Cesare Bartolini, Antonello Calabrò, and Eda Marchetti. Enhancing business process modelling

with data protection compliance: An ontologybased proposal. In Paolo Mori, Steven Furnell, and Olivier Camp, editors, Proceedings of the 5th International Conference on Information Systems Security and Privacy, ICISSP 2019, Prague, Czech Republic, February 23-25, 2019, pages 421–428. SciTePress, 2019. doi:10.5220/ 0007392304210428.

- 10 Davide Basile, Claudio Di Ciccio, Valerio Goretti, and Sabrina Kirrane. Blockchain based resource governance for decentralized web environments. CoRR, abs/2301.06919, 2023. doi: 10.48550/ARXIV.2301.06919.
- 11 Tal Ben-Nun and Torsten Hoefler. Demystifying parallel and distributed deep learning: An in-depth concurrency analysis. ACM Comput. Surv., 52(4), aug 2019. doi:10.1145/3320060.
- 12 Tim Berners-Lee and Mark Fischetti. Weaving the web: the past, present and future of the World Wide Web by its inventor. Orion Business Books, reprinted edition, 2000.
- 13 Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. Scientific American, 284(5):34– 43, 2001. nopublisher: Scientific American, a division of Nature America, Inc. URL: http: //www.jstor.org/stable/26059207.
- 14 Guido Boella, Luigi Di Caro, and Valentina Leone. Semi-automatic knowledge population in a legal document management system. Artif. Intell. Law, 27(2):227–251, 2019. doi:10.1007/S10506-018-9239-8.
- 15 Piero Bonatti, J.L. De Coi, Daniel Olmedilla, and Luigi Sauro. A rule-based trust negotiation system. *IEEE Transactions on Knowledge and Data Engineering*, 22(11):1507–1520, 2010. doi: 10.1109/TKDE.2010.83.
- 16 Grady Booch, Francesco Fabiano, Lior Horesh, Kiran Kate, Jonathan Lenchner, Nick Linck, Andrea Loreggia, Keerthiram Murugesan, Nicholas Mattei, Francesca Rossi, and Biplav Srivastava. Thinking fast and slow in AI. In Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, pages 15042–15046. AAAI Press, 2021. doi:10.1609/ AAAI.V35117.17765.
- 17 Anna Breit, Laura Waltersdorfer, Fajar J. Ekaputra, Marta Sabou, Andreas Ekelhart, Andreea Iana, Heiko Paulheim, Jan Portisch, Artem Revenko, Annette ten Teije, and Frank van Harmelen. Combining machine learning and se-

mantic web: A systematic mapping study. *ACM Comput. Surv.*, mar 2023. Just Accepted. doi: 10.1145/3586163.

- 18 Katrina Brooker. "I Was Devastated": Tim Berners-Lee, the Man Who Created the World Wide Web, Has Some Regrets. Vanity Fair, 2018. URL: https://www.vanityfair.com/news/2018/ 07/the-man-who-created-the-world-wide-webhas-some-regrets.
- 19 Vitalik Buterin. A next-generation smart contract and decentralized application platform. Technical Report 37, Ethereum Foundation, 2014. URL: https://cobesto.com/preview-file/ whitepaper-Ethereum.pdf.
- 20 Elena Cabrio, Alessio Palmero Aprosio, and Serena Villata. These are your rights - A natural language processing approach to automated RDF licenses generation. In The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings, volume 8465 of Lecture Notes in Computer Science, pages 255-269, 2014. doi:10.1007/978-3-319-07443-6_18.
- 21 Cristian Cardellino, Serena Villata, Laura Alonso Alemany, and Elena Cabrio. Information extraction with active learning: A case study in legal text. In Alexander F. Gelbukh, editor, Computational Linguistics and Intelligent Text Processing - 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part II, volume 9042 of Lecture Notes in Computer Science, pages 483-494. Springer, 2015. doi:10.1007/978-3-319-18117-2_36.
- 22 Ilias Chalkidis, Charalampos Nikolaou, Panagiotis Soursos, and Manolis Koubarakis. Modeling and querying greek legislation using semantic web technologies. In The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I, pages 591–606, 2017. doi:10.1007/978-3-319-58068-5_36.
- 23 Shruthi Chari, Oshani Seneviratne, Daniel M. Gruen, Morgan A. Foreman, Amar K. Das, and Deborah L. McGuinness. Explanation ontology: A model of explanations for user-centered AI. In Jeff Z. Pan, Valentina A. M. Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal, editors, The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, volume 12507 of Lecture Notes in Computer Science, pages 228–243. Springer, 2020. doi:10.1007/978-3-030-62466-8_15.
- 24 Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. Technical report, Google Research, 2022. doi: 10.48550/arXiv.2204.02311.
- 25 Giovanni Luca Ciampaglia, Alexios Mantzarlis, Gregory Maus, and Filippo Menczer. Research challenges of digital misinformation: Toward a trustworthy web. AI Mag., 39(1):65–74, 2018. doi:10.1609/AIMAG.V39I1.2783.

- 26 A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In 2022 ACM Conference on Fairness, Accountability, and Transparency, pages 864–876, jun 2022. doi:10.1145/3531146.3533150.
- 27 Uwe Der, Stefan Jähnichen, and Jan Sürmeli. Selfsovereign identity - opportunities and challenges for the digital revolution. *CoRR*, abs/1712.01767, 2017. doi:10.48550/arXiv.1712.01767.
- 28 Johannes Dimyadi, Guido Governatori, and Robert Amor. Evaluating LegalDocML and LegalRuleML as a standard for sharing normative information in the aec/fm domain. In *Proceedings of the 34th International Conference of CIB W78*, volume 1, pages 637–644, 2017. doi:10.24928/JC3-2017/ 0012.
- 29 John Domingue, Aisling Third, Maria-Esther Vidal, Philipp D. Rohde, Juan Cano, Andrea Cimmino, and Ruben Verborgh. Trusting decentralised knowledge graphs and web data at the web conference. In Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben, editors, Companion Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023, pages 1422– 1423. ACM, 2023. doi:10.1145/3543873.3589756.
- 30 Xin Luna Dong. Generations of knowledge graphs: The crazy ideas and the business impact. Proc. VLDB Endow., 16(12):4130-4137, 2023. doi: 10.14778/3611540.3611636.
- 31 Xin Luna Dong. Generations of knowledge graphs: The crazy ideas and the business impact. CoRR, abs/2308.14217, 2023. doi:10.48550/ ARXIV.2308.14217.
- 32 Beatriz Esteves, Harshvardhan J. Pandit, and Víctor Rodríguez-Doncel. ODRL profile for expressing consent through granular access control policies in solid. In *IEEE European Sym*posium on Security and Privacy Workshops, EuroS&P 2021, Vienna, Austria, September 6-10, 2021, pages 298–306. IEEE, 2021. doi:10.1109/ EUROSPW54576.2021.00038.
- 33 Enrico Francesconi. A description logic framework for advanced accessing and reasoning over normative provisions. Artif. Intell. Law, 22(3):291–311, 2014. doi:10.1007/S10506-014-9158-2.
- 34 Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit Haim Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. In *The Eleventh International Conference on Learning Representations, ICLR* 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net, 2023. URL: https://openreview.net/ pdf?id=NAQvF08TcyG.
- 35 Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. Clingo = ASP + control: Preliminary report. CoRR, abs/1405.3694, 2014. URL: http://arxiv.org/abs/1405.3694, arXiv:1405.3694.
- 36 Cindy Gordon. ChatGPT Is The Fastest Growing App In The History Of Web Applications, feb 2023. URL: https://www.forbes.com/sites/ cindygordon/2023/02/02/chatgpt-is-the-

fastest-growing-ap-in-the-history-of-webapplications/?sh=5cee3a30678c.

- 37 Guido Governatori, Mustafa Hashmi, Ho-Pun Lam, Serena Villata, and Monica Palmirani. Semantic business process regulatory compliance checking using LegalRuleML. In Eva Blomqvist, Paolo Ciancarini, Francesco Poggi, and Fabio Vitali, editors, Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, volume 10024 of Lecture Notes in Computer Science, pages 746–761, 2016. doi:10.1007/978-3-319-49004-5_48.
- 38 Paul Groth, Andrew Gibson, and Jan Velterop. The anatomy of a nanopublication. Inf. Serv. Use, 30(1-2):51-56, 2010. doi:10.3233/ISU-2010-0613.
- 39 Susanne Guth, Gustaf Neumann, and Mark Strembeck. Experiences with the enforcement of access rights extracted from odrl-based digital contracts. In Moti Yung, editor, Proceedings of the 2003 ACM workshop on Digital rights management 2003, Washington, DC, USA, October 27, 2003, pages 90–102. ACM, 2003. doi:10.1145/947380.947392.
- 40 Kristian J. Hammond and David B. Leake. Large language models need symbolic AI. In Artur S. d'Avila Garcez, Tarek R. Besold, Marco Gori, and Ernesto Jiménez-Ruiz, editors, Proceedings of the 17th International Workshop on Neural-Symbolic Learning and Reasoning, La Certosa di Pontignano, Siena, Italy, July 3-5, 2023, volume 3432 of CEUR Workshop Proceedings, pages 204– 209. CEUR-WS.org, 2023. URL: https://ceurws.org/Vol-3432/paper17.pdf.
- 41 Dick Hardt. The OAuth 2.0 Authorization Framework. RFC 6749, oct 2012. doi:10.17487/RFC6749.
- 42 Giray Havur, Simon Steyskal, Oleksandra Panasiuk, Anna Fensel, Víctor Mireles, Tassilo Pellegrini, Thomas Thurner, Axel Polleres, and Sabrina Kirrane. Automatic license compatibility checking. In Mehwish Alam, Ricardo Usbeck, Tassilo Pellegrini, Harald Sack, and York Sure-Vetter, editors, Proceedings of the Posters and Demo Track of the 15th International Conference on Semantic Systems co-located with 15th International Conference on Semantic Systems (SE-MANTiCS 2019), Karlsruhe, Germany, September 9th - to - 12th, 2019, volume 2451 of CEUR Workshop Proceedings. CEUR-WS.org, 2019. URL: https://ceur-ws.org/Vol-2451/paper-13.pdf.
- 43 Michal Robert Hoffman, Luis-Daniel Ibáñez, and Elena Simperl. Scholarly publishing on the blockchain - from smart papers to smart informetrics. *Data Sci.*, 2(1-2):291–310, 2019. doi:10.3233/DS-190018.
- 44 Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, Philip S. Yu, and Xuyun Zhang. Membership inference attacks on machine learning: A survey. ACM Comput. Surv., 54(11s), sep 2022. doi:10.1145/3523273.
- 45 Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. In Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Associ*ation for Computational Linguistics: ACL 2023,

Toronto, Canada, July 9-14, 2023, pages 1049–1065. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.FINDINGS-ACL.67.

- 46 Renato Ianella and Serena Villata. ODRL Information Model, 2018. URL: https://www.w3.org/TR/ odrl-model/.
- 47 Eleni Ilkou. Personal Knowledge Graphs: Use Cases in e-learning Platforms. In Companion Proceedings of the Web Conference 2022, pages 344– 348, apr 2022. doi:10.1145/3487553.3524196.
- 48 Lalana Kagal. Rei : A Policy Language for the Me-Centric Project. Technical report, HP Labs, sep 2002. URL: http://www.hpl.hp.com/techreports/ 2002/HPL-2002-270.html.
- 49 Daniel Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, 2011.
- 50 Daniel Kazenoff, Oshani Seneviratne, and Deborah L. McGuinness. Semantic graph analysis to combat cryptocurrency misinformation on the web. In Manolis Koubarakis, Harith Alani, Grigoris Antoniou, Kalina Bontcheva, John G. Breslin, Diego Collarana, Elena Demidova, Stefan Dietze, Simon Gottschalk, Guido Governatori, Aidan Hogan, Freddy Lécué, Elena Montiel-Ponsoda, Axel-Cyrille Ngonga Ngomo, Sofia Pinto, Muhammad Saleem, Raphaël Troncy, Eleni Tsalapati, Ricardo Usbeck, and Ruben Verborgh, editors, Joint Proceedings of Workshops AI4LEGAL2020, NLI-WOD, PROFILES 2020, QuWeDa 2020 and SEMIFORM2020 Colocated with the 19th International Semantic Web Conference (ISWC 2020), Virtual Conference, November, 2020, volume 2722 of CEUR Workshop Proceedings, pages 168-176. CEUR-WS.org, 2020. URL: https://ceurws.org/Vol-2722/semiform2020-paper-3.pdf.
- 51 Ankesh Khandelwal, Jie Bao, Lalana Kagal, Ian Jacobi, Li Ding, and James A. Hendler. Analyzing the AIR language: A semantic web (production) rule language. In Pascal Hitzler and Thomas Lukasiewicz, editors, Web Reasoning and Rule Systems Fourth International Conference, RR 2010, Bressanone/Brixen, Italy, September 22-24, 2010. Proceedings, volume 6333 of Lecture Notes in Computer Science, pages 58–72. Springer, 2010. doi:10.1007/978-3-642-15918-3_6.
- 52 Emre Kiciman, Robert Ness, Amit Sharma, and Chenhao Tan. Causal reasoning and large language models: Opening a new frontier for causality. CoRR, abs/2305.00050, 2023. doi:10.48550/ ARXIV.2305.00050.
- 53 Sabrina Kirrane. Intelligent software web agents: A gap analysis. Journal of Web Semantics, 71:100659, nov 2021. doi:10.1016/ J.WEBSEM.2021.100659.
- 54 Sabrina Kirrane, Javier D. Fernández, Piero Bonatti, Uros Milosevic, Axel Polleres, and Rigo Wenning. The SPECIAL-K personal data processing transparency and compliance platform, 2021. doi:10.48550/arXiv.2001.09461.
- 55 Sabrina Kirrane, Alessandra Mileo, and Stefan Decker. Access control and the Resource Description Framework: A survey. *Semantic Web*, 8(2):311–352, dec 2016. doi:10.3233/SW-160236.
- 56 Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions.

In Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pages 1885–1894, 06–11 August 2017. URL: https: //proceedings.mlr.press/v70/koh17a.html.

- 57 Boshko Koloski, Timen Stepisnik Perdih, Marko Robnik-Sikonja, Senja Pollak, and Blaz Skrlj. Knowledge graph informed fake news classification via heterogeneous representation ensembles. *Neurocomputing*, 496:208–226, 2022. doi:10.1016/ J.NEUCOM.2022.01.096.
- 58 Lars Kotthoff, Chris Thornton, Holger H. Hoos, Frank Hutter, and Kevin Leyton-Brown. Autoweka 2.0: Automatic model selection and hyperparameter optimization in weka. *Journal of Machine Learning Research*, 18(25):1–5, 2017. URL: http://jmlr.org/papers/v18/16-261.html.
- 59 Ziyi Kou, Lanyu Shang, Yang Zhang, and Dong Wang. HC-COVID: A hierarchical crowdsource knowledge graph approach to explainable COVID-19 misinformation detection. Proc. ACM Hum. Comput. Interact., 6(GROUP):36:1–36:25, 2022. doi:10.1145/3492855.
- 60 Tahu Kukutai and Donna Cormack. "Pushing the space" Data sovereignty and self-determination in Aotearoa NZ. In Indigenous Data Sovereignty and Policy, Routledge Studies in Indigenous Peoples and Policy. Routledge, 2021. URL: https://www.taylorfrancis.com/chapters/oaedit/10.4324/9780429273957-2/pushing-spacetahu-kukutai-donna-cormack.
- 61 Freddy Lecue. On the role of knowledge graphs in explainable AI. Semantic Web, 11(1):41–51, jan 2020. doi:10.3233/SW-190374.
- 62 Hanzhou Li, John T Moon, Saptarshi Purkayastha, Leo Anthony Celi, Hari Trivedi, and Judy W Gichoya. Ethics of large language models in medicine and medical research. *The Lancet, Digital Health*, 5, 2023. doi:10.1016/S2589-7500(23)00083-3.
- 63 James Lighthill. Artifical Intelligence: A General Survey. Technical report, UK Science Research Council, 1974. URL: http://www.chiltoncomputing.org.uk/inf/literature/reports/ lighthill_report/p001.htm.
- 64 Lu Liu, Sicong Zhou, Huawei Huang, and Zibin Zheng. From technology to society: An overview of blockchain-based dao. *IEEE Open Journal* of the Computer Society, 2:204–215, 2021. doi: 10.1109/0JCS.2021.3072661.
- 65 James H Lubowitz. Chatgpt, an artificial intelligence chatbot, is impacting medical literature. Arthroscopy, 39(5):1121-1122, 2023. doi: 10.1016/j.arthro.2023.01.015.
- 66 George F. Luger. Modern AI and How We Got Here. In Knowing our World: An Artificial Intelligence Perspective, pages 49–74. Springer International Publishing, 2021. doi:10.1007/978-3-030-71873-2_3.
- 67 Essam Mansour, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, and Tim Berners-Lee. A demonstration of the solid platform for social web applications. In Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao, editors, *Proceedings of*

the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume, pages 223–226. ACM, 2016. doi:10.1145/2872518.2890529.

- 68 Mohit Mayank, Shakshi Sharma, and Rajesh Sharma. DEAP-FAKED: knowledge graph based approach for fake news detection. In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2022, Istanbul, Turkey, November 10-13, 2022, pages 47-51. IEEE, 2022. doi: 10.1109/ASONAM55673.2022.10068653.
- 69 Dhruv Mehrotra. ICE Records Reveal How Agents Abuse Access to Secret Data, apr 2023. URL: https://www.wired.com/story/ice-agentdatabase-abuse-records.
- 70 Merriam-Webster.com Dictionary. Trust. https:// www.merriam-webster.com/dictionary/trust. Accessed 14 Jul. 2023.
- 71 Paolo Missier, Khalid Belhajjame, and James Cheney. The W3C PROV family of specifications for modelling provenance metadata. In Giovanna Guerrini and Norman W. Paton, editors, Joint 2013 EDBT/ICDT Conferences, EDBT '13 Proceedings, Genoa, Italy, March 18-22, 2013, pages 773–776. ACM, 2013. doi:10.1145/ 2452376.2452478.
- 72 Melanie Mitchell and David C. Krakauer. The debate over understanding in ai's large language models. *Proceedings of the National Academy of Sciences*, 120(13):e2215907120, 2023. doi: 10.1073/pnas.2215907120.
- 73 Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. Auditing large language models: a three-layered approach. AI Ethics, 2023. doi:10.1007/s43681-023-00289-2.
- 74 Benjamin Moreau, Patricia Serrano-Alvarado, Matthieu Perrin, and Emmanuel Desmontils. A license-based search engine. In Pascal Hitzler, Sabrina Kirrane, Olaf Hartig, Victor de Boer, Maria-Esther Vidal, Maria Maleshkova, Stefan Schlobach, Karl Hammar, Nelia Lasierra, Steffen Stadtmüller, Katja Hose, and Ruben Verborgh, editors, The Semantic Web: ESWC 2019 Satellite Events - ESWC 2019 Satellite Events, Portorož, Slovenia, June 2-6, 2019, Revised Selected Papers, volume 11762 of Lecture Notes in Computer Science, pages 130–135. Springer, 2019. doi:10.1007/978-3-030-32327-1_26.
- 75 Iman Naja, Milan Markovic, Peter Edwards, and Caitlin Cottrill. A Semantic Framework to Support AI System Accountability and Audit. In *The Semantic Web*, volume 12731, pages 160–176. Springer International Publishing, 2021. Series Title: Lecture Notes in Computer Science. doi: 10.1007/978-3-030-77385-4_10.
- 76 Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. URL: https:// bitcoin.org/bitcoin.pdf.
- 77 María Navas-Loro and Cristiana Santos. Events in the legal domain: first impressions. In Víctor Rodríguez-Doncel, Pompeu Casanovas, Jorge González-Conejero, and Elena Montiel-Ponsoda, editors, Proceedings of the 2nd Workshop on Technologies for Regulatory Compliance co-located with

the 31st International Conference on Legal Knowledge and Information Systems (JURIX 2018), Groningen, The Netherlands, December 12, 2018, volume 2309 of CEUR Workshop Proceedings, pages 45–57. CEUR-WS.org, 2018. URL: https: //ceur-ws.org/Vol-2309/05.pdf.

- 78 Arttu Oksanen, Minna Tamper, Jouni Tuominen, Eetu Mäkelä, Aki Hietanen, and Eero Hyvönen. Semantic Finlex: Transforming, Publishing, and Using Finnish Legislation and Case Law As Linked Open Data on the Web, volume 317 of Frontiers in Artificial Intelligence and Applications, pages 212– 228. IOS Press, 2018. doi:10.3233/FAIA190023.
- 79 Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B. Norton, N. Cameron Russell, Peter Story, Joel R. Reidenberg, and Norman M. Sadeh. Privonto: A semantic framework for the analysis of privacy policies. *Semantic Web*, 9(2):185–203, 2018. doi:10.3233/SW-170283.
- 80 Monica Palmirani, Guido Governatori, Antonino Rotolo, Said Tabet, Harold Boley, and Adrian Paschke. Legalruleml: Xml-based rules and norms. In Frank Olken, Monica Palmirani, and Davide Sottara, editors, Rule-Based Modeling and Computing on the Semantic Web, 5th International Symposium, RuleML 2011- America, Ft. Lauderdale, FL, Florida, USA, November 3-5, 2011. Proceedings, volume 7018 of Lecture Notes in Computer Science, pages 298-312. Springer, 2011. doi:10.1007/978-3-642-24908-2_30.
- 81 Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Legal ontology for modelling GDPR concepts and norms. In Monica Palmirani, editor, Legal Knowledge and Information Systems JURIX 2018: The Thirty-first Annual Conference, Groningen, The Netherlands, 12-14 December 2018, volume 313 of Frontiers in Artificial Intelligence and Applications, pages 91–100. IOS Press, 2018. doi: 10.3233/978-1-61499-935-5-91.
- 82 Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Pronto: Privacy ontology for legal reasoning. In Electronic Government and the Information Systems Perspective - 7th International Conference, EGO-VIS 2018, Regensburg, Germany, September 3-5, 2018, Proceedings, volume 11032 of Lecture Notes in Computer Science, pages 139–152, 2018. doi:10.1007/978-3-319-98349-3_11.
- 83 Harshvardhan J. Pandit, Kaniz Fatema, Declan O'Sullivan, and Dave Lewis. Gdprtext - GDPR as a linked data resource. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, pages 481–495, 2018. doi:10.1007/978-3-319-93417-4_31.
- 84 Eli Pariser. The filter bubble: How the new personalized web is changing what we read and how we think. Penguin, 2011. URL: https://dl.acm.org/doi/10.5555/2361740.
- 85 Neoklis Polyzotis and Matei Zaharia. What can Data-Centric AI Learn from Data and ML Engineering? Arxiv, 2021. doi:10.48550/ ARXIV.2112.06439.

- 86 Shuofei Qiao, Yixin Ou, Ningyu Zhang, Xiang Chen, Yunzhi Yao, Shumin Deng, Chuanqi Tan, Fei Huang, and Huajun Chen. Reasoning with language model prompting: A survey. In Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, pages 5368– 5393. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.294.
- 87 Umair Qudus, Michael Röder, Muhammad Saleem, and Axel-Cyrille Ngonga Ngomo. Hybridfc: A hybrid fact-checking approach for knowledge graphs. In Ulrike Sattler, Aidan Hogan, C. Maria Keet, Valentina Presutti, João Paulo A. Almeida, Hideaki Takeda, Pierre Monnin, Giuseppe Pirrò, and Claudia d'Amato, editors, The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings, volume 13489 of Lecture Notes in Computer Science, pages 462–480. Springer, 2022. doi:10.1007/978-3-031-19433-7_27.
- 88 Enayat Rajabi and Kobra Etminani. Knowledgegraph-based explainable AI: A systematic review. Journal of Information Science, page 016555152211128, sep 2022. doi:10.1177/ 01655515221112844.
- 89 Ariam Rivas, Diego Collarana, Maria Torrente, and Maria-Esther Vidal. A neurosymbolic system over knowledge graphs for link prediction. Semantic Web journal, To appear. URL: https://www.semantic-webjournal.net/content/neuro-symbolic-systemover-knowledge-graphs-link-prediction-0.
- 90 Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. Explaining link prediction systems based on knowledge graph embeddings. In SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022, pages 2062–2075, 2022. doi:10.1145/3514221.3517887.
- 91 Andrei Vlad Sambra, Essam Mansour, Sandro Hawke, Maged Zereba, Nicola Greco, Abdurrahman Ghanem, Dmitri Zagidulin, Ashraf Aboulnaga, and Tim Berners-Lee. Solid: a platform for decentralized social applications based on linked data. Technical report, MIT CSAIL & Qatar Computing Research Institute, Tech. Rep., 2016. URL: http://emansour.com/research/ lusail/solid_protocols.pdf.
- 92 Michael Schmidt, Olaf Görlitz, Peter Haase, Günter Ladwig, Andreas Schwarte, and Thanh Tran. Fedbench: A benchmark suite for federated semantic data query processing. In Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Fridman Noy, and Eva Blomqvist, editors, The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I, volume 7031 of Lecture Notes in Computer Science, pages 585–600. Springer, 2011. doi:10.1007/978-3-642-25073-6_37.
- 93 John Schulman, Barret Zoph, Christina Kim, Jacob Hilton, Jacob Menick, Jiayi Weng, Juan Felipe Ceron Uribe, Liam Fedus, Luke Metz, Michael

Pokorny, et al. Chatgpt: Optimizing language models for dialogue. Accessed: 23/11/2023. URL: https://openai.com/blog/chatgpt.

- 94 Oshani Seneviratne. Data provenance and accountability on the web. Provenance in Data Science: From Data Models to Context-Aware Knowledge Graphs, pages 11–24, 2020. doi:10.1007/978-3-030-67681-0_2.
- 95 Oshani Seneviratne. Blockchain for social good: Combating misinformation on the web with AI and blockchain. In WebSci '22: 14th ACM Web Science Conference 2022, Barcelona, Spain, June 26 - 29, 2022, pages 435–442. ACM, 2022. doi:10.1145/3501247.3539016.
- 96 Lanyu Shang, Ziyi Kou, Yang Zhang, Jin Chen, and Dong Wang. A privacy-aware distributed knowledge graph approach to qois-driven COVID-19 misinformation detection. In 30th IEEE/ACM International Symposium on Quality of Service, IWQoS 2022, Oslo, Norway, June 10-12, 2022, pages 1–10. IEEE, 2022. doi:10.1109/ IWQ0S54832.2022.9812879.
- 97 Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Blockiot: blockchain-based health data integration using iot devices. In AMIA Annual Symposium Proceedings, volume 2021, page 1119. American Medical Informatics Association, 2021. URL: https://knowledge.amia.org/74229-amia-1.462266/t003-1.4626466/t003-1.4626467/3577334-1.4626522/3577640-1.4626519.
- 98 Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Blockiot-retel: Blockchain and iot based readexecute-transact-erase-loop environment for integrating personal health data. In Yang Xiang, Ziyuan Wang, Honggang Wang, and Valtteri Niemi, editors, 2021 IEEE International Conference on Blockchain, Blockchain 2021, Melbourne, Australia, December 6-8, 2021, pages 237-243. IEEE, 2021. doi:10.1109/BLOCKCHAIN53845.2021.00039.
- 99 Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Collaboratively learning optimal patient outcomes using smart contracts in limited data settings. In IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2022, Arlington, VA, USA, November 17-19, 2022, pages 133-137. IEEE, 2022. URL: https://ieeexplore.ieee.org/ document/9983613.
- 100 Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaekermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Andrew Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale R. Webster, Gregory S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Towards expert-level medical question answering with large language models. CoRR, abs/2305.09617, 2023. doi:10.48550/ARXIV.2305.09617.
- 101 Manu Sporny, Amy Guy, Markus Sabadello, Drummond Reed, Orie Steele, and Christopher Allen. Decentralized Identifiers (DIDs), 2022. URL: https://www.w3.org/TR/did-core/.

- 102 Manu Sporny, David Longley, and David Chadwick. Verifiable Credentials Data Model, 2022. URL: https://www.w3.org/TR/vc-data-model/.
- 103 Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, and Xin Luna Dong. Head-to-tail: How knowledgeable are large language models (llm)? A.K.A. will llms replace knowledge graphs? CoRR, abs/2308.10168, 2023. doi:10.48550/ARXIV.2308.10168.
- 104 Ali Sunyaev. Internet Computing Principles of Distributed Systems and Emerging Internet-Based Technologies. Springer, 2020. doi:10.1007/978-3-030-34957-8.
- 105 Nick Szabo. Formalizing and securing relationships on public networks. *First Monday*, 2(9), 1997. doi:10.5210/FM.V2I9.548.
- 106 Allan Third, George Gkotsis, Eleni Kaldoudi, George Drosatos, Nick Portokallidis, Stefanos Roumeliotis, Kalliopi Pafili, and John Domingue. Integrating medical scientific knowledge with the semantically quantified self. In Paul Groth, Elena Simperl, Alasdair J. G. Gray, Marta Sabou, Markus Krötzsch, Freddy Lécué, Fabian Flöck, and Yolanda Gil, editors, The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part I, volume 9981 of Lecture Notes in Computer Science, pages 566–580, 2016. doi:10.1007/978-3-319-46523-4_34.
- 107 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. Artificial Intelligence, 302:103627, jan 2022. doi:10.1016/J.ARTINT.2021.103627.
- 108 Andrzej Uszok, Jeffrey M. Bradshaw, Renia Jeffers, Niranjan Suri, Patrick J. Hayes, Maggie R. Breedy, Larry Bunch, Matt Johnson, Shriniwas Kulkarni, and James Lott. Kaos policy and domain services: Toward a description-logic approach to policy representation, deconfliction, and enforcement. In 4th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2003), 4-6 June 2003, Lake Como, Italy, page 93. IEEE Computer Society, 2003. doi:10.1109/POLICY.2003.1206963.
- 109 Michael van Bekkum, Maaike de Boer, Frank van Harmelen, André Meyer-Vitali, and Annette ten Teije. Modular design patterns for hybrid learning and reasoning systems. Appl. Intell., 51(9):6528– 6546, 2021. doi:10.1007/S10489-021-02394-3.
- 110 Marina De Vos, Sabrina Kirrane, Julian A. Padget, and Ken Satoh. ODRL policy modelling and compliance checking. In Paul Fodor, Marco Montali, Diego Calvanese, and Dumitru Roman, editors, *Rules and Reasoning - Third International Joint Conference, RuleML+RR 2019, Bolzano, Italy, September 16-19, 2019, Proceedings, volume 11784* of Lecture Notes in Computer Science, pages 36–51. Springer, 2019. doi:10.1007/978-3-030-31095-0_3.
- 111 Denny Vrandecic and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. Commun. ACM, 57(10):78–85, 2014. doi:10.1145/2629489.
- 112 Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, Farhad Farokhi, Shi Jin, Tony Q. S. Quek, and H. Vincent Poor. Federated learning with differential privacy: Algorithms

and performance analysis. *IEEE Trans. Inf. Forensics Secur.*, 15:3454–3469, 2020. doi: 10.1109/TIFS.2020.2988575.

- 113 Mika Westerlund. The emergence of deepfake technology: A review. Technology innovation management review, 9(11), 2019. URL: https://www.proquest.com/scholarly-journals/ emergence-deepfake-technology-review/ docview/2329154005/se-2.
- 114 Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, 32(4):791–813, 2023. doi:10.1007/S00778-022-00775-9.
- 115 Gavin Wood. Ethereum: A secure decentralised generalised transaction ledger. Technical report, Ethereum Foundation, 2014. URL: https: //gavwood.com/paper.pdf.
- 116 Adam Z. Wyner and Wim Peters. Lexical semantics and expert legal knowledge towards the identification of legal case factors. In Radboud Winkels, editor, Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference on Legal Knowledge and Information Systems, Liverpool, UK, 16-17 December 2010, volume 223 of Frontiers in Artificial Intelligence and Applications, pages 127–136. IOS Press, 2010. doi:10.3233/978-1-60750-682-9-127.
- 117 Guowen Xu, Hongwei Li, Yun Zhang, Shengmin Xu, Jianting Ning, and Robert H. Deng. Privacypreserving federated deep learning with irregular users. *IEEE Transactions on Dependable*

and Secure Computing, 19(2):1364–1381, 2022. doi:10.1109/TDSC.2020.3005909.

- 118 Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. Technol., 10(2):12:1–12:19, 2019. doi: 10.1145/3298981.
- 119 Bo Yin, Hao Yin, Yulei Wu, and Zexun Jiang. Fdc: A secure federated deep learning mechanism for data collaborations in the internet of things. *IEEE Internet of Things Journal*, 7(7):6348–6359, 2020. doi:10.1109/JI0T.2020.2966778.
- 120 Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. A survey of large language models. CoRR, abs/2303.18223, 2023. doi:10.48550/ARXIV.2303.18223.
- 121 Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G. Parker, and Munmun De Choudhury. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson, editors, Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023, pages 436:1–436:20. ACM, 2023. doi: 10.1145/3544548.3581318.

Multilingual Knowledge Graphs and Low-Resource Languages: A Review

Lucie-Aimée Kaffee¹ $\square \land \bigcirc$ Hasso-Plattner Institut, Potsdam, Germany

Russa Biswas 🖂 🕩 Hasso-Plattner Institut, Potsdam, Germany C. Maria Keet ⊠© University of Cape Town, South Africa

Edlira Kalemi Vakaj 🖂 🗅 Birmingham City University, UK

Gerard de Melo ⊠© Hasso-Plattner Institut, Potsdam, Germany University of Potsdam, Germany

— Abstract -

There is a lack of multilingual data to support applications in a large number of languages, especially for low-resource languages. Knowledge graphs (KG) could contribute to closing the gap of language support by providing easily accessible, machine-readable, multilingual linked data, which can be reused across applications. In this paper, we provide an overview of work in the domain of multilingual KGs with a focus on low-resource languages. We review the current state of multilingual KGs

along with the different aspects that are crucial for creating KGs with language coverage in mind. Special consideration is given to challenges particular to low-resource languages in KGs. We further provide an overview of applications that yield multilingual KG information as well as downstream applications reusing such multilingual data. Finally, we explore open problems regarding multilingual KGs with a focus on low-resource languages.

methodologies \rightarrow Semantic networks

Keywords and phrases knowledge graphs, multilingual, low-resource languages, review

Digital Object Identifier 10.4230/TGDK.1.1.10

Category Vision

Received 2023-06-30 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge



Introduction

For a wide range of applications, including chatbots and search engines, it is important to support a large variety of languages, as this enables more people to access these applications in their native language. However, currently, many applications only provide support for a highly restricted number of languages. While there are over 7,000 languages spoken in the world², applications such as Amazon Alexa or Google Home support 8 or 16, respectively, at the time of writing.³

Link to Amazon Alexa; Link to Google Home.



[©] Lucie-Aimée Kaffee, Russa Biswas, C. Maria Keet, Edlira Kalemi Vakaj, and Gerard de Melo; Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 10, pp. 10:1–10:19

Transactions on Graph Data and Knowledge

¹ Corresponding author

 $[\]mathbf{2}$ https://www.ethnologue.com/

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

10:2 Multilingual Knowledge Graphs

This drastically limits the accessibility for speakers of languages not among this small set of languages. Providing access to information across languages may be crucial whenever one wishes to make high-quality factual information available to people across the globe. However, for most of the world's languages, the lack of coverage in terms of available data, linguistic models (be they rules-based or data-driven), and the broader tooling ecosystem is daunting, and, therefore, there is insufficient information readily available to easily extend existing applications to the language. We refer to such languages as *low-resource languages*. There have been efforts to make low-resource languages more accessible, as exhibited by initiatives such as Masakhane [27] and SaDiLaR⁴, which have created datasets that aim to assist in making language applications available across a greater number of languages. However, the vast majority of language resources are in and for English. For example, on the web at large, 58.8% of websites are estimated to be in English⁵, which is also reflected in the size of available corpora⁶. Low-resource languages are barely represented in these sorts of collections. Therefore, models trained on web data are prone to suffering from a severe lack of information in the majority of languages.

One possible way of addressing the issue of a lack of multilingual data is to rely on knowledge graphs (KGs), which store knowledge as graph-structured data [35]. In a data-to-text generation approach, KGs can be used as a source of information for newly generated text across languages [50]. The central storage of language-agnostic information enables downstream applications to provide knowledge, such as in the form of text, for a wide range of language communities. We describe some of the downstream use-cases of multilingual KGs in Section 4.2.

Despite being machine-readable, knowledge graphs also harbour substantial natural language information. Entities and relationships in a knowledge graph generally have natural language labels and often also natural language descriptions. Indeed, among the most prominent knowledge graphs, many provide such natural language labels and descriptions in a multitude of different languages and are thus also a valuable direct source of multilingual data [49]. For instance, a KG may capture that the chemical element gold is called *gold* in English, *altın* in Turkish, *igolide* in isiXhosa, *bulawan* in Cebuano, and so on. It becomes more challenging when there is no simple 1:1 mapping, such as pet being lexicalized as *pet* in English but de facto only described in isiZulu, for instance, as *isilwane sasekhaya* (an animal that is of the home). Similarly, in agglutinating languages, morphemes are strung together to create a new concept that may be translated as a phrase or sentence, such as *umagwazephindelela* for "one who is not satisfied with a single achievement" (a "persistent fighter")⁷ [17].

By storing information in the KG in a machine-readable form, the KG can hold information about entities, such as the gold chemical symbol, irrespective of whether this information exists in each of the languages. This helps eliminate language barriers and ensures information is available even in languages where it might otherwise be absent.

In this paper, we explore the current state of multilingual data in knowledge graphs (Section 2), particular challenges regarding low-resource languages (Section 3), approaches to increase language coverage in KGs (Section 4.1), applications using multilingual KGs (Section 4.2), and finally we propose open questions regarding the multilingual support in KGs at large (Section 5).

⁴ South African Centre for Digital Language Resources with its language resources repository available at https://repo.sadilar.org/.

⁵ https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/, as of January 2023

⁶ e.g.: https://www.sketchengine.eu/corpora-and-languages/corpus-list/

⁷ From: u-/uma- noun prefix for noun class 1a or u- noun prefix for noun class 1a + -ma (v) "stand", -gwaza (v) "stab", "slaughter", verb with -e ending subjunctive mood, -phinda (v) "repeat", "do again", -phindelela (v) "do again and again" (note the -el- applicative verb extension to -phinda)

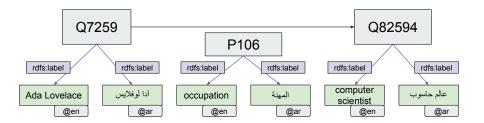


Figure 1 Modelling of multilingual knowledge using a unified knowledge graph with attributes given in multiple different languages, as exemplified by Wikidata. Each entity has a single ID shared across languages. In order to ensure stable and language-neutral IDs, the ID scheme does not include any natural language part intended for human consumption. Instead, a separate entity label is provided for each relevant language (in this example English and Arabic). These are connected using the rdfs:label property in the RDF version of Wikidata.

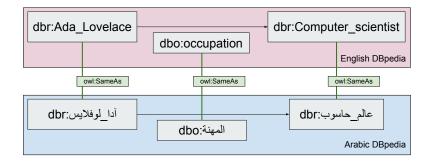


Figure 2 Modelling of multilingual knowledge using separate interconnected entity IDs, as exemplified by DBpedia. A separate entity ID is defined for each language. These are connected by the owl:sameAs property, which indicates that they in fact refer to the same entity. In practice, different languages differ widely in their coverage and hence not all information is mirrored across languages as in this example.

2 State of Multilingual KGs

To describe the challenges and opportunities of multilingual KGs, we first provide a brief overview on the current state of knowledge graphs regarding their language coverage. Gracia et al. describe their vision of a multilingual web of data in 2012 as having "(i) linguistic information for data and vocabularies in different languages, (ii) mappings between data with labels in different languages, and (iii) services to dynamically access and traverse Linked Data across different languages" [31]. Over 10 years later, we seek to understand what and how much of it has been realised. To gain insights into the state of KGs, we observe multilingual data in KGs from different angles. First, we examine how the different modelling choices and ontologies, i.e., the structure-giving elements of the KG, enable or impede inclusion of multilingual information. Further, it is crucial to assess how many and which languages are supported to what extent. Therefore, we provide an overview of work analysing language coverage across different KGs. Finally, we provide insights on the linguistic information available in KGs in the form of lexicographical data.

2.1 Modelling Multilingual Knowledge

There are, depending on the particular KG, different ways to store multilingual information [29]. A common choice is to assume that KG entities can be shared across languages, and thus for a given KG entity, one can provide relevant language-specific information such as labels in multiple

languages. An example of this is illustrated in Figure 1 for the prominent Wikidata KG. An alternative is to essentially create a different KG (or sub-KG) for each language and then interlink these KGs. An example of this is illustrated in Figure 2 for the DBpedia KG.

The availability of labels can differ substantially across languages. Of course, this may simply be due to incomplete coverage, as we discuss in Section 2.3. In some cases, however, this may stem from linguistic differences in lexicalization or naming: A concept or entity may possess a name in one language but not in another, in which case it is sometimes referred to as a lexical gap. For instance, Runyankore does not have a single word for "pet" but only a description, and isiZulu's *ingcula* does not have a translation into English other than a description of the object (a small bladed hunting spear). At the instance level, additional modelling efforts may be required, if Runyankore-speaking people were to be interested in representing in the KG that Lassie and Scooby Doo are pets, say, or for the development of a KG about Shaka Zulu's armoury.

In the literature, there is a range of guidelines indicating how a KG should represent multilingual information. These include the use of standard ontologies, recommendations for the data itself as well as how the data should be modelled. Across the literature, the following guidelines have been identified, as also summarised in previous work [44].

- **Stable identifiers** [7]: Each entity is identified in the KG with an identifier (ID). These are crucial as the way to access information about an entity. Many authors recommend adopting Unique Resource Identifiers (URIs) or Internationalized Resource Identifiers (IRIs), as used on the Web, for better interoperability. Some KGs identify entities in the graph with identifiers incorporating natural language. For instance, DBpedia uses IRIs such as http://dbpedia.org/resource/Ada_Lovelace. Such identifiers are human-readable, and hence easy to interpret for humans, and they also take the function of a label. However, the fact that the natural language portions of such IDs carry meaning can also be a disadvantage. As identifiers are expected to remain stable, such IDs are unable to reflect potential changes in the entity label. For example, if the name of a property changes, the entire structure of the KG would be affected [85]. In contrast, opaque IDs are ones that are not easily readable by humans. For example, they could have a unique identifier for a concept in the form of letters and/or numbers, such as the Wikidata ID Q7259. As such IDs do not reflect the label of the entity, if the entity's name changes, the ID is not affected. Moreover, such entity IDs can be more readily shared across languages.
- **Label coverage** [21, 90, 16]: As natural language labels are the way humans access information in the KG and interact with it, it is important that entities, classes, and properties in the KG are labelled. This ensures that information is human-readable and can be displayed to a user.
- **Language tags** [78]: When labelling an entity, it is important to indicate the language in which this label is provided. Even in monolingual KGs, language tags can be valuable, in order to avoid conflating different languages when fusing information from different language sources, especially when automatically merging KGs or when operating across multiple knowledge graphs in different languages through, e.g., federated queries. Language tags can help applications decide which label should be displayed to which user.
- **Language coverage** [31, 44]: For a multilingual KG it is crucial that entities are labelled across a large number of languages for which a relevant label exists. Only such thorough labelling can ensure access of all users to all information, independent of the languages they speak. It also enables KGs to readily make other applications multilingual.
- Monolingual islands [31]: When parts of the graph are labelled in only one language, they can form monolingual islands. This can happen because knowledge is prevalent in one cultural context and therefore is yet to be translated, such as governmental initiatives publishing structured data in only their native language. Monolingual islands can lead to worse access to a diverse set of knowledge across the graph and should be addressed when working on a multilingual KG.

- **Reusing existing vocabulary**: In the creation of a KG, it is essential to reuse existing vocabulary, or ontology, to describe the schema of the KG. Especially for label and language information in a KG, it is crucial for easier integration of multiple, complementary KGs. For more information on ontologies, see Section 2.2.
- Unambiguity [21]: As there are multiple complementary as well as overlapping ontologies, it is crucial to make sure to avoid ambiguity. According to the Semantic Web standards, it is recommended to use the labelling property rdfs:label to provide natural language labels of entities. Furthermore, it is often recommended to provide a single preferred label in a given language per entity, while using other properties such as *alias* to describe alternative names for a concept.

Conceptual differences between different languages may lead to modelling challenges. An important example is the subclass/superclass relationship (which corresponds to hypo-/hypernymy as a lexical relation). For instance, consider the single concept and word for river in English, whereas in the French language and presumably also the corresponding conceptualisation, one distinguishes rivers that flow into other rivers and those that flow into the sea (*fleuve* versus *rivière*) [62]. This may occur similarly for relational properties [54]. A common solution is to treat incompatible concepts across different languages as distinct entries in the KG. Thus, one can avoid conflating the English concept of river with the two French ones and optionally also explicitly describe how the different entries relate to one another cross-lingually. It may also be the case that in one language only a verbal form of a concept exists, i.e., it is assumed to be only a relational property, and in another language it is nominalised, i.e., exists as a unary object or object type only, such that heterogeneous mappings may be needed [26].

2.2 Multilingual Ontologies

Ontologies define the structure of a KG by setting standards for the different properties or relationships to be used, and, for example, the classes used in a KG and across the web of data [36]. The W3C Web Ontology Language (OWL) is a particularly prominent formalism to define ontologies across different KGs to ensure the interoperability of different ontologies [37].

To be able to create multilingual KGs, it is crucial to understand how multilingualism is addressed in the ontologies, be it in the formalism or in a declarative model associated with it, how natural languages are incorporated, and how translation of entities may be recorded.

Gillis-Webber and Keet [29] survey multilinguality in and for ontologies. They grouped the literature and practices in multilingual ontologies into three types of modelling multilingual information: (1) using multilingual labels, i.e., labelling one entity across languages using a labelling property and mentioning the language of the label in the label string (such as in Wikidata in the example in Figure 1), (2) linguistic models (covered in Section 2.4), and (3) mapping-based approach, i.e., creating one entity per language and linking them across languages to each other with an appropriate property (such as in DBpedia in the example in Figure 2). They observed that there is currently a limited uptake of multilingual ontologies at large. Further, limitations of each of the three described ways of modelling multilingual data, such as "accurate representation of languages that require grammatical features such as inflected forms and gender" [30], are yet to be addressed, which apply at least in part also to KGs (discussed in Section 3.2 below).

As ontologies can be defined with the creation of each new KG anew, a number of monolingual ontologies in different languages exist that could be interlinked so as to be more interoperable. Efforts in this direction align ontologies across languages, making it possible to link different KGs across languages [41] or translate labels across languages to make ontologies interoperable and reusable [22].

2.3 Language Coverage Across KGs

Covering a large number of languages in the data in any KG is crucial to be able to support a wide range of languages in downstream applications. For example, in the domain of Question Answering (QA), using a multilingual KG facilitates easy switching between languages and finding the KG best suited for the language requirements of an application and its user base [46].

A range of studies describes the language coverage of KGs to understand how well they currently serve multilingual users. Studies have found that, across numerous language graphs, there is a lack of non-English information in the form of labels. Further, the most used labelling property is rdfs:label, and across the web of data at large there is a widespread lack of labels, i.e., a large share of entities are not labelled at all. Additionally, most entities are labelled in only one language [49, 21, 44].

Ell et al. [21] and Kaffee and Simperl [49] developed frameworks to analyse language and label coverage in the web of data, in their case a collection of KGs available online. Zaveri et al. [90] survey data quality metrics and describe a metric for *human readable labelling*, which characterises the coverage of entities by labels. This metric was picked up and made actionable by Debattista et al. [16].

Wikidata specifically shows a slightly different distribution than the web of data at large. While there is still an English bias in the KG, there is a higher degree of language diversity overall [47, 44]. This more varied representation of languages could have different potential causes; for one, there is a multilingual community editing the KG, which leads to a larger number of perspectives added to the graph, compared to KGs that are automatically assembled or mostly contributed to by a single community situated in one part of the world [48]. This international community has been recruited to a large extent from the famous sister project of Wikidata, Wikipedia, where an international community already edits knowledge in their respective language [79]. Further, the community of humans is supported by so-called *bots*, which are automated tools that import and edit knowledge on Wikidata across languages [45]. These approaches can be instructive for future projects seeking to create more multilingual KGs.

This opens promising directions for the future development of multilingual KGs. With a dedicated community and specialised tools, better language coverage can be achieved in the future. In Section 4.1, we discuss automated solutions to improve the language coverage of KGs.

2.4 Lexicographical Data

Given that KGs can be used to store diverse kinds of knowledge, their versatile graph structure also facilitates storing information *about* language. Some KGs, such as Lexvo [14] and Glottolog [33], focus specifically on linguistic metadata about languages and dialects as well as scripts and characters.

Lexicographical data in KGs describes a subset of data that expresses information about the lexicon of languages, i.e., information about words, phrases, and other linguistic expressions. For example, this linguistic information could describe how to conjugate verbs across different languages. In this example, the verb may be modelled as an entity in the graph that is connected to its different grammatical forms through its properties (edges in the graph).

To express and standardise this type of linguistic information, *OntoLex Lemon*⁸ was introduced based on the RDF standard, focusing on expressing linguistic information as Linguistic Linked Data [64].

⁸ https://www.w3.org/2016/05/ontolex/

L.-A. Kaffee, R. Biswas, C. M. Keet, E. K. Vakaj, and G. de Melo

The concept of lexemes, a form of describing lexical knowledge, has also been introduced to the KG Wikidata. Wikidata uses a custom community-created schema that only loosely resembles Lemon. However, considering the actual data provided by Wikidata, statistics reveal that there are only a few languages with significant lexeme coverage, while the vast majority of languages have little to no representation [73]. This limited language coverage shows that there is still a long way to go to reuse such linguistic information for applications focusing on low-resource languages.

One of the most successful, widely used knowledge resources is WordNet [65], as it has over the years seen extensive use in a number of NLP applications, such as text summarisation [5] and text categorisation [20]. A large number of similar resources have been created for a multitude of different languages, including low-resource languages, and many of these resources have been interlinked. The Universal WordNet [15] and BabelNet [72] were among the first massively multilingual knowledge graphs, both drawing on WordNet as their backbone. WordNet is based on the notion of synonym sets as linguistic concepts, which are connected by various linguistically inspired semantic relationships.

Despite the widespread use of multilingual lexicographic KGs in NLP, it is important to acknowledge that the coverage is uneven. While high-resource languages are well-represented, many widely-spoken languages of socio-economic importance are covered inadequately. For some low-resource languages, only very basic terminology is covered. Many others are missing entirely. While information about grammar can be useful for languages with sparse training data, the lack of representation in KGs cannot yet fill this gap. However, as argued previously, having a central general-domain storage of this information for low-resource languages can support future applications in broadening language coverage. Therefore, we argue, it would be beneficial to build these resources and maintain and widen existing ones.

3 KG Challenges Regarding Low-resource Languages

For KGs to be part of the solution for low-resource languages (LRL) in data-driven settings, they need to be buildable and deployable in KG-driven information systems. They may also need a better specification of "low-resourced languages". With a full characterisation under way [55], within NLP, it is typically narrowed down to LRLs having just limited online corpora, tools, and computational grammars, or lacking something to build statistics-based NLP applications⁹. LRLs have also been characterised as "less studied, resource scarce, less computerised, less privileged, less commonly taught, or low density, among other denomination" [87], and similarly by others [80, 34]. Some studies have sought to quantify this by counting labelled and unlabelled corpora [42] or other data [34] and tools [6, 58], and conducted audits [67, 82, *inter alia*]. Such characterisations are based mostly on quantities, while neglecting to account for the practicalities of working with limited resources. Such practicalities may be grouped into two: one related to the tools and processes and the other at a "deeper" level on language features of both the representation language and the LRL. We'll discuss each in turn.

3.1 Computational Resources

Within the narrow computational resource-oriented view, a consequence is the existence of blocking interdependencies. As a concrete example related to Figure 1, within Abstract Wikipedia [89], it has been proposed to automatically induce templates for a template-based approach [32] or grammars for the Grammatical Framework-based approach [81], ultimately to facilitate rule-based natural language generation to generate Wikipedia articles from the KG. This, however,

⁹ See, for instance, Felix Laumann's discussion post at https://medium.com/neuralspace/low-resourcelanguage-what-does-it-mean-d067ec85dea5, likely based on Tsvetkov (2017) [87] (slide 26) and repeated in the literature (e.g., [59]).

10:8 Multilingual Knowledge Graphs

presupposes the existence and usability of corpora, of good quality part-of-speech taggers and of morphological analysers, which are rarely available for LRLs. These resources are likewise needed for automated KG construction and use in KG-driven information systems, such as educational question generation and document navigation. Consider KG verbalisation for, e.g., isiZulu, a LRL in South Africa spoken by around 23 million people [53]: nouns had to be pluralised, but there was no pluraliser, and verbs of the object properties needed to be conjugated, yet grammar books were outdated and recent linguistics research is scattered, so all that had to be investigated alongside the actual KG task [8, 52]. Also, popular multilingual realisers, such as SimpleNLG [28], are easily adapted among a selection of well-resourced languages in the Indo-European languages family, but they may not be suitable for a language that needs subword-level management, and so a new modular realisation engine may be needed [60]. Conversely, any prospective KG task helps focus language resource development on a specific, measurable, and achievable segment, which is confidence-building and a way to gradually expand the resources.

These anecdotes, and similar observations, are illustrative of several general issues when creating KGs or using them in applications with LRLs, be this for data-oriented techniques and applications or other KG tasks, being:

- There will be linguistic hurdles (gaps in the linguistic knowledge and sociolinguistics) to overcome, which are on top of the intended KG task or KG as a solution. In addition, computerising the language information takes time and scarce human resources, delaying the KG task.
- Freely available pre-existing data is often imprecise, incorrect, or outdated, and thus not a good basis to rely on [solely/at all], requiring an additional data collection stage in a data-driven KG task; therefore, an expert-driven rules-based approach may be more viable for some tasks.
- A resource shortage also tends to imply a human shortage, both in numbers and capacity/knowledge/skills of the language, limiting the scale of human evaluation and quality of survey or crowdsourced responses.
- The notion of "good" quality is relative and a lower overall size or quality may still mean that the KG task itself performs well (but lower in context due to compounding of less well performing preceding steps).
- Each LRL has its own set of hurdles, and its own history, and how the low resourcedness came about and therewith may need a context-specific incentive to realise the KG task for the LRL.

Also, and separate to KG building, there are power dynamics. Those who build the KG wield power over those who use it and, as Vang argues, "to some degree contests the autonomy of the user" [43]. A multilingual KG would ideally be built *with* the community of prospective users who speak that LRL, not just *for* them. In addition, it is not clear how the notion of KG co-ownership or KG benefit-sharing of extracted and systematised community knowledge, as alluded to in, among others, the groundbreaking San Code of Research Ethics¹⁰, can be realised. Finally, the LRL may have features that do not fit well with the KG language. Since this may be applicable to a subset of LRLs only, we elaborate on this in the next section.

3.2 KG Representation Language Assumptions and Challenges

In this paper, we introduce KGs as a well-suited technology to address limitations in language coverage in downstream applications. There are obvious limitations to this claim. Popular KGs tend to be Eurocentric, in terms of their editors (see for example Wikidata [48]) and data covered.¹¹

¹⁰https://www.globalcodeofconduct.org/affiliated-codes/

¹¹Visualisation of Wikidata entities with geolocation: https://wmde.github.io/wikidata-map/dist/index. html

L.-A. Kaffee, R. Biswas, C. M. Keet, E. K. Vakaj, and G. de Melo

Moreover, they have been developed to a large extent by researchers working in what is known as the *global north*, emphasising the bias in the type of content and how the content is captured and displayed.

The subject-predicate-object paradigm of capturing facts in a knowledge graph, OWL's assumptions about axiomatisations, and the likes of frame-based approaches suit English well, as it typically takes an object-based approach, has limited verb inflections and has a disjunctive orthography, and an SVO ordering of sentences that are statements.

The practices of modelling assume nouns in the singular and predicates in the 3rd person singular, whose string remains fixed irrespective of the domain or range of the property. This is not the case for several LRLs, however, such as at least Niger-Congo B languages [53], nor for several Indo-European languages that have an extensive noun class system where the verb changes contextually depending on the actor. For instance, the "eats" is all the same in (Human, eats, apple), (Elephant, eats, apple), and (Microbe, eats, apple), regardless who or what (in sg.) is doing the eating, where the "eats" property is being reused as it should be. Consider now a direct translation to isiZulu, respectively: (Umuntu, udla, i-apula), (Indlovu, idla, i-apula), and (Igciwane, lidla, i-apula). The 3rd person singular differs. It is not the case that the predicate is different, just the natural language rendering of it is. However, most KG languages by design typically conflate elements with surface rendering [25]. One could use an identifier, as OBO did, and carry over to OBO Foundry ontologies and in Wikidata, but that still requires additional machinery somewhere to complete the -dla verb stem in accordance with the noun class of the noun. Or: to start properly with KG development, one first needs to figure out some sort of extension of, or addition to, RDF. Depending on the LRL, there may well be up to 20 variants, with one for each noun class.

It has also become common practice in "English KGs" to insert prepositions into the property name or label, such as "works *for*" and "part *of*". They may be realised differently in many other languages, such as being affixed to the noun of the class in the range (object) position or infixed in the verb. The affixation to the noun is also specific to the context where it is used, i.e., in which axiom, not the name of the class or individual, for which there is no established KG language yet. An extreme case is the containment relation, typically used in KGs as contained in. The notion of containment in isiZulu is realised through indicating the container, by means of noun affixation for locatives and determined by both the container and the containee such that there is no "contained in" verb or name to put in the predicate position [51]; e.g., a bolus of food (*indilinga yokudla*) that is contained in the stomach (*isisu*) becomes *indilinga yokudla isesiswini*. What would the triple be? (bolus of food, contained in, stomach) maps neatly with the natural language in English and many other well-resourced languages, but neither a (indilinga yokudla, blank_prop, isisu) nor a (indilinga yokudla, inverse(L3951-S4), isisu) are satisfactory solutions.

Therefore, a KG may need to be accompanied by a declarative language model and a set of grammar rules, or a different way of usage to represent all the required inputs, or the KG representation language may need to be revised. This is regardless of the usage scenario, from rendering the content of the KG correctly and understandable to the user to automated KG creation. The W3C community standard OntoLex-Lemon [13] as well as the more expressive *lemon* model [61] that aim to provide such a declarative model address this only in part [29, 10].

These language differences lay bare certain English-oriented assumptions baked into the representation language and in naming conventions. This is complicated further if the conceptualisation or terminology diverges not only for domain knowledge but also for a foundational relation such as parthood (e.g., [23, 54]). This need not impede KGs as a possible contributing solution to LRL applications but is to be borne in mind in both KG construction and in their use.

These limitations need to be addressed to make KGs truly inclusive of a wide range of languages.

4 Applications of Multilingual KGs

Multilingual KGs have emerged as resources that transcend language barriers to enable a wide range of applications in our increasingly linked and multilingual world. Improving language coverage in multilingual KGs is critical for completing these resources and making them more relevant to a wide range of global audiences. Also, they have a wide range of applications across various domains due to their ability to bridge linguistic gaps. This section provides an overview of the existing research in the application areas of multilingual KGs.

4.1 Improving Language Coverage of KGs

Recent research has seen an enormous development in the use of KGs to improve Natural Language Processing tasks such as Natural Language Inference (NLI), Question Answering (QA), and Recommender Systems. Even well-known KGs like DBpedia and Wikidata, which are widely used, are the largest in their English versions despite major human efforts to make them available across languages. Furthermore, geographic area-specific facts are frequently restricted to the KG unique to the region or the native language. The incorporation of Machine Learning (ML) models into several languages is constrained by the scarcity of multilingual knowledge.

Machine Translation (MT) systems have been used [2, 63] to improve the language coverage in multilingual KGs, but these efforts only aim to translate domain-specific KGs from English into a target language. These methods ignore the graph structure of KGs, which is critical for determining the domain in which the word must be translated in the target language. Taking into account the graph structure of KGs can help an MT system identify the proper translation for ambiguous labels. As referred to in the survey [70], Rule-based (RBMT), Example-based (EBMT), and Statistical Machine Translation (SMT) based models have been used in the past to translate Semantic Web Technologies (SWT). However, MT for SWT still remains open research due to: (1) lack of clearly specified object attributes, such as cardinality or reflexiveness, (2) concept blending across thesaurus, vocabulary, and ontology, (3) inaccurate definitions of the domain and range, and (4) using ambiguous annotations [70]. However, domain-specific terms from the medical and financial areas have been translated using a Neural Machine Translation (NMT) architecture, outperforming SMT results [3]. Feng et al. [24] introduced a gated NN strategy for translating English KGs into Chinese, learning continuous triple representations. Source and target triples were mapped in the same semantic vector space using their method, which was extracted from Freebase. Their modified NN strategy increased translation accuracy compared to a strong NMT baseline, highlighting the significance of taking into account KG structure for KG translation and enhancing the quality of disambiguation for ambiguous phrases. Another NMT model, THOTH [69] trains to translate the facts from one language into another, treating the facts (i.e., triples) in a KG as sentences with URIs acting as tokens. It uses KG embeddings and two separate recurrent neural network models to extract bilingual alignments between a source and target KG and then learns the translation.

KG embedding based models have been proposed to enhance multilingual KGs. A more realistic method would draw on the information in several language-specific KGs, keeping in mind that individual KGs have their own strengths and limitations on data quality and coverage. This is a significant challenge since inconsistently expressed facts and a lack of sufficient alignment information can make it difficult for knowledge to be transferred between many independently maintained KGs. KEnS [11] is one such approach that embeds all multilingual KGs in a shared embedding space where the association of entities is captured through self-learning.

LLM-based KG creation has been used recently in an attempt to create monolingual knowledge facts for KGs [77], but multilingual KG construction and enrichment have not yet been attempted at their full strength. Due to the potential for complementary and unequally dispersed

10:11

knowledge stated in many languages, multilingual LLMs can offer richer combined knowledge for multilingual KGs. Prix-LM [91] is a pioneering model for multilingual KG construction and completion that employs both monolingual triples and cross-lingual linkages retrieved from existing multilingual KGs, followed by fine-tuning a multilingual language encoder XLM-R via a causal language modelling objective. Hou et al. [38] proposes another model that leverages adapters to fine-tune LLMs for multilingual KGs, including low-resource languages, enhancing the corresponding downstream tasks. Therefore, leveraging LLMs for the construction and completion of multilingual KGs emerges to be one of the promising future directions of research, considering that their information content has shown promise for knowledge extraction [77].

4.2 Downstream Applications of Multilingual KG Information

In this section, we survey a set of use cases for multilingual KGs. When developing new KGs, new ontologies, or improving the language coverage of existing KGs, it is crucial to keep the use-cases in mind to ensure that the KG can indeed serve knowledge that is useful in the envisioned tasks. Several applications already make use of multilingual KGs, in the following we present a selected few tasks.

Multilingual Knowledge Graph Question Answering (mKGQA) is a task that involves answering a user's questions in a set of languages based on facts stored in a KG. This is a research topic of particular importance, as these systems can bridge the gap between users and pertinent information on the Web. The task should ideally address all languages, making it possible for users to ask questions independent of the languages of the information in the KG. Currently, however, these systems typically define a set of languages in which questions can be answered, relying on information in the corresponding language stored in the target KG. mKGQA is strongly dependent on data, and there is currently a lack of multilingual data on the web, making it accessible to only a fraction of people¹², therefore causing a "cultural gap" on the Web [66]. Further, the task of mKGQA suffers from a lack of multilingual benchmarks.

Perevalov et al. [75, 76] identified 17 mKGQA systems. We have provided in the table below a list of 4 mKGQA systems filtered based on the coverage that they provide for languages besides English (with a focus on lower-resourced languages), along with the KG used, data sets, and the specific languages covered.

mKGQA system	Ref.	KG	Languages	Dataset
WDAqua-core, 2018	[19]	DBpedia, Wikidata, DBLP, MusicBrainz	en, de, es, it	QALD, LC-QuAD 1.0
QAnswer, 2019	[18]	DBpedia, Wikidata, DBLP, MusicBrainz, FreeBase	en, de, fr, it, es, pt, ar, zh	QALD-3-7, LC-QuAD 1.0
Y. Zhou et al.	[92]	DBpedia	en, fa, de, ro, it, ru, fr, nl, es, hi, pt	LC-QuAD 1.0, QALD-9
A. Perevalov et al.	[74]	Wikidata, DBpedia	en, de, fr, ru, uk, lt, be, ba, hy	QALD-9-Plus
BLEU-4	[84]	-	tibetian	TibetanQA+

Table 1 Multilingual Knowledge Graph Question Answering (mKGQA) systems.

¹² 25.9% according to https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-theinternet/, retrieved 2023-02-15

Knowledge Graph Completion aims to add new, missing facts to a KG. This remains a challenging problem that is particularly pronounced in multilingual low-resource languages, given that human annotations are rare and difficult to procure [39]. A number of solutions have been devised to mitigate this problem, such as using KG embeddings [40] or jointly predicting entity alignment across languages and new facts for KG completion [9, 86, 11]. A promising approach is the self-supervised adaptive graph alignment (SS-AGA) method [39], which regards alignment as a new edge type between parallel entities instead of a loss constraint, and fuses KGs from different languages in a single graph. This approach has been evaluated on both the public multilingual DBpedia KG and a newly created industrial multilingual E-commerce KG.

Cross Lingual Fact Extraction (CLFE) is the task of extracting facts from a text to, e.g., store the facts in a structured data format. Extracting facts from source text of different languages has not received as much attention as monolingual fact extraction [83]. KGs can support multilingual fact linking and extraction, which is important in many downstream tasks such as QA. The REFCOGLink model [57] is based on linking facts expressed in a sentence to the corresponding fact labels (i.e., language-specific representation of the fact) in the KG and outperforms standard retrieval + re-ranking. The CLFE (Cross-Lingual Fact Extraction) [1] model demonstrates strong performance in multilingual and cross-lingual fact extraction tasks, specifically in English and seven other LR Indic languages. It achieves an F1 score of 77.46% using two different approaches. The first approach is a classification-based method, where the model first extracts the object or "tail" of the fact and then predicts the relationship between the extracted tail and the subject. The second approach is a generative one, which combines both tasks into a single step. CLFE makes use of the XAlign dataset, which contains 0.45M pairs across 8 languages, of which 5,402 pairs have been manually annotated.

Multilingual Relation Classification is the task of extracting relations (i.e., triples in the context of KGs) from natural language text in various languages. Multilingual relation classification has been explored through the method of prompting, which can receive promising results even for lower-resourced languages [12]. One of the challenges for the task of multilingual relation classification is the lack of multilingual datasets and benchmarks. The dataset RELX [56] aims to close this gap by providing a baseline model and benchmark for English, French, German, Spanish, and Turkish. IndoRE [71] is a comprehensive dataset comprising 21,000 gold-tagged sentences for named entity recognition (NER) and relation extraction (RE) in three Indian languages (English, Bengali, Hindi) as well as English. The dataset provides valuable resources for advancing research in relation classification of Indian languages, which is crucial for KG augmentation and Question Answering systems. The authors employ multilingual BERT and transfer learning techniques, and propose TransRel, a multilingual system for joint named entity recognition (NER) and relation extraction (RE) with interlingual transfer.

Neural Machine Translation translates text from a source to a target language using deep learning architectures. The performance of Neural Machine Translation (NMT) systems can be enhanced through the integration of KGs, particularly when translating domain-specific expressions and named entities. For example, Moussallem et al. [68] introduce an approach incorporating the KG DBpedia into NMT models, resulting in considerable improvements in performance of these NMT models.

Automatic KG creation for low-resource languages has been leveraged as another possible route to tackle the lack of low-resource language KGs. HKC has been recently introduced as a framework that constructs a knowledge graph for the Hindi language [88] using various NLP techniques. FarsBase, a Persian multi-source knowledge graph [4] is another example of such an application. To construct such a KG, the authors apply a number of techniques to integrate data from Wikipedia and both structured and unstructured data from the web. Overall, KGs offer immense potential in supporting multilingualism, which in the future will likely also increasingly benefit low-resource languages. By constructing multilingual KGs, aligning entities across languages, and employing techniques such as machine translation and cross-lingual extraction, KGs can effectively bridge the knowledge gap and facilitate access to information in low-resource languages. These advancements not only empower speakers of all languages but also contribute to a more inclusive and diverse knowledge ecosystem.

5 Open Problems

We have introduced existing approaches to close the language gap in the domain of KGs. We described multilingual KGs, their challenges with regard to low-resource languages, and approaches to improve and use them. However, there remain a number of open challenges, which we will describe here to build a foundation for future work and point out some of the pressing issues in the domain of multilingual KGs.

5.1 Regarding the State of Multilingual KGs

The lack of multilingual information in KGs should be addressed in the future, and while there are a number of approaches seeking to mitigate this problem, none have thus far succeeded in addressing this important challenge. One of the notable issues with current KGs is that they are typically not aligned across languages, given the usage of different language ontologies. This misalignment poses a large challenge when merging knowledge across languages. Addressing the cross-lingual alignment of ontologies is one of the crucial challenges that will build the foundation for the wide availability of cross-lingual information.

Merging KGs across languages, even with aligned ontologies, remains a non-trivial task. Not only is there a need to identify alignments at the entity and class level. Future studies are needed to understand which languages are covered to which extent in the different KGs and how these different KGs with potentially different topic-focus could be merged harmonically without centering one language in the approach.

Another open problem, especially for community-edited KGs, is the question of how to interest a larger, more diverse community in the contribution to the knowledge stored in the KG. If a diverse community is contributing, naturally more languages will be covered. Creating incentives and showing how the information could be used, such as building tools based on KG information, could be a way to address this challenge. However, future work will have to better understand existing incentives and broaden them.

5.2 Regarding Low-resource Languages

While there have been approaches suggested to address the lack of languages covered in KGs (see Section 4), there is currently a glaring gap when it comes to low-resource languages. Having very little available information in any language is a challenge for all applications, especially neural- or deep-learning-driven ones, such as large language models. This then also raises the question of optimal strategies for KG construction, such as human-in-the-loop procedures due to limited and overly noisy data and appropriate incentive strategies for manual modelling. KGs could contribute to closing this gap to some extent by providing central storage of multilingual, linked, reusable data. One of the advantages of using KGs may be that while a fact is not yet translated to a lower-resourced language, it already exists in another language, and by linking the data across languages, these facts can be reused even before translation. However, the challenges described in Section 3.2 in the form of assumptions of language modelling need to be urgently addressed

10:14 Multilingual Knowledge Graphs

for under-represented languages to be able to catch up in the content representation. From an ontology engineering perspective, it is important to focus on representing the diversity of language structures accurately, lest the technique limits its use to applications in a few languages only rather than the breadth of opportunities across the world.

5.3 Regarding Applications of Multilingual KGs

For the reuse of multilingual KGs in downstream tasks, there are still many opportunities to better integrate multilingual KG information with approaches from the field of NLP and with multi-modal approaches. Reusing the knowledge graph information is not only beneficial to the applications that use them but can also create an incentive to create better, more diverse KGs. To ensure this, the current information needs to be accurately curated, and multilingual information should be put into focus for future approaches in increasing KG coverage. Understanding coverage in different low-resource languages as well as reusing this information is a promising path forward for future work that can have a real-world impact.

6 Conclusion

In this paper, we reviewed different aspects of multilingual KGs; we established the state of multilingual KGs by summarising the guidelines on creating multilingual KGs; established the challenges with regards to low-resource languages; described applications for and using multilingual KGs; and finally pointed out open problems derived from our survey of multilingual KGs. The literature provides clear guidelines and ontologies for multilingual information and lexicographic data. Yet, there is a severe lack of multilingual information in existing KGs and a bias toward English-language information. Particularly for low-resource languages, KGs could be useful for closing information gaps. However, we concluded that there are currently major challenges that need to be addressed regarding low-resource language integration into KGs, such as the English-centric structure and content of existing KGs.

Current approaches to improve language coverage of KGs, such as machine translation of KG labels or leveraging KG embeddings to align monolingual KGs across languages, are a promising direction to make the KG more diverse. However, these approaches are currently limited to a small set of languages and need to be explored for low-resource languages with a focus on the challenges described. Future work has to focus on the inclusion of non-European languages from the very structure of KGs, including them in the considerations of how knowledge is modelled in KGs. Expanding language inclusivity holds immense importance as it paves the way for a more accessible and all-encompassing digital landscape, enabling the internet to cater to a diverse array of communities. While currently language coverage, especially for low-resource languages, is limited, there are viable avenues for progress. If these pathways are pursued, KGs could serve as a technology to realise the vision of a more equitable and inclusive internet, facilitating the exchange of knowledge across language communities.

— References -

Tushar Abhishek, Shivprasad Sagare, Bhavyajeet Singh, Anubhav Sharma, Manish Gupta, and Vasudeva Varma. Xalign: Cross-lingual fact-totext alignment and generation for low-resource languages. In *Companion Proceedings of the Web Conference 2022*, pages 171–175, 2022. doi: 10.1145/3487553.3524265.

² Mihael Arcan and Paul Buitelaar. Ontology Label Translation. Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA, pages 40-46, 2013. URL: https://aclanthology.org/N13-2006/.

- 3 Mihael Arcan and Paul Buitelaar. Translating Domain-Specific Expressions in Knowledge Bases with Neural Machine Translation. CoRR, abs/1709.02184, 2017. doi:10.48550/arXiv.1709.02184.
- 4 Majid Asgari-Bidhendi, Ali Hadian, and Behrouz Minaei-Bidgoli. Farsbase: The persian knowledge graph. Semantic Web Journal, 10(6):1169–1196, 2019. doi:10.3233/SW-190369.
- 5 Kedar Bellare, Anish Das Sarma, Atish Das Sarma, Navneet Loiwal, Vaibhav Mehta, Ganesh Ramakrishnan, and Pushpak Bhattacharyya. Generic Text Summarization Using WordNet. In Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal. European Language Resources Association, 2004. URL: http://www.lrec-conf.org/proceedings/ lrec2004/summaries/342.htm.
- 6 V. Berment. Méthodes pour informatiser des langues et des groupes de langues peu dotées. Phd thesis, J. Fourier University – Grenoble I, may 2004. URL: https://theses.hal.science/ tel-00006313/document.
- 7 Tim Berners-Lee. Cool URIs don't change, 1998. Accessed on 05.07.2023. URL: https://www.w3. org/Provider/Style/URI.html.
- 8 Joan Byamugisha, C. Maria Keet, and Langa Khumalo. Pluralising nouns in isiZulu and related languages. In A. Gelbukh, editor, *Pro*ceedings of CICLing'16, volume 9623 of LNCS, pages 271–283. Springer, 2018. doi:10.1007/ 978-3-319-75477-2_18.
- 9 Soumen Chakrabarti, Harkanwar Singh, Shubham Lohiya, Prachi Jain, and Mausam. Joint completion and alignment of multilingual knowledge graphs. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 11922–11938. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.EMNLP-MAIN.817.
- 10 Catherine Chavula and C. Maria Keet. Is lemon sufficient for building multilingual ontologies for Bantu languages? In C. Maria Keet and Valentina Tamma, editors, Proceedings of the 11th OWL: Experiences and Directions Workshop (OWLED'14), volume 1265 of CEUR-WS, pages 61-72, 2014. Riva del Garda, Italy, Oct 17-18, 2014. URL: https://ceur-ws.org/Vol-1265/ owled2014_submission_10.pdf.
- 11 Xuelu Chen, Muhao Chen, Changjun Fan, Ankith Uppunda, Yizhou Sun, and Carlo Zaniolo. Multilingual knowledge graph completion via ensemble knowledge transfer. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 3227–3238. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.FINDINGS-EMNLP.290.
- 12 Yuxuan Chen, David Harbecke, and Leonhard Hennig. Multilingual relation classification via efficient and effective prompting. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, edit-

ors, Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 1059–1075. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.EMNLP-MAIN.69.

- 13 Philipp Cimiano, John P. McCrae, and Paul Buitelaar. Lexicon model for ontologies: Community report. Final community group report, 10 may 2016, W3C, 2016. URL: https://www.w3.org/2016/05/ ontolex/.
- 14 Gerard de Melo. Lexvo.org: Language-related information for the Linguistic Linked Data cloud. Semantic Web Journal, 6(4):393-400, aug 2015. doi:10.3233/SW-150171.
- 15 Gerard de Melo and Gerhard Weikum. Towards a universal wordnet by learning from combined evidence. In David Wai-Lok Cheung, Il-Yeol Song, Wesley W. Chu, Xiaohua Hu, and Jimmy J. Lin, editors, Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009, pages 513–522, New York, NY, USA, 2009. ACM. doi:10.1145/1645953.1646020.
- 16 Jeremy Debattista, Sören Auer, and Christoph Lange. Luzzu - A methodology and framework for linked data quality assessment. ACM J. Data Inf. Qual., 8(1):4:1-4:32, 2016. doi:10.1145/2992786.
- 17 G. R. Dent and C. L. S. Nyembezi. Scholar's Zulu Dictionary. Shuter & Shooter Publishers, 4 edition, 2009.
- 18 Dennis Diefenbach, Pedro Henrique Migliatti, Omar Qawasmeh, Vincent Lully, Kamal Singh, and Pierre Maret. QAnswer: A Question Answering prototype bridging the gap between a considerable part of the LOD cloud and end-users. In *The World Wide Web Conference*, pages 3507–3510, 2019. doi:10.1145/3308558.3314124.
- 19 Dennis Diefenbach, Kamal Singh, and Pierre Maret. WDAqua-core1: A Question Answering service for RDF Knowledge Bases. In Companion Proceedings of the The Web Conference 2018, pages 1087–1091, 2018. doi:10.1145/3184558.3191541.
- 20 Zakaria Elberrichi, Abdellatif Rahmoun, and Mohamed Amine Bentaallah. Using wordnet for text categorization. Int. Arab J. Inf. Technol., 5(1):16– 24, 2008.
- 21 Basil Ell, Denny Vrandecic, and Elena Simperl. Labels in the web of data. In Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Fridman Noy, and Eva Blomqvist, editors, The Semantic Web ISWC 2011 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I, volume 7031 of Lecture Notes in Computer Science, pages 162–176. Springer, 2011. doi:10.1007/978-3-642-25073-6_11.
- 22 Mauricio Espinoza, Asunción Gómez-Pérez, and Eduardo Mena. Enriching an ontology with multilingual information. In Sean Bechhofer, Manfred Hauswirth, Jörg Hoffmann, and Manolis Koubarakis, editors, The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008, Proceedings, volume 5021 of Lecture Notes in Computer

Science, pages 333–347. Springer, 2008. doi: 10.1007/978-3-540-68234-9_26.

- 23 Chen-Chieh Feng and David M. Mark. Crosslinguistic research on landscape categories using geonet names server data: A case study for indonesia and malaysia. The Professional Geographer, 69(4):567–578, 2017. doi:10.1080/ 00330124.2017.1288575.
- 24 Xiaocheng Feng, Duyu Tang, Bing Qin, and Ting Liu. English-Chinese Knowledge Base Translation with Neural Network. COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan, pages 2935-2944, 2016. URL: https://aclanthology. org/C16-1276/.
- 25 P. R. Fillottrani and C. M. Keet. An analysis of commitments in ontology language design. In B. Brodaric and F. Neuhaus, editors, 11th International Conference on Formal Ontology in Information Systems 2020 (FOIS'20), volume 330 of FAIA, pages 46–60. IOS Press, 2020. doi: 10.3233/FAIA200659.
- 26 Pablo R. Fillottrani and C. Maria Keet. Patterns for Heterogeneous TBox Mappings to Bridge Different Modelling Decisions. In E. Blomqvist et al., editors, *Proceeding of the 14th Extended Semantic Web Conference (ESWC'17)*, volume 10249 of *LNCS*, pages 371–386. Springer, 2017. 30 May 1 June 2017, Portoroz, Slovenia. doi: 10.1007/978-3-319-58068-5_23.
- 27 ∀, Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddee Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages. Findings of EMNLP, 2020. doi:10.18653/v1/2020.findings-emnlp.195.
- 28 A. Gatt and E. Reiter. Simplenlg: A realisation engine for practical applications. In E. Krahmer and M. Theune, editors, *Proceedings of the* 12th European Workshop on Natural Language Generation (ENLG'09), pages 90–93. ACL, 2009. March 30-31, 2009, Athens, Greece. URL: https: //aclanthology.org/W09-0613.pdf.
- 29 Frances Gillis-Webber and C. Maria Keet. A review of multilingualism in and for ontologies. CoRR, abs/2210.02807, 2022. doi:10.48550/ARXIV.2210. 02807.
- 30 Frances Gillis-Webber and C. Maria Keet. A survey of multilingual OWL ontologies in bioportal. In Katy Wolstencroft, Andrea Splendiani, M. Scott Marshall, Chris Baker, Andra Waagmeester, Marco Roos, Rutger A. Vos, Rianne Fijten, and Leyla Jael Castro, editors, 13th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, SWAT4HCLS 2022, Virtual Event, Leiden, The Netherlands, January 10th to 14th, 2022, volume 3127 of CEUR Workshop Proceedings, pages 87–96. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3127/paper-11.pdf.
- 31 Jorge Gracia, Elena Montiel-Ponsoda, Philipp Cimiano, Asunción Gómez-Pérez, Paul Buitelaar,

and John P. McCrae. Challenges for the multilingual web of data. J. Web Semant., 11:63-71, 2012. doi:10.1016/J.WEBSEM.2011.09.001.

- 32 Ariel Gutman and C. Maria Keet. Template language for wikifunctions, 2022. URL: https:// meta.wikimedia.org/wiki/Abstract_Wikipedia/ Template_Language_for_Wikifunctions.
- 33 Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. *Glottolog. Ver*sion 4.8. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2023. doi:10.5281/zenodo. 8131084.
- 34 Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. A survey on recent approaches for natural language processing in low-resource scenarios. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2545–2568. Association for Computational Linguistics, jun 2021. doi:10.18653/V1/2021.NAACL-MAIN.201.
- 35 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. Number 22 in Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool, 2021. doi:10.2200/ S01125ED1V01Y202109DSK022.
- 36 Ian Horrocks. Ontologies and the semantic web. Communications of the ACM, 51(12):58-67, 2008. doi:10.1007/3-540-45810-7.
- 37 Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From SHIQ and RDF to OWL: the making of a web ontology language. J. Web Semant., 1(1):7-26, 2003. doi:10.1016/J.WEBSEM. 2003.07.001.
- 38 Yifan Hou, Wenxiang Jiao, Meizhen Liu, Carl Allen, Zhaopeng Tu, and Mrinmaya Sachan. Adapters for Enhanced Modeling of Multilingual Knowledge and Text. Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 3902–3917, 2022. doi: 10.18653/V1/2022.FINDINGS-EMNLP.287.
- 39 Zijie Huang, Zheng Li, Haoming Jiang, Tianyu Cao, Hanqing Lu, Bing Yin, Karthik Subbian, Yizhou Sun, and Wei Wang. Multi-lingual knowledge graph completion with self-supervised adaptive graph alignment. arXiv pre-print arXiv:2203.14987, 2022. doi:10.48550/arXiv.2203.14987.
- 40 Zijie Huang, Zheng Li, Haoming Jiang, Tianyu Cao, Hanqing Lu, Bing Yin, Karthik Subbian, Yizhou Sun, and Wei Wang. Multilingual knowledge graph completion with self-supervised adaptive graph alignment. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 474–485. Association for Com-

putational Linguistics, 2022. doi:10.18653/V1/ 2022.ACL-LONG.36.

- 41 Shimaa Ibrahim, Said Fathalla, Jens Lehmann, and Hajira Jabeen. Toward the multilingual semantic web: Multilingual ontology matching and assessment. *IEEE Access*, 11:8581–8599, 2023. doi:10.1109/ACCESS.2023.3238871.
- 42 Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. The state and fate of linguistic diversity and inclusion in the NLP world. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6282-6293, Online, jul 2020. Association for Computational Linguistics. doi:10.18653/V1/2020.ACL-MAIN.560.
- 43 K. Juel Vang. Ethics of Google's Knowledge Graph: some considerations. Journal of Information, Communication and Ethics in Society, 11(4):245– 260, 2013. doi:10.1108/JICES-08-2013-0028.
- 44 Lucie-Aimée Kaffee. Multilinguality in Knowledge Graphs. PhD thesis, University of Southampton, 2021. URL: https://eprints.soton.ac.uk/ 456783/.
- 45 Lucie-Aimée Kaffee, Kemele M. Endris, and Elena Simperl. When humans and machines collaborate: cross-lingual label editing in wikidata. In Björn Lundell, Jonas Gamalielsson, Lorraine Morgan, and Gregorio Robles, editors, Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019, Skövde, Sweden, August 20-22, 2019, pages 16:1–16:9. ACM, 2019. doi:10.1145/3306446.3340826.
- 46 Lucie-Aimée Kaffee, Kemele M. Endris, Elena Simperl, and Maria-Esther Vidal. Ranking knowledge graphs by capturing knowledge about languages and labels. In Mayank Kejriwal, Pedro A. Szekely, and Raphaël Troncy, editors, Proceedings of the 10th International Conference on Knowledge Capture, K-CAP 2019, Marina Del Rey, CA, USA, November 19-21, 2019, pages 21–28. ACM, 2019. doi:10.1145/3360901.3364443.
- 47 Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. A glimpse into babel: An analysis of multilinguality in wikidata. In Lorraine Morgan, editor, Proceedings of the 13th International Symposium on Open Collaboration, OpenSym 2017, Galway, Ireland, August 23-25, 2017, pages 14:1–14:5. ACM, 2017. doi:10.1145/3125433.3125465.
- 48 Lucie-Aimée Kaffee and Elena Simperl. Analysis of editors' languages in wikidata. In Proceedings of the 14th International Symposium on Open Collaboration, OpenSym 2018, Paris, France, August 22-24, 2018, pages 21:1-21:5. ACM, 2018. doi:10.1145/3233391.3233965.
- 49 Lucie-Aimée Kaffee and Elena Simperl. The human face of the web of data: A cross-sectional study of labels. In Anna Fensel, Victor de Boer, Tassilo Pellegrini, Elmar Kiesling, Bernhard Haslhofer, Laura Hollink, and Alexander Schindler, editors, Proceedings of the 14th International Conference on Semantic Systems, SEMANTiCS 2018, Vienna, Austria, September 10-13, 2018, volume 137 of Procedia Computer Science, pages 66–77. Elsevier, 2018. doi:10.1016/J.PR0CS.2018.09.007.

- 50 Lucie-Aimée Kaffee, Pavlos Vougiouklis, and Elena Simperl. Using natural language generation to bootstrap missing wikipedia articles: A humancentric perspective. Semantic Web, 13(2):163–194, 2022. doi:10.3233/SW-210431.
- 51 C. M. Keet. Representing and aligning similar relations: parts and wholes in isizulu vs english. In J. Gracia, F. Bond, J. McCrae, P. Buitelaar, C. Chiarcos, and S. Hellmann, editors, *Language*, *Data, and Knowledge 2017 (LDK'17)*, volume 10318 of *LNAI*, pages 58–73. Springer, 2017. 19-20 June, 2017, Galway, Ireland. doi:10.1007/ 978-3-319-59888-8_5.
- 52 C. M. Keet and L. Khumalo. Grammar rules for the isiZulu complex verb. Southern African Journal of Language and Linguistics, 35(2):183–200, 2017. doi:10.2989/16073614.2017.1358097.
- 53 C. Maria Keet and Langa Khumalo. Toward a knowledge-to-text controlled natural language of isiZulu. Language Resources and Evaluation, 51(1):131–157, 2017. doi:10.1007/ S10579-016-9340-0.
- 54 C. Maria Keet and Langa Khumalo. Parthood and part-whole relations in zulu language and culture. *Applied Ontology*, 15(3):361–384, 2020. doi:10.3233/A0-200230.
- 55 C. Maria Keet and Langa Khumalo. Contextualising levels of language resourcedness affecting digital processing of text. *CoRR*, abs/2309.17035, 2023. doi:10.48550/ARXIV.2309.17035.
- 56 Abdullatif Köksal and Arzucan Özgür. The RELX dataset and matching the multilingual blanks for cross-lingual relation classification. In Trevor Cohn, Yulan He, and Yang Liu, editors, Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020, volume EMNLP 2020 of Findings of ACL, pages 340–350. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020. FINDINGS-EMNLP.32.
- 57 Keshav Kolluru, Martin Rezk, Pat Verga, William W. Cohen, and Partha Talukdar. Multilingual Fact Linking, 2021. doi:10.48550/arXiv.2109.14364.
- 58 Steven Krauwer. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. In Proceedings of the 2003 International Workshop Speech and Computer SPECOM'03, volume 2003, pages 8–15, 2003. Moscow, Russia, 2003. URL: http://www.elsnet. org/dox/krauwer-specom2003.pdf.
- 59 A. Magueresse, V. Carles, and E. Heetderks. Lowresource languages: A review of past work and future challenges. *CoRR*, abs/2006.07264, 2020. doi:10.48550/arXiv.2006.07264.
- 60 Zola Mahlaza. Foundations for reusable and maintainable surface realisers for isiXhosa and isiZulu. Phd thesis, Department of Computer Science, University of Cape Town, South Africa, nov 2022. URL: https://adeebnqo.github.io/files/ Thesis.pdf.
- **61** John McCrae, Guadalupe Aguado-de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr,

and Tobias Wunner. Interchanging lexical resources on the semantic web. Language Resources and Evaluation, 46(4):701–719, 2012. doi: 10.1007/S10579-012-9182-3.

- 62 John McCrae, Guadalupe Aguado de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, and Tobias Wunner. The lemon cookbook. Technical report, Monnet Project, jun 2012. URL: https://www.lemon-model.net/ learn/cookbook.php.
- 63 John P. McCrae, Mihael Arcan, Kartik Asooja, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. Domain adaptation for ontology localization. Journal of Web Semantics, 36:23–31, 2016. doi:10.1016/J.WEBSEM.2015.12.001.
- 64 John Philip McCrae, Christian Chiarcos, Francis Bond, Philipp Cimiano, Thierry Declerck, Gerard de Melo, Jorge Gracia, Sebastian Hellmann, Bettina Klimek, Steven Moran, Petya Osenova, Antonio Pareja-Lora, and Jonathan Pool. The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud. In Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016), pages 2435-2441, Paris, France, 2016. URL: http://www.lrec-conf.org/ proceedings/lrec2016/pdf/851_Paper.pdf.
- 65 George A. Miller. Wordnet: A lexical database for english. Commun. ACM, 38(11):39–41, 1995. doi:10.1145/219717.219748.
- 66 Marc Miquel-Ribé and David Laniado. Wikipedia culture gap: quantifying content imbalances across 40 language editions. Frontiers in Physics, 6:54, 2018. doi:10.3389/fphy.2018.00054.
- 67 Carmen Moors, Ilana Wilken, Karen Calteaux, and Tebogo Gumede. Human language technology audit 2018: Analysing the development trends in resource availability in all south african languages. In Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists, SAICSIT '18, pages 296-304, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3278681.3278716.
- 68 Diego Moussallem, Axel-Cyrille Ngonga Ngomo, Paul Buitelaar, and Mihael Arcan. Utilizing knowledge graphs for neural machine translation augmentation. In *Proceedings of the 10th international* conference on knowledge capture, pages 139–146, 2019. doi:10.1145/3360901.3364423.
- 69 Diego Moussallem, Tommaso Soru, and Axel-Cyrille Ngonga Ngomo. THOTH: neural translation and enrichment of knowledge graphs. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtech Svátek, Isabel F. Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part I, volume 11778 of Lecture Notes in Computer Science, pages 505–522. Springer, 2019. doi:10.1007/978-3-030-30793-6_29.
- 70 Diego Moussallem, Matthias Wauer, and Axel-Cyrille Ngonga Ngomo. Machine Translation using Semantic Web Technologies: A Survey. Journal

of Web Semantics, 51:1-19, 2018. doi:10.1016/J. WEBSEM.2018.07.001.

- 71 Arijit Nag, Bidisha Samanta, Animesh Mukherjee, Niloy Ganguly, and Soumen Chakrabarti. A Data Bootstrapping Recipe for Low Resource Multilingual Relation Classification. arXiv preprint, 2021. doi:10.48550/arXiv.2110.09570.
- 72 Roberto Navigli and Simone Paolo Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217– 250, 2012. doi:10.1016/J.ARTINT.2012.07.001.
- 73 Finn Årup Nielsen. Lexemes in wikidata: 2020 status. In Maxim Ionov, John P. McCrae, Christian Chiarcos, Thierry Declerck, Julia Bosque-Gil, and Jorge Gracia, editors, Proceedings of the 7th Workshop on Linked Data in Linguistics, LDL@LREC 2020, Marseille, France, May 2020, pages 82-86. European Language Resources Association, 2020. URL: https://aclanthology.org/2020.ldl-1.12/.
- 74 Aleksandr Perevalov, Andreas Both, Dennis Diefenbach, and Axel-Cyrille Ngonga Ngomo. Can machine translation be a reasonable alternative for multilingual question answering systems over knowledge graphs? In *Proceedings of the ACM Web Conference 2022*, pages 977–986, 2022. doi: 10.1145/3485447.3511940.
- 75 Aleksandr Perevalov, Andreas Both, and Axel-Cyrille Ngonga Ngomo. Multilingual question answering systems for knowledge graphs – A survey. Semantic Web, 2023. URL: https://www.semantic-web-journal.net/ system/files/swj3417.pdf.
- 76 Aleksandr Perevalov, Axel-Cyrille Ngonga Ngomo, and Andreas Both. Enhancing the accessibility of knowledge graph question answering systems through multilingualization. In 2022 IEEE 16th International Conference on Semantic Computing (ICSC), pages 251–256. IEEE, 2022. doi: 10.1109/ICSC52841.2022.00048.
- 77 Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, pages 2463– 2473. Association for Computational Linguistics, 2019. doi:10.18653/V1/D19-1250.
- 78 A. Phillips and M. Davis. Tags for Identifying Languages, sep 2009. URL: https://www.rfc-editor. org/info/bcp47.
- 79 Alessandro Piscopo, Christopher Phethean, and Elena Simperl. Wikidatians are born: Paths to full participation in a collaborative structured knowledge base. In Tung Bui, editor, 50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017, pages 1–10. ScholarSpace / AIS Electronic Library (AISeL), 2017. doi:10. 24251/HICSS.2017.527.

- 80 S. Ranathunga, E-S. A. Lee, M.P. Skenduli, R. Shekar, M. Alam, and R. Kaur. Neural machine translation for low-resource languages: A survey. *CoRR*, abs/2106.15115, 2021. doi: 10.48550/arXiv.2106.15115.
- 81 Aarne Ranta. Multilingual Text Generation for Abstract Wikipedia in Grammatical Framework: Prospects and Challenges, pages 125–149. Springer International Publishing, Cham, 2023. doi:10. 1007/978-3-031-21780-7_6.
- 82 Georg Rehm and Andy Way, editors. European Language Equality: A Strategic Agenda for Digital Language Equality. Cognitive Technologies. Springer, 2023. doi:10.1007/978-3-031-28819-7.
- 83 Bhavyajeet Singh, Pavan Kandru, Anubhav Sharma, and Vasudeva Varma. Massively Multilingual Language Models for Cross Lingual Fact Extraction from Low Resource Indian Languages. arXiv preprint, 2023. doi:10.48550/arXiv.2302. 04790.
- 84 Yuan Sun, Yan Zhuang, Sisi Liu, and Xiaobing Zhao. Low-resource language question generation based on key sentence and knowledge graph. Available at SSRN 4560896, 2023. doi: 10.2139/ssrn.4560896.
- 85 Thomas Pellissier Tanon and Lucie-Aimée Kaffee. Property label stability in wikidata: Evolution and convergence of schemas in collaborative knowledge bases. In Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis, editors, Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, April 23-27, 2018, pages 1801–1803. ACM, 2018. doi:10.1145/3184558.3191643.
- 86 Vinh Tong, Dat Quoc Nguyen, Trung Thanh Huynh, Tam Thanh Nguyen, Quoc Viet Hung Nguyen, and Mathias Niepert. Joint multilingual knowledge graph completion and alignment. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for*

Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 4646-4658. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022. FINDINGS-EMNLP.341.

- 87 Y. Tsvetkov. Opportunities and challenges in working with low-resource languages, 2017. URL: https://www.cs.cmu.edu/ ytsvetko/jsaltpart1.pdf.
- 88 Preeti Vats, Nonita Sharma, and Deepak Kumar Sharma. Hkg: A novel approach for low resource indic languages to automatic knowledge graph construction. ACM Transactions on Asian and Low-Resource Language Information Processing, 2023. doi:10.1145/3611306.
- 89 Denny Vrandecic. Building a multilingual wikipedia. Communications of the ACM, 64(4):38-41, 2021. doi:10.1145/3425778.
- 90 Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. Quality assessment for linked data: A survey. Semantic Web, 7(1):63–93, 2016. doi: 10.3233/SW-150175.
- 91 Wenxuan Zhou, Fangyu Liu, Ivan Vulic, Nigel Collier, and Muhao Chen. Prix-LM: Pretraining for Multilingual Knowledge Base Construction. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 5412–5424, 2022. doi:10.18653/V1/2022.ACL-LONG.371.
- 92 Yucheng Zhou, Xiubo Geng, Tao Shen, Wenqiang Zhang, and Daxin Jiang. Improving zeroshot cross-lingual transfer for multilingual question answering over knowledge graph. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5822–5834, 2021. doi:10.18653/V1/2021. NAACL-MAIN.465.

How Does Knowledge Evolve in Open Knowledge **Graphs?**

Axel Polleres 🖂 🏠 💿

Vienna University of Economics and Business, Austria Complexity Science Hub Vienna, Austria

Angela Bonifati 🖂 🏠 💿 Lyon 1 University, CNRS LIRIS, France IUF, France

Daniil Dobriy 🖂 🏠 💿 Vienna University of Economics and Business, Austria

Lorena Etcheverry 🖂 🗈 Universidad de la República, Uruguay

Katja Hose 🖂 🏠 🝺 TU Wien, Austria

Matteo Lissandrini 🖂 🕩 Aalborg University, Denmark

Riccardo Tommasini 🖂 🏠 回 INSA Lyon, CNRS LIRIS, France

Romana Pernisch¹ 🖂 🏠 💿

Vrije Universiteit Amsterdam, The Netherlands Discovery Lab, Elsevier, Amsterdam, The Netherlands

Daniele Dell'Aglio 🖂 🏠 💿 Aalborg University, Denmark

StefaniaDumbrava 🖂 🏠 🔎

ENSIIE, France SAMOVAR, IP Paris, France

Nicolas Ferranti 🖂 回

Vienna University of Economics and Business, Austria

Ernesto Jiménez-Ruiz 🖂 🏠 💿

City, University of London, UK SIRIUS, University of Oslo, Norway

Ansgar Scherp 🖂 🏠 🕩 Ulm University, Germany

Johannes Wachs 🖂 🗈

Corvinus University of Budapest, Hungary Centre for Economic and Regional Studies, Hungary Complexity Science Hub Vienna, Austria

— Abstract -

Openly available, collaboratively edited Knowledge Graphs (KGs) are key platforms for the collective management of evolving knowledge. The present work aims t o provide an analysis of the obstacles related to investigating and processing specifically this central aspect of evolution in KGs. To this end, we discuss (i) the dimensions of evolution in KGs, (ii) the observability of evolution in existing, open, collaboratively constructed Knowledge Graphs over

time, and (iii) possible metrics to analyse this evolution. We provide an overview of relevant state-ofthe-art research, ranging from metrics developed for Knowledge Graphs specifically to potential methods from related fields such as network science. Additionally, we discuss technical approaches - and their current limitations - related to storing, analysing and processing large and evolving KGs in terms of handling typical KG downstream tasks.

2012 ACM Subject Classification Information systems \rightarrow Graph-based database models; Information systems \rightarrow Data streaming; Information systems \rightarrow Web data description languages

Keywords and phrases KG evolution, temporal KG, versioned KG, dynamic KG

Digital Object Identifier 10.4230/TGDK.1.1.11

Category Survey

Funding Axel Polleres: supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 957402 (Teaming.AI).

Daniil Dobriy: supported by Jubiläumsfonds der Stadt Wien (project: Collaborative Knowledge Management through Knowledge Graphs: The Case of Managing Historical Knowledge about Vienna).

Corresponding author



© Axel Polleres, Romana Pernisch, Angela Bonifati, Daniele Dell'Aglio, Daniil Dobriy, Stefania Dumbrava, Lorena Etcheverry, Nicolas Ferranti, Katja Hose, Ernesto Jiménez-Ruiz, Matteo Lissandrini, Ansgar Scherp, Riccardo Tommasini, and Johannes Wachs; licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 11, pp. 11:1-11:59 Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



11:2 How Does Knowledge Evolve in Open Knowledge Graphs?

Ernesto Jiménez-Ruiz: supported by the SIRIUS Centre for Scalable Data Access (Research Council of Norway, project 237889).

Johannes Wachs: supported by the Center for Collective Learning (101086712-LearnData-HORIZON-WIDERA-2022-TALENTS-01 financed by EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)) Acknowledgements We would like to thank the reviewers for their invaluable comments that helped to improve our manuscript.

Received 2023-06-30 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

1 Introduction

Knowledge Graphs (KGs) [112] are graph-structured representations intended to capture the semantics about how entities relate to each other, used as a general tool for the symbolic representation and integration of knowledge in a structured manner. The actual *semantics* or schema of such graphs can be formally described using expressive logic-based languages such as the Web Ontology Language (OWL) [101], as well as in terms of constraint languages such as the Shapes Constraint Language (SHACL) [135] or Shape Expressions (ShEx) [195]. Thanks to the expressivity provided by such formalisations, KGs have become a de-facto standard data model for integrating information across organisations and public institutions. It also facilitates the collaborative construction of structured knowledge on the Web by dispersed communities. In other words, KGs serve as intermediate layers of abstraction between raw data and decision support systems. Raising the level of abstraction has allowed us to ask more sophisticated questions, integrate data from heterogeneous sources, and spark collaborations between groups with different perspectives and views on business problems.

As a result of their function as a basis for knowledge integration, KGs are rarely produced in a single one-shot process. Instead, KGs are often collaboratively built and accessed over time. As such, KGs have become a significant driver for the collaborative management of evolving knowledge, integrating knowledge provided by different actors and multiple stakeholders: use cases range from the collaborative collection of factual base knowledge in general-purpose Open KGs such as Wikidata [242] to capturing specialised collaborative knowledge about engineering processes in manufacturing [110].

However, the sheer scale of - in particular - openly available, collaborative KGs has exacerbated the challenge of managing their evolution, be it in terms of (i) the size and temporal nature of the data, (ii) heterogeneity and evolution of the communities of their contributors, or (iii) the development of information, knowledge, and semantics captured within these graphs over time.

Even though analysis of the content, nature, and quality of KGs has already attracted a vast amount of research (i. e. [192, 104, 202] and references therein), these works focus less on how their structure and contents change over time, indeed how these systems evolve.

With the present article, we aim to shift the focus on precisely this matter. In particular, we try to answer the following main questions:

- **RQ1** Which publicly accessible, open KGs are observable in a manner that would allow a longitudinal analysis of their evolution and how? That is, how could we obtain historical data about their development, or which infrastructures and techniques would we need to monitor their growth and changes in the future?
- **RQ2** Which metrics could be used to compare the evolution and structure over time, and how could existing static metrics be adapted accordingly? Here, we are particularly interested in approaches from other adjacent fields, such as network science, and how those could be adapted and applied to specifically analyse the evolution of knowledge graphs.

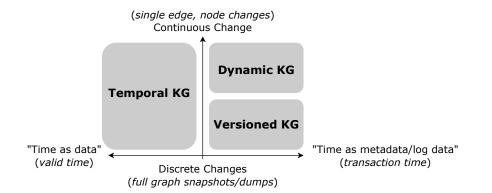


Figure 1 Perspectives on evolving knowledge graphs. We distinguish between considering time as data or metadata (x-axis) vs. at which "granularity" changes range from full static "bulk" snapshots at different points in time to continuous reporting of changes at the level of single edges or nodes (y-axis). We explicitly note that these perspectives are *not* mutually exclusive: the visualisation does not capture the possible overlaps of these perspectives, nor the nature of *how* changes are produced and captured, e.g. in terms of underlying collaborative processes.

RQ3 Finally, do we have the right techniques to process evolving KGs, both in terms of scaling monitoring and computing the necessary metrics, but also in terms of enabling longitudinal queries, or other downstream tasks such as reasoning and learning in the context of change – facing the rapid growth and evolution of existing KGs?

To approach these questions, the remainder of this article surveys existing approaches and works and raises open questions in four directions: observing, studying, managing and spreading KG evolution. Before elaborating on these directions, we first discuss the different *dimensions of evolution* in Section 2, introducing relevant terminology. In Section 3, we discuss to what extent data about the evolution of open KGs (like Wikidata or DBpedia) is available and what evolution trends have been observed so far in prior literature. In Section 4, we discuss different types of metrics to study evolving KGs; starting from state-of-the-art graph and ontology metrics, we also discuss metrics related to quality and consistency, as well as potentially valuable works and metrics from the area from network science. In Section 5, we discuss data management problems for evolving knowledge graphs, i.e. data models that capture temporality as well as storage approaches and schema mappings for versioned and dynamic KGs. In Section 6, we focus on downstream tasks on KGs in the specific context of evolution. More precisely, we discuss how querying, reasoning, and learning approaches can be tailored for evolving KGs. We also address the exploration of KGs, an essential aspect of evolving KGs. We conclude with a summary of the main research challenges we currently see unaddressed (or only partially addressed) in Section 7.

2 Dimensions of Evolution

The temporal evolution of graphs, knowledge graphs (KGs), and collaboratively edited KGs has multiple dimensions that we outline in this section, along with relevant terminology. That is to say, there are multiple coherent perspectives we can use to talk about the "*evolution*" of KGs, ranging from considering time and evolution as being part of the data itself to considering evolution and change over time on a meta-level. We illustrate these perspectives in Figure 1.

11:4 How Does Knowledge Evolve in Open Knowledge Graphs?

Temporal KGs: Time as data

The first perspective considers time, or – more concretely, the temporal validity of information in a KG – as part of the KG itself; we call this the "Temporal KG" perspective. In this context, the evolution depicted by the data pertains to the changes in the "world" it represents, not the evolution of the data itself. Following database terminology, this temporal validity of information in a KG is typically referred to as *valid time*; see, for instance, [103]. A very simple example of a temporal KG is illustrated in Figure 2, which contains the year of production of Picasso's "Guernica", as a slightly simplified subgraph DBpedia [146].¹

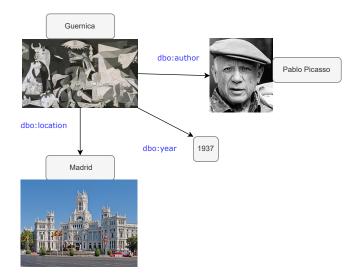


Figure 2 A simple KG containing temporal information as data (literal).

Time and temporality may be represented with a single temporal literal – as illustrated here a year or a timestamp, or likewise an interval: for instance, the production of "Guernica" itself was not a one-shot process, but its painting took place over a longer period. For instance, the production period of "Guernica" was carried out between 1937-05-01 and 1937-06-04, as illustrated in Figure 3, a simplified graph inspired by the Linked Art project.²

We note here that capturing intervals typically requires extensions of the "flat" directed labelled graph model used to represent simple knowledge graphs, as shown in Figure 3: contextual information about simple statements (such as in this case, the start and end time of a production interval), can be modelled in various ways, either

1. in terms of adding intermediate nodes to a flat graph model, also often referred to as "reification", or alternatively

2. in terms of bespoke, extended graph models such as so-called *property graphs*

Let us refer to Section 5.1 for a more in-depth discussion of different data models to capture time and temporality in KGs.

Time-varying KGs: Time as meta-data

The second perspective on evolution is scoped by the time granularity of change in the KG itself; in other words, by how the temporal aspect of the data, i.e. nodes, edges, and structure, of the KG is evolving. We call this the "Time-varying KG" perspective. Again, using database terminology, such changes in data are typically referred to as *transaction time* [103].

¹ https://www.dbpedia.org/

² https://linked.art/model/

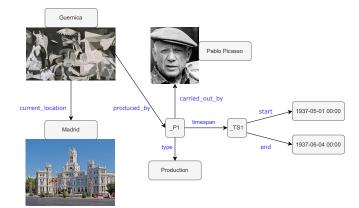


Figure 3 A KG containing temporal information in terms of intervals using a simple form of reification.

We present an example from the arts. Paintings like "Guernica" and information about their artists and other attributes have been added dynamically to Knowledge Graphs like Wikidata over time. The entry for "Guernica" (Q175036) in the Wikidata [242] KG was created on 28 November 2012,³ while its creator "Pablo Picasso" (Q5593) was added on 1 November 2012⁴. Of course, both of these dates are independent of the birth or production dates of the referred entities themselves. As we will further discuss in Section 3 and also Section 5 below, the granularity and manner of how such changes are stored affect the observability and analysis of a KG's evolution.

In terms of granularity, we can differentiate between two types of knowledge graphs based on how they are stored:

- Dynamic KGs which allow access to all observable atomic *changes* in the knowledge graph.
- Versioned KGs which provide static snapshots of the materialised state of the knowledge graph at specific points in time.

These represent opposite ends of the granularity spectrum. Figures 4 and 5 show two examples of how the changing information regarding the location of "Guernica" over time⁵ could be represented in terms of versions or dynamic changes, respectively.

For instance, as discussed above, Wikidata embodies continuous change, accessible through the entities' edit histories at the level of real-time modifications. At the same time, DBpedia represents both the spectrum's discrete end, releasing snapshot updates,⁶ as well as offering small-scale releases with DBpedia Live⁷ on minute level. Observe that in both cases, the temporal information about neither the materialisation time of a DBpedia snapshot or the edits of single statement claims on Wikidata are available in terms of the (RDF) graph materialisations of these KGs themselves, but only in terms of the publication metadata or edit histories, which is why we may also speak of "time as meta-data".

We note that this distinction is hardly clear-cut. The difference between dynamic and versioned temporalities is marked by the technical means by which particular KGs evolve. In particular, this boundary is shaped by differences in technical infrastructures supporting these evolutionary processes rather than general characteristics of the KG and the kind of knowledge it captures.

³ https://www.wikidata.org/w/index.php?title=Q175036&action=history&dir=prev

⁴ https://www.wikidata.org/w/index.php?title=Q5593&action=history&dir=prev

⁵ The painting was first exhibited in Paris in 1937, and moved to an exhibition in New York in 1939. Since 1992 "Guernica" is displayed in Museo Reina Sofía in Madrid.

⁶ https://www.dbpedia.org/resources/snapshot-release/

⁷ https://www.dbpedia.org/resources/live/

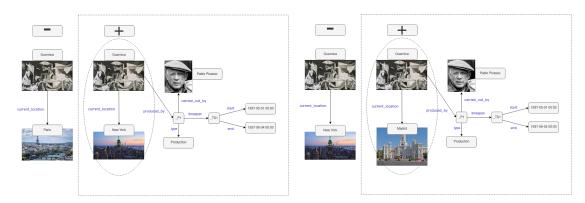


Figure 4 Representation of Dynamic Knowledge Graph, with updates at edge level, i. e. deletions (left) and future additions (right).



Figure 5 Representation of Versioned Knowledge Graph, with snapshots sorted by time.

For example, on the one hand, while changes in Wikidata may be recorded down to the level of single statements, Wikibase⁸ also supports interfaces for bulk updates. Likewise, each single statement change in Wikidata's overall edit history may be theoretically materialised in terms of sequential snapshots. On the other hand, DBpedia's extraction framework constructing a KG from Wikipedia may be analogously applied to any materialised point in time of the fine-granular page edit history of Wikipedia, or even per page [80]. DBpedia's model has also changed over the past years from irregular, approximately annual, snapshots published in its beginnings, to enable more dynamic publishing (monthly) cycles [111] through the DBpedia Databus.⁹

Lastly, we note that analogously to the examples in Figures 2 and 3 both *timestamps* and *time intervals* can be used to represent not only validity but also transaction and versions, i. e. snapshots of the entire graph in the context of KGs. However, depending on which dimension is considered, it will have an impact on how data should be managed, whether evolution is observable, and how the information about evolution is spread into downstream tasks, see the further discussions in Sections 5 and 6 below.

⁸ https://wikiba.se, Wikidata's underlying software framework.

⁹ https://www.dbpedia.org/resources/databus/

Both of the aforementioned perspectives can serve the purpose of monitoring the evolution of KGs along different yet interrelated (sub-)dimensions. We outline these dimensions in the following subsections. First, according to Section 2.1, the *structural evolution* of KGs can be observed through the temporal information captured in them; here, KGs present a distinction between changes on the data and schema levels. Second, one can analyse the dynamics or velocity of evolution in KG over time, see Section 2.1. Finally, when considering the collaborative processes involved in KG editing and evolution, one can analyse the structure and dynamics of these collaborations, see Section 2.2. After exploring these dimensions in detail, we then discuss concrete metrics in Section 4.

2.1 Structural Evolution, Dynamics, Timeliness, and Monotonicity

In the context of evolving KGs (hereafter EKGs), we may consider different forms of change related to the graph structure, dynamics of change or its nature (monotonic or with deletions), and alternative notions of time. The following will briefly elaborate on our running example in Figure 6.

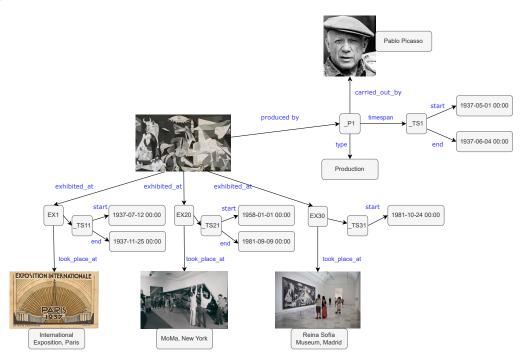


Figure 6 A sample KG containing temporal information about the production (static) and exhibitions (dynamic) of paintings.

Structural Evolution

The first dimension to measure on a graph is essentially related to its structure: descriptive statistics about nodes and edge distributions, centrality, connectedness, density, and modularity. In KGs, similar static metrics can also be observed concerning the *schema*, typically the node and edge *types*, and – if additionally axiomatic knowledge on the schema-level is considered – the complexity of this schema.

For all of these structural properties (both on the instance-level and schema-level), we may also be interested in their development over time, i.e. in quantifying their changes. The existing concrete metrics for this dimension will be discussed in more detail in Section 4 below.

11:8 How Does Knowledge Evolve in Open Knowledge Graphs?

Notably, longitudinal investigations of structural properties are not restricted to the timevarying KG perspective: depending on whether temporal information is present in the KG itself, one may also be interested in analysing and comparing structural evolution in terms of "temporal slices".

Dynamics

Dynamics for KGs refers to characteristics such as growth and change frequencies over time and per time interval). These may be observed overall but also in terms of subgraphs or topic-wise components of a KG. For instance, one may consider comparing the change dynamics of entities related to different topic areas, such as "arts" and "sports" within a particular KGs like Wikidata. Again, these dynamics may be observed concerning the KG schema. Referring to a concrete elaboration of our running example in Figure 6, we can derive that properties related to the *production* of paintings evolve more slowly than properties relating to exhibitions. Notably, dynamics and temporal granularity may again be compared and analysed both from secular and time-varying perspectives.

Timeliness

Timeliness, from a data quality perspective, refers to the "freshness" of the data concerning the occurrence of change, the current time, or the time of processing. Timeliness directly links to query answering (or processing in general), as it establishes the value of the retrieved answer considering some requirements. More specifically, the timeliness of data in a KG can be interpreted as

- "out-of-date" or "stale" information: i.e. in terms of recency of temporal information concerning the current time;
- "out-of-sync" or "delayed" information, i.e. in terms of the difference between valid times and transaction times of items in the KG, i.e. the interplay between these temporal and time-varying perspectives.

Regarding the former case, considering Figure 6, the question "Where is Guernica currently?" obtains a different answer at different times. While historical events such as the creation of "Guernica" lie far in the past, even far before Wikidata was founded, the location of paintings is an important dimension to analyse over time as it changes with exhibitions or purchases. If neglecting such variations is an issue for the users, e.g. when an accurate current location is needed to recommend a museum visit, then we witness a data quality problem related to timeliness.

A "drastic" example of the latter, i. e. extended out-of-sync information from the art domain is documented in Rembrandt's "Portrait of a Young Woman" (Q85523581 in Wikidata) from 1632, which was added to Wikidata only in February 2020, after it was recently confirmed to be an authentic Rembrandt.¹⁰ Users who have asked for the number of Rembrandt paintings before 2020 would have received a stale answer.

Monotonicity

Monotonicity refers to the nature of changes, i.e. if they are positive changes only augmenting the content of the graphs, or if they take the form of an update which may include *deletions* of past information.

Continuing our examples in the domain of painting, we consider rectifying a painting's attribution to its artist, which happens repeatedly in arts. A documented case is the painting "Girl with a Flute" (Q3739200) in Wikidata, originally attributed to the Dutch painter Vermeer

 $^{^{10}\,\}tt https://news.artnet.com/art-world/pennsylvania-museum-rembrandt-discovery-1773954$

but later confirmed to be the work of another painter.¹¹ Similar, non-monotonic changes may arise when temporal information itself changes in the KG: imagine, following our running example, that subsequent research may reveal Guernica was actually created in 1936, not 1937.

From this combination of dynamics (i.e. the study of changes), timeliness, and monotonicity (i.e. the frequency of deletions and, therefore, errors and rectifications of incorrect information in a KG), it is also possible to estimate the frequency of future transactions. Together they form an essential dimension of evolving KGs, both in the context of the ability to process evolution technically but in terms of its impact on the validity of updated results of downstream tasks Section 6: as KGs are meant to support sophisticated decision-making tasks, it is often paramount to guarantee up-to-date information and provide answers before they become obsolete.

2.2 Evolution in Collaboration

Knowledge evolution is driven by different types of collaborations [190, 5]. As described by Piscopo et al. [190], collaborative KGs rely on experts for specific types of activities, defining rules and processes for how and by whom some activities should be carried out, or provide tools to facilitate such collaboration.

In the context of KG evolution, we may thus want to analyse the behaviours of single users or user groups over time. To classify the collaboration types, we can distinguish the following roles of users/agents:

- Anonymous users: These are Users who do not have a registered account or a consistent identity within a project (e.g. anonymous Wikibase users)
- Registered users: similarly, these are *Users* who have a registered account or a consistent identity within a project (e.g. registered Wikibase users), ideally also combined with additional information or characteristics which allow to classify such users (e.g. country of origin or other demographic attributes)
- Authoritative users: These are Users characterised by in-depth domain knowledge or knowledge engineering expertise. This group represents vetted knowledge engineers, domain experts, and moderators.
- **Bots:** These are automated agents performing recurring tasks (e.g. Wikibase bot accounts).

Longitudinal analyses of the contributions of such users may include changes in their behaviours (e. g. in terms of edit frequencies), interests (e. g. in terms of editing particular parts or topics of KGs), or role changes. Additionally, based on the aforementioned roles, various collaboration types can be potentially recognised when analysing the evolution of edits in collaboratively edited KGs [191]:

- **Expert-driven collaboration:** this type of collaboration involves *Authoritative users* developing schemas or editing data on the instance-level (creating mapping rules, as in the case of DBpedia, would be an example of such schema-level expert collaboration, whereas the instance data, origins from Wikipedia, thus following another collaboration model).
- **Crowd-sourced collaboration:** this type of collaboration involves many *Users* not considered *Authoritative users* performing basic editing tasks which neither requires in-depth domain or knowledge engineering expertise nor coordination between the editors (for instance, any users being allowed to edit Wikipedia could be understood as such a crowd-sourced collaboration model, if a more moderated process did not govern it, see below).

¹¹ https://www.wikidata.org/w/index.php?title=Q3739200&oldid=803621750

11:10 How Does Knowledge Evolve in Open Knowledge Graphs?

KG	Expert-driven	Crowd- sourced	Resource- dependent	Community- driven	Bot-assisted
Wikidata [242]		\checkmark		\checkmark	\checkmark
DBpedia [146]	\checkmark		\checkmark		(\checkmark)
YAGO [153]		\checkmark			

Table 1 Types of Collaboration in Open-domain Knowledge Graphs.

- **Resource-dependent collaboration:** This type of collaboration is based on integrating information from external resources, potentially governed by different heterogeneous collaboration models (indeed, DBpedia's extraction of instance data from Wikipedia may be understood as such a resource-dependent "collaboration").
- **Community-driven collaboration:** this type of collaboration relies on self-moderating communities of *Users* characterised by deep involvement in the project, collective discussion, and decision making (e.g. Item/Property discussions characteristic for Wikidata, but also characteristic for the curation process in Wikipedia).
- **Bot-assisted collaboration:** this type of collaboration is characterised by *Bots* performing repetitive tasks alongside *Users* (i.e. curation tasks, e.g. checking property constraints on Wikidata, but also, indirectly in DBpedia, via bot interactions in Wikipedia, cf. [254, 50]).

Table 1 describes the common collaboration models of some existing, collaboratively maintained open general-purpose KGs, according to the literature. We note that the list of KGs shown here is not meant to be exhaustive and that such metrics could be further extended and refined in more fine-grained longitudinal analyses. As described in Section 4.4, for example, topologically identified groups of collaborators could be used to predict outcomes. A concrete methodology to analyse the composition of the collaborators within the KG and assess their effects on quality has been suggested in [189]. Further investigation can also include the different evolution and collaboration approaches and how these influence the possibility of analysing evolution. For example: does the relatively small DBpedia ontology and the limited frequency of updates via mapping changes make the analysis of the evolution of its ontology easier than the direct ontology editing model of Wikidata? Does the extraction and mapping mechanism and changes to the rules that drive them make ontology evolution in turn less flexible for the community in DBpedia? Likewise, does the free-for-all collaboration approach in Wikidata render a structured analysis of ontology evolution impossible, or what are the methods to handle this challenge? For instance (i) can one define "checkpoints" of limited changes that can be used as anchor points to produce useful analyses, or (ii) does it make sense to investigate the evolution of vocabularies specifically scoped to editors' sub-communities? Another avenue for investigation is a more effective utilisation of machine learning in supporting the collaborative evolution of KGs and their schemas. Specifically, it would be interesting to learn how this evolution is affected and affects the interaction of automated extraction (DBpedia), extraction by statistical learning (YAGO), or in leveraging or improving bots (Wikidata): that is, can ontology extraction rules or curation pipelines be improved by observing and learning from the collaboration and evolution processes over time?

2.3 Semantic Drift

Semantic drift is a crucial concept of evolution in language. It refers to the *change in meaning* of a concept over time [246, 218] independently from the downstream tasks like querying or reasoning. Before detecting semantic drift, one needs to identify the two concepts to compare between versions. Although early work on identifying semantic drift focused on the definition of the

identity of a concept [246], when a concept changes meaning, it might also change its identifying information. Therefore, it is not always possible to rely only on identity-based approaches to understand semantic drift. In such cases, morphing chain-based strategies are more suitable [90]. The morphing chain approach presents the user with a comparison of a concept to all the concepts between the versions of an ontology and lets the user choose or chooses heuristically which is the most likely concept that a previous one evolved into.

For KGs, Meroño-Peñuela et al. [158] studied semantic drift in DBpedia concepts, while Stavropoulous et al. [219] studied semantic drift in the context of the Dutch Historical Consensus and the BBC Sports Ontology. SemaDrift [218] takes a morphing-chain approach, where three aspects are used to identify concepts that have potentially evolved from another: label, intention, and extension. The advantage of this approach is that every concept in a new version will have evolved from some previous concept. Unfortunately, the identity of concepts, such as URI, is not used in SemaDrift. OntoDrift [44] uses a hybrid approach and can be considered an extension of SemaDrift [218]. Additionally to using the label, intention, and extension aspects of concepts, it also considers the subclass relations. The drawback of this approach is that rules need to be defined for every type of predicate, as demonstrated by OntoDrift.

The notion of logical difference [136] between KGs can also be used to evaluate the semantic drift of the KG concepts. The logical difference focuses on the entailments or facts that follow from one KG but not from the other, and vice versa. Jiménez-Ruiz et al. [126] proposed an approach to evaluate the logical difference among different versions of the same ontology. Considering the new logical entailments/axioms involving a given entity, one could define a metric. The entity's role within the entailment (i. e. the entity is being defined vs. the entity referenced) may also impact the metric.

Potential approaches in the future could make additional use of embeddings, representing concepts in vector space and assessing their neighbourhoods. Pernisch et al. [181] showed that comparing two embeddings to each other is complex, and the similarity between concepts is, e. g. around 0.5 for FB15k-237 with TransE; Verkijk et al. [240] further discuss the difficulties with this approach, especially comparing it to concept shift in natural language. Finally, the lack of domain-specific benchmarks for semantic drift makes comparing methods difficult. For instance, OntoDrift and SemaDrift return very different numbers when detecting drift, but we cannot tell which ones are closer to the truth. Also, the number of studies that look at semantic drift is limited. Not many KGs have been studied, and even though the phenomenon is known, it has not been investigated extensively so far [158, 219].

3 Observe and Analyse the Evolution

This section discusses how far evolution can be observed and analysed along the dimensions defined above in various existing KGs. KGs come in very different flavours and structures, and in particular, we may also assume that their evolution shows very diverse characteristics.

Below, we first characterise different kinds of graphs. In Section 3.1, we discuss tools to observe the historical longitudinal data on the evolution of the most important existing KGs. Section 3.2 provides a respective overview of available studies to analyse and track the dynamics of some of these KGs. We consider both monitoring and analysing the evolution of the instance-level of graph data as well as the schema-level.

Without claiming completeness, we distinguish the following kinds of KGs:

General-purpose Open Knowledge Graphs: publicly available open-domain (or, resp., cross-domain) KGs such as DPpedia [146] and Wikidata [242] as two of the most prominent KGs have been developed since more than a decade by now, covering a wide range of comprehensive

11:12 How Does Knowledge Evolve in Open Knowledge Graphs?

knowledge. Yet, they differ fundamentally in the process in which knowledge is maintained and developed within the KG: whereas DBpedia relies on extractors to collect data from Wikipedia's infoboxes regularly, Wikidata comprises a completely collaboratively evolving schema and factbases that, by themselves, feed back into Wikipedia. In particular, we observed significant growth and dynamics in both the instance-level and schema-level of Wikidata over the past years. Collections of structured RDF data and microdata (e.g. schema.org [102] metadata) from Web pages through openly available Web crawls, such as made available regularly by the Webdatacommons¹² project [159], may indeed also be perceived as evolving, general purpose, real-world Knowledge Graphs.

- **Domain-specific Special-purpose Open Knowledge Graphs**: Many open knowledge graphs available to the public are often overlooked. These graphs are collaboratively developed and serve narrow, special-purpose topics or use cases. An example is Semantic MediaWiki (SMW)[138], which has been around for almost 20 years and is still actively developed and used in various community projects. SMW can be considered a predecessor of Wikibase, the underlying platform for Wikidata. Wikibase is increasingly being used in separate, special-purpose community projects. Other examples of domain-specific knowledge graphs include the UMLS Metathesaurus [34], as well as the ontologies in the OBO Foundry [121], and BioPortal [248]. These graphs focus on the schema and are assumed to have significantly different evolution characteristics [182].
- **Task-specific Knowledge Graphs**: One category of Knowledge Graphs that some authors identify is task-specific Knowledge Graphs [122]. These graphs, often used in benchmarks, are typically subsets of larger KGs created to support a specific application or may result from a downstream application (e.g. DBP15K as a subset of DBpedia for cross-lingual entity alignment). However, since these KGs are usually artificially limited and static (i.e. subset of specific snapshots), compared to real-world evolving KGs, we will not discuss them separately in this paper. We note, however, that principled approaches to *create evolving subsets* of KGs for specific benchmarking tasks are sorely needed to better understand these tasks "in evolution".
- **Large (and Small) Enterprise Knowledge Graphs** Lastly, we see many companies reportedly using and adopting Knowledge Graph technologies in their operations and businesses over the past years, including large firms like Google, Amazon, Facebook, and Apple, as well as many other smaller examples. What these KGs typically have in common is that due to their commercial value, they are non-observable to the community and we may only speculate about their sizes and structures using white papers [170, 209, 117], high-level announcements, and to some extent through industry track reports in conference series such as ISWC (e.g. [97]), SEMANTiCS (e.g. [204]), or recently the Knowledge Graph conference series. Given these limitations, we exclude enterprise KGs from the scope of the present paper.

Except for the latter two cases then, it appears that the research community has built up a large number of publicly accessible and observable KGs that vary in characteristics, and purpose, with unique communities of maintainers that seek to capture a rich variety of knowledge artefacts in evolving graph-like structures. In the remainder of this section, we specifically focus on Open General-purpose KGs rather than attempt to cover all types of KGs.

	Level	Queryable	Collaborative	Formats	Protocol	Metadata	Temporality	Timeliness
Wikidata	V	Yes	Yes	NT, TTL, HDT, JSON	HTTP, SPARQL	schema.org	No	2-3 Days
	S	Yes	Yes	NT, TTL, JSON	HTTP, SPARQL	schema.org	No	2-3 Days
-	CL	Yes	Yes	JSON	SSE	No	Event TS	Seconds
а	V	Yes	Partial	NT	HTTP,	No	No	Quarterly
$\mathbf{DBpedia}$	\mathbf{S}	Yes	Yes	RDF	SPARQL HTTP, SPARQL	No	No	Daily
Η	CL	Yes	Yes	RDF	HTTP	No	${\rm Graph} \ {\rm TS}$	Daily
0	V	Yes	No	RDF all	HTTP	No	No	NA
YAGO	\mathbf{S}	Yes	No	RDF	HTTP	No	No	NA
X	CL	No	No			No		
	V	Yes	Yes	RDF all	HTTP, SPARQL	Some	No	NA
LOD	S	Yes	Yes	RDF, OWL	HTTP, SPARQL	No	No	NA
	CL	Depends on individual datasets						
GDELT	V	Yes	No	SQL Big Query	HTTP	No	Yes	15 min
	S	No	No	CSV, JSON, XML	HTTP	No	No	No
•	CL	No	No	CSV	HTTP	No	Event TS	15 min

Table 2 Availability of Open KG Versions (V), Schema (S), and Change logs (CL).

3.1 Availability of Graph Data

In the following, we start by assessing how and where historical longitudinal data about existing open KGs and their evolution can be found. We specifically focus on KGs that are still available and, therefore, do not include KGs like Freebase [36] and OpenCyc [156]. These two KGs are no longer maintained but are considered pioneering work and predecessors of the KGs investigated in this subsection. Therefore, it is generally possible for KGs to go dark, e.g. through neglect or malign actions.

Here, we give an overview of the datasets regarding the availability of their versions, their schema, and their changelogs in Table 2. The table captures if the versions, schema, or changelogs are queryable and collaborative. Queryable in this context captures if the KG answers queries in any way or form specifically over (historical) versions, schema as well as change logs, for which we then further specify the protocol (HTTP, SPARQL, etc.); for possible temporal queries over RDF archives that should be enabled over evolving KGs, we refer to, for instance, the categorisation in [84, Section 3.2]. Collaborativeness in Table 2 refers to the possibility of reconstructing user information on the different levels. For example, on the changelog level, a "yes" refers to having user information for individual changes. Wikidata and DBpedia allow anonymous edits, which potentially limits a reconstruction of the editing history, indicated with "Partial" in the table.

¹² http://webdatacommons.org/

11:14 How Does Knowledge Evolve in Open Knowledge Graphs?

Further information on formats (RDF, JSON, etc.) is given. Temporality refers to the ability of the KG to capture temporal information for example through reification or other means. With "Event TS", we indicate that the KG allows for events to be timestamped, whereas with "Graph TS", we refer to the whole graph having timestamps. Lastly, timeliness refers to how often the part of the KG is updated.

Wikidata is an open KG read and edited by humans and machines and is hosted by the Wikimedia Foundation. Intuitively, the considerable level of automation and collaboration on Wikidata, and its scale¹³ present significant challenges in Wikidata evolution maintenance.

As for direct *queryability*, Wikidata's public SPARQL endpoint¹⁴ provides query access to the current, regularly synced snapshot; it is undisputed that due to its scale, querying Wikidata in the light of its rapid growth – even on static snapshots – is currently reaching its limits in terms of regular SPARQL engines, as well documented for instance in [13]. Yet, there are various ways to access and potentially – given the respective infrastructure – query the historic versions and change data about Wikidata: Wikidata Entities dumps are available in JSON in a single JSON array, or RDF (using Turtle and N-triples) with Full RDF dumps are available for download¹⁵ every 2-3 days, and historically for approximately a month. Schema.org metadata is used to describe the dump that contains additional helpful metadata such as the entity revision counter (schema:version), last modification time (schema:dateModified), and the link to the entity node with (schema:about).

As a subset, also *truthy dumps* are provided, which are limited to direct, truthy statements – since Wikidata offers (validTime) temporal annotations for statements, as well as provenance annotated statements, this "truthy" subset contains only currently valid or preferred ranked statements, where however additional metadata such as qualifiers, ranks, and references are consequently left out. The truthy dump could, therefore, be perceived as a "current truth" snapshot of Wikidata. In contrast, the entire dump also contains outdated (valid time) or disputed (in terms of being lower-ranked alternative statements by particular contributors).

RDF HDT¹⁶ hosts roughly annual HDT [83] snapshots of Wikidata's complete dumps. In addition to these hosted RDF dumps, obtaining the statement-level change log from Wikidata's aggregated entity and editing history, which are also available via respective APIs, would be possible.

Finally, Wikimedia offers changes (of both Wikipedia and Wikidata) through the **Wikimedia Event Streams**¹⁷ Web service that exposes continuous streams of JSON event data. It uses chunked transfer encoding following the Server-Sent Events protocol (SSE) and emits changes events, including Wikidata entity creations, updates, page moves, etc. The usage of edit history and event stream data, apart from RDF dumps, also has the advantage of making (where available) user/contributor information visible, which is helpful for collaboration analyses. Pelisser and Suchanek [225] have presented a prototype to provide this additional information in RDF via a SPARQL interface.

Wikidata Schema/Ontology. Wikidata does not follow a pre-defined formal ontology, meaning it does not formally differentiate between classes and instances. Instead, the terminology is derived from the relationships between the items in the graph and is collectively created by the editors. In other words, Wikidata (deliberately) does not make a formal commitment to the logical meaning of its properties and classes, which could be, for instance, roughly defined as the objects of the P31 (instance of) property.

¹³ with over 15B triples at the time of writing: https://w.wiki/7iez.

 $^{^{14}\, {\}tt query.wikidata.org}$

¹⁵ https://dumps.wikimedia.org/wikidatawiki/entities/

¹⁶https://www.rdfhdt.org/datasets/

¹⁷ https://stream.wikimedia.org/

As a consequence, Wikidata's schema is evolving entirely in parallel with its data – and analogous considerations for the availability of data about its historic evolution apply as mentioned above. This has been reported to pose significant data quality challenges [190]; moreover, as a primary consequence of such an informal, collaborative process, Wikidata's ontology may change quickly. In practice, this does not impact the evolution of the graph itself, but it poses an obstacle to downstream tasks and analyses. We note that prior attempts to map the user-defined terminological vocabulary of Wikidata to RDFS and OWL, such as [105], could be used to partially map Wikidata to more standard ontology languages and conduct (approximate) analyses on a logical level. In contrast, we should note that theoretically, OWL/RDFS "mappable" properties could evolve independently in Wikidata.

DBpedia is an openly available KG encoded in RDF, which evolves alongside Wikipedia. It has four releases per year (approximately the 15th of January, April, June, and September, with a five-day tolerance), named using the same date convention as the Wikipedia Dumps that served as the basis for the release.¹⁸ **DBpedia Latest Core Releases**¹⁹ are published separately as small subsets of the total DBpedia release. Its extraction is fully automated using MARVIN [111] and then catalogued. The standard release is available on the 15th of each month, five days after Wikimedia releases Wikipedia dumps. **DBpedia Databus**²⁰ is a platform designed for data developers and consumers to catalogue and version data, not only restricted to DBpedia alone. It enables the smooth release of new data versions and promotes a shift towards more frequent and regular releases. DBpedia takes advantage of this functionality to promptly publish the most up-to-date DBpedia datasets, generating approximately 5,500 triples per second and 21 billion triples per release every month. **DBpedia Live**²¹ is a changelog stream accessible in a pull manner. DBpedia Live monitors edits on Wikipedia and extracts the information of an article after it was changed. A synchronisation API is available to transfer updates to a dedicated online SPARQL endpoint, whereas temporal evolution as such is not directly queryable from that endpoint.

DBpedia Ontology (DBO), the core schema of DBpedia, is currently crowd-sourced by its community: *DBpedia mappings* are contributed and made automatically available daily, where DBO is generated every time changes in the mappings Wiki have been made. Notably, DBpedia Latest Core and DBpedia Live are based on the latest DBO snapshot available at the point of generation, i.e. one should consider the evolutions of data (Wikipedia edits), schema (mappings), and also the various releases of the actual DBpedia KG, separately.

Finally, we note that a fine-grained historical development, in terms of reproducing any DBpedia page at any point in time in the past, and thereby reconstructing a fine-grained RDF "history" would be theoretically possible by combining DBpedia's mappings with the Wikipedia edit history API. A prototypical implementation of this approach, the "DBpedia Wayback Machine" – inspired by the Web Archive's Wayback machine – has been presented by Fernández et al. [80].

YAGO is a large multilingual KG with general knowledge about people, cities, countries, movies, and organisations [220]. At the time of writing, there are six versions of YAGO. In its latest version, 4.5, YAGO combines Wikidata and Schema.org. Older versions integrate different sources such as Wikipedia, WordNet, and GeoNames but are independent of the most recent ones. YAGO places a strong emphasis on data extraction quality, achieving a precision rate of 95% through manual evaluation [198]. One of YAGO's unique features is its inclusion of spatial and temporal

¹⁸https://www.dbpedia.org/resources/snapshot-release/

¹⁹ https://www.dbpedia.org/resources/latest-core/

²⁰ https://databus.dbpedia.org/

²¹ https://www.dbpedia.org/resources/live/

11:16 How Does Knowledge Evolve in Open Knowledge Graphs?

information for many facts, enabling users to query the data across different locations and time periods. Since version 4, YAGO combines Schema.org's structured typing and constraints with Wikidata's rich instance data. It contains 2 billion type-consistent triples for 64 million entities, providing a consistent ontology for semantic reasoning with OWL 2 description logics. Temporal information in YAGO 4 is sourced from Wikidata qualifiers, which annotate facts with validity periods and other metadata. YAGO 4 adopts the RDF* model for representing temporal scopes, enabling precise assertions about facts within specific timeframes. This approach ensures accurate temporal modelling without implying current states [180]. YAGO can be accessed in different RDF formats, but little information is provided on its evolution or the changes in its schema.

The <u>LOD Cloud</u>,²² is, although regularly re-published and maintained since 2007, a collection/catalogue of (interlinked) Knowledge Graphs, rather than a KG on its own. Due to its decentralised nature, anyone can submit a dataset, and the evolution of the respective constituent KGs is not observable from this source directly. While many of its catalogues KGs are accessible via dumps or even SPARQL endpoints, at the same time, many of its datasets have disappeared over time and are no longer (or irregularly available).

As for queryability, the LOD-a-LOT dataset,²³ which has been created as an attempt to clean and crawl all accessible datasets of the LOD cloud and make it available in HDT [83] compressed form [82] – to the best of our knowledge this remains to date a static, once-off effort. While this dataset has also been re-used in other works, for instance, to analyse cross-linkage and ontology-reuse within the LOD Cloud [104], such investigations are lacking a longitudinal analysis of development over time. Likewise, little is known about the evolution of its schema expressivity: a once-off study from 2012 on the Billion Triple Challenge sample from different LOD Cloud datasets has found for instance that hardly any OWL2 constructs had been used at the time [95], and most of the ontologies in Linked Data had used only a moderately expressive fragment of OWL, which had been called OWL LD in this study. A subsequent or even continuous assessment over time with respect to changes or uptake of OWL constructs in LOD over time is to the best of our knowledge still missing. We note that, while the evolution of the LOD Cloud schema itself was partially studied, e. g. the changes and interlinkage of the RDF vocabularies [1, 2], this study did not include expressivity as such.

Unfortunately, such longitudinal analyses over the LOD cloud's evolution as a whole are hardly reproducible or observable a posteriori, since, by its nature, availability of versions, separate schemata and change logs, as well as information about temporality and timeliness is highly heterogeneous across the LOD Cloud datasets. Only summary statistics about the individual states of available datasets at the time of updates are available; i. e. the LOD Cloud service as such does not capture the LOD's historical development itself and older versions of the data itself are typically not provided. External initiatives have attempted to address this problem:

- the Billion Triples Challenge (BTC)²⁴ initiative that, starting from a certain set of seeds, collected billions of triples on the LOD using the popular LDspider [118] framework. The first BTC snapshot of the LOD Cloud from 2009 contained about 1B triples. The crawls have been repeated in irregular year-based intervals. The largest version is from 2014, with about 4B triples.
- The Dynamic Linked Data Observatory (DyLDO) [140]²⁵, initiated in 2012, partially overcomes this limitation by providing weekly snapshots of about 90,000 URIs using the same crawler as the BTC dataset, resembling about 150 to 205 million triples per week. Key characteristics

 $^{^{22}}$ https://lod-cloud.net/

²³http://lod-a-lot.lod.labs.vu.nl/

²⁴ https://www.aifb.kit.edu/web/BTC

²⁵http://km.aifb.kit.edu/projects/dyldo/

of the dataset are that the weekly crawls are stored as so-called snapshots using the N-Quad format [45]. This means that the full graph data collected per week is available in a single data dump. The variance of the collected data reflects the changes in the LOD Cloud. The main drawback of this approach in evolution analysis is that the seed URLs have not changed since the start of the data collection; this initiative is apparently the longest-running collection of a subset of the LOD Cloud.

While well-known, publicly available Knowledge Graphs (KGs) such as DBpedia and Wikidata play a significant role in the realm of structured knowledge, there are other, perhaps less widely recognised, but equally substantial KGs that deal with highly dynamic data. Two notable examples are the GDELT Global Knowledge Graph²⁶ and Diffbot.

The **GDELT** project has been providing an integrated event stream for media news events since 2013, and it has evolved into a comprehensive event KG. It separates events and associated entities such as individuals, organisations, locations, emotions, themes, and event counts into a continuously updating KG. The GDELT 1.0 Global Knowledge Graph, initiated on April 1, 2013, consisted of two data streams – one encoding the complete KG and the other focusing on counts of predefined categories (e. g. protester numbers, casualties). GDELT 2.0's Global Knowledge Graph (GKG)²⁷ enhances this with additional features, incorporates 65 translated languages, and updates every 15 minutes. Notably, mappings of GDELT into RDF stream were proposed, yet it is limited to only the event graphs and the GKG [235, 236].

As for *queryability*, GDELT can be accessed via Google's BigQuery²⁸ in its current state [235], updated every 15 minutes in real-time with temporal information available at the event level at different granularities, with a fixed schema.

Being updated in an automated manner from news sources, this stream KG is not in the same sense *collaboratively* evolving as Wikidata or DBpedia, in the sense of individual users contributing changes by their edits, but rather from curated news sources. While, to some extent, these sources could also be interpreted as "collaborative" agents contributing to the KG on the one hand, on the other hand, the act of changes has not collaborative nature in the sense that one of these actors could overwrite or undo others' additions.

Similar to GDELT, Diffbot offers a commercially available Knowledge Graph²⁹ that combines dynamic event data with information about products, events, and organisations. This Knowledge Graph is only available as a commercial service, wherefore we do not discuss it here in more detail.

3.2 Monitoring Trends

The LOD cloud can be seen as a network of open interconnected KGs, the most prominent of which are Wikidata, DBpedia, DBLP and YAGO. As such, a key part of its evolution has been the open community's continuous maintenance of these KGs. Indeed, their growth has been central to the expansion of the LOD cloud from $\approx 6.7B$ triples and 90 RDF datasets [20], in 2009, to ≈ 28 B triples and more than 1,200 datasets [177], by 2020.

With the growth of the LOD cloud comes the desire to analyse its temporal changes and track trends and evolution. Below, we first discuss approaches to analyse at the instance-level the changes in the LOD cloud. Subsequently, we take the perspective of the schema-level and consider methods and works analysing the changes of the LOD cloud in terms of the vocabulary.

 $^{^{26}\,{\}tt https://blog.gdeltproject.org/gdelt-global-knowledge-graph/$

²⁷ https://www.gdeltproject.org/data.html

²⁸ https://console.cloud.google.com/marketplace/product/the-gdelt-project/gdelt-2-events

²⁹https://www.diffbot.com/products/knowledge-graph/

11:18 How Does Knowledge Evolve in Open Knowledge Graphs?

3.2.1 Instance-level Monitoring

Several works have sought to capture and understand the nature of KG evolution. One such seminal initiative is DyLDO (see Section 3.1), which has been monitoring Linked Data on the Web since 2012, by collecting continuous LOD snapshots and examining them in terms of their document-level and RDF-level dynamics. The original paper [139] is based on the analysis of 86,696 Linked Data documents for 29 weeks and reveals that $\approx 62\%$ of the documents available during that time were, in fact, unchanged. In the remaining, the changes occurred mainly very infrequently, $\approx 23\%$, or very frequently, $\approx 8\%$, with very few documents reporting changes in between. The same polarising trend is recorded for very static domains, $\approx 44\%$, change very infrequently at the level of object literals, while schema changes (involving predicates and rdf:type values) were very infrequent, often related to time stamps, and very rarely involved the creation of fresh links.

Analyses of the DyLDO dataset include the work of Nishioka and Scherp [166] who applied time-series clustering over the temporal changes of the DyLDO snapshots and determined the most likely periodicities of the changes using an algorithm from Elfeky et al. [75]. This resulted in the finding of patterns in the evolution of the graph data. Although 78% of the first three considered years of DyLDO snapshots do not change at all, the remaining nodes could be organised into seven clusters of various sizes and periodicity. The latter ranges from periodicity prediction every week to once every half a year or year. Information-theoretic analyses have also been applied to analyse pairwise changes in graph snapshots of the DyLDO dataset [167]. Time-series clustering allowed us to organise the evolution into segments of similar behaviour. The study reveals that nodes of the same type show a similar evolution, even if these nodes are defined in different pay-level domains, i. e., different organisations. Finally, Gottron and Gottron analysed the same dataset but applied perplexity to explain the evolution of graph data [98].

At the level of the individual LOD cloud KGs, Wikidata is an especially interesting example of an evolving KG, having 90M entities and 1.4B revisions by more than 20K users.³⁰ The recent Wikidated 1.0 dataset [208] records the fine-grained organic evolution of Wikidata from its inception in 2012 until June 2021. The statistical characteristics of Wikidated 1.0 reveal a linear growth in the number of entities, which has been slightly accentuated after the Freebase integration in 2015. Also, almost all entities have less than 100 revisions, with half having less than 10. In terms of revision speed, the analysis highlights that most entities are edited frequently. Specifically, 60% of the revisions of a given entity occurred less than a month after a previous revision of the same entity. Inspecting the types of revisions, the paper indicates that most revisions consist of atomic changes, with approximately 90% containing less than 10 triple additions; moreover, 80% of revisions do not feature triple deletions. Another interesting trend indicates that half of the triples are added less than a day after the creation of their entity, while deletions take much longer, with over half involving triples that are deleted more than 6 months after they have been added. Although the vast majority of Wikidata triples are never deleted, \approx 10% are deleted only once and less than $\approx 1\%$ are deleted repeatedly after being added again. The CorHist dataset [224] is also built from Wikidata's edit histories, although with a focus on constraint violations and their corrections. The study shows that users are more likely to accept corrections for familiar constraints and certain types of constraints favour over-represented entities. highlighting the impact of biases. The evolution of Wikidata has also been studied in terms of editor engagement [207] and impact [191], as well as the quality of provenance information [188].

³⁰ According to https://www.wikidata.org/wiki/Wikidata:Statistics.

The work in [169] analyses the changes in Wikidata KG from a topological perspective. As such, it establishes that the evolution of the number of nodes and edges resembles a power law [147], similar to those commonly observed in social network graphs; based on this, it proposes classifiers that verify whether changes are correct.

Levels of Granularity. Alloatti et al. [10] propose to analyse KG evolution trends by capturing their changes across different snapshots at three levels of granularity: *atomic* focuses on operations at the resource level, *local* targets the evolution of a resource within its community, and *global* detects communities at the level of the entire graph. At the level of atomic evolutions, given a set of atomic updates performed between two snapshots, the authors distinguish between statistical changes, quantifiable in terms of the mean and variance with respect to a normal distribution, and so-called *noteworthy* ones, which capture snapshot features that diverge from the expected KG evolution with respect to a given threshold that is dataset-specific. An example of the former type would be quantifying the number of citations of a paper, while an exceptionally high number of new citations would illustrate the latter. Local evolution would also account for community-level features, such as graph density. As such, a publication may be noteworthy only at the level of its community, and communities themselves may be identified as noteworthy based on specific features, such as topological ones. At the global level, community detection methods can provide insights into the general behaviour of the different entities in the KG. When considering KGs as multi-community networks, various detection algorithms can be applied using custom network metrics, as reviewed in [193, 87]. When it comes to investigating KG evolution at a global level, studies have applied metrics transferred from different disciplines, such as databases [70], information theory [167, 98], web data crawling [68] and machine learning [168, 169].

Future Directions

Even with the large number of analyses already done in the past, there are many avenues to investigate further when it comes to monitoring, but especially analysing evolving KGs at instancelevel. One such direction involves exploring the commonality of data sources across different open KGs. For example, knowledge graphs like YAGO3 and Wikidata draw extensively from various language editions of Wikipedia. Investigating the extent of shared data sources and how this commonality has evolved can provide valuable insights into the collaborative dynamics of KG development. By understanding the overlaps and changes in data sources, researchers can gain a more comprehensive understanding of how this influences evolution; for example, an investigation of link evolution and cross-references between KGs over time could deliver new insights here.

Another compelling area for analysis pertains to the role of programmatic intervention in the development of knowledge bases. Many knowledge graphs, including YAGO and DBpedia, rely on automated processes for data extraction and transformation, including, in the case of YAGO, statistical learning. Likewise, Wikidata's data generation, while predominantly carried out by its users, also relies partially on programs that extract information from external sources through bots. Delving into the balance between manual curation and automated data extraction and its impact on KG growth and quality can offer valuable insights into the mechanisms that drive their evolution.

These future directions in KG analysis provide exciting opportunities to deepen our understanding of how these structures evolve, the factors influencing their development, and their crucial role in the dissemination of structured knowledge. Addressing these challenges will contribute to the ongoing advancement of knowledge representation and dissemination in the digital age.

11:20 How Does Knowledge Evolve in Open Knowledge Graphs?

3.2.2 Schema-level Monitoring

All the aforementioned studies of the evolution of Web graphs focused on the instance-level of the graph *data*, i. e., the nodes modelling the entities in the domain. Only a few works also considered **analysing the evolution of the** *schema-level* of the graph. An early study by Dividino et al. [70] shows that indeed, the schema of a node changes over time when one considers how the available RDF properties and RDF types are combined to a set of edge labels and node types to model a node. We call this set of properties and types the *schematic structure* of a node. Over one year in the DyLDO dataset, the authors analysed the schema structures of the nodes in terms of both the outgoing properties as well as types. They found that in each snapshot between 20% and 90% of the schema structures change from one version to the next. This means that more or fewer nodes have the same schema structures, nodes with new schema structures are observed, and some schema structures are not used anymore. There are also some combinations of properties and types where the schema structure of the nodes is very stable, i. e. the set of nodes with that specific schema structure did not change for one year [166, 70].

Just like new data nodes appear and change in the Web graph, the vocabularies used to model such data also change, but at a much slower speed. New vocabulary terms are coined to cover additional requirements or reflect changes in the domain. Other existing terms are modified or even deprecated. Previous work analysed the amount and frequency of changes in vocabularies based on different snapshots of the Billion Triples Challenge, DyLDO and Wikidata datasets [1]. Although the evolution of vocabularies is slow [1, 140], i. e., they happen on average a few changes every year only, a change may still have a significant impact due to the large amount of distributed graph data on the Web.

Another insight is that, in the course of an evolving vocabulary, the update of new terms from released vocabulary versions varies greatly and ranges from a few days to years. It is not surprising that even deprecated terms are still used by data publishers. Moreover, it is important to analyse both the change in the vocabulary, as well as how the various terms are used in combination. This can be seen at the schema-level: one can observe changes in the node and property shapes (e. g. SHACL shapes), as well as in their prevalence. For example, a recent study [196] compared the property shapes extracted from two Wikidata snapshots (one from 2015 and one from 2021). The analysis reported that the number of RDF classes increased from 13K to 82K and the number of predicates from 4,906 to 9,017, while the number of distinct property shapes increased from 202K to more than 2M. This calls for an in-depth study of how the different elements of the vocabulary evolve, not only in isolation but also together at the schema-level.

Finally, similar to the LOD Cloud showing the dependencies of different Web graph datasets, one may also consider the Network of Linked Vocabularies (NeLO) where the nodes are the vocabularies and the edges model vocabulary reuse [2]. Vocabulary reuse is generally encouraged, as it improves the interoperability of data, but at the same time, it also introduces dependencies between vocabularies that are to be resolved when vocabulary terms in the network change, are deprecated, or deleted. The NeLO network has been analysed over a history of 17 years based on the data from the Linked Open Vocabulary (LOV) service³¹ with respect to standard network metrics, such as size, density, degree and importance [2]. LOV collects the temporal information from hundreds of RDF vocabularies added to the service through a review-based process. The evolution of this schema-level graph has been analysed with respect to the impact of vocabulary term changes, term reuse and vocabulary importance [1, 2].

³¹ https://lov.linkeddata.es/dataset/lov/

Future Directions

Exploring the schema-level dynamics of open KGs reveals several promising avenues for future research and analysis. These areas of inquiry offer valuable insights into the evolving nature of knowledge graphs and their impact on knowledge representation.

One important aspect of KG analysis pertains to understanding how schemas are structured and evolve within graphs, but also how re-use between graphs evolves. Many open KGs, including Wikidata and DBpedia, make use of RDFS and OWL to organise their ontologies. However, the specific integration of schemas into the data varies. For instance, some graphs incorporate their ontologies directly into the data, while others maintain separate ontology files. Investigating the consequences of these schema design choices on knowledge graph evolution is another possible research direction. Additionally, assessing how expressive power and intended meaning in these schemas evolve and potentially influence KG development is of strong interest.

KGs exhibit varying degrees of semantic underpinnings, ranging from basic RDFS to more complex representations like OWL. Some, like Wikidata, may have intricate intended meanings and collaboratively evolving schema constructs that go beyond OWL's expressivity, which may necessitate advanced logics for interpretation (for instance the constantly evolving set of Wikidata's property constraints). Analysing the gap between intended, implied and supported semantics in KGs and its implications for their evolution is a further promising area of investigation. Overall debates within the Semantic Web and Knowledge Graph communities, about additional complex ontology features and the evolution of ontology languages as such, may also raise questions about the role of evolving ontology expressiveness in shaping knowledge graph structures over time.

Comparing the rates of schema/ontology evolution vs instance/data evolution in different knowledge graphs in depth is another potential future direction: preliminary observations may suggest that in some cases, the evolution of ontology structures lags behind changes in the data. Such temporal misalignment raises questions about how it affects the overall coherence and semantics of knowledge graphs over time; as a concrete example, let us again name constraints in Wikidata, which partially become outdated (and even explicitly deprecated) by their actual use – which could indeed be understood as a form of semantic drift.

Comparative analyses between knowledge graphs, especially those with similar characteristics or shared data sources, can provide valuable insights into ontology evolution, schema design and knowledge representation choices. By examining similarities and differences in their evolution processes, researchers can identify best practices and challenges in crowd-sourced ontology development.

These future directions in schema-level analysis offer opportunities to gain a deeper understanding of how knowledge graphs evolve structurally and semantically. By addressing these challenges, researchers can contribute to advancing our knowledge of knowledge representation dynamics and the evolving landscape of open KGs.

4 Study the Evolution

In this section, we discuss methods for studying the evolution of KGs. First, we introduce some relevant static graphs and KG metrics, as they have been defined to inform KG quality and are sometimes used to analyse KG evolution. Second, we address measures that concern consistency and quality specifically using constraints, as opposed to the simple metrics introduced first. In the third part, we discuss measures specifically developed to capture and quantify evolution, and we finish this section with a focus on how network science approaches could be used in the future for the study of KG evolution.

11:22 How Does Knowledge Evolve in Open Knowledge Graphs?

4.1 Basic Graph and Knowledge Graph Metrics

This section introduces metrics designed initially to study the properties of graphs and specifically knowledge graphs, which have been used to assess ontology quality [11, 142, 91, 37, 213, 227, 205] and that has also been used to study KG evolution [250, 252, 73, 71, 172]. Table 3 summarises such metrics, which – however – do not take an evolving KG as input for their calculation as they consider only one graph at a time. We can broadly group these static metrics into two groups: graph metrics and knowledge graph metrics.

Graph metrics are applied to a graph version of the KG or adapted to work on the KG. Examples of these metrics include *average depth* [71, 73, 91, 142], *number of paths* [142], *tangledness* [11, 91, 142] and *absolute leaf cardinality* [11, 91, 142]. In the work of Alm et al. [11], Gangemi et al. [91] and Lantow et al. [142], the metrics are applied only to the isA graph, whereas Djedidi et al. [71] apply the average depth on the OWL graph, the same as Duque-Ramos et al. [73].

Knowledge Graph metrics can be distinguished from graph metrics based on the idea of taking semantics into account. However, each approach, metric or paper specifies what type of semantics (RDF, RDFS, OWL or other) are considered and if the metrics are applied to materialised KGs or not. We do not make this specification here but leave it up to the interested reader to follow the cited sources. While instance-level analyses focus on the data graph, schemalevel analyses focus on the semantic information [33]. Therefore, we divide the metrics into three groups:

- Primitive metrics focus on a single aspect of the KG; for instance, they are used to characterise the *number of entities* of a KG [37, 142]
- Schema metrics focus on the schema or T-Box of the KG. Examples of such metrics include Property Class Ratio [250, 252, 172, 73], Depth of Inheritance Tree [250, 172, 73] and Inheritance Richness [71, 73]. For example, most of these metrics are used in the OQuaRE quality assessment by Duque-Ramos [73] to inform about varying quality (sub-)characteristics.
- **Data metrics** or A-Box metrics mostly combine an aspect of the A-Box with one from the T-Box. Examples of such metrics include *Average Population* [73] and *Instance Comprehension* [71]. Due to their simplicity, data metrics give only a partial view of KG quality and often need to be contextualised for a complete evaluation [73].

In summary, KGs have been analysed by calculating static metrics like the ones in Table 3 on linear/nonlinear series of consecutive snapshots: by combining these measures over some time, as done for instance in [73, 33, 182, 71, 172], one obtains time series data (a versioned or dynamic KG) that allows (and is currently primarily used) for calculating descriptive statistics (e.g. central tendencies, dispersion, distribution) that partially describe the KG evolution over time.

Future Directions

While static metrics can provide valuable insights at little cost, we argue that designing specific metrics and combining those with more sophisticated time-series analyses can lead to more precise monitoring of KG evolution. In particular – for any of the above-mentioned static metrics – investigating time-series trends in *metrics variations* such as seasonality or stationarity or even more complex models [214] can provide further insights about the KG evolution. We illustrate some ideas for such future metrics by the example questions listed below:

- **Trends:** How has the average degree of nodes or centrality developed in KGs such as Wikidata over the past N years? How interconnected is the KG becoming over time?
- Seasonality: Are there recurring periods of increased or decreased growth in the size (number of nodes or edges)? Is there any correlation with specific events?

Table 3 Overview of general graph metrics and specific Knowledge Graph metrics from the literature: metrics are only included if there are at least 3 papers (graph metrics) using and defining a measure (excluded 98 metrics). We excluded some of the graph metrics cited by the same three papers (5 metrics); for knowledge graph data metrics we also included those with 2 citations – any of these static metrics and changes would seem worthwhile to be also investigated in a longitudinal manner over time.

		Metric	Description	Used/Defined in
		Absolute depth	$d_a =$ sum over the cardinality of each path in a set of paths in graph	IsA graph [11, 91, 142, 250]
		Average depth	$d_a / paths $	IsA graph [11, 91, 142, 250], graph [37, 71], OWL schema [73, 71]
Graph		Maximal depth	longest path	IsA graph [11, 91, 142], graph [37]
5 U		Number of paths	paths	DAG [142, 250, 252]
-		Tengledness	$\frac{n_G}{t}$, n_G = cardinality of G, t = cardinality of the set of nodes with more than one ingoing IsA arc in G	IsA graph [11, 91, 142]
		Degree Distribution	mean-square deviation of the de- gree of graph nodes	graph [37, 67, 142]
		Entities	number of entities, classes and instances	graph [37, 142], IsA graph [91], OWL [213], DAG [250]
	ives	Properties	number of unique properties or relations	OWL schema [172], OWL [227, 229, 231], DAG [252]
	Primitives	Classes	C = number of classes (concepts)	OWL [227, 231], DAG [250, 252]
	Ъ.	Instances	I = number of individuals	OWL [227, 229, 231]
raph		Object properties	P_o = number of object proper- ties (non-inheritancE)	Schema [142], OWL [213, 229]
ge G	ma	Depth of Inheritance Tree		Tree [172], OWL [73, 205, 227], DAG [250]
led	che	Property Class Ratio	$\frac{ P }{ C }_{ H }$, H = inheritance relations	OWL [229, 172, 73], DAG [250]
Knowledge Graph	T-Box/Schema	Inheritance Richness	$\frac{ H }{ C }, H =$ inheritance relations	OWL [71, 73, 205, 227], Schema [142]
<u> </u>	Г. В	Attribute Richness	$\frac{ P_d }{ C }, P_d = \text{datatype properties}$	OWL [71, 227], Schema [142]
		Class Property Ratio	$\overline{ P }$	Onto [11, 91, 142]
	ata	Average Population	$\frac{ I }{ C }$	OWL [73, 227], Onto [91]
	x/D:	Cohesion	number of connected compon- ents	OWL [71, 227]
	A-Box/Data	Average Class Connectiv- ity	mean((c1, p, c2) where c1 and c2 are instances of classes	OWL [205, 227]
	A-B	0		OWL [205, 227]

- *Moving Averages*: How does the moving average of additions (new triples) or deletions (removed triples) over 12 months compare to the monthly new triples values? Are there evolutionary anomalies?
- Autocorrelation: Is there autocorrelation in the time series data of a given ratio metric (e.g. Property Class ratio, etc.) in the KG?
- *Stationarity*: Do structural changes in the KG (for instance, lengths of certain paths or other structural metrics) follow a stationary process?

11:24 How Does Knowledge Evolve in Open Knowledge Graphs?

So far, time series analyses with static metrics for LOD characterisation have been traditionally restricted to descriptive statistics, e.g. in [129, 182, 73]. We argue that this is an opportunity for the Semantic Web and Knowledge Graph research community to rethink more sophisticated metrics designed to precisely measure KG dynamics and change overall and in a modular fashion (e.g. instance data vs. schema dynamics, etc.). Likewise, we see a lack of tools and calculation frameworks geared specifically towards running such more complex time series analytics on evolving KGs at scale.

4.2 Consistency-Based Quality Metrics

Assessing data quality within a KG presents significant challenges that worsen if the aim extends to monitoring, ensuring, or improving such quality over time. Consistency-based quality metrics play a crucial role in assessing many dimensions of data quality, for example, measuring the integrity, coherence and general consistency of KGs [245]. Paulheim and Gangemi [176] estimated inconsistency in DBpedia by clustering conflicting statements; they limit their evaluation to a given snapshot, neglecting the evolution of these inconsistencies.

Various languages have been developed to express and represent constraints in KGs, yet not all are equally suited to "measure" consistency and quality. That is, while formal ontology languages such as OWL [101] and the respective underlying Description Logics [21] allow one to determine inconsistency of the whole KG, typically, due to their expressivity, they suffer from ambiguity between pinpointing and counting violations. Earlier work has used rule-based fragments of OWL, OWL RL to – again statically – quantify and repair inconsistencies [113].

More recent specific standards for KG constraint languages have revived the research on quantifying constraint violations. Specifically, the relatively new W3C standard SHACL [135], and similarly ShEx [195], allows validation and counting violations in a KG, w.r.t. a set of (integrity) constraints and target node/edge definitions. Yet, we only see both formal ontology languages such as OWL, e.g. [95], and these novel constraint languages being only slowly, if ever, adopted in (openly available) KGs.

In the following, we dive deeper into the measurability of quality metrics, focusing on consistency. Consistency metrics evaluate the coherence and absence of contradictions within a KG. Constraints can be used to specify rules regarding relationships between entities, ensuring that the graph remains internally consistent. Inconsistencies, such as conflicting assertions or logical contradictions, can be identified with these metrics. There is a trade-off between measuring consistency and simply measuring missing information. However, this trade-off will be explored as part of defining assessment frameworks.

As a first approach towards monitoring consistency w.r.t. constraints over time, Wikidata has leveraged constraint modelling to enhance data quality and usability. Within the Wikidata ecosystem, the Schemas project³² uses ShEx to define schemas for modelling various Wikidata classes. Additionally, Wikidata uses its own representation model to define constraints on its properties, known as Wikidata property constraints.³³ These property constraints serve as valuable guidelines for the community of users, aiding in maintaining data integrity and the development of violations is documented over time in Wikidata's own published database reports.³⁴ In a recent work, Ferranti et al. [86] have attempted to formalise the respective constraints in SHACL and SPARQL, in order to enable generating such violation reports in a standardised manner, on the fly, which may be viewed as a starting point to enable monitoring constraint violation over time.

³² https://www.wikidata.org/wiki/Wikidata:WikiProject_Schemas

³³ https://www.wikidata.org/wiki/Help:Property_constraints_portal

³⁴ https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/Summary

An alternative approach to quantify violations is to attach the number of violations (n_{C_i}) for each violated denial constraint (C_i) to nodes and edges in the KG. The counting can be done in a bag or set semantics by considering the duplicates in the constraint violations or not. Provenance polynomials can be built by summing the monomials given by $C_i n_{C_i}$. The obtained polynomials and corresponding degrees of quality can be leveraged during query evaluation to characterise the quality of the query results further. Although this approach has been conceived for static relational data [119, 120], the temporal aspects of inconsistency are still largely unexplored.

Despite these starting points, the question of how to measure and monitor quality in terms of consistency in a systematic manner for particular KGs over time seems to be still an open question that opens up engaging scenarios. For example, the presence of time in evolving KGs adds a dynamic perspective to constraint enforcement, facilitating ongoing improvements in the KG through data repairs, as proposed by [57]. Moreover, the analysis of constraints over time can also provide significant insights into the occurrence of semantic drift (see Section 2.3) within the schema layer of a KG. When historical constraint definitions are compared with the current state, it becomes possible to identify schema modifications, shifts in the focus of the schema layer and potential mismatches between the evolving semantics and the intended scope.

Future Directions

As outlined above, consistency is a big factor when assessing the quality of KGs. Hence, we see several potential directions of analyses in the future using constraints to learn more about knowledge evolution concerning quality. For example, before even analysing evolution, an investigation into which KGs use RDFS, SHACL and ShEX but also how expressive their ontologies are and which are entirely based on external data sources. Such questions directly tie into an investigation of quality based on consistency and constraints and how these evolve. First, measures and frameworks must be developed to support these kinds of investigations as they require handling KGs at scale. At the same time, the tradeoff between measuring quality and consistency vs. measuring missing information must be considered in greater detail before applying such approaches to any open general-purpose KGs, as these KGs operate with an open-world assumption.

The analysis directions align well with the dimensions of evolution (dynamics, timeliness and monotonicity), but each requires different approaches or solutions. Thus, we urge the community to use constraint-based metrics to analyse the consistency of the evolution of KGs, the change (trends, seasonality, etc.) of completeness, data freshness, data recency and temporal completeness. Precisely, the last three need to regard time as data rather than meta-data.

4.3 Methods for Quantifying Evolution

In this section, we want to give space to metrics specifically introduced to capture the evolution of a KG, which require pairs of (consecutive) graphs as input in the form of a versioned or dynamic KG, according to the classification presented in Figure 1. Most of the works introduced below study the changes between two (consecutive) versions of a graph, that is, two snapshots, such as [69, 177, 182, 181], making them specifically applicable to versioned KGs rather than dynamic ones. Pernisch et al. [182] propose several metrics to capture evolution on the materialisation and also provide their implementation in a Protégé plugin [183]. The evolution metrics capture the amount of change between two snapshots using simple counts of deltas between the snapshots. Pelgrin et al. [177] developed a framework to analyse various properties of versioned KGs based on changesets computed over pairwise versions of DBpedia, YAGO and Wikidata. Their framework consists of multiple evolution metrics such as growth rate and dynamicity. The authors also measure high-level changes, such as the number of entities changed between a pair of versions,

11:26 How Does Knowledge Evolve in Open Knowledge Graphs?

using the metrics we discussed in Section 4.1, but relating them directly to the evolution. The metrics capture the changes between a pair of snapshots but do not directly reflect KG evolution over multiple snapshots, i.e. a sequence of snapshots. Instead, pairwise comparison sequences can be considered to identify trends in evolution. Lastly, Dividino et al. [69] developed a monotonic measure for KG evolution that aggregates the amount of data changes over a sequence of snapshots. This results in a function measuring the evolution of the graph by approximating the actual evolution with an aggregation of absolute infinitesimal changes. When a KG evolves, such as Wikidata, most of the additions and deletions may be valid changes reflecting the nature of the entities modelled. However, collaborative KGs can also receive erroneous changes, be it due to vandalism or carelessness. Evolution information is exploited to assess which changes in a KG are correct [169]. Based on the features for Web data caching [168], several triple features are employed on the subject, predicate and object URIs, including additional information about the age and last edit. Notably, this improvement is achieved by purely employing information about KG evolution and not requiring historical information about the editors who perform changes on the collaborative graph.

Future Directions

As is evident from the studies mentioned above, there are not many metrics specifically developed for the study of KG evolution. This, we identify as a research gap as it is necessary to introduce measures capturing different dimensions and aspects of KG evolution. Following the examples above, measures need to capture the different aspects of evolution while at the same time being outlier-resistant. Approaches from time series analysis can be fruitful to kick-start this future direction and enable the further development of methods and metrics to study KG evolution. In the future, it is important to move from snapshot analysis to more continuous approaches capturing fine-grained evolution at the time of individual edits. We can also potentially borrow approaches from network science, as they also analyse the evolution of networks, even though the networks have a simpler representation than KGs.

4.4 Metrics and Methods from Network Science

Network science has developed tools to map and analyse complex systems, suggesting the possibility of adopting them to study the structural properties of KGs. Researchers have discovered that regularities in domains such as transportation systems, scientific communities, economic sectors, or communication systems can be fruitfully represented and studied as networks. Indeed, there are remarkable regularities in such domains that play an important role in how these systems function and evolve. For example, networks tend to have very heterogeneous degree distributions, which means there are "hub" nodes with orders of magnitude more connectivity than the typical node [7]. Social networks tend to have many triangles, as suggested by the saying that a friend of a friend is likely to be a friend. Scientific community networks often have modular structures [87], reflecting coherent subcommunities of nodes in a larger system. Empirical networks tend to be sparse (i. e. given a network on n nodes, there are far fewer than the possible n(n-1)/2 edges). But they also have short paths connecting all pairs of nodes (i. e., low diameters) [247].

Although recent work on multiplex or multi-layer networks considers data with multiple kinds of objects or links between them, most networks studied are generally simpler than those observed in the Semantic Web community. For instance, ordinary networks usually consist of a homogeneous set of nodes (i. e. airports) and relationships between them (i. e. direct flights between airports). Multilayer networks consist of the same nodes and different kinds of relationships they might have. For example, people who may communicate via email and telephone. Studies using this

kind of multi-layer data tend rather to just generalise the methods applied to ordinary networks described in this section than to invent new ones [26, 25]. On the other hand, knowledge graphs are multi-dimensional by design. Although undoubtedly useful, such complexity presents an obstacle to studying their evolution using methods from network science. Therefore, to apply these methods to study the evolution of KGs, we must first simplify the data. However, any simplification must be driven by a substantive question to make it meaningful, and it must be significant in the sense that it discards a significant amount of data, to be tractable.

Once a simple network has been constructed, the temporal dimension of the data can be integrated by slicing data into time periods (for instance, as in [143]). Measures of the network, for instance, its diameter, the mean and variance of its degree distribution, the modularity of a community detection exercise, or the prevalence of clustering can be calculated for each slice and then plotted over time. However, the choice of the width of the time slice can have major implications for subsequent analyses [211].

The stylized facts about networks described above have important implications for things that happen to them or to them. They predict the robustness of a network, i.e. how well it holds when its nodes are removed. They predict how quickly things like information or diseases spread. Network structure plays an important role in its navigability: if you do not have a map of the network, can you still find your way from a node you know to another specific node in a reasonably short amount of steps [134, 215]? Network scientists are naturally interested in how changes in a network are captured by these measures and, in turn, how they influence things that happen within networks [165].

Network scientists have two broad solutions for the comparability issue between networks of different sizes. The first is to propose a generative model that captures many of the key properties of the network in question [38, 39], and to instantiate random graphs from this model. Next, one calculates the same statistics on this randomised version of the graph and uses it as a kind of benchmark or normalisation factor. The most simple generative model is the Erdős-Rényi model, in which edges are randomly added between nodes with a fixed probability p. Given two empirical networks of different sizes, one can create corresponding random networks with the same number of nodes and edges for each. Calculating the clustering on these random networks allows us to scale or normalise the clustering observed in the corresponding empirical networks, which then become more comparable. More sophisticated models like the Barabasi-Albert model [7] (which generates networks with heterogeneous degree distributions, i. e. hubs) and the Watts-Strogatz model [247] (which generates "small world networks" that have both short paths and high clustering) can also be used in this way, depending on the research question.

The second way to make network measures comparable between networks of different sizes and over time is to create randomised versions of empirical networks, sometimes called *null models* [128, 206]. Such randomisation typically takes place among the edges, which are randomly rewired or shuffled subject to constraints depending on context. For example, a randomization of links between Wikipedia editors and the articles they touch creates a "random" version of Wikipedia preserving editor activity counts and article edit counts. Such randomisations are similar to statistical Monte Carlo simulations and can be computationally intensive, but the resulting randomised versions of the empirical graph can provide a useful benchmark to compare against the original graph. Although these methods require both a drastic simplification of the data contained in KGs and the deployment of complicated methods such as generative models or null models, they present a significant opportunity to create more robust estimates of the dynamics of KGs. Given the degree of simplification this process requires, a clear research question about the structure and dynamics of KGs is an essential first step.

11:28 How Does Knowledge Evolve in Open Knowledge Graphs?

Future directions

We see the potential of using network science to investigate the collaborative nature of many open general-purpose knowledge graphs. Not only does knowledge evolve, but the way it evolves is intertwined with the editing network, for which network science and its approaches to analysing its changes over time would be beneficial. For example, if one wanted to study whether Wikidata editors were becoming more or less collaborative over time, how could one define a reasonable notion of collaborative behaviour? Could one define collaboration between two editors as a function of their using the same properties or working on the same entities? Should a pair of editors both using the most widely used property be as thickly connected as two editors using a more rarely used property? Network science offers tools to carry out such an analysis, but the researcher must make choices in pursuit of a question. Question-driven modelling of KGs as simplified "networks" can move us beyond a descriptive analysis of KG evolution.

5 Manage the Evolution

Although dynamic/versioned and temporal KGs can be considered as two alternative approaches, they introduce different challenges in their management. In the case of temporal KGs, the main challenges lie in how the temporal information is captured and represented. We discuss different approaches in Section 5.1. Although, when time is not part of the data, the KGs do not require specific data models. The temporal information lies in the updating process itself; they often publish complementary changelog streams that may or may not be represented in RDF. However, time as metadata raises a different set of challenges for KGs, including the representation of the evolution and storage options, discussed in Sections 5.1, 5.2, respectively.

5.1 Data Models for Temporal Knowledge Graphs

The two main approaches for implementing KGs are RDF and labelled property graphs (LPG). In the rest of this section, we describe how researchers and practitioners modelled temporal KGs in these two approaches. In the last part, we elaborate on open challenges with regard to capturing and then analysing the evolution of knowledge in Temporal KGs.

Temporality in RDF

The problem of how to model time-related information has been intensively studied. Amongst the multitude of proposed solutions, a broad distinction can be made by representing time in the data vs. in the metadata.

In the former case, entities can be part of statements together with their temporal properties. The Time Ontology and the Sensor, Observation, Sample, and Actuator (SOSA) ontology implements this idea, e.g. an observation can have a relation sosa:phenomenonTime with a time:TemporalEntity individual.

In the latter case, the temporal annotation applies to RDF statements (or graphs). A common method to implement it is reification, which involves annotating triples. In [109], various reification schemes were examined:

- Standard Reification uses a resource to represent a statement, such that it can be used in other RDF statements to add annotations (including temporal ones).
- N-ary Relations represent relationships using resources, stating subject involvement, value, and qualifiers. Instead of stating that a subject has a given value, it states that the subject is involved in a relationship that has a value and qualifiers.

- The *Singleton Properties* approach involves creating a property that is only used for a single statement. The resource representing the statement is annotated with this property to add more information.
- RDF 1.1 introduced the notion of *Named Graphs*, which can, for example, be serialised in N-Quads. One can annotate the named graphs, e.g. associating the same temporal annotation to all the statements contained in the graph.
- RDF-star [107] extends RDF through embedded triples, i.e., an RDF statement can be the subject or object in another RDF statement. Just as standard RDF can be queried via the SPARQL query language, RDF-star can be queried using SPARQL-star (formerly SPARQL*), allowing users to query both standard and nested triples.

There is no single way to represent contextual information in RDF graphs, and the different mechanisms have advantages and disadvantages. Reification and n-ary relationships model complex facts in RDF. However, adding reification triples for each reified triple increases the data volume, making metadata queries cumbersome due to the need for additional subexpressions to match the corresponding reification triples. Other methods, such as singleton properties and named graphs, reduce the number of extra triples. However, these approaches require verbose constructs in queries, introducing artefacts to associate triples with their metadata [171]. RDF-star is more compact and adds facilities to the query language via SPARQL-star but does not achieve the levels of flexibility as some previous approaches. Of the strategies presented, named graphs are the most flexible since they allow assigning one annotation to sets of statements; RDF-star is the least flexible option since it cannot capture different sets of contextual values on an edge [112].

Temporality in Labelled Property Graphs

Labelled Property Graphs (LPGs) are another popular solution to represent KGs. The problem of the representation of evolution, particularly temporality, has also been addressed in their context. However, while in principle LPGs allow direct attachment of temporal information to edges in the graph, there is no consensus on a single approach for temporal LPGs. Similar to core RDF approaches, works in LPG in the literature differ in supported time dimensions (valid time, transaction time or both/bitemporal), types of possible changes to graph structure and properties, and representation as either a series of graph snapshots or a single graph reflecting changes over time.

The Temporal Property Graph Model (TPGM) [201] extends the Extended Property Graph Model (EPGM) to support analytical operators on directed graphs that evolve in Gradoop. TPGM adds support for two different time dimensions, valid and transaction time, to differentiate between the evolution of the graph data concerning the application and managing the data. This approach offers a flexible representation of temporal graphs with bitemporal time semantics. TPGM expands EPGM with four new time attributes as mandatory for vertices, edges, and logical graphs: two for transaction time intervals and two for valid time intervals.

Debrouvier et al. [60] apply temporal database concepts to graph databases to model, store, and query temporal graphs for historical data tracking. The focus is on the *Interval-labelled Property Graphs* data model, which timestamps nodes, relationships, and node properties with temporal validity intervals, allowing for heterogeneous graphs with different types of relationships. This model enables richer queries and supports two path semantics: *Continuous Path Semantics* and *Consecutive Path Semantics*.

Andriamampianina et al. [12] propose a conceptual model to represent temporal property graphs and define a set of operators to perform queries on these. The model establishes various concepts to represent objects, their relationships, and their evolution over time. It manages time

11:30 How Does Knowledge Evolve in Open Knowledge Graphs?

through valid time intervals to track changes and occurrences in the real world. To describe an object, the model introduces the notion of *temporal entity*, comprising a set of states to represent different versions of the entity over time. Each state includes attributes, attribute values, and a valid time interval. A temporal relationship, analogous to a temporal entity, describes the link between two entity states.

Future Directions

Despite RDF and LPGs originating in different contexts, the two approaches are valid for creating and representing KGs. Several graph database vendors support both approaches to offer their customers flexibility and choice. In this context, an ongoing research direction lies in the *interoperability* between the approaches. Despite the active research [4, 15, 144], to the best of our knowledge, there is no study on the RDF-LGP interoperability in the context of temporal KGs. The challenge lies in the way the time can be represented in both RDF and LPGs: the multitude of different approaches leads to many possible conversion procedures. We argue that reference models are needed to unify the existing approaches and to set the basis for standardisation initiatives that will ease the creation, storage and processing of temporal knowledge graphs in different engines.

Another direction relates to query languages for temporal KGs. SPARQL and the LGP query languages consider temporal annotations as any other type of annotations. As such, query writers need to understand how time is represented in the graph and write the query accordingly. However, temporal annotations enable specific time-related operations, such as creating selection criteria based on Allen's relations [9]. Encoding such relations in the queries is not trivial and often errorprone. Treating time as a first-class citizen in the data models can lead to query languages with specific time-related operators, simplifying the query writing process and constructing dedicated query engines that can efficiently evaluate such operators. While this idea has been investigated in the context of continuous query processing over RDF streams (see section 6.2), it has not yet been deeply investigated for temporal knowledge graphs.

Interoperability between the two models would also further enable the possible application of analysis frameworks, existing and future ones. The same can also be said about SPARQL integrations, as in the past analyses have made use of SPARQL. Therefore, a SPARQL extension for temporality (of any dimensions) would further support efforts into KG evolution analysis.

5.2 Storage Methods

Since in temporal KGs the time dimension is managed as part of the data, temporal information integrates naturally in the data model and can therefore be captured using standard methods as outlined in Section 5.1. In the case of dynamic and versioned KGs (Figure 1), alternative approaches have been proposed capturing temporal information outside the data model itself.

An intuitive way of storing versioned KGs is to store each complete version of the KG as a new copy, often referred to as the *Independent Copies* approach [81]. While this can even be implemented using standard triple stores with named graphs, it has scalability issues regarding the number of named graphs (one for each version) and the required storage space for larger KGs. An advantage of this approach is that all queries to be executed on a single full version of a KG can be executed very efficiently since no additional computation (see below) is needed to retrieve the complete version of a graph to execute the query on. IC approaches are generally very useful for small knowledge graphs [177].

To reduce the storage overhead, *Change-Based* approaches store several full versions of the KG as snapshots but only sets of changes (deltas) for the versions in between. This makes them a hybrid solution between versioned and dynamic KGs. In this setup, querying versions

that correspond to snapshots is again very efficient since the full KG is readily available. The disadvantage of this approach is that for the versions between snapshots, chains of deltas have to be applied on the preceding snapshot to recreate full intermediate versions [222, 19, 5, 179, 178]. An important aspect is then to identify which versions to materialise as snapshots and which ones to capture as deltas.

Instead of capturing entire versions of complete KGs, dynamic KGs annotate individual triples with timestamps, so-called *Timestamp-Based* approaches. In such a setting, it is then of course expensive to recreate particular versions of a KG since this requires filtering all triples based on their temporal validity. On the other hand, it becomes efficient to look up the temporal validity of each triple.

Future Directions

While most systems implement only one of the above-mentioned storage methods [223], there are hybrid approaches that can be configured to resemble one or the other. In this sense, one direction of future work is to investigate how to exploit the strengths of different storage techniques for certain use cases and develop adaptive approaches that choose and adjust the storage layout based on how the data is used.

Building upon existing approaches for the above-mentioned storage models, one of the main challenges is scalability. On the one hand, we need to develop more efficient storage methods to reduce the storage overhead of capturing information about versions and temporal validity. On the other hand – and this is very much determined and influenced by how the data is stored – future work needs to develop efficient methods for querying that can not only retrieve complete versions of a KG but also allow efficient query processing over certain versions of a graph (see also Section 6.1).

Finally, it is worth noting that the way the data is stored affects the type of possible analyses on KG evolution. For example, if one wants to run time-series analyses (as described in Section 4.1), change-based approaches are ideal due to their focus on changes. Independent copies may not contain enough fine-grained information to perform such analysis. However, metrics based on consistency metrics (as described in Section 4.2) may not work in change-based approaches as some intermediate changes may affect the consistency of the KG. Therefore, we envision storage solutions able to store KGs following different approaches, with the ability to perform a wide range of analytics tasks on KG evolution in efficient ways.

5.3 Mapping Schemas

Supporting KG versions is a key approach to ensure the stability of downstream applications for KGs. Therefore, it is essential to capture the evolution on the schema-level by sets of schema changes that typically occur in collaborative and decentralised processes.

Schema evolution requirements have been discussed in the past, in particular with respect to ontology evolution [28]. The availability of expressive and declarative mappings specifying the evolution between an original version of a schema S and an evolved version S' makes it possible to cater for the automatic propagation of the changes on the corresponding instances.

There exist two inherent problems with mappings between schemas. The first problem corresponds to the (semi-)automatic computation of the schema mappings by leveraging schema matchings and Diff(erence) computation [197]. Schema matchings can be defined as one-to-one correspondences between two different versions of a schema, and they can be coupled with a confidence value. On the other hand, schema mappings are declarative specifications, typically expressed in a subset of First-Order logic, representing the transformation between two different

11:32 How Does Knowledge Evolve in Open Knowledge Graphs?

versions of the underlying data. Schema mappings are typically expressed as source-to-target tuple generating dependencies (s-t tgds), whose left-hand side is a conjunctive query and right-hand side is a conjunctive query enhanced with existential variables, which lead to value creation. In the case of schema evolution, schema mappings are adapted after schema changes and meta modelling abstractions serve the need of providing high-level programming interfaces than other techniques [31].

The second problem concerning mappings between schemas is the so-called schema mapping or data exchange problem [78, 28], consisting of computing the transformed target instance (also called target solution) by applying the source-to-target tuple-generating dependencies between source and target schemas. In the case of schema evolution, the target schema might undergo some changes, thus entailing the propagation of these changes to both the mappings (s-t tgds) and the corresponding target solution.

The most expressive schemas for KGs are ontologies, which allow conceptualising a domain. They provide a steerable vocabulary for a given domain of interest, defining the ontology concepts as well as the properties and relationship between these concepts. Several research approaches study collaborative ontology evolution and ontology matching, as surveyed in [108, 77]. Without going into the details of these approaches, we point out that in the last decade after the above approaches, schemas for graphs have profoundly evolved, thus bringing more open challenges for KG mappings and transformations.

Finally, often KGs originate from external databases that can contain graph data in different formats or even other data models, such as relational or document databases. There are approaches, such as R2RML [55], to facilitate the latter, but the mappings from relational data to RDF have to be (manually) adapted whenever the native (or the integrated) schema changes.

Future Directions

Recent schemas for KGs range from RDFS [42], SHACL [135], and ShEX [22] to PG-Schema [14] and their evolution, as well as the mapping problems related to computation of schema mappings and computation of the target solution, are not yet studied. The first three schemas are applicable to mapping RDF data, while the latter is applicable to mapping property graphs [38]. One relevant future direction consists of studying the automatic generation of schema mapping transformations and the data exchange problem for the above models in a time-varying context thus exploring schema evolution and versioning for evolving knowledge graphs under recent schema languages.

Another important direction concerns the mappings from RDF to property graphs or the other way round [132, 15] in order to pay attention to producing incremental or comparable schemas in comparison to previous versions. Especially complex constructs have alternative translations into the other model. Hence, small changes can have big structural impacts on the integrated result. It therefore remains mostly unclear how to appropriately capture and measure schema evolution caused by schema changes in the input data. Although some proposals, such as the OneGraph vision [144], propose to achieve graph interoperability by allowing users to use Cypher or SPARQL independently from whether RDF or property graphs were chosen as the data model, this only means that users are free to choose the query language that they prefer or that is more appropriate for a different use case; the underlying challenges of how to capture evolution in the underlying graph model remain the same.

Typical tasks to process KGs include querying, reasoning, and machine learning. When we move from static to evolving KGs, one should consider the temporal dimension. In Sections 6.1 and 6.2, we discuss two classical operations on knowledge graphs: querying and reasoning. Next, we discuss learning techniques in Section 6.3. We conclude by discussing evolving KG exploration in Section 6.4.

6.1 Query Processing

We introduced data models for temporal KGs in Section 5.1. As the temporal information can be modelled in standard RDF (e.g. through named graphs or reification), in RDF-star and LPG, it follows that their relative query languages, such as SPARQL (or SPARQL-star), can be used to retrieve data from them. However, as we explain in Section 6.1, several researchers proposed ad-hoc query languages where time is a first-class citizen. Next, we discuss querying for versioned KGs in Section 6.1, focusing on the solutions to extract and query a specific KG version. Finally, we introduce continuous queries in Section 6.1 to monitor changes and to evaluate a query on evolving data continuously.

Temporal Querying

Temporal queries refer to languages and operators that offer native support for retrieving and manipulating time-referenced data. The semantics of a temporal query language are usually closely coupled to a temporal data model that defines the underlying data abstractions (see Section 5.1).

Despite the growing popularity of temporal data in KGs, this research area is still in its infancy. Exciting proposals (with a few exceptions) represent the graphs using either RDF or LPG and approaching change as a snapshot sequence. Thus, their query-answering capabilities are limited to those possible under the snapshot reproducibility principles, i.e. answering a temporal query over a database is equivalent to taking the union of all the answers obtained by evaluating the non-temporal variants of the query for each database state [35]. For example, τ -SPARQL [226], SPARQL^T [251] propose syntactic extension meant to access RDF triples annotated with a timestamp. Zhang et al. [253] went one step further with their proposal, SPARQL[t], extending the annotation with an interval-based validity time. Raising the expressivity bar, Arenas et al. [18] studied Temporal Regular Path Queries (TRPQ) to interrogate reachability over time over property graphs extended with time intervals of validity. Intervals of validity represent consecutive time points during which no change occurred for a node or an edge in terms of its existence or property values. Their approach, similar to T-GQL [61] and the Temporal Graph Algebra [161], is designed for Labelled Property Graphs. The main drawback of such a query model is the lack of support for operations that explicitly reference temporal information [18]. Therefore, an extension of this query model that propagates temporal information across snapshots has been proposed [66].

Querying Versions

Querying archives is not straightforward; since there is no well-defined or commonly accepted standard, archiving engines typically propose customised solutions for querying their data. AnQL [256] and SPARQL-T [92], for instance, are SPARQL extensions based on quad patterns – where the fourth component indicates the version over which the given query should be executed. T-SPARQL [100] instead is a SPARQL extension where groups of triple patterns are annotated with constraints regarding temporal validity supporting time ranges and timestamps. Other extensions go beyond the temporal dimension and include geospatial constraints [30, 185]. Some archiving engines [178, 179] also use the GRAPH clause of SPARQL to denote specific versions.

11:34 How Does Knowledge Evolve in Open Knowledge Graphs?

Apart from different approaches on how to formulate queries syntactically, one can distinguish different types of queries over archives based on the way they access the available versions of the knowledge graph [81, 177]. Two basic retrieval tasks are to extract a specific full version of a KG from storage (*Version Materialisation*) and to extract deltas (changesets) between pairs of versions (*Delta Materialisation*). In addition, we can distinguish different types of queries; the most commonly supported type of queries on evolving KGs are those where a SPARQL query is to be evaluated over a specified full version of the KG (*Single Version*). Another type of query aims at comparing answers to full SPARQL queries on different versions of a KG (*Cross Version*, e. g. which of the current countries was not in the original list of UN members. Instead of retrieving the answers to a SPARQL query, one can also aim to retrieve the specific versions in which a given SPARQL query yields (specific) results (*Version*), e. g. in which revisions did the USA and Cuba have a diplomatic relationship?

While the literature also introduces queries on deltas (single delta and cross delta), where queries can be evaluated on the changesets only, we argue that these types of queries can be considered subsumed by the above-mentioned types on full versions of a KG and assume that the archiving engine will detect during query optimization whether a complete version of the KG needs to be retrieved of a retrieving a changeset is sufficient.

Continuous Querying

Continuous queries (CQs), also known as standing queries, differ from other query processing tasks due to their never-ending nature. Indeed, they are typically used to analyse evolving data, including evolving KGs, to identify patterns, trends and outliers. With respect to the running example, one may write a query to monitor the movements of artworks between galleries. While the artwork is displayed in New York, the continuous query returns New York when specifically queried for the "current location". When the artwork is moved to Madrid and consequently the KG is updated, the query's result changes to Madrid as soon as the information changes.

The most relevant trait of CQs is the time-varying nature of the answers. Indeed, a query evaluated under continuous semantics produces a series of responses as if it was evaluated for every time instant. In practice, continuous-query evaluation is either periodic or based on custom conditions, e.g. the occurrence of an event or data change. Although several proposals exist for relational data [237], their potential in the Knowledge Graph world remains substantially unexpressed.

The Semantic Web literature has explored continuous queries for Streaming Linked Data [41] proposing several SPARQL extensions, e.g. C-SPARQL, CQELS, SPARQL_{stream}, including some able to combine different modalities [184]. Such languages have been reconciled by Dell'Aglio et al. [64], who explained their continuous query semantics using three families of operators adapted to RDF from [17]. RSP-QL describes how, despite syntactical differences, the existing languages all use window operators to cope with the infinite nature of the input data, usually modelled as a partially ordered sequence of timestamped RDF graphs. On a parallel line of research, EP-SPARQL [16], DOTR [155], and OBEP [233] have explored the approach for detecting event patterns in RDF streams. Such languages leverage time-aware operators and can be evaluated using regular expressions. Although the SPARQL query is entirely supported semantically, such proposals have given little attention to subgraph matching and navigational/exploratory continuous queries. Notably, queries involving (regular) path expressions that cover more than 99% of all recursive queries found in massive Wikidata query logs [40].

Regarding navigational continuous queries, Pacaci et al. [174, 175] modelled the graph as an ever-growing sequence of timestamped edges. Moreover, they studied two query models, Regular Path Queries (RQP) and Union of Conjunctive RPQs. Such query models are analysed with and without explicit deletions as a form of the materialised view.

A. Polleres et al.

Finally, continuous subgraph-matching (CSM) is a particular case of the foundational subgraphmatching problem, where the target graph is subject to updating (either append-only or with explicit deletions). Sun et al. [221] recently surveyed the existing exact approaches, modelling the CSM problem as incremental view maintenance.

Future Directions

Besides an investigation of which approaches have been applied to which general-purpose open KGs and how they perform, we distinguish two main directions for what concerns querying evolving knowledge graphs, i.e. addressing the open challenges related to each query model and a more general challenge that goes in the direction of a unified query model.

Temporal Querying for EKG has built upon the adoption of a single temporal model and snapshot reducibility. Future work requires relaxing such assumptions. The simultaneous application of multiple temporal models relates to the heterogeneous nature of graph data. Indeed, KGs are often referred to as a way to address data variety and perform data integration. However, such variety is not allowed within the temporal model, given an entailed complexity exposition. Going beyond the snapshot reducibility means allowing explicit temporal reference within the query settings. Such an approach reduces the temporal-navigational mismatch in the query language, allowing for posing complex questions over hybrid graph data models.

As explained above, querying versions of a KG often entails evaluating queries on a specific version of a KG or multiple ones. Naturally, the storage layout and available indexes determine how efficiently a query can be answered. Hence, developing appropriate indexing, storage layout, and efficient query optimisation techniques exploiting them are important aspects of future work.

The challenge related to continuous queries over EKGs relates to the central role of windowing in Streaming Linked Data, which poses serious limitations to the adoption and the optimisation of continuous queries. Users must know the temporal context of the interested phenomenon to choose an appropriate windowing policy. Moreover, aggregation-optimised windowing, which is well-known for relational data, was not studied for graphs. On the other hand, navigational continuous queries, and in general continuous subgraph matching, were little studied. Their relationship with knowledge evolution is noticeable and further investigation is required.

Finally, searching for a unifying query model that could make the best of the existing one is open and motivated by the specific need to migrate from one model to another when necessary. Currently, the users must pick one data and query model, and thus, their query ability is limited by the design choice of such languages. Instead, a formally verified language for EKG data that can express queries about time, through time, and in time is still missing.

6.2 Reasoning

Reasoning over large KGs layered with an OWL ontology to describe their schema may be prohibitive when using the full power of OWL. However, reasoning within the OWL 2 profiles [137] brings very interesting computational properties. Indeed, state-of-the-art reasoners over KGs typically focus on fragments of OWL (e.g. [164, 46, 238, 29]). For example, OWL 2 RL axioms can directly be translated into Datalog rules [162] enabling the use of efficient Datalog engines (e.g. [164]) that will expand the KG with implicit facts following from the OWL ontology and the KG data. Reasoning also enables the use of the notion of logical difference [136], which can be essential to understanding the evolution of a KG in terms of new entailed facts. For example, diff(KG, KG') represents the (entailed) facts in KG' not present in KG.

11:36 How Does Knowledge Evolve in Open Knowledge Graphs?

Reasoning in Evolving Knowledge Graphs

Rule-based systems typically perform materialisation (i.e. precomputation of the consequences after reasoning) before queries over the KG are evaluated. Changes in the KG require recomputing the materialisation so that query results are up-to-date concerning the changes. This process may be expensive for very large KGs and rule sets, especially if they constantly evolve. Most systems adopt Incremental materialisation when changes are to be reflected as soon as they occur (e.g. [241, 163]). These systems focus only on the part of the KG affected by the changes and implement optimised solutions to perform efficient incremental reasoning. In addition, there have been efforts in the literature to enhance incremental reasoning via modular materialisation (e.g. [114]) and enable distributed materialisation via data partitioning (e.g. [6]).

The evolution of a KG may also require the integration with other KGs as described in Section 5.3. The compatibility of integrating multiple KGs has been extensively evaluated from the ontology alignment perspective. In the literature, several approaches aim at identifying logical errors and unintended logical consequences derived from the alignment of the KGs (e.g. [157, 79, 216]). To the best of our knowledge, at the moment, no studies are focusing on how KG evolution affects consistency in alignment tasks. We believe that this is an important future direction because the effect of changes on reasoning can be substantial [182] and can also unexpectedly impact alignment tasks [183].

Reasoning for Studying Evolution

Logic-based reasoning, as discussed in previous sections, can play a key role in conducting constraint validation and can contribute to the definition of robust metrics to measure KG evolution. For example, the *semantic drift* described in Section 2.3 can be tackled via the logical difference [136] between two versions of a KG after materialisation. diff(KG_i, KG_{i+1}) represents the new (materialised) facts in KG_{i+1} not present in KG_i, while diff(KG_{i+1}, KG_i) represents the facts that were lost in the new version of the KG KG_{i+1}. An analysis of the impact of changes on the materialisation in the case of \mathcal{EL}^{++} ontologies in the biomedical domain was analysed in previous work [182], where the authors quantified the change in the materialisation to learn how ontologies evolve over time.

Efficient rule-based reasoning can also be leveraged to evaluate the evolution of the knowledge graph in terms of the conformance of the data with respect to the ontology and available constraints. This conformance evaluation can complement the related *quality metrics* (see Section 4.2). For example, Kharlamov et al. [131] interpreted some OWL 2 axioms involving cardinalities and ranges as integrity constraints and represented them as Datalog rules to identify violations of those constraints. For example, the following OWL axiom (α):

is transformed into the following Datalog rules:

$$Art_pieces_carried_out_by_artists(?x) \leftarrow carried_out_by(?x, ?p) \land Artist(?p)$$
(2)

 $Violation(?p, \alpha) \leftarrow MasterPiece(?p) \land not Art_piece_carried_out_by_artists(?p)$ (3)

In the example above, it is expected that *MasterPiece* in the KG have at least an explicitly associated *Artist*.

A. Polleres et al.

Stream Reasoning

When the KGs evolve at a high pace, and the information needs to focus on extracting novel and recent information, we enter the realm of stream reasoning [62]. Stream reasoning combines knowledge representation with stream processing techniques [52] to process evolving ontologies and KGs in a continuous and responsive fashion[64]. Stream reasoning cases relate to Timeliness (Section 2.1), i.e. the inference is needed before data are no longer useful.

Firstly, several research groups worked on defining data models and vocabularies to capture data streams through KGs and ontologies. Zhang and Stuckenschmidt [115] introduce the notion of linear version space to define a sequence of ontologies. Such a notion was later adapted by Ren and Pan [199] to define ontology streams as a sequence of timestamped ontologies. An alternative model for data streams is RDF streams, defined as a sequence of timestamped statements (as in [24, 186]) or graphs (as in [63]).

Reasoning task extensions over streams, such as consistency check and closure, were first studied with a focus on adapting reasoning algorithms to the streaming settings. For example, Barbieri et al. [24] extend the incremental reasoning algorithms DReD for stream reasoning with sliding windows. The authors exploit the knowledge derived from the sliding window operator to calculate when assertions must be deleted and use such information to improve the performance of the materialisation algorithm. Ren and Pan [199] propose a truth maintenance system implemented in the TrOWL reasoner that builds a graph to track the derivations. When the assertion changes, the system incrementally maintains the graph and consequently updates the materialisation.

Over time, the focus moved to the application of temporal logic for stream reasoning: here, Beck et al. proposed the Logic-based framework for Analysing Reasoning over Streams (LARS) [27]. LARS combines temporal logic operators with specific operators to reason over streams, such as the window operator. Tiger and Heintz [230] propose P-MTL, an extension of the Metric Temporal Logic with probabilities to model the state uncertainty. P-MTL allows the use of probabilities in the logic formulas and to use them in the inference process. One of the most recent studies is from Walega et al. [243], who researched DatalogMTL in the context of stream reasoning. They study the conditions to guarantee that no infinite materialisation occurs and show that reasoning over the fragment of DatalogMTL that satisfies such conditions is not more complicated than reasoning over Datalog, i.e. ExpTime-complete for combined complexity.

Lastly, several researchers and practitioners studied stream reasoning applications. One area where stream reasoning found considerable interest is smart cities and traffic management. Lecue et al. propose STAR-CITY [152], a system to analyse streaming heterogeneous data by combining ontological reasoning, rule-based reasoning, and machine learning. Eiter et al. [74] designed a stream reasoning solution based on Answer Set Programming (ASP) to optimise traffic control systems. Le Phuoc, Eiter, and Le-Tuan [187] use stream reasoning to integrate streams of images from car cameras and data streams to reason over them.

Stream reasoning has also found application in other domains. For example, Barbieri et al. [23] applied stream reasoning techniques to social media streams for personalised recommendations; Kharlamov et al. [130] propose stream reasoning in the context of monitoring failures in an industrial setting; De Leng and Heintz [59] integrated stream reasoning techniques in the Robot Operating System (ROS) to reason on the input IoT data and determine the most appropriate configuration. A recent survey discusses the maturity level of knowledge representation and reasoning within the lifecycle of existing stream reasoning applications [41]

11:38 How Does Knowledge Evolve in Open Knowledge Graphs?

Future Directions

Much attention is still required concerning logical reasoning to analyse and spread the evolution in state-of-the-art open knowledge graphs. As discussed, performing reasoning may be prohibitive in modern knowledge graphs if the full expressiveness of the underlying ontology is used. State-ofthe-art solutions focus on tractable fragments (e.g. OWL 2 profiles) to scale with large knowledge graphs and ontologies; however, coping with these KGs still poses essential challenges in terms of scale completeness and errors in the data. To assess how far the current approaches can take us, a comprehensive analysis of reasoning methods with a combination of general-purpose open KGs is necessary to understand the limitations in real-world settings. The combination of deductive and inductive techniques [65], as discussed in Section 6.3, is key to tackling these challenges as it leads to data and knowledge-driven techniques to, e.g., complement the evolving knowledge graph and to identify and correct potentially wrong new facts [48].

Stream reasoning is a candidate to have a central role in making sense of evolving knowledge graphs. In particular, expressive stream reasoners like Laser and LARS are candidates as formalisms to capture the complex interrelations between dynamic, versioned, and temporal KGs (cf Section 1). Similarly, it needs to be verified if existing languages like RSP-QL [63] are adequate for defining transformation across EKG types. Moreover, as we envision a more prominent role for events [99], agent-based reasoning methods are an important direction towards efficient methods to spread and handle the evolution [234]. Finally, from an application/engineering standpoint, different reasoning tasks may benefit from alternative KG encoding. Therefore, solutions like RSP4J [232], ChImp [183], or the SR PlayGround [210] need to evolve to welcome EKGs as first-class citizens.

6.3 Learning

In machine learning, KGs or ontologies are often transformed into vector space known as embeddings before use. KG embeddings are low-dimensional vector representations of entities and relationships within a KG. Typical tasks over such embeddings are link prediction, KG completion, node classification, query answering and data integration. Overall, we can distinguish two main families of graph embedding approaches: transductive and inductive. In transductive approaches, all nodes and relations are seen during training while new edges among seen nodes can be predicted at inference time. Inductive approaches instead allow to train on one version of the graph and then perform inference even with new nodes and edges introduced at testing time [8]. Therefore, when dealing with evolving KGs, we can distinguish between approaches that try to adapt transductive embedding methods to the case of dynamic or evolving graphs [43, 228, 249] and inductive methods that try to learn from contextual information and metadata, e. g. attributes or recurrent structures, high-level patterns that should allow inference even when the underlying data changes [89, 255, 58].

In the following, we first discuss existing continual learning approaches for embeddings of time-varying KGs, which could potentially be used to analyse the evolution of KGs in the future.

Next, we discuss temporal embeddings, where instead of embedding changes to the KG, the objective is to embed temporal information in vector space as well, therefore having a temporal KG as input. This type of method inherently requires a different KG, one with temporal information. Lastly, we discuss some applications of learning for KGs with the evolving nature in mind.

We aim to provide a high-level overview of learning with regard to evolving KGs but do not claim to provide an in-depth survey of approaches. We specifically want to highlight known open challenges at the end of this section.

Continuous Embedding Learning

PuTransE provides a self-contained model, based on TransE, which builds on a metaphor of "parallel universes" [228]. It trains several parallel embedding spaces using different subgraphs. The retraining is then limited to some of the parallel universes instead of relearning the entire representation. DKGE is another self-contained model [249]. In this approach, the embedding of an entity consists of two parts, the embedding of the entity itself and its context embedding. Both puTransE and DKGE deal with the changing graph as a whole, but their scalability to larger graphs is limited. Song et al. [217] was one the first efforts regarding dynamic KG embeddings, focusing on the addition of new triples on translation-based models, which the authors refer to as enrichment. Cui et al. [53] present a transfer-based strategy for embedding generation for newly introduced entities. This self-contained model is based on auto-encoders and scales well with large graphs. Daruna et al. [54] extends and reformulates the principles of five main types of continual learning methods not specific to KGs. These criteria are applied to KG embedding models, each requiring a different kind of adjustment to fit the continual learning problem. All three methods [217, 53, 54] can only deal with additions and not with deletions or modifications. Lastly, the objective of Hamaguchi et al. [106] is slightly different. They rely on GNNs to generate embeddings for unseen entities at testing time but do not update and reuse the embedding for subsequent use.

All the methods above have drawbacks and there does not exist a go-to method so far to embed KGs continuously. The big challenges are (1) deterioration of the task performance as the embedding is updated and (2) dealing with deletions of triples or nodes.

Temporal Knowledge Graph Embeddings

The goal of temporal KG embeddings is to represent a time-annotated KG in a vector space. As such, these methods are completely different from the methods dealing with evolving snapshots of a KG. Many methods have been proposed for embedding temporal KGs and can be roughly separated into four categories: geometric, matrix factorisation, deep learning, and model-agnostic methods. There are some methods that are meant for dynamic temporal knowledge graphs; however, they only consider additions, arguing that deletions are not necessary for temporal knowledge graphs [148].

Geometric methods use geometrical transformations, such as translations and rotations, to represent the KG elements, e.g. HyTE [56] as an extension of TransE for temporal knowledge graphs: it incorporating time in the entity-relation space through a hyper-plane for each timestamp. TeRo [125] and ChronoR [203] use rotation transformations by creating multiple representations over time and creating time-dependent embeddings for relations respectively.

Matrix factorisation methods produce embeddings by decomposition tensors representing the KG. While a KG is usually represented in a 3rd-order tensor, a temporal KG can be represented in a 4th-order tensor, with the additional dimension representing time. For example, TNTComplEx [141] extends ComplEx. One of the main peculiarities of the method is that it distinguishes between non-temporal predicates and temporal facts.

Deep-learning methods exploit neural networks to learn the embeddings. For example, RE-Net [127] learns temporal KG embeddings using a recurrent neural network, while [151] uses convolutional neural networks to capture the time interaction between facts.

Finally, model-agnostic methods can be applied to time-agnostic KG embedding methods to add the temporal dimension. For example, the Diachronic Embeddings [96] represent the entity as a function of time and entity, while [145] provides a framework to extend methods to deal with arbitrary time granularities.

11:40 How Does Knowledge Evolve in Open Knowledge Graphs?

Applications of Learning on Evolving Knowledge Graphs

Learning on evolving KGs has been extensively used for completion and data integration tasks. Here, we aim to present some examples, not a complete overview.

Completion. Completion is the problem of inferring missing links in a knowledge graph. In recent years, many approaches have been proposed to address completion through KG embeddings. There, the completion problem can be targeted through the link prediction task, i.e. finding a missing element of a statement given the other two, or question answering, i.e. discovering unseen links through approximate query answering. However, KG completion also includes other tasks. namely triple completion, node classification, and relation prediction [212]. Many of the methods presented above have been proposed for the purpose of KG completion and also tested with that task specifically. Shen et al. [212] provide an up-to-date overview of approaches in this area without considering KG evolution. They divide the existing approaches into those only relying on structural information (the knowledge graph) and those that also make use of additional resources. Additionally, some more specialised approaches deal with temporal KGs and their embeddings. commonsense KG, and hyper-relational KGs. Since our goal is not to provide such an overview, we refer to the work of Shen et al. [212]. Other surveys, which might not cover all of KG completion like Rossi et al. [200] who only focus on link prediction or Wang et al. [244] who focus more on the embedding methods and their application. Lastly, Gesese et al. [94] gave an overview of approaches which specifically deal with literals.

Question answering. Then there are the approaches that are more specific for approximate query answering, though they can also be seen as KG completion approaches. When not using the graph information directly, it is possible to answer queries approximately by making use of implicit information, the same as with KG completion. These can be presented in a transductive [160, 49] or inductive setting [88]. There are emerging question-answering systems that target time-related questions. For example, Jia et al. [123] propose TEQUILA, a system that enriches question-answering systems with temporal question-answer capabilities. Three years later, Jia et al. [124] created EXAQT, which answers questions using graph convolutional networks enhanced with time-aware entity embeddings. Otte et al. [173] propose a question-answering system that exploits an ensemble of diachronic temporal KG embeddings.

Data integration. An important practical application of graph embeddings lies in their usage for data integration tasks on KGs. This has been particularly impactful in bio-medicine, where data has been accumulating at an unprecedented rate and where efficient solutions for uniformly integrating and processing them are particularly needed. The work in [72] introduces a semantic KG embedding approach for biomedical data. As such, the authors focus on integrating biomedical literature, e. g. MedLine and PubMed, with ontologies used to contextualise KG entities. At a larger scale, a case in point of KG data integration with embeddings is the Bioteque knowledge graph [85]. This integrates data from 150 sources and comprises 450K biological entities and 30M relationships. To reduce dimensionality, while still capturing the various types of relationships between entities, specific node embeddings are defined.

Future Directions

When it comes to continuous learning of KG embeddings, in light of an evolving KG as input, there are three main challenges still open. From previously published approaches, the deterioration of task performance is a known problem when continuously learning as new information arrives. Here we can also draw parallels to catastrophic forgetting in other continual learning tasks without KGs. Additionally, most approaches currently available for the continuous learning of embeddings, do not consider deletions but only additions. Therefore, being able to handle all manners of

A. Polleres et al.

changes when embedding evolving KGs is an open challenge. Lastly, studies presented often only deal with a small number of updates to a KG, and hence, investigations are limited and need to be investigated at scale.

Embedding temporal knowledge graphs gained attention in recent years, and it is not at the same level of maturity of embedding techniques for knowledge graphs. One challenge lies in the definition of *temporal* knowledge graph, which is not standardised. Existing studies on the topic consider knowledge graphs where the temporal information is represented differently (see Section 5.1) and can have different semantics, e. g. time intervals where the fact is true or time instant where an event starts. Moreover, there is no set of well-defined and shared tasks, e. g. most studies focus on slightly different variations of the completion tasks, where time can or cannot be predicted. As a consequence, the existing methods are hardly comparable. Therefore, we envision the creation of de-facto standard datasets and tasks, which can help consolidate existing techniques and drive this research trend.

In parallel, as also mentioned in Section 6.2, there is an opportunity to enrich temporal knowledge graph embedding methods with deductive techniques. Specifically, in future, we expect novel research that combines embeddings, which are effective in capturing the structural information stored in a knowledge graph, with temporal logics, which have proven a robust solution to manage and reason on the temporal information.

By embedding a KG into a vector representation, we can potentially learn more about the evolution of the KG and conduct longitudinal analyses, e.g. of concept drift. However, due to the stochastic nature of the learning process, this remains a large open challenge, until the stochasticity problem is resolved to some extent [181]. We see a large number of open challenges when it comes to applications relying on embeddings of evolving KGs. Currently, we lack techniques and approaches for embedding-dependent tasks to be able to handle the changing KG without losing in performance or requiring complete recalculations. We can, however, also look at this from a slightly different perspective, that of the impact of evolution on these applications. When these applications first involve the learning of an embedding, it becomes extremely difficult to judge and capture the impact of evolution [181]. However, judging impact should not only be based on benchmark performance but rather the real impact in terms of changes to predictions. Therefore, we see an open challenge in analysing the performance of evolving tasks not in terms of metrics like mean-reciprocal-rank or accuracy, but rather the changes to the individual predictions. Approaches like inter-rater agreements may be useful for analysing localised changes in predictions [93].

6.4 Exploring Evolving Knowledge Graphs

When it comes to managing and analysing KGs, their heterogeneity constitutes both a defining characteristic and a challenge. In particular, both the contents and the schemas of these graphs have become less and less familiar even to domain experts and almost impenetrable to first-time users, leading to a rising need for exploratory methods for knowledge graphs [149, 150]. *Knowledge* graph exploration [149] is the machine-assisted and progressive process of analysis of a KG leading to (1) the understanding of the structure and nature of the graph, (2) the identification of which portion of the KG can satisfy the current information need, and (3) the extraction of insights that enable the formulation of novel research questions and hypotheses. These goals translate to three main tasks: (i) summarization and profiling, (ii) exploratory data analytics, and (iii) exploratory search. Looking at the dimension of evolution (Figure 1), we see that time adds a new dimension to the data to be explored and becomes a subject of exploration by itself when we explore how the structure (and not only the content) of the KG evolves and can provide new information that can then in turn guide the exploration.

11:42 How Does Knowledge Evolve in Open Knowledge Graphs?

Data profiling is the simplest form of exploration providing descriptive statistics and analysis about a given dataset. Typically, profiling tasks include counting the number of classes and their instances, summarising value distributions for specific (numerical) attributes, and they also identify important descriptors of the structure of the graph, e.g. node degree distribution [154]. There are also structural summarization [47] and pattern mining tasks [257, 194] to facilitate understanding the structure of the graph and to obtain concise representations of the most salient features of their contents. Profiling an evolving KG will provide insights into its structural changes through time, yet, only a few works scratch the surface of profiling KG evolution [76, 32]. They focus on analysing the statistical dataset characteristics at different snapshots [76], while more recent work started proposing algorithms to incrementally compute and update structural graph summaries defined as equivalence relations [32]. Therefore, to date, how to extend existing methods to tackle the challenges of scalable and continuous profiling of evolving KGs is still an open question. Moreover, as described above (Section 4.1), we are missing methods that can concisely summarise the results of a longitudinal analysis of the evolution of the schema and main characteristics of the dataset.

Exploratory analytics, is similar to data profiling since it is an iterative process of extracting aggregate information from portions of the graph, similar to a localised data profiling task [3, 51, 116]. The typical focus is to provide functionalities equivalent to those of *multi-dimensional analysis* that exist for relational data. Here, we see the need for analytical methods that can effectively include the temporal dimension in exploratory analytics, both when time is part of the data, as well as when time is treated as metadata. In this regard, we have recently witnessed a proposal to allow aggregation both at the attribute and at the time dimension [133, 239]. This is especially relevant since it offers the opportunity to employ exploration strategies that can guide the user through the evolution of the graph based on the identification of time intervals of significant growth, shrinkage, or stability of certain attribute values.

Finally, *Exploratory search* supports information needs that can be answered by retrieving specific entities, relationships, or paths. Exploratory queries change the traditional semantics of the search input: *instead of strictly prescribing the conditions that the desired result set must satisfy, they provide a hint of what is relevant* [149]. In these cases, the system should become an active agent able to suggest or infer query reformulations, refinements, and suggestions to help the user in their navigation. On the one hand, we see the need to help users explore the evolution of a given entity, e.g. identifying the most relevant changes w.r.t. a given stable state. On the other hand, the question is whether tapping into the analysis of the evolution of the KG, this information could be used to provide better suggestions or refinements. Overall, the methods designed to allow for query processing over evolving graphs (see Section 6.1) can still be used under the hood to enable exploratory search. Yet, to date, no method actively accounts for the rate and evolution of given entities and substructures when computing query suggestions to help the user in their exploration.

Future Directions

In summary, we identify both the need for new exploratory techniques that take into account the temporal dimension, and at the same time we highlight how existing techniques need to face the computational challenges posed by a KG that is not static anymore but dynamic. In particular, we postulate the need for new KG profiling techniques that apply longitudinal analysis to the data model in the KG through its lifespan. Furthermore, they see the need for methods that can understand trends in graph-centric measures and can efficiently compute and measure their evolution over time while the graph evolves. Finally, we ask which signals can be extracted from the observation of the evolution of the graph that can be exploited as a signal to help users identify interesting information and to identify methods to assist users in navigating more easily through an unfamiliar KG.

7 Summary and Conclusions

While KGs are gaining attention overall, the analysis and management of their evolution is still a "less conquered" territory in research. The present paper encourages us to look closely at KG evolution and make it a more prominent subject in our research. After emphasising that different types of KGs likely have very different change and evolution characteristics, we motivated various dimensions of looking at the evolution of KGs. We started investigating how known static structural analyses of KGs can be considered in a dynamic context, exploring the evolution of quality and consistency over time, to specific aspects related to dynamic collaboration processes of KG contributors, and finally, semantic drift in KGs. We provided an overview of publicly available KGs and, specifically, the availability of historical longitudinal data about their evolution that could serve as a starting point for analyses, as well as an overview of already existing studies.

We identified a research gap in terms of specific metrics for studying KG evolution in different dimensions; here, in the future, we will need to address concerns regarding the application and adaption of static metrics for longitudinal and time-series analyses on KGs. In particular, regarding the analysis of KG consistency over time, we have sketched viable approaches in Section 4.2; however, these have not yet been applied in an analysis of KG evolution, presenting a notable research gap.

Finally, we had a detailed discussion about the metrics and techniques that can be applied to analyse KGs. We suggested exploring more methods not commonly used in our community but well-established in other fields, such as network science. This field has a long-standing tradition of analysing large-scale networks' structural and dynamic aspects. Given the extensive reach and rapid growth of KGs, it is imperative to implement similar methods in our field. However, we should remember that these methods may require adaptations due to the "multi-level" network characteristic of KGs, as they can be viewed as overlaid networks encompassing all their properties.

We further discussed challenges related to different graph representation models and storage strategies for the extraction/construction of dynamic KGs. They focus mainly on the interoperability of the different ways time is captured in evolving KGs, different schemas and their mapping to each other, and how these could be integrated in the future, for instance in standardised ways to query evolving KGs. Regarding storage, currently, different storage solutions facilitate different types of analyses. Still, in the future, we hope to see storage solutions enabling the storage of dynamic and versioned graphs to enable all kinds of analyses.

The popular downstream tasks when using knowledge graphs, such as querying, reasoning, and learning, can benefit from considering the evolution of knowledge more explicitly. Considering the temporal dimension as a first-class citizen at the query level opens the possibility to specific operators for retrieving data about time, through time, and in time. In the future, we can expect further extensions of SPARQL and other LPG-specific query languages to support these operators, ideally combining temporal, versioned, and continuous flavours in more comprehensive query languages. Similarly, reasoning is affected by evolving knowledge. On the one hand, there are new algorithmic challenges, e.g. how to maintain a materialisation incrementally and reactively (on time). On the other hand, considering temporal logics at a fundamental level could enhance reasoning over evolving KGs and their schemas over time. KG evolution can also provide additional signals for training machine learning models, capturing dynamic processes. However, respective approaches that for instance capture updates in learned embeddings, are still lacking in performance and scalability to be helpful in practical analytical use cases. Finally, we envision querying, reasoning, and learning to be fruitfully combined to overcome individual weaknesses for managing, processing and analysing evolving KGs, eventually creating new applications and services. While such combinations have been studied for static KGs, we expect and hope to see more studies in the future that consider the evolving knowledge case.

11:44 How Does Knowledge Evolve in Open Knowledge Graphs?

In the following list, we summarise the most important future directions and open challenges, in particular about learning more about and understanding how knowledge evolves in open, general-purpose KGs:

- Systematic analysis of open general-purpose KGs along various dimensions of evolution such as dynamics, timeliness and monotonicity, but also structural, semantic and collaborative aspects making use of approaches such as time-series analysis and network science.
- Principled approaches to *create evolving subsets* of KGs in evolution for specific benchmarking tasks would be dearly needed to better understand these tasks "in evolution".
- Further development of *metrics for measuring and understanding knowledge evolution in KGs*, specifically capable of handling outliers and the complexity and size of the known KGs.
- Interoperability between different KG models, mainly RDF and LPG, and query languages that support these to enable better and complementary analyses of temporal KGs.
- Development of adaptive approaches and respective querying capabilities to store dynamic and versioned KGs simultaneously, making it possible to apply any analysis (time-series and constraint-based) on the evolving KGs.
- The combination of deductive and inductive techniques [65] is necessary to tackle challenges with reasoning (scale, completeness, errors) as it leads to data and knowledge-driven techniques. For example, one may complement the evolving knowledge graph and identify and correct potentially wrong new facts.
- Development of novel continuous embedding approaches and methods for embedding temporal KGs, i.e., the study of concept drift with large evolving KGs from different perspectives becomes a new open challenge.
- **T**ackling the computational challenges of existing *exploratory techniques* and the development of new ones specifically facilitating longitudinal analysis through, e.g. graph-centric measures to help navigate the evolution of an unfamiliar KG.

In summary, we have performed an extensive survey of evolution in KGs - significantly more extensive than initially expected. From this survey we conclude that KG's evolution is apparently a field that – while having already attracted a lot of attention – remains to have various open questions. The authors hope we motivated the readers to work jointly on more in-depth investigations and more standardised, agreed-upon methods of capturing and dealing with Knowledge (Graph) Evolution as well as newer methods for analysis as identified in this work.

— References

- 1 Mohammad Abdel-Qader, Ansgar Scherp, and Iacopo Vagliano. Analyzing the evolution of vocabulary terms and their impact on the LOD cloud. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings, volume 10843 of Lecture Notes in Computer Science (LNCS), pages 1–16. Springer, 2018. doi:10.1007/ 978-3-319-93417-4_1.
- 2 Mohammad Abdel-Qader, Iacopo Vagliano, and Ansgar Scherp. Analyzing the Evolution of Linked Vocabularies. In Maxim Bakaev, Flavius Frasincar, and In-Young Ko, editors, Web Engineering -19th International Conference, ICWE 2019, Daejeon, South Korea, June 11-14, 2019, Proceed-

ings, volume 11496 of Lecture Notes in Computer Science (LNCS), pages 409-424. Springer, 2019. doi:10.1007/978-3-030-19274-7_29.

- 3 Alberto Abelló, Oscar Romero, Torben Bach Pedersen, Rafael Berlanga Llavori, Victoria Nebot, María José Aramburu Cabo, and Alkis Simitsis. Using semantic web technologies for exploratory OLAP: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):571–588, 2015. doi:10.1109/TKDE.2014.2330822.
- 4 Ghadeer Abuoda, Daniele Dell'Aglio, Arthur Keen, and Katja Hose. Transforming rdf-star to property graphs: A preliminary analysis of transformation approaches. In Muhammad Saleem and Axel-Cyrille Ngonga Ngomo, editors, Proceedings of the QuWeDa 2022: 6th Workshop on Storing, Querying and Benchmarking Knowledge Graphs co-located with 21st International

Semantic Web Conference (ISWC 2022), Hangzhou, China, 23-27 October 2022, volume 3279 of CEUR Workshop Proceedings, pages 17-32. CEUR-WS.org, 2022. URL: https://ceur-ws. org/Vol-3279/paper2.pdf.

- 5 Christian Aebeloe, Gabriela Montoya, and Katja Hose. ColChain: Collaborative Linked Data Networks. In *Proceedings of the Web Conference 2021*, pages 1385–1396, Ljubljana Slovenia, apr 2021. ACM. doi:10.1145/3442381.3450037.
- 6 Temitope Ajileye and Boris Motik. Materialisation and data partitioning algorithms for distributed RDF systems. *Journal of Web Semantics*, 73:100711, 2022. doi:10.1016/J.WEBSEM.2022. 100711.
- 7 Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews* of Modern Physics, 74(1):47–97, jan 2002. doi: 10.1103/RevModPhys.74.47.
- 8 Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Mikhail Galkin, Sahand Sharifzadeh, Asja Fischer, Volker Tresp, and Jens Lehmann. Bringing Light Into the Dark: A Large-Scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8825–8845, dec 2022. doi: 10.1109/TPAMI.2021.3124805.
- 9 James F. Allen. Maintaining knowledge about temporal intervals. Communications of the ACM, 26(11):832-843, 1983. doi:10.1145/182.358434.
- 10 Francesca Alloatti, Riley Capshaw, Molka Dhouib, María G. Buey, Ismail Harrando, Jaime Salas, and Claudia d'Amato. Towards an automatic detection of evolution in knowledge graphs. In Valentina Anita Carriero, Luigi Asprino, and Russa Biswas, editors, Knowledge Graphs Evolution and Preservation - A Technical Report from ISWS 2019, volume abs/2012.11936 of Computing Research Repository (CoRR). arXiv, 2020. doi:10.48550/arXiv.2012.11936.
- 11 Rebekka Alm, Sven Kiehl, Birger Lantow, and Kurt Sandkuhl. Applicability of quality metrics for ontologies on ontology design patterns. In Joaquim Filipe and Jan L. G. Dietz, editors, KEOD 2013 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Vilamoura, Algarve, Portugal, 19-22 September, 2013, pages 48–57. SciTePress, 2013. doi:10.5220/0004541400480057.
- 12 Landy Andriamampianina, Franck Ravat, Jiefu Song, and Nathalie Vallès-Parlangeau. Querying Temporal Property Graphs. In Xavier Franch, Geert Poels, Frederik Gailly, and Monique Snoeck, editors, Advanced Information Systems Engineering - 34th International Conference, CAiSE 2022, Leuven, Belgium, June 6-10, 2022, Proceedings, volume 13295 of Lecture Notes in Computer Science (LNCS), pages 355–370. Springer, 2022. doi:10.1007/978-3-031-07472-1_21.
- 13 Renzo Angles, Carlos Buil Aranda, Aidan Hogan, Carlos Rojas, and Domagoj Vrgoč. WDBench: A Wikidata Graph Query Benchmark. In Ulrike Sattler, Aidan Hogan, Maria Keet, Valentina Presutti, João Paulo A. Almeida, Hideaki Takeda, Pierre Monnin, Giuseppe Pirrò, and Claudia d'Amato,

editors, *The Semantic Web – ISWC 2022*, pages 714–731, Cham, 2022. Springer International Publishing. doi:10.1007/978-3-031-19433-7_41.

- 14 Renzo Angles, Angela Bonifati, Stefania Dumbrava, George Fletcher, Alastair Green, Jan Hidders, Bei Li, Leonid Libkin, Victor Marsault, Wim Martens, Filip Murlak, Stefan Plantikow, Ognjen Savkovic, Michael Schmidt, Juan Sequeda, Slawek Staworko, Dominik Tomaszuk, Hannes Voigt, Domagoj Vrgoc, Mingxi Wu, and Dusan Zivkovic. PG-Schema: Schemas for Property Graphs. Proceedings of the ACM on Management of Data, 1(2):198:1–198:25, 2023. doi:10.1145/3589778.
- 15 Renzo Angles, Harsh Thakkar, and Dominik Tomaszuk. Mapping RDF Databases to Property Graph Databases. *IEEE Access*, 8:86091–86110, 2020. doi:10.1109/ACCESS.2020.2993117.
- 16 Darko Anicic, Paul Fodor, Sebastian Rudolph, and Nenad Stojanovic. EP-SPARQL: a unified language for event processing and stream reasoning. In Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors, Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011, pages 635–644. ACM, 2011. doi:10.1145/1963405.1963495.
- 17 Arvind Arasu, Shivnath Babu, and Jennifer Widom. The CQL continuous query language: semantic foundations and query execution. VLDB Journal, 15(2):121–142, 2006. doi:10.1007/ S00778-004-0147-Z.
- 18 Marcelo Arenas, Pedro Bahamondes, Amir Aghasadeghi, and Julia Stoyanovich. Temporal regular path queries. In 38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022, pages 2412– 2425. IEEE, 2022. doi:10.1109/ICDE53745.2022. 00226.
- 19 Natanael Arndt, Patrick Naumann, Norman Radtke, Michael Martin, and Edgard Marx. Decentralized Collaborative Knowledge Management Using Git. Journal of Web Semantics, 54:29–47, jan 2019. doi:10.1016/J.WEBSEM.2018.08.002.
- 20 Sören Auer. The emerging web of linked data. In Ayman Jameel Alnsour and Shadi A. Aljawarneh, editors, Proceedings of the 2nd International Conference on Intelligent Semantic Web-Services and Applications, ISWSA 2011, Amman, Jordan, April 18-20, 2011, page 1. ACM, 2011. doi:10.1145/1980822.1980823.
- 21 Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook.* Cambridge University Press, Cambridge, UK, 2 edition, 2007. doi:10.1017/CB09780511711787.
- 22 Thomas Baker and Eric Prud'hommeaux. Shape Expressions (ShEx) 2.1 Primer. Technical report, W3C Shape Expressions Community Group, oct 2019. URL: http://shex.io/shex-primer/.
- 23 Marco Balduini, Irene Celino, Daniele Dell'Aglio, Emanuele Della Valle, Yi Huang, Tony Kyung-il Lee, Seon-Ho Kim, and Volker Tresp. Reality mining on micropost streams. *Semantic Web*, 5(5):341– 356, 2014. doi:10.3233/SW-130107.

11:46 How Does Knowledge Evolve in Open Knowledge Graphs?

- 24 Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, and Michael Grossniklaus. Incremental reasoning on streams and rich background knowledge. In Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache, editors, The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part I, volume 6088 of Lecture Notes in Computer Science (LNCS), pages 1–15. Springer, 2010. doi:10.1007/978-3-642-13486-9_1.
- 25 Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, aug 2020. doi:10.1016/j.physrep.2020.05.004.
- 26 Federico Battiston, Vincenzo Nicosia, and Vito Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3):032804, mar 2014. doi:10.1103/PhysRevE.89.032804.
- 27 Harald Beck, Minh Dao-Tran, Thomas Eiter, and Michael Fink. LARS: A logic-based framework for analyzing reasoning over streams. In Blai Bonet and Sven Koenig, editors, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA, pages 1431–1438. AAAI Press, 2015. doi:10.1609/AAAI.V29I1.9408.
- 28 Zohra Bellahsene, Angela Bonifati, and Erhard Rahm, editors. Schema Matching and Mapping. Data-Centric Systems and Applications. Springer, 2011. doi:10.1007/978-3-642-16518-4.
- 29 Luigi Bellomarini, Davide Benedetto, Georg Gottlob, and Emanuel Sallinger. Vadalog: A modern architecture for automated reasoning with large knowledge graphs. *Information Sys*tems, 105:101528, 2022. doi:10.1016/J.IS.2020. 101528.
- 30 Konstantina Bereta, Panayiotis Smeros, and Manolis Koubarakis. Representation and Querying of Valid Time of Triples in Linked Geospatial Data. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, *The Semantic Web: Semantics and Big Data*, volume 7882, pages 259–274, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. doi:10.1007/978-3-642-38288-8_18.
- 31 Philip A. Bernstein. Applying model management to classical meta data problems. In First Biennial Conference on Innovative Data Systems Research, CIDR 2003, Asilomar, CA, USA, January 5-8, 2003, Online Proceedings. www.cidrdb.org, 2003. URL: http://www-db.cs.wisc.edu/cidr/ cidr2003/program/p19.pdf.
- 32 Till Blume, David Richerby, and Ansgar Scherp. Incremental and parallel computation of structural graph summaries for evolving graphs. In

Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, pages 75–84, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3340531.3411878.

- 33 Till Blume and Ansgar Scherp. Indexing data on the web: A comparison of schema-level indices for data search. In Sven Hartmann, Josef Küng, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, editors, Database and Expert Systems Applications - 31st International Conference, DEXA 2020, Bratislava, Slovakia, September 14-17, 2020, Proceedings, Part II, volume 12392 of Lecture Notes in Computer Science (LNCS), pages 277–286. Springer, 2020. doi:10.1007/978-3-030-59051-2_18.
- 34 Olivier Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic acids research, 32(Database-Issue):267-270, 2004. doi:10.1093/NAR/GKH061.
- 35 Michael H. Böhlen, Anton Dignös, Johann Gamper, and Christian S. Jensen. Temporal data management - an overview. In Esteban Zimányi, editor, Business Intelligence and Big Data - 7th European Summer School, eBISS 2017, Bruxelles, Belgium, July 2-7, 2017, Tutorial Lectures, volume 324 of Lecture Notes in Business Information Processing, pages 51–83. Springer, 2017. doi: 10.1007/978-3-319-96655-7_3.
- 36 Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM* SIGMOD International Conference on Management of Data, SIGMOD '08, pages 1247–1250, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1376616.1376746.
- 37 ES Bolotnikova, TA Gavrilova, and VA Gorovoy. To a method of evaluating ontologies. Journal of Computer and Systems Sciences International, 50(3):448–461, 2011. doi:10.1134/ S1064230711010072.
- 38 Angela Bonifati, George H. L. Fletcher, Hannes Voigt, and Nikolay Yakovets. *Querying Graphs*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2018. doi:10.2200/ S00873ED1V01Y201808DTM051.
- 39 Angela Bonifati, Irena Holubová, Arnau Prat-Pérez, and Sherif Sakr. Graph Generators: State of the Art and Open Challenges. ACM Computing Surveys, 53(2):1–30, mar 2021. doi:10.1145/ 3379445.
- 40 Angela Bonifati, Wim Martens, and Thomas Timm. An analytical study of large SPARQL query logs. VLDB Journal, 29(2-3):655–679, 2020. doi:10.1007/S00778-019-00558-9.
- 41 Pieter Bonte and Riccardo Tommasini. Streaming linked data: A survey on life cycle compliance. Journal of Web Semantics, 77:100785, 2023. doi:10.1016/J.WEBSEM.2023.100785.
- 42 Dan Brickley and Ramanathan Guha. RDF Schema 1.1. Technical report, The World Wide Web Consortium (W3C), feb 2014. URL: http://www.w3.org/TR/2014/ REC-rdf-schema-20140225/.
- 43 Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. Temporal know-

ledge graph completion: A survey. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 6545–6553. International Joint Conferences on Artificial Intelligence Organization, aug 2023. Survey Track. doi: 10.24963/IJCAI.2023/734.

- 44 Giuseppe Capobianco, Danilo Cavaliere, and Sabrina Senatore. Ontodrift: a semantic drift gauge for ontology evolution monitoring. In Fabrizio Orlandi, Damien Graux, Maria-Esther Vidal, Javier D. Fernández, and Jeremy Debattista, editors, Proceedings of the 6th Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW) co-located with the 19th International Semantic Web Conference (ISWC 2020), Virtual event (instead of Athens, Greece), November 1st, 2020, volume 2821 of CEUR Workshop Proceedings, pages 1–10. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2821/xpreface.pdf.
- 45 Gavin Carothers. RDF 1.1 N-Quads A line-based syntax for RDF datasets. Technical report, The World Wide Web Consortium (W3C), 2014. URL: http://www.w3.org/TR/2014/ REC-n-quads-20140225/.
- 46 David Carral, Irina Dragoste, Larry González, Ceriel J. H. Jacobs, Markus Krötzsch, and Jacopo Urbani. VLog: A Rule Engine for Knowledge Graphs. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtech Svátek, Isabel F. Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, The Semantic Web - ISWC 2019
 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II, volume 11779 of Lecture Notes in Computer Science (LNCS), pages 19–35. Springer, 2019. doi:10.1007/978-3-030-30796-7_2.
- 47 Sejla Cebiric, François Goasdoué, Haridimos Kondylakis, Dimitris Kotzinos, Ioana Manolescu, Georgia Troullinou, and Mussab Zneika. Summarizing semantic graphs: a survey. VLDB Journal, 28(3):295–327, 2019. doi:10.1007/ S00778-018-0528-3.
- 48 Jiaoyan Chen, Ernesto Jiménez-Ruiz, Ian Horrocks, Xi Chen, and Erik Bryhn Myklebust. An assertion and alignment correction framework for large scale knowledge bases. *Semantic Web*, 14(1):29–53, 2023. doi:10.3233/SW-210448.
- 49 Xuelu Chen, Ziniu Hu, and Yizhou Sun. Fuzzy logic based logical query answering on knowledge graphs. In Thirty-Sixth Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, AAAI 2022, February 22 March 1, 2022, pages 3939–3948. AAAI Press, 2022. doi:10.1609/AAAI.V36I4.20310.
- 50 Maxime Clément and Matthieu J. Guitton. Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia. *Computers in Human Behavior*, 50:66–75, 2015. doi: 10.1016/J.CHB.2015.03.078.
- 51 Dario Colazzo, François Goasdoué, Ioana Manolescu, and Alexandra Roatis. RDF analytics: lenses over semantic graphs. In Chin-Wan Chung, Andrei Z. Broder, Kyuseok Shim, and Torsten Suel, editors, 23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea,

April 7-11, 2014, pages 467–478. ACM, 2014. doi:10.1145/2566486.2567982.

- 52 Gianpaolo Cugola and Alessandro Margara. Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys*, 44(3):15:1–15:62, 2012. doi:10.1145/2187671.2187677.
- 53 Yuanning Cui, Yuxin Wang, Zequn Sun, Wenqiang Liu, Yiqiao Jiang, Kexin Han, and Wei Hu. Lifelong embedding learning and transfer for growing knowledge graphs. In Brian Williams, Yiling Chen, and Jennifer Neville, editors, *Thirty-Seventh Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, AAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 4217–4224. AAAI Press, 2023. doi:10.1609/AAAI.V37I4.25539.
- 54 Angel Andres Daruna, Mehul Gupta, Mohan Sridharan, and Sonia Chernova. Continual learning of knowledge graph embeddings. *IEEE Robotics and Automation Letters*, 6(2):1128–1135, 2021. doi:10.1109/LRA.2021.3056071.
- 55 Souripriya Das, Seema Sundara, and Richard Cyganiak. R2RML: RDB to RDF mapping language. Technical report, The World Wide Web Consortium (W3C), sep 2012. URL: http://www.w3.org/ TR/2012/REC-r2rml-20120927/.
- 56 Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha P. Talukdar. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, pages 2001–2011. Association for Computational Linguistics, 2018. doi:10.18653/V1/D18-1225.
- 57 Robert David, Shqiponja Ahmetaj, Mantas Šimkus, and Axel Polleres. Repairing SHACL constraint violations using answer set programming. In Proceedings of the 21st International Semantic Web Conference (ISWC 2022), volume 13489 of Lecture Notes in Computer Science (LNCS), pages 375–391, Virtual Conference (Hangzhou, China), oct 2022. Springer. doi:10.1007/ 978-3-031-19433-7_22.
- 58 Daniel Daza, Michael Cochez, and Paul Groth. Inductive Entity Representations from Text via Link Prediction. In *Proceedings of the Web Conference 2021*, WWW '21, pages 798–808, New York, NY, USA, jun 2021. Association for Computing Machinery. doi:10.1145/3442381.3450141.
- 59 Daniel de Leng and Fredrik Heintz. DyKnow: A dynamically reconfigurable stream reasoning framework as an extension to the robot operating system. In 2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots, SIMPAR 2016, San Francisco, CA, USA, December 13-16, 2016, pages 55–60. IEEE, 2016. doi:10.1109/SIMPAR.2016.7862375.
- 60 Ariel Debrouvier, Eliseo Parodi, Matías Perazzo, Valeria Soliani, and Alejandro A. Vaisman. A model and query language for temporal graph databases. VLDB J., 30(5):825–858, 2021. doi: 10.1007/S00778-021-00675-4.

11:48 How Does Knowledge Evolve in Open Knowledge Graphs?

- 61 Ariel Debrouvier, Eliseo Parodi, Matías Perazzo, Valeria Soliani, and Alejandro A. Vaisman. A model and query language for temporal graph databases. VLDB Journal, 30(5):825–858, 2021. doi:10.1007/S00778-021-00675-4.
- 62 Emanuele Della Valle, Stefano Ceri, Frank van Harmelen, and Dieter Fensel. It's a streaming world! Reasoning upon rapidly changing information. *IEEE Intelligent Systems*, 24(6):83–89, 2009. doi:10.1109/MIS.2009.125.
- 63 Daniele Dell'Aglio, Emanuele Della Valle, Jean-Paul Calbimonte, and Óscar Corcho. RSP-QL semantics: A unifying query model to explain heterogeneity of RDF stream processing systems. International Journal on Semantic Web and Information Systems (IJSWIS), 10(4):17–44, 2014. doi:10.4018/IJSWIS.2014100102.
- 64 Daniele Dell'Aglio, Emanuele Della Valle, Frank van Harmelen, and Abraham Bernstein. Stream reasoning: A survey and outlook. *Data Science*, 1(1-2):59–83, 2017. doi:10.3233/DS-170006.
- 65 Lauren Nicole Delong, Ramon Fernández Mir, Matthew Whyte, Zonglin Ji, and Jacques D. Fleuriot. Neurosymbolic AI for reasoning on graph structures: A survey. *Computing Research Repository (CoRR)*, abs/2302.07200, 2023. doi: 10.48550/ARXIV.2302.07200.
- 66 Anton Dignös, Michael H. Böhlen, and Johann Gamper. Temporal alignment. In K. Selçuk Candan, Yi Chen, Richard T. Snodgrass, Luis Gravano, and Ariel Fuxman, editors, Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012, Scottsdale, AZ, USA, May 20-24, 2012, pages 433–444. ACM, 2012. doi:10.1145/2213836.2213886.
- 67 Yuehang Ding, Hongtao Yu, Ruiyang Huang, and Yunjie Gu. Complex network based knowledge graph ontology structure analysis. In 2018 1st IEEE international conference on hot informationcentric networking (HotICN), pages 193–199, 2018. doi:10.1109/HOTICN.2018.8606002.
- 68 Renata Queiroz Dividino, Thomas Gottron, and Ansgar Scherp. Strategies for Efficiently Keeping Local Linked Open Data Caches Up-To-Date. In Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, and Steffen Staab, editors, The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II, volume 9367 of Lecture Notes in Computer Science (LNCS), pages 356-373. Springer, 2015. doi:10.1007/ 978-3-319-25010-6 24.
- 69 Renata Queiroz Dividino, Thomas Gottron, Ansgar Scherp, and Gerd Gröner. From Changes to Dynamics: Dynamics Analysis of Linked Open Data Sources. In Elena Demidova, Stefan Dietze, Julian Szymanski, and John G. Breslin, editors, Proceedings of the 1st International Workshop on Dataset PROFILing & fEderated Search for Linked Data co-located with the 11th Extended Semantic Web Conference, PROFILES@ESWC 2014, Anissaras, Crete, Greece, May 26, 2014, volume 1151 of CEUR Workshop Proceedings.

CEUR-WS.org, 2014. URL: https://ceur-ws.org/Vol-1151/paper4.pdf.

- 70 Renata Queiroz Dividino, Ansgar Scherp, Gerd Gröner, and Thomas Grotton. Change-a-lod: Does the schema on the linked data cloud change or not? In Olaf Hartig, Juan F. Sequeda, Aidan Hogan, and Takahide Matsutsuka, editors, Proceedings of the Fourth International Workshop on Consuming Linked Data, COLD 2013, Sydney, Australia, October 22, 2013, volume 1034 of CEUR Workshop Proceedings. CEUR-WS.org, 2013. URL: https://ceur-ws.org/Vol-1034/ DividinoEtAl_COLD2013.pdf.
- 71 Rim Djedidi and Marie-Aude Aufaure. ONTO-EVO A L an ontology evolution approach guided by pattern modeling and quality evaluation. In International symposium on foundations of information and knowledge systems, pages 286–305, 2010. doi:10.1007/978-3-642-11829-6_19.
- 72 Jens Dörpinghaus and Marc Jacobs. Semantic knowledge graph embeddings for biomedical research: Data integration using linked open data. In Mehwish Alam, Ricardo Usbeck, Tassilo Pellegrini, Harald Sack, and York Sure-Vetter, editors, Proceedings of the Posters and Demo Track of the 15th International Conference on Semantic Systems co-located with 15th International Conference on Semantic Systems (SEMANTiCS 2019), Karlsruhe, Germany, September 9th - to - 12th, 2019, volume 2451 of CEUR Workshop Proceedings. CEUR-WS.org, 2019. URL: https://ceur-ws. org/Vol-2451/paper-10.pdf.
- 73 Astrid Duque-Ramos, Manuel Quesada-Martínez, Miguela Iniesta-Moreno, Jesualdo Tomás Fernández-Breis, and Robert Stevens. Supporting the analysis of ontology evolution processes through the combination of static and dynamic scaling functions in OQuaRE. Journal of Biomedical Semantics, 7(1):63, oct 2016. doi:10.1186/S13326-016-0091-Z.
- 74 Thomas Eiter, Andreas A. Falkner, Patrik Schneider, and Peter Schüller. ASP-Based signal plan adjustments for traffic flow optimization. In Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang, editors, ECAI 2020 - 24th european conference on artificial intelligence, 29 august-8 september 2020, santiago de compostela, spain, august 29 - september 8, 2020 - including 10th conference on prestigious applications of artificial intelligence (PAIS 2020), volume 325 of Frontiers in artificial intelligence and applications, pages 3026–3033. IOS Press, 2020. doi:10.3233/FAIA200478.
- 75 Mohamed G. Elfeky, Walid G. Aref, and Ahmed K. Elmagarmid. Periodicity Detection in Time Series Databases. *IEEE Transactions on Knowledge and Data Engineering*, 17(7):875–887, 2005. doi: 10.1109/TKDE.2005.114.
- 76 Ivan Ermilov, Michael Martin, Jens Lehmann, and Sören Auer. Linked open data statistics: Collection and exploitation. In Pavel Klinov and Dmitry Mouromtsev, editors, Knowledge Engineering and the Semantic Web - 4th International Conference, KESW 2013, St. Petersburg, Russia, October 7-9, 2013. Proceedings, volume 394

of Communications in Computer and Information Science, pages 242–249. Springer, 2013. doi: 10.1007/978-3-642-41360-5_19.

- 77 Jérôme Euzenat and Pavel Shvaiko. Ontology Matching, Second Edition. Springer, 2013.
- 78 Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *Theoretical Computer Science*, 336(1):89–124, 2005. doi:10.1016/J.TCS.2004. 10.033.
- 79 Daniel Faria, Catia Pesquita, Emanuel Santos, Matteo Palmonari, Isabel F. Cruz, and Francisco M. Couto. The AgreementMakerLight Ontology Matching System. In Robert Meersman, Hervé Panetto, Tharam S. Dillon, Johann Eder, Zohra Bellahsene, Norbert Ritter, Pieter De Leenheer, and Dejing Dou, editors, On the Move to Meaningful Internet Systems: OTM 2013 Conferences - Confederated International Conferences: CoopIS, DOA-Trusted Cloud, and ODBASE 2013, Graz, Austria, September 9-13, 2013. Proceedings, volume 8185 of Lecture Notes in Computer Science (LNCS), pages 527-541. Springer, 2013. doi:10.1007/978-3-642-41030-7_38.
- 80 Javier D. Fernández, Patrik Schneider, and Jürgen Umbrich. The DBpedia wayback machine. In Axel Polleres, Tassilo Pellegrini, Sebastian Hellmann, and Josiane Xavier Parreira, editors, Proceedings of the 11th International Conference on Semantic Systems, SEMANTiCS 2015, Vienna, Austria, September 15-17, 2015, pages 192–195. ACM, 2015. doi:10.1145/2814864.2814889.
- 81 Javier D. Fernández, Jürgen Umbrich, Axel Polleres, and Magnus Knuth. Evaluating query and storage strategies for RDF archives. In *Proceedings* of the 12th International Conference on Semantic Systems, SEMANTICS 2016, pages 41–48, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2993318.2993333.
- 82 Javier D. Fernández, Wouter Beek, Miguel A. Martínez-Prieto, and Mario Arias. LOD-a-lot. In Claudia d'Amato, Miriam Fernandez, Valentina Tamma, Freddy Lecue, Philippe Cudré-Mauroux, Juan Sequeda, Christoph Lange, and Jeff Heflin, editors, *The Semantic Web – ISWC 2017*, Lecture Notes in Computer Science (LNCS), pages 75–83, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-68204-4_7.
- 83 Javier D. Fernández, Miguel A. Martínez-Prieto, Claudio Gutierrez, Axel Polleres, and Mario Arias. Binary RDF representation for publication and exchange (HDT). Web Semantics, 19:22–41, 2013. doi:10.1016/J.WEBSEM.2013.01.002.
- 84 Javier D. Fernández, Jürgen Umbrich, Axel Polleres, and Magnus Knuth. Evaluating query and storage strategies for RDF archives. *Semantic Web*, 10(2):247–291, 2019. doi:10.3233/SW-180309.
- 85 Adrià Fernández-Torras, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque. *Nature Communications*, 13, sep 2022. doi:10.1038/s41467-022-33026-0.
- 86 Nicolas Ferranti, Axel Polleres, Jairo Francisco de Souza, and Shqiponja Ahmetaj. Formalizing property constraints in wikidata. In Lucie-

Aimée Kaffee, Simon Razniewski, Gabriel Amaral, and Kholoud Saad Alghamdi, editors, Proceedings of the 3rd Wikidata Workshop 2022 co-located with the 21st International Semantic Web Conference (ISWC2022), Virtual Event, Hanghzou, China, October 2022, volume 3262 of CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3262/paper1.pdf.

- 87 Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75-174, feb 2010. doi: 10.1016/j.physrep.2009.11.002.
- 88 Michael Galkin, Zhaocheng Zhu, Hongyu Ren, and Jian Tang. Inductive logical query answering in knowledge graphs. In Advances in Neural Information Processing Systems, 2022. doi:10.48550/ ARXIV.2210.08008.
- 89 Mikhail Galkin, Etienne G. Denis, Jiapeng Wu, and William L. Hamilton. Nodepiece: Compositional and parameter-efficient representations of large knowledge graphs. In The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net, 2022. URL: https: //openreview.net/forum?id=xMJWUKJnFSw.
- 90 João Gama, Indre Zliobaite, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. ACM Computing Surveys, 46(4):44:1–44:37, 2014. doi:10.1145/ 2523813.
- 91 Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Modelling ontology evaluation and validation. In York Sure and John Domingue, editors, The Semantic Web: Research and Applications, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings, volume 4011 of Lecture Notes in Computer Science (LNCS), pages 140–154. Springer, 2006. doi:10.1007/11762256_ 13.
- 92 Shi Gao, Jiaqi Gu, and Carlo Zaniolo. RDF-TX: A Fast, User-Friendly System for Querying the History of RDF Knowledge Bases. In 19th International Conference on Extending Database Technology (EDBT), 2016. doi:10.5441/002/EDBT.2016. 26.
- 93 Robert Geirhos, Kristof Meding, and Felix A. Wichmann. Beyond accuracy: quantifying trial-bytrial behaviour of CNNs and humans by measuring error consistency. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 2020. doi:10.48550/arXiv.2006.16736.
- 94 Genet Asefa Gesese, Russa Biswas, and Harald Sack. A Comprehensive Survey of Knowledge Graph Embeddings with Literals: Techniques and Applications. In Mehwish Alam, Davide Buscaldi, Michael Cochez, Francesco Osborne, Diego Reforgiato Recupero, and Harald Sack, editors, Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019) Co-located with the 16th Extended Semantic Web Conference 2019 (ESWC 2019), Portoroz, Slovenia, June 2,

2019, volume 2377 of CEUR Workshop Proceedings, pages 31-40. CEUR-WS.org, 2019. URL: https://ceur-ws.org/Vol-2377/paper_4.pdf.

- 95 Birte Glimm, Aidan Hogan, Markus Krötzsch, and Axel Polleres. OWL: yet to arrive on the web of data? In Christian Bizer, Tom Heath, Tim Berners-Lee, and Michael Hausenblas, editors, WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012, volume 937 of CEUR Workshop Proceedings. CEUR-WS.org, 2012. URL: https://ceur-ws. org/Vol-937/ldow2012-paper-16.pdf.
- 96 Rishab Goel, Seyed Mehran Kazemi, Marcus A. Brubaker, and Pascal Poupart. Diachronic embedding for temporal knowledge graph completion. In The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, new york, NY, USA, february 7-12, 2020, pages 3988–3995. AAAI Press, 2020. doi:10.1609/AAAI.V34I04.5815.
- 97 Rafael S. Gonçalves, Matthew Horridge, Rui Li, Yu Liu, Mark A. Musen, Csongor I. Nyulas, Evelyn Obamos, Dhananjay Shrouty, and David Temple. Use of owl and semantic web technologies at pinterest. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web – ISWC 2019*, pages 418–435, Cham, 2019. Springer International Publishing. doi: 10.1007/978-3-030-30796-7_26.
- 98 Thomas Gottron and Christian Gottron. Perplexity of Index Models over Evolving Linked Data. In The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings 11, volume 8465, pages 161–175. Springer, 2014. doi:10.1007/978-3-319-07443-6_12.
- 99 Simon Gottschalk and Elena Demidova. Eventkg - the hub of event knowledge on the web - and biographical timeline generation. *Semantic Web*, 10(6):1039–1070, 2019. doi:10.3233/SW-190355.
- 100 Fabio Grandi. T-SPARQL: A tsql2-like temporal query language for RDF. In Mirjana Ivanovic, Bernhard Thalheim, Barbara Catania, and Zoran Budimac, editors, Local Proceedings of the Fourteenth East-European Conference on Advances in Databases and Information Systems, Novi Sad, Serbia, September 20-24, 2010, volume 639 of CEUR Workshop Proceedings, pages 21–30. CEUR-WS.org, 2010. URL: https://ceur-ws.org/Vol-639/021-grandi.pdf.
- 101 OWL Working group. OWL 2 web ontology language overview (second edition). Technical report, The World Wide Web Consortium (W3C), dec 2012. URL: http://www.w3.org/TR/2012/ REC-owl2-overview-20121211/.
- 102 Ramanathan V Guha, Dan Brickley, and Steve Macbeth. Schema.org: evolution of structured data on the web. Communications of the ACM, 59(2):44-51, 2016. doi:10.1145/2844544.
- 103 Claudio Gutierrez, Carlos A. Hurtado, and Alejandro Vaisman. Introducing Time into RDF. *IEEE Transactions on Knowledge and Data Engineering*, 19(2):207–218, 2007. doi:10.1109/TKDE. 2007.34.

- 104 Armin Haller, Javier D. Fernández, Maulik R. Kamdar, and Axel Polleres. What Are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web. Journal of Data and Information Quality, 12(2):1–34, jun 2020. doi:10.1145/3369875.
- 105 Armin Haller, Axel Polleres, Daniil Dobriy, Nicolas Ferranti, and Sergio J. Rodríguez Méndez. An analysis of links in Wikidata. In 19th European Semantic Web Conference, ESWC 2022. Springer, may 2022. doi:10.1007/978-3-031-06981-9_2.
- 106 Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. Knowledge Transfer for Out-of-Knowledge-Base Entities : A Graph Neural Network Approach. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pages 1802–1808, Melbourne, Australia, aug 2017. International Joint Conferences on Artificial Intelligence Organization. doi:10.24963/IJCAI.2017/250.
- 107 Olaf Hartig, Pierre-Antoine Champin, Gregg Kellogg, and Andy Seaborne. RDF-star and SPARQL-star. Technical report, W3C RDF-DEV Community Group, dec 2021. URL: https://w3c.github.io/rdf-star/cg-spec/2021-12-17.html.
- 108 Michael Hartung, James F. Terwilliger, and Erhard Rahm. Recent Advances in Schema and Ontology Evolution. In Schema Matching and Mapping, pages 149–190. Springer, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-16518-4_6.
- 109 Daniel Hernández, Aidan Hogan, and Markus Krötzsch. Reifying RDF: What Works Well With Wikidata? In Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2015), volume 1457 of CEUR Workshop Proceedings. CEUR-WS.org, 2015. URL: https://ceur-ws.org/Vol-1457/ SSWS2015_paper3.pdf.
- 110 Thomas Hoch, Bernhard Heinzl, Gerald Czech, Maqbool Khan, Philipp Waibel, Stefan Bachhofner, Elmar Kiesling, and Bernhard Moser. Teaming.ai: Enabling human-ai teaming intelligence in manufacturing. In Martin Zelm, Andrés Boza, Ramona Diana León, and Raúl Rodríguez-Rodríguez, editors, Proceedings of Interoperability for Enterprise Systems and Applications Workshops colocated with 11th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA 2022), Valencia, Spain, March 23-25, 2022, volume 3214 of CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https: //ceur-ws.org/Vol-3214/WS5Paper6.pdf.
- 111 Marvin Hofer, Sebastian Hellmann, Milan Dojchinovski, and Johannes Frey. The new DBpedia release cycle: Increasing agility and efficiency in knowledge extraction workflows. In Eva Blomqvist, Paul Groth, Victor de Boer, Tassilo Pellegrini, Mehwish Alam, Tobias Käfer, Peter Kieseberg, Sabrina Kirrane, Albert Meroño-Peñuela, and Harshvardhan J. Pandit, editors, Semantic Systems. In the Era of Knowledge Graphs 16th International Conference on Semantic Systems, SEMANTICS 2020, Amsterdam, The Netherlands, September 7-10, 2020, Proceedings, volume 12378 of Lecture Notes in Computer Science

(LNCS), pages 1–18. Springer, 2020. doi:10.1007/978-3-030-59833-4_1.

- 112 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. ACM Computing Surveys, 54(4):71:1–71:37, jul 2021. doi:10.1145/3447772.
- 113 Aidan Hogan, Antoine Zimmermann, Jürgen Umbrich, Axel Polleres, and Stefan Decker. Scalable and distributed methods for entity matching, consolidation and disambiguation over linked data corpora. Journal of Web Semantics (JWS), 10:76–110, jan 2012. doi:10.1016/J.WEBSEM.2011.11.002.
- 114 Pan Hu, Boris Motik, and Ian Horrocks. Modular materialisation of datalog programs. Artificial Intelligence, 308:103726, 2022. doi:10.1016/ J.ARTINT.2022.103726.
- 115 Zhisheng Huang and Heiner Stuckenschmidt. Reasoning with multi-version ontologies: A temporal logic approach. In Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, editors, The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings, volume 3729 of Lecture Notes in Computer Science (LNCS), pages 398–412. Springer, 2005. doi:10.1007/11574620_30.
- 116 Dilshod Ibragimov, Katja Hose, Torben Bach Pedersen, and Esteban Zimányi. Towards exploratory OLAP over linked open data A case study. In Malú Castellanos, Umeshwar Dayal, Torben Bach Pedersen, and Nesime Tatbul, editors, Enabling Real-Time Business Intelligence International Workshops, BIRTE 2013, Riva del Garda, Italy, August 26, 2013, and BIRTE 2014, Hangzhou, China, September 1, 2014, Revised Selected Papers, volume 206 of Lecture Notes in Business Information Processing, pages 114–132. Springer, 2014. doi:10.1007/978-3-662-46839-5_8.
- 117 Ihab F. Ilyas, Theodoros Rekatsinas, Vishnu Konda, Jeffrey Pound, Xiaoguang Qi, and Mohamed A. Soliman. Saga: A platform for continuous construction and serving of knowledge at scale. In Zachary G. Ives, Angela Bonifati, and Amr El Abbadi, editors, SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 17, 2022, pages 2259–2272. ACM, 2022. doi:10.1145/3514221.3526049.
- 118 Robert Isele, Jürgen Umbrich, Christian Bizer, and Andreas Harth. LDspider: An Open-source Crawling Framework for the Web of Linked Data. In Axel Polleres and Huajun Chen, editors, Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts, Shanghai, China, November 9, 2010, volume 658 of CEUR Workshop Proceedings. CEUR-WS.org, 2010. URL: https://ceur-ws.org/Vol-658/paper495.pdf.
- 119 Ousmane Issa, Angela Bonifati, and Farouk Toumani. Evaluating top-k queries with inconsistency degrees. Proceedings of the VLDB Endowment, 13(11):2146-2158, 2020. doi:10.14778/3407790. 3407815.

- 120 Ousmane Issa, Angela Bonifati, and Farouk Toumani. INCA: inconsistency-aware data profiling and querying. In SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021, pages 2745–2749, 2021. doi:10.1145/3448016.3452760.
- 121 Rebecca Jackson, Nicolas Matentzoglu, James A Overton, Randi Vita, James P Balhoff, Pier Luigi Buttigieg, Seth Carbon, Melanie Courtot, Alexander D Diehl, Damion M Dooley, William D Duncan, Nomi L Harris, Melissa A Haendel, Suzanna E Lewis, Darren A Natale, David Osumi-Sutherland, Alan Ruttenberg, Lynn M Schriml, Barry Smith, Christian J Stoeckert Jr., Nicole A Vasilevsky, Ramona L Walls, Jie Zheng, Christopher J Mungall, and Bjoern Peters. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. Database Journal on Biological Databases Curation, 2021, 2021. doi:10.1093/DATABASE/BAAB069.
- 122 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Net*works and Learning Systems, 33(2):494–514, feb 2022. doi:10.1109/TNNLS.2021.3070843.
- 123 Zhen Jia, Abdalghani Abujabal, Rishiraj Saha Roy, Jannik Strötgen, and Gerhard Weikum. TEQUILA: Temporal question answering over knowledge bases. In Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang, editors, Proceedings of the 27th ACM international conference on information and knowledge management, CIKM 2018, torino, italy, october 22-26, 2018, pages 1807–1810. ACM, 2018. doi:10.1145/3269206. 3269247.
- 124 Zhen Jia, Soumajit Pramanik, Rishiraj Saha Roy, and Gerhard Weikum. Complex temporal question answering on knowledge graphs. In Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong, editors, CIKM '21: The 30th ACM international conference on information and knowledge management, virtual event, queensland, australia, november 1 - 5, 2021, pages 792-802. ACM, 2021. doi:10.1145/3459637. 3482416.
- 125 Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui. Encoding temporal information for time-aware link prediction. In Jian Su, Xavier Carreras, and Kevin Duh, editors, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, pages 2350–2354. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1260.
- 126 Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Ian Horrocks, and Rafael Berlanga Llavori. Supporting concurrent ontology development: Framework, algorithms and tool. *Data & Knowledge Engineering*, 70(1):146–164, 2011. doi:10.1016/J.DATAK. 2010.10.001.

11:52 How Does Knowledge Evolve in Open Knowledge Graphs?

- 127 Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, Proceedings of the 2020 conference on empirical methods in natural language processing, EMNLP 2020, online, november 16-20, 2020, pages 6669–6683. Association for Computational Linguistics, 2020. doi: 10.18653/V1/2020.EMNLP-MAIN.541.
- 128 M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2):025102, feb 2011. doi:10.1103/PhysRevE. 83.025102.
- 129 Abbas Keshavarzi and Krys J. Kochut. Kgdiff: Tracking the evolution of knowledge graphs. In 21st International Conference on Information Reuse and Integration for Data Science, IRI 2020, Las Vegas, NV, USA, August 11-13, 2020, pages 279– 286. IEEE, 2020. doi:10.1109/IRI49571.2020. 00047.
- 130 Evgeny Kharlamov, Sebastian Brandt, Martin Giese, Ernesto Jiménez-Ruiz, Steffen Lamparter, Christian Neuenstadt, Özgür Lütfü Özçep, Christoph Pinkel, Ahmet Soylu, Dmitriy Zheleznyakov, Mikhail Roshchin, Stuart Watson, and Ian Horrocks. Semantic access to siemens streaming data: the optique way. In Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), Bethlehem, PA, USA, October 11, 2015, volume 1486 of CEUR Workshop Proceedings. CEUR-WS.org, 2015. URL: https://ceur-ws.org/Vol-1486/paper_23.pdf.
- 131 Evgeny Kharlamov, Bernardo Cuenca Grau, Ernesto Jiménez-Ruiz, Steffen Lamparter, Gulnar Mehdi, Martin Ringsquandl, Yavor Nenov, Stephan Grimm, Mikhail Roshchin, and Ian Horrocks. Capturing Industrial Information Models with Ontologies and Constraints. In 15th International Semantic Web Conference, pages 325–343, 2016. doi:10.1007/978-3-319-46547-0_30.
- 132 Shahrzad Khayatbashi, Sebastián Ferrada, and Olaf Hartig. Converting property graphs to RDF: a preliminary study of the practical impact of different mappings. In Proceedings of the 5th ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA), pages 1–9, Philadelphia Pennsylvania, jun 2022. ACM. doi:10.1145/3534540.3534695.
- 133 Udayan Khurana and Amol Deshpande. Storing and Analyzing Historical Graph Data at Scale. In Evaggelia Pitoura, Sofian Maabout, Georgia Koutrika, Amélie Marian, Letizia Tanca, Ioana Manolescu, and Kostas Stefanidis, editors, Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016, pages 65–76. OpenProceedings.org, 2016. doi:10.5441/002/EDBT.2016. 09.

- 134 Jon M. Kleinberg. Navigation in a small world. Nature, 406(6798):845–845, aug 2000. doi:10. 1038/35022643.
- 135 Holger Knublauch and Dimitris Kontokostas. Shapes constraint language (shacl). Technical report, The World Wide Web Consortium (W3C), jul 2017. URL: https://www.w3.org/TR/2017/ REC-shacl-20170720/.
- 136 Boris Konev, Dirk Walther, and Frank Wolter. The Logical Difference Problem for Description Logic Terminologies. In Alessandro Armando, Peter Baumgartner, and Gilles Dowek, editors, Automated Reasoning, 4th International Joint Conference, IJCAR 2008, Sydney, Australia, August 12-15, 2008, Proceedings, volume 5195 of Lecture Notes in Computer Science (LNCS), pages 259–274. Springer, 2008. doi:10.1007/ 978-3-540-71070-7_21.
- 137 Markus Krötzsch. OWL 2 Profiles: An Introduction to Lightweight Ontology Languages. In Thomas Eiter and Thomas Krennwallner, editors, Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School 2012, Vienna, Austria, September 3-8, 2012. Proceedings, volume 7487 of Lecture Notes in Computer Science (LNCS), pages 112–183. Springer, 2012. doi:10.1007/ 978-3-642-33158-9_4.
- 138 Markus Krötzsch, Denny Vrandecic, and Max Völkel. Semantic MediaWiki. In Isabel F. Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Michael Uschold, and Lora Aroyo, editors, The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings, volume 4273 of Lecture Notes in Computer Science (LNCS), pages 935– 942. Springer, 2006. doi:10.1007/11926078_68.
- 139 Tobias Käfer, Ahmed Abdelrahman, Jürgen Umbrich, Patrick O'Byrne, and Aidan Hogan. Observing Linked Data Dynamics. In Philipp Cimiano, Óscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, The Semantic Web: Semantics and Big Data, 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings, volume 7882 of Lecture Notes in Computer Science (LNCS), pages 213–227. Springer, 2013. doi: 10.1007/978-3-642-38288-8_15.
- 140 Tobias Käfer, Jürgen Umbrich, Aidan Hogan, and Axel Polleres. DyLDO: Towards a Dynamic Linked Data Observatory. In Christian Bizer, Tom Heath, Tim Berners-Lee, and Michael Hausenblas, editors, WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012, volume 937 of CEUR Workshop Proceedings. CEUR-WS.org, 2012. URL: https:// ceur-ws.org/Vol-937/ldow2012-paper-14.pdf.
- 141 Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. Tensor decompositions for temporal knowledge base completion. In 8th international conference on learning representations, ICLR 2020, addis ababa, ethiopia, april 26-30, 2020. OpenReview.net, 2020. URL: https://openreview.net/forum?id=rke2P1BFwS.

- 142 Birger Lantow and Kurt Sandkuhl. An analysis of applicability using quality metrics for ontologies on ontology design patterns. *Intelligent Systems in* Accounting, Finance and Management, 22(1):81– 99, 2015. doi:10.1002/ISAF.1360.
- 143 Andri Lareida, Romana Pernischova, Bruno Bastos Rodrigues, and Burkhard Stiller. Abstracting .torrent content consumption into two-mode graphs and their projection to content networks (ConNet). In 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pages 151– 159, may 2017. doi:10.23919/INM.2017.7987275.
- 144 Ora Lassila, Michael Schmidt, Olaf Hartig, Brad Bebee, Dave Bechberger, Willem Broekema, Ankesh Khandelwal, Kelvin Lawrence, Carlos Manuel Lopez Enriquez, Ronak Sharda, and Bryan Thompson. The OneGraph vision: Challenges of breaking the graph model lock-in. Semantic Web, 14(1):125–134, nov 2022. doi:10. 3233/SW-223273.
- 145 Julien Leblay, Melisachew Wudage Chekol, and Xin Liu. Towards Temporal Knowledge Graph Embeddings with Arbitrary Time Precision. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, pages 685–694, New York, NY, USA, oct 2020. Association for Computing Machinery. doi:10.1145/3340531.3412028.
- 146 Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, and others. Dbpedia – a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195, 2015. doi:10.3233/SW-140134.
- 147 Jure Leskovec, Jon M. Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In Robert Grossman, Roberto J. Bayardo, and Kristin P. Bennett, editors, Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, Illinois, USA, August 21-24, 2005, pages 177–187. ACM, 2005. doi:10.1145/1081870.1081893.
- 148 Siyuan Liao, Shangsong Liang, Zaiqiao Meng, and Qiang Zhang. Learning Dynamic Embeddings for Temporal Knowledge Graphs. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21, pages 535–543, New York, NY, USA, mar 2021. Association for Computing Machinery. doi: 10.1145/3437963.3441741.
- 149 M. Lissandrini, D. Mottin, K. Hose, and T. B. Pedersen. Knowledge graph exploration systems: Are we lost? In Proc. of the 12th Conference on Innovative Data Systems Research, CIDR 2022, volume 22, pages 10-13, 2022. URL: https://www.cidrdb.org/cidr2022/ papers/p40-lissandrini.pdf.
- 150 Matteo Lissandrini, Davide Mottin, Themis Palpanas, and Yannis Velegrakis. Data Exploration Using Example-Based Methods. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2018. doi:10.2200/ S00881ED1V01Y201810DTM053.

- 151 Yu Liu, Wen Hua, Kexuan Xin, and Xiaofang Zhou. Context-Aware Temporal Knowledge Graph Embedding. In Reynold Cheng, Nikos Mamoulis, Yizhou Sun, and Xin Huang, editors, Web Information Systems Engineering – WISE 2019, Lecture Notes in Computer Science (LNCS), pages 583– 598, Cham, 2019. Springer International Publishing. doi:10.1007/978-3-030-34223-4_37.
- 152 Freddy Lécué, Simone Tallevi-Diotallevi, Jer Hayes, Robert Tucker, Veli Bicer, Marco Luca Sbodio, and Pierpaolo Tommasi. Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City. Journal of Web Semantics, 27-28:26–33, 2014. doi:10.1016/J.WEBSEM.2014.07.002.
- 153 Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. Yago3: A knowledge base from multilingual wikipedias. In 7th biennial conference on innovative data systems research. CIDR Conference, 2014. URL: http://cidrdb.org/cidr2015/ Papers/CIDR15_Paper1.pdf.
- 154 Sofía Maiolo, Lorena Etcheverry, and Adriana Marotta. Data profiling in property graph databases. ACM J. Data Inf. Qual., 12(4):20:1–20:27, 2020. doi:10.1145/3409473.
- 155 Alessandro Margara, Gianpaolo Cugola, Dario Collavini, and Daniele Dell'Aglio. Efficient temporal reasoning on streams of events with DOTR. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, The Semantic Web 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings, volume 10843 of Lecture Notes in Computer Science (LNCS), pages 384–399. Springer, 2018. doi:10.1007/978-3-319-93417-4_25.
- 156 Cynthia Matuszek, John Cabral, Michael Witbrock, and John DeOliveira. An introduction to the syntax and content of cyc. In Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-05, Stanford, California, USA, March 27-29, 2006, pages 44-49. AAAI, 2006. URL: http://www.aaai.org/Library/Symposia/Spring/2006/ss06-05-007.php.
- 157 Christian Meilicke. Alignment incoherence in ontology matching. PhD Thesis, University of Mannheim, 2011. URL: https://ub-madoc.bib. uni-mannheim.de/29351.
- 158 Albert Meroño-Peñuela, Efstratios Kontopoulos, Sándor Darányi, and Yiannis Kompatsiaris. A study of intensional concept drift in trending DBpedia concepts. In SEMANTICS workshops, volume 2063 of CEUR workshop proceedings. CEUR-WS.org, 2017. URL: https://ceur-ws. org/Vol-2063/dal-paper3.pdf.
- 159 Robert Meusel, Petar Petrovski, and Christian Bizer. The WebDataCommons Microdata, RDFa and Microformat Dataset Series. In Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig A. Knoblock, Denny Vrandecic, Paul Groth, Natasha F. Noy, Krzysztof Janowicz, and Carole A. Goble, editors, The Semantic Web - ISWC 2014 -

13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I, volume 8796 of Lecture Notes in Computer Science (LNCS), pages 277–292. Springer, 2014. doi:10.1007/978-3-319-11964-9_18.

- 160 Pasquale Minervini, Erik Arakelyan, Daniel Daza, and Michael Cochez. Complex query answering with neural link predictors (extended abstract). In Luc De Raedt, editor, Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022, pages 5309–5313. ijcai.org, 2022. doi:10.24963/IJCAI.2022/741.
- 161 Vera Zaychik Moffitt and Julia Stoyanovich. Temporal graph algebra. In Tiark Rompf and Alexander Alexandrov, editors, Proceedings of The 16th International Symposium on Database Programming Languages, DBPL 2017, Munich, Germany, September 1, 2017, pages 10:1–10:12. ACM, 2017. doi:10.1145/3122831.3122838.
- 162 Boris Motik, Yavor Nenov, Robert Piro, Ian Horrocks, and Dan Olteanu. Parallel Materialisation of Datalog Programs in Centralised, Main-Memory RDF Systems. In Carla E. Brodley and Peter Stone, editors, Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada, pages 129–137. AAAI Press, 2014. doi: 10.1609/AAAI.V28I1.8730.
- 163 Boris Motik, Yavor Nenov, Robert Edgar Felix Piro, and Ian Horrocks. Incremental Update of Datalog Materialisation: the Backward/Forward Algorithm. In Blai Bonet and Sven Koenig, editors, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA, pages 1560–1568. AAAI Press, 2015. doi:10.1609/AAAI.V29I1.9409.
- 164 Yavor Nenov, Robert Piro, Boris Motik, Ian Horrocks, Zhe Wu, and Jay Banerjee. RD-Fox: A Highly-Scalable RDF Store. In Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, and Steffen Staab, editors, The Semantic Web - ISWC 2015 - 14th International Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II, volume 9367 of Lecture Notes in Computer Science (LNCS), pages 3–20. Springer, 2015. doi:10.1007/978-3-319-25010-6_1.
- 165 Mark Newman, Albert-László Barabási, and Duncan J. Watts. The Structure and Dynamics of Networks. Princeton University Press, dec 2011. doi:10.1515/9781400841356.
- 166 Chifumi Nishioka and Ansgar Scherp. Temporal Patterns and Periodicity of Entity Dynamics in the Linked Open Data Cloud. In Ken Barker and José Manuél Gómez-Pérez, editors, Proceedings of the 8th International Conference on Knowledge Capture, K-CAP 2015, Palisades, NY, USA, October 7-10, 2015, pages 22:1–22:4. ACM, 2015. doi:10.1145/2815833.2816948.
- 167 Chifumi Nishioka and Ansgar Scherp. Informationtheoretic Analysis of Entity Dynamics on the Linked Open Data Cloud. In Elena Demidova, Stefan Dietze, Julian Szymanski, and John G.

Breslin, editors, Proceedings of the 3rd International Workshop on Dataset PROFIling and fEderated Search for Linked Data (PROFILES '16) co-located with the 13th ESWC 2016 Conference, Anissaras, Greece, May 30, 2016, volume 1597 of CEUR Workshop Proceedings. CEUR-WS.org, 2016. URL: https://ceur-ws.org/Vol-1597/ PROFILES2016_paper2.pdf.

- 168 Chifumi Nishioka and Ansgar Scherp. Keeping linked open data caches up-to-date by predicting the life-time of RDF triples. In Amit P. Sheth, Axel Ngonga, Yin Wang, Elizabeth Chang, Dominik Slezak, Bogdan Franczyk, Rainer Alt, Xiaohui Tao, and Rainer Unland, editors, *Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017*, pages 73– 80. ACM, 2017. doi:10.1145/3106426.3106463.
- 169 Chifumi Nishioka and Ansgar Scherp. Analysing the Evolution of Knowledge Graphs for the Purpose of Change Verification. In 12th IEEE International Conference on Semantic Computing, ICSC 2018, Laguna Hills, CA, USA, January 31 February 2, 2018, pages 25–32. IEEE Computer Society, 2018. doi:10.1109/ICSC.2018.00013.
- 170 Natasha F. Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale knowledge graphs: Lessons and challenges. ACM Queue, 17(2):20, 2019. doi: 10.1145/3329781.3332266.
- 171 Fabrizio Orlandi, Damien Graux, and Declan O'Sullivan. Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF. In 2021 IEEE 15th International Conference on Semantic Computing (ICSC), pages 233– 240, jan 2021. ISSN: 2325-6516. doi:10.1109/ ICSC50631.2021.00049.
- 172 Anthony M Orme, Haining Yao, and Letha H Etzkorn. Indicating ontology data quality, stability, and completeness throughout ontology evolution. Journal of Software Maintenance and Evolution: Research and Practice, 19(1):49–75, 2007. doi:10.1002/SMR.341.
- 173 Kristian Otte, Kristian Simoni Vestermark, Huan Li, and Daniele Dell'Aglio. Towards A question answering system over temporal knowledge graph embedding. In Mehwish Alam and Michael Cochez, editors, Proceedings of the workshop on deep learning for knowledge graphs (DL4KG 2022) co-located with the 21th international semantic web conference (ISWC 2022), virtual conference, online, october 24, 2022, volume 3342 of CEUR workshop proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3342/paper-1.pdf.
- 174 Anil Pacaci, Angela Bonifati, and M. Tamer Özsu. Regular path query evaluation on streaming graphs. In David Maier, Rachel Pottinger, An-Hai Doan, Wang-Chiew Tan, Abdussalam Alawini, and Hung Q. Ngo, editors, Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020, pages 1415–1430. ACM, 2020. doi:10.1145/3318464. 3389733.
- 175 Anil Pacaci, Angela Bonifati, and M. Tamer Özsu. Evaluating complex queries on streaming graphs. In 38th IEEE International Conference on Data

Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022, pages 272-285. IEEE, 2022. doi:10.1109/ICDE53745.2022.00025.

- 176 Heiko Paulheim and Aldo Gangemi. Serving dbpedia with DOLCE more than just adding a cherry on top. In Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, and Steffen Staab, editors, The Semantic Web ISWC 2015 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I, volume 9366 of Lecture Notes in Computer Science (LNCS), pages 180–196. Springer, 2015. doi:10.1007/978-3-319-25007-6_11.
- 177 Olivier Pelgrin, Luis Galárraga, and Katja Hose. Towards fully-fledged archiving for RDF datasets. Semantic Web Journal, 12(6):903–925, 2021. doi:10.3233/SW-210434.
- 178 Olivier Pelgrin, Ruben Taelman, Luis Galárraga, and Katja Hose. GLENDA: Querying RDF Archives with full SPARQL. In Proceedings of the 25th Extended Semantic Web Conference (ESWC), 2023. doi:10.1007/978-3-031-43458-7_14.
- 179 Olivier Pelgrin, Ruben Taelman, Luis Galárraga, and Katja Hose. Scaling Large RDF Archives To Very Long Histories. In 2023 IEEE 17th International Conference on Semantic Computing (ICSC), pages 41–48, Laguna Hills, CA, USA, feb 2023. IEEE. doi:10.1109/ICSC56153.2023.00013.
- 180 Thomas Pellissier Tanon, Gerhard Weikum, and Fabian Suchanek. YAGO 4: A Reason-able Knowledge Base. In Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez, editors, *The Semantic Web*, pages 583–596. Springer, 2020. doi:10.1007/ 978-3-030-49461-2_34.
- 181 Romana Pernisch, Daniele Dell'Aglio, and Abraham Bernstein. Toward Measuring the Resemblance of Embedding Models for Evolving Ontologies. In Proceedings of the 11th on Knowledge Capture Conference, pages 177–184, Virtual Event USA, dec 2021. ACM. doi:10.1145/3460210.3493540.
- 182 Romana Pernisch, Daniele Dell'Aglio, and Abraham Bernstein. Beware of the hierarchy An analysis of ontology evolution and the materialisation impact for biomedical ontologies. Journal of Web Semantics, 70:100658, jul 2021. doi: 10.1016/J.WEBSEM.2021.100658.
- 183 Romana Pernisch, Daniele Dell'Aglio, Mirko Serbak, Rafael S. Gonçalves, and Abraham Bernstein. Visualising the effects of ontology changes and studying their understanding with ChImp. *Journal of Web Semantics*, 74:100715, oct 2022. doi:10.1016/J.WEBSEM.2022.100715.
- 184 Romana Pernischová, Florian Ruosch, Daniele Dell'Aglio, and Abraham Bernstein. Stream processing: The matrix revolutions. In Thorsten Liebig, Achille Fokoue, and Zhe Wu, editors, Proceedings of the 12th International Workshop on Scalable Semantic Web Knowledge Base Systems co-located with 17th International Semantic Web Conference, SSWS@ISWC 2018, Monterey,

California, USA, October 9, 2018, volume 2179 of CEUR Workshop Proceedings, pages 15-27. CEUR-WS.org, 2018. URL: https://ceur-ws. org/Vol-2179/SSWS2018_paper2.pdf.

- 185 Matthew Perry, Prateek Jain, and Amit P. Sheth. SPARQL-ST: Extending SPARQL to Support Spatiotemporal Queries. In Naveen Ashish and Amit P. Sheth, editors, *Geospatial Semantics and the Semantic Web*, volume 12, pages 61–86, Boston, MA, 2011. Springer US. doi:10.1007/ 978-1-4419-9446-2_3.
- 186 Danh Le Phuoc, Minh Dao-Tran, Josiane Xavier Parreira, and Manfred Hauswirth. A native and adaptive approach for unified processing of linked streams and linked data. In Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, Abraham Bernstein, Lalana Kagal, Natasha Fridman Noy, and Eva Blomqvist, editors, The semantic web - ISWC 2011 - 10th international semantic web conference, bonn, germany, october 23-27, 2011, proceedings, part I, volume 7031 of Lecture Notes in Computer Science (LNCS), pages 370–388. Springer, 2011. doi:10.1007/978-3-642-25073-6_24.
- 187 Danh Le Phuoc, Thomas Eiter, and Anh Lê Tuán. A scalable reasoning and learning approach for neural-symbolic stream fusion. In AAAI, pages 4996–5005. AAAI Press, 2021. doi:10.1609/AAAI. V3516.16633.
- 188 Alessandro Piscopo, Lucie-Aimée Kaffee, Chris Phethean, and Elena Simperl. Provenance information in a collaborative knowledge graph: An evaluation of wikidata external references. In Claudia d'Amato, Miriam Fernández, Valentina A. M. Tamma, Freddy Lécué, Philippe Cudré-Mauroux, Juan F. Sequeda, Christoph Lange, and Jeff Heflin, editors, The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part I, volume 10587 of Lecture Notes in Computer Science (LNCS), pages 542–558. Springer, 2017. doi:10.1007/978-3-319-68288-4_32.
- 189 Alessandro Piscopo, Chris Phethean, and Elena Simperl. What makes a good collaborative knowledge graph: Group composition and quality in wikidata. In Giovanni Luca Ciampaglia, Afra J. Mashhadi, and Taha Yasseri, editors, Social Informatics 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I, volume 10539 of Lecture Notes in Computer Science (LNCS), pages 305–322. Springer, 2017. doi:10.1007/978-3-319-67217-5_19.
- 190 Alessandro Piscopo and Elena Simperl. Who models the world?: Collaborative ontology creation and user roles in wikidata. Proc. ACM Hum. Comput. Interact., 2(CSCW):141:1-141:18, 2018. doi:10.1145/3274410.
- 191 Alessandro Piscopo and Elena Simperl. Who Models the World?: Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings* of the ACM on Human-Computer Interaction, 2(CSCW):1–18, nov 2018. doi:10.1145/3274410.
- 192 Alessandro Piscopo and Elena Simperl. What we talk about when we talk about wikidata quality: a literature survey. In Björn Lundell, Jonas Gamalielsson, Lorraine Morgan, and

Gregorio Robles, editors, Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019, Skövde, Sweden, August 20-22, 2019, pages 17:1–17:11. ACM, 2019. doi:10.1145/ 3306446.3340822.

- 193 Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. Communities in Networks. Computing Research Repository (CoRR), abs/0902.3788, 2009. arXiv:0902.3788.
- 194 Giulia Preti, Matteo Lissandrini, Davide Mottin, and Yannis Velegrakis. Mining patterns in graphs with multiple weights. *Distributed Parallel Databases*, 39(2):281–319, 2021. doi:10.1007/ S10619-019-07259-W.
- 195 Eric Prud'hommeaux, Iovka Boneva, Jose Emilio Labra Gayo, and Gregg Kellogg. Shape expressions language 2.1. Technical report, W3C Shape Expressions Community Group, oct 2019. URL: http://shex.io/shex-semantics/.
- 196 Kashif Rabbani, Matteo Lissandrini, and Katja Hose. Extraction of validating shapes from very large knowledge graphs. Proceedings of the VLDB Endowment, 16(5):1023–1032, 2023. doi:10.14778/3579075.3579078.
- 197 Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. VLDB Journal, 10(4):334–350, 2001. doi: 10.1007/S007780100057.
- 198 Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames. In The Semantic Web-ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15, pages 177–185. Springer, 2016. doi:10.1007/978-3-319-46547-0_19.
- 199 Yuan Ren and Jeff Z. Pan. Optimising ontology stream reasoning with truth maintenance system. In Craig Macdonald, Iadh Ounis, and Ian Ruthven, editors, Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011, pages 831–836. ACM, 2011. doi:10.1145/2063576.2063696.
- 200 Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis. ACM Transactions on Knowledge Discovery from Data, 15(2):14:1–14:49, jan 2021. doi:10.1145/3424672.
- 201 Christopher Rost, Kevin Gómez, Matthias Täschner, Philip Fritzsche, Lucas Schons, Lukas Christ, Timo Adameit, Martin Junghanns, and Erhard Rahm. Distributed temporal graph analytics with GRADOOP. VLDB Journal, 31(2):375–401, 2022. doi:10.1007/S00778-021-00667-4.
- 202 Anisa Rula, Amrapali Zaveri, Elena Simperl, and Elena Demidova. Editorial: Special issue on quality assessment of knowledge graphs dedicated to the memory of amrapali zaveri. Journal of Data and Information Quality, 12(2), may 2020. doi:10.1145/3388748.
- 203 Ali Sadeghian, Mohammadreza Armandpour, Anthony Colas, and Daisy Zhe Wang. ChronoR:

Rotation Based Temporal Knowledge Graph Embedding. Proceedings of the AAAI Conference on Artificial Intelligence, 35(7):6471–6479, may 2021. Number: 7. doi:10.1609/AAAI.V3517.16802.

- 204 Christian Sageder. Annotating entities with finegrained types in austrian court decisions. In Further with Knowledge Graphs - Proceedings of the 17th International Conference on Semantic Systems, SEMANTICS 2017, Amsterdam, The Netherlands, September 6-9, 2021, volume 53 of Studies on the Semantic Web, pages 139–153. IOS Press, 2021. doi:10.3233/SSW210041.
- 205 Satya S. Sahoo, Christopher Thomas, Amit Sheth, William S. York, and Samir Tartir. Knowledge modeling and its application in life sciences: A tale of two ontologies. In Proceedings of the 15th International Conference on World Wide Web, WWW '06, pages 317–326, New York, NY, USA, 2006. Association for Computing Machinery. doi:10.1145/1135777.1135826.
- 206 Fabio Saracco, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini. Randomizing bipartite networks: the case of the World Trade Web. Scientific Reports, 5(1):10595, jun 2015. doi:10.1038/srep10595.
- 207 Cristina Sarasua, Alessandro Checco, Gianluca Demartini, Djellel Eddine Difallah, Michael Feldman, and Lydia Pintscher. The evolution of power and standard wikidata editors: Comparing editing behavior over time to predict lifespan and volume of edits. *Computer Supported Cooperative Work (CSCW)*, 28(5):843–882, 2019. doi:10.1007/S10606-018-9344-Y.
- 208 Lukas Schmelzeisen, Corina Dima, and Steffen Staab. Wikidated 1.0: An evolving knowledge graph dataset of wikidata's revision history. In Lucie-Aimée Kaffee, Simon Razniewski, and Aidan Hogan, editors, Proceedings of the 2nd Wikidata Workshop (Wikidata 2021) co-located with the 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24, 2021, volume 2982 of CEUR Workshop Proceedings. CEUR-WS.org, 2021. URL: https://ceur-ws. org/Vol-2982/paper-11.pdf.
- 209 Stefan Schmid, Cory Henson, and Tuan Tran. Using knowledge graphs to search an enterprise data lake. In Pascal Hitzler, Sabrina Kirrane, Olaf Hartig, Victor de Boer, Maria-Esther Vidal, Maria Maleshkova, Stefan Schlobach, Karl Hammar, Nelia Lasierra, Steffen Stadtmüller, Katja Hose, and Ruben Verborgh, editors, The Semantic Web: ESWC 2019 Satellite Events -ESWC 2019 Satellite Events, Portorož, Slovenia, June 2-6, 2019, Revised Selected Papers, volume 11762 of Lecture Notes in Computer Science (LNCS), pages 262–266. Springer, 2019. doi: 10.1007/978-3-030-32327-1_46.
- 210 Patrik Schneider, Daniel Alvarez-Coello, Anh Le-Tuan, Manh Nguyen Duc, and Danh Le Phuoc. Stream reasoning playground. In Paul Groth, Maria-Esther Vidal, Fabian M. Suchanek, Pedro A. Szekely, Pavan Kapanipathi, Catia Pesquita, Hala Skaf-Molli, and Minna Tamper, editors, The Semantic Web 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29 June 2, 2022, Proceedings, volume 13261

of Lecture Notes in Computer Science (LNCS), pages 406-424. Springer, 2022. doi:10.1007/ 978-3-031-06981-9_24.

- 211 Vedran Sekara, Arkadiusz Stopczynski, and Sune Lehmann. Fundamental structures of dynamic social networks. *Proceedings of the national* academy of sciences, 113(36):9977–9982, 2016. doi:10.1073/pnas.1602803113.
- 212 Tong Shen, Fu Zhang, and Jingwei Cheng. A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems*, 255:109597, nov 2022. doi:10.1016/J.KNOSYS.2022.109597.
- 213 Ying Shen, Daoyuan Chen, Buzhou Tang, Min Yang, and Kai Lei. EAPB: entropy-aware pathbased metric for ontology quality. *Journal of Biomedical Semantics*, 9(1):20, 2018. doi:10.1186/ s13326-018-0188-7.
- 214 Robert H Shumway, David S Stoffer, and David S Stoffer. Time series analysis and its applications, volume 3. Springer, 2000. doi:10.1007/ 978-3-319-52452-8.
- 215 Philipp Singer, Denis Helic, Behnam Taraghi, and Markus Strohmaier. Detecting Memory and Structure in Human Navigation Patterns Using Markov Chain Models of Varying Order. *PLoS ONE*, 9(7):e102070, jul 2014. doi:10.1371/journal. pone.0102070.
- 216 Alessandro Solimando, Ernesto Jiménez-Ruiz, and Giovanna Guerrini. Minimizing conservativity violations in ontology alignments: algorithms and evaluation. Knowledge and Information Systems (KAIS), 51(3):775–819, 2017. doi:10.1007/ S10115-016-0983-3.
- 217 Hyun-Je Song and Seong-Bae Park. Enriching Translation-Based Knowledge Graph Embeddings Through Continual Learning. *IEEE Access*, 6:60489–60497, 2018. Conference Name: IEEE Access. doi:10.1109/ACCESS.2018.2874656.
- 218 T. G. Stavropoulos, S. Andreadis, E. Kontopoulos, and I. Kompatsiaris. SemaDrift: A hybrid method and visual tools to measure semantic drift in ontologies. *Journal of Web Semantics*, jun 2018. doi:10.1016/j.websem.2018.05.001.
- 219 Thanos G. Stavropoulos, Efstratios Kontopoulos, Albert Meroño-Peñuela, Stavros Tachos, Stelios Andreadis, and Yiannis Kompatsiaris. Crossdomain semantic drift measurement in ontologies using the semadrift tool and metrics. In Jeremy Debattista, Jürgen Umbrich, Javier D. Fernández, Anisa Rula, Amrapali Zaveri, Anastasia Dimou, and Wouter Beek, editors, Joint proceedings of the 3rd Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW 2017) and the 4th Workshop on Linked Data Quality (LDQ 2017) co-located with 14th European Semantic Web Conference (ESWC 2017), Portorož, Slovenia, May 28th-29th, 2017, volume 1824 of CEUR Workshop Proceedings, pages 59–72. CEUR-WS.org, 2017. URL: https://ceur-ws. org/Vol-1824/mepdaw_paper_5.pdf.
- 220 Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web, WWW '07, pages 697–706, New York, NY, USA, 2007. Association

for Computing Machinery. doi:10.1145/1242572. 1242667.

- 221 Xibo Sun, Shixuan Sun, Qiong Luo, and Bingsheng He. An in-depth study of continuous subgraph matching. *Proceedings of the VLDB Endowment*, 15(7):1403–1416, 2022. doi:10.14778/ 3523210.3523218.
- 222 Ruben Taelman, Thibault Mahieu, Martin Vanbrabant, and Ruben Verborgh. Optimizing storage of RDF archives using bidirectional delta chains. *Semantic Web*, 13(4):705–734, may 2022. doi:10.3233/SW-210449.
- 223 Ruben Taelman, Miel Vander Sande, Joachim Van Herwegen, Erik Mannens, and Ruben Verborgh. Triple storage for random-access versioned querying of RDF archives. Journal of Web Semantics, 54:4–28, jan 2019. doi:10.1016/J.WEBSEM.2018.08.001.
- 224 Thomas Pellissier Tanon, Camille Bourgaux, and Fabian M. Suchanek. Learning how to correct a knowledge base from the edit history. In Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia, editors, *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 1465–1475. ACM, 2019. doi:10.1145/3308558.3313584.
- 225 Thomas Pellissier Tanon and Fabian M. Suchanek. Querying the edit history of wikidata. In Pascal Hitzler, Sabrina Kirrane, Olaf Hartig, Victor de Boer, Maria-Esther Vidal, Maria Maleshkova, Stefan Schlobach, Karl Hammar, Nelia Lasierra, Steffen Stadtmüller, Katja Hose, and Ruben Verborgh, editors, The Semantic Web: ESWC 2019 Satellite Events - ESWC 2019 Satellite Events, Portorož, Slovenia, June 2-6, 2019, Revised Selected Papers, volume 11762 of Lecture Notes in Computer Science (LNCS), pages 161–166. Springer, 2019. doi:10.1007/978-3-030-32327-1_32.
- 226 Jonas Tappolet and Abraham Bernstein. Applied temporal RDF: efficient temporal querying of RDF data with SPARQL. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, editors, The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31-June 4, 2009, Proceedings, volume 5554 of Lecture Notes in Computer Science (LNCS), pages 308–322. Springer, 2009. doi:10.1007/978-3-642-02121-3 25.
- 227 Samir Tartir, Ismailcem Arpinar, and Amit Sheth. Ontological Evaluation and Validation, pages 115–130. Springer, sep 2010. doi:10.1007/ 978-90-481-8847-5_5.
- 228 Yi Tay, Anh Luu, and Siu Cheung Hui. Nonparametric estimation of multiple embeddings for link prediction on dynamic knowledge graphs. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA, 31(1), feb 2017. doi:10.1609/aaai.v31i1.10685.
- 229 Christoph Tempich and Raphael Volz. Towards a benchmark for semantic web reasoners - an analysis of the DAML ontology library. In York Sure and Óscar Corcho, editors, *EON2003, Evaluation*

of Ontology-based Tools, Proceedings of the 2nd International Workshop on Evaluation of Ontologybased Tools held at the 2nd International Semantic Web Conference ISWC 2003, 20th October 2003 (Workshop day), Sundial Resort, Sanibel Island, Florida, USA, volume 87 of CEUR Workshop Proceedings. CEUR-WS.org, 2003. URL: https: //ceur-ws.org/Vol-87/E0N2003_Tempich.pdf.

- 230 Mattias Tiger and Fredrik Heintz. Stream reasoning using temporal logic and predictive probabilistic state models. In *TIME*, pages 196–205. IEEE Computer Society, 2016. doi:10.1109/TIME.2016.28.
- 231 Stein L. Tomassen and Darijus Strasunskas. An ontology-driven approach to web search: Analysis of its sensitivity to ontology quality and search tasks. In Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services, iiWAS '09, pages 130–138, New York, NY, USA, 2009. Association for Computing Machinery. doi:10.1145/1806338. 1806368.
- 232 Riccardo Tommasini, Pieter Bonte, Femke Ongenae, and Emanuele Della Valle. RSP4J: an API for RDF stream processing. In Ruben Verborgh, Katja Hose, Heiko Paulheim, Pierre-Antoine Champin, Maria Maleshkova, Óscar Corcho, Petar Ristoski, and Mehwish Alam, editors, The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings, volume 12731 of Lecture Notes in Computer Science (LNCS), pages 565–581. Springer, 2021. doi:10.1007/978-3-030-77385-4_34.
- 233 Riccardo Tommasini, Pieter Bonte, Emanuele Della Valle, Femke Ongenae, and Filip De Turck. A query model for ontology-based event processing over RDF streams. In Catherine Faron-Zucker, Chiara Ghidini, Amedeo Napoli, and Yannick Toussaint, editors, Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings, volume 11313 of Lecture Notes in Computer Science (LNCS), pages 439–453. Springer, 2018. doi:10.1007/978-3-030-03667-6_28.
- 234 Riccardo Tommasini, Davide Calvaresi, and Jean-Paul Calbimonte. Stream reasoning agents: Blue sky ideas track. In Edith Elkind, Manuela Veloso, Noa Agmon, and Matthew E. Taylor, editors, Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019, pages 1664–1680. International Foundation for Autonomous Agents and Multiagent Systems, 2019. URL: http://dl.acm.org/citation.cfm? id=3331894.
- 235 Riccardo Tommasini, Mohamed Ragab, Alessandro Falcetta, Emanuele Della Valle, and Sherif Sakr. Bootstrapping the publication of linked data streams. In Mari Carmen Suárez-Figueroa, Gong Cheng, Anna Lisa Gentile, Christophe Guéret, C. Maria Keet, and Abraham Bernstein, editors, Proceedings of the ISWC 2019 Satellite Tracks (Posters & Demonstrations, Industry, and Outrageous Ideas) co-located with 18th International Semantic Web Conference (ISWC 2019), Auck-

land, New Zealand, October 26-30, 2019, volume 2456 of CEUR Workshop Proceedings, pages 29-32. CEUR-WS.org, 2019. URL: https://ceur-ws. org/Vol-2456/paper8.pdf.

- 236 Riccardo Tommasini, Mohamed Ragab, Alessandro Falcetta, Emanuele Della Valle, and Sherif Sakr. A first step towards a streaming linked data life-cycle. In Jeff Z. Pan, Valentina A. M. Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal, editors, The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II, volume 12507 of Lecture Notes in Computer Science (LNCS), pages 634–650. Springer, 2020. doi:10.1007/978-3-030-62466-8_39.
- 237 Riccardo Tommasini, Sherif Sakr, Emanuele Della Valle, and Hojjat Jafarpour. Declarative languages for big streaming data. In Angela Bonifati, Yongluan Zhou, Marcos Antonio Vaz Salles, Alexander Böhm, Dan Olteanu, George H. L. Fletcher, Arijit Khan, and Bin Yang, editors, Proceedings of the 23rd International Conference on Extending Database Technology, EDBT 2020, Copenhagen, Denmark, March 30 April 02, 2020, pages 643–646. OpenProceedings.org, 2020. doi: 10.5441/002/EDBT.2020.84.
- 238 Efthymia Tsamoura, David Carral, Enrico Malizia, and Jacopo Urbani. Materializing Knowledge Bases via Trigger Graphs. Proceedings of the VLDB Endowment, 14(6):943–956, 2021. doi:10.14778/3447689.3447699.
- 239 Evangelia Tsoukanara, Georgia Koloniari, and Evaggelia Pitoura. Graph-tempo: An aggregation framework for evolving graphs. In Proceedings 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023. doi:10.48786/edbt.2023.79.
- 240 Stella Verkijk, Ritte Roothaert, Romana Pernisch, and Stefan Schlobach. Do you catch my drift? In Proceedings of the 12th on Knowledge Capture Conference, Pensecola, FL, USA, dec 2023. ACM. doi:10.1145/3587259.3627555.
- 241 Raphael Volz, Steffen Staab, and Boris Motik. Incrementally Maintaining Materializations of Ontologies Stored in Logic Databases. *Journal on Data Semantics II*, 2:1–34, 2005. doi:10.1007/ 978-3-540-30567-5_1.
- 242 Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10):78–85, 2014. doi: 10.1145/2629489.
- 243 Przemyslaw Andrzej Walega, Mark Kaminski, Dingmin Wang, and Bernardo Cuenca Grau. Stream reasoning with DatalogMTL. Journal of Web Semantics, 76:100776, 2023. doi:10.1016/J. WEBSEM.2023.100776.
- 244 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, dec 2017. doi:10.1109/TKDE. 2017.2754499.
- 245 Richard Y Wang, Mostapha Ziad, and Yang W Lee. Data quality, volume 23. Springer Science & Business Media, 2006. doi:10.1007/b116303.

- 246 Shenghui Wang, Stefan Schlobach, and Michel Klein. Concept drift and how to identify it. Journal of Web Semantics, 9(3):247–265, sep 2011. doi:10.1016/J.WEBSEM.2011.05.003.
- 247 Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440-442, jun 1998. doi:10.1038/30918.
- 248 Patricia L. Whetzel, Natalya Fridman Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen. Bio-Portal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*, 39(Web-Server-Issue):541–545, jul 2011. doi:10.1093/NAR/ GKR469.
- 249 Tianxing Wu, Arijit Khan, Melvin Yong, Guilin Qi, and Meng Wang. Efficiently embedding dynamic knowledge graphs. *Knowledge-Based Sys*tems, 250:109124, 2022. doi:10.1016/J.KNOSYS. 2022.109124.
- 250 Zhe Yang, Dalu Zhang, and Chuan Ye. Evaluation metrics for ontology complexity and evolution analysis. In 2006 IEEE International Conference on e-Business Engineering (ICEBE 2006), 24-26 October 2006, Shanghai, China, pages 162–170. IEEE Computer Society, 2006. doi:10.1109/ICEBE.2006.48.
- 251 Carlo Zaniolo, Shi Gao, Maurizio Atzori, Muhao Chen, and Jiaqi Gu. User-friendly temporal queries on historical knowledge bases. *Information* and Computation, 259(3):444–459, 2018. doi: 10.1016/J.IC.2017.08.012.
- 252 Dalu Zhang, Chuan Ye, and Zhe Yang. An evaluation method for ontology complexity analysis in ontology evolution. In Steffen Staab and Vojtech Svátek, editors, Managing Knowledge in a World of Networks, 15th International Conference, EKAW 2006, Podebrady, Czech Republic, October

- 253 Fu Zhang, Ke Wang, Zhiyin Li, and Jingwei Cheng. Temporal data representation and querying based on RDF. *IEEE Access*, 7:85000-85023, 2019. doi:10.1109/ACCESS.2019.2924550.
- 254 Lei (Nico) Zheng, Christopher M. Albano, Neev M. Vora, Feng Mai, and Jeffrey V. Nickerson. The roles bots play in wikipedia. Proceedings of the ACM on Human Computer Interaction (HCI), 3(CSCW), nov 2019. doi:10.1145/3359317.
- 255 Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural Bellman-Ford networks: A general graph neural network framework for link prediction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 29476–29490. Curran Associates, Inc., 2021. doi: 10.48550/arXiv.2106.06935.
- 256 Antoine Zimmermann, Nuno Lopes, Axel Polleres, and Umberto Straccia. A general framework for representing, reasoning and querying with annotated Semantic Web data. *Journal of Web Semantics*, 11:72–95, mar 2012. doi:10.1016/J. WEBSEM.2011.08.006.
- 257 Mussab Zneika, Claudio Lucchese, Dan Vodislav, and Dimitris Kotzinos. Summarizing linked data RDF graphs using approximate graph pattern mining. In Evaggelia Pitoura, Sofian Maabout, Georgia Koutrika, Amélie Marian, Letizia Tanca, Ioana Manolescu, and Kostas Stefanidis, editors, Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016, pages 684–685. Open-Proceedings.org, 2016. doi:10.5441/002/EDBT. 2016.86.

Structural Summarization of Semantic Graphs Using Quotients

Ansgar Scherp 🖂 🏠 🝺 Ulm University, Germany

David Richerby 🖂 🗅 University of Essex, UK

Till Blume ⊠© Ernst and Young Research, Berlin, Germany

Michael Cochez 🖂 🏠 💿

Vrije Universiteit Amsterdam, The Netherlands Elsevier Discovery Lab, Amsterdam, The Netherlands

JannikRau 🖂 回

Ulm University, Germany

— Abstract -

Graph summarization is the process of computing a compact version of an input graph while preserving chosen features of its structure. We consider semantic graphs where the features include edge labels and label sets associated with a vertex. Graph summaries are typically much smaller than the original graph. Applications that depend on the preserved features can perform their tasks on the summary, but much faster or with less memory overhead, while producing the same outcome as if they were applied on the original graph.

In this survey, we focus on structural summaries based on quotients that organize vertices in equivalence classes of shared features. Structural summaries are particularly popular for semantic graphs and have the advantage of defining a precise graphbased output. We consider approaches and algorithms for both static and temporal graphs. A common example of quotient-based structural summaries is bisimulation, and we discuss this in detail. While there exist other surveys on graph summarization, to the best of our knowledge, we are the first to bring in a focused discussion on quotients, bisimulation, and their relation. Furthermore, structural summarization naturally connects well with formal logic due to the discrete structures considered. We complete the survey with a brief description of approaches beyond structural summaries.

2012 ACM Subject Classification Mathematics of computing \rightarrow Graph algorithms; Theory of computation \rightarrow Graph algorithms analysis; General and reference \rightarrow Surveys and overviews

Keywords and phrases graph summarization, quotients, stratified bisimulation

Digital Object Identifier 10.4230/TGDK.1.1.12

Category Survey

Funding Ansgar Scherp: Co-funded by the CodeInspector project (No. 504226141) of the DFG, German Research Foundation.

Michael Cochez: Partially funded by the Graph-Massivizer project, funded by the Horizon Europe programme of the European Union (grant 101093202).

Received 2023-07-03 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge



© Ansgar Scherp, David Richerby, Till Blume, Michael Cochez, and Jannik Rau; licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 12, pp. 12:1–12:25 Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

12:2 Structural Summarization of Semantic Graphs Using Quotients

1 Introduction

Representing data as a graph is increasingly popular [45, 48], though the idea dates back to at least the 1960s [5]. The strength of graphs as a data representation lies in their general applicability and their focus on relationships between data points rather than just the properties of the individual data points [45]. The same graph formalism can be used in various domains ranging from relations in social networks, drug and protein interactions, connections between terminals in a telecommunications network, pages on the World Wide Web, linked data on the Semantic Web, and many others [45].

1.1 What are Semantic Graphs and Why Graph Summarization?

We are agnostic to the specific representation of graphs. We use the umbrella term *semantic graph* for the kinds of graphs considered in this work. We assume that a graph is a collection of vertices connected by directed edges. The vertices and/or edges may be labeled and one may perform some semantic inference on them, e.g., generalizations and specializations [16].

▶ **Definition 1** (Semantic Graph). A semantic graph is a finite, directed, labeled graph $G = (V, E, \ell_V, \ell_E)$. Here $E \subseteq V \times V$ is the edge relation, and the functions $\ell_V \colon V \to \Sigma_V$ and $\ell_E \colon E \to \Sigma_E$ label the vertices and edges, respectively.

We also refer to the value $\ell_V(v)$ as the *type set* of a vertex, and we use the function ℓ_E to define the *property set* of a vertex, which is defined over its outgoing edges. Some variants of graph summarization also consider incoming edges (we discuss these in Section 3).

▶ **Definition 2** (Property Set). The property set of a vertex v in a semantic graph G is the set $\{\ell_E(v, w) \mid (v, w) \in E\}$ of labels on its outgoing edges. We abuse notation and write $\ell_E(v)$ for the property set of v.

A semantic graph may be represented as a Resource Description Framework (RDF) graph [24], labeled property graph (LPGs), or some other approach. RDF graphs do not directly support vertex labels, but rdf:type edges can be used to simulate these, so RDF graphs and LPGs can be transformed into one another [8]. A comprehensive overview of different kinds of semantic graphs is given by Hogan et al. [48] under the term "Knowledge Graphs". The term was coined in 2012 by Google as part of its knowledge representation and web search service extended by contextual knowledge such as mapping queries to persons, companies, etc.

Classical graph algorithms focus on finding structures such as shortest paths or minimum spanning trees, or invariants such as treewidth or chromatic number. In the context of graph-structured data, the focus of algorithms shifts from the graph *per se* to the data it represents. Typical tasks on such graphs include mere querying of the data, but also estimating cardinalities for queries in graph databases [77], subgraph-based indices for data search [59], data modeling recommendation [87], schema induction [99], data exploration [80], data visualization [41], and related entity retrieval [22].

The motivation for graph summaries lies in the growing size of semantic graphs. As the graphs can be extremely large, tasks become computationally expensive and might require a large amount of memory. Structural graph summaries have been developed as useful abstractions of large graphs to solve tasks more efficiently. A summary of a graph G is a smaller graph S, which retains the information from G that is required to perform the desired tasks, but which discards information that is not needed, and which may represent the retained information more compactly.

Example

Suppose we wish to count solutions to queries on a graph database G such as finding all students studying the same course, or all books in the same genre by the same set of authors. Many vertices in the graph G may be equivalent from the point of view of the queries. Hence, we can summarize the graph by merging each set X of equivalent vertices into a single vertex v_X that is labeled with the cardinality of X. In doing this, we lose the ability to answer certain queries that could be answered on the original database. We can no longer distinguish between vertices that have been merged, and we no longer know their identities. But we can still answer our class of counting queries, and we can do so by processing a much smaller graph – graph summaries are typically orders of magnitude smaller than the original graph, measured in numbers of edges [19].

Summary graphs can be constructed in several ways. The survey by Čebirić et al. [19] classifies existing techniques into structural, pattern-mining, statistical, and hybrid approaches. A broad overview of these summarization approaches can be found there. In this paper, we focus on structural approaches based on quotients, due to their versatility and popularity for summarizing semantic graphs. The idea here is to partition the vertices into equivalence classes, assigning each vertex to exactly one equivalence class. These equivalence classes are used as the vertices of the summary. A structural summary is again a graph that can be used to answer a given application task exactly [19], i. e., as if it were executed on the original graph. Structural graph summaries defined using quotients [20] are closely related to k-bisimulation [53,88]. Many summary approaches for semantic graphs are in fact stratified bisimulations.

Such structural summaries are "lossless" with respect to the features defined in a summary model: selected features of the original graph are accurately preserved to allow the task to be exactly computed on the summary. The features preserved by the graph summary are defined in the so-called *graph summary model*. Summaries can also be "lossy", only allowing the task to be approximated on the summary.

1.2 What is a Task, a Graph Summary, and a Graph Summary Model?

Since different works and communities deal with graph summarization from different perspectives, we first need to provide a high-level clarification of the basic concepts. This shall provide the reader with an intuition about the nature of graph summaries and how they are used. The basic definitions are of a *task* (in the context of a graph summary) as well as the *graph summary* and its *model*, based on Blume, Scherp, and Richerby [10].

▶ Definition 3 (Task). Given a graph G, a task T applied on G is a parameterized function $T_{\Psi}: G \to Y$ that maps the input graph G to a task-specific range of values Y.

The set of parameters Ψ and the output of the function T, (the range Y) are specific to a given application domain.

Example Task

For cardinality estimations of queries on the graph G the range Y is \mathbb{N}^+ and the parameter Ψ is the graph-based query Q for which the cardinality is to be estimated [77].

12:4 Structural Summarization of Semantic Graphs Using Quotients

▶ Definition 4 (Graph Summary Model). A graph summary model is a tuple $M = (EQR, \Phi, PAY)$, where EQR specifies an equivalence relation on graph vertices, Φ are model parameters, and PAY is a set of task-specific payload functions. They are computed during summarization and their output is included in the summary.

Structural approaches summarize a graph G w.r.t. an equivalence relation EQR $\subseteq V \times V$ defined on the vertices V of G [14, 19]. The vertices V(S) of the resulting summary graph S correspond to equivalence classes of the equivalence relation EQR, and to equivalence classes of subsidiary equivalence relations used in the definition of EQR – see Figure 1.

Example Summary Models

Example graph summary models are the attribute-based collection [18], where the EQR defines vertex equivalence based on having the same set of labels of the outgoing edges. In contrast, the class-based collection [18] is defined by an EQR that groups vertices sharing the same vertex labels. There are also summary models that require both attribute and class equivalence [22, 59].

The model parameters Φ are applied to control the output. Φ can, e.g., limit the maximum summary size (in terms of number of vertices or edges), the weights assigned to some graph elements (preference for certain edges or vertices), or the minimum support of subgraph structures summarized by S [19]. PAY describes what information must be stored about the summarized vertices to allow the task T to be answered using the summary.

Example Payload Functions

In a data search task, such as in the first example, the set PAY contains one payload function to compute cardinalities of the search results. Another commonly used payload function for data search is computing the set of data sources (URIs where the results can be found on the web) [43].

Different summary models serve different tasks. While a structural summary is a lossless representation of the input graph with respect to the features defined in a summary model, different summaries (with different payloads) will be needed for different tasks. Thus, the choice of features and the tasks have to be aligned. Finally, based on the graph summary model, we define the *Graph Summary*:

▶ Definition 5 (Graph Summarization). A graph summarization is a parameterized function $C_M: G \to S$ that computes a representation of the input graph based on a graph summary model $M = (EQR, \Phi, PAY)$. The summary S preserves (selected) features of G in accordance with M. Note that S also contains the output of the payload functions PAY.

For the full definition of a graph summary with respect to a graph summary model, see [10].

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

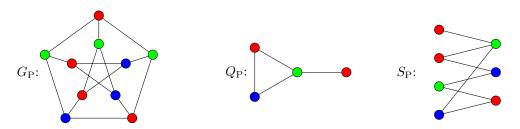


Figure 1 A colored Petersen graph G_P (left) quotiented (center, Q_P ; see Section 3) and summarized (right, S_P). The graph summary model defines two vertices of G_P to be equivalent if they have the same color and are adjacent to the same set of colors.

Example Summary of the Petersen Graph

In Figure 1, the semantic graph $G_{\rm P}$ to be summarized is a colored version of the Petersen graph. In this case, the graph is undirected and has vertex labels (the colors) but no edge labels. We use a graph summary model that defines two vertices of $G_{\rm P}$ to be equivalent if they have the same color and are adjacent to the same set of colors. There are four equivalence classes: the red vertex with only green neighbors (the top vertex), the three red vertices with green and blue neighbors, the three green vertices with red and blue neighbors, and the three blue vertices with red and green neighbors.

In the summary graph $S_{\rm P}$, these "primary" equivalence classes are on the left-hand side; the vertices on the right are "secondary" equivalence classes of the "same color" relation, which is used to define equivalence of the neighbors. In this example, there is no payload, but payload would be stored as labels of additional vertices connected to the primary equivalence classes.

1.3 What is *Not* a Graph Summary!

Graph summarization is distinct from other related concepts, some of which also use the term *summarization*, but to mean something different.

The goal of *compression* is to allow the original graph to be exactly recovered (in the case of lossless compression, for example, RDF HDT [38]), or approximately recovered (lossy compression). Thus, compression must retain (or approximately retain) all information in the graph. Another term in this context is "corrections" [58, 89]. Again, the goal is to (incrementally) compute a lossless compression of a graph by determining a so-called corrections set, i. e., a set of edges that must be added or removed to reconstruct the original graph [58, 89]. Another work on graph compression is by Hajiabadi et al. [44], who propose an approach to reconstruct a graph exactly (lossless) or with a small error (lossy). In contrast to compression and correction, summarization only retains the information that is needed for specific tasks.

An intermediate operation between summarization and compression is graph *contraction* [34], which uses the graph-theoretic concept of contraction minors [25]. Contraction makes graphs smaller by replacing regular structures such as cliques and paths with "supernodes". Each supernode is annotated with a payload and a mechanism for recovering the original graph. This can also be done hierarchically, to produce even smaller graphs [35]. Compared to summarization, this only operates on parts of the graph; on the other hand, the entire graph can be recovered if needed, so the contracted graph is not task-specific (albeit that tasks that can be performed using the payload of the supernodes, without expanding out the graph, will run faster).

12:6 Structural Summarization of Semantic Graphs Using Quotients

Further related research areas are graph transformation systems, graph rewriting, and modeldriven engineering [30]. These approaches have in common that they have rules and mappings to manipulate input graphs to a desired output graph, e.g., to merge adjacent vertices by preserving (hyper-)edge structures [57,81]. In general, the difference between structured graph summaries and the communities of graph transformation and graph rewriting is as follows: The goal of summarization is to provide concise representations of the input graph G while preserving specific, defined features (e.g., structural features like which edges are attached to a vertex). The focus of graph rewriting is to perform operations on the graph to transform G from one state to another via rewrite rules. Nonetheless, graph summarization is an example of graph rewriting in the most general sense.

1.4 Structure

The article is organized as follows: First, we consider applications of structural graph summaries in Section 2. In Section 3, we introduce features and models of structural graph summaries for static graphs that are based on quotients.

Bisimulation is a versatile and popular technique among structural graph summarization models, which we discuss in further detail in Section 4. Logics are a natural partner of structural summaries and this connection is discussed in Section 5. We consider graph summarization approaches for temporal graphs in Section 6.

Finally, in Section 7, we briefly discuss alternative approaches to graph summaries that are not based on graph quotients, such as pattern mining, and statistical approaches. We conclude this article with a brief reflection, as well as an outlook on future directions and open questions.

2 Applications of Structural Graph Summaries

Many different structural graph summaries have been developed to solve different tasks [6,18,22, 23,41,53,59,71,72,77,87,88,91,95]. In the following, we describe four common applications of structural (semantic) graph summaries in detail. These applications are typical applications for managed graph database systems but also for the decentralized Semantic Web [1].

These applications showcase the universal nature of structural summaries. We connect the applications with the basic notion of graph summarization by providing the specific definitions of the graph summary model M being used in each application. These are the equivalence relation EQR its parameters Ψ and equally important the set of payload functions PAY (see Definition 4).

2.1 Semantic Entity Retrieval

One application of structural graph summaries is to find semantically related entities in the Semantic Web [22]. Entities on the Semantic Web are represented using vertices, each identified with a unique IRI [29], and labeled edges indicate relations between them. Entities can have a set of types, each of which is indicated by an edge labeled rdf:type to a vertex representing that type. This graph data is typically stored as an RDF graph [1].

If two entities share the same set of RDF types and RDF properties (i.e., labels of outgoing edges other than rdf:type), we can say they are semantically related [22]. Hence, we can create a summary where the equivalence relation EQR represents that; i.e., it puts vertices in the same partition if their RDF types and set of RDF properties are the same.

This structural graph summary summarizes vertices (entities), based on such structural subgraph features. To find semantically related entities, we need to memorize the vertex identifiers of each summarized vertex in the computed structural graph summary, this index gets created by

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

a function from PAY and is part of the output of the graph summary. We can now use that index for immediate retrieval of related vertices since all semantically related vertices are summarized together.

A variation of this task would be when the graph is dynamic. In that case, there is an interesting trade-off between recomputing the index, which can be costly, versus updating it in place, which requires more index information to be kept and leads to a more complex implementation.

2.2 Cardinality Computation

Cardinality computation is often desired in databases [77, 84]. For graph databases, graph summaries can be used as an index to look up how many vertices will be returned by certain queries. If the graph summary is smaller than the original graph, as it usually is, the query runs faster on the graph summary than on the database. This task is related to the semantic entity retrieval task described above. However, for this purpose, we only need to memorize the number of summarized vertices rather than all vertex identifiers. Hence, EQR would be the same, but the function in PAY would only compute the counts, rather than keeping the full index.

A graph summary that memorizes the number of summarized vertices enables fast implementations of query size estimation [77]. Analogously to the semantically related entities task above, expensive re-computation of the graph summary from scratch when the database changes can yield an unwanted performance overhead, and hence a way to deal with online updates is necessary.

Knowledge bases often have a data schema or ontology that defines how entities should be modeled. Structural graph summaries can help determine how many entities strictly follow that schema, match the schema partially, or even contradict the schema. The stored numbers of summarized vertices can be used as an indicator of data completeness [84]. It is often desired to evaluate the evolution of data quality over time [84]; also in that case an incremental update mechanism is needed.

2.3 Data Source Search

As part of *data source search*, one needs to find (sub)graphs in the Semantic Web that match a given schema structure [43]. Structural graph summaries can be used as an index that memorizes the location of summarized vertices on the Web. This is illustrated in Figure 2.

The EQR of the graph summary model in this scenario is defined such that it summarizes vertices that have the same set of RDF types, and are connected by edges with the same labels to target vertices with the same set of types, known as SchemEX [59]. The set of payload functions PAY is in this case a cardinality count and memorizing the data source URIs.

With a graph query, the structural graph summary is queried to get the URLs of relevant data sources. Then, the data sources are accessed to download the graphs matching the query. Search systems like LODatio [43], LODeX [6], Loupe [71], and LODatlas [80] rely on structural graph summaries to offer a search for relevant data sources or exploration of data sources.

To implement this task, we need to memorize the locations where each summarized vertex appears, which is computed by a function in PAY. As the data on the Web changes [51], the summaries need to be updated as well. In contrast to the previous two tasks, for data source search we neither memorize vertex identifiers nor the number of summarized vertices but only their location.

12:8 Structural Summarization of Semantic Graphs Using Quotients

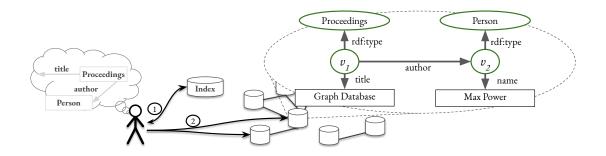


Figure 2 Finding data sources on the Web using an index based on graph summaries (from [8]). A structural query is executed over an index to identify relevant data sources (1). Subsequently, the data sources are accessed to retrieve actual vertices (2).

2.4 Training Graph Neural Networks

Another application of graph summaries can be found in recent work by Generale, Blume, and Cochez [39], who address scalability issues when training graph neural networks (GNNs) on larger graphs. They suggest training the GNN on a summary, rather than on the original graph. The summaries used in this case are actually quotient graphs (see Section 3), using either k-bisimulation (see Section 4) or direct vertex attributes for EQR.

The machine learning task solved with this GNN is the prediction of vertex types. In the normal training setting, the training set is a subset of the vertices of known type. The model parameters are then optimized such that the outputs of the GNN, when passed through a classifier, predict the correct vertex type for the test vertices. To evaluate whether the model works, one uses it to predict the types of vertices outside the training set and compares these predictions to the correct answer. An issue while training on the summary is that multiple vertices from the original graph are in the same equivalence class and hence mapped to the same vertex in the summary. Now, the difficulty is that we do not have a clear label for this vertex; it is not the case that all nodes in the equivalence class have the same label, since this is not taken into account when summarizing. The authors [39] suggest using a weighted multi-label classification task during training where the labels are weighted by their frequency. To compute this label, we need to collect the frequency of the labels of the vertices of each equivalence class, which is one of the functions in PAY.

After training on the summary graph, the weights of the model are transferred back to the original graph, where inference is performed to predict the types of vertices from the test set. Inference can be done on the full graph because it is much faster than training and requires much less memory. In some cases, this method of training can not only provide reasonable results, but also the results can be improved if the model is further trained on the original graph. This two-stage process can give better results than training only on the original graph, without summarizing. To transfer the model weights, we include the index with the equivalence classes as part of PAY.

A follow-up work by Bollen et al. [13] provides a theoretical foundation for the earlier work. They prove that it is possible to create a specific graph summary that for a specific class of GNN performs the same message-passing steps as on the original graph. In effect, this means that the same outcome is obtained at inference time. The prior work is not covered by this proof because the summary did not retain information about the cardinality of edges in the quotient graph, and therefore the conditions for equivalence are not met. Error bounds for the approximation remain open.

3 Structural Summarization of Static Graphs based on Quotients

In Section 1, we introduced structural graph summaries as a condensed representation of a graph based on some summary model. Our focus is on quotient-based summaries. For context, we first define graph quotients.

▶ **Definition 6** (Graph Quotient). Consider a labeled graph G, and an equivalence relation \equiv on G's vertices. The quotient of G w.r.t. \equiv is the graph Q defined as follows. The vertices are the equivalence classes of \equiv . Q has an edge labeled p from class c to class c' iff G contains an edge labeled p from some vertex in c to some vertex in c'.

Example

Recall Figure 1. There, vertices of the colored Petersen graph $G_{\rm P}$ are defined to be equivalent if they have the same color and their neighbors have the same set of colors. This leads to four equivalence classes among $G_{\rm P}$'s vertices: the blue vertices (which are all adjacent to both red and green), the green vertices (which are all adjacent to red and blue), and what we will call type-1 (adjacent to green and blue) and type-2 red vertices (adjacent only to green). These are the four vertices of the quotient $Q_{\rm P}$, with the type-1 red vertex on the left and the type-2 on the right. The edges can be read off from the descriptions of the equivalence classes, noting that each green vertex is adjacent to a blue vertex, a type-1 red and the type-2 red.

Quotient-based summaries are constructed along similar lines, but the summary stores relationships between multiple equivalence relations, instead of just between the classes of a single one. For example, the equivalence relation described above for the colored Petersen graph $G_{\rm P}$ defines the overall equivalence on vertices with reference to a second equivalence relation: that which only considers the vertex's color and not its neighbors. Blume, Richerby, and Scherp [10] define a language, FLUID, for specifying such combinations of equivalence relations. We omit the details here, as the following intuition suffices.

The input semantic graph is summarized using a *primary equivalence relation*, which is defined in FLUID by a logic-like expression that combines *secondary equivalence relations*. The secondary equivalence relations may themselves be defined by combining other secondary equivalence relations.

▶ Definition 7 (Summary Graph). The summary S of a graph G w.r.t. a summary model M has a vertex for each equivalence class of each equivalence relation used in the definition of EQR in M. Primary and secondary vertices correspond to primary and secondary equivalence relations, respectively. Between these vertices, edges are added similarly to quotient graphs. Suppose equivalence relation R is defined by combining equivalence relations, one of which is R'. Then there is an edge labeled p from class c of R to class c' of R' in the summary graph if and only if there is an edge labeled p in the input graph from some vertex in c to some vertex in c'.

This can be seen in Figure 1. The primary vertices are on the left of $S_{\rm P}$ (equivalence classes of "same color and same colored neighbors") and the secondary vertices ("same color") are on the right. All edges are between a primary vertex and a secondary vertex.

In this approach to structural graph summarization, each vertex v of the input graph G is represented by exactly one primary vertex p_v in the summary graph S. The structure around vin G that is used to determine v's equivalence class is represented by the edges between p_v and secondary vertices, and edges between the secondary vertices. Structure around v that is not used to determine v's equivalence class is not represented in the summary.

12:10 Structural Summarization of Semantic Graphs Using Quotients

Bisimulation is a very common feature among structural graph summarization approaches, but often the works on structural summarization using quotients and bisimulation do not explicitly refer to each other. We discuss bisimulation in detail in Section 4, where we explain its relation to summary models introduced in this section.

There are many structural graph summary models based on quotients, defining the features that shall be captured by the summary, often targeted for solving one specific task [6, 22, 43, 71, 77, 80, 87,91,92]. Based on Blume et al. [10], we analyze existing structural graph summaries with respect to the captured schema structure, i. e., what features of the input graph are used to summarize vertices. This analysis complements existing surveys covering graph summaries [14, 19, 55, 66] and the taxonomy is summarized in Figure 3. We distinguish features that only use triple information (triple features), features that define how features of multiple vertices are combined (subgraph features), and features that define explicit semantic rules such as joining and inference (semantic rule features). Each group of features adds another level of complexity, i. e., intuitively, the computational complexity of computing summaries grows when features of different groups are used. There is no single graph summary model that supports all features. However, we see common combinations of features. In the following, we summarize the graph summary models along with the identified features.

Other features in quotient-based summaries include dependent compression [50], which summarizes vertices v_1 and v_2 if and only if v_1 is adjacent only to v_2 , or vice versa [50]. In another variant of dependent compression, a set of vertices is grouped if they are connected to the same set of other vertices in G [65].

3.1 Triple Features

Triple features are solely based on outgoing triples of vertices. A triple corresponds to a directed edge between two vertices, namely the subject connecting to the object, which is labeled with the predicate.

Edge Labels

To compute the equivalence of two vertices s and s' of G, we compare the triples where the subject is s with those where the subject is s'. The most commonly used feature in structural graph summaries is using properties to compute the schema of vertices. More specifically, for each vertex s in the data graph the set of edge labels $\ell_E(s)$ is compared. For example, the graph summary model attribute-based collections, due to Campinas et al. [18], solely relies on the sets of edge labels to compute the graph summary. If vertices s and s' share the same property set, i.e., $\ell_E(s) = \ell_E(s')$, they are considered equivalent, so are summarized together.

Vertex Labels

Another common feature is using the vertices' labels to compute the summary. Here, for each vertex s in the data graph, the so-called type set $\ell_V(s)$ is compared. If vertices s and s' share the same type set, i. e., $\ell_V(s) = \ell_V(s')$, they are considered equivalent. For example, class-based collections (again due to Campinas et al. [18]) is a summary model, which uses only the vertices' label sets to compute the schema. These are used, along with the attribute-based collections described above, to implement a query recommendation system for SPARQL,¹ that facilitates working with heterogeneous datasets, especially when the schema structure is unknown.

¹ https://www.w3.org/TR/sparql11-overview/

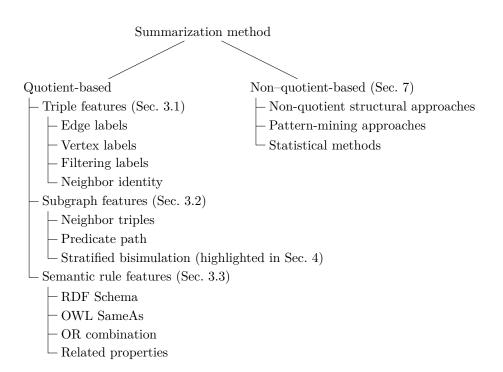


Figure 3 The taxonomy of features that are used in structural graph summarization, based on Čebirić et al. [19]. These features are described in detail in the sections indicated, and can be summarized in both static and temporal graphs. Section 3 introduces the different features of quotient-based methods. Section 4 discusses stratified bisimulation methods in detail, as they are most prominent among quotient-based methods. Section 5 further reflects on quotient-based summaries and their relation to logics, which are by their nature orthogonal to the taxonomy. Quotient-based summaries for temporal graphs are discussed in Section 6. The overview of summarization methods is complemented with a brief discussion of non–quotient-based summarization methods in Section 7.

Filtering Labels

Tran et al. [95] proposed the feature of label parameterization for graph summaries. With the label parameterization, only a subset of all edge labels is used to compute the schema. More precisely, one defines a set of predicates P_l , the so-called label set, which are ignored when determining the equivalence of vertices. Tran et al.'s graph summary combines property sets $\ell_E(s)$ with label sets. Furthermore, they combine this with k-bisimulation (see Section 4 on k-bisimulation).

Neighbor Identity

The final triple feature uses the identity of outgoing neighbors $\Gamma^+(s) = \{v \mid (s,v) \in E\}$ to determine the equivalence of vertices. It appears that no existing graph summary summarizes vertices solely by comparing the neighbor identities. However, SemSets [22] summarize vertices that share the same outgoing predicates, which are linked to the same vertices. To check if vertices s and s' are equivalent under SemSets, all triples where s or s' are the subject vertices are compared. For each triple $(s, p, o) \in G$ there has to be a triple $(s', p, o) \in G$, and vice versa. Thus, they combine neighbor vertex identifiers $\Gamma^+(s)$ with predicate paths (see below).

12:12 Structural Summarization of Semantic Graphs Using Quotients

3.2 Subgraph Features

The neighbor vertex identifier is the most direct approach to incorporate neighbor information and leads to a wider range of summary models that consider neighbor information, e.g., vertices in $\Gamma^+(s)$. We classify features as subgraph features when they combine triple features of multiple vertices.

Neighbor Triples

SchemEX [59], SchemEX+U+I [11], ABSTAT [91], LODeX [6], and Loupe [71] summarize vertices s and s' based on having a common type set and common edge labels linking to vertices with the same type sets. This means that, to compute the schema of one vertex s, also the type sets of outgoing neighbors $\Gamma^+(s)$ are required to be equivalent, i. e., we compare neighbor triples. In contrast to SemSets [22], these approaches use not the neighbor vertex identifiers $\Gamma^+(s)$ but the type set $\ell_V(o)$ for each $o \in \Gamma^+(s)$. SchemEX [59], SchemEX+U+I [11], ABSTAT [91], LODeX [6], and Loupe [71] combine type sets $\ell_V(s)$, property sets $\ell_E(s)$, and neighbor type sets $\ell_V(o)$ for $o \in \Gamma^+(s)$ using predicate paths, introduced next. This mapping can aid in recommending related queries and generally for finding relevant data sources [43].

Predicate Path

Almost all analyzed graph summaries that use neighbor information combine the schema structures using predicate paths, i.e., they compare which predicates link to which neighbors. Predicate paths are compared based on the edge labels and type sets that appear along paths. For example, SchemEX [59], SchemEX+U+I [11], ABSTAT [91], LODeX [6], and Loupe [71] consider which property links to which type set. TermPicker [87] follows a different strategy to integrate the schema of neighboring vertices. TermPicker summarizes vertices s based on having the same type set $\ell_V(s)$, the same property set $\ell_E(s)$, and the same set of types among the neighbors, $\{\ell_V(o) \mid o \in \Gamma^+(s)\}$. Consequently, TermPicker's graph summaries compress all type sets of all neighbors into a single type set. Thus, TermPicker's graph summaries do not contain information about which *specific* property linked to which neighbor.

Stratified Bisimulation

Many graph summaries compute the schema of vertices by taking into account the schema of neighbors over multiple hops [53, 59, 83, 95]. This is commonly defined as a bisimulation. Bisimulation operates on state transition systems and defines an equivalence relation over states [86]. Two states are equivalent (or bisimilar) if they change into equivalent states with the same type of transition. Inductively, this means that applying an arbitrary sequence of transitions to two bisimilar states will result in bisimilar states. Interpreting a labeled graph as a representation of a state transition system allows us to apply bisimulation on graph data to discover structurally equivalent parts.

In practice, many graph summary models define a **stratified** k-**bisimulation**, e.g., [53, 59, 83, 95]. When states are k-bisimilar, applying any sequence of k transitions to them will result in equivalent states, but applying more than k transitions may lead to inequivalent states. Thus, stratified bisimulation only considers paths of lengths up to k when determining equivalence. This increases the chance that two vertices are considered equivalent.

Some graph summaries combine the feature of using only incoming or only outgoing properties with the k-bisimulation feature [23, 72, 88]. This is referred to as **backward** k-bisimulation and forward k-bisimulation, respectively [41]. The T-index of Milo and Suciu [72] supports path

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

queries in semi-structured databases. This summarizes vertices s based on having the same set of incoming property-paths, i. e., they use k-bisimulation only on incoming property sets $\ell_E^-(s)$. Consens et al. [23] propose a structural graph summary model to support navigational SPARQL queries, so-called Extended Property Paths (EPPs). They summarize vertices s based on having the same set of outgoing property-paths, i. e., they use k-bisimulation only on outgoing property sets $\ell_E(s)$. In addition, for each hop, the type sets $\ell_V(s)$ have to be equivalent.

Stratified k-bisimulation is very popular in the summary models found in the literature. Often, however, it is not referred to as such and/or the summary model is considering only the case of the 1-bisimulation. Due to its importance for structural graph summarization, we discuss and reflect on the k-bisimulation feature in detail in Section 4.

3.3 Semantic Rule Features in Graph Summaries

The last group of features for structural graph summaries defines explicit semantic rules. It deals with RDF Schema (RDFS) reasoning, OWL's owl:sameAs reasoning, as well as inference on property sets in an OR-like fashion and via the inclusion of related properties.

RDF Schema

Several semantic structural graph summaries use RDF Schema inference to enhance their summaries. ABSTAT [91] exploits RDF Schema type hierarchies to compute so-called minimal patterns. They select the minimal number of types, i.e., they only keep the most specific types from the RDF Schema type hierarchy. Goasdoué et al. [41] exploit RDF Schema type hierarchies, property hierarchies, and RDF Schema domain and RDF Schema range. With domain and range, types for the subject vertex and the object vertex can be inferred.

OWL's SameAs

SchemEX+U+I [11] also uses the full RDF Schema inference but also exploits the semantics of the owl:sameAs property. This property is part of W3C's OWL [70], which is heavily used in the context of RDF graphs. The owl:sameAs property defines an equivalence relation [70, Section 4.2], intended to identify vertices that represent the same real-world entity. To compute the schema structure of one vertex v, the schema structures of all vertices v' in the weakly connected components in an owl:sameAs-labeled subgraph of G are merged (see Ding et al. [26] for details on owl:sameAs networks).

OR Combination

Goasdoué et al. [41] define the Weak Summary using an OR-like combination. In the Weak Summary, two vertices s and s' are equivalent if they have the same outgoing property set and/or the same incoming property set. This is not necessarily transitive, so the transitive closure is taken as the equivalence relation. The authors also define a Typed Weak Summary, which combines the Weak Summary based on properties with vertex types. There is also a variant of a Strong Summary, which does not consider the OR-like combination. Details can be found in Goasdoué et al. [41], see also the discussion in Blume et al. [10].

Related Properties

Goasdoué et al. [41] also propose to include property relations. Two properties p and p' are sourcerelated if they co-occur in any property set $\ell_E(s)$ of any vertex s and they are target-related if they co-occur in any incoming property set $\ell_E^-(s)$ of any vertex s (i.e., the set $\{\ell_E(v,s) \mid (v,s) \in E\}$).

4 Structural Graph Summarization by k-Bisimulation

We introduced bisimulation in Section 3.2 as a means of defining equivalence of vertices in a graph. Bisimulations are specific kinds of equivalence relations that classify vertices v and w as equivalent if, for each edge (v, v') with label p, there is an edge (w, w'), also with label p, where w' is, recursively, equivalent to v'. Complete bisimulation extends this recursively to all distances from v and w, whereas k-bisimulation only requires equivalence out to distance k. Forward bisimulation considers outgoing edges to determine equivalence, as described above; backward bisimulation is analogous but uses incoming edges; backward-forward bisimulation uses both. In addition to edge labels, bisimulation may also be based on vertex labels or both, but this makes no principal difference.

In complete bisimulation, for the vertices v and w to be bisimilar, their in-/out-neighbors must be bisimilar as well. This recursive definition is, essentially, an equivalence relation defined in terms of itself. Since there is only one equivalence relation, this naturally lends itself to a quotient representation, as exemplified in Section 2.4. However, stratified bisimulations are more commonly used for graph summarization, and lead naturally to quotient-based graph summaries. as we discuss below.

4.1 Stratified Bisimulation to Paths of Length k

As real-world graphs are quite heterogeneous, there may be only a few bisimilar vertices [19] if we consider full bisimulation. Thus, k-stratified bisimulation is often used, restricting the paths to length k. This increases the possibility that two vertices are bisimilar and, overall, reduces the size of the summary.

A k-bisimulation on a graph G considers features a distance at most k from a vertex that is to be decided equivalent to another vertex [85]. Formally, this can be defined for the forward bisimulation on the outgoing edges as follows (based on [85]). Backward bisimulation is defined similarly.

▶ **Definition 8** (Stratified Forward Bisimulation based on Edge Labels). *The* forward *k*-bisimulation $\approx_{\text{fw}}^k \subseteq V \times V$ with $k \in \mathbb{N}$ is defined as follows:

- $\begin{array}{l} & u \approx_{\mathrm{fw}}^{0} v \ for \ all \ u, v \in V, \\ & u \approx_{\mathrm{fw}}^{k+1} v \ iff \ u \approx_{\mathrm{fw}}^{k} v \ and, \ for \ every \ edge \ (u, u'), \ there \ exists \ an \ edge \ (v, v') \ with \ the \ same \ label \ u \in V, \end{array}$ such that $u' \approx_{\text{fw}}^{k} v'$, and vice-versa.

For bisimulation with vertex labels, we modify the base case of the definition so that $u \approx_{\text{fw}}^{0} v$ iff u and v have the same labels. Note that stratified bisimulation defines a hierarchy of equivalence relations $\approx_{fw}^0, \approx_{fw}^1, \ldots \approx_{fw}^k$, in contrast to complete bisimulation, which defines a single equivalence relation, recursively in terms of itself. Thus, stratified bisimulation is best suited to summarization, rather than quotienting. When producing a summary for k-bisimulation, the relation \approx_{fw}^{k} will be the primary equivalence relation, and $\approx_{fw}^{0}, \ldots, \approx_{fw}^{k-1}$ are secondary.

Bisimulation stratified to paths of length k is a popular technique to compute structural graph summaries, though often k = 1 is used. We give examples in Table 1 and discuss them in detail in this section. Note that TermPicker [87] is a relaxed version of bisimulation. Conventional bisimulation requires the same edge label to the same type of neighbor, whereas TermPicker just requires the same edge labels and the same neighbor types, without the correlation.

Examples of Stratified Bisimulation for Graph Summarization 4.2

Efficient algorithms for bisimulation have been developed by Paige and Tarjan [78], Kaushik et al. [53], Dovier et al. [27], Schätzle et al. [88], and others.

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

Summary model	Depth	Labels used	Direction	Application
Class collection [18]	0	vertex	forward	query recommendation
Attribute collection [18]	1	edge	forward	query recommendation
LODex [6], Loupe [71]	1	vertex and edge	forward	data exploration
SchemEX [43, 59]	1	vertex and edge	forward	data search
TermPicker [87]	1	vertex and $edge^*$	forward	modeling recommendation
Tran et al. [95]	k	edge	backward-forward	entity search

backward

backward

forward

general purpose

general-purpose

data indexing

Table 1 Example graph summary models based on bisimulation.

(* relaxed version of bisimulation; see text)

k

k

k

vertex

vertex

edge

A(k)-index [53]

Schätzle et al. [88]

T-index [72]

A notion of k-bisimulation w.r.t. graph indices is introduced by seminal works such as the k-RO index [76] and the T-index summaries [72]. Milo and Suciu's T-index [72], the A(k)-Index by Kaushik et al. [53], and others summarize graphs using backward k-bisimulation. Qun et al. [83] extend the A(k)-Index to a D(k)-Index, which is also based on bisimulation but focuses on query optimization. To this end, the D(k)-Index dynamically adapts its structure according to the current query load. Another work using stratified bisimulation is by Fan et al. [33], for reachability and graph pattern matching on large graphs.

Conversely, the k-RO index, the Extended Property Paths of Consens et al. [23], the SemSets model of Ciglan et al. [22], Buneman et al.'s RDF graph alignment [17], and the work of Schätzle et al. [88] are based on forward k-bisimulation. Buneman et al. use forward k-bisimulation to summarize the union of two consecutive versions $G_{\text{union}} = G_1 \cup G_2$ of an RDF graph with respect to k-bisimulation, which puts vertices to be aligned in the same partition. As well as to k-bisimulation, they use a similarity measure to further refine the initial k-bisimulation partition. as it does not capture all vertices to be aligned. The focus of their work is the optimization of the alignment process so that every node pair (v_1, v_2) , with $v_1 \in G_1$ and $v_2 \in G_2$, which have to be aligned is identified and not the construction of a k-bisimulation-based partition of G. Schätzle et al. compute a forward k-bisimulation on RDF graphs in sequential and distributed settings [88]. For a small synthetic dataset ($\sim 1M$ RDF-triples) the sequential algorithm slightly outperforms the distributed one; for larger datasets, the distributed algorithm clearly outperforms the sequential one.

Tran et al. compute a structural index for graphs based on backward-forward k-bisimulation [95]. Moreover, they parameterize their notion of bisimulation to a forward-set L_1 and a backward-set L_2 , so that only labels $l \in L_1$ are considered for forward bisimulation and labels $l \in L_2$ for backward bisimulation. However, similar to Buneman et al., the particular focus of their work is not the construction of the bisimulation partition. Rather, they evaluate how one can efficiently optimize query processing on semi-structured data using an index graph based on bisimulation.

There are also structural summarization approaches that determine vertex equivalence only based on local information ($k \leq 1$). As shown in Table 1, many of the summary models introduced in Section 3.2 are actually very shallow bisimulations.

4.3 **Distributed and Parallel Bisimulation**

Luo et al. [67, 68] examine structural graph summarization by forward k-bisimulation in a distributed, external-memory model. They empirically observe that, for values of k > 5, the summary graph's partition blocks change little or not at all. Therefore they state that, for summarizing a graph with respect to k-bisimulation, it is sufficient to summarize up to a value of k = 5.

12:16 Structural Summarization of Semantic Graphs Using Quotients

Martens et al. [69] introduce a parallel bisimulation algorithm for massively parallel devices such as GPU clusters. Their approach is tested on a single GPU with 24 GB RAM, which limits its use on large datasets. Nonetheless, their proposed blocking mechanism could be combined with our vertex-centric approach to further improve performance.

5 Structural Graph Summarization and Logics

Logics are a natural framework for defining properties of graphs, queries on graphs, and transformations between graphs. More specifically, logics can be used to define equivalence relations on the set of vertices V of a graph, partitioning V into disjoint sets [10]. Thus, logics provide a natural framework for defining structural graph summaries. The partitions and the relationships between them are interpreted as summaries of the vertices, either in the sense of a quotient or a structural graph summary. The question is what logics should be used to define graph summaries.

Since Fagin's discovery [31] that existential second-order logic defines exactly the properties of graphs in the complexity class **NP**, the field of descriptive complexity [49] has sought to understand the relationship between the features of a logical language (e.g., the logical operators and quantifiers it contains) and the computational complexity of the graph properties it can define. A paradigmatic tension in descriptive complexity theory is the trade-off between a logic's expressive power and the computational cost of evaluating formulas [49,64]. Thus, we seek logics that are expressive enough to define interesting summaries but not so expressive that summaries cannot be computed in a reasonable time. To allow summarization of large graphs, it is essential that formulas can be evaluated in polynomial time. For web-scale graphs, we need even more efficient evaluation, which can be obtained by syntactically restricting the formulas that can be written, e.g., by restricting the number of variables in formulas or using guarded fragments of the logics. For example, evaluating a formula like $\exists y P(y)$ requires searching the whole graph for a vertex satisfying the predicate P, whereas, for a given vertex x, the guarded formula $\exists y (E(x, y) \land P(y))$ only requires us to search among x's neighbors. We note that existing graph summaries are typically expressible using guarded formulas, such as the neighborhood feature in FLUID [10] or constraints on requiring certain vertex labels to appear with specific edge labels [99].

First-order logic (FO) is powerful but can also be evaluated efficiently. However, from analyzing the existing graph summary approaches, we see that extensions to FO are needed to express more complex summaries. For example, extensions with counting quantifiers [49,64] can count vertices and edges, and define vertex equivalences based on the number of neighbors of particular kinds, rather than just the existence of such neighbors. This is required for pattern mining methods with a min-supp threshold or summarizing vertices based on the number of edges having the same label. Counting quantifiers and their expressive power have been extensively studied in descriptive complexity [49,64] and similar ideas have also been introduced in graph pattern matching [36]. One can also express iterative constructs such as loops – as required for computing bisimulations – by extending FO with fixed-point operators [49] or other recursive mechanisms such as those in Datalog [21].

6 Summarization of Temporal Graphs

Existing structural graph summarization algorithms are often designed and/or evaluated using static graphs only [6,22,66,77,87,91]. Few quotient-based structural graph summaries are designed for evolving graphs [40,59].

There are two ways in which one could consider summarizing temporal graphs. First, we may have a summary model that is not aware of time, and desire an algorithm that can update the summary as the data graph changes, considering the changes as a sequence of versions of the

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

The literature predominantly uses the first approach. Thus, we discuss algorithms that efficiently recompute quotient-based summaries as a graph changes over time. Subsequently, we present related approaches to incremental graph indexing and schema discovery. Indices are often based directly on quotients, often as bisimulations. Schema discovery classifies vertices according to their properties, naturally defining equivalence classes that can be quotiented.

6.1 Incremental Graph Summarization

Konrath et al. [59] compute their graph summary over a stream of vertex-edge-vertex triples. They can deal with the addition of new vertices and edges to the graph but not the deletion of vertices or edges, or modification of their labels. Similarly, Goasdoué et al. [40] present the summarization tool RDFQuotient, which only supports iterative additions of vertices and edges to their structural graph summaries, and does not handle deletions. Thus, these approaches are not suited to updating structural summaries of evolving graphs. Goasdoué et al. also do not support payload information, which is needed for tasks such as cardinality estimations and data search. The purpose of their summaries is to visualize them to a human viewer. Finally, Blume, Richerby, and Scherp [9] propose an incremental algorithm to update graph summaries that also takes deletions and payload into account. Experiments on benchmark datasets show that using the incremental algorithm is beneficial even if up to half of the graph has changed from one version to the next.

6.2 Incremental Subgraph Indices

Often, graph databases use path indices, tree indices, and subgraph indices [46]. A seminal approach to computing subgraph indices is DataGuide [42], implemented in the Lore database management system (DBMS) [102]. A DataGuide is a graph index built incrementally while executing queries on an XML database. It indicates to the query engine if and how a specific path defined in the query can be reached. To this end, a DataGuide represents all possible paths between two vertices in an XML file. While DataGuides operate on semi-structured data in terms of XML trees, a guide is essentially quotienting the graph.

Tran et al. [95,96] took up this idea and applied it to quotienting RDF graphs. Representative Objects (ROs) by Nestorov et al. [76] take up the ideas of DataGuides and are also implemented in the Lore DBMS with focus on path queries, query optimization, and schema discovery. While the Full ROs capture a description of the global structure of the graph, the authors also introduce a notion of a k-RO. which only considers paths of length up to k. These are examples of bisimulation in graph summarization.

We now consider incremental subgraph indices based on frequent pattern mining. These techniques group graph patterns, similarly to structural summarization. For example, Yuan et al. [104] (see also extensions in [52, 105]) propose an index based on mining frequent and discriminative features in subgraphs. Their algorithm minimizes index lookups for a given query and regroups subgraphs based on newly added features.

A work directly based on quotients is Qiao et al. [82] who compute an index of isomorphic subgraphs in an unlabeled, undirected graph G. The goal is to find the set of subgraphs in G that are isomorphic to a given query pattern. The result is a compression of the original graph that can be used to answer, e.g., cardinality queries regarding subgraphs. This is for static graphs,

12:18 Structural Summarization of Semantic Graphs Using Quotients

but the algorithm of Fan et al. [32] can deal with graph changes for the subgraph isomorphism problem. Their incremental computation of an index for isomorphic subgraphs is closely related to structural graph summarization but, unlike in summarization, the graph pattern p is an input to the algorithm, not the output.

Min et al. [73] propose an algorithm for continuous subgraph matching using a summary-like data structure that stores the intermediate results between a query graph and a dynamic data graph. They consider undirected graphs where only vertices are labeled. Dynamic graphs are updated through a sequence of edge insertions and edge deletions. The TipTap [74] algorithm computes approximations of the frequent subgraphs on up to k vertices w.r.t. a given threshold. This is similar to a quotient, but vertices may appear in multiple subgraphs. It does this to count occurrences of subgraphs in large, evolving graphs, modeled as a stream of updates on an existing graph. Tesseract [7] is a distributed framework for executing general graph mining algorithms on evolving graphs. It uses a vertex-centric approach to distribute updates to different workers. It assumes that most changes affect only local graphs, so few duplicate updates need to be detected. Tesseract supports the quotient-like ideas of k-clique enumeration, graph keyword search, motif counting, and frequent subgraph mining.

Another area of incremental subgraph indices considers an evolving set of queries over a static data graph. Duong et al. [28] propose a streaming algorithm using approximate pattern matching to determine subgraph isomorphisms. They use k-bisimulation to determine equivalent subgraphs and store them in an index. However, this index is computed offline for a static graph only and their algorithm considers a stream of graph queries as input.

6.3 Incremental Schema Discovery

Another area related to quotient-based summarization is incremental schema discovery. Vertices with the same schema are naturally equivalence classes that can be quotiented. Völker and Niepert mine logical patterns in the Web Ontology Language from static RDF graphs [99]. Wang et al. [100] incrementally discover attribute-based schemata from JSON documents. The schema is computed incrementally as more documents are processed. Baazizi et al. [4] also compute schemata from JSON objects, focusing on optional and mandatory attributes.

In addition to document-oriented formats like JSON, schema discovery is also used for graph data. For example, XStruct [47] follows a heuristic approach to incrementally extract the XML schema of XML documents. However, such schema discovery approaches cannot deal with modifications or deletions of nodes in the XML tree. Other schema discovery approaches focus on generating (probabilistic) dataset descriptions. Kellou-Menouer and Kedad [54] apply density-based hierarchical clustering on vertex and edge labels in a graph database. This computes profiles that can be used to visualize the schema of the graph. Recently, Bouhamoum et al. [15] used density-based clustering to extract schema information from an RDF graph and incrementally update the schema when new RDF instances arrive. While the work can deal with additions, the deletion of edges and vertices is not considered.

7 Non-Quotient Graph Summaries

The methods for structural graph summarization discussed so far focus on analyzing the graph based on pre-defined structural features such as paths and subgraphs encountered in the graph [10, 19]. These structural graph summaries are based on quotient graphs [19] (see Section 3).

There are also structural summary models that are not formed from quotient graphs, which we discuss below. Subsequently, we consider approaches for summarization based on pattern mining and statistical approaches. The organization of the methods in these categories is taken from Čebirić et al. [20].

7.1 Non-Quotient Structural Graph Summaries

Non-quotient summaries do not use equivalence relations to summarize a graph. Rather, the summary graph is composed of vertex summaries vs, which group together vertices v of the original graph G according to certain criteria [19].

The main difference from quotient summaries discussed above is that, in the non-quotient summaries, a vertex v can belong to zero, one, or multiple vertex summaries vs. In contrast, in quotient summaries every vertex v has exactly one corresponding vertex summary vs, which is the equivalence class of v under ~ [19].

Early work on non-quotient summarization includes that of Goldman and Widom [42], who created a vertex summary vs for every labeled path in the original graph G. A vertex v of G is associated with a vertex summary vs if it is reachable by the corresponding label path. The summary graph is used as a path index, as well as a tool for understanding the schema structure in semi-structured databases, and hence finds application in query formulation and query optimization. Revisiting the summarization tool SchemEX [59], its first layer – the RDF class layer – consists of vertex summaries vs_{c_i} representing all the classes c_j present in the input RDF graph G. A vertex v of G is associated with a vertex summary $v_{s_{c_i}}$, iff v is of the corresponding type c_j . Since a vertex v can have multiple types c_{j_1}, c_{j_2}, \ldots , it is possible that v is associated with several vertex summaries $vs_{c_{j_1}}, vs_{c_{j_2}}, \ldots$ and therefore the index's RDF class layer is considered a non-quotient summary. Kellou-Menouer and Kedad [54] perform schema extraction by using density-based clustering to establish a partition of the vertices based on type profiles. For each type T_j , a type profile $TP_j = \{(label_1, \alpha_1), (label_2, \alpha_2), \ldots\}$ is constructed, consisting of tuples of edge labels for outgoing edges (v, w) and incoming edges (w, v), with $v \in T_i$. The associated probabilities α_i denote how likely it is that a vertex $v \in T_i$ has an edge with the respective label_i. If a type profile TP_j contains all entries $(label_i, \alpha_i)$ of another type profile TP_k and every α_i is greater than a certain threshold θ (e.g., $\theta = 0.6$), then the vertices in T_k are added to T_i to create overlapping classes. Clustering can be found in more structural non-quotient approaches [56, 63, 75, 98].

Other structural graph summarization methods that do not use quotients are based on structural measures such as centrality. They identify the most important vertices, cliques, and others, and connect them in the summary [19]. The difference from quotient-based summaries is that some graph vertices may not be represented in the summary, i.e., the summaries are approximate. Examples of summary methods using structural features are [79,97,107]. These guide the summaries using vertex centrality measures [12] such as vertex (in/out) degree, betweenness (how often a vertex lies on the shortest path between two other vertices), (k, h)-cores, and the well-known information retrieval measures PageRank and HITS (based on eigenvalue analysis over vertices), as well as further measures of vertex centrality such as those applied by Pappas et al. [79].

7.2 Pattern-Mining Approaches for Graph Summarization

Pattern-mining approaches identify frequent patterns in the input graph G, which are then used to construct the summary graph SG [19]. Song et al. [90] construct *d*-summaries to summarize a knowledge graph G. A summary P, which is a graph pattern found in G, is considered a *d*-summary, iff all the summary vertices $u \in P$ are *d*-similar (R_d) to all their respective original vertices $v \in V$. Informally, uR_dv iff (1) u and v share the same label and (2) for every neighbor $u' \in P$ of uconnected over an edge with label p there exists a respective neighbor $v' \in V$ connected via the same edge label and $u'R_{d-1}v'$. Their definition of *d*-similarity is very similar to *k*-bisimulation (Section 4) and mainly differs in the domain on which it is defined, namely summary vertices and original vertices.

12:20 Structural Summarization of Semantic Graphs Using Quotients

Pattern mining methods for graph summarization discover frequently occurring patterns in the data [19]. Various algorithms exist for graph pattern mining, based either on the well-known Apriori principle (e.g., [61]) or pattern-growth algorithms (e.g., [103]). These define a minimum support (min-supp) threshold over subgraphs $X \subseteq G$. Only patterns that are frequent enough are included in the summary. Pattern mining methods are approximate summaries of the input graph G, as they do not include subgraphs that occur infrequently (thus, the summarization function is no longer homomorphic). However, setting min-supp = 1 usually produces a lossless summary, equivalent to a structural summary computation, as no subgraphs are omitted. Pattern mining methods also have interesting features such as automatically mining specific types of subgraphs like cliques, (bipartite) cores, stars, and chains [37,60]. While star-shaped subgraphs and chains are in principle also supported by quotient summaries, the difference here is that the selection of edges in star patterns and the length of the chains is determined in a data-driven way, rather than being pre-defined in a summary model. Finally, some pattern mining methods are also approximate because they use approximate methods such as locality-sensitive hashing (LSH) to assign graph vertices to the summary [65]. This is an inaccuracy introduced by the LSH function but not a characteristic of the underlying frequent pattern mining algorithms.

7.3 Statistical Methods for Graph Summarization

Statistical methods for graph summarization summarize the contents of a graph quantitatively such as by counting occurrences of edge labels or computing histograms over the labels [19] and define further constraints on the summary models. For example, in k-SNAP [94], the number k of summary vertices in a summary graph can be specified by the user, which controls the size of the summaries. The summarization operation k-SNAP [93] minimizes a function based on occurrences of user-selected edge labels to produce a summary graph SG, which contains exactly k vertex summaries. In its top-down approach, it starts by partitioning the graph based on user-selected vertex attributes. Afterward, the algorithm splits elements (vertex summaries) of the partition based on the aforementioned function, until the partition's size is k. Combining the first step, partitioning vertices by label, and the second step, minimizing a function that considers edge labels, k-SNAP can be considered a hybrid approach, combining structural and statistical concepts. CANAL is an extension of k-SNAP that supports numeric edge attributes [106]. Summarizing edges labeled with numeric values is approached by bucketing the values into predefined categories. Thus, the problem of supporting unbounded numerical values is reduced to summarizing graphs with discrete categories only.

8 Conclusion and Outlook

8.1 Conclusion

We delved into the domain of graph summarization, a process aimed at generating concise representations of input graphs while preserving specific structural attributes. Particularly, we focus on structural graph summaries that can be applied to semantic graphs, i. e., labeled graphs such as in the RDF or provided as labeled property graphs.

Different approaches and algorithms have been developed to address graph summarization. Our focus has been on exploring the state-of-the-art methods for structural graph summarization based on quotients. We have examined the relationship of structural summarization with other pertinent fields, like the well-known k-bisimulation. A noteworthy observation is the natural connection between structural summarization and logics, owing to the discrete structures under consideration.

A. Scherp, D. Richerby, T. Blume, M. Cochez, and J. Rau

Finally, there are also hybrid approaches for graph summarization. They combine features of quotient-based structural and non-quotient statistical and pattern mining techniques [19]. One such example is the combination of bisimulation with clustering [3]. Hierarchical complete-link clustering is applied over vertex features to group the vertices. Similar, Wang et al. cluster paths in a graph to compute an approximate summary [101].

8.2 Outlook

This research contributes to a deeper understanding of graph summarization techniques and opens avenues for future advancements in this domain. As directions of future research, we see:

- 1. Multi summaries: Existing works summarize a graph with respect to a single, defined summary model only. Multi-summaries compute multiple condensed representations of the input graph at once, stored in a joined data structure.
- 2. Summaries on temporal graphs: while there are already graph summarization approaches for temporal graphs, including incremental summarization [9], there is still a lot of work to be done. For example, we are currently missing approaches for summarization of rapidly evolving graphs such as social media graphs.
- **3.** Task-specific learned summaries: Graph summarization addresses different application needs. Example applications are outlined in Section 2. Interesting future work would be to automatically learn which features of a summary model are most relevant to a task (in general) or to the workload of a task (e.g., the kinds of queries executed in the data search scenario).
- 4. Exploiting modern hardware such as GPUs: We have already briefly reflected on parallel and distributed computation of graph summaries in Section 4.3. First steps have been taken on using GPUs but there is still a lack of research in this direction.
- 5. PyGraphSum: Each graph summarization model and algorithm typically comes with its own implementation, datasets, and evaluation measures. Comparing different algorithms and methods is difficult and cumbersome, as a common standard library for efficient distributed summarization of static and temporal graphs is missing. A standardized library that is used among industry and researchers alike will contribute not only to more transparency and comparability of the different approaches but also accelerate research in the field by facilitating reuse.

— References

- 1 Dean Allemang and James A. Hendler. Semantic Web for the Working Ontologist - Effective Modeling in RDFS and OWL, Second Edition. Morgan Kaufmann, 2011. URL: http://www.elsevierdirect.com/product. jsp?isbn=9780123859655.
- 2 James F. Allen. Maintaining knowledge about temporal intervals. Commun. ACM, 26(11):832-843, 1983. doi:10.1145/182.358434.
- 3 Anas Alzogbi and Georg Lausen. Similar structures inside RDF-graphs. In Workshop on Linked Data on the Web. CEUR-WS.org, 2013. URL: http://ceur-ws.org/Vol-996/ papers/ldow2013-paper-05.pdf.
- 4 M. A. Baazizi, H. Ben Lahmar, D. Colazzo, G. Ghelli, and C. Sartiani. Schema inference for massive JSON datasets. In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT*, pages 222–233. OpenProceedings.org, 2017. doi:10.5441/002/EDBT.2017.21.

- 5 Charles W. Bachman. Data structure diagrams. Data Base, 1(2):4–10, 1969. doi:10.1145/1017466. 1017467.
- 6 Fabio Benedetti, Sonia Bergamaschi, and Laura Po. Exposing the underlying schema of LOD sources. In Web Intelligence (WI), pages 301–304. IEEE, 2015. doi:10.1109/WI-IAT.2015.99.
- 7 Laurent Bindschaedler, Jasmina Malicevic, Baptiste Lepers, Ashvin Goel, and Willy Zwaenepoel. Tesseract: distributed, general graph pattern mining on evolving graphs. In *European Conf. on Comp. Systems (EuroSys)*, pages 458–473. ACM, 2021. doi:10.1145/3447786.3456253.
- 8 Till Blume. Semantic structural graph summaries for evolving and distributed graphs. PhD thesis, University of Ulm, Germany, 2022. URL: https://nbn-resolving.org/urn:nbn:de: bsz:289-oparu-46050-1.
- 9 Till Blume, David Richerby, and Ansgar Scherp. Incremental and parallel computation of structural graph summaries for evolving graphs. In

Int. Conf. on Inform. and Knowledge Management (CIKM), pages 75-84. ACM, 2020. doi: 10.1145/3340531.3411878.

- 10 Till Blume, David Richerby, and Ansgar Scherp. FLUID: A common model for semantic structural graph summaries based on equivalence relations. *Theoretical Computer Science*, 854:136–158, 2021. doi:10.1016/J.TCS.2020.12.019.
- 11 Till Blume and Ansgar Scherp. Indexing data on the web: A comparison of schema-level indices for data search. In *Database and Expert Systems Applications (DEXA)*, pages 277–286, 2020. doi:10.1007/978-3-030-59051-2_18.
- 12 Paolo Boldi and Sebastiano Vigna. Axioms for centrality. Internet Math., 10(3-4):222-262, 2014. doi:10.1080/15427951.2013.865686.
- 13 Jeroen Bollen, Jasper Steegmans, Jan Van den Bussche, and Stijn Vansummeren. Learning graph neural networks using exact compression. In Proceedings of the 6th Joint Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA), pages 8:1–8:9. ACM, 2023. doi:10.1145/3594778.3594878.
- 14 A. Bonifati, S. Dumbrava, and H. Kondylakis. Graph summarization. CoRR, abs/2004.14794, 2020. arXiv:2004.14794.
- 15 Redouane Bouhamoum, Zoubida Kedad, and Stéphane Lopes. Incremental schema discovery at scale for RDF data. In *The Semantic Web* - 18th International Conference, ESWC, volume 12731 of Lecture Notes in Computer Science, pages 195–211. Springer, 2021. doi:10.1007/ 978-3-030-77385-4_12.
- 16 D. Brickley and R.V. Guha. RDF Schema 1.1, 2014. URL: https://www.w3.org/TR/2014/ REC-rdf-schema-20140225/.
- 17 Peter Buneman and Slawek Staworko. RDF graph alignment with bisimulation. Proc. VLDB Endow., 9(12):1149–1160, 2016. doi:10.14778/2994509. 2994531.
- 18 Stéphane Campinas, Thomas Perry, Diego Ceccarelli, Renaud Delbru, and Giovanni Tummarello. Introducing RDF graph summary with application to assisted SPARQL formulation. In *Database and Expert Systems Applications (DEXA)*, pages 261–266. IEEE, 2012. doi:10.1109/DEXA.2012.38.
- 19 Šejla Čebirić, François Goasdoué, Haridimos Kondylakis, Dimitris Kotzinos, Ioana Manolescu, Georgia Troullinou, and Mussab Zneika. Summarizing semantic graphs: a survey. VLDB J., 28(3):295–327, 2019. doi:10.1007/S00778-018-0528-3.
- 20 Šejla Čebirić, François Goasdoué, and Ioana Manolescu. A framework for efficient representative summarization of RDF graphs. In Int. Semantic Web Conf. (ISWC). CEUR-WS.org, 2017. URL: http://ceur-ws.org/Vol-1963/paper512.pdf.
- 21 S. Ceri, G. Gottlob, and L. Tanca. What you always wanted to know about Datalog (but never dared to ask). *Trans. Knowl. Data Eng.*, 1(1):146– 166, 1989. doi:10.1109/69.43410.
- 22 Marek Ciglan, Kjetil Nørvåg, and Ladislav Hluchý. The SemSets model for ad-hoc semantic list search. In World Wide Web Conf. (WWW), pages 131– 140, 2012. doi:10.1145/2187836.2187855.

- 23 Mariano P. Consens, Valeria Fionda, Shahan Khatchadourian, and Giuseppe Pirrò. S+EPPs: Construct and explore bisimulation summaries, plus optimize navigational queries; all on existing SPARQL systems. *PVLDB*, 8(12):2028–2031, 2015. doi:10.14778/2824032.2824128.
- 24 Richard Cyganiak, David Wood, and Markus Lanthaler. RDF 1.1 Concepts and Abstract Syntax, 2014. URL: http://www.w3.org/TR/2014/ REC-rdf11-concepts-20140225.
- 25 Reinhard Diestel. Graph Theory, volume 173 of Graduate Texts in Mathematics. Springer, 5th edition, 2016. doi:10.1007/978-3-662-53622-3.
- 26 Li Ding, Joshua Shinavier, Zhenning Shangguan, and Deborah L. McGuinness. SameAs networks and beyond: Analyzing deployment status and implications of owl:sameAs in Linked Data. In Int. Semantic Web Conf. (ISWC), pages 145–160. Springer, 2010. doi:10.1007/ 978-3-642-17746-0_10.
- 27 Agostino Dovier, Carla Piazza, and Alberto Policriti. An efficient algorithm for computing bisimulation equivalence. *Theor. Comput. Sci.*, 311(1-3):221–256, 2004. doi:10.1016/S0304-3975(03) 00361-X.
- 28 Chi Thang Duong, Dung Hoang, Hongzhi Yin, Matthias Weidlich, Quoc Viet Hung Nguyen, and Karl Aberer. Efficient streaming subgraph isomorphism with graph neural networks. VLDB Endow., 14(5):730–742, 2021. doi:10.14778/3446095. 3446097.
- 29 M. Dürst and M. Suignard. RFC 3987 int.ized resource identifiers (IRIs), 2005. URL: https: //www.ietf.org/rfc/rfc3987.txt.
- 30 Gregor Engels, Claus Lewerentz, Wilhelm Schäfer, Andy Schürr, and Bernhard Westfechtel. Graph transformations and model-driven engineering. In Graph Transformations and Model-Driven Engineering, pages 1–5. Springer, 2010. doi:10.1007/ 978-3-642-17322-6_1.
- 31 Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity* of *Computation*, volume 7, pages 43–73, 1974.
- 32 W. Fan, J. Li, J. Luo, Z. Tan, X. Wang, and Y. Wu. Incremental graph pattern matching. In *Management of Data (SIGMOD)*, pages 925–936. ACM, 2011. doi:10.1145/1989323.1989420.
- 33 Wenfei Fan, Jianzhong Li, Xin Wang, and Yinghui Wu. Query preserving graph compression. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 157–168. ACM, 2012. doi:10.1145/2213836.2213855.
- 34 Wenfei Fan, Yuanhao Li, Muyang Liu, and Can Lu. Making graphs compact by lossless contraction. In SIGMOD '21: International Conference on Management of Data, pages 472–484. ACM, 2021. doi:10.1145/3448016.3452797.
- 35 Wenfei Fan, Yuanhao Li, Muyang Liu, and Can Lu. A hierarchical contraction scheme for querying big graphs. In SIGMOD '22: International Conference on Management of Data, pages 1726–1740. ACM, 2022. doi:10.1145/3514221.3517862.
- 36 Wenfei Fan, Yinghui Wu, and Jingbo Xu. Adding counting quantifiers to graph patterns. In *Management of Data (SIGMOD)*, pages 1215–1230. ACM, 2016. doi:10.1145/2882903.2882937.

- 37 Yixiang Fang, Xin Huang, Lu Qin, Ying Zhang, Wenjie Zhang, Reynold Cheng, and Xuemin Lin. A survey of community search over big graphs. *VLDB J.*, 29(1):353–392, 2020. doi:10.1007/ S00778-019-00556-X.
- 38 Javier D. Fernández, Miguel A. Martínez-Prieto, Claudio Gutierrez, Axel Polleres, and Mario Arias. Binary RDF representation for publication and exchange (HDT). J. Web Semant., 19:22–41, 2013. doi:10.1016/J.WEBSEM.2013.01.002.
- 39 Alessandro Generale, Till Blume, and Michael Cochez. Scaling R-GCN training with graph summarization. In *Companion of The Web Conference 2022*, pages 1073–1082. ACM, 2022. doi: 10.1145/3487553.3524719.
- 40 François Goasdoué, Pawel Guzewicz, and Ioana Manolescu. Incremental structural summarization of RDF graphs. In Proceedings of the International Conference on Extending Database Technology (EDBT), pages 566–569. OpenProceedings.org, 2019. doi:10.5441/002/EDBT.2019.57.
- 41 François Goasdoué, Pawel Guzewicz, and Ioana Manolescu. RDF graph summarization for firstsight structure discovery. VLDB J., 29(5):1191– 1218, 2020. doi:10.1007/S00778-020-00611-Y.
- 42 Roy Goldman and Jennifer Widom. DataGuides: Enabling query formulation and optimization in semistructured databases. In VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases, pages 436-445. Morgan Kaufmann, 1997. URL: http://www.vldb.org/conf/ 1997/P436.PDF.
- 43 Thomas Gottron, Ansgar Scherp, Bastian Krayer, and Arne Peters. LODatio: using a schema-level index to support users in finding relevant sources of linked data. In *Knowledge Capture (K-CAP)*, pages 105–108. ACM, 2013. doi:10.1145/2479832. 2479841.
- 44 Mahdi Hajiabadi, Venkatesh Srinivasan, and Alex Thomo. Dynamic graph summarization: Optimal and scalable. In *IEEE International Conference on Big Data*, pages 545–554. IEEE, 2022. doi:10.1109/BIGDATA55660.2022.10020422.
- 45 William L. Hamilton. Graph Representation Learning. Morgan and Claypool, 2020. URL: https: //www.cs.mcgill.ca/~wlh/grl_book/.
- 46 Wook-Shin Han, Jinsoo Lee, Minh-Duc Pham, and Jeffrey Xu Yu. iGraph: A framework for comparisons of disk-based graph indexing techniques. Proceedings of the International Conference on Very Large Data Bases (VLDB) Endowment, 3(1):449– 459, 2010. doi:10.14778/1920841.1920901.
- 47 Jan Hegewald, Felix Naumann, and Melanie Weis. XStruct: Efficient schema extraction from multiple and large XML documents. In Proceedings of the International Conference on Data Engineering Workshops (ICDE), page 81. IEEE Computer Society, 2006. doi:10.1109/ICDEW.2006.166.
- 48 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, and others. Knowledge graphs. CoRR, abs/2003.02320, 2020. arXiv:2003.02320.
- 49 Neil Immerman. Descriptive Complexity. Springer, 1999. doi:10.1007/978-1-4612-0539-5.
- 50 Xiaowei Jiang, Xiang Zhang, Feifei Gao, Chunan Pu, and Peng Wang. Graph compression

strategies for instance-focused semantic mining. In Chinese Semantic Web Symposium/Chinese Web Science Conf., pages 50–61. Springer, 2013. doi:10.1007/978-3-642-54025-7_5.

- 51 Tobias Käfer, Ahmed Abdelrahman, Jürgen Umbrich, Patrick O'Byrne, and Aidan Hogan. Observing linked data dynamics. In Proceedings of the Extended Semantic Web Conference (ESWC), volume 7882 of Lecture Notes in Computer Science, pages 213–227. Springer, 2013. doi:10.1007/978-3-642-38288-8_15.
- 52 A. Kansal and F. Spezzano. A scalable graph-coarsening based index for dynamic graph databases. In Int. Conf. on Information and Knowledge Management (CIKM), pages 207–216. ACM, 2017. doi:10.1145/3132847.3133003.
- 53 Raghav Kaushik, Pradeep Shenoy, Philip Bohannon, and Ehud Gudes. Exploiting local similarity for indexing paths in graph-structured data. In Int. Conf. on Data Engineering (ICDE), pages 129–140. IEEE, 2002. doi:10.1109/ICDE.2002.994703.
- 54 K. Kellou-Menouer and Z. Kedad. Schema discovery in RDF data sources. In Int. Conf. on Conceptual Modeling (ER), volume 9381 of Lecture Notes in Computer Science, pages 481–495. Springer, 2015. doi:10.1007/978-3-319-25264-3_36.
- 55 Arijit Khan, Sourav S. Bhowmick, and Francesco Bonchi. Summarizing static and dynamic big graphs. VLDB Endowment, 10(12):1981–1984, 2017. doi:10.14778/3137765.3137825.
- 56 Kifayat-Ullah Khan, Waqas Nawaz, and Young-Koo Lee. Set-based approximate approach for lossless graph summarization. Computing, 97(12):1185–1207, 2015. doi:10.1007/ S00607-015-0454-9.
- 57 Aleks Kissinger, Alex Merry, and Matvey Soloviev. Pattern graph rewrite systems. In Int. Workshop on Developments in Computational Models (DCM), pages 54–66, 2012. doi:10.4204/EPTCS. 143.5.
- 58 Jihoon Ko, Yunbum Kook, and Kijung Shin. Incremental lossless graph summarization. In Conf. on Knowledge Discovery and Data Mining (SIGKDD), pages 317–327. ACM, 2020. doi: 10.1145/3394486.3403074.
- 59 M. Konrath, T. Gottron, S. Staab, and A. Scherp. SchemEX - efficient construction of a data catalogue by stream-based indexing of linked data. J. Web Semant., 16:52–58, 2012. doi:10.1016/J. WEBSEM.2012.06.002.
- 60 Danai Koutra, U Kang, Jilles Vreeken, and Christos Faloutsos. Summarizing and understanding large graphs. Stat. Anal. Data Min., 8(3):183–202, 2015. doi:10.1002/SAM.11267.
- 61 Michihiro Kuramochi and George Karypis. Frequent subgraph discovery. In Int. Conf. on Data Mining (ICDM), pages 313–320. IEEE, 2001. doi: 10.1109/ICDM.2001.989534.
- 62 Kostis Kyzirakos, Manos Karpathiotakis, Konstantina Bereta, George Garbis, Charalampos Nikolaou, Panayiotis Smeros, Stella Giannakopoulou, Kallirroi Dogani, and Manolis Koubarakis. The spatiotemporal RDF store Strabon. In Advances in Spatial and Temporal Databases - 13th International Symposium,

volume 8098 of Lecture Notes in Computer Science, pages 496–500. Springer, 2013. doi:10.1007/978-3-642-40235-7_35.

- 63 Kristen LeFevre and Evimaria Terzi. Grass: Graph structure summarization. In SIAM International Conference on Data Mining, SDM, pages 454–465. SIAM, 2010. doi:10.1137/1.9781611972801.40.
- 64 Leonid Libkin. Elements of Finite Model Theory. Springer, 2004. doi:10.1007/ 978-3-662-07003-1.
- 65 Xingjie Liu, Yuanyuan Tian, Qi He, Wang-Chien Lee, and John McPherson. Distributed graph summarization. In Int. Conf. on Information and Knowledge Management (CIKM), pages 799–808. ACM, 2014. doi:10.1145/2661829.2661862.
- 66 Yike Liu, Tara Safavi, Abhilash Dighe, and Danai Koutra. Graph summarization methods and applications: A survey. ACM Comput. Surv., 51(3):62:1– 62:34, 2018. doi:10.1145/3186727.
- 67 Yongming Luo, George H. L. Fletcher, Jan Hidders, Paul De Bra, and Yuqing Wu. Regularities and dynamics in bisimulation reductions of big graphs. In *First International Workshop on Graph Data Management Experiences and Systems*, *GRADES*, page 13. CWI/ACM, 2013. doi:10.1145/2484425.2484438.
- 68 Yongming Luo, George H. L. Fletcher, Jan Hidders, Yuqing Wu, and Paul De Bra. External memory k-bisimulation reduction of big graphs. In 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, pages 919– 928. ACM, 2013. doi:10.1145/2505515.2505752.
- 69 Jan Martens, Jan Friso Groote, Lars B. van den Haak, Pieter Hijma, and Anton Wijs. A linear parallel algorithm to compute bisimulation and relational coarsest partitions. In *Formal Aspects of Component Software (FACS)*, volume 13077 of *LNCS*, pages 115–133. Springer, 2021. doi:10.1007/978-3-030-90636-8_7.
- 70 D. L. McGuinness and F. van Harmelen. OWL Web Ontology Language, 2014. [Online, accessed: December 18, 2023]. URL: https://www.w3.org/ TR/2004/REC-owl-features-20040210/.
- 71 Nandana Mihindukulasooriya, María Poveda-Villalón, Raúl García-Castro, and Asunción Gómez-Pérez. Loupe – an online tool for inspecting datasets in the linked data cloud. In *Int. Semantic Web Conf. (ISWC)*. CEUR-WS.org, 2015. URL: http://ceur-ws.org/Vol-1486/paper_113.pdf.
- 72 Tova Milo and Dan Suciu. Index structures for path expressions. In Int. Conf. Database Theory (ICDT), pages 277–295. Springer, 1999. doi: 10.1007/3-540-49257-7_18.
- 73 Seunghwan Min, Sung Gwan Park, Kunsoo Park, Dora Giammarresi, Giuseppe F. Italiano, and Wook-Shin Han. Symmetric continuous subgraph matching with bidirectional dynamic programming. *Proc. VLDB Endow.*, 14(8):1298–1310, 2021. doi:10.14778/3457390.3457395.
- 74 Muhammad Anis Uddin Nasir, Cigdem Aslay, Gianmarco De Francisci Morales, and Matteo Riondato. Tiptap: Approximate mining of frequent k-subgraph patterns in evolving graphs. ACM Trans. on Knowledge Discovery from Data, 15(3), apr 2021. doi:10.1145/3442590.

- 75 Saket Navlakha, Rajeev Rastogi, and Nisheeth Shrivastava. Graph summarization with bounded error. In ACM SIGMOD International Conference on Management of Data, pages 419–432. ACM, 2008. doi:10.1145/1376616.1376661.
- 76 Svetlozar Nestorov, Jeffrey D. Ullman, Janet L. Wiener, and Sudarshan S. Chawathe. Representative objects: Concise representations of semistructured, hierarchial data. In *Proceedings 13th Intl Conference on Data Engineering*, pages 79–90. IEEE Computer Society, 1997. doi:10.1109/ICDE. 1997.581741.
- 77 T. Neumann and G. Moerkotte. Characteristic sets: Accurate cardinality estimation for RDF queries with multiple joins. In *Int. Conf. on Data Engineering (ICDE)*, pages 984–994. IEEE, 2011. doi:10.1109/ICDE.2011.5767868.
- 78 Robert Paige and Robert Endre Tarjan. Three partition refinement algorithms. SIAM J. Comput., 16(6):973–989, 1987. doi:10.1137/0216062.
- 79 Alexandros Pappas, Georgia Troullinou, Giannis Roussakis, Haridimos Kondylakis, and Dimitris Plexousakis. Exploring importance measures for summarizing RDF/S KBs. In *The Semantic Web ESWC*, pages 387–403. Springer, 2017. doi:10.1007/978-3-319-58068-5_24.
- 80 Emmanuel Pietriga, Hande Gözükan, Caroline Appert, Marie Destandau, Sejla Čebirić, François Goasdoué, and Ioana Manolescu. Browsing linked data catalogs with LODAtlas. In Int. Semantic Web Conf. (ISWC), pages 137–153. Springer, 2018. doi:10.1007/978-3-030-00668-6_9.
- 81 Detlef Plump. Essentials of term graph rewriting. Electron. Notes Theor. Comput. Sci., 51:277-289, 2001. doi:10.1016/S1571-0661(04)80210-X.
- 82 Miao Qiao, Hao Zhang, and Hong Cheng. Subgraph matching: on compression and computation. Proceedings of the International Conference on Very Large Data Bases (VLDB) Endowment, 11(2):176–188, 2017. doi:10.14778/ 3149193.3149198.
- 83 Chen Qun, Andrew Lim, and Kian Win Ong. D(k)-index: An adaptive structural summary for graph-structured data. In *Management of Data* (SIGMOD), pages 134–144. ACM, 2003. doi: 10.1145/872757.872776.
- 84 Mohammad Rashid, Giuseppe Rizzo, Nandana Mihindukulasooriya, Marco Torchiano, and Óscar Corcho. KBQ - A tool for knowledge base quality assessment using evolution analysis. In Proceedings of Workshops and Tutorials of the International Conference on Knowledge Capture (K-CAP), volume 2065 of CEUR Workshop Proceedings, pages 58-63. CEUR-WS.org, 2017. URL: http://ceur-ws.org/Vol-2065/paper13.pdf.
- 85 Jannik Rau, David Richerby, and Ansgar Scherp. Computing k-bisimulations for large graphs: A comparison and efficiency analysis. In Graph Transformation - 16th International Conference, ICGT 2023, volume 13961 of Lecture Notes in Computer Science, pages 223–242. Springer, 2023. doi:10.1007/978-3-031-36709-0_12.
- 86 Davide Sangiorgi. On the origins of bisimulation and coinduction. ACM Trans. Program. Lang. Syst., 31(4):15:1-15:41, 2009. doi:10.1145/ 1516507.1516510.

- 87 Johann Schaible, Thomas Gottron, and Ansgar Scherp. TermPicker: Enabling the reuse of vocabulary terms by exploiting data from the Linked Open Data cloud. In *The Semantic Web (ESWC)*, pages 101–117. Springer, 2016. doi:10.1007/ 978-3-319-34129-3_7.
- 88 Alexander Schätzle, Antony Neu, Georg Lausen, and Martin Przyjaciel-Zablocki. Large-scale bisimulation of RDF graphs. In Semantic Web Information Management Workshop, pages 1–8. ACM, 2013. doi:10.1145/2484712.2484713.
- 89 Kijung Shin, Amol Ghoting, Myunghwan Kim, and Hema Raghavan. SWeG: Lossless and lossy summarization of web-scale graphs. In World Wide Web Conf. (WWW), pages 1679–1690. ACM, 2019. doi:10.1145/3308558.3313402.
- 90 Qi Song, Yinghui Wu, and Xin Luna Dong. Mining summaries for knowledge graph search. In *ICDM*, pages 1215–1220. IEEE, 2016. doi:10.1109/ICDM. 2016.0162.
- 91 Blerina Spahiu, Riccardo Porrini, Matteo Palmonari, Anisa Rula, and Andrea Maurino. AB-STAT: ontology-driven linked data summaries with pattern minimalization. In *The Semantic Web (ESWC)*, pages 381–395, 2016. doi: 10.1007/978-3-319-47602-5_51.
- 92 Giorgio Stefanoni, Boris Motik, and Egor V. Kostylev. Estimating the cardinality of conjunctive queries over RDF data using graph summarisation. In *Proceedings of the 2018 World Wide Web Conference*, pages 1043–1052. ACM, 2018. doi:10.1145/3178876.3186003.
- 93 Yuanyuan Tian, Richard A. Hankins, and Jignesh M. Patel. Efficient aggregation for graph summarization. In ACM SIGMOD International Conference on Management of Data, SIGMOD, pages 567-580. ACM, 2008. doi:10.1145/1376616. 1376675.
- 94 Yuanyuan Tian and Jignesh M. Patel. Interactive graph summarization. In Link Mining: Models, Algorithms, and Applications, pages 389–409. Springer, 2010. doi:10.1007/ 978-1-4419-6515-8_15.
- 95 Thanh Tran, Günter Ladwig, and Sebastian Rudolph. Managing structured and semistructured RDF data using structure indexes. *Trans. Knowl. Data Eng.*, 25(9):2076–2089, 2013. doi: 10.1109/TKDE.2012.134.
- 96 Thanh Tran, Haofen Wang, Sebastian Rudolph, and Philipp Cimiano. Top-k exploration of query candidates for efficient keyword search on graphshaped (RDF) data. In *Proceedings of the 25th*

International Conference on Data Engineering, pages 405–416. IEEE Computer Society, 2009. doi:10.1109/ICDE.2009.119.

- 97 Georgia Troullinou, Haridimos Kondylakis, Evangelia Daskalaki, and Dimitris Plexousakis. Ontology understanding without tears: The summarization approach. Semantic Web, 8(6):797–815, 2017. doi:10.3233/SW-170264.
- 98 Octavian Udrea, Andrea Pugliese, and V. S. Subrahmanian. GRIN: A graph based RDF index. In AAAI, pages 1465–1470. AAAI Press, 2007. URL: http://www.aaai.org/Library/AAAI/2007/ aaai07-232.php.
- 99 Johanna Völker and Mathias Niepert. Statistical schema induction. In Extended Semantic Web Conf. (ESWC), pages 124–138. Springer, 2011. doi:10.1007/978-3-642-21034-1_9.
- 100 Lanjun Wang, Oktie Hassanzadeh, Shuo Zhang, Juwei Shi, Limei Jiao, Jia Zou, and Chen Wang. Schema management for document stores. Proceedings of the International Conference on Very Large Data Bases (VLDB) Endowment, 8(9):922– 933, 2015. doi:10.14778/2777598.2777601.
- 101 Qiu Yue Wang, Jeffrey Xu Yu, and Kam-Fai Wong. Approximate graph schema extraction for semistructured data. In Advances in Database Technology (EDBT), pages 302–316. Springer, 2000. doi:10.1007/3-540-46439-5_21.
- 102 Jennifer Widom. Data management for XML: research directions. *IEEE Data Eng. Bull.*, 22(3):44-52, 1999. URL: http://sites.computer.org/ debull/99sept/jennifer.ps.
- 103 Xifeng Yan and Jiawei Han. gSpan: Graph-based substructure pattern mining. In Int. Conf. on Data Mining (ICDM), pages 721–724. IEEE, 2002. doi:10.1109/ICDM.2002.1184038.
- 104 D. Yuan, P. Mitra, H. Yu, and C. L. Giles. Iterative graph feature mining for graph indexing. In Int. Conf. on Data Engineering (ICDE), pages 198–209. IEEE, 2012. doi:10.1109/ICDE.2012.11.
- 105 D. Yuan, P. Mitra, H. Yu, and C. L. Giles. Updating graph indices with a one-pass algorithm. In *Management of Data (SIGMOD)*, pages 1903– 1916. ACM, 2015. doi:10.1145/2723372.2746482.
- 106 Ning Zhang, Yuanyuan Tian, and Jignesh M. Patel. Discovery-driven graph summarization. In Int. Conf. on Data Engineering (ICDE), pages 880–891. IEEE, 2010. doi:10.1109/ICDE.2010.5447830.
- 107 Xiang Zhang, Gong Cheng, and Yuzhong Qu. Ontology summarization based on RDF sentence graph. In World Wide Web (WWW), pages 707– 716. ACM, 2007. doi:10.1145/1242572.1242668.

Autonomy in the Age of Knowledge Graphs: Vision and Challenges

Jean-Paul Calbimonte 🖂 🖻

University of Applied Sciences and Arts Western Switzerland HES-SO, Sierre, Switzerland

Andrei Ciortea 🖂 回 Institute of Computer Science, University of St.Gallen, Switzerland

Timotheus Kampik 🖂 🕩

Department of Computing Science, Umeå University, Sweden SAP Signavio, Berlin, Germany

Simon Mayer ⊠© Institute of Computer Science, University of St.Gallen, Switzerland

Terry R. Payne 🖂 🗅 Department of Computer Science, University of Liverpool, United Kingdom

Valentina Tamma¹ 🖂 🗅 Department of Computer Science, University of Liverpool, United Kingdom

Antoine Zimmermann 🖂 回

Mines Saint-Etienne, Université Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, France

— Abstract -

In this position paper, we propose that Knowledge Graphs (KGs) are one of the prime approaches to support the programming of autonomous software systems at the knowledge level. From this viewpoint, we survey how KGs can support different dimensions of autonomy in such systems: For example, the autonomy of systems with respect to their environment, or with respect to organisations; and we discuss related practical and research challenges. We emphasise that KGs need to be able to support systems of autonomous software agents that are themselves *highly heterogeneous*, which limits how these systems may use KGs. Furthermore, these heterogeneous software agents may populate highly dynamic environments, which im-

plies that they require adaptive KGs. The scale of the envisioned systems – possibly stretching to the size of the Internet - highlights the maintainability of the underlying KGs that need to contain large-scale knowledge, which requires that KGs are maintained jointly by humans and machines. Furthermore, autonomous agents require procedural knowledge, and KGs should hence be explored more towards the provisioning of such knowledge to augment autonomous behaviour. Finally, we highlight the importance of *modelling choices*, including with respect to the selected abstraction level when modelling and with respect to the provisioning of more expressive constraint languages.

2012 ACM Subject Classification Computing methodologies \rightarrow Multi-agent systems; Computing methodologies \rightarrow Intelligent agents; Computer systems organization \rightarrow Self-organizing autonomic computing; Computing methodologies \rightarrow Knowledge representation and reasoning; Information systems \rightarrow Semantic web description languages

Keywords and phrases Knowledge graphs, Autonomous Systems

Digital Object Identifier 10.4230/TGDK.1.1.13

Category Vision

¹ Corresponding author



© Jean-Paul Calbimonte, Andrei Ciortea, Timotheus Kampik, Simon Mayer, Terry R. Payne, Valentina Tamma, and Antoine Zimmermann; licensed under Creative Commons License CC-BY 4.0



Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 13, pp. 13:1-13:22 Transactions on Graph Data and Knowledge TGDK Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

13:2 Autonomy in the Age of Knowledge Graphs

Funding This research has received funding from the Swiss National Science Foundation under grant No. 189474 (*HyperAgents*), grant No. 213369 (*StreamDKG*), from the European Union's Horizon 2020 research and innovation program under grant No. 957218 (*IntellIoT*), and from the Horizon Europe program under grant No. 101092908 (*SmartEdge*).

Acknowledgements We would like to thank Chiara Ghidini for the useful discussions on modelling processes using KGs to support process execution, mining and discovery.

Received 2023-09-18 Accepted 2023-11-17 Published 2023-12-19

Editors Aidan Hogan, Ian Horrocks, Andreas Hotho, and Lalana Kagal

Special Issue Trends in Graph Data and Knowledge

Author Contributions All authors participated in the planning and continuous refinements of this work, and hence contributed to the abstract, *Introduction* and *Conclusions* sections. JPC primarily contributed to Sections 3.5, 3.1, and 5. AC primarily contributed to Sections 3.1 and 3.4. TK primarily contributed to Sections 3.2, 3.3, 3.4, and 5. SM primarily contributed to Sections 3 and 5, as well as to the main tenet of the paper. TRP contributed to Section 1, 3, and 4, and to the general arguments in the paper. VT primarily contributed to Section 2 and Section 4, as well as to the main tenet of the paper.

1 Introduction: Autonomy

Knowledge exists in different forms and for different purposes. Some of this knowledge may describe how the world or the universe works, and thus is *ontological* in nature: It describes what entities exist, how they can be categorised, and how they relate to each other. This knowledge, coupled with propositional knowledge (i.e., beliefs or assertions, which themselves may be true or false), can be used by computational entities (e.g., autonomous agents), to take informed decisions as to whether action is currently required. However, such knowledge does not typically describe what an agent needs to do in order to achieve a specific goal, and cannot be directly used by an agent to decide on what their next action or task should be. Conversely, knowledge may hence be procedural, in that it describes how an agent can achieve a task, either in isolation or through collaboration with other agents or services within its environment. This is analogous to, for example, the use of a user-manual that contains information about how customers can assemble their furniture by themselves, or the instructions for an espresso machine that allows the user to independently diagnose and rectify malfunctions. Knowledge may also be normative [20] (i.e., expressing notions such as prohibition, permission, obligation, or dispensation), which can be used to describe communities of agents and their inherent organisation (for example, when describing institutions [63]). This knowledge may be represented by, for example, a charter for a group that defines the mission of that group and the different roles that members may adopt, together with corresponding obligations that constrain their autonomy to actions that are in conformance with the group's objectives; or the terms of use of a service such as an auction house that defines the rules that must be followed when bidding for goods without explicitly hard-coding this behaviour, thus allowing bidders to decide whether or not to comply with the rules – or to understand the sanctions that may be imposed if the rules are violated.

These different types of knowledge sources hold much of the information necessary for agents – natural or artificial – to facilitate their *autonomy*. To express such knowledge, the use of *Knowledge Graphs (KGs)* [38] has gained prominence across numerous research areas and application domains, due to its interconnected nature, expressivity, and the increasing variety of tools and techniques to support querying and reasoning at scale. KGs build upon a graph representation whereby objects (e.g., things, actions, tasks, affordances, methods, and capabilities) are represented as *nodes* and the relationships between them as *edges*. This type of representation, instantiated in the Resource

J.-P. Calbimonte et al.

Description Framework (RDF) data model [22], has been used not only to represent *ontological* definitions and *propositional* definitions, but also *normative* knowledge [33], as we discuss in Section 4. As such, we claim that KGs are particularly useful in facilitating *autonomy*; KGs can be used to identify the specific vocabulary needed to represent autonomy and the types of reasoning tasks that are supported by the associated ontological models. This can be facilitated through the development of KG schemas that model autonomous behaviour, including the definition of entities (in terms of modules or fragments of hierarchical knowledge) that represent both static (e.g., *agent*, *object*, or *scene*) and dynamic (e.g., *task*, *goal*, or *action*) aspects of autonomy. The possibility to reuse and augment KGs promotes interoperability, which is further supported through the existence of highly interconnected and expressive schemata.

Sharing the goal of making data machine-readable and machine-understandable, KGs are well-aligned with the Semantic Web, which today consists of a plethora of knowledge sources that are interlinked and discoverable through a variety of different mechanisms. This provides more than simple access to static knowledge: Online services, data sources, and IoT devices that perform actions and perceive their environment all make up a rich and dynamic resource that is inherently human focused. This human focus is primarily due to the Web's hypermedia-driven interaction model that has already made much of the (human) Web content intuitively explorable and usable by humans without specialised training. The emergence of the Semantic Web has facilitated machine-oriented access and comprehension to much of this knowledge and, through the integration of KGs, it has recently become an increasingly rich knowledge base that can support autonomous behaviour by artificial agents, i.e., autonomous software systems, on the Web. Already today, the Web enables a variety of heterogeneous stakeholders to achieve their goals; whether they are reactive agents that are triggered to perform actions based on a set of preconditions, or humans who meet, transact, and interact. Furthermore, experiences by and interactions between these heterogeneous agents, through services (such as marketplaces and electronic banking), or social media (such as collaborative software development or data curation) can be interpreted as early signs of life of an emerging global social machine [78].

In this paper, we argue that autonomy and KGs are mutually supportive and beneficial, with the potential for both to gain from the resulting synergy. While many of our observations also hold for intelligent agents in general (including humans), our paper is focused on Web-based software agents. And we argue that there is a need to create systems that support the programming of such artificial agents at the knowledge level. Our central claim is that many of the underlying tenets of autonomy can be represented using KGs, autonomous agents may emerge that can transact across the Web – without the need for developers to encode specific protocols that the agents should use to interact with Web resources, specific information about other agents that they may encounter, or specific information about normative relationships. Likewise, by providing the procedural frameworks that permit agents to interact with services and institutions that they may encounter, the agents can potentially operate with a similar sense of autonomy that human users exhibit. Furthermore, since this Web-embedded knowledge is not read-only, autonomous Web Agents² may construct and expose new knowledge, and they may evolve knowledge that they find to be stale. We explore how any autonomous agent on the Web – human as well as artificial – could benefit from such knowledge, and how they may evolve this knowledge in turn. We examine how KGs provide a natural abstraction for representing the knowledge required by autonomous agents on an open, dynamic Web, and consider the current limitations of using such a formalism for autonomy. Our aim is not to provide an exhaustive treatment of KGs, but rather to provide the foundations for discussing how KGs have been used or could be used to support different aspects

² See https://www.w3.org/community/webagents/

13:4 Autonomy in the Age of Knowledge Graphs

of autonomy. This allows us to map the different characteristics of autonomy to the knowledge needed to support them and their representation, thereby highlighting the role that KGs play in current autonomous applications and pave the way for identifying challenges and open problems.

We first provide a brief introduction to KGs and the way they are represented. This is followed, in Section 3, by a discussion of how KGs can be used to support autonomy with respect to its individual dimensions as identified by Castelfranchi and Falcone [17]; we also discuss the evolution of KGs themselves in such open environments. While we argue that KGs are well suited to overcome many of the obstacles on the way to the programming of agents on the knowledge level, many hurdles still remain, as identified in Section 4. We therefore present the limitations of KGs with respect to supporting autonomy – at the moment and in principle – in Section 5, and we distill open questions from our discussion in Section 6.

2 Knowledge Graphs from an Autonomous Agents' Perspective

Autonomous software agents can be characterised as computational systems that acquire data from their environment (both sensory and from interactions with others), and reason about some course of action that may bring them closer to achieving their goals. These actions are then realised; this could be through the usage of some service, collaboration with other agents (through an appropriate communication mechanism), the creation of new knowledge, or the curation or modification of existing knowledge. This cycle repeats, typically without the direct intervention of humans in the goal-achievement process, which results in autonomous behaviour that can be proactive as well as reactive. However, for each step, the agents require a variety of different types of knowledge that may describe: their environment (see Section 3.1); other agents and their abilities (see Section 3.2); the relations, roles, and any form of normative organisation that they or others may be part of (see Section 3.3); and the structure of the problem domain (see Section 3.4). Although this knowledge is typically encoded implicitly or assumptions are made by many closed multi-agent systems, the ability to share this knowledge (e.g., for service discovery or to facilitate adherence to a normative institution) and to together evolve it (see Section 3.5) facilitates interoperability and reuse across different agent communities, which in turn can also increase reliability, veracity, and scalability for numerous agent tasks. By representing and sharing this knowledge in the form of KGs, we argue that the synergy and opportunities for agents within a knowledge- and service-rich environment are greatly increased, especially when the agents interact in open environments such as on the Web.

The use of KGs is today becoming more prominent across research areas and application domains, and within systems where graph-oriented data representation is used alongside relational databases or proprietary format files. Yet despite their increasing popularity, there is no single accepted definition of a KG [28]. In this context, we define a KG as a directed, edge-labelled multi-graph (whereby there may be different edges between two nodes that may have different relationships or types). In a KG, the *nodes* (or *vertices*) are labelled and represent real-world entities of interest (e.g., objects, actions, tasks, affordances, methods, and capabilities) whereas directed edges between a source and a target node are labelled with names that represents the relationships between these entities. This can lead to the simple formalisation of a KG as a set of triples "node–edge–node", which is typically instantiated in the RDF data model [22]. However, other more involved structures may exist, such as property graphs [61] that allow nodes and edges to have a structure of their own. Whilst there is no commonly accepted definition of KGs, there is some agreement on the characteristics they should exhibit, based on the initial characterisation proposed by Paulheim [70], in that they:

J.-P. Calbimonte et al.

- 1. represent (semantically-described) real-world entities and their interrelations by means of links, which are organised in a graph;
- 2. define entities and their relations by means of data models referred to as *ontologies* (vocabularies or schemata);
- 3. support potentially interrelating arbitrary entities with each other;
- 4. cover a variety of domains that span different topics and are often expressed at different levels of granularity, ranging from common-sense knowledge to domain- and task-specific knowledge.

KGs integrate both static data collections and stream data from multiple sources, thereby supporting sophisticated querying and exploration of this data in a seamless way [52]. This helps overcome the inherent abstraction gap that exists in the environment between raw sensor data and the higher-level knowledge needed by artefacts that produce and consume complex data to capture realistic models of the world. Thus, KGs have been used in a number of different contexts, including life sciences and linguistics, and a number of representations have been proposed. A symbolic language representation has typically been used by those KGs that aim to support autonomy, particularly those with precise semantics that support agreement on the vocabulary used and the reasoning services that infer new facts and check the consistency of the dataset. Semantic Web languages such as RDF³ and OWL [69] have become prominent candidates as: 1) they combine machine-readable syntax with a clear formal semantics; 2) they are sufficiently expressive to model vocabularies for disparate domains with sufficient detail; and 3) they provide extensive reasoning capabilities that range from type inheritance and type inference (RDFS) to richer reasoning services, e.g., classification, consistency checking, entailment checking, explanations and instance realisation (OWL), facilitated by the existence of different reasoner implementations.

Depending on the expressivity required by the downstream application, KGs can be distinguished as those that only possess relational statements in the form of RDF triples with no schema, and those that conform to an explicit schema (i.e., an *ontology*) that is expressed in RDFS or OWL. In the latter case, the expressivity of the schema depends on the constructors used in its axiom definitions.⁴ Expressive ontologies are highly interconnected, promote semantic interoperability between systems, and better support collaborative decision-making [1, 13]. However, despite efforts to promote the establishment of semantic mappings between entities [49, 66] and interlinking of data, data sources are often siloed, existing in isolation and are not interlinked with other sources, or they may be schema-less (or use a proprietary, unpublished, schema), with limited reuse or mappings to other data sources [36].

In addition to representing data and their schema in KGs using formal languages (i.e., RDF for data and RDFS or OWL for the schema), the Semantic Web stack also provides a standard for validating graphs in order to ensure the quality of data. The Shapes Constraint Language (SHACL) [46] is used to validate RDF graphs against a set of conditions described as "shapes" of data over nodes and edges on top of a KG schema expressed in OWL. Shapes are applied to "targets" and form a shapes graph defining constraints, i.e., the conditions against which a data graph is evaluated. A data graph is deemed valid if all the targets conform to their related shapes; furthermore, the result of the validation is a report expressed in RDF that can also be processed and queried using the same software used to manipulate the data graphs. SHACL has recently been used to model constraints of various type in virtual institutions and organisations [20]. The constraint components supported by SHACL are of two types: *core constraint* and *SPARQL-based constraint*. Core constraint components are those supported by SHACL-compliant systems, and can be divided into the following categories [68]:

³ RDF Syntax [77] and RDF Schema (RDFS) [15]

⁴ The direct semantics of OWL are expressed in Description Logic style and we refer the interested reader to Krötzsch et.al. [48] for a self-contained introduction to Description Logic.

13:6 Autonomy in the Age of Knowledge Graphs

- Graph Structure Constraints: These are constraints evaluated over the statements in the data graph and include, for example, minimum and maximum cardinality constraints satisfied by a focus node for some given paths, type of inference with respect to a focus node, and property pair equality and disjointedness, verifying whether two sets of nodes reachable from different paths are equal or disjoint.
- Filter Constraints: These constraints are typically evaluated at node level, they are independent from the statements in the data graph and verify whether a focus node is of a particular type (e.g., IRI, literal or blank node); a given RDF term or a literal that satisfies certain conditions, e.g., a literal that is of type xsd:integer.
- Logical Constraints: These are standard logical operators (conjunction, disjunction, and negation) defined over other constraints.

SPARQL constraints support the ability to embed queries in SHACL constraints, therefore enhancing their expressivity. However, the inclusion of constraints that are potentially arbitrarily complex can come to the detriment of their usability and cause performance issues.

3 Knowledge Graphs and the Dimensions of Autonomy

Following our introduction of KGs from the perspective of autonomous agents, we next discuss in greater detail the different *types of knowledge* that support autonomous behaviour – in principle, and specifically on the Web. We structure this discussion according to the dimensions of autonomy that are introduced by Castelfranchi and Falcone [17] and concern the *environment* (see Section 3.1), other agents (see Section 3.2), organisations (see Section 3.3), and the agent's own goal-setting and execution (see Section 3.4). These are aligned with other meta-models for autonomous (multi-agent) systems, such as the JaCaMo meta-model [12] that distinguishes between the *environment*, agent, and organisation dimensions. We furthermore discuss the construction and evolution of knowledge by agents (see Section 3.5).

3.1 Knowledge supports Autonomy in Relation to the Environment

The first dimension of autonomy that we consider is the autonomy of an agent in relation to its environment. According to Castelfranchi and Falcone [17], this relates to agents being able to act in an environment based on their perception of this environment while maintaining autonomy from the perceived stimuli and from the environment itself. This is illustrated in [17] with with a billiard ball – which is not autonomous with respect to its environment – and it is argued that agents who indeed are independent from their environment instead feature teleonomic behaviour, i.e., that they actively perceive and interpret their environment and the effects of their actions, that they orient themselves towards the input, and that their behaviour depends on internal states with their own evolution principles – rather than merely receiving an environmental "force" [17] as the billiard ball does. With respect to the *environment* dimension, the autonomy of an agent in an environment is increased when this agent gains knowledge about artifacts that are present (or likely present) in that environment. This includes knowing how it could successfully carry out low-level interactions with these artifacts – that is, information about the user interface or API of the artifact; it furthermore encompasses knowledge about how to *discover* artifacts in an environment (e.g., through an artifact registry or through broadcasting), and knowledge that permits an agent to establish that a specific artifact is *relevant* to an agent's course of action i.e., that using the artifact would be beneficial for the agent to achieve its goal.

The benefits of creating agents that are more autonomous from their environment are sought after across many fields: In *software engineering*, this relates to the creation of software systems that emphasise the non-functional properties of *adaptability* and *resilience* – that is, systems that

J.-P. Calbimonte et al.

are still able to perform their core function even if their environment changes. These properties are typically – and, often, consciously – traded-off against *performance*: In a system that emphasises performance over adaptability, there is a high incentive to *tightly couple* even low-level interactions and to forgo run-time discovery capabilities by rather hard-coding device addresses and service APIs. Increasing the autonomy of a system with respect to its environment thus requires, explicitly or implicitly, to implement such systems in a way that does not couple their function to the interfaces of specific artifacts in the environment (e.g., by hard-coding a specific HTTP request); one layer above such interface coupling, the same goal motivates the creation of systems whose functions are not even coupled to specific *artifacts* – such systems hence need to be enabled to discover artifacts at run time [80], and then require an (internal or external) way to evaluate whether using these artifacts would contribute to the agents achieving their goals (e.g., a planner). The same fundamental principle is mirrored in *education*, a seemingly vastly different domain: In a similar way as it might be desirable to implement a software system that has enough knowledge to reason about the usage of specific artifacts while not being hard-coded to use them, we typically teach children in a way that increases their autonomy from *specific* environmental artifacts. To illustrate this, it is perfectly feasible to instruct a child – at the *interface* level – to use a specific means of public transport at a specific time and location to reach a specific destination – "To get to your Judo class, you need to always use bus line #2 at 5:37pm". However, parents who emphasise autonomy of their child in relation to the environment will rather opt for explaining to the child – at the knowledge level – how a public transport system is used in general, which involves knowledge about public transportation, ticketing, routes, schedules, and possibly even how to read maps or use digital tools for navigation. This provides the child with information about how to navigate their environment effectively even when the environment is dynamic, and hence increases the child's autonomy in the environment – e.g., when bus #2 is cancelled.

We argue that KGs are highly suited to support this type of autonomy of agents from their environment, i.e., with respect to the specifics of how to interact with things, by furnishing them with information about the artifacts in the environment and where this information is compatible with the information in the agents' own knowledge base – in this case, this knowledge can be readily integrated and processed by the agent. Information about how to interact with (heterogeneous) artifacts in an environment would build on top of a basic (homogeneous) interaction ability; on the Web, this ability is hypermedia-based interaction, i.e., that an agent is able to use hypermedia controls (including unparameterised hyperlinks as well as forms). The first steps towards this have already been made, and have even been standardised by the World Wide Web Consortium (W3C): the goal of the W3C Web of Things Thing Description⁵ (W3C WoT TD) is precisely to permit agents that understand this standard to use any W3C WoT TD-described device they encounter in their environment. This is accomplished by conveying details about the usage of the interface of this device at run time. This is not only academically interesting, but has convincing practical application: For example, it means that a W3C WoT TD-described sensor can be upgraded to a better model without the requirement to update clients of the sensor, as long as the two sensors' outputs are compatible with respect to the clients' common knowledge. Further, while W3C WoT TDs today remain limited, they may be generalised to what is referred to as Artifact Profiles: "structured data describing the artifact through signifiers and general (domain- and application-specific) metadata" [87]. Such profiles hence describe features of an environmental artifact beyond its interface, and can be used by agents at run time to determine whether an artifact is interesting to them in their current situation – again increasing the agents' environmental autonomy. It has also been proposed that such profiles could be brokered to agents

⁵ https://www.w3.org/TR/wot-thing-description11/

13:8 Autonomy in the Age of Knowledge Graphs

based on the current situation and abilities of the agent; this information is in this case shared in the form of an Agent Profile [54, 87]. In this scope, the SHACL vocabulary can be used to define behaviour specifications that describe to agents how to use an artifact (e.g., how to adjust a lamp's brightness through an HTTP API) [83].

3.2 Knowledge supports Autonomy in Relation to other Agents

Regarding the second dimension of autonomy – social autonomy – the autonomy of an agent in its environment is also increased when this agent has sufficient knowledge about other agents that are present in this environment, which should not depend on the hard-wiring of agents' social capabilities (either directly or through their environment). Similar to autonomy from the environment, we also consider social autonomy when training autonomous agents – this is less prevalent in software engineering, where software programs (themselves) are not yet being considered primarily as social entities, but it is highly visible in human education. In fact, this includes core tenets of education itself, including the ability to speak (multiple languages) and read, as well as neurological maturation. For example, the development of Theory of Mind (cf. [72]) in infants and, with it, the knowledge that one's own beliefs, desires, emotions, and thoughts are different from those of others, is considered crucial for successful everyday social interactions among *heterogeneous* humans.

Following Castelfranchi and Falcone [17], we may distinguish between *autonomy as independence* – the ability of an agent to reach its goals without the help of other specific agents – and *autonomy in delegation*, i.e., an agent's ability to achieve delegated tasks without detailed imperative instructions (colloquially: without *micro-management*). We argue that both of these sub-dimensions of autonomy can benefit from KGs in similar ways, as they require similar technical capabilities:

- i) self-sufficiency is a key requirement that provides the basic abilities necessary to achieve a goal or complete a delegated task, irrespective of the particular environment and the agents therein; and
- ii) in either case, generally applicable social communication abilities enable the utilisation of other agents' abilities and resources in order to complete the goal or task *given* the current specifics of the environment.

In Section 3.1, we discussed how interface specifications can avoid the hard-wiring of two specific system interfaces to each other (including the tight coupling of agents to environment specifics), and introduced that this supports autonomy – specifically, by facilitating adaptability and resilience through dynamic (re)composability. In the same way, agents may avoid tight coupling with other agents by utilising knowledge about their social environment together with generic abilities to discover new knowledge by means of interacting with other agents. Concretely, and mirroring the environment dimension, such agents would not be hard-coded to interact with specific other agents in specific ways, but would rather be equipped with more broadly applicable knowledge about other agents – this knowledge might include that other agents might have specific abilities, that the agent might consult with them or delegate tasks to them, and that the likelihood of other agents to engage in an interaction is determined through social relationships between agents. An agent that is equipped with this knowledge – much like a human who understands social relationships – is much better suited to navigate a social situation, and hence features increased social autonomy.

KGs are very well suited to implement this type of knowledge-based programming, as they can augment an agent's interactive capabilities and hence decrease its dependency on other *specific* agents. As in the case of autonomy from the environment, autonomous agents in this context may also benefit from both accessing external KGs and maintaining their own ones. Agents may then use internal KGs to maintain knowledge that is specific to the agent and knowledge that is

J.-P. Calbimonte et al.

sufficiently important for being logically grouped with a given agent, thus facilitating this agent's independence. Some of the internal knowledge may be strictly private, while other knowledge can be shared with other agents, or be entirely public. For example, the agent may maintain lookup services of its own capabilities, used for provisioning and planning, many of which can be shared publicly, which facilitates collaboration with other agents. In contrast, knowledge about other specific agents that the agent maintains, e.g., for the purpose of strategic reasoning and negotiation, might remain private to start with. It may be disclosed eventually however, if by doing so, this is useful in the context of a specific objective, e.g., during a multi-agent dialogue. Finally, privacy-relevant knowledge such as health information about a natural person will typically remain private, but may be shared with authorised agents on a need-to-know basis. When KGs are applied within an agent, they can be seen as knowledge representation tools that are used to persist an agent's belief base, or snapshots thereof, considering that it may not be feasible to capture all belief revision operations that are executed in the course of the agent's reasoning cycle. In case external KGs are provided as part of the environment, agents might use them for stigmergy, i.e., for coordination through environmental artefacts that are created or manipulated by the agents [73, 76] (see Section 3.1). This can facilitate multi-agent coordination and relax the coupling between agent instances, thus facilitating resilience and ability to respond to change.

3.3 Knowledge supports Autonomy in Relation to Organisations

Going beyond the interactions of agents where agent goals are generally aligned, *organisations* provide a deontic framework for agents to operate in by specifying joint objectives, norms, and policies. On the meta-level, they also permit agents to participate in organisational governance. The idea of virtual organisations as a means to facilitate autonomous (multi-agent) systems is a research direction that emerged at the turn of the century [18, 32, 67] and has since continued as a vibrant area of research [44]. While applications of multi-agent organisations in the sense of the holistic academic vision are nascent, most real-world organisations have by now some (often crucially important) form of electronic representation, and substantial efforts are incurred to ensure organisational compliance, e.g., with respect to privacy regulations [23].

In the context of organisational governance, KGs already exist in industry organisations for providing shared domain models that can be used across software systems and application scenarios, thus facilitating interoperability and centralising some of the knowledge maintenance efforts [40]. Beyond such current scenarios, at the organisational level, KGs are well-suited to enable agents to adjust to and switch between organisational contexts with greater independence, covering a broad range of concerns in the context of organisational autonomy. Goal-driven agents may act in a system containing several organisations, and *interact* with these organisations considering agent-internal goals. We claim that an agent that is aware of organisations (in general) and can hence better understand specific organisations it encounters at run time is able to reach higher autonomy. Such an agent would, for instance, know that another agent is deontically bound to performing a specific service, and that it may hence interact with that agent towards fulfilling a goal that requires this service. This is again similar to human interactions: Humans learn that they may trust a bus driver to share accurate information about bus routes. Crucially, they further know that this trust is modulated not by the bus driver (as an autonomous agent) but it is due to the bus driver being bound by the organisational principles of the public transport company they work for. In addition to interacting with organisations, agents might *join* organisations to then achieve an alignment of potentially conflicting knowledge or goals. That is, an agent's decision to join a specific organisation and to adopt a particular organisational role can be informed by knowledge that the agent obtains about the organisation, prior to joining.

13:10 Autonomy in the Age of Knowledge Graphs

Technically, this may be enabled through agents exchanging knowledge about organisations through KGs directly, or by making this information discoverable through environmental artifacts; the latter case also can be extended to permit (sufficiently authorised) agents to modify organisational knowledge at run time and hence allows groups of agents to evolve organisations together. Furthermore, KGs can provide agents that join an organisation with fundamental knowledge that is required for successful (inter-)action in the context of this organisation; i.e., knowledge about the available roles and objectives, and even knowledge that helps an agent decide about whether it should commit to a specific role in the organisation. To this end, KGs can provide deontic specifications of the organisation's objectives, as well as of the policies and norms that agents must and should, respectively, comply with in order to reach the organisational objectives or agent-specific goals. Since these goals may conflict with organisational objectives and with the goals of other agents, knowledge about sanctioning mechanisms may also be shared, which again leads to an increase in an agent's autonomy with respect to the organisation. Equipped with this knowledge, it can now make an informed decision about whether or not to break norms that are set by the organisation. Here, one can view the interaction of agents with the organisation (or with each other through the organisation) from a governance perspective: the organisation governs agent behaviour through norms and policies, and the organisation's agents govern the organisation, e.g., by changing organisational policies or by gradually evolving norms. Then, KGs can provide a transparent representation not only of the objectives, policies, and norms, but also about the meta-framework that governs them, thus enabling agents to affect organisational change in a structured manner. With respect to technologies, constraints on KGs – which can be represented using languages such as SHACL - can play an important role for organisations of autonomous agents. For example: constraints that are to be satisfied in the future may model organisational goals; constraints that have to be satisfied amount to policies; and constraints that should typically be satisfied may represent norms.

3.4 Knowledge supports Internal Autonomy

So far, we have discussed how knowledge can support autonomy from an *external* perspective: an agent is typically situated in an *environment* (see Section 3.1) and interacts with artifacts and *other agents* (see Section 3.2), and potentially as part of an *organisation* (see Section 3.3). Across these dimensions, we argued that a formal, explicit, knowledge-level representation of relevant entities allows the agent to reason and decide about its interactions with the rest of the system.

We now finally survey the agent's internal perspective, where autonomy relates to the agent's freedom of selecting the goals it is working towards, and of executing a course of action for achieving selected goals. Following Castelfranchi and Falcone [17], an agent has goal autonomy if: i) it is endowed with its own goals (e.g., at design time); or ii) it adopts goals received from other agents only when the adopted goals enable some of its own goals. The latter relates to the discussion in Section 3.2 about autonomy in delegation, but from an internal perspective that focuses on the receiving agent's ability to select its own goals. Goal autonomy hence is compatible with social ability: If an agent interacts with other agents and can adopt goals from them, then the agent is susceptible to influence – but the agent still preserves its goal autonomy if it can reason about the relations between candidate goals and its own goals, and if it can select specific goals to be adopted based on this reasoning. We argue that, similar to how KGs enable an agent to achieve autonomy with respect to the environment, other agents, and organisations, KGs may also help an agent in this very reasoning process, and hence help it *achieve* goal autonomy. KGs would do this by representing explicitly or implicitly relations between candidate goals and the agent's own goals; they may also represent commonsense knowledge (e.g., see [42]), thereby providing a foundation for the agent's reasoning about goal adoption.

J.-P. Calbimonte et al.

After selecting a goal, an agent will typically strive to execute a course of action towards achieving that goal. To do this, it requires *procedural knowledge*, which can take various forms: For example, procedural knowledge can be defined in terms of *plans* that specify sequences of actions and are programmed by a developer. Conversely, within a marketplace, an auctioneer may need to declare formally the dimensions along which the *auction protocols* may vary (ascending or descending, sealed-bid or open-cry, single-shot or continuous, first or second price, etc.) to allow agents to reason whether or not they should engage in the auction, as well as providing evidence to bidders of the auction outcome [6]. In the case of a reinforcement learning agent, procedural knowledge takes the form of a *policy* that is learned through repeated interactions with the environment. In other systems, from a practical/industry-oriented perspective, procedural knowledge could be represented through workflows, processes, or standard procedures [5, 74]. Agents' goal achievement autonomy – that is, their executive autonomy – is increased if agents are able to acquire procedural knowledge at run time. However, the means to represent procedural knowledge depend on the internals of the agents using the knowledge. In some cases, KGs might provide a suitable solution for representing such knowledge themselves [88]. In the more general case, KGs may provide a uniform way of describing and linking to procedural knowledge such that agents can then acquire the knowledge in the context of their operation. We hence propose that KGs may not only directly enable executive autonomy, but that they may – across many scenarios - provide a uniform interface to procedural knowledge that itself is represented in diverse formats.

3.5 Knowledge Construction and Evolution

We propose that KGs provide a solid foundation for supporting autonomy at different levels, including environmental, social, and deontic aspects as well as an agent's internal goal-selection and execution. However, this knowledge is neither static nor immutable [24]: In the case of knowledge regarding the agents' environment, this may change over the course of time, and represent variations in the conditions and stimuli accessible to the agents. Individual and communal knowledge evolves according to the changes in the internal state of the agents, as well as according to the different interactions that they may engage in. Moreover, this knowledge is not simply available for querying as a set of *read-only* resources, but it is actually subject to modifications by the very autonomous agents that use it. Using KGs to represent this knowledge caters to these requirements: KGs – and the contained entities – feature unique identifiers and can hence be easily and unambiguously addressed and, if need be, versioned, in centralised as well as decentralised settings.

The way these KGs are constructed may differ in terms of approach and/or underlying technologies [89]: Individual autonomous agents might develop their own knowledge independently of their peers, where they will in the case of KGs typically use at least a common standard ontology or controlled vocabulary, in order to allow interoperability at the time of inter-agent interactions. On the other hand, if agents are limited to updating existing knowledge with new information (e.g., in the case of new observations from a sensor), the knowledge structure does not change often, but the observations may rapidly become stale [59]. This is often the case in IoT environments, where sensors and actuators rely on knowledge that takes the form of dynamic streams of data [79]. Agents need to react in a timely fashion to these data streams, so as to update their beliefs or revise their intentions. Although certain extensions exist for allowing KGs to integrate stream reasoning [26] and continuous querying capabilities [25, 51, 9], the usage of streaming KGs by autonomous entities is still incipient. Beyond traditional pipelines of data streams, the employment of streaming KGs in multi-agent systems and in general by decentralised autonomous entities has the potential to boost the exploitation of local and personal information captured by devices on the edge and other sensing agents.

13:12 Autonomy in the Age of Knowledge Graphs

Moreover, autonomous systems need to be able to represent and reason with knowledge about external entities (the environment, other agents, organisations) but also with internal knowledge about their goals, tasks and capabilities. The former is primarily knowledge that naturally emerges from heterogeneous agents cooperating to reach some common goal. Knowledge is hence created and exchanged following agent communication and social mechanisms [16]. This knowledge may emerge from the agreements and deliberations that arise during the lifetime of autonomous entities [55]. Ontologies may then be constructed from commonly learned facts contributed by autonomous agents, thus constituting an interaction-driven KG. Given that this knowledge is built from diverse and potentially contradicting sources, different methods of reconciliation and mediation have been explored; for example by relying on ontology alignment techniques [14, 31], or distributed ontology evolution [56].

In this kind of scenario, knowledge is heterogeneous and distributed, and therefore interaction mechanisms and protocols cannot assume that agents share common knowledge beyond what is required to establish their interactions. This heterogeneity also implies that agents can use both (onto)logical or, more broadly, symbolic knowledge as well as numerical, sub-symbolic knowledge representation. The evolution of this type of knowledge cannot happen in a siloed manner, since each type of knowledge representation is essential to cover different aspects; i.e., abstracting from and making sense of stimuli from the environment. Furthermore, autonomous and cooperative decision-making implies that all the systems inhabiting the environment have to respond to the same type of demands, therefore it does not make sense for them to react to these demands in isolation, especially when they are cooperating towards a common goal.

Any knowledge evolution model should hence consider the abilities of the autonomous agents to: i) reason and evolve their own knowledge, which may be modelled using disparate formalisms; ii) reason about the different demands they perceive from the environment they are immersed in; iii) deliberate about the knowledge to evolve, given conflicting demands (and ultimately knowledge); and iv) interact with other autonomous systems to communicate and cooperatively decide the best course of action.

4 KGs and Autonomy

Section 3 identifies how knowledge supports different dimensions of autonomy, with respect to the environment, other agents, organisations, and internally, regarding executive and goal-setting autonomy. This knowledge, as discussed in Section 3.5, is dynamically constructed and evolves as a result of decision making by autonomous systems that have a common goal. The use of KGs in autonomous systems is quite disparate, and KGs are often used to support only some aspects of autonomy, e.g., knowledge about the environment in W3C WoT TD, or about the capabilities of other autonomous systems, e.g., as in the MARIO reference architecture for social robots [4].

Following from the discussion in the previous section, we argue that all of these dimensions of autonomy can benefit from the use of KGs, where the use of Semantic Web languages such as RDFS and OWL overcomes syntactic interoperability issues and makes the vocabulary to use in communicative acts (and any constraints on its use) explicit.

In the remainder of this section we identify the characteristics that KGs should exhibit in order to enable the holistic view of autonomy we discussed so far: i.e., the expressivity of the representation language used to model the KG (in terms of the primitives and the consequent reasoning services) and the coverage of vocabulary used. These should promote autonomous behaviour by supporting the mechanisms used for interaction, governance with respect to organisational autonomy, and an explicit representation of how to perform a given task.

4.1 Expressivity and Coverage of KGs Supporting Autonomy

A recent systematic review [2] surveyed a number of knowledge-enabled approaches to develop and deploy autonomous robots in order to assess whether current state of the art ontology-based approaches model those concepts and relations that are necessary to represent knowledge about other autonomous entities and the environment; i.e., the ability to model and reason about the system's own capabilities and the capabilities of others. This study argues that the key factors for the success of knowledge-based autonomous systems is the use of explicit ontological models to support autonomous systems in understanding and reasoning about their environment, their own internal knowledge and knowledge about other agents and organisations. The systematic review is based on Langley's characterisation of autonomy with respect to abstract reasoning tasks such as recognition and categorisation, decision-making and choice, problem solving and planning, as well as perception and situation assessment [50]. Langley distinguishes between: i) what, i.e., the functional capabilities that an autonomous system should exhibit; and ii) the way in which these functional modules should interconnect (how). Functional capabilities and their interconnections determine the specific vocabulary needed to represent autonomy and the types of reasoning tasks that are supported by different ontological models. These interconnections are supported by the exchange of data and knowledge that is often both produced and consumed by autonomous systems.

KG schemata modelling autonomous behaviour should include the definition of entities (in terms of modules or fragments of hierarchical knowledge) that represent both static (e.g., *agent*, *object*, or *scene*) and dynamic (e.g., *task*, *goal*, or *action*) aspects of autonomy. We follow the same characterisation and we discuss the ways KGs are currently used to support autonomy, as a way to identify the limitations (which are introduced in Section 5).

As discussed in Section 2, there is no accepted definition of KG, and as a result the term KG is used to refer to different data sources, and for example, to model both real-time data generated by sensors that does not conform to some schema, or a graph of interrelated data that is modelled according to some explicit schemata that often includes standard foundational (upper-level) and domain-specific ontologies, as well as organisation-specific schemata (which might be proprietary). Foundational ontologies such as SUMO [62] or Dolce [35] and its simplified version in OWL, Dolce+DnS Ultralite⁶ are used to model domain-independent concepts, e.g., time, object or event, and support interoperability at the highest level of abstraction. Domain-specific vocabularies have emerged in many domains to provide common ground for communication and have become standards to model data for the specific domain. Examples of such standards are the Semantic Sensor Network (SSN) ontology⁷ [37], and the provenance ontology Prov-O⁸ [53]. SSN is a modular ontology that also includes a self-contained core ontology modelling Sensors, Observations, Samples, and Actuators (SOSA). SSN and SOSA can support a variety of applications and use cases, including satellite imagery, large-scale scientific monitoring, and use cases across the Web of Things [43]. PROV-O is an OWL ontology that describes the classes, properties and restrictions describing the provenance data model, used to integrate provenance information in different systems and based on different contexts. More recently, the W3C WoT TD [43] standard provides the vocabulary for describing the metadata and interfaces for a "Thing", i.e., an abstraction of an entity (physical or virtual) that interacts with and participates in the Web of Things, thereby endowing agents with higher autonomy with respect to their environment (see Section 3.1).

⁶ http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

⁷ https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/

⁸ https://www.w3.org/TR/prov-o/

13:14 Autonomy in the Age of Knowledge Graphs

4.2 KGs Supporting Autonomous Agents' Governance

Regarding organisational autonomy (see Section 3.3), a similar issue manifests itself when representing and reasoning with obligations and norm governance in general [44]. The ODRL standard⁹ is a formal language, expressed in RDF, for modelling policies regulating the use of digital rights. However, the ODRL core that models the components of a policy expression does not provide any formal specification of its semantics, and it provides only informal semantics described in English [81]. Recently, some extensions have been proposed, based on the work in Normative MAS, where autonomy determines the decision-making processes [20]. For example, Fornara and Colombetti [33] aim to fully capture the formalisation of norms used to express obligations, permissions and prohibitions by modelling the operational semantics of conditional obligations using OWL, possibly extended with SWRL [39] rules to overcome OWL's limited expressivity with respect to the state of an interaction and the rules for computing the state of obligations [33]. SWRL adds Horn-style rules to OWL, thus permitting the modelling of additional domain knowledge that cannot be done by OWL alone. Both OWL and SWRL adopt the Open World assumption, whereby the absence of a statement does not necessarily mean that it can be inferred to be true or false. This is in contrast with SHACL, which adopts the Closed World Assumption (a statement that is not known to be true must be false). More recently, SHACL has been proposed to model more complex type of norms, such as compensatory norms [75].

4.3 KGs Supporting Agent Interaction

Several conceptual architectures and programming platforms that support autonomous behaviour are currently available (e.g., JaCaMo [12] or KnowRob [85]), and they rely on domain and foundational ontologies and KGs for representing and sharing declarative knowledge. However, many of the reasoning tasks underlying executive autonomy (see Section 3.4) involve procedural knowledge that is not typically captured by KGs. The execution of complex tasks involving several autonomous entities in open environments requires them to interact with each other and the environment, and to coordinate their executions; e.g., to support autonomy in delegation. Interaction protocols (also called *choreographies* within the service community) are formal specifications of these interactions, which describe the roles played by the autonomous systems, the allowed actions and choices, and the states and termination conditions [8]. Similarly, communication protocols specify the rules of interactions governing a dialogue between autonomous systems [29, 34, 71]. Typically, these protocols are expressed in some protocol language based on different abstractions and operational assumptions that are difficult to compare. An exhaustive description of these protocols is outside the scope of this paper, but we refer to the work by Chopra and colleagues for a review of the main formalisms and a comparison framework [21]. These protocols ground the meaning of the different concepts involved in some ontology, therefore implicitly establishing an a-priori commitment to a shared understanding of the terms being communicated [41], where the procedural part of the model and the execution traces are typically not included in KGs.

4.4 How to: KGs for Supporting Agents in Performing Tasks

Various efforts in robotics have also attempted to model and execute procedural knowledge about some tasks, e.g., *kitting*, where a robot places (a set of) parts on a tray and carries them to an assembly cell, using a knowledge-enabled approach. For example, the KnowRob [85] architecture has been extended with OWL based assembly planning, where assembled products are modelled

⁹ https://www.w3.org/TR/odrl-model/

J.-P. Calbimonte et al.

according to ontologies, which are then used to plan the next action according to faulty and missing assertions in the robot's beliefs about an ongoing assembly task [11]. Other knowledge-enabled approaches, however, convert OWL descriptions into PDDL [58] specifications [7, 47].

Multi-dimensional event knowledge graphs (EKGs) [30] have been proposed as a way to represent process events in a way that supports the execution and inspection of process traces. Recently, EKGs have also been proposed to model process logs [45] in order to support process analysis.

KGs have also been explored as a possible source of knowledge to be injected in reinforcement learning approaches [1, 3], where autonomous agents use sequential decisions to learn and adapt through interactions with their environment by trying to find (near-) optimal policies to perform an intended task [84]. In these approaches, commonsense KGs such as Wikidata [90] or CSKG, the Commons Sense Knowledge Graph [42] are exploited to model commonsense notions such as spatial, part-whole, or temporal relations, and even affordances [19]. In line with our discussion in Section 3.5, we propose that such usage of KGs represents only the beginning of a movement to the more active curation of KGs by autonomous software agents themselves.

Recently, both KGs and architectures for autonomous agents have emerged as promising facilitators of autonomy in the context of generative AI and specifically Large Language Models (LLMs) [65, 82]. In this context, KGs are typically envisioned as providers of reliable, "hard" knowledge that LLMs are not capable of producing. While this emphasises the relevance and timeliness of research on KGs as facilitators of autonomy, we deliberately choose to present perspectives in this paper that can be considered orthogonal to LLMs and generative AI in order to avoid a tunnel view that is focused on current hype topics.

5 Limitations of Knowledge Graphs for Supporting Autonomy

Whilst there are successful examples of KGs that facilitate autonomy across the different dimensions of autonomy, employing KGs as facilitators of autonomy may bring substantial challenges with it, depending on context and scale. This section describes where KGs are in principle not suitable or are not sufficient to support autonomy, as well as where they are not *yet* suitable. From the resulting list of limitations, we identify challenges that exist today, and provide initial sketches of how these challenges could be addressed. We claim that the following factors can limit the applicability of KGs as facilitators of autonomy:

- 1. Dynamics of participating systems. Autonomous software systems can be expected to appear in highly dynamic environments, and they may frequently change, often without direct human involvement (cf. Section 4.4). Consequently, KGs need to be able to accommodate and adapt to frequent changes in the knowledge they provide; otherwise, knowledge consumers risk basing their decisions and actions on stale or deprecated knowledge that no longer reflects the environment they act in. A key question in this context is to what extent KGs can maintain models of dynamic behaviour as first-class abstractions, making it easier to adjust to change and to mitigate the resulting uncertainty.
- 2. Scale of knowledge. The content of knowledge-based systems (such as KGs) is notoriously effortful to maintain. While the maintenance issues on schema-level are alleviated by (schema-free) KGs (in contrast to traditional knowledge-based systems), the challenge of maintaining the contained knowledge at scale remains. Specifically, a common challenge in addition to stale knowledge, as mentioned above is that ensuring consistency will become an increasingly tedious task with growing KGs. Currently, it is not clear which use cases are favourable to large-scale knowledge maintenance of KG content and which are not. Our assumption is that given current practices, not all potential use cases of KGs for autonomous agents are viable

13:16 Autonomy in the Age of Knowledge Graphs

at scale and that viability can depend on a number of factors, such as the economic and technical resources available for knowledge maintenance and the extent to which the knowledge is open and considered generally useful beyond a specific target use case. For example, we expect that open knowledge that is managed by long-lived communities is more scalable than domain-specific knowledge that remains in silos. On a per-case basis, the maintainability of a KG will partially depend on the selected technologies and hence on the diversity and maturity of the KG technology ecosystem (cf. Section 4.1).

- **3. Knowledge maintenance by humans and machines.** A key challenge that might, if overcome, mitigate the scalability issue mentioned above, is to have knowledge maintained jointly by humans and machines, serving the needs of both (as hinted at in Section 4.4). In particular, knowledge that machines can efficiently consume may be hard to interpret for humans, and vice versa. At the same time, duplication of knowledge should be avoided and joint knowledge maintenance can facilitate efficient knowledge maintenance, if accessibility and interpretability of both humans and machines can be guaranteed.
- 4. Heterogeneity of participating systems. A Web of autonomous agents will in practice be heterogeneous from architectural and technological perspectives; a substantial part of this heterogeneity will stem from legacy software that does not follow the idiomatic principles of conceptual meta-models (cf. Section 4.3) that will ideally emerge for future Web-based systems. These legacy software systems will not be able to utilise KGs in a comprehensive manner and their ability to be active and passive facilitators of autonomy will be limited: without tight coupling, these legacy systems will not be able to explore the Web, nor will they be straightforwardly discoverable by others. Also, interoperability issues can be expected to render governance (cf. Section 4.2) of legacy software-heavy agents a difficult task. Proposals for facilitating autonomy must account for such limitations, i.e., they must not assume that all systems relevant in a given application context are "good citizens" given idealistic architectural requirements.
- 5. Knowledge modelling and representation choices. KGs allow for a plethora of ways to model knowledge (cf. Section 4.1). However, inappropriate knowledge modelling may overcomplicate KG usage. For example, certain ontologies are too detailed and too complicated for use cases that do not benefit from the level of precision and granularity that these ontologies provide. An agent specialised for working in a domain such as healthcare may benefit from access to the corresponding ontologies; for simpler use cases, it may be better to merely expose a subset of the knowledge in a more straightforward manner, and hence to modularise the underlying knowledge [64]. Towards addressing the same challenge of real-world applicability, the KG community should consider more actively managing interfaces to systems that (for organisational or technological reasons) are not willing to adopt elements of the Semantic Web stack, and instead prefer shallow representations such as plain JSON. Such systems can still actively participate in a global knowledge organisation as envisioned in this paper, but the interfaces need to be maintained.
- **6.** Modelling of procedural knowledge. As mentioned in Section 3.4, the provision of procedural knowledge is key to enabling and augmenting autonomous behaviour. However, while several works propose and study the management of procedural knowledge with KGs (cf. Subsection 4.4), KGs do not provide first-class abstractions of procedural knowledge; and in industry applications, the management of procedural knowledge such as business process specification on the one hand and the maintenance of domain models in KGs and ontologies on the other are seen as orthogonal tasks. Integration of these types of knowledge is still an understudied challenge.

J.-P. Calbimonte et al.

Given the challenges presented above, we identify the following research directions that can help mitigate them, although each of them will require substantial further investigation.

- Hybrid representation of knowledge. The flexible nature of KGs allows for the integration of data in various ways that does not reflect the graph structure; e.g., knowledge does not necessarily need to be represented in a particular, rigid manner (think of RDF triples), but may as well be embedded as literals or be referenced through *Internationalized Resource Identifiers* (*IRIs*), as shown for example in [60]. Such flexibility mitigates the unsuitability of KGs to certain kind of information, such as procedural knowledge and constraints (Challenge 6) and allows for nuanced and pragmatic modelling decisions (Challenge 5), thus facilitating KG maintenance at scale (Challenge 2). Still, open questions remain regarding the trade-offs that need to be made when compromising on knowledge representation (from a KG perspective), and when implementing reasoning and decision-making capabilities. With these trade-offs, one can properly interpret external knowledge or knowledge represented as literals, which essentially requires lifting basic knowledge representation and reasoning tasks to the business logic layer.
- **Ontology summarisation.** Knowledge in KGs may not be particularly accessible for some agents, either because it is hard to find given the agents' limitations (Challenge 4), or because it is not modelled on the abstraction level that the agent requires (Challenge 5). Ontology summarisation approaches [86] can be applied to extract concise aggregations of knowledge that is relevant in a given context from a KG, thus making the content in a KG easier to consume for humans or machines.
- **Modularity of knowledge.** In many scenarios, agents may only be able to work with a sub-graph of a given KG, either because of partial semantic interoperability or partial agreement and trust [27]. It can therefore make sense to take a modular view on KGs, for example by modularising knowledge into run-time-loadable "knowledge packs" such as proposed in [57]. In such systems, an individual agent may decide to work only with the sub-graph it can understand and considers reliable in the broader sense, whereas a group of agents may agree on a sub-graph as a joint ground truth. Thus, modular KGs can facilitate applicability in the context of heterogeneous, partially interoperable sub-systems (Challenges 4 and 5) and account for heterogeneity of content, e.g., regarding the frequency of updates (Challenge 1) and the general state of maintenance (Challenge 3).
- (Semi-)Automating knowledge graph maintenance. To increase the feasibility of large-sale KGs (Challenge 2), their maintenance can be partially automated by utilising software agents of various degrees of autonomy and technical sophistication. Agents in the broader sense i.e., simple bots for information crawling and the propagation of simple updates already play a role in maintaining KGs today. Thus, social machines of humans and artificial agents that jointly maintain knowledge could profit from the use of organisational abstractions from multi-agent systems research, as shown in [91]. How to best advance the joint human-machine maintenance of knowledge is, however, an open question (Challenge 3), which must be carefully approached considering current hypes and trends: the emergence of LLM-based bots and the resulting threat of low-quality, technically wrong, intransparent, or otherwise undesirable "knowledge" [10] makes structured approaches to knowledge management and maintenance even more important.

13:18 Autonomy in the Age of Knowledge Graphs

6 Conclusions

In open systems, the management of knowledge is a crucial enabler of autonomous interaction among autonomous agents, including humans and software systems. We provide a conceptual vision of how KGs can be key facilitators of such autonomy. We propose that KGs enable the programming of artificial agents at the knowledge level – thus alleviating the need for low-level hard-wiring across the dimensions of autonomy – and discuss their merit with a special focus on the Web as an environment where interactions between autonomous agents take place. Based on an overview of current and fundamental limitations of KGs, we furthermore outline a set of challenges affecting the use of KGs as facilitators of autonomy; in particular, it is crucial to enable the joint human-machine maintenance of knowledge in highly dynamic and heterogeneous Web-scale systems. We then briefly discuss potential approaches, such as hybrid and modular knowledge representation, to address the challenges. Overall, this paper informs cross-community efforts to move towards reliable and resilient autonomous systems at Web scale.

— References

- Andrea Agiollo, Andrea Rafanelli, Matteo Magnini, Giovanni Ciatto, and Andrea Omicini. Symbolic knowledge injection meets intelligent agents: QoS metrics and experiments. Autonomous Agents and Multi-Agent Systems, 37(2), jun 2023. doi: 10.1007/S10458-023-09609-6.
- 2 Alberto Olivares Alarcos, Daniel Beßler, Alaa M. Khamis, P-J. Sequeira Gonçalves, Maki K. Habib, Julita Bermejo-Alonso, Marcos E. Barreto, Mohammed Diab, Jan Rosell, João Quintas, Joanna Isabelle Olszewska, Hirenkumar Chandrakant Nakawala, Edison Pignaton de Freitas, Amelie Gyrard, Stefano Borgo, G. Alenyà, Michael Beetz, and Howard Li. A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review*, 34, 2019. doi:10.1017/ S0269888919000237.
- 3 Prithviraj Ammanabrolu and Matthew J. Hausknecht. Graph constrained reinforcement learning for natural language action spaces. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020. URL: https://openreview.net/forum?id=B1x6w0EtwH.
- 4 Luigi Asprino, Paolo Ciancarini, Andrea Giovanni Nuzzolese, Valentina Presutti, and Alessandro Russo. A reference architecture for social robots. *Journal of Web Semantics*, 72:100683, 2022. doi:10.1016/J.WEBSEM.2021.100683.
- 5 Stefan Bachhofner, Elmar Kiesling, Kate Revoredo, Philipp Waibel, and Axel Polleres. Automated process knowledge graph construction from BPMN models. In Christine Strauss, Alfredo Cuzzocrea, Gabriele Kotsis, A. Min Tjoa, and Ismail Khalil, editors, *Database and Expert Systems Applications*, pages 32–47, Cham, 2022. Springer International Publishing. doi:10.1007/978-3-031-12423-5_3.
- 6 Wei Bai, Emmanuel Tadjouddine, and Terry Payne. A dialectical approach to enable decision making in online trading. In Michael Rovatsos, George Vouros, and Vicente Julian, editors, Multi-Agent Systems and Agreement Technologies, pages 203-

218. Springer International Publishing, 2016. doi: 10.1007/978-3-319-33509-4_17.

- 7 Stephen Balakirsky, Zeid Kootbally, Thomas Kramer, Anthony Pietromartire, Craig Schlenoff, and Satyandra Gupta. Knowledge driven robotics for kitting applications. *Robotics and Autonom*ous Systems, 61(11):1205–1214, 2013. Ubiquitous Robotics. doi:10.1016/J.ROBOT.2013.04.006.
- 8 Matteo Baldoni, Cristina Baroglio, Amit K. Chopra, Nirmit Desai, Viviana Patti, and Munindar P. Singh. Choice, interoperability, and conformance in interaction protocols and service choreographies. In Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2009, volume 2, pages 843–850. International Foundation for Autonomous Agents and Multiagent Systems, 2009. URL: https://dl. acm.org/doi/abs/10.5555/1558109.1558129.
- 9 Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, and Michael Grossniklaus. C-SPARQL: a continuous query language for RDF data streams. International Journal of Semantic Computing, 4(01):3-25, 2010. doi: 10.1142/S1793351X10000936.
- 10 Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, pages 610–623, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3442188.3445922.
- 11 Daniel Beßler, Mihai Pomarlan, and Michael Beetz. OWL-Enabled Assembly Planning for Robotic Agents. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, pages 1684–1692. International Foundation for Autonomous Agents and Multiagent Systems, 2018. URL: https://dl.acm.org/ doi/10.5555/3237383.3237950.
- 12 Olivier Boissier, Rafael H. Bordini, Jomi Fred Hübner, Alessandro Ricci, and Andrea Santi. Multiagent oriented programming with JaCaMo. Sci-

ence of Computer Programming, 78(6):747-761, 2013. doi:10.1016/J.SCIC0.2011.10.004.

- 13 Rafael H. Bordini, Amal El Fallah Seghrouchni, Koen Hindriks, Brian Logan, and Alessandro Ricci. Agent programming in the cognitive era. Autonomous Agents and Multi-Agent Systems, 34(2), may 2020. doi:10.1007/S10458-020-09453-Y.
- 14 Yasser Bourahla, Manuel Atencia, and Jérôme Euzenat. Knowledge improvement and diversity under interaction-driven adaptation of learned on-tologies. In AAMAS '21: Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, pages 242–250. International Foundation for Autonomous Agents and Multiagent Systems, 2021. URL: https://dl.acm.org/doi/10.5555/3463952.3463986, doi: 10.5555/3463952.3463986.
- 15 Dan Brickley and Ramanathan Guha. RDF Schema 1.1. W3C recommendation, World Wide Web Consortium, feb 2014. URL: https://www. w3.org/TR/2014/REC-rdf-schema-20140225/.
- 16 John Cantwell. A formal model of multi-agent belief-interaction. Journal of Logic, Language and Information, 15:303–329, 2006. doi:10.1007/ S10849-006-3776-3.
- 17 Cristiano Castelfranchi and Rino Falcone. From Automaticity to Autonomy: The Frontier of Artificial Agents, pages 103–136. Springer US, Boston, MA, 2003. doi:10.1007/978-1-4419-9198-0_6.
- 18 Adam Cheyer and David Martin. The Open Agent Architecture. Autonomous Agents and Multi-Agent Systems, 4(1-2):143-148, 2001. doi: 10.1023/A:1010091302035.
- 19 Prateek Chhikara, Jiarui Zhang, Filip Ilievski, Jonathan Francis, and Kaixin Ma. Knowledgeenhanced Agents for Interactive Text Games, 2023. doi:10.48550/arXiv.2305.05091.
- 20 Amit Chopra, Leendert van der Torre, Harko Verhagen, and Serena Villata. Handbook of Normative Multiagent Systems. College Publications, Rickmansworth, 2018.
- 21 Amit K. Chopra, Samuel H. Christie V, and Munindar P. Singh. An evaluation of communication protocol languages for engineering multiagent systems. *Journal of Artificial Intelligence Research*, 69:1351– 1393, dec 2020. doi:10.1613/JAIR.1.12212.
- 22 Richard Cyganiak, David Wood, and Markus Lanthaler. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, World Wide Web Consortium, February 25 2014. URL: https://www.w3. org/TR/2014/REC-rdf11-concepts-20140225/.
- 23 Marina De Vos, Sabrina Kirrane, Julian Padget, and Ken Satoh. ODRL Policy Modelling and Compliance Checking. In Paul Fodor, Marco Montali, Diego Calvanese, and Dumitru Roman, editors, *Rules and Reasoning*, pages 36–51, Cham, 2019. Springer International Publishing. doi: 10.1007/978-3-030-31095-0_3.
- 24 Emanuele Della Valle, Stefano Ceri, Frank Van Harmelen, and Dieter Fensel. It's a streaming world! reasoning upon rapidly changing information. *IEEE Intelligent Systems*, 24(6):83–89, 2009. doi:10.1109/MIS.2009.125.
- 25 Daniele Dell'Aglio, Jean-Paul Calbimonte, Emanuele Della Valle, and Oscar Corcho. Towards a Unified Language for RDF Stream Query

Processing. In The Semantic Web: ESWC 2015 Satellite Events - ESWC 2015 Satellite Events Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers, pages 353–363. Springer, 2015. doi:10.1007/978-3-319-25639-9_48.

- 26 Daniele Dell'Aglio, Emanuele Della Valle, Frank van Harmelen, and Abraham Bernstein. Stream reasoning: A survey and outlook. Data Science, 1(1-2):59-83, 2017. doi:10.3233/DS-170006.
- 27 Paul Doran, Valentina Tamma, Terry R. Payne, and Ignazio Palmisano. Dynamic selection of ontological alignments: a space reduction mechanism. In *Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09) (10/07/09 16/07/09)*, jul 2009. URL: http://ijcai.org/Proceedings/09/Papers/334.pdf.
- 28 Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. In International Conference on Semantic Systems, 2016. URL: https://ceur-ws.org/Vol-1695/paper4.pdf.
- 29 Ulrich Endriss, Nicolas Maudet, Fariba Sadri, and Francesca Toni. Aspects of protocol conformance in inter-agent dialogue. In Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pages 982–983, 2003. doi:10.1145/860575.860756.
- 30 Stefan Esser and Dirk Fahland. Multi-dimensional event data in graph databases. Journal on Data Semantics, 10:109–141, 2020. doi:10.1007/ S13740-021-00122-1.
- 31 Jérôme Euzenat. Interaction-based ontology alignment repair with expansion and relaxation. In IJCAI 2017-26th International Joint Conference on Artificial Intelligence, pages 185–191. AAAI Press, 2017. doi:10.24963/IJCAI.2017/27.
- 32 Jacques Ferber and Olivier Gutknecht. A metamodel for the analysis and design of organizations in multi-agent systems. In Yves Demazeau, editor, Proceedings of the Third International Conference on Multiagent Systems, ICMAS 1998, Paris, France, July 3-7, 1998, pages 128–135. IEEE Computer Society, 1998. doi:10.1109/ICMAS.1998. 699041.
- 33 Nicoletta Fornara and Marco Colombetti. Using semantic web technologies and production rules for reasoning on obligations, permissions, and prohibitions. Ai Communications, 32(4):319–334, 2019. doi:10.3233/AIC-190617.
- 34 Foundation for Intelligent Physical Agents. FIPA ACL Message Structure Specification. Technical report, Foundation for Intelligent Physical Agents - FIPA ACL Message Structure Specification, 2002. URL: http://www.fipa.org/specs/ fipa00061/SC00061G.pdf.
- 35 Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with DOLCE. In International conference on knowledge engineering and knowledge management, pages 166–181. Springer, 2002. doi:10.1007/3-540-45810-7_18.
- 36 Amélie Gyrard, Pankesh Patel, Amit Sheth, and Martin Serrano. Building the web of knowledge with smart IoT applications. *IEEE Intelligent Systems*, 31(5):83–88, 2016. doi:10.1109/MIS.2016.
 81.

13:20 Autonomy in the Age of Knowledge Graphs

- 37 Armin Haller, Krzysztof Janowicz, Simon J. D. Cox, Maxime Lefrançois, Kerry L. Taylor, Danh Le-Phuoc, Joshua Lieberman, Raúl García-Castro, Robert Atkinson, and Claus Stadler. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web, 10:9–32, 2018. doi:10.3233/SW-180320.
- 38 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. Number 22 in Synthesis Lectures on Data, Semantics, and Knowledge. Springer, 2021. doi:10.2200/S01125ED1V01Y202109DSK022.
- 39 Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, S. Tabet, Benjamin Grosof, and Mike Dean. SWRL: A semantic web rule language combining OWL and RuleML. Technical report, W3C Member Submission, 2004. URL: https: //www.w3.org/submissions/SWRL/.
- 40 Thomas Hubauer, Steffen Lamparter, Peter Haase, and Daniel Markus Herzig. Use cases of the industrial knowledge graph at Siemens. In Marieke van Erp, Medha Atre, Vanessa López, Kavitha Srinivas, and Carolina Fortuna, editors, Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018, volume 2180 of CEUR Workshop Proceedings. CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2180/paper-86.pdf.
- 41 Michael N. Huhns and Munindar P. Singh. Ontologies for agents. *IEEE Internet computing*, 1(6):81–83, 1997. doi:10.1109/4236.643942.
- 42 Filip Ilievski, Pedro Szekely, and Bin Zhang. CSKG: The commonsense knowledge graph. In The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18, pages 680–696. Springer, 2021. doi:10.1007/978-3-030-77385-4_41.
- 43 Sebastian Käbisch, Ege Korkan, and Michael McCool. Web of things (WoT) thing description 1.1. W3C proposed reccommendation, World Wide Web Consortium, jul 2023. URL: https://www.w3.org/TR/2023/ PR-wot-thing-description11-20230711/.
- 44 Timotheus Kampik, Adnane Mansour, Olivier Boissier, Sabrina Kirrane, Julian A. Padget, Terry R. Payne, Munindar P. Singh, Valentina A. M. Tamma, and Antoine Zimmermann. Governance of autonomous agents on the web: Challenges and opportunities. ACM Trans. Internet Techn., 22(4):104:1–104:31, 2022. doi:10.1145/ 3507910.
- 45 Shahrzad Khayatbashi, Olaf Hartig, and Amin Jalali. Transforming event knowledge graph to object-centric event logs: A comparative study for multi-dimensional process analysis. In Proceedings of the 42nd International Conference on Conceptual Modeling, pages 220–238.

Springer Nature Switzerland, 2023. doi:10.1007/978-3-031-47262-6_12.

- 46 Holger Knublauch and Dimitris Kontokostas. Shapes Constraint Language (SHACL), W3C Recommendation 20 July 2017. W3c recommendation, World Wide Web Consortium, July 20 2017. URL: https://www.w3.org/TR/ 2017/REC-shacl-20170720/.
- 47 Zeid Kootbally, Craig I. Schlenoff, Christopher R. Lawler, Thomas R. Kramer, and Satyandra K. Gupta. Towards robust assembly with knowledge representation for the planning domain definition language (PDDL). *Robotics and Computer-Integrated Manufacturing*, 33:42–55, 2015. Special Issue on Knowledge Driven Robotics and Manufacturing. doi:10.1016/j.rcim.2014.08.006.
- 48 Markus Krötzsch, Frantisek Simancik, and Ian Horrocks. A description logic primer, 2013. doi: 10.48550/arXiv.1201.4089.
- 49 Loredana Laera, Ian Blacoe, Valentina Tamma, Terry Payne, Jérôme Euzenat, and Trevor Bench-Capon. Argumentation over ontology correspondences in mas. In Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, pages 1–8, 2007. doi: 10.1145/1329125.1329400.
- 50 Pat Langley, John E. Laird, and Seth Rogers. Cognitive architectures: Research issues and challenges. Cognitive Systems Research, 10(2):141–160, 2009. doi:10.1016/J.COGSYS.2006.07.004.
- 51 Danh Le-Phuoc, Minh Dao-Tran, Josiane Xavier Parreira, and Manfred Hauswirth. A native and adaptive approach for unified processing of linked streams and linked data. In International Semantic Web Conference, pages 370–388. Springer, 2011. doi:10.1007/ 978-3-642-25073-6_24.
- 52 Danh Le-Phuoc, Hoan Nguyen Mau Quoc, Hung Ngo Quoc, Tuan Tran Nhat, and Manfred Hauswirth. The Graph of Things: A step towards the Live Knowledge Graph of connected things. Journal of Web Semantics, 37–38:25–35, 2016. doi:10.1016/J.WEBSEM.2016.02.003.
- 53 Tim Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. PROV-O: The PROV ontology: W3C recommendation 30 April 2013, apr 2013. URL: https://www.w3.org/TR/prov-o/.
- 54 Jérémy Lemée, Danai Vachtsevanou, Simon Mayer, and Andrei Ciortea. Signifiers for Affordancedriven Multi-Agent Systems. In International Workshop on Engineering Multi-Agent Systems (EMAS) at the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2022. URL: https://www.alexandria. unisg.ch/handle/20.500.14171/108730.
- 55 Michael Luck and Peter McBurney. Computing as interaction: Agent and agreement technologies. In *IEEE International Conference on Distributed Human-Machine Systems. IEEE Press*, 2008. URL: https://nms.kcl.ac.uk/michael. luck/resources/dhms08.pdf.
- 56 Alexander Maedche, Boris Motik, Ljiljana Stojanovic, Rudi Studer, and Raphael Volz. An infrastructure for searching, reusing and evolving distributed

ontologies. In Proceedings of the 12th international conference on World Wide Web, pages 439–448, 2003. doi:10.1145/775152.775215.

- 57 Simon Mayer, Jack Hodges, Dan Yu, Mareike Kritzler, and Florian Michahelles. An open semantic framework for the industrial internet of things. *IEEE Intelligent Systems*, 32(1):96–101, 2017. doi:10.1109/MIS.2017.9.
- 58 Drew McDermott, Malik Ghallab, Adele Howe, Craig Knoblock, Ashwin Ram, Manuela Veloso, Daniel Weld, and David Wilkins. PDDL – The Planning Domain Definition Language. Technical Report CVC TR98003/DCS TR1165, Yale Center for Computational Vision and Control, oct 1998. URL: https://www.cs.cmu.edu/afs/cs/ user/mmv/www/papers/98aips-PDDL.pdf.
- 59 Sascha Meckler, Harald Steinmüller, and Andreas Harth. Building a knowledge graph with inference for a production machine using the web of things standard. In Advances and Trends in Artificial Intelligence. From Theory to Practice: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26-29, 2021, Proceedings, Part II 34, pages 240-251. Springer, 2021. doi:10.1007/978-3-030-79463-7_20.
- 60 Katarina Milenkovic, Simon Mayer, Konrad Diwold, and Josef Zehetner. Enabling knowledge management in complex industrial processes using semantic web technology, 2019. URL: https://www.alexandria.unisg.ch/handle/20. 500.14171/98451.
- 61 Justin J. Miller. Graph Database Applications and Concepts with Neo4j. In Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA March 23rd-24th, 2013, pages 141-147. Association for Information Systems, 2013. URL: https://aisel.aisnet.org/ sais2013/24.
- 62 Ian Niles and Adam Pease. Towards a standard upper ontology. In Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001, pages 2–9, 2001. doi:10.1145/505168.505170.
- 63 Pablo Noriega. Agent Mediated Auctions: The Fishmarket Metaphor. PhD thesis, Universitat Autònoma de Barcelona, 1997. URL: https://www.iiia.csic.es/research/ thesis-details?pastphd_id=85.
- 64 Ignazio Palmisano, Valentina Tamma, Terry Payne, and Paul Doran. Task oriented evaluation of module extraction techniques. In Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *The Semantic Web - ISWC 2009*, pages 130–145, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. doi:10.1007/978-3-642-04930-9_9.
- 65 Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. Large language mod-

els and knowledge graphs: Opportunities and challenges, 2023. doi:10.48550/arXiv.2308.06374.

- 66 Massimo Paolucci, Takahiro Kawamura, Terry R Payne, and Katia Sycara. Semantic matching of web services capabilities. In *The Semantic* Web—ISWC 2002: First International Semantic Web Conference Sardinia, Italy, June 9–12, 2002 Proceedings 1, pages 333–347. Springer, 2002. doi:10.1007/3-540-48005-6_26.
- 67 Massimo Paolucci, Onn Shehory, Katia Sycara, Dirk Kalp, and Anandeep Pannu. A planning component for RETSINA agents. In Nicholas R. Jennings and Yves Lespérance, editors, *Intelligent Agents VI. Agent Theories, Architectures,* and Languages, pages 147–161, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. doi:10.1007/ 10719619_11.
- 68 Paolo Pareti and George Konstantinidis. A review of SHACL: From data validation to schema reasoning for RDF graphs. In *Reasoning Web. Declarative Artificial Intelligence: 17th International Summer School 2021, Leuven, Belgium, September 8–15, 2021, Tutorial Lectures, pages 115–144, Berlin, Heidelberg, 2021. Springer-Verlag. doi:10.1007/978-3-030-95481-9_6.*
- 69 Bijan Parsia, Peter Patel-Schneider, and Boris Motik. OWL 2 web ontology language structural specification and functional-style syntax (second edition). W3C recommendation, World Wide Web Consortium, dec 2012. https://www.w3.org/TR/2012/REC-owl2syntax-20121211/.
- 70 Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. Semant. Web, 8(3):489–508, jan 2017. doi: 10.3233/SW-160218.
- 71 Jeremy Pitt and Abe Mamdani. A protocolbased semantics for an agent communication language. In *IJCAI*, volume 99, pages 486– 491, 1999. URL: https://dl.acm.org/doi/abs/ 10.5555/1624218.1624288.
- 72 David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, 1(4):515-526, 1978. doi:10.1017/S0140525X00076512.
- 73 Alessandro Ricci, Andrea Omicini, Mirko Viroli, Luca Gardelli, and Enrico Oliva. Cognitive stigmergy: Towards a framework based on agents and artifacts. In Danny Weyns, H. Van Dyke Parunak, and Fabien Michel, editors, *Environments* for Multi-Agent Systems III, volume 4389, pages 124–140, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. doi:10.1007/978-3-540-71103-2_7.
- 74 Dennis M. Riehle, Sven Jannaber, Patrick Delfmann, Oliver Thomas, and Jörg Becker. Automatically annotating business process models with ontology concepts at design-time. In Sergio de Cesare and Ulrich Frank, editors, Advances in Conceptual Modeling, pages 177–186, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-70625-2_17.
- 75 Livio Robaldo, Sotirios Batsakis, Roberta Calegari, Francesco Calimeri, Megumi Fujita, Guido Governatori, Maria Concetta Morelli, Francesco Pacenza, Giuseppe Pisano, Ken Satoh, Ilias Tachmazidis, and Jessica Zangar. Compliance check-

ing on first-order knowledge with conflicting and compensatory norms: a comparison among currently available technologies. *Artificial Intelligence and Law*, jun 2023. doi:10.1007/ s10506-023-09360-z.

- 76 Daniel Schraudner. Stigmergic multi-agent systems in the semantic web of things. In Ruben Verborgh, Anastasia Dimou, Aidan Hogan, Claudia d'Amato, Ilaria Tiddi, Arne Bröring, Simon Mayer, Femke Ongenae, Riccardo Tommasini, and Mehwish Alam, editors, *The Semantic Web: ESWC 2021 Satellite Events*, pages 218–229, Cham, 2021. Springer International Publishing. doi:10.1007/978-3-030-80418-3_34.
- 77 Guus Schreiber and Fabien Gandon. RDF 1.1 XML syntax. W3C recommendation, World Wide Web Consortium, feb 2014. URL: https://www.w3.org/ TR/2014/REC-rdf-syntax-grammar-20140225/.
- 78 Paul R. Smart and Nigel R. Shadbolt. Social Machines, pages 6855–6862. IGI Global, jul 2015. doi:10.4018/978-1-4666-5888-2.ch675.
- 79 John Soldatos, Nikos Kefalakis, Manfred Hauswirth, Martin Serrano, Jean-Paul Calbimonte, Mehdi Riahi, Karl Aberer, Prem Prakash Jayaraman, Arkady Zaslavsky, Ivana Podnar Žarko, et al. Openiot: Open source internet-of-things in the cloud. In Interoperability and Open-Source Solutions for the Internet of Things: International Workshop, FP7 OpenIoT Project, Held in Conjunction with SoftCOM 2014, Split, Croatia, September 18, 2014, Invited Papers, pages 13–25. Springer, 2015. doi:10.1007/978-3-319-16546-2_3.
- 80 Sebastian Stein, Terry R. Payne, and Nicholas R. Jennings. Flexible provisioning of web service workflows. ACM Trans. Internet Technol., 9(1):1–45, feb 2009. doi:10.1145/1462159.1462161.
- 81 Simon Steyskal and Axel Polleres. Towards formal semantics for ODRL policies. In Nick Bassiliades, Georg Gottlob, Fariba Sadri, Adrian Paschke, and Dumitru Roman, editors, *Rule Technologies: Foundations, Tools, and Applications*, pages 360– 375, Cham, 2015. Springer International Publishing. doi:10.1007/978-3-319-21542-6_23.
- 82 Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents, 2023. doi: 10.48550/ARXIV.2309.02427.
- 83 Yousouf Taghzouti, Danai Vachtsevanou, Simon Mayer, and Andrei Ciortea. A step toward semantic content negotiation. In 23rd International Conference on Knowledge Engineering and

Knowledge Management (EKAW). CEUR-WS, 2022. URL: https://ceur-ws.org/Vol-3256/paper5.pdf.

- 84 Matthew E. Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. J. Mach. Learn. Res., 10:1633–1685, dec 2009. doi:10.5555/1577069.1755839.
- 85 Moritz Tenorth and Michael Beetz. KnowRob: A knowledge processing infrastructure for cognitionenabled robots. *The International Journal of Robotics Research*, 32(5):566–590, 2013. doi: 10.1177/0278364913481635.
- 86 Georgia Troullinou, Haridimos Kondylakis, Evangelia Daskalaki, and Dimitris Plexousakis. Ontology understanding without tears: The summarization approach. Semant. Web, 8(6):797–815, jan 2017. doi:10.3233/SW-170264.
- 87 Danai Vachtsevanou, Andrei Ciortea, Simon Mayer, and Jérémy Lemée. Signifiers as a first-class abstraction in hypermedia multi-agent systems. In Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS '23, pages 1200–1208, Richland, SC, 2023. International Foundation for Autonomous Agents and Multiagent Systems. doi:10.5555/3545946. 3598763.
- 88 Danai Vachtsevanou, Philip Junker, Andrei Ciortea, Iori Mizutani, and Simon Mayer. Longlived agents on the web: Continuous acquisition of behaviors in hypermedia environments. In *Companion Proceedings of the Web Conference* 2020, WWW '20, pages 185–189, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3366424.3383537.
- 89 Dylan Van Assche, Thomas Delva, Gerald Haesendonck, Pieter Heyvaert, Ben De Meester, and Anastasia Dimou. Declarative RDF graph generation from heterogeneous (semi-) structured data: A systematic literature review. Journal of Web Semantics, page 100753, 2022. doi:10.1016/J. WEBSEM.2022.100753.
- 90 Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10):78–85, 2014. doi: 10.1145/2629489.
- 91 Antonia Wild, Andrei Ciortea, and Simon Mayer. Designing social machines for tackling online disinformation. In *Companion Proceedings of the Web Conference 2020*, WWW '20, pages 650–654, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3366424.3385770.