DAGSTUHL
REPORTS

Volume 14, Issue 3, March 2024

Dagstuhl Reports, Vol. 14, Issue 3 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
https://www.dagstuhl.de/dagpub/2192-5283

Publication date
September, 2024

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`.

Digital Object Identifier: 10.4230/DagRep.14.3.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
https://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 24101

Robust Query Processing in the Cloud

Goetz Graefe*[1], Allison Lee*[2], and Caetano Sauer*[3]

1 **Google – Madison, US.** goetz.graefe@gmail.com
2 **Snowflake – San Mateo, US.** allison@snowflake.com
3 **Salesforce – München, DE.** caetano.sauer@salesforce.com

Abstract

The Dagstuhl Seminar on "Robust Query Processing in the Cloud" (24101), held from March 3 to March 8, 2024, brought together researchers from academia and industry to discuss robustness in database management systems. This seminar was a continuation of previous seminars on the topic of Robust Query Processing, where we focused in particular on cloud computing and also discussed aspects that have not been addressed by the previous instances of the seminar. This article summarizes the main discussion topics, and presents the summary of the outputs of five working groups that discussed: i) robustness benchmarking, ii) economics of query processing in the cloud, iii) storage architectures, iv) out-of-memory query operators, and v) indexing for data warehousing.

Seminar March 3–8, 2024 – https://www.dagstuhl.de/24101
2012 ACM Subject Classification Information systems → Data management systems; Information systems → Storage architectures
Keywords and phrases database, execution, hardware, optimization, performance, query
Digital Object Identifier 10.4230/DagRep.14.3.1

1 Executive Summary

Goetz Graefe (Google – Madison, US)
Allison Lee (Snowflake – San Mateo, US)
Caetano Sauer (Salesforce – München, DE)

The Dagstuhl Seminar on "Robust Query Processing in the Cloud" (24101) assembled researchers from industry and academia for the fifth time to discuss robustness issues in database query performance, this time with a focus on Cloud Computing. The seminar gathered researchers around the world working on indexing, storage, plan generation and plan execution in database query processing, and in cloud-based massively parallel systems with the purpose to address the open research challenges with respect to the robustness of database management systems. Delivering robust query performance is well known to be a difficult problem for database management systems. All experienced DBAs and database users are familiar with sudden disruptions in data centers due to poor performance of queries that have performed perfectly well in the past. The goal of the seminar was to discuss the current state-of-the-art, to identify specific research opportunities in order to improve the state-of-affairs in query processing, and to develop new approaches or even solutions for these opportunities, building upon successes of the past Dagstuhl Seminars [2, 3, 5, 1, 4, 6]. The organizers (Goetz Graefe, Allison Lee, and Caetano Sauer) this time attempted to have a

[*] Editor / Organizer

focused subset of topics that the participants discussed and analyzed in more depth. From the proposed topics on algorithm choices, join sequences, storage architectures, database utilities, modern storage hardware, cloud database economics, and benchmarking for robust query processing, the participants formed five work groups: i) robustness benchmarking, ii) economics of query processing in the cloud, iii) storage architectures, iv) out-of-memory query operators, and v) indexing for data warehousing. Upon choosing the topics of interest, the organizers then guided the participants to approach the topic through a set of steps: by first considering related work in the area; then introducing metrics and tests that will be used for testing the validity and robustness of the solution; after metrics, the focus was on proposing specific mechanisms for the proposed approaches; and finally the last step focused on the implementation policies. At the end of the week, each group presented their progress with the hope to continue their work towards a research publication. The reports of work groups are presented next.

References

1 Peter A. Boncz, Yannis Chronis, Jan Finis, Stefan Halfpap, Viktor Leis, Thomas Neumann, Anisoara Nica, Caetano Sauer, Knut Stolze, and Marcin Zukowski. SPA: economical and workload-driven indexing for data analytics in the cloud. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*, pages 3740–3746. IEEE, 2023.

2 Renata Borovica-Gajic, Stratos Idreos, Anastasia Ailamaki, Marcin Zukowski, and Campbell Fraser. Smooth scan: Statistics-oblivious access paths. In Johannes Gehrke, Wolfgang Lehner, Kyuseok Shim, Sang Kyun Cha, and Guy M. Lohman, editors, *ICDE*, pages 315–326. IEEE Computer Society, 2015.

3 Renata Borovica-Gajic, Stratos Idreos, Anastasia Ailamaki, Marcin Zukowski, and Campbell Fraser. Smooth scan: robust access path selection without cardinality estimation. *VLDB J.*, 27(4):521–545, 2018.

4 David Justen, Daniel Ritter, Campbell Fraser, Andrew Lamb, Nga Tran, Allison Lee, Thomas Bodner, Mhd Yamen Haddad, Steffen Zeuch, Volker Markl, and Matthias Boehm. POLAR: adaptive and non-invasive join order selection via plans of least resistance. *Proc. VLDB Endow.*, 17(6):1350–1363, 2024.

5 Martin L. Kersten, Alfons Kemper, Volker Markl, Anisoara Nica, Meikel Poess, and Kai-Uwe Sattler. Tractor pulling on data warehouses. In Goetz Graefe and Kenneth Salem, editors, *DBTest*, page 7. ACM, 2011.

6 Lukas Vogel, Daniel Ritter, Danica Porobic, Pinar Tözün, Tianzheng Wang, and Alberto Lerner. Data pipes: Declarative control over data movement. In *13th Conference on Innovative Data Systems Research, CIDR 2023, Amsterdam, The Netherlands, January 8-11, 2023*. www.cidrdb.org, 2023.

2 Table of Contents

3 Working groups

3.1 Benchmarking Robustness Using Perturbed Workloads

Manos Athanassoulis (Boston University, US), Carsten Binnig (TU Darmstadt, DE), Yannis Chronis (Google – Sunnyvale, US), John Cieslewicz (Google – Mountain View, US), Sudipto Das (Amazon Web Services – Seattle, US), Stefan Halfpap (Technische Universität Berlin, DE), Allison Lee (Snowflake – San Mateo, US), Bernhard Seeger (Universität Marburg, DE), and Nga Tran (InfluxData – Boston, US)

The robustness of database systems is an important and extensively discussed topic in academia and industry. While various new approaches have been proposed to improve the robustness of database systems – including new robust query operators such as smooth scan – surprisingly, there is still no method that allows us to holistically analyze and quantify the robustness of a database system. In this paper, we define the very first method that allows us to quantify the robustness of a database system. Intuitively, our method analyzes whether small changes in the execution of a workload lead to only small observed changes in the performance of the database system. Building on this robustness definition, we propose a new benchmark framework that tests the robustness of a database system. This framework can take any existing workload as input (e.g., TPC-H or TPC-DS) and by injecting small perturbations targeting, for instance, query and data characteristics, quantify the robustness of a system. Importantly, along with the benchmark, we propose a new robustness metric that allows us to quantify robustness at the system level and enable comparison across systems or versions of the same system.

3.2 From ASAP to JUST Pricing: The Economics of Robust Query Performance in the Cloud

Thomas Bodner (Hasso-Plattner-Institut, Universität Potsdam, DE), Peter A. Boncz (CWI – Amsterdam, NL), Goetz Graefe (Google – Madison, US), Viktor Leis (TU München – Garching, DE), Danica Porobic (Oracle Switzerland – Zürich, CH), and Caetano Sauer (Salesforce – München, DE)

We propose the concept of performance SLAs (Service Level Agreements) for SQL workloads, as an enabler for new pricing models in cloud databases. We identify two new research areas necessary for making this possible: (i) reliable methods to determine SLA pricing and associated financial risks and profits, (ii) technical innovations in cloud database engines that take into account query workload deadlines, and employ elastic resource allocation to make these deadlines. In all, we think that this new model, on the one hand finally offers customers a real handle on workload SLAs, while on the other hand enables both customer and provider cost saving as well as reduced hardware resource usage (and thus reduced carbon footprint).

3.3 The Query Exchange: Auctioning and Bidding on Query Workloads for a Competitive Cloud Database Market

Thomas Bodner (Hasso-Plattner-Institut, Universität Potsdam, DE), Peter A. Boncz (CWI – Amsterdam, NL), Goetz Graefe (Google – Madison, US), Viktor Leis (TU München – Garching, DE), Danica Porobic (Oracle Switzerland – Zürich, CH), and Caetano Sauer (Salesforce – München, DE)

In today's cloud analytics market, customers choose providers to serve their query workloads for an extended period of time. In contrast, an efficient choice per query (or per workload) would enable and encourage innovation for efficient database systems. Building on shared storage in the public cloud, we propose the concept of a query exchange. A query exchange implements a competitive market, brokering queries to the provider with the lowest bid for execution satisfying the customer requirements. Using open table formats and appropriate access governance, alternative providers can read and process the same data sources. Relying on specifications of data, metadata and workload SLAs, the query exchange aligns customers' incentive to lower costs with providers' incentive to process queries for which they have a competitive advantage, thus achieving efficient, scalable, and robust query performance in the cloud.

3.4 SAUNA – Saved Aggregates Unleash Novel Acceleration

Nicolas Bruno (Microsoft – Redmond, US), Campbell Fraser (Google – Mountain View, US), Kyoungmin Kim (EPFL – Lausanne, CH), Anisoara Nica (SAP SE – Waterloo, CA), Immanuel Trummer (Cornell University – Ithaca, US), and Juliane Waack (Snowflake – Berlin, DE)

The Dagstuhl Seminar "Database Indexing and Query Processing" in March 2022 produced the SPA paper [1], which presented a general framework for improving the performance of scanning modern column stores. SPA introduced several interesting building blocks, such as incremental building of intermediate structures and using economic principles to guide the construction (and deconstruction) of such structures. We explored several methods inspired by that framework to improve the performance of query processing in a robust way. More precisely, we explored the following ideas, in a new framefork SAUNA – Saved Aggregates Unleash Novel Acceleration.

Single table aggregate result caching. We can see the original block pruning technique as creating index structures that can answer boolean aggregate queries about whether any tuple exists in a block. We generalize the approach to return the actual results from evaluating single table aggregates with filters, resulting in an incremental/partial materialized view algorithm that follows the workload.

Using single table aggregate results to prune blocks . When single table aggregates return MIN/MAX values, these values can be used to generate runtime constraints over each block, and conceptually reason whether the original filter plus the constraint results in a contradiction, which results in novel ways to do block-level pruning.

Block pruning for fact-dimension table joins where dimension tables change slowly over time. We extend the index structures on the blocks of a fact table to include precomputed information about joins over small dimension tables that change rarely. In that way, we complement bloom filters with more precise information that can skip blocks that are guaranteed not to join with any row in the dimension tables. When executing a query plan, we can keep track of block ids that remain after each operator, detect the join operators that are selective and prune most of the blocks compared to sub-operators, and adaptively index the patterns to our index (join patterns as keys and lists of block ids as values). This idea can be easily extended to cover arbitrary query plans and operators including aggregations.

Reinforcement learning with partial configurations . State-of-the-art methods for automated database tuning rely on reinforcement learning. To find optimal solutions, such approaches need to try out various tuning options. This can lead to non-robust performance from the user's perspective. E.g., to find optimal index configurations, current reinforcement learning methods must create and drop indexes. This can lead to significant performance variations, even when running the same query repeatedly. Considering partial configurations, e.g., indexes that cover only data subsets, can help to improve robustness in such scenarios. On the other hand, benchmarking configurations that differ only in their configuration for small data subsets may make it harder to reliably identify optima. To explore those tradeoffs, a proof-of-concept prototype was created, using reinforcement learning to find optimal index configurations in a search space of partial configurations. First experimental results show that this approach still finds optimal indexing solutions while improving performance robustness very significantly.

3.5 Coping with Out-of-Memory Situations in Distributed Join Processing

Matthias Böhm (TU Berlin, DE), Periklis Chrysogelos (Oracle Switzerland – Zürich, CH), Thanh Do (Celonis Inc. – New York, US), Jan Finis (Salesforce – München, DE), Alfons Kemper (TU München – Garching, DE), Thomas Neumann (TU München – Garching, DE), Knut Stolze (Ocient – Jena, DE), and Marcin Zukowski (Snowflake – San Mateo, US)

A major challenge in distributed query processing in the cloud is the handling of mispredicted input cardinalities and skew. In order to mitigate this problem, our group set out to devise graceful strategies for unexpected out-of-memory situations in distributed joins, group-bys, and window aggregates. After detailed discussions of existing system architectures and operational challenges, we defined a framework for detecting and mitigating such out-of-memory situations. At its core, we create virtual partitions, and optimize configurations including the mapping of virtual partitions to nodes, build-side broadcasting, probe-side materialization, as well as the number of utilized nodes. Furthermore, our group explored a broad list and classification of other unexpected performance problems of user-facing issues (e.g., LIKE predicates, UDFs, and APIs), query compilation issues (e.g., generated huge queries and expression trees), as well as runtime challenges (e.g., parallelism for pipelines with small inputs, loads, tail latency, resource isolation, large strings).

3.6 The 5 minute rule for the cloud: When are caches needed?

Andrew Lamb (InfluxData – Boston, US), Angelos Christos Anadiotis (Oracle Switzerland – Zürich, CH), Kira Isabel Duwe (EPFL – Lausanne, CH), Lucas Lersch (Amazon Web Services – East Palo Alto, US), Boaz Leskes (MotherDuck – Amsterdam, NL), Daniel Ritter (SAP SE – Walldorf, DE), Kai-Uwe Sattler (TU Ilmenau, DE), and Pinar Tözün (IT University of Copenhagen, DK)

Many modern cloud database systems use disaggregated architectures, separating the computations and the underlying object storages (e.g S3). Traditionally, database systems have used caches to improve performance to improve the data on local SSDs, essentially following the so called 5-minute rule of thumb to decide when to cache in-memory or read directly from local storage. As data moves to object stores, which have highly unpredictable tail latency and explicit costs per access, the question naturally arises whether these new disaggregated architectures need comparable caches. In practice, many systems do use caches between object storage and compute, however caches can introduce new overall system robustness challenges (e.g. cache misses) as well as add non-trivial expense both in terms of engineering and operational overhead. In this paper, we review the requirements that lead to object store caching layers, analyze the design space, and propose new rules of thumb to help system designers determine under what circumstances they should introduce caches instead of reading directly from cloud object storage.

Participants

- Angelos Christos Anadiotis
Oracle Switzerland – Zürich, CH
- Manos Athanassoulis
Boston University, US
- Carsten Binnig
TU Darmstadt, DE
- Thomas Bodner
Hasso-Plattner-Institut, Universität Potsdam, DE
- Matthias Böhm
TU Berlin, DE
- Peter A. Boncz
CWI – Amsterdam, NL
- Nicolas Bruno
Microsoft – Redmond, US
- Yannis Chronis
Google – Sunnyvale, US
- Periklis Chrysogelos
Oracle Switzerland – Zürich, CH
- John Cieslewicz
Google – Mountain View, US
- Sudipto Das
Amazon Web Services – Seattle, US
- Thanh Do
Celonis Inc. – New York, US
- Kira Isabel Duwe
EPFL – Lausanne, CH
- Jan Finis
Salesforce – München, DE
- Campbell Fraser
Google – Mountain View, US
- Goetz Graefe
Google – Madison, US
- Stefan Halfpap
Technische Universität Berlin, DE
- Alfons Kemper
TU München – Garching, DE
- Kyoungmin Kim
EPFL – Lausanne, CH
- Andrew Lamb
InfluxData – Boston, US
- Allison Lee
Snowflake – San Mateo, US
- Viktor Leis
TU München – Garching, DE
- Lucas Lersch
Amazon Web Services – East Palo Alto, US
- Boaz Leskes
MotherDuck – Amsterdam, NL
- Thomas Neumann
TU München – Garching, DE
- Anisoara Nica
SAP SE – Waterloo, CA
- Danica Porobic
Oracle Switzerland – Zürich, CH
- Daniel Ritter
SAP SE – Walldorf, DE
- Kai-Uwe Sattler
TU Ilmenau, DE
- Caetano Sauer
Salesforce – München, DE
- Bernhard Seeger
Universität Marburg, DE
- Knut Stolze
Ocient – Jena, DE
- Pinar Tözün
IT University of Copenhagen, DK
- Nga Tran
InfluxData – Boston, US
- Immanuel Trummer
Cornell University – Ithaca, US
- Juliane Waack
Snowflake – Berlin, DE
- Marcin Zukowski
Snowflake – San Mateo, US

Report from Dagstuhl Seminar 24102

Shapes in Graph Data: Theory and Implementation

Shqiponja Ahmetaj[*1], **Slawomir Staworko**[*2], **Jan Van den Bussche**[*3], **and Maxime Jakubowski**[†4]

1 **TU Wien, AT.** shqiponja.ahmetaj@tuwien.ac.at
2 **relationalAI – Berkeley, US.** slawek.staworko@relational.ai
3 **Hasselt University, BE.** jan.vandenbussche@uhasselt.be
4 **Hasselt University, BE.** maxime.jakubowski@uhasselt.be

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar "Shapes in Graph Data: Theory and Implementation" (24102). The seminar brought together active expert and junior researchers, both from academia and industry, to discuss the many open problems and research directions that arise from shapes in graph data, and, more generally, flexible and expressive schema and constraint languages for graph databases. The participants informed each other on how we perceive the research area, reported on the most recent results, discussed open problems and future directions, and in particular, four working groups were formed with promising intentions to work on new research and vision papers.

Seminar March 3–8, 2024 – https://www.dagstuhl.de/24102
2012 ACM Subject Classification Theory of computation → Data modeling; Information systems → Integrity checking; Theory of computation → Logic and databases; Information systems → Semantic web description languages
Keywords and phrases constraint languages, data for the semantic web, graph data, schema languages
Digital Object Identifier 10.4230/DagRep.14.3.9

1 Executive Summary

Shqiponja Ahmetaj (TU Wien, Austria, shqiponja.ahmetaj@tuwien.ac.at)
Slawomir Staworko (relationalAI – Berkeley, US, slawek.staworko@relational.ai)
Jan Van den Bussche (Hasselt University, BE, jan.vandenbussche@uhasselt.be)

Research Area and Goals of the Seminar

One of the main reasons for the success of graph databases is that they do not require an elaborate database schema, with accompanying integrity constraints, to be set up in advance. In these classical applications, constraints and schemas are mainly *descriptive*, having as purpose to support the mental map from the real world to the data to be managed in the database. However, the emergence of graph databases is accompanied by a paradigm shift towards new applications where schemas and constraints are used for a *prescriptive* purpose. Here, the goal is to establish a contract between the database and its users, which provides guarantees on the structure and form of data provided. This shift has led to the development of a new class of formalisms based on the notion of *shapes*. Shapes are constraints on nodes

[*] Editor / Organizer
[†] Editorial Assistant / Collector

in the graph that impose or forbid structural patterns (involving paths, edges, labels, and constant values). Naturally, then, a novel, prescriptive notion of schema emerges, consisting of a set of shapes, together with a targeting mechanism that specifies which nodes should satisfy which shapes. In the world of RDF graphs, two main shape-based formalisms have been proposed: *SHACL* (Shapes Constraint Language), standardized by the W3C, and *ShEx* (Shape Expression schemas). In the world of property graphs (PGs), different systems have their own data definition languages, such as Cypher or GSQL. Moreover, there are recent formal approaches to define schemas for property graphs such as PG-Schema and PG-Keys. The main aim of the Dagstuhl Seminar was to bring together active researchers, both from academia and industry, to report on the most recent results, to discuss the many open problems and research directions that arise from shapes, constraints, and schemas for graph databases, and to initiate new research.

Organization and Outcomes

The organisers created a schedule based on the entries from a Google document set up before the seminar, inviting participants to add talks, demos, and research topics. The seminar began with a round of introductions, where participants also asked questions they wanted to be answered during the seminar. The final schedule included 18 contributed talks and 6 short presentations on potential research and discussion topics.

As a major result from the seminar, four working groups were formed on the topics:

1. *What is used in practice for graph data abstractions? What is needed in practice for graph data abstractions?* The group formation was inspired by related questions posed by many participants during the opening introductory round on the first day of the seminar. Several research challenges were discussed and addressing them will call for opening new human-centered research lines in the data management community and beyond.
2. *Repairs and explanations in knowledge graph data management systems in the presence of shape constraints.* The group discussed the problem of assessing and managing data quality in knowledge graphs (KGs). This is a long-standing issue that attracts significant attention both in industry and academia. The new proposals on schemas and shape languages for KGs have introduced new challenges, which involve new methods to verify their validity, to deal with inconsistency, and repair the inconsistent data.
3. *Relating 6NF (Sixth Normal Form) and PG-Schema.* In this working group, two main questions were discussed: (1) Can we show in a systematic manner how schemas for property graphs, as expressed in the proposals of PG-Schema and PG-Keys, can be represented relationally, obtaining highly decomposed (6NF) schemas with key constraints and inclusion constraints such as foreign keys? (2) Can the intent of a graph database application be formalized in a suitable variant of EER (extended Entity-Relationship) diagrams?
4. *Convergence of graph data models and schemas.* The goal of the group was to understand the commonalities and differences between RDF and LPG (labelled property graphs), and their corresponding schema languages, ShEx and SHACL for RDF, and PG-Schema for LPG. The aim is to identify a common core (a small but useful common sublanguage, easily expressible in all three formalisms) and a common superlanguage (a language that captures all three formalisms, yet remains manageable).

The organisers regard the seminar as a very successful scientific event. Members of each working group expressed a clear commitment to staying connected to further investigate these topics. The first two groups specify a vision paper as a specific goal and the result of the group's future efforts and the second two groups aim to produce research papers.

The organisers are grateful to the Scientific Directorate and to the staff for supporting in making this seminar possible.

2 Table of Contents

Executive Summary

Overview of Talks

3 Overview of Talks

3.1 Explanations and Repairs for Non-Validation in SHACL

Shqiponja Ahmetaj (TU Wien, AT)

Joint work of Shqiponja Ahmetaj, Robert David, Magdalena Ortiz, Axel Polleres, Bojken Shehu, Mantas Simkus
Main reference Shqiponja Ahmetaj, Robert David, Magdalena Ortiz, Axel Polleres, Bojken Shehu, Mantas Simkus: "Reasoning about Explanations for Non-validation in SHACL", in Proc. of the 18th International Conference on Principles of Knowledge Representation and Reasoning, KR 2021, Online event, November 3-12, 2021, pp. 12–21, 2021.
URL https://doi.org/10.24963/KR.2021/2

The Shapes Constraint Language (SHACL) is a W3C standardized language for describing and validating constraints over RDF graphs. The SHACL specification describes the so-called validation reports, which are meant to explain to the users the outcome of validating an RDF graph against a collection of shape constraints. Specifically, explaining the reasons why the input graph does not satisfy the constraints is challenging. Inspired by works on logic-based abduction and database repairs, we study in [1] the problem of explaining non-validation of SHACL constraints. In particular, in our framework non-validation is explained using the notion of a repair, i.e., a collection of additions and deletions whose application on an input graph results in a repaired graph that does satisfy the given SHACL constraints. We define a collection of decision problems for reasoning about explanations, possibly restricting to explanations that are minimal with respect to cardinality or set inclusion. We provide a detailed characterization of the computational complexity of those reasoning tasks, including the combined and the data complexity. We then propose in [2] an algorithm to compute repairs for non-recursive SHACL, the largest fragment of SHACL that is fully defined in the specification. More precisely, we encode the explanation problem – using Answer Set Programming (ASP) – into a logic program, the answer sets of which correspond to repairs. We then study a scenario where it is not possible to simultaneously repair all the targets, which may be often the case due to overall unsatisfiability or conflicting constraints. We introduce a relaxed notion of validation, which allows to validate a (maximal) subset of the targets and adapt the ASP translation to take into account this relaxation. Our implementation in Clingo is – to the best of our knowledge – the first implementation of a repair generator for SHACL.

References

1 Shqiponja Ahmetaj, Robert David, Magdalena Ortiz, Axel Polleres, Bojken Shehu, and Mantas Šimkus, *Reasoning about Explanations for Non-validation in SHACL*, in *18th International Conference on Principles of Knowledge Representation and Reasoning*, pp. 12–21, 2021, doi: 10.24963/KR.2021/2.

2 Shqiponja Ahmetaj, Robert David, Axel Polleres, and Mantas Šimkus, *Repairing SHACL Constraint Violations Using Answer Set Programming*, in *21st International Semantic Web Conference*, pp. 375–391, Springer, 2022, doi: 10.1007/978-3-031-19433-7_22.

3.2 SHACL shapes extraction

Anastasia Dimou (KU Leuven, BE)

Joint work of Anastasia Dimou, Xuemin Duan, David Chaves-Fraga
Main reference Xuemin Duan, David Chaves-Fraga, Olivier Derom, Anastasia Dimou: "SCOOP All the Constraints' Flavours for Your Knowledge Graph", in Proc. of the The Semantic Web – 21st International Conference, ESWC 2024, Hersonissos, Crete, Greece, May 26-30, 2024, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 14665, pp. 217–234, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-60635-9_13

Defining shapes for the validation of RDF graphs is a non-trivial endeavour. While in most cases the shapes are manually defined, various methods were proposed for the extraction of shapes. In this talk, we went through the methods for extracting shapes and discussed open challenges related to the integration of shapes extracted from different sources.

Shapes are typically mined from RDF graphs [1, 2, 3, 4], and thus, their effectiveness is inherently influenced by the size and complexity of the RDF graph. However, these systems often overlook the constraints imposed by individual artifacts which contributed to the construction of RDF graphs.

RDF graphs are often constructed by applying ontology terms to heterogeneous data according to a set of mapping rules. Methods were proposed to extract SHACL shapes from the data schema [6, 7], the ontology [5] or the mapping rules [8]. However, these approaches lead to limited or incomplete constraints.

Methods were also proposed that exploit all artifacts associated with the construction of RDF graphs. SCOOP extract shapes from data schemas, ontologies, and mapping rules, and integrates the shapes extracted from each artifact into a unified shapes graph. SCOOP's implementation was configured to extract shapes from XML Schema [9] using XSD2SHACL [6], OWL Ontologies [10] using Astrea [5], and RML mapping rules [11] using RML2SHACL [8].

SCOOP was applied to real-world use cases and experimental results. So far methods that exploit all artifacts associated with the construction of RDF outperform methods that extract shapes from RDF graphs. However, the integration of shapes often leads to inconsistences. Such inconsistences were discussed during the talk as well as strategies to deal with them.

References

1 Kashif Rabbani, Matteo Lissandrini, Katja Hose (2023) *Extraction of Validating Shapes from Very Large Knowledge Graphs*, VLDB Endowment, doi: 10.14778/3579075.3579078.

2 Daniel Fernandez-Álvarez, Jose Emilio Labra-Gayo, and Daniel Gayo-Avello, *Automatic extraction of shapes using sheXer*, Knowledge-Based Systems, vol. 238, 2022. doi: 10.1016/j.knosys.2021.107975.

3 Blerina Spahiu, Andrea Maurino, and Matteo Palmonari,*Towards Improving the Quality of Knowledge Graphs with Data-driven Ontology Patterns and SHACL*, in *Studies on the Semantic Web*, vol. 36: *Emerging Topics in Semantic Technologies*, pp. 52–66, 2018, IOS Press, doi: 10.3233/978-1-61499-894-5-103.

4 Nandana Mihindukulasooriya, Mohammad Rifat Ahmmad Rashid, Giuseppe Rizzo, Raul Garcia-Castro, Oscar Corcho, and Marco Torchiano, *RDF Shape Induction using Knowledge Base Profiling*, in *Proceedings of the 33rd ACM/SIGAPP Symposium On Applied Computing*, 2017, doi: 10.1145/3167132.3167341.

5 Andrea Cimmino, Alba Fernández-Izquierdo, and Raúl García-Castro, *Astrea: Automatic Generation of SHACL Shapes from Ontologies*, in *European Semantic Web Conference*, pp. 497–513, Springer, 2020, doi: 10.1007/978-3-030-49461-2_29.

6 Xuemin Duan, David Chaves-Fraga, and Anastasia Dimou, *XSD2SHACL: Capturing RDF Constraints from XML Schema*, in *Proceedings of the 12th Knowledge Capture Conference 2023*, pp. 214–222, Association for Computing Machinery, 2023, doi: 10.1145/3587259.3627565.

7 Herminio Garcia-Gonzalez and Jose Emilio Labra-Gayo, *XMLSchema2ShEx: Converting XML validation to RDF validation*, *Semantic Web*, vol. 11, no. 2, pp. 235–253, 2020, publisher: IOS Press.

8 Thomas Delva, Birte De Smedt, Sitt Min Oo, Dylan Van Assche, Sven Lieber, and Anastasia Dimou, *RML2SHACL: RDF Generation Taking Shape*, in *Proceedings of the 11^{th} on Knowledge Capture Conference*, pp. 153–160, ACM, 2021, doi: 10.1145/3460210.3493562.

9 David Fallside, & Priscilla Walmsley, XML Schema Part 0: Primer Second Edition. (W3C,2004,10), https://www.w3.org/TR/xmlschema-0/

10 Conrad Bock, Achille Fokoue, Peter Haase, Rinke Hoekstra, Ian Horrocks, Alan Ruttenberg, Uli Sattler, & Michael Smith. *OWL 2 Web Ontology Language – Structural Specification and Functional-Style Syntax (Second Edition).* (World Wide Web Consortium (W3C),2012, http://www.w3.org/TR/owl2-syntax/

11 Ana Iglesias-Molina, Dylan Van Assche, Julian Arenas-Guerrero, Ben De Meester, Christophe Debruyne, Sam Jozashoori, Maria Poveda, Michel Frank, David Chaves-Fraga, & Anastasia Dimou. *The RML Ontology: A Community-Driven Modular Redesign After a Decade of Experience in Mapping Heterogeneous Data to RDF.* The Semantic Web – ISWC 2023. pp. 152-175, 2023.

12 Xuemin Duan, David Chaves-Fraga, Oliver Derom, & Anastasia Dimou. *SCOOP all the Constraints' Flavours for your Knowledge Graph.* The Semantic Web – ESWC2024, 2024.

3.3 Schema Discovery for Property Graphs

Stefania Dumbrava (ENSIIE – Paris, FR) and Angela Bonifati (Université Claude Bernard – Lyon, FR & IUF – Paris, FR)

Joint work of Angela Bonifati, Stefania Dumbrava, Emile Martinez, Nicolas Mir, Hanâ Lbath, Fatemeh Ghasemi, Malo Jaffré, Pacome Luton, Thomas Pickles

Main reference Angela Bonifati, Stefania Dumbrava, Nicolas Mir: "Hierarchical Clustering for Property Graph Schema Discovery", in Proc. of the 25th International Conference on Extending Database Technology, EDBT 2022, Edinburgh, UK, March 29 – April 1, 2022, pp. 2:449–2:453, OpenProceedings.org, 2022.

URL https://doi.org/10.48786/EDBT.2022.39

Property graphs are becoming pervasive in various graph processing applications using interconnected data. They allow encoding multi-labeled nodes and edges, as well as their properties, represented as key/value pairs. Although property graphs are widely used in several open-source and commercial graph databases, their schema definition is not as well-understood as that of their relational counterparts. The property graph schema discovery problem consists of extracting the underlying schema concepts and types from such graph datasets. The talk provides an overview of two recent schema discovery methods for property graphs.

The first method, MRSchema [1], builds upon Cypher queries to extract the node and edge serialization of a property graph, then leverages a MapReduce type inference system to obtain subtype and supertype information for nodes, and analyzes these to compute node hierarchies. The second method, GMMSchema [2], relies on hierarchical clustering using a Gaussian Mixture Model, which accounts for both node labels and properties, unlike

MRSchema. This allows for preventing the discovery of spurious types and achieving efficient performance without accuracy loss. Moreover, the approach supports efficient incremental schema maintenance, as showcased in the corresponding DiscoPG tool [3].

DiscoPG allows users to perform schema discovery for both static and dynamic graph datasets. Suitable visualization layouts and dedicated dashboards enable the user perception of the static and dynamic inferred schema on the node clusters, as well as the differences in runtimes and clustering quality. To our knowledge, DiscoPG is the first system to tackle the property graph schema discovery problem. As such, it supports the insightful exploration of the graph schema components and their evolving behavior, while revealing the underpinnings of the clustering-based discovery process.

References

1 Hanâ Lbath, Angela Bonifati, Russ Harmer: Schema Inference for Property Graphs. EDBT 2021: 499-504

2 Angela Bonifati, Stefania Dumbrava, Nicolas Mir: Hierarchical Clustering for Property Graph Schema Discovery. EDBT 2022: 2:449-2:453

3 Angela Bonifati, Stefania-Gabriela Dumbrava, Emile Martinez, Fatemeh Ghasemi, Malo Jaffré, Pacome Luton, Thomas Pickles: DiscoPG: Property Graph Schema Discovery and Exploration. Proc. VLDB Endow. 15(12): 3654-3657 (2022)

3.4 SHACL and SPARQL to detect inconsistencies in Wikidata

Nicolas Ferranti (Wirtschaftsuniversität Wien, AT)

Joint work of Nicolas Ferranti, Jairo Francisco de Souza, Shqiponja Ahmetaj, Axel Polleres
Main reference Nicolas Ferranti, Jairo Francisco de Souza, Shqiponja Ahmetaj, Axel Polleres: "Formalizing and Validating Wikidata's Property Constraints using SHACL and SPARQL", in Semartic Web Journal, 2024
URL https://www.semantic-web-journal.net/system/files/swj3611.pdf

In this talk, I delve into the crucial role of constraints in maintaining data integrity in knowledge graphs with a specific focus on Wikidata, one of the most extensive collaboratively maintained open data knowledge graphs on the Web. The World Wide Web Consortium (W3C) recommends SHACL as the constraint language for validating Knowledge Graphs, which comes in two different levels of expressivity, SHACL-Core, as well as SHACL-SPARQL. Despite the availability of SHACL, Wikidata currently represents its property constraints through its own RDF data model, which relies on Wikidata's specific reification mechanism. This talk discusses whether and how the semantics of Wikidata property constraints, can be formalized using SHACL-Core, SHACL-SPARQL, as well as directly as SPARQL queries.

3.5 PG-Keys: An Introduction

George Fletcher (TU Eindhoven, NL)

Main reference Renzo Angles, Angela Bonifati, Stefania Dumbrava, George Fletcher, Keith W. Hare, Jan Hidders, Victor E. Lee, Bei Li, Leonid Libkin, Wim Martens, Filip Murlak, Josh Perryman, Ognjen Savkovic, Michael Schmidt, Juan F. Sequeda, Slawek Staworko, Dominik Tomaszuk: "PG-Keys: Keys for Property Graphs", in Proc. of the SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021, pp. 2423–2436, ACM, 2021.
URL https://doi.org/10.1145/3448016.3457561

We give an introduction to PG-Keys, which together with PG-Schemas forms a community consensus proposal for property graph schema standards. PG-Keys enable the definition of key constraints on property graphs under different modes, which are combinations of basic restrictions that require the key to be exclusive, mandatory, and singleton. Further, PG-Keys can be defined on nodes, edges, and properties since in practice these all can represent valid entities. PG-Keys was an outcome of the Linked Data Benchmark Council's Property Graph Schema Working Group, consisting of members from industry, academia, and ISO GQL standards group, representing the past, present, and future of the science and practice of property graph data management.

3.6 Scalable Extraction of Shapes from Large Knowledge Graphs

Katja Hose (TU Wien, AT)

Joint work of Kashif Rabbani, Matteo Lissandrini, Katja Hose
Main reference Kashif Rabbani, Matteo Lissandrini, Katja Hose: "SHACTOR: Improving the Quality of Large-Scale Knowledge Graphs with Validating Shapes", in Proc. of the Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023, pp. 151–154, ACM, 2023.
URL https://doi.org/10.1145/3555041.3589723

Shapes, may they be formulated in SHACL or ShEx, have become an important instrument for validating knowledge graphs and ensuring that their content adheres to a set of well-defined constraints. Building upon such constraints, downstream applications incl. machine learning can benefit from the ensured or increased quality of knowledge graphs. Despite their usefulness in a broad range of situations, the adoption of shapes is hampered by the fact that there so far is a lack of tools that enable efficient mining of meaningful shapes. This motivated us to develop an approach that can automatically mine shapes from very large knowledge graphs [2] while providing an effective means to identify meaningful shapes based on support and confidence. As an extension, we have developed SHACTOR [1], which offers users a graphical interface and the opportunity to directly edit and update the underlying knowledge graph based on violations and inconsistencies identified after mining the shapes. While the current approach focuses on a subset of SHACL, we are working on extending the scope by enabling ShEx and a broader range of constraints.

References

1 Kashif Rabbani, Matteo Lissandrini, Katja Hose. *SHACTOR: Improving the Quality of Large-Scale Knowledge Graphs with Validating Shapes.*. SIGMOD Conference Companion. pp. 151-154, 2023

2 Kashif Rabbani, Matteo Lissandrini, Katja Hose. *Extraction of Validating Shapes from very large Knowledge Graphs.* Proc. VLDB Endow. 16(5), pp. 1023–1032, 2023

3.7 Data Provenance for SHACL

Maxime Jakubowski (Hasselt University, BE)

Joint work of Maxime Jakubowski, Thomas Delva, Anastasia Dimou, Jan Van den Bussche
Main reference Thomas Delva, Anastasia Dimou, Maxime Jakubowski, Jan Van den Bussche: "Data Provenance for SHACL", in Proc. of the Proceedings 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023, pp. 285–297, OpenProceedings.org, 2023.
URL https://doi.org/10.48786/EDBT.2023.23

Using SHACL, we present the notion of neighborhood of a node v satisfying a given shape in a graph G. This neighborhood is a subgraph of G, and provides data provenance of v for the given shape. We establish a correctness property for the obtained provenance mechanism, by proving that neighborhoods adhere to the Sufficiency requirement articulated for provenance semantics for database queries. As an additional benefit, neighborhoods allow a novel use of shapes: the extraction of a subgraph from an RDF graph, the so-called shape fragment. We compare shape fragments with SPARQL queries. We discuss implementation strategies for computing neighborhoods, and present initial experiments demonstrating that our ideas are feasible.

References

1 Thomas Delva, Anastasia Dimou, Maxime Jakubowski, Jan Van den Bussche. *Data Provenance for SHACL*. Proceedings 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28–31, 2023

3.8 Decision Problems in SHACL

George Konstantinidis (University of Southampton, GB) and Fabio Mogavero (University of Naples, IT)

Joint work of George Konstantinidis, Fabio Mogavero, Paolo Pareti
Main reference Paolo Pareti, George Konstantinidis, Fabio Mogavero: "Satisfiability and containment of recursive SHACL", J. Web Semant., Vol. 74, p. 100721, 2022.
URL https://doi.org/10.1016/J.WEBSEM.2022.100721

The Shapes Constraint Language (SHACL) is a W3C recommendation language used for validating RDF data by examining specific shapes within graphs. While previous research has largely focused on validation, the standard decision problems of satisfiability and containment have only been investigated for simplified versions of SHACL. In this talk, we offer a view of SHACL's diverse features and introduce Shape Constraint Logic (SCL), an extension of a first-order language that accurately captures SHACL's semantics. Additionally, we present MSCL, a second-order extension of SCL, which allows us to define, in a unified formal logic framework, the main recursive semantics of SHACL. Using this framework, we provide a detailed analysis of (un)decidability and complexity for the satisfiability and containment decision problems across different SHACL fragments. Notably, while both problems are undecidable for the complete language, we identify decidable combinations of interesting features, even amidst recursion.

References

1 Paolo Pareti, George Konstantinidis, Fabio Mogavero, Timothy J. Norman. *SHACL Satisfiability and Containment*. ISWC (1) 2020: 474-493

2 Paolo Pareti, George Konstantinidis, Fabio Mogavero. *Satisfiability and Containment of Recursive SHACL*. J. Web Semant. 74: 100721 (2022)

3.9 Introduction to ShEx

José Emilio Labra Gayo (University of Oviedo, ES)

Main reference Eric Prud'hommeaux, José Emilio Labra Gayo, Harold R. Solbrig: "Shape expressions: an RDF validation and transformation language", in Proc. of the 10th International Conference on Semantic Systems, SEMANTiCS 2014, Leipzig, Germany, September 4-5, 2014, pp. 32–40, ACM, 2014.
URL https://doi.org/10.1145/2660517.2660523

Shape Expressions (ShEx) is a concise and human-readable language to describe and validate RDF data. In this talk, we present an introduction to the ShEx language, describing the main features of the language and how it can be used for RDF validation. We started with a short motivation about the need of ShEx and we later presented the notion of shape in ShEx as well as the evolution of the language and its main motivation and features.

References

1 Eric Prud'hommeaux, Jose E. Labra Gayo, Harold Solbrig, *Shape expressions: an RDF validation and transformation language*. 10th International Conference on Semantic Systems, 32-40, Leizing, Germany

2 Jose E. Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, Dimitris Kontokostas, *Validating RDF data*. Springer Nature, 2017

3.10 ShEx and SHACL compared

José Emilio Labra Gayo (University of Oviedo, ES)

Main reference José Emilio Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, Dimitris Kontokostas: "Validating RDF Data", Morgan & Claypool Publishers, 2017.
URL https://doi.org/10.2200/S00786ED1V01Y201707WBE016

In this talk, we presented a comparison between ShEx (Shape Expressions) and SHACL (Shapes Constraint Language). Although they have several common features, there are several differences. An important one is the motivation for their design: while ShEx has more emphasis on Description and Validation, SHACL has more emphasis on Constraints and Validation which makes ShEx schemas more similar to a grammar that defines RDF data topologies, which SHACL shapes are more similar to a conjunction of constraints about RDF data.

References

1 Jose E. Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, Dimitris Kontokostas *Validating RDF data*. Springer Nature, 2018

3.11 Learning Schemas from Typed Graphs

Aurélien Lemay (INRIA Lille, FR)

In this talk, I present a learning algorithm for learning simple Shex expressions from typed graphs. These expressions do not include disjunction, counting, or negation. We demonstrate that while learning from a single typed graph is straightforward, the problem becomes more intricate for multi-typed graphs. This complexity arises partly due to the introduction of types. We isolate specific cases that lead to these difficulties and identify conditions under which the problem remains tractable (polynomial time) or becomes intractable.

References

1 Benoît Groz, Aurélien Lemay, Slawek Staworko, Piotr Wieczorek: Inference of Shape Graphs for Graph Databases. ICDT 2022: 14:1–14:20

3.12 PG-Schema: An introduction

Filip Murlak (University of Warsaw, PL)

Main reference 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, Dusan Zivkovic: "PG-Schema: Schemas for Property Graphs", CoRR, Vol. abs/2211.10962, 2022.
URL https://doi.org/10.48550/ARXIV.2211.10962

Property graphs have reached a high level of maturity, witnessed by multiple robust graph database systems as well as the ongoing ISO standardization effort aiming at creating a new standard Graph Query Language (GQL). Yet, despite documented demand, schema support is limited both in existing systems and in the first version of the GQL Standard. It is anticipated that the second version of the GQL Standard will include a rich DDL. Aiming to inspire the development of GQL and enhance the capabilities of graph database systems, we propose PG-Schema, a simple yet powerful formalism for specifying property graph schemas. It features PG-Types with flexible type definitions supporting multi-inheritance, as well as expressive constraints based on the recently proposed PG-Keys formalism.

3.13 An epistemic approach to model uncertainty in RegGXPath data-graphs

Nina Pardal (University of Sheffield, GB)

Joint work of Sergio Abriola, Santiago Cifuentes, Maria Vanina Martinez, Nina Pardal, Edwin Pin
Main reference Sergio Abriola, Santiago Cifuentes, Maria Vanina Martinez, Nina Pardal, Edwin Pin: "An epistemic approach to model uncertainty in data-graphs", Int. J. Approx. Reason., Vol. 160, p. 108948, 2023.
URL https://doi.org/10.1016/J.IJAR.2023.108948

Graph databases are becoming widely successful as data models that allow to effectively represent and process complex relationships among various types of data. Data-graphs are particular types of graph databases whose representation allows both data values in the paths and in the nodes to be treated as first class citizens by the query language. As with any other type of data repository, data-graphs may suffer from errors and discrepancies with respect to the real-world data they intend to represent. In this work, we explore the notion of probabilistic unclean data-graphs, in order to capture the idea that the observed (unclean) data-graph is actually the noisy version of a clean one that correctly models the world but that we know only partially. As the factors that lead to such a state of affairs may be many, e.g., all different types of clerical errors or unintended transformations of the data, and depend heavily on the application domain, we assume an epistemic probabilistic model that describes the distribution over all possible ways in which the clean (uncertain) data-graph could have been polluted. Based on this model we define two computational problems: data cleaning and probabilistic query answering and study for both of them their corresponding complexity when considering that the polluting transformation of the data-graph can be caused by either removing (subset), adding (superset), or modifying (update) nodes and edges. For data cleaning, we explore restricted versions when the transformation only involves updating data-values on the nodes.

3.14 The different "Shapes" of RDF(S) and OWL: a fragmented history

Axel Polleres (Wirtschaftsuniversität Wien, AT)

Since the introduction of the Semantic Web in the late 90s, schema and ontology languages to describe the schema of what we now call "Knowledge Graphs" have played a central role. The semantic basis of these ontology languages have – historically – been based on formalisms such as Frame Logic, Description Logics as well as Datalog. The syntactic representations of Schema axioms is integrated in Knowledge Graphs by representations of axioms in RDF, using the W3C standardised RDF, RDFS and OWL vocabularies. OWL and RDFS can therefore be both seen as logical languages, but also simply as RDF vocabularies, a constrained use of which allows us to "encode" terminological axioms as part of an RDF (knowledge) graph. Yet, an unconstrained use of these vocabularies yields obviously "unintuitive" graphs. In this short talk/paper we would like to discuss two questions, namely: (a) is there too much syntactic freedom in RDF and OWL? (b) (how) can useful syntactic fragments of OWL and RDFS usage be captured by constraints and shapes? In the course of (b) we also aim at providing an "historical" overview of (semantic and syntactic) OWL and RDFS fragments from the literature.

3.15 SHACL vs PG-Schema

Ognjen Savkovic (Freie Universität Bozen, IT)

We define SHACL abstract syntax and propose two semantics for the recursive case, and show their differences. We identify four selected issues that reflect differences between PG-Schema and SHACL. In particular, we discuss: (1) differences in labels for SHACL shape expressions and PG-Schema labels; (2) the support for negation in expressions; (3) quantification over edges and cardinality constraints; and finally (4) differences in open and closed constraints for both formalisms.

3.16 Towards a SHACL Validator under the Well-founded Semantics

Mantas Simkus (TU Wien, AT) and Cem Okulmus (University of Umeå, SE)

W3C has recently introduced SHACL as a new standard for integrity constraints on RDF graphs. Unfortunately, the standard defines the semantics of *non-recursive* constraints only, which has spurred recent research efforts into finding a suitable, mathematically crisp semantics for constraints with cyclic dependencies. To this end, Corman et al. [4] introduced a semantics related to *supported models* known in logic programming, while Andreşel et al. [1] presented a semantics based on *stable models* known in *Answer Set Programming (ASP)*. In [2], the authors argue that recursive SHACL can be naturally equipped with a semantics inspired in the *well-founded semantics* for recursive logic programs with default negation [2]. This semantics is not only intuitive, but it is also computationally tractable, unlike the previous proposals. In this talk and demo, we review the well-founded semantics of SHACL and present an implementation of a new validator for this semantics. The implementation combines multiple technologies in order to obtain good efficiency and coverage of the features of the SHACL standard. In particular, our ShaWell[1] system uses a sophisticated strategy to issue a series of SPARQL queries over an RDF triple store, whose results are post-processed to obtain a validation outcome. For non-recursive SHACL constraints, this yields a system that is largely compliant to the SHACL standard. In case of recursion, ShaWell additionally employs a deductive database engine DLV to evaluate a logic program that is produced as part of the validation process. In this way, the SHACL validation task is reduced to the problem of evaluating an ordinary logic program under the well-founded semantics. This method is similar to the one in [3], where a SAT solver was used for handling the supported model semantics from [4].

References

1 Medina Andreşel, Julien Corman, Magdalena Ortiz, Juan L. Reutter, Ognjen Savkovic, and Mantas Šimkus. Stable model semantics for recursive SHACL. In *WWW '20: The Web Conference 2020*, pages 1570–1580. ACM / IW3C2, 2020.

[1] `https://github.com/cem-okulmus/shawell`

2 Adrian Chmurovič and Mantas Šimkus. Well-founded semantics for recursive SHACL. In Mario Alviano and Andreas Pieris, editors, *Proceedings of the 4th International Workshop on the Resurgence of Datalog in Academia and Industry (Datalog-2.0 2022), Genova-Nervi, Italy, September 5, 2022*, volume 3203 of *CEUR Workshop Proceedings*, pages 2–13. CEUR-WS.org, 2022.

3 Julien Corman, Fernando Florenzano, Juan L. Reutter, and Ognjen Savkovic. Validating SHACL constraints over a SPARQL endpoint. In *Proc. of ISWC 2019*, volume 11778 of *LNCS*, pages 145–163. Springer, 2019.

4 Julien Corman, Juan L. Reutter, and Ognjen Savkovic. Semantics and validation of recursive SHACL. In *Proc. of ISWC 2018*, volume 11136 of *LNCS*, pages 318–336. Springer, 2018.

3.17 How the Wikidata Community Uses ShEx

Katherine Thornton (Yale University Library – New Haven, US)

The Wikidata community announced the debut of the schema namespace in 2019. Wikidata editors contribute schemas in the Shape Expressions language to the schema namespace. As of February 2024, editors have contributed more than four hundred schemas describing domains from the life sciences, to the humanities, to computing.

Wikidata editors use schemas to communicate data models and to validate entity data. Each schema page contains a link to the ShEx2 Simple Online Validator so that editors can identify which Wikidata items are currently in conformance with a schema and which items require changes in order to be brought into conformance. Each schema has a unique identifier and support for labels and descriptions in the human languages Wikidata accommodates.

Wikidata editors have written schemas leveraging many feature of ShEx including recursion and importing one schema into another.

As awareness of Wikidata's schema namespace grows, we anticipate more editors will author schemas, more editors will develop ShEx-based tooling for the community, and that the ecosystem of schemas that anyone can reuse and edit will continue to thrive.

3.18 Introduction to SHACL

Jan Van den Bussche (Hasselt University, BE)

Joint work of Jan Van den Bussche, Bart Bogaerts, Maxime Jakubowski
Main reference Bart Bogaerts, Maxime Jakubowski, Jan Van den Bussche: "Expressiveness of SHACL Features and Extensions for Full Equality and Disjointness Tests", Log. Methods Comput. Sci., Vol. 20(1), 2024.
URL https://doi.org/10.46298/LMCS-20(1:16)2024

We give an introduction to the W3C-recommended Shapes Constraint Language, SHACL. We present the syntax based on description logics introduced by Corman et al., and extended to full core SHACL by Jakubowski. We point out that SHACL views an RDF graph as an edge-labeled graph. This is interesting, since a working group during the seminar proposed edge-labeled graphs (with types) as a "greatest common lower bound" for RDF and property graphs. We present a generalization of SHACL in terms of general inclusions among shapes.

Under the natural semantics, and excluding closedness constraints, we remark that this generalization does not add expressive power compared to real SHACL where left-hand shapes must be targets of specific kinds only. We discuss expressiveness issues, and approaches to recursion. We touch briefly upon SHACL engines and systems research on the topic, and mention applications of shapes beyond validation.

3.19 Primer on GQL graph types

Hannes Voigt (Neo4j – Leipzig, DE)

ISO/IEC 39075:2024 – GQL defines "a database language for modeling structured data as a graph, and for storing, querying, and modifying that data in a graph database or other graph store". GQL was published in April 2024. Part of GQL is the concept of a graph type. A graph type as a GQL-object type describing a graph in terms of restrictions on its labels, properties, nodes, edges, and topology. The purpose of a graph type is to constrain the set of nodes and edges that can be contained in a graph. The talk gave an introduction into GQL graph types, covering the basics of the concept, DDL operations on graph types, syntax and semantics as well as the advanced topics of key label sets and structural consistency.

4 Working groups

4.1 "What is used in practice for graph data abstractions? What is needed in practice for graph data abstractions?": Working group report

Angela Bonifati (Université Claude Bernard – Lyon, FR & IUF – Paris, FR), Anastasia Dimou (KU Leuven, BE), Stefania Dumbrava (ENSIIE – Paris, FR), George Fletcher (TU Eindhoven, NL), Katja Hose (TU Wien, AT), George Konstantinidis (University of Southampton, GB), Aurélien Lemay (INRIA Lille, FR), Wim Martens (Universität Bayreuth, DE), Nina Pardal (University of Sheffield, GB), Liat Peterfreund (The Hebrew University of Jerusalem, IL), Katherine Thornton (Yale University Library – New Haven, US), Maria-Esther Vidal (TIB – Hannover, DE), and Hannes Voigt (Neo4j – Leipzig, DE)

Usability is a perennial topic in the study of data and knowledge systems. If we look at the first volume of the ACM Transactions on Database Systems, we find already McGee's criteria for usability for data abstractions (i.e., data models, query languages, schema languages): ease of comprehension and learning; ease of information modeling; ease of data definition and programming; and, ease of formalizability and theoretical study (McGee 1976). During the opening introductory round on the first day of the seminar, many participants mentioned these criteria as they apply to graph shapes. This cross-cutting concern led to the formation of a working group for the seminar, focusing on the questions: What is used in practice for graph data abstractions, and how can we find this out? What is needed in practice for graph data abstractions?

Graph data abstractions were recognized early as being close to "mental structures underlying human thinking" and hence beneficial for usability of data systems (Sowa 1976). In the working group, we discussed several motivational use cases around the challenge of usability of graph data abstractions. Some of these use cases were inspired from open-source healthcare data (Johnson 2023) with different usability needs. Other use cases focused on the freely available Wikidata knowledge graph on which communities of contributors have specific knowledge-intensive tasks. A third use case is provided by the analysis of open DBpedia query logs (Bonifati et al. 2020). A fourth use case is provided by a survey on how shapes are generated and adopted by the community (Rabbani et al. 2022).

We also focused on the terminology to indicate the actors of usability, distinguishing between data graph builders, analysts, and consumers (Li et al. 2024). Are they end users? Are they a broader group of people with diverse domain expertises (and beyond, people impacted by our work in society at large)? This entails the question of identifying the right usability level for each group of people depending on several factors, such as their familiarity with graph-based human-computer interfaces, or graph-based formalisms expressing high-level abstractions, such as data models, query languages and schema languages.

To advance towards human-centered graph abstractions, we also need to look into aspects of fairness and responsibility. Graphs are conceptualizations of a domain of interest and they encode human biases that potentially have beneficial and/or harmful impacts on people. We identified two main points of bias: bias in the graph conceptualization and abstraction level, e.g., the schema of a graph; and, bias in data instances. We must develop solutions that study, detect and mitigate bias on both levels. This is especially important as it is our community that is mainly responsible for the design and engineering of these abstractions, a very impactful dimension.

This opens several new research challenges. The first is to study the suitability and impact of current graph data abstractions for human benefits, at the technical, application, and societal levels. Second, and related to this, the limitations of current approaches must be studied from human-centered perspectives. Third, to adequately address these challenges, we must also, as a research community, make use of research methodologies typically deployed in other areas, such as AI fairness, (design and use) of programming languages, the Visualization and the HCI communities. We need to look in these communities for both quantitative and qualitative methodologies. For qualitative methods we should also look outside our own discipline to fields such as social sciences and cognitive psychology. Finally, we must understand how to translate the answers of these questions to the next generation of computer and data scientists. This will require us to investigate data systems education and update existing and design new curricula at all levels of education.

Addressing these challenges will call for opening new human-centered research lines in the data management community and beyond, as we experience a broader turn in the scientific community towards placing people and society more centrally in our work.

References

1 Angela Bonifati, Wim Martens, Thomas Timm. An analytical study of large SPARQL query logs. VLDB J. 29(2–3): 655–679 (2020).

2 Angela Bonifati, Ugo Comignani, Emmanuel Coquery, Romuald Thion. Interactive Mapping Specification with Exemplar Tuples. SIGMOD Conference 2017: 667–682.

3 A.E.W. Johnson, L. Bulgarelli, L. Shen, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 10, 1 (2023).

4 Paul Juillard, Angela Bonifati, Andrea Mauri. Interactive Graph Repairs for Neighborhood Constraints. EDBT 2024.

5 Harry X. Li, Gabriel Appleby, Camelia Daniela Brumar, Remco Chang, Ashley Suh. Knowledge Graphs in Practice: Characterizing their Users, Challenges, and Visualization Opportunities. IEEE Trans. Vis. Comput. Graph. 30(1): 584-594 (2024).

6 Matteo Lissandrini, Davide Mottin, Katja Hose, Torben Bach Pedersen. Knowledge Graph Exploration Systems: are we lost? CIDR 2022.

7 William C. McGee. On user criteria for data model evaluation. ACM Trans. Database Syst. 1(4): 370–387 (1976).

8 Kashif Rabbani, Matteo Lissandrini, Katja Hose. SHACL and ShEx in the Wild: A Community Survey on Validating Shapes Generation and Adoption. WWW (Companion Volume) 2022: 260-263.

9 John F. Sowa: Conceptual Graphs for a Data Base Interface. IBM J. Res. Dev. 20(4): 336-357 (1976).

4.2 Repairs and explanations in knowledge graph data management systems in the presence of shape constraints

Anastasia Dimou (KU Leuven, BE), Shqiponja Ahmetaj (TU Wien, AT), Nicolas Ferranti (Wirtschaftsuniversität Wien, AT), Maxime Jakubowski (Hasselt University, BE), José Emilio Labra Gayo (University of Oviedo, ES), Cem Okulmus (University of Umeå, SE), Nina Pardal (University of Sheffield, GB), Ognjen Savkovic (Freie Universität Bozen, IT), and Mantas Simkus (TU Wien, AT)

Context. The problem of assessing and managing data quality in knowledge graphs (KGs) is a long-standing issue that attracts significant attention both in industry and academia. The new proposals on schemas and constraints languages (such as SHACL, ShEx, PG-schema) for knowledge graphs (KGs), so-called shapes, have introduced new challenges. Since these new languages allow users to easily express complex properties over KGs, this requires new methods to verify them, deal with inconsistency, and repair the inconsistent data.

Setting. KGs are created throughout different processes and to identify the root causes of violations, we consider the so-called knowledge-based data management (KBDM) setting, which describes the creation and integration of data into the KG, a possible ontology of the KG, schema shapes, and possible relevant queries. The implementation of these systems in real-world applications yields a non-trivial situation, as several components need to be considered: the original data, the ontology, the mapping rules, the data graph, and the shape constraints. All of these components can contribute to violations in the final data graph and as such are potential candidates to be repaired. This opens several new research challenges, which need to be addressed separately.

Approach. Traditional approaches to data quality usually focus on the proposals that describe how to fix the data, i.e., which facts are missing and which facts should be deleted. That would be a viable approach in many settings where data is inserted manually and where the information about sources is not available. On the other hand, fixing the data may not often be sufficient and one may need to look at the sources of violation that may be rooted in inaccurate source data or poor design of the mapping rules or shape constraints. Finally, one may consider scenarios where one is given queries of interests, and while data may not be of sufficient fitness overall it may be sufficient to answer the given queries.

Questions. The group discussed the possible challenges that arise when trying to obtain high-quality KGs: How do we achieve high-quality KGs? How can we obtain more tailored constraints? How to describe and repair the violations obtained through the reasoning process? How to manufacture meaningful explanations that allow us to explain what was violated and how the repair was addressed? When is it meaningful to propose fixes on data, and when on mappings, shapes, or even queries? How can we introduce a probabilistic model into the KDBM, either on data or schemas, that may provide better ways of ranking violations and repairs?

Plans. As a result of the exchange of views and taking into consideration the different backgrounds of all the participants, the group has decided to pursue a comprehensive discussion and study of the state-of-the-art inconsistency management tasks for graph data, providing as well use-cases that may shed light on the motivation behind different research questions that remain open and which may be overlooked in academic environments. We will stay in touch for further collaboration, having a vision paper as a specific horizon and the result of the group's future efforts.

4.3 Relating 6NF and PG-Schema

Benoit Groz (University Paris-Saclay – Orsay, FR), Jan Hidders (Birkbeck, University of London, GB), Nina Pardal (University of Sheffield, GB), Slawomir Staworko (relationalAI – Berkeley, US), and Piotr Wieczorek (University of Wroclaw, PL)

Experience has shown us that relational database schemas are very versatile and can model a variety of data modeling approaches, including property graphs. Moreover, novel techniques in query processing, such as worst-case optimal joins, or Datalog query processing, indicate that a relational approach to graph database applications is viable.

In this working group, we discussed two questions:

1. Can we show in a systematic manner how schemas for property graphs, as expressed in the proposals of PG-Schema and PG-Keys, can be represented relationally, obtaining highly decomposed (6NF) schemas with key constraints and inclusion constraints such as foreign keys?
2. Can the intent of a graph database application be formalized in a suitable variant of EER (extended Entity-Relationship) diagrams?

The established semantics of EER diagrams is through a mapping to relational schemas with constraints. The group discussed various use cases of EER diagrams with additional constructs inspired by PG-Schema, such as multivalued or optional properties. By observing the relational schemas that are obtained for these use cases, one may form a rough picture of the classes of relational constraints needed to support various graph database applications. While we believe that primary and foreign keys are two classes of constraints that can be enforced very efficiently, we have carefully identified features of EER models that require more expressive constraints, which leaves as an important open question if they can also be efficiently enforced.

The group has found that there is a lack of consensus on the interpretation of EER diagrams, with multiple semantics proposed that differ on finer points that in the context of graph databases can have important ramifications. For instance, do relationships in EER diagrams allow for multiple links between the same pair of entity instances (multi-graphs)? Can we assume that all entities have object identifiers, thus assuming that all entities are subclasses of a single top superclass?

We pointed out that a good understanding of mappings from property graphs to EER, and from EER to relational, may yield as an added bonus, a method to visualize a class of relational database schemas as PG-Schema with PG-Keys, as well as a way to visualize PG-Schemas in the more familiar notation of EER diagrams. Members of the working group will continue to stay in touch with each other to further investigate this line of research.

4.4 Convergence of graph data models and schemas

Filip Murlak (University of Warsaw, PL) and Jan Hidders (Birkbeck, University of London, GB)

Joint work of Shqiponja Ahmetaj, Iovka Boneva, Jan Hidders, Maxime Jakubowski, Jose Emilio Labra Gayo, Leonid Libkin, Wim Martens, Fabio Mogavero, Filip Murlak, Cem Okulmus, Axel Polleres, Ognjen Savković, Mantas Šimkus

4.4.1 Objective

The goal is to understand the commonalities and differences between RDF and LPG (labelled property graphs), and their corresponding schema languages, ShEx and SHACL for RDF, and PG-Schema for LPG. We aim to identify a *common core* (a small but useful common sublanguage, easily expressible in all three formalisms) and a *common superlanguage* (a language that captures all three formalisms, yet remains manageable).

4.4.2 How do we compare schema languages?

Schema languages can be used for different purposes: prescriptive (e.g., refuse certain updates, refuse a certain graph as input), descriptive (e.g., defining a vocabulary), deriving information, or defining patterns (types or shapes) to be used as as part of a query language. From this, we have abstracted two concrete tasks: defining sets of valid graphs, and defining sets of nodes in a graph.

4.4.3 How to compare schema languages for different data models?

RDF and LPG are similar enough to make the comparison of their schema languages viable, but different enough to make direct comparison impossible. There are several workarounds. The simplest approach is to define restrictions on RDF and LPG that result in isomorphic data models. A more refined approach is to consider a canonical encoding of RDF in LPG, and vice versa. Then, the question is which constraints/patterns in the sourse schema language can be expressed in the target schema language over the encoded instances, and which constraints/patterns over encodings can be translated back to the source schema language. Finally, one could consider a whole class of (relatively simple) encodings of one data model in the other.

4.4.4 Plans

A group of participants coming from all three communities has declared interest in pursuing these goals together, aiming to produce a research paper within a half-year horizon. We plan to identify a common restriction of LPG and RDF, give the semantics of the three formalisms over the restricted data model, identify a common core and evaluate it against known usecases, and design a manageable common superlanguage.

Participants

- Shqiponja Ahmetaj
TU Wien, AT
- Iovka Boneva
Université de Lille I, FR
- Angela Bonifati
Université Claude Bernard – Lyon, FR & IUF – Paris, FR
- Anastasia Dimou
KU Leuven, BE
- Stefania Dumbrava
ENSIIE – Paris, FR
- Nicolas Ferranti
Wirtschaftsuniversität Wien, AT
- George Fletcher
TU Eindhoven, NL
- Benoit Groz
University Paris-Saclay – Orsay, FR
- Jan Hidders
Birkbeck, University of London, GB
- Katja Hose
TU Wien, AT
- Maxime Jakubowski
Hasselt University, BE
- George Konstantinidis
University of Southampton, GB
- José Emilio Labra Gayo
University of Oviedo, ES
- Aurélien Lemay
INRIA Lille, FR
- Leonid Libkin
University of Edinburgh, GB
- Wim Martens
Universität Bayreuth, DE
- Fabio Mogavero
University of Naples, IT
- Filip Murlak
University of Warsaw, PL
- Cem Okulmus
University of Umeå, SE
- Nina Pardal
University of Sheffield, GB
- Liat Peterfreund
The Hebrew University of Jerusalem, IL
- Axel Polleres
Wirtschaftsuniversität Wien, AT
- Ognjen Savkovic
Freie Universität Bozen, IT
- Mantas Simkus
TU Wien, AT
- Slawomir Staworko
relationalAI – Berkeley, US
- Katherine Thornton
Yale University Library – New Haven, US
- Jan Van den Bussche
Hasselt University, BE
- Maria-Esther Vidal
TIB – Hannover, DE
- Hannes Voigt
Neo4j – Leipzig, DE
- Piotr Wieczorek
University of Wroclaw, PL

Report from Dagstuhl Seminar 24111

Logics for Dependence and Independence: Expressivity and Complexity

Juha Kontinen*[1], Jonni Virtema*[2], Heribert Vollmer*[3], Fan Yang*[4], and Nicolas Fröhlich†[5]

1 **University of Helsinki, FI.** juha.kontinen@helsinki.fi
2 **University of Sheffield, GB.** j.t.virtema@sheffield.ac.uk
3 **Leibniz Universität Hannover, DE.** vollmer@thi.uni-hannover.de
4 **Utrecht University, NL.** fan.yang.c@gmail.com
5 **Leibniz Universität Hannover, DE.** nicolas.froehlich@thi.uni-hannover.de

Abstract

This report documents the programme and outcomes of Dagstuhl Seminar "Logics for Dependence and Independence: Expressivity and Complexity" (24111). This seminar served as a follow-up seminar to the highly successful seminars "Dependence Logic: Theory and Applications" (13071), "Logics for Dependence and Independence" (15261) and "Logics for Dependence and Independence" (19031). A key objective of the seminar was to bring together researchers working in dependence logic and in application areas (for this edition with a particular emphasis on the areas of hyperproperties and formal linguistics), so that they can communicate state-of-the-art advances and embark on a systematic interaction. The goal was especially to reach those researchers who have recently started working in this thriving area, as well as researchers working on several aspects of complexity studies of team-based logics as well as expressivity issues, in particular in the just mentioned application areas. In particular, bringing together researchers from areas of theoretical studies with the application areas aimed at enhancing the synergy between the different communities working on dependence logic.

Seminar March 10–15, 2024 – https://www.dagstuhl.de/24111
2012 ACM Subject Classification Theory of computation → Complexity theory and logic; Theory of computation → Logic and databases
Keywords and phrases finite model theory, formal linguistics, hyperproperties, information theory, team semantics
Digital Object Identifier 10.4230/DagRep.14.3.31

1 Executive Summary

Heribert Vollmer (Leibniz Universität Hannover, DE)
Juha Kontinen (University of Helsinki, FI)
Jonni Virtema (University of Sheffield, GB)
Fan Yang (Utrecht University, NL)

Dependence and independence are interdisciplinary notions that are pervasive in many areas of science. They appear in domains such as mathematics, computer science, statistics, quantum physics, and game theory. The systematic development of logical and semantical structures for these notions via the logics of dependence and independence has exposed surprising connections between these areas.

* Editor / Organizer
† Editorial Assistant / Collector

Logics for Dependence and Independence: Expressivity and Complexity, *Dagstuhl Reports*, Vol. 14, Issue 3, pp. 31–51
Editors: Juha Kontinen, Jonni Virtema, Heribert Vollmer, and Fan Yang
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Logics for dependence and independence are new tools for modeling dependencies and interaction in dynamical scenarios. Reflecting this, these logics often have higher expressive power and complexity than classical logics used for these purposes previously. During the past decade, pioneering results on logics for dependence and independence has been disseminated in a spectrum of respected international conferences such as LICS, MFCS, JELIA, LPAR, CSL, AiML, and FSTTCS, and in top journals in the areas of logic and theoretical computer science. Although significant progress has been made in understanding the computational side of these novel logics (see Section 2 for some examples) still many central questions remain unsolved so far. In addition to addressing the open questions, the seminar also aims at boosting the exchange of ideas and techniques between team-based logics and the application areas.

The complexity and expressivity aspects of logics in propositional, modal and first-order team semantics have been studied extensively during the past decade. Recently, the complexity theoretic focus has turned to the (parameterized) complexity of logically defined counting and enumeration problems as well as algebraic complexity of probabilistic and real-valued logics. Furthermore, the expressivity and complexity of the novel temporal team logics are also not yet well understood.

Logics for real valued data and probabilistic reasoning

Algorithmically, first-order dependence and independence logic correspond exactly to the complexity class NP and to the existential fragment of second-order logic (ESO) while inclusion logic corresponds to the complexity class P over ordered finite structures. Recent discoveries on the connections between so-called probabilistic independence logic and logics on real valued data have revealed similar fundamental connections to a computation paradigm that uses real numbers as primitive entities (so-called BSS paradigm). These probabilistic logics have fascinating connections to the area of information theory via the notion of entropy, which can be adopted as a dependency in the probabilistic team semantics framework.

Applications to Hyperproperties and Formal Verification

An emerging area of applications for team semantics is the area of Hyperproperties. In the field of formal verification an execution of a system is modeled by a trace depicting the evolution of the system over discrete time. Traceproperties, ubiquitous in formal verification, are properties of systems that boil down to verifying that each trace of the system satisfies that property. Hyperproperties on the other hand are properties of systems that cannot be reduced to checking properties of individual execution traces of the system in isolation, but are instead properties of sets of traces. These properties are of vital importance in applications concerning security and information flow. A canonical example here is bounded termination; one cannot check whether there exists a uniform time bound for some action by checking computation traces in isolation. Other examples include security policies such as *non-interference* and *secure information flow.*

Applications to Formal Linguistics

Team semantics was also proven to be a fruitful tool for formal linguistics, especially for *inquisitive semantics* and the study of *free choice inferences.* Inquisitive semantics is a unified formal framework for analyzing both statements and questions in natural language. It is known that inquisitive logic essentially adopts team semantics and can thus be viewed as a variant of propositional dependence logic. This connection has already sparked a significant amount of interest and new research at the interface of the two fields. On a different line of research, recently a bilateral modal logic based on team semantics, called BSML, was developed to model free choice inferences in natural language, where an atom NE studied in the context of propositional team logics plays a central role. Very recent works have studied the logical properties of BSML, and promising broader applications of the team semantics method along this line are yet to be explored.

Organization of the Seminar and Activities

The seminar brought together 42 researchers from mathematical logic, natural language semantics, and theoretical computer science. The participants consisted of both senior and junior researchers, including a number of postdoctoral researchers and advanced graduate students.

Participants were invited to present their work and to communicate state-of-the-art advances. Over the five days of the seminar, 29 talks of various lengths took place. Introductory and tutorial talks of 60 minutes were scheduled prior to the seminar. The remaining slots were filled with shorter talks, mostly scheduled after the seminar commenced. Furthermore the seminar included an open problem session and a concluding perspectives address.

The tutorial talks took place in the beginning of the week in order to establish a common background for the different communities that came together for the seminar. The presenters and topics were:

- Jonni Virtema: Introduction to Team Semantics
- Erika Ábrahám: (Probabilistic) Hyperproperties
- Maria Aloni: Logic and Language: Linguistic Applications of Team Semantics
- Till Miltzow: Existential Theory of the Reals
- Cheuk Ting Li: The Undecidability of Probabilistic Conditional Independence Implication

In addition, the seminar consisted of 24 shorter contributed talks, addressing various topics concerning expressibility, complexity and applications of team-based logics.

A one hour long open problem session was held on Wednesday morning, just before the hike ("Open Problem Walk"). It was moderated by Juha Kontinen. The session was announced already on Monday morning to give participants the opportunity to register for the session. Besides a couple of shorter contributions on decidability of Team-LTL (by Martin Zimmerman), expressivity of different forms of implications when added to inclusion (predicate) logic (by Jouko Väänänen), and expressivity of propositional independence logic (by Fan Yang), the session consisted of three longer introductions of the following open problems:

- Is PosSLP, the question if a given straight-line program (over the integers with operations of addition, multiplication and subtraction) computes a positive number, solvable in polynomial time? It is conjectured that NP with an oracle to PosSLP equals the complexity class $\exists\mathbb{R}$. (*Till Miltzow*)
- Are all probabilistic conditional independence implications derivable from information inequalities? (*Milan Studený*)
- Identify tractable fragment for model checking for dependence logic, that is, fragments with an effective syntax that are "natural" and "useful" in the sense that they can express interesting computational problems in a relatively straightforward way, have strictly higher expressive power than first-order logic FO, and have a polynomial-time model checking in data complexity. (*Phokion Kolaitis*)

The participants were asked to contribute more open problems to a collection in form of an Overleaf project.

The seminar ended with a perspectives address given by Jouko Väänänen just before Friday lunch.

Concluding Remarks

The seminar achieved its aim of bringing together researchers from various related communities to share state-of-the-art research. Considerable exchange took place between researchers in the application areas of hyperproperties and formal semantics and those working more theoretically on complexity and expressivity questions of team-based logics. The organizers left ample time for interaction outside of this schedule of talks and, as a result, many fruitful discussions between participants took place throughout the afternoons and evenings.

The organizers regard the seminar as a significant success. Bringing together researchers from different areas fostered valuable interactions and led to fruitful discussions. Feedback from the participants was very positive as well.

Finally, the organizers wish to express their gratitude to the Scientific Directorate of the Center for its support of this Dagstuhl Seminar.

2 Table of Contents

Executive Summary

Overview of Talks

Open problems

3 Overview of Talks

3.1 (Probabilistic) Hyperproperties

Erika Ábrahám (RWTH Aachen, DE)

Joint work of Erika Ábrahám, Borzoo Bonakdarpour, Ezio Bartocci, Gianfranco Ciardo, Oyendrila Dobe, Lina Gerlach, Christof Löding, Eshita Zaman

Main reference Erika Ábrahám, Borzoo Bonakdarpour: "HyperPCTL: A Temporal Logic for Probabilistic Hyperproperties", in Proc. of the Quantitative Evaluation of Systems - 15th International Conference, QEST 2018, Beijing, China, September 4-7, 2018, Proceedings, Lecture Notes in Computer Science, Vol. 11024, pp. 20–35, Springer, 2018.

URL http://dx.doi.org/10.1007/978-3-319-99154-2_2

Four decades ago, Lamport used the notion of trace properties as a means to specify the correctness of individual executions of concurrent programs. This notion was later formalized and classified by Alpern and Schneider to safety and liveness properties. Temporal logics like LTL and CTL were built based on these efforts to give formal syntax and semantics to requirements of trace properties. Subsequently, verification algorithms were developed to reason about individual executions of a system.

However, it turns out that many interesting requirements are not trace properties. For example, important information-flow security policies (e.g. noninterference, observational determinism) or service level agreements (e.g. mean response time, percentage uptime) cannot be expressed as properties of individual execution traces of a system. Rather, they are properties of sets of execution traces, also known as hyperproperties. Temporal logics such as HyperLTL and HyperCTL* have been proposed to provide a unifying framework to express and reason about hyperproperties.

This talk focussed on a special class of hyperproperties: we asked the question what are hyperproperties in the context of systems with random behavior. We discussed what are relevant probabilistic relations between independent executions of a system, how we can formally express them in a temporal logic, and how we can decide the truth of such statements.

3.2 Logic and Language - Linguistic applications of team semantics

Maria Aloni (University of Amsterdam, NL)

Joint work of Maria Aloni, Marco Degano

In the first part of the talk I surveyed linguistic applications of team-based logics including IF-logic (branching and exceptional scope of indefinites) [Hintikka and Sandu 1998], inquisitive (epistemic) logic (questions, attitude verbs) [Ciardelli, Groenendijk and Roelofsen 2018] and Bilateral State-Based Modal Logic (non-classical inference, including free choice)[Aloni 2022]. In the second part I presented an application of a two-sorted dependence logic [Väänänen 2007] to capture cross-linguistic variations in the expression of specificity [Aloni and Degano 2022].

Indefinites are known to give rise to different scopal (specific vs nonspecific) and epistemic (known vs unknown) uses. Farkas and Brasoveanu [2020] explained these specificity distinctions in terms of stability vs. variability in value assignments of the variable introduced by the indefinite. Typological research [Haspelmath, 1997] showed that indefinites have different functional distributions with respect to these uses. In the talk I presented a two-sorted dependence logic, with dependence, inclusion and variation atoms, where Farkas and Brasoveanu [2020]'s ideas can be rigorously formalised. I further applied the framework to explain typological variety of indefinites, their restricted distribution and licensing conditions, and some diachronic developments of indefinite forms.

References

1 Maria Aloni. Logic and conversation: the case of free choice. *Semantics and Pragmatics*, 5, 2022.

2 Maria Aloni and Marco Degano. (Non-)specificity across languages: constancy, variation, v-variation. *Semantic and Linguistic Theory* (SALT) 32, 2022.

3 Ivano Ciardelli, Jeroen Groenendijk and Floris Roelofsen. *Inquisitive Semantics*. Oxford University Press, 2018.

4 Donka Farkas and Adrian Brasoveanu. Kinds of (Non)Specificity. *The Wiley Blackwell Companion to Semantics*, pages 1–26, 2020.

5 Martin Haspelmath. *Indefinite Pronouns*. Oxford University Press, 1997.

6 Jakko Hintikka and Gabriel Sandu, Informational Independence as a Semantical Phenomenon, in *Logic, Methodology and Philosophy of Science*, Vol. 8, J. E. Fenstad, I. T. Frolov, and R. Hilpinen (eds.), Amsterdam: Elsevier, pp. 571–589, 1989.

7 Jouko Väänänen. *Dependence Logic: A New Approach to Independence Friendly Logic*, volume 70. Cambridge University Press, 2007.

3.3 Deep inference sequent calculi for propositional logics with team semantics

Aleksi Ilari Anttila (University of Amsterdam, NL)

Joint work of Aleksi Ilari Anttila, Rosalie Iemhoff, Fan Yang

While natural deduction- and Hilbert-style axiomatizations of logics employing team semantics such dependence logic and inquisitive logic have been extensively studied, the development of sequent calculi and proof theory in general for these logics appears to have been hindered by the factors such as the fact these logics are not closed under uniform substitution. The sequent calculi which have been constructed for these logics thus far have been labelled or multi-type systems. We propose a different approach: by appending a few deep inference-style rules (rules which can act not only on the immediate subformulas and main connectives of formulas, but also on subformulas and connectives deeper within a formula) to a standard Gentzen-style calculus, we obtain a very simple system for at least one of these logics.

3.4 Conditionals for union-closed languages

Fausto Barbero (University of Helsinki, FI)

Joint work of Fausto Barbero, Fan Yang

Are team logics logics? In the 70's, a similar criticism was addressed to quantum logic, on the basis that the latter lacks a (proof-theoretically) well-behaved conditional operator. Similarly, many of the languages considered in the literature on team semantics lack (and fail to define) an adequate conditional, and as a consequence they are not amenable to simple Hilbert-style axiomatizations. This situation was cleared by the discovery that downward closed languages can be enriched with a well-behaved conditional (the inquisitive implication), in many cases without increase in expressive power (propositional case) or, at least, without losing the property of downward closure.

It is less clear what could be taken as a conditional for languages that are not downward closed. As an interesting special case, the literature is rich with union-closed languages, e.g. inclusion logics and languages with possibility operators or relevant disjunctions. I will show that there is no conditional that preserves union closure and satisfies the key proof-theoretical constraints of modus ponens and the deduction theorem. Such an operator is missing not only for the whole class of union-closed languages, but also when trying to extend some specific individual languages.

I will then consider four candidates for the role of conditional for union closed languages: four binary operators that preserve union closure, respect restricted forms of modus ponens and the deduction theorem, and share many properties with what is usually considered a "conditional".

3.5 Hyperteams for compositionality and determinacy in logics for games

Dario Della Monica (University of Udine, IT)

Joint work of Dylan Bellier, Massimo Benerecetti, Dario Della Monica, Fabio Mogavero
Main reference Dylan Bellier, Massimo Benerecetti, Dario Della Monica, Fabio Mogavero: "Good-for-Game QPTL: An Alternating Hodges Semantics", ACM Trans. Comput. Log., Vol. 24(1), pp. 4:1–4:57, 2023.
URL http://dx.doi.org/10.1145/3565365

Team semantics is the compositional infrastructure at the basis of so-called logics for (in)dependencies [1, 2, 3], making it possible to specify the set of first-order variables upon which another first-order variable value is allowed to depend on, overruling the standard linear dependence relation imposed by the quantification prefix.

On the one hand, this ability comes in handy for modeling games with incomplete information, where moves of one player may be dependent on only some of the moves of the other player, while being independent of the other ones. On the other hand, teams (i.e., sets of assignments) are meant to collect the uncertainty about possible evaluations of only a subset of variables, e.g., the universally quantified ones; it is thus possible to only specify dependencies for the remaining variables, e.g., the existentially quantified ones.

To overcome this asymmetry, which strides with the symmetric treatment of players in games, we extend the notion of teams with the one of hyperteams (i.e., sets of teams), thus allowing for a symmetric treatment of existentially and universally quantified variables.

We study variants of both Quantified Propositional Temporal Logic (QPTL) and First-Order Logic (FOL), replacing their standard semantics with one based on hyperteams, thus obtaining Good-for-Game QPTL (GFG-QPTL) [4] and Alternating Dependence/Independence-Friendly Logic (ADIF) [5]. Both these logics enjoy determinacy, which makes them particularly apt to model 2-player zero-sum games. We provide both compositional and game-theoretic semantics, and study the complexity of satisfiability (for GFG-QPTL) and model checking (for both GFG-QPTL and ADIF) problems.

References

1 J. Hintikka and G. Sandu. *Informational Independence as a Semantical Phenomenon*, in: International Congress on Logic, Methodology, and Philosophy of Science, Elsevier, pp. 571–589, 1989.

2 W. Hodges. *Compositional Semantics for a Language of Imperfect Information*. Logic Journal of the IGPL 5, 539–563, 1997.

3 J.A. Väänänen. *Dependence Logic: A New Approach to Independence Friendly Logic*, volume 70 of London Mathematical Society Student Texts, Cambridge University Press, 2007.

4 D. Bellier, M. Benerecetti, D. Della Monica, and F. Mogavero. *Good-for-Game QPTL: An Alternating Hodges Semantics*. Transactions On Computational Logic 24, 4:1–57, 2023.

5 D. Bellier, M. Benerecetti, D. Della Monica, and F. Mogavero. *Alternating (In)Dependence-Friendly Logic*. Annals of Pure and Applied Logic, 174(10), 2023.

3.6 Modular SAT-based techniques for reasoning tasks in team semantics

Arnaud Durand (Paris Cité University, FR), Juha Kontinen (University of Helsinki, FI), and Jouko Väänänen (University of Helsinki, FI)

We study the complexity of reasoning tasks for logics in team semantics. Our main focus is on the data complexity of model checking but we also derive new results for logically defined counting and enumeration problems. Our approach is based on modular reductions of these problems into the corresponding problems of various classes of Boolean formulas. We illustrate our approach via several new tractability/intractability results.

3.7 The propositional logic of teams

Fredrik Engström (University of Gothenburg, SE)

Joint work of Fredrik Engström, Orvar Lorimer Olsson
Main reference Fredrik Engström, Orvar Lorimer Olsson: "The propositional logic of teams", CoRR, Vol. abs/2303.14022 (2023).
URL https://arxiv.org/abs/2303.14022

Logics based on team semantics often fail to be substitutional, limiting any algebraic treatment, and rendering schematic uniform proof systems impossible. This shortcoming can be attributed to the flatness principle, commonly adhered to when generating team semantics.

Investigating the formation of team semantics from algebraic semantics, and disregarding the flatness principle, we present the logic of teams, LT, a substitutional logic in which important propositional team logics are axiomatisable as fragments. Starting from classical propositional logic and Boolean algebras, we give a semantics for LT by considering the algebras that are powersets of Boolean algebras B, i.e., of the form P(B), equipped with internal (pointwise) and external (set theoretic) connectives. Furthermore, we present a well-motivated complete and sound labelled natural deduction system for LT.

3.8 Second-Order Hyperproperties

Hadar Frenkel (CISPA - Saarbrücken, DE)

Joint work of Raven Beutner, Bernd Finkbeiner, Hadar Frenkel, Niklas Metzger
Main reference Raven Beutner, Bernd Finkbeiner, Hadar Frenkel, Niklas Metzger: "Second-Order Hyperproperties", in Proc. of the Computer Aided Verification - 35th International Conference, CAV 2023, Paris, France, July 17-22, 2023, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 13965, pp. 309–332, Springer, 2023.
URL http://dx.doi.org/10.1007/978-3-031-37703-7_15

We introduce Hyper2LTL, a temporal logic for the specification of hyperproperties that allows for second-order quantification over sets of traces. Unlike first-order temporal logics for hyperproperties, such as HyperLTL, Hyper2LTL can express complex epistemic properties like common knowledge, Mazurkiewicz trace theory, and asynchronous hyperproperties. The model checking problem of Hyper2LTL is, in general, undecidable. For the expressive fragment where second-order quantification is restricted to smallest and largest sets, we present an approximate model-checking algorithm that computes increasingly precise under- and overapproximations of the quantified sets, based on fixpoint iteration and automata learning. We report on encouraging experimental results with our model-checking algorithm, which we implemented in the tool HySO.

3.9 Strongly First Order Dependencies and Dual Negation in Team Semantics

Pietro Galliani (University of Insubria - Varese, IT)

Much (although not all) of the work in the area of Team Semantics assumes that all expressions are in negation normal form (and, in particular, that no negated dependence atoms can appear).

I will argue that, even though this choice has valid historical reasons and is appropriate for certain logics based on Team Semantics, in general there is no compelling reason not to introduce (dual) negation; and I will re-examine the question of finding out which dependency families lead to logics that are reducible to First Order Logic in this more general setting.

3.10 Hidden Variables in Quantum Mechanics and Logics for Dependence and Independence

Erich Grädel (RWTH Aachen, DE)

Joint work of Rafael Albert, Erich Grädel
Main reference Rafael Albert, Erich Grädel: "Unifying hidden-variable problems from quantum mechanics by logics of dependence and independence", Ann. Pure Appl. Log., Vol. 173(10), p. 103088, 2022.
URL http://dx.doi.org/10.1016/J.APAL.2022.103088

We study hidden-variable models from quantum mechanics, and their abstractions in purely probabilistic and relational frameworks, by means of logics of dependence and independence, based on team semantics. We show that common desirable properties of hidden-variable models can be defined in an elegant and concise way in dependence and independence logic. The relationship between different properties, and their simultaneous realisability can thus been formulated and a proved on a purely logical level, as problems of entailment and satisfiability of logical formulae. Connections between probabilistic and relational entailment in dependence and independence logic allow us to simplify proofs. In many cases, we can establish results on both probabilistic and relational hidden-variable models by a single proof, because one case implies the other, depending on purely syntactic criteria. We also discuss the 'no-go' theorems by Bell and Kochen-Specker and provide purely logical variants of the latter, introducing non-contextual choice as a team-semantical property.

3.11 Temporal Team Semantics Revisited

Jens Gutsfeld (TU Braunschweig, DE)

Joint work of Jens Oliver Gutsfeld, Arne Meier, Christoph Ohrem, Jonni Virtema
Main reference Jens Oliver Gutsfeld, Arne Meier, Christoph Ohrem, Jonni Virtema: "Temporal Team Semantics Revisited", in Proc. of the LICS '22: 37th Annual ACM/IEEE Symposium on Logic in Computer Science, Haifa, Israel, August 2 - 5, 2022, pp. 44:1–44:13, ACM, 2022.
URL http://dx.doi.org/10.1145/3531130.3533360

In this talk, we study a novel approach to asynchronous hyperproperties by reconsidering the foundations of temporal team semantics. We consider three logics: TeamLTL, TeamCTL and TeamCTL*, which are obtained by adding quantification over so-called time evaluation functions controlling the asynchronous progress of traces. We then relate synchronous to our new logics and show how it can be embedded into them. We discuss that the model checking problem for with Boolean disjunctions is highly undecidable by encoding recurrent computations of non-deterministic 2-counter machines. Finally, we present a translation from to Alternating Asynchronous Büchi Automata and obtain decidability results for the path checking problem as well as restricted variants of the model checking and satisfiability problems.

3.12 Conditional independence on semiring relations

Miika Hannula (University of Helsinki, FI)

Main reference Miika Hannula: "Conditional Independence on Semiring Relations", in Proc. of the 27th International Conference on Database Theory, ICDT 2024, March 25-28, 2024, Paestum, Italy, LIPIcs, Vol. 290, pp. 20:1–20:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.
URL http://dx.doi.org/10.4230/LIPICS.ICDT.2024.20

Conditional independence plays a foundational role in database theory, probability theory, information theory, and graphical models. Many properties of conditional independence are shared across various domains, and to some extent these commonalities can be studied through a measure-theoretic approach. In this talk we consider an alternative approach via semiring relations, defined by extending database relations with tuple annotations from some commutative semiring. Integrating various interpretations of conditional independence in this context, we investigate how the choice of the underlying semiring impacts the corresponding axiomatic and decomposition properties.

3.13 Implication problems for some qualitative and quantitative dependencies

Minna Eveliina Hirvonen (University of Helsinki, FI)

Main reference Minna Hirvonen: "The Implication Problem for Functional Dependencies and Variants of Marginal Distribution Equivalences", in Proc. of the Foundations of Information and Knowledge Systems - 12th International Symposium, FoIKS 2022, Helsinki, Finland, June 20-23, 2022, Proceedings, Lecture Notes in Computer Science, Vol. 13388, pp. 130–146, Springer, 2022.
URL http://dx.doi.org/10.1007/978-3-031-11321-5_8

A dependence or independence atom is a statement that some variables are in some sense (in)dependent. A set of atoms S is said to logically imply another atom s if every suitable object (e.g. a database or a distribution) that satisfies all of the atoms in S also satisfies the atom s. An implication problem is the task of deciding whether a given set of atoms S logically implies another given atom s.

In this talk, I will present some axiomatization results for implication problems for classes that combine different types of qualitative and quantitative atoms. We consider two different implication problems that combine well-known qualitative and quantitative atoms with two lesser-known quantitative atoms: unary marginal identity and unary marginal distribution equivalence. A unary marginal identity states that two variables x and y are identically distributed. A unary marginal distribution equivalence states that the multiset consisting of the marginal probabilities of all the values for variable x is the same as the corresponding multiset for y.

The first one of these implication problems combines unary marginal identity and unary marginal distribution equivalence with functional dependency. The second implication problem combines the two atoms with probabilistic independence. Both implication problems have a sound and complete axiomatization and a polynomial-time algorithm.

3.14 Approximate dependence atoms

Matilda Häggblom (University of Helsinki, FI)

Joint work of Matilda Häggblom, Åsa Hirvonen

In the team semantic setting, Väänänen (2017) defined and axiomatized approximate dependence atoms suitable for cases when "almost dependence" is permittable. An approximate dependence atom is satisfied by a team if there exists a large enough subteam that satisfies the corresponding usual dependence atom.

We aim to define and axiomatize approximate versions of other dependence atoms, such as exclusion and inclusion. Depending on the properties of the atom, such as downward closure or the lack thereof, different definitions of the atoms' approximate versions might be needed. We present some preliminary results regarding axiomatizations of approximate exclusion and anonymity atoms.

References

1 Jouko Väänänen. *The Logic of Approximate Dependence*. In C. Basket, L. Moss, & R. Ramanujam (Eds.), Rohit Parikh on Logic, Language and Society (Vol. 11, pp. 227-234), Springer, 2017.

3.15 Approximate Implication for Probabilistic Graphical Models

Batya Kenig (Technion - Haifa, IL)

Main reference Batya Kenig: "Approximate Implication for Probabilistic Graphical Models", CoRR, Vol. abs/2310.13942, 2023.
URL http://dx.doi.org/10.48550/ARXIV.2310.13942

The graphical structure of Probabilistic Graphical Models (PGMs) represents the conditional independence (CI) relations that hold in the modeled distribution. The premise of all current systems-of-inference for deriving conditional independence relations in PGMs, is that the set of CIs used for the construction of the PGM hold exactly. In practice, algorithms for extracting the structure of PGMs from data discover approximate CIs that do not hold exactly in the distribution. In this work, we ask how the error in this set propagates to the inferred CIs read off the graphical structure. More precisely, what guarantee can we provide on the inferred CI when the set of CIs that entailed it hold only approximately? In this talk, I will describe new positive and negative results concerning this problem.

3.16 Expressive power: BSML and Propositional Independence Logic

Søren Brinck Knudstorp (University of Amsterdam, NL)

In recent work, Aloni, Anttila and Yang (2023) present two extensions of the modal team logic BSML, demonstrating their expressive completeness for all properties [invariant under bounded bisimulation] and all union-closed properties, respectively, and leave open the problem of characterizing the expressive power of BSML. Continuing this line of work, we solve this problem by showing that BSML is expressively complete for all convex, union-closed properties.

We then shift our focus to independence logic, characterizing the expressive power of propositional independence logic with the inquisitive disjunction.

3.17 Characterizing Data Dependencies

Phokion G. Kolaitis (University of California - Santa Cruz, US)

Joint work of Marco Console, Phokion G. Kolaitis, Andreas Pieris
Main reference Marco Console, Phokion G. Kolaitis, Andreas Pieris: "Model-theoretic Characterizations of Rule-based Ontologies", in Proc. of the PODS'21: 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Virtual Event, China, June 20-25, 2021, pp. 416–428, ACM, 2021.
URL http://dx.doi.org/10.1145/3452021.3458310

Data dependencies are integrity constraints that the data at hand ought to obey. After functional dependencies were introduced by E.F. Codd in the 1970s, several other kinds of data dependencies were defined and extensively studied. Eventually, it was realized that essentially all data dependencies are either equality generating dependencies (egds) or tuple generating dependencies (tgds); the former generalize functional dependencies, while the latter generalize inclusion dependencies, multi-valued dependencies, and full tgds. In 1987, Makowsky and Vardi characterized full tgds and egds in terms of their structural properties, such as domain independence, closure under direct products, and modularity. The aim of this talk is to present characterizations of arbitrary tgds and egds that employ a new notion of locality. This is joint work with Marco Console and Andreas Pieris.

3.18 The undecidability of probabilistic conditional independence implication

Cheuk Ting Li (The Chinese University of Hong Kong, HK)

Main reference Cheuk Ting Li: "Undecidability of Network Coding, Conditional Information Inequalities, and Conditional Independence Implication", IEEE Trans. Inf. Theory, Vol. 69(6), pp. 3493–3510, 2023.
URL http://dx.doi.org/10.1109/TIT.2023.3247570

The probabilistic conditional independence implication problem is to decide whether a given statement on the conditional independence among several random variables is implied by a given list of such statements. This problem was shown to be undecidable, that is, there is no algorithm that is guaranteed to solve this problem. We will also describe the relation between this problem and the periodic tiling problem, and its implication on network coding.

References

1 C. T. Li, "Undecidability of Network Coding, Conditional Information Inequalities, and Conditional Independence Implication," IEEE Trans. Inf. Theory 69(6): 3493-3510, 2023.
2 C. T. Li, "The Undecidability of Conditional Affine Information Inequalities and Conditional Independence Implication With a Binary Constraint," IEEE Trans. Inf. Theory 68(12): 7685-7701, 2022.

3.19 Parameterized Complexity of Weighted Team Definability

Yasir Mahmood (Universität Paderborn, DE)

Joint work of Juha Kontinen, Yasir Mahmood, Arne Meier, Heribert Vollmer
Main reference Juha Kontinen, Yasir Mahmood, Arne Meier, Heribert Vollmer: "Parameterized complexity of weighted team definability", Mathematical Structures in Computer Science, p. 1–15, 2024.
URL http://dx.doi.org/10.1017/S0960129524000033

This Talk is based on our recent article with the same title: Parameterized Complexity of Weighted Team Definability In this article, we study the complexity of weighted team definability for logics with team semantics. This is a natural analogue of one of the most studied problems in parameterized complexity, the notion of weighted Fagin-definability, which is formulated in terms of satisfaction of first-order formulas with free relation variables. We focus on the parameterized complexity of weighted team definability for a fixed formula φ of central team-based logics. Given a first-order structure A and the parameter value $k \in N$ as input, the question is to determine whether $(A, T) \models \varphi$ for some team T of size k. We show several results on the complexity of this problem for dependence, independence, and inclusion logic formulas. Moreover, we also relate the complexity of weighted team definability to the complexity classes in the well-known W-hierarchy as well as paraNP.

3.20 A Parameterized View on the Complexity of Dependence Logic

Arne Meier (Leibniz Universität Hannover, DE)

Joint work of Juha Kontinen, Arne Meier, Yasir Mahmood
Main reference Juha Kontinen, Arne Meier, Yasir Mahmood: "A parameterized view on the complexity of dependence and independence logic", J. Log. Comput., Vol. 32(8), pp. 1624–1644, 2022.
URL http://dx.doi.org/10.1093/LOGCOM/EXAC070

In this talk, we give an overview of different parameterisations in propositional dependence logic as well as for first-order dependence logic. We start with a short primer on parameterised complexity theory and then dive into the results for the two dependence logic variants mentioned.

3.21 Existential Theory of the Reals

Till Miltzow (Utrecht University, NL)

Joint work of Jean Cardinal, Till Miltzow, Marcus Schaefer

During the seminar at Dagstuhl, I gave a presentation about ER-complete problems. We gave many examples of ER-complete problems from different domains. We showed techniques to show ER-completeness, that are different from reductions for NP-completeness. Furthermore, we explained some common phenomena, like high solution precision and topological universality.

3.22 A Set Based Semantics for Asynchronous TeamLTL

Max Sandström (University of Sheffield, GB)

Joint work of Juha Kontinen, Max Sandström, Jonni Virtema

In this talk I will present a relaxed set based variant of asynchronous linear temporal logic under team semantics. Linear temporal logic (LTL) is used in system verification to write formal specifications for reactive systems. However, some relevant properties, e.g. non-inference in information flow security, cannot be expressed in LTL. A class of such properties that has recently received ample attention is known as hyperproperties. Team semantics offers an avenue to capture such hyperproperties. The asynchronous variant I will present gives a bottom-up approach to capturing hyperproperties, as the asynchronous variant is expressively equivalent with LTL, but it grows in expressivity quickly with extensions. I will introduce the extensions of TeamLTL with the Boolean disjunction and a fragment of the extension of TeamLTL with the Boolean negation, where the negation cannot occur in the left-hand side of the Until-operator or within the Global-operator. I will present complexity results of the model checking problem, as well as some results relating the logics with other logics capturing hyperproperties.

Based on Joint work with Juha Kontinen and Jonni Virtema.

3.23 Lattices of abstract conditional independence models and their implication-based description

Milan Studený (The Czech Academy of Sciences - Prague, CZ)

Joint work of Tobias Boege, Janneke H. Bolt, Milan Studený
Main reference Tobias Boege, Janneke H. Bolt, Milan Studený: "Self-adhesivity in lattices of abstract conditional independence models", CoRR, Vol. abs/2402.14053v1, 2024.
URL https://arxiv.org/abs/2402.14053v1

In a recent manuscript [2], an abstract approach to *conditional independence* (CI) structures has been introduced. The approach was inspired by the idea of deriving further probabilistically valid implications among CI statements based on the idea of *self-adhesion* of CI models. That particular method is an abstraction of an information-theoretical method to derive non-Shannon information-theoretical inequalities, based on the so-called *copy lemma* [9, 4, 8, 3].

The talk was, however, devoted solely to a limited sub-topic of the manuscript. Specifically, it dealt with an important class of abstract CI frames which are closed under 3 basic operations: copying, marginalization and intersection. Many of standard classes of CI structures, including classic probabilistic CI structures fall into this class of CI frames.

The point is that then, for every variable set N, the family of respective CI models forms a lattice relative to inclusion of models. Therefore, one can apply the results from lattice theory [1] and *formal concept analysis* [5] to describe such lattices of models in terms of the corresponding abstract functional dependence relation [7]. Particular concepts of *pseudo-closed sets* relative to a Moore family and *canonical implicational basis* [6] then offer a kind of simplest standard description in terms of implications among CI statements. Such description can then be interpreted as a kind of "axiomatization" of the respective CI models.

In the talk, the above mentioned concept of an abstract CI frame was introduced, including three basic examples. Substantial related results from lattice theory and the formal concept analysis were then recalled and illustrated by a simple running example.

References

1 G. Birkhoff: *Lattice Theory (Third edition)*. AMS Colloquium Publications 25, American Mathematical Society, Providence 1995.

2 T. Boege, J. H. Bolt, M. Studený: T. Boege, J. H. Bolt, M. Studený: Self-adhesivity in lattices of abstract conditional independence models. A 2024 manuscript available online at `https://arxiv.org/abs/2402/14053v1`.

3 L. Csirmaz: One-adhesive polymatroids. *Kybernetika 56* (2020) 886-902.

4 R. Dougherty, C. Freiling, K. Zeger: Non-Shannon information inequalities in four random variables. A 2011 manuscript available online at `https://arxiv.org/abs/1104.3602v1`.

5 B. Ganter, R. Wille: *Formal Concepts Analysis: Mathematical Foundations*, Springer, Berlin 1999.

6 J.-L. Guigues, V. Duquenne: Familles minimales d'implications in-formatives résultant d'un tableau de données binaires. (in French) *Mathématiques et Sciences Humaines 95* (1986) 5-18.

7 F. Matúš: Abstract functional dependency structures. *Theoretical Computer Science 81* (1991) 117-126.

8 F. Matúš: Adhesivity of polymatroids. *Discrete Mathematics 307* (2007) 2464-2477.

9 Z. Zhang, R.W. Yeung: On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory 44* (1998) 1440–1450.

3.24 Guarded Hybrid Team Logics

Marius Tritschler (TU Darmstadt, DE)

Main reference Marius Tritschler: "Guarded Hybrid Team Logics", in Proc. of the 32nd EACSL Annual Conference on Computer Science Logic, CSL 2024, February 19-23, 2024, Naples, Italy, LIPIcs, Vol. 288, pp. 48:1–48:22, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.
URL http://dx.doi.org/10.4230/LIPICS.CSL.2024.48

Hybrid team logics are inspired by hybrid modal logics where binders can be used to bind (and later reference) assignments. In the team setting, bound teams may be referenced as regular relations. We find that Hybrid Team Logic and its positive and negative fragments are equivalent to well-known team logics. Further, we can analyze guarded versions of these logics to find that guarded hybrid logics share some prominent properties of classical guarded logics, while also overcoming some of the limitations of atom-based guarded team logics.

3.25 Introduction to team semantics

Jonni Virtema (University of Sheffield, GB)

In this talk I will give an introduction to team semantics. I will start by covering motivation, basic definitions, and results from the classical era of team semantics (2007-2017). I will cover the most important aspects first-order team semantics, and briefly discuss how this approach can be adapted to the propositional and modal contexts. In the second half of my talk, I will introduce two of the most important recent advancements in team semantics relevant to this meeting; namely probabilistic team semantics and temporal team semantics.

3.26 Consistent Query Answering with Respect to Primary Keys: Past Research and Future Challenges

Jef Wijsen (University of Mons, BE)

Joint work of Paraschos Koutris, Jef Wijsen
Main reference Paraschos Koutris, Jef Wijsen: "Consistent Query Answering for Primary Keys in Datalog", Theory Comput. Syst., Vol. 65(1), pp. 122–178, 2021.
URL http://dx.doi.org/10.1007/S00224-020-09985-6

We consider database instances that may violate their primary key constraints. A repair of such a database instance is an inclusion-maximal consistent subset of it. For a fixed Boolean query q, the decision problem CERTAINTY(q) takes a database instance as input, and asks whether q holds true in every repair. It is known that for every self-join-free Boolean conjunctive query q, CERTAINTY(q) falls in one of three complexity classes: FO, L-complete, or coNP-complete; furthermore, it can be decided, given q, which of the three cases holds. However, the complexity classification of CERTAINTY(q) for Boolean conjunctive queries q with self-joins remains a notorious open problem. This presentation provides an overview of previous research and the ongoing challenges concerning the study of CERTAINTY(q) for Boolean conjunctive queries.

4 Open problems

4.1 Model checking LTL with team semantics

Martin Zimmermann (Aalborg University, DK)

Joint work of Andreas Krebs, Arne Meier, Jonni Virtema, Martin Zimmermann
Main reference Andreas Krebs, Arne Meier, Jonni Virtema, Martin Zimmermann: "Team Semantics for the Specification and Verification of Hyperproperties", in Proc. of the 43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK, LIPIcs, Vol. 117, pp. 10:1–10:16, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
URL http://dx.doi.org/10.4230/LIPICS.MFCS.2018.10

Linear Temporal Logic (LTL) is arguably the most important specification language for trace properties, properties specifying requirements on individual execution traces of a system. With team semantics, LTL is naturally able to express properties of multiple traces, which allows to express information-flow properties. This extends the expressiveness of plain LTL.

The model-checking problem asks whether a given system satisfies a given specification. For LTL (and hence also LTL under team semantics without splitjunctions), model checking is in PSPACE [1] while it is highly undecidable if one allows Boolean negation [2]. However, the decidability of LTL under team semantics with splitjunctions but without Boolean negation is still unresolved.

In this talk, I will introduce team semantics for LTL, present some useful results, highlight some challenges one has to overcome when trying to solve the problem, and end with a call to arms: let us solve the problem.

References

1 Andreas Krebs, Arne Meier, Jonni Virtema, Martin Zimmermann. *Team Semantics for the Specification and Verification of Hyperproperties*. MFCS 2018, LIPIcs vol. 117, Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

2 Martin Lück. *On the complexity of linear temporal logic with team semantics*. Theor. Comput. Sci. vol. 837.

Participants

- Erika Ábrahám
RWTH Aachen, DE
- Maria Aloni
University of Amsterdam, NL
- Aleksi Ilari Anttila
University of Amsterdam, NL
- Christel Baier
TU Dresden, DE
- Fausto Barbero
University of Helsinki, FI
- Timon Barlag
Leibniz Universität Hannover, DE
- Dario Della Monica
University of Udine, IT
- Arnaud Durand
Paris Cité University, FR
- Fredrik Engström
University of Gothenburg, SE
- Hadar Frenkel
CISPA – Saarbrücken, DE
- Nicolas Fröhlich
Leibniz Universität Hannover, DE
- Pietro Galliani
University of Insubria – Varese, IT
- Erich Grädel
RWTH Aachen, DE
- Jens Gutsfeld
TU Braunschweig, DE
- Matilda Häggblom
University of Helsinki, FI
- Miika Hannula
University of Helsinki, FI
- Peter Harremoës
Niels Brock Copenhagen Business College, DK
- Lauri Hella
Tampere University, FI
- Minna Eveliina Hirvonen
University of Helsinki, FI
- Batya Kenig
Technion – Haifa, IL
- Søren Brinck Knudstorp
University of Amsterdam, NL
- Phokion G. Kolaitis
University of California – Santa Cruz, US
- Juha Kontinen
University of Helsinki, FI
- Cheuk Ting Li
The Chinese University of Hong Kong, HK
- Martin Lück
OHB System – Bremen, DE
- Yasir Mahmood
Universität Paderborn, DE
- Alessio Mansutti
IMDEA Software Institute – Madrid, ES
- Arne Meier
Leibniz Universität Hannover, DE
- Till Miltzow
Utrecht University, NL
- Christoph Ohrem
Universität Münster, DE
- Ana Oliveira da Costa
IST Austria – Klosterneuburg, AT
- Martin Otto
TU Darmstadt, DE
- Nina Pardal
University of Sheffield, GB
- Max Sandström
University of Sheffield, GB
- Milan Studený
The Czech Academy of Sciences – Prague, CZ
- Marius Tritschler
TU Darmstadt, DE
- Jouko Väänänen
University of Helsinki, FI
- Jonni Virtema
University of Sheffield, GB
- Heribert Vollmer
Leibniz Universität Hannover, DE
- Jef Wijsen
University of Mons, BE
- Fan Yang
Utrecht University, NL
- Martin Zimmermann
Aalborg University, DK

Report from Dagstuhl Seminar 24112

EU Cyber Resilience Act: Socio-Technical and Research Challenges

Mila Dalla Preda*[1], Serge Egelman*[2], Anna Maria Mandalari*[3], Volker Stocker†[4], Juan Tapiador†[5], and Narseo Vallina-Rodriguez*[6]

1 **University of Verona, IT.** `mila.dallapreda@univr.it`
2 **ICSI – Berkeley, US.** `egelman@cs.berkeley.edu`
3 **University College London, GB.** `a.mandalari@ucl.ac.uk`
4 **TU-Berlin, DE.** `vstocker@inet.tu-berlin.de`
5 **UC3M – Madrid, ES.** `jestevez@inf.uc3m.es`
6 **IMDEA Networks Institute – Madrid, ES.** `narseo.vallina@imdea.org`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar *"EU Cyber Resilience Act: Socio-Technical and Research Challenges"* (24112). This timely seminar brought together experts in computer science, tech policy, and economics, as well as industry stakeholders, national agencies, and regulators to identify new research challenges posed by the EU Cyber Resilience Act (CRA), a new EU regulation that aims to set essential cybersecurity requirements for digital products to be permissible in the EU market.

The seminar focused on analyzing the proposed text and standards for identifying obstacles in standardization, developer practices, user awareness, and software analysis methods for easing adoption, certification, and enforcement. Seminar participants noted the complexity of designing meaningful cybersecurity regulations and of aligning regulatory requirements with technological advancements, market trends, and vendor incentives, referencing past challenges with GDPR and COPPA adoption and compliance. The seminar also emphasized the importance of regulators, marketplaces, and both mobile and IoT platforms in eliminating malicious and deceptive actors from the market, and promoting transparent security practices from vendors and their software supply chain. The seminar showed the need for multi-disciplinary and collaborative efforts to support the CRA's successful implementation and enhance cybersecurity across the EU.

Seminar March 10–13, 2024 – https://www.dagstuhl.de/24112

2012 ACM Subject Classification Security and privacy → Human and societal aspects of security and privacy

Keywords and phrases Cyber Resilience Act, Software Testing, Software Analysis, IoT, Security Regulations, Security Economics

Digital Object Identifier 10.4230/DagRep.14.3.52

* Editor / Organizer
† Editorial Assistant / Collector

EU Cyber Resilience Act: Socio-Technical and Research Challenges, *Dagstuhl Reports*, Vol. 14, Issue 3, pp. 52–74
Editors: Mila Dalla Preda, Serge Egelman, Anna Maria Mandalari, Volker Stocker, Juan Tapiador, and Narseo Vallina-Rodriguez
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Mila Dalla Preda
Serge Egelman
Anna Maria Mandalari
Narseo Vallina-Rodriguez

Introduction and Motivation

The increasing number of cyberattacks affecting digital products has caused significant security and financial costs to societies. For example, the Mirai attack in 2016 compromised millions of Internet of Things (IoT) devices by exploiting default usernames and passwords, turning them into a botnet army that launched a massive Distributed Denial of Service (DDoS) attack. This attack significantly impacted critical Internet services, causing major outages and disruptions on platforms like Twitter and Netflix [1].

The European Commission has proposed in 2022 the EU Cyber Resilience Act (CRA) to define the legislative framework of essential cybersecurity requirements that product manufacturers must meet when placing any product with digital elements on the internal market, while empowering users to make better security-aware decisions when purchasing and deploying digital products. Following its adoption in 2024, manufacturers will have two years to comply with the new rule, with specific deadlines for different types of products.

The roadmap for CRA adoption follows a multi-phased approach, focusing on high-risk products first and progressively expanding to cover a broader range of digital products over the next few years, aiming to ensure robust cybersecurity standards across the EU. Specifically, during the first year, the focus will be on raising awareness among stakeholders and providing guidance on compliance requirements. The European Commission and national authorities will offer support and resources to help manufacturers understand the new obligations. Then, during the second year, manufacturers and developers will need to ensure that their products meet CRA requirements. This includes implementing necessary security measures, conducting risk assessments, and updating product documentation.

In this scenario, device and software analysis methods – from formal methods to black-box testing – are essential for facilitating compliance at different stages of the product life-cycle, but also for self-attestation and independent verification and certification. However, the rapid evolution and increasing complexity of new technologies and other socio-technical factors such as developers' awareness and incentives for compliance may add further challenges and barriers to adoption.

On the one hand, it is essential to understand whether regulatory requirements are realistic, unambiguous, and whether they are partially misaligned with technology trends, manufacturers' incentives and goals, and with users' privacy and security awareness. For example, research evidence has shown that many developers do not fully comply with the General Data Protection Regulation (GDPR) and the USA Children Online Privacy Protection Act (COPPA) requirements due to their dependency on obscure third-party components for development support and advertising, economic incentives, poor software engineering habits, or even a lack of awareness about the regulations' existence and scope (and hence their compliance obligations). On the other hand, we need to assess to which extent existing device and software analysis methods are fit for aiding developers and manufacturers in assessing compliance, but also for independent certification by third-parties and regulatory enforcement. Yet, current software and device analysis techniques (e.g., black-box testing) often over-simplify the complexity of digital products and present various scalability and coverage limitations that prevent them from reliably auditing and testing whether observed software properties in digital products comply with regulatory requirements.

This Dagstuhl Seminar united a multidisciplinary group of tech and legal academics, industry actors, and policy experts to share their knowledge and experience to collaboratively explore the complex landscape of research and socio-technical challenges for the adoption and enforcement of the CRA. These challenges arise from developer practices and incentives, user awareness, and the feasibility of existing software analysis methods for certification and enforcement.

Seminar Structure

The seminar had a dynamic structure during the 3 days, combining dedicated presentations, panels, and multi-disciplinary working groups to encourage active participation and dialogue between different communities and stakeholders. Arriving on Sunday and starting with a welcome dinner at Schloss Dagstuhl. The three-day seminar activities were structured as follows:

- **Day 1.** The first morning was dedicated to participant introductions, setting common ground on seminar objectives through short elevator pitches by participants, followed by two seminar-like talks and guided discussions. This engaging round of introductions provided a comprehensive overview of the diverse knowledge and skills present in the room, setting the scene for collaborative and constructive discussions. Following these introductions, the seminar continued with an introductory talk by the organizers, a key presentation by Christin Hartung-Kümmerling and Anna Schwendicke from the BSI on the fundamentals, goals, and roadmap of the CRA, and a talk by Vicent Toubina (CNIL) on their experiences with GDPR implementation and enforcement. Following these, participants engaged in open discussions to identify sub-problems of interest. At the end of the first day, participants formed multidisciplinary discussion groups to summarize seminar outputs and a brainstorm session for identifying three key topics for further discussion: (*i*) Understanding and Aiding the Developer Ecosystem; (*ii*) Standardization Efforts; and (*iii*) Tools for Regulatory Enforcement.
- **Day 2.** The second day continued with the interactive group discussions, finalizing with a final all-hands group to consolidate the outputs of the discussions. The day ended with a social activity involving a guided visit to the Völklingen Ironworks, and a dinner in Saarbrücken.
- **Day 3.** The final day involved several all-hands sessions to identify the main outcomes of the seminar, and research challenges for easing CRA adoption and compliance, ensuring continued progress beyond the seminar.

The full seminar agenda is available at: `https://www.dagstuhl.de/24112/schedule.pdf`.

References

1 Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis, et al. Understanding the mirai botnet. In *26th USENIX security symposium (USENIX Security 17)*, pages 1093–1110, 2017.

2 Jukka Ruohonen and Kalle Hjerppe. The gdpr enforcement fines at glance. *Information Systems*, 106:101876, 2022.

3 Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe's transparency and consent framework. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 791–809. IEEE, 2020.

4 Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, Serge Egelman, et al. "won't somebody think of the children?" examining coppa compliance at scale. In *The 18th Privacy Enhancing Technologies Symposium (PETS 2018)*, 2018.

5 Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, 2016.

6 Noura Alomar and Serge Egelman. Developers say the darnedest things: Privacy compliance processes followed by developers of child-directed apps. *Proceedings on Privacy Enhancing Technologies*, 2022.

7 Michael Backes, Sven Bugiel, and Erik Derr. Reliable third-party library detection in android and its security applications. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 356–367, 2016.

8 Andrei Costin, Jonas Zaddach, Aurélien Francillon, and Davide Balzarotti. A {Large-scale} analysis of the security of embedded firmwares. In *23rd USENIX security symposium (USENIX Security 14)*, pages 95–110, 2014.

9 Gianluca Anselmi, Anna Maria Mandalari, Sara Lazzaro, and Vincenzo De Angelis. *COPSEC: Compliance-Oriented IoT Security and Privacy Evaluation Framework*. Association for Computing Machinery, New York, NY, USA, 2023.

10 Aniketh Girish, Tianrui Hu, Vijay Prakash, Daniel J Dubois, Srdjan Matic, Danny Yuxing Huang, Serge Egelman, Joel Reardon, Juan Tapiador, David Choffnes, et al. In the room where it happens: Characterizing local communication and threats in smart homes. In *Proceedings of the 2023 ACM on Internet Measurement Conference*, pages 437–456, 2023.

2 Table of Contents

3 Expected Objectives

Mila Dalla Preda, Serge Egelman, Anna Maria Mandalari, Narseo Vallina-Rodriguez

The goal of this multi-disciplinary seminar on the CRA was to foster comprehensive understanding, collaboration, and strategic planning among stakeholders and research disciplines to aid effective implementation and compliance with the upcoming cybersecurity regulations. Specifically:

1. **Bridging the gap between policy, users and developers.** New tech regulations are often perceived as too late and too difficult to enforce. Regulations can be ambiguous and difficult to interpret and implement by non-legal experts like software developers, even for large companies with legal support [2, 3]. We also note that legal experts may not be able to appropriately capture technical challenges and concepts in the law. Evidence and experience show that even core aspects of the GDPR such as informed consent were interpreted differently by national Data Protection Agencies, thus leading to confusion across developers and facilitating abuse. The EU single market should foster harmonized enforcement across all member states. We wanted to review the legal framework conditions and the research literature to identify shortcomings of existing tech policies and regulations at the compliance and enforcement level. Specifically, we wanted to cover the EU GDPR and EU CRA, but also related international efforts like COPPA and NIST's Cybersecurity Framework in the USA and industry certification frameworks like the IoXT Alliance. This analysis allowed us to assess regulation's aptness and effectiveness at achieving their core objectives, as well as potential barriers for (1) compliance; and (2) both self- and independent certification schemes like certification authorities and regulatory bodies. In fact, one particular discussion of interest was about the effectiveness and shortcomings of self-certification schemes and the processes followed by certification authorities in order to identify procedures and protocols to avoid malicious and deceptive actors from cheating or giving a false sense of compliance [4, 5]. Discussing the legal context from a socio-technical perspective is key to (1) identifying barriers to adoption due to regulations' misalignment with developers' expectations and incentives, and users' preferences (Topic 2), and (2) mapping the requirements and scope of certification frameworks to testing methods for compliance and enforcement (Topic 3).

2. **Understanding development and consumption habits.** Software development practices, industry incentives, and the lack of strict enforcement actions are known barriers to the adoption of the regulation. Additionally, user awareness is key not only to pressure industry actors to comply with regulations but also to pressure policymakers in the development of stricter policies and demanding enforcement actions. Unfortunately, regulatory requirements are often misaligned with developer's development paradigms and incentives [4]. Some developers may not be fully aware of how to comply with the rule or may not be familiar with the principles of privacy- and security-by-default engineering when creating new products [6]. In some cases, developers may introduce harmful components in their programs and products due to their dependency on third-party service providers and libraries (i.e., the supply chain) [7, 8], or they may need to cause privacy harm to enhance the security of their programs (e.g., anti-fraud measures). In this

topic, we presented and discussed developer- and user-studies, metrics and methodologies to understand whether existing software development practices are aligned with regulatory requirements, and if users are aware of their digital rights and the potential threats inherent to the use of connected devices and software.

3. **Technology for compliance, certification, and enforcement.** This block aimed to explore the gap between legal requirements and software analysis. Software analysis and verification methods can play a fundamental role in aiding developers to make their products compliant with regulation, but also in enabling certification and enforcement actions by validating program and device security and privacy properties without any access to device code and specifications using black-box testing methods. We wanted to first evaluate to which extent regulatory requirements can be automatically verified without human involvement, and which ones are ambiguous and open to interpretation. A fundamental concern is about the fitness of current testing methods – proposed by academia as well as by industry – to automatically verify and certify all the properties and security requirements of regulatory frameworks, at scale [9]. This is a complex and hard problem to solve, as there are open research and technical challenges to enable fully automated software testing, even more if it must be done from a regulatory perspective. In fact, most prior work neglects the highly interconnected and complex nature of modern programs, which often interact with neighbouring devices and with their environment [10]. In this seminar, we wanted to integrate the perspective of regulators and cybersecurity agencies, cybersecurity researchers, software engineering researchers, and industry to gather their opinions about how software testing can enable compliance, certification, and enforcement. We wanted to put a special focus on efforts targeting privacy and security analysis of consumer-oriented mobile applications and IoT products, and discuss the applicability, limitations, and strengths of both white- and black-box testing methods for pre- and post-release analysis. We wanted to discuss a research agenda to develop new methodologies that can effectively aid developers at the design, development, and release stages (white- and gray-box testing), and both regulators and certification authorities (black-box testing).

The search for answers to the above technical questions was also intended to help to generally illuminate other orthogonal questions that relate more to the future research agenda in this field and to future policy-making and regulatory enforcement actions.

4 Seminar Participants

Mila Dalla Preda, Serge Egelman, Anna Maria Mandalari, Narseo Vallina-Rodriguez

We designed our seminar structure (discussed in Section 2.5) and the list of invited participants with different backgrounds and expertise to create the right environment for discussing these three intertwined socio-technical topics.

The diverse set of participants covers a broad range of research areas and stakeholders like industry and regulators which are relevant for CRA implementation and enforcement: (i) black-box testing, formal methods, and runtime compliance to help address technical aspects of the EU Cyber Resilience Act; (ii) supply chain analysis, vulnerability detection, and attribution to provide insights into securing complex, multi-component systems; and (iii) human factors, patching, and automatic updates to discuss practical implications for end-users.

Thanks to this multidisciplinary set of participants, the seminar has benefited from a balanced perspective, fostering discussions that bridge technical solutions and policy requirements with research efforts for the adoption and enforcement of the Cyber Resilience Act.

5 Overview of the Talks

This section describes the three talks of day 1.

5.1 Seminar Introduction

On the first day, seminar organizers delivered a talk to introduce the seminar motivation and goals, highlighting the critical importance of the CRA and the research challenges it opens. This talk emphasized the seminar's goal of fostering collaboration and generating a constructive and multi-disciplinary analysis of the EU Cyber Resilience Act and its challenges, leveraging the experience gained with previous regulations such as the EU GDPR and COPPA.

5.2 The EU Cyber Resilience Act

Christin Hartung-Kümmerling (BSI – Freital, DE) and Anna Schwendicke, (BSI – Freital, DE)

In today's world, many products with digital elements are affected by cyberattacks as they lack cybersecurity. The provision of security updates is often inconsistent and insufficient. Additionally, users often do not have the needed access to information that would enable them to choose products that are more cyber-secure. The upcoming Cyber Resilience Act (CRA) therefore addresses these problems as it regulates the market access in form of horizontal European cybersecurity requirements for a broad range of digital products and services. Cybersecurity will be addressed throughout a product's lifecycle – from development until the end of the support period. CRA is part of the New Legislative Framework, a framework meant to improve the internal market by setting up rules for market surveillance and conformity assessment. The CRA extends said framework from safety to security for the first time on a broad basis. It is demanding compliance to security requirements relating to the properties of products with digital elements, extensive information for users, as well as vulnerability management throughout a defined support phase. As security is not a stable state, continuous monitoring is necessary in order to ensure that a product's vulnerabilities are handled in time before they can be used as gateways for cyberattacks. Without knowing a product's contents it is, however, impossible to make any statement regarding its security. As a means to have more clarity about the software components of products, the CRA requires manufacturers to draw up a software bill of materials (SBOM) in order to facilitate their vulnerability handling. The SBOM does not have to be published, but market surveillance authorities can request them. The German Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik) has published technical guidance on SBOM by defining formal and technical requirements, which will help manufacturers to draft one. Once the CRA is adopted, there will be an implementation period in order to set up the necessary infrastructure and to give manufacturers time to prepare. Manufacturers have to report actively exploited vulnerabilities and severe security incidents starting 21 months after the CRA has entered into force and fulfill all other CRA requirements 36 months after the date of entry into force. As certain products require a third-party conformity assessment Member States have to ensure that there are enough notified bodies available to carry out this task. Therefore, each Member State has to have established the required notification infrastructure 18 months after the regulation has entered into force and ought to have enough notified bodies available 24 months after the CRA has come into effect.

5.3 Experiences from GDPR adoption and enforcement

Vincent Toubiana, CNIL – Paris, FR)

The GDPR, implemented in May 2018, will soon turn six years old. In this talk, Vincent Toubiana – Head of LINC (CNIL's Digital Innovation Lab – gave a quick introduction to GPRD enforcement process at CNIL: describing the enforcement chain (complaint, audit/-controls, sanctions) and discussing the challenges that emerged, how they were handled and the success in enforcement. The talk identified key lessons from 6 years of GDPR enforcement that could be applied to the EU Cyber Resilience Act. Several topics, in fact, were discussed: (i) the need and challenges to synchronize and foster collaboration with other DPAs and other authorities that get new responsibilities and competences, (ii) the adaptation to a changing jurisprudence (on concepts such as personal data and what happens when certain cases are elevated to the European Court of Justice), (iii) the estimation of economical impact, which can be intertwined with data protection and market competition challenges, and (iv) technical challenges for easing enforcement and regulatory control in cases such as dark patterns or guidelines for the correct use of web cookies.

6 Breakout Sessions

The identified topics for the three working groups were:

1. Analyzing the developer ecosystem and their incentives for compliance, including communication channels for responsible disclosures and developer obligations towards them and supply chain concerns.
2. The status of existing standardization efforts relevant for CRA compliance.
3. Regulatory compliance and enforcement, including independent assessment and product life-cycle management.

Seminar participants rotated between these breakout groups to better capture their different perspectives and experiences in these three aspects, particularly with regards to the implementation and enforcement of prior tech policies. After each breakout session, we organized a plenary meeting to present the conclusions of the different groups and identify (i) synergies between them, (ii) research challenges and (iii) potential areas for discussion.

6.1 Working Group 1: Developer Ecosystem

The objective of this session was to delve into the intricacies of the developer ecosystem concerning the EU Cyber Resilience Act (CRA). The discussions aimed to identify and address the challenges developers may face in complying with the Act, particularly in areas such as software development practices, vulnerability disclosure, lifecycle management, and secure-by-default standards.

Developer Awareness

During the first year after its implementation, CRA's primary goal is to promote developer awareness. This campaign must be performed at a global scale, as CRA will impact on any manufacturer or software developer targeting the EU market. Although the provisions of CRA might appear vague and high-level (as we will discuss in the context of current standards), their consequences are broad and global. In fact, it is important to first understand whether developers will perceive CRA as a challenge or a barrier that may impact their processes and business. Since the CRA makes a distinction in compliance obligations based on the risk-category/type of product/service and not the size of the firm that offers those, the CRA is asymmetric but only with regards to product type, not in terms of compliance obligations and resulting cost. This could lead to a crowding out of small firms in some contexts (e.g., limited resources and mechanisms to comply). These aspects highlight the need for agencies to start awareness campaigns promptly, drawing from lessons learned from previous regulations like GDPR by CNIL and other EU Data Protection Agencies.

This first awareness stage opens interesting research opportunities to measure the effectiveness of these campaigns to raise developer awareness. It is still unclear which channels and mechanisms regulators will use to effectively raise global awareness. We propose utilizing platforms like mobile and IoT platform app stores and technical development forums (e.g., StackOverflow) for raising awareness and facilitating communication between regulators and developers. Seminar participants emphasized that these efforts must be accompanied by well-grounded and pragmatic standardization processes – still an ongoing process – , to provide comprehensive technical documentation that can facilitate compliance and understanding of current regulatory requirements. By addressing these key areas and questions, regulators and developers can work together to ensure effective compliance with CRA, fostering a more secure and resilient software development ecosystem.

Facilitating Self-Attestation and Testability

The CRA requires vendors to perform self-attestation on certain security properties (e.g., use of cryptographic functions),[1] meaning they must internally assess and document their compliance with the Act's security requirements throughout its life-cycle; i.e., from the designing to the post-release stages of the product. However, these requirements vary depending on the criticality of the product, and they do not seem well-specified. Consequently,

[1] `https://www.enisa.europa.eu/publications/cyber-resilience-act-requirements-standards-mapping`

clear definitions of the responsibilities and liabilities of developers and vendors are essential to avoid uncertainty and ensure compliance but rule requirements are still too high level. One discussion revolved around whether these points may become more concrete as standards get developed. This clarity could be enhanced through collaboration with platforms like GitHub, developer forums and events, and other development repositories, and involving more actively the research community in this process.

The base of secure-by-default assessments raises questions: *Will it be done via checklists or based on a set of standards?* This approach has proven ineffective in prior regulatory contexts such as the USA COPPA rule, even when assisted by organizations forming part of certification schemes [7]. We suggest that we need more process-oriented approaches, focusing on a methodology rather than a simple checklist, similar to the SSDF NIST SP800-218 framework.[2] Creating guidelines and clear procedures is vital for enabling seamless compliance and fostering a secure software development environment. These frameworks, tools, and processes – currently non-existent – could be integrated into Integrated Development Environment (IDE) tools, which would indeed open exciting research and market opportunities. For instance, formal verification tools or tools for automatic Software Bill of Material (SBOM) generation and management could be developed and integrated into IDEs, providing centralized repositories of libraries to help developers manage dependencies effectively. Yet, while self-attestation encourages vendors to proactively identify and fix security issues and comply with regulatory requirements, it is possible that certain vendors may have incentives to avoid compliance and elude external controls. Consequently, this opens an interesting research problem that involves developer studies and empirical approaches to understand developer incentives and the effectiveness of various self-attestation approaches to reduce the number of vulnerable devices in the EU market. st." SSDF NIST SP800-218 (`https://csrc.nist.gov/projects/ssdf`).

Software Supply Chain (SBOM)

A significant amount of seminar time and effort focused on **Software Bill of Materials (SBOMs)**[3] extraction and generation. SBOMs are a key concept in the CRA because they provide a comprehensive inventory of all components and dependencies in a digital product, which is crucial for identifying and managing vulnerabilities. SBOMs enhance transparency and security by allowing regulators and developers to trace, verify, and address potential risks throughout the software supply chain. Therefore, ensuring the accuracy and trustworthiness of SBOMs is essential but, *can SBOMs released by developers and vendors be entirely accurate? Should developers' self-disclosed SBOMs be trusted?*

While the German Federal Office for Information Security (BSI) and other EU national agencies have released guidelines for SBOM generation and management,[4] it is known that vendors and developers face challenges in keeping track of all dependencies when integrating open-source tools and third-party code, as many of their dependencies can be proprietary black-boxes outside their control [5, 3, 6, 8]. Additionally, as several research studies show, developers also struggle at maintaining effectively their dependencies: prior work results show that app developers only slowly adapt new library versions, exposing their end-users to large windows of vulnerability [1].

[2] `https://csrc.nist.gov/pubs/sp/800/218/final`

[3] `https://www.cisa.gov/sbom`

[4] `https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03183/TR-03183_node.html`

This complex scenario is compounded by the high number of supply-chain attacks that can negatively impact the security guarantees of a digital product. However, one discussion point was about the level of information expected to be disclosed by developers about their dependencies, and whether this will be effective at tackling issues such as those described by Backes *et al.* [1]. This opened a discussion about the need for proper labeling [3], and how to include versioning (and customization, as for example in open-source libraries) of dependencies in the SBOMs throughout the whole lifecycle of the product, as libraries and dependencies can experience multiple changes and releases along the supply chain. This may have direct consequences on users' security. For example, *will all developers using open-source tools maintain and send upstream patches to their products to fix vulnerabilities they have found in their products? Will they have the right mechanisms to ensure that all their customers get the patches?*

Establishing a reliable method to gain control over the supply chain and verify the correctness of SBOMs using both black- [5] and gray-box [6] testing techniques is essential, especially when considering the potential lack of developer awareness or in cases of potentially deceptive developers. Microsoft has developed open-source tools for SBOM generation.[7] [8] These tools could be integrated into IDEs to raise awareness and would benefit from a community-built repository of third-party library fingerprints for detection. However, we note that creating such repository is a daunting effort (particularly in terms of maintaining it due to versioning and – in the case of commercial libraries offering analytics and advertising SDKs – due to company merges and acquisitions). Moreover, it is also known that fingerprint-based methods with static analysis methods can easily introduce false positives (e.g., identifying libraries that may not be actually integrated in the code), and false negatives (e.g., not identifying libraries in obfuscated programs), as the research literature in the mobile domain has shown [9, 5]. One challenge would be how to extract SBOMs from software running in the cloud, either partially or entirely as in the case of Amazon Skills [4]. In these scenarios, the platform could do the dependency checks, however there are lambda functions that can be used by deceptive developers to avoid scrutiny.

Shadow libraries and dependencies, i.e., where developers (partially) copy-paste someone's code to take responsibility, might make it difficult to fix critical code as these developers may be using under-resourced/homebrew code without investing in its testing/development/-maintenance.

In fact, many vulnerabilities manifest across connections between chunks of code. Therefore, it is necessary to manage the exposure to responsibilities for these problems. For instance, an incorrect use of methods provided by a cryptographic library can have devastating consequences for software security. Unfortunately, the concept of SBOM may only reveal that a particular library is used but not such development errors.

Participants pointed out that most scenarios will require SBOM extraction and generation from binary files since developers may introduce libraries during compilation time or even compiled products with incomplete SBOMs [2]. There is a risk of an entity not knowing what chain of dependencies are in their software, and automation could reveal unexpected hidden

[5] Black-box testing or closed-box testing is a form of software testing that is performed with no knowledge of a system's internals, and it can be carried out to evaluate the functionality, security, performance, and other aspects of an application.

[6] Gray-box testing is a method you can use to debug software and evaluate vulnerabilities with some but limited knowledge of the workings of the component being tested.

[7] `https://github.com/microsoft/component-detection`

[8] `https://github.com/microsoft/sbom-tool`

risks. Yet, it is an open research challenge to check the correctness of SBOMs from binaries, especially for regulators needing to verify vendors' and developers claims. This may have legal consequences given EU laws prohibiting reverse engineering as we will discuss in §6.1.

Several tooling systems are available to aid this process, such as OWASP CycloneDX[9] and SPDX[10] by the Linux Foundation. It is important to ensure developers do not run scared if there is too little information available, making the compliance process too daunting. CRA may result in more awareness of what elements the code is linking to. A kind of 'dependency amnesty' could encourage those down the chain to provide an SBOM. Otherwise, each developer needs to know all the code they use, even if it's in a library made by someone else. ol and the high number of supply chain attacks that can negatively impact on the security guarantees of a digital product. It is essential establishing a reliable method to gain control over the supply chain and verify the correctness of SBOMs using both gray- and black-box testing techniques as questions arise about when to trust an SBOM disclosed by the developer due to lack of developer awareness or in the case of potentially deceptive developers. Another discussion point was regarding how to do labeling and versioning of SBOMs throughout the whole lifecycle of the product as libraries and dependencies can experience multiple releases.

Effective and Transparent Vulnerability Disclosure Processes

Vulnerability disclosures are critical in the context of the CRA as they ensure that all identified security weaknesses are promptly reported and addressed to the software developer/vendor when identified, thereby reducing the risk of exploitation. By mandating transparent and timely disclosure of vulnerabilities, the CRA aims to foster a culture of accountability and continuous improvement in software security, thus enhancing the resilience of digital products and protecting consumers and businesses from cyber threats.

The research community, including organizations like OWASP,[11] has established several best practices for responsible disclosures to ensure that security vulnerabilities are addressed effectively and ethically. Generally, these practices involve a coordinated process where researchers privately notify the affected vendors about the discovered vulnerabilities, providing them with detailed information and a reasonable time frame to develop and deploy fixes. The standard time frame for fixing a security issue is around 90 days, although this can vary depending on the severity of the vulnerability and the complexity of the fix required and the challenges to demonstrate evidence of in-the-wild exploitation. Researchers are encouraged to maintain confidentiality and offer assistance during this period. In practice, there are many challenges and hidden incentives that often impede proceeding according to these best practices. As discussed in the seminar, vendors and software developers are not always proactive in releasing patches for their products, and researchers often struggle at finding the right communication channel or contact point at a particular vendor. Additionally, defining what constitutes a vulnerability is complex, with debates on whether all vulnerabilities require a CVE [12] and how to handle vendors that do not acknowledge some vulnerabilities (i.e., *"won't fix"*). In fact, the CVE format, while widely used by the security community, is not always the best option. Examples are attacks related to personal data access or privacy concerns.

[9] `https://cyclonedx.org/`

[10] `https://spdx.dev/`

[11] `https://cheatsheetseries.owasp.org/cheatsheets/Vulnerability_Disclosure_Cheat_Sheet.html`

[12] CVE stands for Common Vulnerabilities and Exposures. CVE is a glossary that classifies vulnerabilities. The glossary analyzes vulnerabilities and then uses the Common Vulnerability Scoring System (CVSS) to evaluate the threat level of a vulnerability.

When a developer claims that a reported vulnerability is actually a feature and decides not to fix it, it presents a significant challenge in the responsible disclosure process. This situation requires careful handling to ensure security concerns are addressed without dismissing valid reports. Researchers should document their findings thoroughly and provide a clear explanation of why the identified issue constitutes a vulnerability, including potential risks and impact on users. They can escalate the matter by seeking a second opinion from independent security experts and even openly disclose the vulnerability to raise broader awareness (e.g., through interactions with the press or academic publications). In some cases, involving regulatory bodies or industry standards organizations may be necessary to resolve the dispute and ensure that security vulnerabilities are not overlooked under the guise of being a feature. This approach helps maintain a balance between legitimate security concerns and the developer's design choices, ensuring that user safety remains a priority. CRA articles 32-34 and ENISA's "Good Practices for Vulnerability Disclosure"[13] set minimum requirements for vulnerability disclosures and offer detailed guidance on handling and disclosing vulnerabilities, including scenarios where there might be disagreements on whether an issue is a vulnerability. In the USA, CISA has also defined a new framework for documenting vulnerabilities. It will be important to monitor if CRA will have a positive impact on the responsible disclosure process to ensure that vendors effectively take measures.

Yet, regulators face the challenge of ensuring that every vulnerability is disclosed, clearly defined, and ultimately patched in the products. Moreover, the risk of criminalizing researchers who find vulnerabilities exists: while reverse engineering is usually allowed for achieving interoperability, research, and security analysis, companies can decide to take legal action against security researchers. Additionally, there is the potential exploitation of identified vulnerabilities by authorities as part of this process that could be used for cyberwarfare, such as zero-day exploits. It would be necessary to facilitate mechanisms to increase the transparency of these processes and the interactions between researchers and vendors, hence increasing public awareness on patched and unpatched vulnerabilities.

CRA opens multiple interesting research and industrial opportunities in the context of vulnerability disclosure and product lifecycle management. For example, better defining what a vulnerability constitutes (would privacy threats fall in this context?), finding effective means for balancing intellectual property protection and security – a challenging socio-technical problem in the context of the CRA – , or conducting empirical measurements to see whether CRA has indeed contributed to fix vulnerabilities in digital products. Performing such empirical analysis would be similar in objectives to those measurements showing whether GDPR has contributed to protect users' privacy on the web. Consequently, measuring the impact of regulations in-the-wild is inherently hard to measure as there is a diffusion process of impacts across digital ecosystems and jurisdictions.

Product Life Cycle Management

Product life cycle management is essential to ensure that security measures are maintained throughout the entire lifespan of a product, from development to end-of-life. This continuous oversight helps in (promptly) addressing and patching emerging vulnerabilities and maintaining compliance with security standards and requirements over time. However, product life cycle support may vary significantly depending on manufacturer/developer incentives

[13] https://www.enisa.europa.eu/publications/vulnerability-disclosure

and security standards, the product type (e.g., a smart bulb vs. a smartphone), so the CRA should include clear and specific vendor requirements tailored to different product categories and their inherent threats.[14]

Seminar participants did not observe strict requirements on how vendors should acknowledge vulnerabilities, the confidentiality clauses (as it can be abused by vendors to avoid fixing issues), and the time required for fixing the vulnerability (Annex I states "without delay"), as well as usability requirements to inform users about the security properties of a product and means to reduce threats or patch products. Based on GDPR experience, it is unclear whether vendors will diligently and clearly inform users, and what will be the temporal support of vendors over their products as, for example in the case of smartphone manufacturers, they will have to dedicate engineering teams to support various product lines with specific hardware requirements. This has been identified as another problem that could be better studied through developer studies and large-scale empirical measurements.

6.2 Working Group 2: Standardization Efforts

Standardization efforts are crucial for facilitating the adoption of the CRA provisions. To ensure compliance, seminar participants consider CRA requirements need to be carefully translated into harmonized (yet realistic and clear) standards that manufacturers can adhere to. Yet, this is still an ongoing effort, so drawing conclusions at this stage is hard and potentially premature.

During the first day, the group began discussing the differences between the existing CE standard (safety-oriented) [15] and the requirements necessary for CRA certification (security-oriented). Specifically, the CRA falls under the New Legislative Framework (NLF),[16] which governs market access and surveillance. This framework comprises several modules that vendors can choose from, such as one that checks the product type and another that documents the development process. Insights from the existing CE marking process tell us that documentation must be held and shown to an authority upon request.

A significant challenge with the CRA is that compliance is much more difficult to measure compared to existing CE rules, which cover safety and measurable aspects like voltage and electromagnetic emissions. Participants discussed whether all CRA requirements can indeed be testable and measurable. The same concerns hold for vendors' compliance.

Some interesting research questions arise from this discussion:

- How can all standardization requirements be operationalized and implemented, and then measured and tested? What are the expectations?
- How does a standard get approved as *"CRA compliant"*? This process involves three recognized European standards organizations – CEN,[17] CENELEC,[18] and ETSI[19] – which write and approve standards according to rules where member states are represented.

[14] CRA article 6 states: *"When placing a product with digital elements on the market, and for the expected product lifetime or for a period of five years from the placing of the product on the market, whichever is shorter, manufacturers shall ensure that vulnerabilities of that product are handled effectively and in accordance with the essential requirements set out in Section 2 of Annex I."*

[15] The CE marking stands for Conformité Européene, or European Conformity marking for a range of product regulations.

[16] `https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en`

[17] European Committee for Standardization

[18] European Committee for Electrotechnical Standardization

[19] European Telecommunications Standards Institute

- How do we combine different standards? Looking ahead, we can imagine that we will have a set of established standards, with at least one per sector, possibly leading to competition between standards. These will be complemented by horizontal standards that span multiple sectors.
- How can we re-evaluate standards at regular intervals based on empirical research?
- How do standards adapt to the constant evolution of technology, innovation, threat models and new applications?
- How can we write standards so that compliance is easy, affordable, and automatically verifiable?

ENISA standard

The second day, the group discussed the latest[20] ENISA's standardization guidelines. These aim at identifying relevant existing cybersecurity standards for each CRA requirement, analyzing their scope, and highlighting potential gaps to be addressed. The guidelines help in integrating standards into development processes, ensuring that developers follow secure-by-default principles throughout the product lifecycle.[21]. The file containing the updated version (April 4th 2024, released after the seminar) of the standardization guidelines is available **here**. During the seminar we had access to the previous version of these guidelines and the general observations that we made about the scope and coverage of these standards apply also to the updated version.

As regarding the security guidelines relating to the properties of products with digital elements the group wonders whether there will be tools available to help lower-resourced entities to meet these standards due to asymmetric approaches that may increase the burden to small-size organizations based on the type of product. Moreover, the group collected several observations on the proposed guidelines. For example, the document refers to *"product with digital element"*, this definition should be better specified as in its current form it, for example, also applies to a 1970s radio alarm clock with a digital display. The group also discussed the high level scope and description of requirements, which could be potentially abused by vendors to avoid scrutiny and may not apply to specific types of products and sectors.

The general consensus among participants is that, although the guidelines offer some direction for developing security standards, they still leave several aspects open to interpretation and their scope must be extended to the broad range of digital products, while being flexible enough to catch up any technical development, use-case, and innovation.

Standards and Software Developers

In the latest discussion, the group examined existing standards, including the ETSI *Cyber Security for Consumer Internet of Things: Baseline Requirements* from 2020.[22] Participants noted some outdated recommendations and problematic choices, such as entrusting the threat model to the manufacturer. For example, Provision 5.5-5 states *"device functionality that allows security-relevant changes in configuration via a network interface shall only be accessible after authentication. The exception is for network service protocols that are relied*

[20] As of March 2024
[21] `https://www.cyberresilienceact.eu/the-cyber-resilience-act-annex-eu/`
[22] `https://www.etsi.org/deliver/etsi_en/303600_303699/303645/02.01.01_60/en_303645v020101p.pdf`

upon by the device and where the manufacturer cannot guarantee what configuration will be required for the device to operate. NOTE 3: Protocols that are an exception include ARP, DHCP, DNS, ICMP, and NTP." In this particular example, this standard leaves fundamental network security improvements by ignoring encrypted versions of some of these protocols such as DNS-over-HTTPS (DoH) and the encrypted version of NTP, known as NTPS.

This approach could allow vendors to avoid liability for certain vulnerabilities or claim that patching is difficult due to the device's nature. The group emphasized that standardization efforts must consider developers' software development approaches, needs, and the complexities of the supply chain. They should also be updated more frequently to incorporate newer security measures and protocols, and unknown exploits.

The group also discussed the need for standards for SBOM (Software Bill of Materials) management and clear definitions of what constitutes a vulnerability and guidelines for the disclosure process that may protect security researchers and consumers from potential enterprise interests. In conclusion, the general opinion is that the definition of appropriate standards may be a challenging task due to the complexity of existing technologies and the constantly evolving threat landscape: standards can quickly become obsolete and give final consumers a wrong sense of security.

6.3 Working Group 3: Regulatory Enforcement

This working group focuses on the discussions and insights related to regulatory enforcement within the context of privacy, compliance, and standardization. Prior research experience from seminar participants in relevant projects and research efforts focused on developing technologies for regulatory compliance and testing, and from their daily activities as software developers, regulators and policymakers were essential for informing this discussion.

This working group discussions are divided into three key subsections: (*i*) Security-by-design and Security-by-default, (*ii*) Measuring and Assessing Compliance, and (*iii*) Enforcement and Standardization. Each subsection delves into critical aspects of regulatory enforcement, exploring the challenges, potential solutions, and strategic approaches necessary for effective implementation.

Security-by-design and Security-by-default

The principles of security-by-design and security-by-default are essential in the CRA, which should be effectively captured in upcoming standards. Specific sectoral security requirements must be clearly defined to establish a comprehensive catalogue for certification and also for a complete catalogue of threat models. Participants consider that reference threat models and Key Performance Indicators (KPIs) are critical for facilitating and measuring CRA adoption. Additionally, understanding the costs of regulatory enforcement and vendors' compliance in relation to their available technical, engineering and human resources is vital.

Mandatory technical documentation for compliance verification, along with third-party assessments or endorsements by trusted Certificate Authorities can help to enhance compliance. However, self-certification can be exploited by deceptive actors, particularly if this process does not involve code reviews or software testing as occurred in the case of organizations participating in the U.S. Federal Trade Commission COPPA Safe Harbor program.[23]

[23] https://www.ftc.gov/enforcement/coppa-safe-harbor-program

While attestation at the level of Member States is essential, incentives for data sharing across jurisdictions could improve enforcement and control, particularly benefiting those Member States with less human and technical resources, also reducing the arbitrary decisions that have been observed in enforcing GDPR in certain countries. There are questions about the feasibility and scalability of general auditing tools for devices and whether the Cyber Resilience Act (CRA) provides new tools for enforcement. The duration of attestation and the time required to remove potentially vulnerable devices from the EU market also needs consideration.

The intersection of Intellectual Property (IP) protection laws and the CRA also raises concerns, which may require changes to current intellectual property laws to facilitate reverse engineering for compliance and independent testing. Vendor liability and certification authority accountability, especially if cheating occurs, need examination and consideration. The distribution of responsibilities among supply chain actors for a given service or product, and the exclusion of cloud services from the CRA's scope, focusing solely on device software and hardware, are points for consideration from a research perspective.

To conclude the following questions and observations need to be addressed:

- What are the (broader) economic impacts of CRA enforcement? Which metrics can be used or need to be developed to measure this impact?
- How will the CRA change developers' incentives (e.g., with regard to their R&D efforts)? How will developers perceive the CRA and how will they respond to it? Will they view it as a burden or a beneficial regulation, and will they try to avoid regulatory scrutiny as seen with GDPR?
- Implementation of security-by-design principles in complex devices like IoT.
- Consideration of manufacturer disclaimers to avoid scrutiny and limit liability.
- Development of device and software verification standards and guidelines, both for device manufacturers and for independent certification authorities (and labs) or regulatory bodies. [24]
- Concerns that developers might perceive the CRA as a burden, similar to GDPR, potentially leading to non-cooperation with regulatory investigations, to avoid regulatory scrutiny (e.g., anti-testing) or barriers for small firms to enter the market and innovate, thus potentially stifling innovation by such firms.

Measuring Compliance and Enforcement

In this discussion, group members examined the methods and challenges associated with measuring CRA compliance and standard adoption. The discussion covered various approaches and tools for verification, as well as the potential benefits and drawbacks of different strategies.

Effective measurement of compliance with standard adoption requires the development of testable guidelines for technologies that ease the integration of Continuous Integration (CI) tools for standard verification. Software-based compliance verification is proposed as a potentially more effective alternative to traditional certification authorities, based on check-lists. Additionally, the potential for IoT Industry alliances (e.g., IoxT[25]) and their certifications to satisfy CRA requirements needs careful consideration. Establishing clear guidelines by device categories and addressing the gap between certification, adoption guidelines, and enforcement is crucial for compliance.

[24] https://owasp.org/www-project-iot-security-verification-standard/
[25] https://www.ioxtalliance.org/

Enforcement and standardization are closely intertwined, particularly within the framework of the CRA and its voluntary certification program. Effective regulatory enforcement necessitates the alignment of these efforts with standardization processes, while also addressing the challenges posed by Intellectual Property laws. The complexity of identifying trustworthy certification programs is compounded by varying national accreditation systems and the emerging market for certification schemes. Yet, the creation of standards discussed earlier, also necessitates sufficient resources and authoritative bodies, with the current reliance on industry proposals posing a limitation. The US system, such as NIST's Cybersecurity CMMC[26] could serve as a viable model.

While the tools for CRA enforcement are yet to be determined, they are anticipated to be clarified with the law's approval. However, these regulatory efforts must be aware of the limitations of current software testing methodologies, particularly in terms of scope and scale (not to mention the testability of specific regulatory requirements). Proposals include self-assessment for low-risk products and a EU-wide certification framework for high-risk products, with the European Commission expected to publish a list of high-risk categories. This requires observing these developments and analyzing their alignment with current white- and black-box testing capabilities for device and software security analysis.

The evolving interplay between regulatory enforcement, standardization, and economic impacts underscores the need for precise definitions of enforceable properties, robust evaluation and compliance tools, and comprehensive certification processes. Moving forward, participants believe that the success of the Cyber Resilience Act (CRA) will depend on aligning these efforts with existing software engineering and analysis frameworks, and adapting intellectual property laws like the EU Copyright Act. These are necessary measures to ensure thorough and effective regulatory enforcement.

- How can we measure the compliance of standard adoption and regulatory compliance in various sectors, particularly complex consumer IoT products integrated in multi-party and multi-agent environments?
- What are the differences between the scope of existing guidelines and CRA ultimate goals, and how can guidelines be tailored by device categories (e.g., IoT) or sectors?
- (How) Can the efforts started by IoT industrial alliances (e.g., IoxT), including standardization and certification processes, be leveraged for CRA compliance?
- Can standard and CRA requirements be integrated into CI tools to automate verification prior release?
- How can the balance between intellectual property protection and CRA compliance be managed? Are changes to intellectual property laws necessary to facilitate CRA enforcement and independent verification?
- What are the liability implications for vendors and certification authorities (and labs) if they are found to be non-compliant?
- What role will gatekeeping intermediaries and stakeholders, such as platform operators, e-commerce platforms and software distribution channels, play in removing non-compliant software/products from their platforms?
- What tools or frameworks can complement certification authorities with software-based test cases (black-box) for automated and independent compliance verification?
- What white-box and black-box tools can aid self-assessment for compliance based on risk, particularly for low-risk and high-risk products and the supply chain?
- What principles and methodologies of current certification processes from other regulated markets, such as food safety regulations or aerospace, be adapted to CRA?

[26] https://www.nist.gov/mep/successstories/2020/leading-way-cmmc-compliance

7 Conclusions

In the final plenary meeting, all seminar participants gathered to focus on the main challenges related to standardization, the developer ecosystem, and enforcement of the CRA. Through this discussion, participants collectively identified the research challenges and opportunities described in § 7.1.

We note that addressing these challenges requires collaboration across various disciplines and stakeholders. In fact, a few weeks after this Dagstuhl seminar, the European Commission has updated the CRA requirements, partially addressing some of the concerns raised by the participants.[27]

7.1 Opportunities

Developing and Monitoring the Development of CRA Standards:

Standardization bodies must establish and refine comprehensive standards and guidelines for CRA quickly. These should offer guidance on the scope of the regulation and consider sector-specific or device-specific requirements. The research community must evaluate whether these requirements align with threats and risks identified by the community on consumer-oriented products to effectively protect consumers and identify vendors with deceptive and insecure practices. It is also fundamental to investigate how digital platforms and software distribution platforms can help mitigate the impact of malicious actors through guidelines and publication policies as in the case of the COPPA and GDPR regulations (e.g., collection of unique identifiers and other data types in children-oriented apps).

Informing the Development of Standardization Efforts and Guidelines:

The development of standards and guidelines is fundamental for CRA adoption and compliance. We must have a multi-disciplinary debate to develop these standards and guidelines, and to analyze and discuss the scope of new standards, including the necessity for sector-specific standards. Regulators should actively promote these standards and provide clear, specific guidance on compliance, learning from the adoption and enforcement pitfalls of GDPR: standardization efforts should go hand in hand with tools for assessing compliance by vendors. The research community could inform these efforts and assess their scope and effectiveness, drawing on their research experience with consumer IoT devices and cybersecurity.

Understanding Developer Awareness and Compliance:

It is essential to conduct longitudinal developer-oriented studies (e.g., surveys) to gauge developer awareness, readiness, and incentives for compliance with CRA requirements. To maximize success, these efforts could be done in collaboration with regulatory bodies and digital platforms and marketplaces. We also consider important to encourage contributions to open-source projects by providing incentives, and addressing legal and IP issues to balance security and independent certification, with innovation. In fact, transparency regarding security guarantees (e.g., vulnerability patching) and obligations from vendors to users is crucial. Considerations of usability and incentives are also important (for instance, if someone

[27] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0130_EN.pdf

buys an expensive smart refrigerator that is later found to be vulnerable, do users have incentives to isolate it as it will become just a regular fridge?). What legal expectations are there for such situations? It is important to consider the impact of CRA on the economy, including consumer expectations, market adoption, and international implications.

Creating Methodologies and Tools:

Once standards are in place, it is key to develop methodologies and tools to assist developers with compliance, including tools for SBOM generation and vulnerability management. We also advocate for and support research in formal verification methods to ensure the accuracy and reliability of SBOMs and other compliance measures that developers can use during the design and development of their products. This is especially important for resource-constrained organizations with limited financial resources. Yet, these tools must be complemented with others for independent testers (e.g., regulators, certification authorities, researchers) to facilitate external certification and assessment, recognizing that self-attestation and check-list based approaches may not always be effective. These efforts must investigate the feasibility and scalability of developing a general tool for auditing devices under CRA, and assessing the testeability of certain regulatory requirements. We must not ignore usability considerations and developer incentives for using these tools and maintaining product security throughout their life cycle. While there are companies and research efforts already offering tools for managing SBOMs, the experts expressed concerns regarding the technical challenges of SBOM generation and the aptness of regulatory requirements, which only require developers to disclose high-level dependencies. We consider essential creating effective black-box analysis tools for library version detection as it is critical for pin-pointing specific program vulnerabilities.

Public Outreach and Transparency:

Increase public outreach efforts to enhance transparency regarding security guarantees and mitigations offered by vendors, considering usability and incentives for users. Furthermore, regulators need to actively inform and guide vendors on CRA compliance requirements to avoid the pitfalls experienced with GDPR.

Multi-Disciplinary Longitudinal Analysis:

We consider key to study the overall economic impact of CRA enforcement and develop metrics to evaluate this impact. This will allow us to assess developer perceptions of CRA to understand to which extent it is seen as a burden and in which respects it is seen as a necessary and appropriate regulation by vendors. We recommend performing active scans of the EU's Internet Protocol (IP) address space to monitor the deployment of legacy non-compliant devices with public IP addresses and assess CRA's impact on replacing or isolating them.

Several participants urged for a Systematization of Knowledge (SoK) on CRA, focusing on informing the research community about this new regulation and the identified cross-disciplinary research challenges. Additionally, we considered the need to establish an EU MSCA-ITN (Marie Skłodowska-Curie Actions Innovative Training Network) to train future CRA experts with the necessary multidisciplinary background and skills.

Acknowledgements

We would like to thank and acknowledge the support that the staff of Schloss Dagstuhl has provided us before and after the seminar, as well as for their hospitality. We would also like to mention the active contribution of all the seminar participants, who have helped to ground the implications of a broad regulation with significant and global impact on software development, markets, and cybersecurity, in a scientific research agenda that will need to be addressed in the coming years. The contributions by Narseo Vallina-Rodriguez and Juan Tapiador were possible thanks to the EU H2020 TRUST aWARE project (GA 101021377). The work of Anna Maria Mandalari was supported by the EPSRC PETRAS (EP/S035362/1). The contributions of Simon Parkin were supported by the INTERSCT project, Grant No. NWA.1160.18.301, and RAPID project (Grant No. CS.007), both funded by Netherlands Organisation for Scientific Research (NWO). The contributions by Guillermo Suarez-Tangil were possible thanks to TED2021-132900A-I00 funded by MICIU/AEI /10.13039/501100011033 and the EU NextGenerationEU/ PRTR. The contribution by Volker Stocker was possible thanks to funding from the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII131 (Weizenbaum-Institut für die vernetzte Gesellschaft – Das Deutsche Internet-Institut).

References

1 Michael Backes, Sven Bugiel, and Erik Derr. Reliable third-party library detection in android and its security applications. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 356–367, 2016.

2 Tingting Bi, Boming Xia, Zhenchang Xing, Qinghua Lu, and Liming Zhu. On the way to sboms: Investigating design issues and solutions in practice. *ACM Transactions on Software Engineering and Methodology*, 2023.

3 Peter Caven and L Jean Camp. Towards a more secure ecosystem: Implications for cybersecurity labels and sboms. *Available at SSRN 4527526*, 2023.

4 Jide S Edu, Xavier Ferrer-Aran, Jose Such, and Guillermo Suarez-Tangil. Skillvet: automated traceability analysis of amazon alexa skills. *IEEE Transactions on Dependable and Secure Computing*, 20(1):161–175, 2021.

5 Álvaro Feal, Julien Gamba, Juan Tapiador, Primal Wijesekera, Joel Reardon, Serge Egelman, and Narseo Vallina-Rodriguez. Don't accept candy from strangers: An analysis of third-party mobile sdks. *Data Protection and Privacy: Data Protection and Artificial Intelligence*, 13:1, 2021.

6 Julien Gamba, Mohammed Rashed, Abbas Razaghpanah, Juan Tapiador, and Narseo Vallina-Rodriguez. An analysis of pre-installed android software. In *2020 IEEE symposium on security and privacy (SP)*, pages 1039–1055. IEEE, 2020.

7 Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, Serge Egelman, et al. "won't somebody think of the children?" examining coppa compliance at scale. In *The 18th Privacy Enhancing Technologies Symposium (PETS 2018)*, 2018.

8 Trevor Stalnaker, Nathan Wintersgill, Oscar Chaparro, Massimiliano Di Penta, Daniel M German, and Denys Poshyvanyk. Boms away! inside the minds of stakeholders: A comprehensive study of bills of materials for software systems. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, pages 1–13, 2024.

9 Yafei Wu, Cong Sun, Dongrui Zeng, Gang Tan, Siqi Ma, and Peicheng Wang. {LibScan}: Towards more precise {Third-Party} library identification for android applications. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 3385–3402, 2023.

Participants

- Rainer Böhme
Universität Innsbruck, AT
- Mila Dalla Preda
University of Verona, IT
- Daniel J. Dubois
Northeastern University – Boston, US
- Carolyn Egelman
Google – Mountain View, US
- Serge Egelman
ICSI – Berkeley, US
- Hamed Haddadi
Imperial College London, GB
- Christin Hartung-Kümmerling
BSI – Freital, DE
- François Hublet
ETH Zürich, CH
- Martina Lindorfer
TU Wien, AT
- Anna Maria Mandalari
University College London, GB
- Federica Maria Francesca Paci
University of Verona, IT
- Simon Parkin
Delft University of Technology, NL
- Sergio Pastrana
Carlos III University of Madrid, ES
- Joel Reardon
University of Calgary, CA
- Anna Schwendicke
BSI – Freital, DE
- Ben Stock
CISPA – Saarbrücken, DE
- Volker Stocker
Weizenbaum Institut – Berlin, DE
- Guillermo Suárez-Tangil
IMDEA Networks Institute – Madrid, ES
- Juan Tapiador
Carlos III University of Madrid, ES
- Vincent Toubiana
CNIL – Paris, FR
- Narseo Vallina-Rodriguez
IMDEA Networks Institute – Madrid, ES

Report from Dagstuhl Seminar 24121

Trustworthiness and Responsibility in AI – Causality, Learning, and Verification

Vaishak Belle^{*1}**, Hana Chockler**^{*2}**, Shannon Vallor**^{*3}**, Kush R. Varshney**^{*4}**, Joost Vennekens**^{*5}**, and Sander Beckers**$^{\dagger 6}$

1 **University of Edinburgh, GB.** vaishak@ed.ac.uk
2 **King's College London, GB.** hana.chockler@kcl.ac.uk
3 **University of Edinburgh, GB.** svallor@ed.ac.uk
4 **IBM Research – Yorktown Heights, US.** krvarshn@us.ibm.com
5 **KU Leuven, BE.** joost.vennekens@kuleuven.be
6 **University of Amsterdam, NL.** srekcebrednas@gmail.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24121 "Trustworthiness and Responsibility in AI – Causality, Learning, and Verification". How can we trust autonomous computer-based systems? Since such systems are increasingly being deployed in safety-critical environments while interoperating with humans, this question is rapidly becoming more important. This Dagstuhl Seminar addressed this question by bringing together an interdisciplinary group of researchers from Artificial Intelligence (AI), Machine Learning (ML), Robotics (ROB), hardware and software verification (VER), Software Engineering (SE), and Social Sciences (SS); who provided different and complementary perspectives on responsibility and correctness regarding the design of algorithms, interfaces, and development methodologies in AI.

The purpose of the seminar was to initiate a debate around both theoretical foundations and practical methodologies for a "Trustworthiness & Responsibility in AI" framework that integrates quantifiable responsibility and verifiable correctness into all stages of the software engineering process. Such a framework will allow governance and regulatory practices to be viewed not only as rules and regulations imposed from afar, but instead as an integrative process of dialogue and discovery to understand why an autonomous system might fail and how to help designers and regulators address these through proactive governance.

In particular, we considered how to reason about responsibility, blame, and causal factors affecting the trustworthiness of the system. More practically, we asked what tools we can provide to regulators, verification and validation professionals, and system designers to help them clarify the intent and content of regulations down to a machine interpretable form. While existing regulations are necessarily vague, and dependent on human interpretation, we asked:

How should they now be made precise and quantifiable? What is lost in the process of quantification? How do we address factors that are qualitative in nature, and integrate such concerns in an engineering regime?

In addressing these questions, the seminar benefitted from extensive discussions between AI, ML, ROB, VER, SE, and SS researchers who have experience with ethical, societal, and legal aspects of AI, complex AI systems, software engineering for AI systems, and causal analysis of counterexamples and software faults.

Seminar March 17–22, 2024 – https://www.dagstuhl.de/24121
2012 ACM Subject Classification General and reference → General conference proceedings; Social and professional topics → Codes of ethics
Keywords and phrases responsible AI, trustworthy AI, causal machine learning, autonomous systems
Digital Object Identifier 10.4230/DagRep.14.3.75

* Editor / Organizer
† Editorial Assistant / Collector

Trustworthiness and Responsibility in AI – Causality, Learning, and Verification, *Dagstuhl Reports*, Vol. 14, Issue 3, pp. 75–91
Editors: Vaishak Belle, Hana Chockler, Shannon Vallor, Kush R. Varshney, and Joost Vennekens
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Hana Chockler (King's College London, GB, hana.chockler@kcl.ac.uk)
Shannon Vallor (University of Edinburgh, GB, svallor@ed.ac.uk)
Kush R. Varshney (IBM Research – Yorktown Heights, US, krvarshn@us.ibm.com)
Joost Vennekens (KU Leuven, BE, joost.vennekens@kuleuven.be)

Motivation and research area

How can we trust autonomous computer-based systems? Widely accepted definitions of autonomy take the view of being "independent and having the power to make your own decisions." While many AI systems fit that description, they are often assembled by integrating many heterogenous technologies – including machine learning, symbolic reasoning or optimization – and correspondingly the notion of trust is fragmented and bespoke for the individual communities. However, given that automated systems are increasingly being deployed in safety-critical environments whilst interoperating with humans, a system would not only need to be able to reason about its actions, but a human user would need to additionally externally validate the behavior of the system. This seminar tackled the issue of trustworthiness and responsibility in autonomous systems by considering: notions of cause, responsibility and liability, and tools to verify the behavior of the resulting system.

In the last few years, we have observed increasing contributions in terms of manifestos, position papers, and policy recommendations issued by governments and learned societies, touching on interdisciplinary research involving AI ethics. This has primarily focused on "Fairness, Accountability, and Transparency" (FAT) with a majority focus on fairness, as individual and group fairness seems relatively easier to define precisely. On the other hand, DARPA's XAI agenda has led to a resurgence in diagnostic explanations, but also ignited the question of interpretability and transparency in machine learning models, especially deep learning architectures. Our high-level motivation is that governance and regulatory practices can be viewed not only as rules and regulations imposed from afar but instead as an integrative process of dialogue and discovery to understand why an autonomous system might fail and how to help designers and regulators address these through proactive governance. But before that agenda can be approached, we need to resolve an important low-level question: how can we understand trust and responsibility of the components that make up an AI system? Autonomous systems will make 'mistakes', and accidents will surely happen despite best efforts. How should we reason about responsibility, blame, and causal factors affecting trustworthiness of the system? And if that is considered, what tools can we provide to regulators, verification and validation professionals, and system designers to help them clarify the intent and content of regulations down to a machine interpretable form? Existing regulations are necessarily vague, depending on the nuance of human interpretation for actual implementation. How should they now be made more precise and quantifiable?

The purpose of the seminar was to initiate a debate around these theoretical foundations and practical methodologies with the overall aim of laying the foundations for a "Trustworthiness & Responsibility in AI" framework – a framework for systems development methodology that integrates quantifiable responsibility and verifiable correctness into all stages of the software engineering process. As the challenge, by nature, is multidisciplinary, addressing it must involve experts from different domains, working on creating a coherent, jointly agreed framework. The seminar brought together researchers from Artificial Intelligence (AI), Machine Learning (ML), Robotics (ROB), hardware and software verification (VER), Software Engineering (SE) and the Humanities (HUM), especially Philosophy (PHI), who provided different and complementary perspectives on responsibility and correctness regarding the design of algorithms, interfaces, and development methodologies in AI. From the outset, we wished to especially focus on understanding correctness for AI systems that integrate or utilize data-driven models (i.e., ML models), and to anchor our discussions by appealing to causality (CAU). Causality is widely used in the natural sciences to understand the effect of interventions on observed correlations, allowing scientists to design physical and biological laws. In ML too, increasingly there is recognition that conventional models focus on statistical associations, which can be misleading in critical applications demanding human-understandable explanations. The concept of causality is central to defining a notion of responsibility, and thus was a key point in our discussions.

Directions identified and discussed

The seminar involved extensive discussions between AI, ML, ROB, VER, SE, PHI and HUM researchers who have experience in the following research topics:

- Ethical aspects of AI & ML algorithms: explainability and interpretability in AI algorithms, bias & fairness, accountability, moral responsibility. For example, there were discussions on large language models, their black box nature, and capabilities. There was also quite a bit of work on how explanations and causality might be related. Relevant papers that the participants identified included [10, 1].
- The moral and legal concepts of responsibility that underpin trust in autonomous systems, and how these relate to or can be aided by explainability or causal models of responsibility.
- Technical aspects of AI & ML algorithms: explainability and interpretability in AI algorithms, bias & fairness, accountability, quantification of responsibility. There were discussions regarding how visual input and human-in-the-loop models could provide the next frontier of explainability. Relevant papers identified by the participants included [11].
- Complex AI systems: robotics, reinforcement learning, integrated task and motion planning, mixed-initiative systems. There were discussions that suggest that incorporating high-level specifications from humans could considerably enhance the literature. Examples include recent loss function-based approaches and program induction-related directions for reinforcement policies [5, 4].
- Software engineering for AI systems: development methodologies, specification synthesis, formal verification of ML models, including deep learning architectures, software testing, causality. Outside of a range of recent approaches and looking at verifying the robustness properties of newer networks, there was a discussion on enhancing these perspectives by modeling trust. In fact, what exactly trustworthy machine learning might look like and the components it might involve were also discussed. Examples of relevant work include [9, 12, 8].
- Causal analysis of counterexamples and software faults. Causality was a central topic in the discussion, anchoring some of the key perspectives on how trustworthy AI, as well as explanations, could be addressed along with more nuanced notions such as harm. Following Joseph Halpern's talk on how harm could be formalized and related discussions, a number of relevant papers were identified as promising starting points for causal analysis [2, 3].

- Social aspects of AI & society, AI & law, AI & ethics. Examples of related literature include ideas on the types of ethical robots, the ironies of automation, and the notion of how empathy should apply to explainability among other related topics [7, 6]

Open questions

Discussions between researchers from these different areas of expertise allowed us to explore topics at the intersection between the main areas, and to ask (and obtain partial answers on) the following questions:

- What sorts of explanations, and more generally, correctness notions are users looking for (or may be helpful for them)? How should these be generated and presented?
- How should we reason about responsibility, blame and causal factors affecting trustworthiness in individual components? How should that be expanded to the overall AI system?
- How do we define and quantify trust? Is trust achieved differently depending on the type of the user? Can trust in AI be achieved only using technology, or do we need societal changes?
- How do users reason about and handle responsibility, blame and cause in their day-to-day activities, and how do we interface those concepts with that of the AI system?
- Do our notions of responsibility and explanations increase user's trust in the technology?
- Who are the users of the technology? We envision different types of users, from policy makers and regulators to developers of the technology, to laypeople – the end-users. Should we differentiate the type of analysis for different categories of users?
- What tools can we provide to regulators, verification and validation professionals and system designers to help them clarify the intent and content of regulations down to a machine interpretable form?
- What tools are available to verify ML components, and do they cover the scope of "correct behavior" as understood by users and regulators?
- What SE practices are relevant for interfacing, integrating and challenging the above notions?
- How can properties of AI systems that are of interest be expressed in languages that lend themselves to formal verification or quantitative analysis?
- What kinds of user interfaces are needed to scaffold users to scrutinise the way AI systems operate?
- What frameworks are needed to reason about blame and responsibility in AI systems?
- How do we integrate research in causal structure learning with low-level ML modules used in robotics?
- How do we unify tools from causal reasoning and verification for assessing the correctness of complex AI systems?
- What challenges arise in automated reasoning and verification when considering the above mixed-initiative systems?
- Given a falsification of a specification, what kind of automated diagnosis, proof-theoretic and causal tools are needed to identify problematic components?
- How broadly will counterfactual reasoning (i.e., "what-if" reasoning) be useful to tackle such challenges?

References

1 Lisanne Bainbridge. Ironies of automation. *Automatica*, 19, 1983.

2 Sander Beckers, Hana Chockler, and Joseph Halpern. A causal analysis of harm. *Advances in Neural Information Processing Systems*, 35:2365–2376, 2022.

3 Ilan Beer, Shoham Ben-David, Hana Chockler, Avigail Orni, and Richard Trefler. Explaining counterexamples using causality. *Formal Methods in System Design*, 40:20–40, 2012.

4 Vaishak Belle and Andreas Bueff. Deep inductive logic programming meets reinforcement learning. In *The 39th International Conference on Logic Programming*. Open Publishing Association, 2023.

5 Craig Innes and Subramanian Ramamoorthy. Elaborating on learned demonstrations with temporal logic specifications. *arXiv preprint arXiv:2002.00784*, 2020.

6 William Kidder, Jason D'Cruz, and Kush R Varshney. Empathy and the right to be an exception: What llms can and cannot do. *arXiv preprint arXiv:2401.14523*, 2024.

7 Bran Knowles, Jason D'Cruz, John T. Richards, and Kush R. Varshney. Humble ai. *Commun. ACM*, 66(9):73–79, aug 2023.

8 Ekaterina Komendantskaya and Guy Katz. Towards a certified proof checker for deep neural network verification. *Logic-Based Program Synthesis and Transformation*, page 198.

9 Madsen and Gregor. Measuring human-computer trust. In *11th Australasian Conference on Information Systems*, 2000.

10 James H. Moor. Four types of ethical robot. *Philosophy Now*, 2009.

11 Spinner, Schlegel, Schäfer, and El-Assady. explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE Trans. on Visualization and Computer Graphics*, 2020.

12 Kush R. Varshney. *Trustworthy Machine Learning.*

2 Table of Contents

Executive Summary

Overview of Talks

Summary of Breakout Session on "AI in 20 years: 6 Ambitions"

3 Overview of Talks

3.1 Responsible AI Control

Nadisha-Marie Aliman (Utrecht University, NL)

This talk on "Responsible AI Control" elucidates why when confronted with inconsistent human-level AI/AGI/ASI achievement claims, AI researchers can respond responsibly by rigorously formulating scientific impossibility statements (as has e.g. analogously already been practiced in the Large Hadron Collider Safety case) and developing scientific evaluation frameworks that constrain those achievement claims. For example, related work already introduced diverse AI-related impossibility statements grounded in thermodynamical, biological, cognitive-science-linked and hardware-verification-related explanations. The talk introduces a novel epistemic paradigm termed "cyborgnetic invariance" that entails multiple new impossibility statements. For illustration, a simple new scientific evaluation framework for automated quantity superintelligence achievement claims is discussed. Simply put, the framework extends the tasks of interest for ASI assessment to asymmetrical intelligence/creativity/consciousness levels of civilizations. The cyborgnetic invariance paradigm consists of two postulates: invariance of maximal quantity superintelligence and impossibility of reliable stupidity-based construction. Thereby, asymmetrically measurable intelligence/creativity/-consciousness is non-algorithmic (but it involves physical computation). To build an AGI "from scratch" is at least as hard as physically building a new baby universe. To build such a non-controllable but value-alignable creature, humanity would have to at least first become superintelligent in relation to its current self. In the meantime, one can build controllable but non-value-alignable "AI" tools encapsulated in human-centered units of cyborgnetic control loops to deepen critical thinking and broaden human creativity via so-called artificial EM repeaters, EDM miners and EDE generators in order to tackle global risks. The talk ends by stressing that present-day "AI" should not be underestimated either since its use and misuse is currently linked to an "AI"-related epistemic security threat landscape which subsumes multiple novel global/existential risks for a civilization like present-day humanity.

3.2 Moral Responsibility for AI Systems

Sander Beckers (University of Amsterdam, NL)

Main reference Sander Beckers: "Moral responsibility for AI systems", in Proc. of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., 2024.
URL https://dl.acm.org/doi/10.5555/3666122.3666312

As more and more decisions that have a significant ethical dimension are being outsourced to AI systems, it is important to have a definition of moral responsibility that can be applied to AI systems. Moral responsibility for an outcome of an agent who performs some action is commonly taken to involve both a causal condition and an epistemic condition: the action should cause the outcome, and the agent should have been aware – in some form or other – of the possible moral consequences of their action. In this talk I present a formal definition of both conditions within the framework of causal models. I compare my approach to the existing approaches of Braham and van Hees (BvH) and of Halpern and Kleiman- Weiner (HK). I then generalize my definition into a degree of responsibility.

3.3 Are We Correct To Ascribe Conversational Agency to LLM-Based Chatbots?

Jan M. Broersen (Utrecht University, NL)

To trust AIs and give correct assessments of responsibility in situations where they interact with humans, we need to understand their agency. We need to understand if their agency differs from human agency, and if so, what the differences are. For this talk, I will focus on the conversational agency of LLMs.

3.4 The Simpson and Bias Amplification Paradoxes

Yanai Elazar (AI2 – Seattle, US)

Joint work of Preethi Seshadri, Sameer Singh, Yanai Elazar
Main reference Preethi Seshadri, Sameer Singh, Yanai Elazar: "The Bias Amplification Paradox in Text-to-Image Generation", CoRR, Vol. abs/2308.00755, 2023.
URL https://doi.org/10.48550/ARXIV.2308.00755

The Simpson's paradox (and the Sex Bias in Graduate Admission) is a classic example that illustrates the challenges in evaluation data – originating from the real world or AI models. I will introduce Simpson's paradox briefly and how alternative views of the same data can lead to different conclusions. Then, I will describe our recent work on the Bias Amplification Paradox in the text-to-image models. I argue that bias amplification is highly dependent on the evaluation procedure and sensitive to confounding factors that influence the implications of naive evaluations.

3.5 Trustworthy Autonomy

Michael Fisher (University of Manchester, GB)

Joint work of Dennis, Louise, A., Fisher, Michael
Main reference Louise A. Dennis, Michael Fisher: "Verifiable Autonomous Systems: Using Rational Agents to Provide Assurance about Decisions Made by Machines", Cambridge University Press, 2023.
URL https://doi.org/10.1017/9781108755023

Autonomous Systems have the ability to make their own decisions and potentially to take their own actions, and to do both without direct human intervention. When we deploy these systems, especially in important or even critical situations, do we know what this use of autonomy will result in? And can we trust it to always work "well"?

I discuss issues around the development of Trustworthy Autonomy, including reliability (does it work?), beneficiality (is it working for our benefit?), and the verification of these both before and after deployment.

This will highlight that not only are there distinct forms of AI, each with different benefits and drawbacks, but that combining these in a heterogeneous way can be beneficial. Such combinations are alternatively termed "hybrid" or "neuro-symbolic" systems.

By utilising a specific hybrid “agent” architecture, where our agents are logical and able to represent and implements concepts such as “belief” and “intention”, we are able to expose the reasons for decisions – i.e: “why did you do that”. Furthermore, we can formally verify this agent decision-making to prove whether the agent, and hence the autonomous system, will never choose to do anything “bad”.

This exposure of decision-making processes also has an impact on the broader issues of these autonomous systems, for example around ethical decision-making and responsibility.

References

1 Bremner, Paul, Dennis, Louise A., Fisher, Michael, Winfield, Alan F.T.: On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots. Proceedings of the IEEE pp. 1–21 (2019). https://doi.org/10.1109/JPROC.2019.2898267

2 Chatila, Raja, Dignum, Virginia, Fisher, Michael, Giannotti, Fosca, Morik, Katharina, Russell, Stuart, Yeung, Karen. Trustworthy AI. Pages 13–39 of: Braunschweig, Bertrand, and Ghallab, Malik (eds), *Reflections on Artificial Intelligence for Humanity*. Springer, 2021. https://doi.org/10.1007/978-3-030-69128-8_2

3 Dennis, Louise A., Bentzen, Martin, M., Lindner, Felix, Fisher, Michael: Verifiable Machine Ethics in Changing Contexts. Proceedings of the AAAI Conference on Artificial Intelligence **35**(13), 11470–11478 (May 2021), https://ojs.aaai.org/index.php/AAAI/article/view/17366

4 Dennis, Louise, A., Fisher, Michael, Slavkovik, Marija, Webster, Matthew, P.: Formal Verification of Ethical Choices in Autonomous Systems. Robotics and Autonomous Systems **77**, 1–14 (2016). https://doi.org/10.1016/j.robot.2015.11.012

5 Dennis, Louise, A., Fisher, Michael, Webster, Matthew, P., Bordini, Rafael, H.: Model Checking Agent Programming Languages. Automated Software Engineering **19**(1), 5–63 (2012). https://doi.org/10.1007/S10515-011-0088-X

6 Dennis, Louise, A., Fisher, Michael: Verifiable Autonomous Systems – Using Rational Agents to Provide Assurance about Decisions Made by Machines. Cambridge University Press, 2023. https://doi.org/10.1017/9781108755023

7 Fisher, Michael, Mascardi, Viviana, Rozier, Kristin Yvonne, Schlingloff, Bernd-Holger, Winikoff, Michael, and Yorke-Smith, Neil. Towards a Framework for Certification of Reliable Autonomous Systems. *Autonomous Agents and MultiAgent Systems*, **35**(1), 2021. https://doi.org/10.1007/s10458-020-09487-2

8 Koeman, Vincent, Dennis, Louise, Webster, Matthew, P., Fisher, Michael, Hindriks, Koen. The “Why Did You Do That?” Button: Answering Why-Questions for End Users of Robotic Systems. In *Engineering Multi-Agent Systems*, pages 152–172. Springer, 2020. https://doi.org/10.1007/978-3-030-51417-4_8

3.6 AI Safety and The EU AI Act

Leon Kester (TNO Netherlands – The Hague, NL)

Joint work of Nadisha-Marie Aliman, Leon Kester
Main reference Nadisha-Marie Aliman, Leon Kester: “4. Moral programming”, pp. 63–80, Wageningen Academic, 2022.
URL https://doi.org/10.3920/978-90-8686-922-0_4

Risk Management aiming at harm minimization and systemic risk mitigation is required for Trustworthy AI compatible with the EU AI Act. Moreover, for a meaningful AI control, there is a need for a rigorous harm model such as e.g. via Augmented Utilitarianism to safely encapsulate AI systems in a human-centric socio-technological feedback-loop. In this talk, I also explain why one should not overestimate present-day AI since it is linked to a comprehension bottleneck. For instance, as in science, the ethical value alignment among people can include the creation of new unknown better chains of explanations that present-day AI cannot understand. However, taking the case of so-called "deepfake science attacks" as illustration for a systemic risk, I discuss why one should also not underestimate present-day AI. Here, by way of example, it becomes clear why in a risk-aware approach, instead of asking "was this contribution generated by present-day AI or by a human?" a better suited question would be "does this material encode a better new scientific chain of explanations in comparison to the ones that are already available?". In conclusion, a future risk-aware Trustworthy AI research which is compatible with the EU AI Act should include the AI-aided augmentation of both human critical thinking and human scientific creativity.

References

1 Aliman, Nadisha-Marie and Kester, Leon. 4. Moral Programming. in: *Moral design and technology*, Wageningen Academic, 63–80, 2022.

2 Aliman, Nadisha-Marie and Kester, Leon. VR, Deepfakes and epistemic security. 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), IEEE, 93–98, 2022.

3.7 Specification-based Falsification and Repair of DNN Controllers

Stefan Leue (Universität Konstanz, DE)

Joint work of Fabian Bauer-Marquart, David Boetius, Stefan Leue, Christian Schilling
Main reference Fabian Bauer-Marquart, David Boetius, Stefan Leue, Christian Schilling: "SpecRepair: Counter-Example Guided Safety Repair of Deep Neural Networks", in Proc. of the Model Checking Software – 28th International Symposium, SPIN 2022, Virtual Event, May 21, 2022, Proceedings, Lecture Notes in Computer Science, Vol. 13255, pp. 79–96, Springer, 2022.
URL https://doi.org/10.1007/978-3-031-15077-7_5

I sketch the SpecRepair approach towards specification-based repair of Deep Neural Networks. It implements a counterexample-guided repair approach which includes optimzation-based counterexample finding, counterexample-based retraining of the network and finally the certification of the desired property by a complete DNN verifier.

3.8 Kaspar Causally Explains

Mohammad Reza Mousavi (King's College London, GB)

The Kaspar robot has been used with great success to work as an education and social mediator with children with autism spectrum disorder. Enabling the robot to automatically generate causal explanations is key to enrich the interaction scenarios for children and promote trust in the robot. In our research, we analysed the human-robot interactions in which causal explanations can contribute substantially to the child's understanding of Visual Perspective Taking (VPT). The results helped us identify multiple interaction categories that benefit from causal explanation [3]. Subsequently, we developed a theory of causal explanation to be embedded in Kaspar and built a causal model and an analysis method to calculate causal explanations. We implemented our method to automatically generate causal explanations spoken by Kaspar [2]. We validated our explanations for user satisfaction and brought the robot to a school. The results revealed that children improved their VPT abilities significantly when the robot provided causal explanations [1].

References

1 M. Sarda Gou, G. Lakatos, P. Holthaus, B. Robins, S. Moros, L. Jai Wood, H. Araujo, C. A. E. deGraft-Hanson, M. R. Mousavi, F. Amirabdollahian. Kaspar Explains: The Effect of Causal Explanations on Visual Perspective Taking Skills in Children with Autism Spectrum Disorder. Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), IEEE, 2023.

2 H. Araujo, P. Holthaus, M, Sarda Gou, G, Lakatos, G. Galizia, L. Wood, B. Robins, M.R. Mousavi, and F. Amirabdollahian. Kaspar Causally Explains, Proceedings of the 14th International Conference on Social Robotics, Lecture Notes in Computer Science, Springer, 2022.

3 M. Sarda Gou, G. Lakatos, P. Holthaus, L. Jai Wood, M.R. Mousavi, B. Robins, and F. Amirabdollahian. Towards understanding causality – a retrospective study of using explanations in interactions between a humanoid robot and autistic children. Proceedings of the 31st IEEE International Conference on Robot & Human Interactive Communication (RO-MAN 2022), IEEE, 2022.

3.9 BRIO: A Bias and Risk Assessment Tool

Giuseppe Primiero (University of Milan, IT)

Joint work of Greta Coraglia, Fabio Aurelio D’Asaro, Francesco Antonio Genco, Davide Giannuzzi, Davide Posillipo, Giuseppe Primiero, Christian Quaggio

Main reference Greta Coraglia, Fabio Aurelio D’Asaro, Francesco Antonio Genco, Davide Giannuzzi, Davide Posillipo, Giuseppe Primiero, Christian Quaggio: “BRIOxAlkemy: a Bias Detecting Tool”, in Proc. of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023), Rome, Italy, November 6, 2023, CEUR Workshop Proceedings, Vol. 3615, pp. 44–60, CEUR-WS.org, 2023.

URL https://ceur-ws.org/Vol-3615/paper4.pdf

Phenomena of bias by AI systems based on machine learning methods are well known, and largely discussed in the literature. A variety of tools are being developed to assess these undesirable behaviours. In this short talk I present a bias and risk assessment tool [5] developed within the BRIO Research Project (https://sites.unimi.it/brio/). The tool is based on various formal logics developed in [1, 2, 3, 4]. The tool works on the I/O data of a ML system remaining agnostic on the model itself. The user can choose one of two distinct modules to evaluate either the difference in behaviour that the model displays on outputs produced by subclasses of inputs, or to evaluate against a desirable output. The type of distance and the threshold for admissibile distance from the target distribution can also be selected. The result is a set of all the features and combinations thereof that produce violations with respect to the target distribution. These features can be fed into a risk function which computes an overall value weighting them on parameters such as size of the population and number of features involved, mapping naturally into notions of group and individual fairness.

References

1 F. D'Asaro, G. Primiero, *Probabilistic typed natural deduction for trustworthy computations*, in: Proceedings of the 22nd International Workshop on Trust in Agent Societies (TRUST2021@ AAMAS), 2021.

2 Giuseppe Primiero, Fabio Aurelio D'Asaro: Proof-checking Bias in Labeling Methods. BEWARE@AI*IA 2022: 9-19

3 Fabio Aurelio D'Asaro, Giuseppe Primiero: Checking Trustworthiness of Probabilistic Computations in a Typed Natural Deduction System. CoRR abs/2206.12934 (2022)

4 Francesco A. Genco, Giuseppe Primiero: A Typed Lambda-Calculus for Establishing Trust in Probabilistic Programs. CoRR abs/2302.00958 (2023)

5 Greta Coraglia, Fabio Aurelio D'Asaro, Francesco A. Genco, Davide Giannuzzi, Davide Posillipo, Giuseppe Primiero and Christian Quaggio, *BRIOxAlkemy: A Bias detecting tool*, in Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023), pp. 44–60. 2024.

3.10 Trustworthiness for Medical Diagnostics: What and How?

Ajitha Rajan (University of Edinburgh, GB)

The rapidly advancing field of Explainable Artificial Intelligence (XAI) aims to tackle the issue of trust regarding the use of complex black-box deep learning models in real-world applications. Existing post-hoc XAI techniques have recently been shown to have poor performance on medical data, producing unreliable explanations which are infeasible for clinical use. To address this, we propose an ante-hoc approach based on concept bottleneck models that introduces for the first time clinical concepts into the classification pipeline, allowing the user valuable insight into the decision-making process. On a large public dataset of chest X-rays and associated medical reports, we focus on the binary classification task of lung cancer detection. Our approach yields improved classification performance on lung cancer detection when compared to baseline deep learning models (F1 > 0.9), while also generating clinically relevant and more reliable explanations than existing techniques. We evaluate our approach against post-hoc image XAI techniques LIME and SHAP, as well as CXR-LLaVA, a recent textual XAI tool that operates in the context of question answering on chest X-rays.

3.11 Some Challenges on the Path to Certifying AI-Enabled Autonomy

Subramanian Ramamoorthy (University of Edinburgh, GB)

URL https://web.inf.ed.ac.uk/tas

The increasing use of AI in autonomous systems has made the problem of certifying such systems hard. While the difficulties are associated with broad questions of AI safety, AI-enabled autonomous systems raise certain uniquely challenging questions. This includes the problem of characterising the dynamic behaviour of adaptive systems in open and human-centred environments. This talk surveys work done within the UKRI Research Node on Trustworthy Autonomous Systems Governance and Regulation (https://web.inf.ed.ac.uk/tas), with a focus on the AV certification case study. Within this, we outline results from work on specification gaps [1], scenario generation and sampling with multiple representations [2], [3], and active learning methods for risk-sensitive design [4].

References

1 Abeywickrama DB, Bennaceur A, Chance G, Demiris Y, Kordoni A, Levine M, Moffat L, Moreau L, Mousavi MR, Nuseibeh B, Ramamoorthy S. On specifying for trustworthiness. Communications of the ACM. 2023 Dec 21;67(1):98-109.

2 Innes C, Ramamoorthy S. Testing rare downstream safety violations via upstream adaptive sampling of perception error models. In 2023 IEEE International Conference on Robotics and Automation (ICRA) 2023 May 29 (pp. 12744-12750).

3 Innes C, Ireland A, Lin Y, Ramamoorthy S. Anticipating Accidents through Reasoned Simulation. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems 2023 Jul 11 (pp. 1-11).

4 Corso A, Katz S, Innes C, Du X, Ramamoorthy S, Kochenderfer MJ. Risk-driven design of perception systems. Advances in Neural Information Processing Systems. 2022 Dec 6;35:9894-906.

3.12 A Causal Analysis of Harm

Joseph Halpern (Cornell University – Ithaca, US)

Joint work of Sander Beckers, Hana Chockler, Joseph Halpern

It has proved notoriously difficult to define harm. Indeed, it has been claimed that the notion of harm is a "Frankensteinian jumble" that should be replaced by other well-behaved notions. On the other hand, harm has become increasingly important as concerns about the potential harms that may be caused by AI systems grow. For example, the European Union's draft AI act mentions "harm" over 25 times and points out that, given its crucial role, it must be defined carefully.

I start by defining a qualitative notion of harm that uses causal models and is based on a well-known definition of actual causality. The key features of the definition are that it is based on contrastive causation and uses a default utility to which the utility of actual outcomes is compared. I show that our definition is able to handle the problematic examples from the literature. I extend the definition to a quantitative notion of harm, first in the case of a single individual, and then for groups of individuals. I show that the "obvious" way of doing this (just taking the expected harm for an individual and then summing the expected harm over all individuals) can lead to counterintuitive or inappropriate answers, and discuss alternatives, drawing on work from the decision-theory literature.

References

1 Beckers, S., Chockler, H., and Halpern, J.Y. A Causal Analysis of Harm, Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), pp. 2365–2376, 2022. `https://dl.acm.org/doi/abs/10.5555/3600270.3600442`

2 Beckers, S., Chockler, H., and Halpern, J.Y. Quantifying harm, Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023), pp. 363–371. 2023. https://www.ijcai.org/proceedings/2023/0041

3.13 AI Governance and Agential Power: How Can We Make Systems Answer?

Shannon Vallor (University of Edinburgh, GB)

Joint work of Shannon Vallor and Bhargavi Ganesh
Main reference Shannon Vallor, Bhargavi Ganesh: "Artificial intelligence and the imperative of responsibility: Reconceiving AI governance as social care". In M. Kiener (ed.), The Routledge Handbook of Philosophy of Responsibility (Routledge: NY), p. 395-406, 2023
URL https://doi.org/10.4324/9781003282242

The accelerating development of artificial intelligence (AI) systems has generated acute and interlinked challenges for social trust, responsibility ascription, and governance. While today's AI tools lack the type of agency that can bear responsibility, they are deployed in ways that create novel configurations and social appearances of agential power. That is, they allow new things to be done by us, for us, and to us, in ways that do not easily fit our existing practices for governing moral and legal responsibility. This is commonly referred to as the problem of AI "responsibility gaps".

We confront this challenge by framing normative responsibility for AI actions in a new way: not as a metaphysical fact about agents to be discovered, nor a set of criteria that responsible agents must fully satisfy, but as a set of constructed social practices in the exercise of agential power, that make agential powers answerable for their impact on others' vulnerabilities and interests. The construction and use of such practices for new or changed agential powers is an essential precondition of social trust.

Drawing from historical examples in steamboat engineering, consumer finance, and environmental governance, we highlight how responsibility gaps have generated the moral and political imperative to construct new forms of responsible agency and governance to balance novel agential powers, of which AI is merely the latest iteration. We conclude with observations about the two general classes of available AI governance strategies, agential obligation and agential constraint, that must be balanced in order to secure public trust in AI technologies that represent new agential powers.

4 Summary of Breakout Session on "AI in 20 years: 6 Ambitions"

4.1 AI Broadens Out

AI needs to broaden out in two directions; within AI *and* without. Within AI, it needs to be taught that learning AI requires more than just machine learning; other techniques are needed to complement ML and to continue growth and exploration beyond the existing paradigm.

AI must also broaden to incorporate necessary knowledge from other fields: philosophy, law, neuroscience, HCI and design, for example. AI researchers will increasingly need critical thinking skillsets that take them beyond technical work and allow for better evaluation of AI methods and applications.

Suitably broadened, AI itself needs to be a fundamental "core" area of knowledge for university graduates; in the future, understanding how the world works will be impossible without some understanding of AI and its uses. Could AI Studies be a core educational requirement? Something like this has been tried in the Netherlands before, using a broad interdisciplinary model (for example, in Utrecht in the 90s).

In 20 years time, we hope to see first year interdisciplinary courses like "Intro to AI" that all students can take – but how to overcome institutional and disciplinary resistance/tradition? Some countries are very resistant to curriculum change, and this would require retraining of academic staff in universities across disciplines. How can we make this kind of change possible? We can learn from the successes and failures of other interdisciplinary studies created in the last 50 years: Science and Technology Studies, Environmental Studies, Bioethics.

4.2 Knowing *How* to Use AI vs Knowing *About* AI

Both are going to be essential. We might see AI trade schools in 20 years, to teach the areas that create new, attractive, well-paying jobs without needing theoretical foundations. Potential career paths include:

- AI User Specialist (domain specific)
- AI Data Quality Officer
- AI Prompt Engineer
- AI Error and Bias Controller
- AI Ombudsperson
- AI UI Specialist

4.3 Greater Professionalisation of the AI Community

Professionalisation and accreditation can be mechanisms to prescribe certain educational requirements and also diversify the field into a more balanced set of specializations. We might also consider the "Nuclear Option": using licensing/certification of AI Professionals for safety-critical industries and applications, in the way that we have seen work in medicine and certain areas of engineering. We do this in medicine and many areas of engineering because they are highly dangerous professions as well as beneficial ones. AI is now *also* a highly dangerous (and beneficial) profession.

4.4 Effective and Balanced AI Regulation

Regulation can advance AI further in a number of ways, beyond just making AI safer for people to engage with. It can serve as another incentive for broadening the field of AI – as with privacy regulation, it can require fulfillment of certain roles and create incentives for corporations to invest in more types of AI expertise. Regulators and professional societies might be able to coordinate incentives strategically if not captured by industry. For example, testing and licensing could be an incentive embedded in procurement standards, liability caps, etc. for safety-critical AI development or application. Regulation could help drive the acquisition and normalization of these areas of expertise and more:

- AI Ethics
- AI Law and Policy
- AI Security
- AI Safety
- AI Privacy
- AI Auditor

We note that If no one is willing to take responsibility for an AI system in a high-stakes environment, it arguably should not be deployed in that domain – the burden needs to be on organisations to demonstrate that they have assigned specific and adequate duties to competent, empowered and accountable professional(s).

4.5 Standardisation of Responsible AI Design and Development Practices

With a more professionalised and well-governed AI ecosystem, we expect to see better ways to standardise the conversion of responsible policy choices into design and engineering choices – right now that falls on AI Developers that aren't trained to formalize values and either aren't doing it or end up doing it poorly.

Professionalisation and standardisation of ethical design principles and processes will also shield individual professionals from being unfairly held personally accountable for unavoidable harms/failures; AI Developers today are disincentivized to make explicit moral choices, for which they will then be personally on the hook if the outcome isn't ideal. No technology can be made risk-free and we need to shield developers from liability or at least cap their liability for making responsible choices that follow best professional practice.

4.6 A Mature and Collaborative AI Culture

In 20 years we hope to see AI research, learner and practitioner environments that embody openness to interdisciplinarity, effective translation, co-construction and communication of AI knowledge, and intellectual charity. We can start early by moving the learning of AI to earlier phases, before the "hard/soft" skills division (which itself must be challenged and rethought in the next decades), beyond STEM/not STEM, so that AI is marked by a culture of shared intellectual community rather than knowledge hoarding and turf-defending.

Participants

- Nadisha-Marie Aliman
  Utrecht University – NL
- Emma Beauxis-Aussalet
  VU Amsterdam – NL
- Sander Beckers
  University of Amsterdam – NL
- Vaishak Belle
  University of Edinburgh – GB
- Jan M. Broersen
  Utrecht University – NL
- Georgiana Caltais
  University of Twente – Enschede, NL
- Hana Chockler
  King's College London – GB
- Jens Claßen
  Roskilde University – DK
- Sjur K. Dyrkolbotn
  West. Norway Univ. of Applied Sciences – Bergen, NO
- Yanai Elazar
  AI2 – Seattle, US
- Esra Erdem
  Sabanci University – Istanbul, TR
- Michael Fisher
  University of Manchester – GB
- Sarah Alice Gaggl
  TU Dresden – DE
- Leilani H. Gilpin
  University of California – Santa Cruz, US
- Gregor Goessler
  INRIA – Grenoble, FR
- Joseph Y. Halpern
  Cornell University – Ithaca, US
- Till Hofmann
  RWTH Aachen University – DE
- David Jensen
  University of Massachusetts – Amherst, US
- Leon Kester
  TNO Netherlands – The Hague, NL
- Ekaterina Komendantskaya
  Heriot-Watt University – Edinburgh, GB
- Stefan Leue
  Universität Konstanz – DE
- Joshua Loftus
  London School of Economics and Political Science – GB
- Mohammad Reza Mousavi
  King's College London – GB
- Giuseppe Primiero
  University of Milan – IT
- Ajitha Rajan
  University of Edinburgh – GB
- Subramanian Ramamoorthy
  University of Edinburgh – GB
- Kilian Rückschloß
  LMU München – DE
- Judith Simon
  Universität Hamburg – DE
- Luke Stark
  University of Western Ontario – London, CA
- Daniel Susser
  Cornell University – Ithaca, US
- Shannon Vallor
  University of Edinburgh – GB
- Kush R. Varshney
  IBM Research – Yorktown Heights, US
- Joost Vennekens
  KU Leuven – BE
- Felix Weitkämper
  LMU München – DE

Report from Dagstuhl Seminar 24122

Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications

Dmitry Kobak[*1], **Fred A. Hamprecht**[*2], **Smita Krishnaswamy**[*3], **Gal Mishne**[*4], **and Sebastian Damrich**[†5]

1 **Universität Tübingen, DE.** dmitry.kobak@uni-tuebingen.de
2 **Universität Heidelberg, DE.** fred.hamprecht@iwr.uni-heidelberg.de
3 **Yale University – New Haven, US.** smita.krishnaswamy@yale.edu
4 **University of California, San Diego – La Jolla, US.** gmishne@ucsd.edu
5 **Universität Tübingen, DE.** sebastian.damrich@uni-tuebingen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar "Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications" (24122). Low-dimensional embeddings are widely used for unsupervised data exploration across many scientific fields, from single-cell biology to artificial intelligence. These fields routinely deal with high-dimensional characterization of millions of objects, and the data often contain rich structure with hierarchically organized clusters, progressions, and manifolds. Researchers increasingly use 2D embeddings (t-SNE, UMAP, autoencoders, etc.) to get an intuitive understanding of their data and to generate scientific hypotheses or follow-up analysis plans. With so many scientific insights hinging on these visualizations, it becomes urgent to examine the current state of these techniques mathematically and algorithmically.

This Dagstuhl Seminar brought together machine learning researchers working on algorithm development, mathematicians interested in provable guarantees, and practitioners applying embedding methods in biology, chemistry, humanities, social science, etc. The aim of the seminar was to (i) survey the state of the art; (ii) identify critical shortcomings of existing methods; (iii) brainstorm ideas for the next generation of methods; and (iv) forge collaborations to help make these a reality.

Seminar March 17–22, 2024 – https://www.dagstuhl.de/24122
2012 ACM Subject Classification Computing methodologies → Machine learning
Keywords and phrases dimensionality reduction, high-dimensional, visualization
Digital Object Identifier 10.4230/DagRep.14.3.92

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Low-Dimensional Embeddings of High-Dimensional Data: Algorithms and Applications, *Dagstuhl Reports*, Vol. 14, Issue 3, pp. 92–115
Editors: Dmitry Kobak„ Fred Hamprecht, Smita Krishnaswamy, and Gal Mishne
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

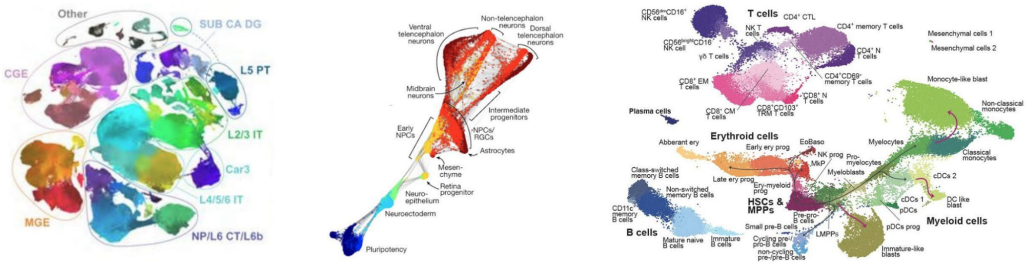


Figure 1 Example applications in single-cell transcriptomics. Left: cortical neurons [8], sample size $n = 1.2$M. Middle: human brain organoid development [9], $n = 43$K. Right: human blood and bone marrow cells in leukaemia [10], $n = 70$K. Figures from original publications.

1 Executive Summary

Dmitry Kobak (Universität Tübingen, DE)
Sebastian Damrich (Universität Tübingen, DE)
Fred A. Hamprecht (Universität Heidelberg, DE)
Smita Krishnaswamy (Yale University – New Haven, US)
Gal Mishne (University of California, San Diego – La Jolla, US)

2D embeddings in science

In recent years, high-dimensional "big" data have become commonplace in multiple academic fields. To give some examples, single-cell transcriptomics routinely produces datasets with sample sizes in hundreds of thousands and dimensionality in tens of thousands [1]; single-cell mass spectrometry deals with millions of samples [2]; genomic datasets quantifying single-nucleotide polymorphisms can deal with many millions of features [3]; behavioural physiology produces high-dimensional datasets with tens of thousands of samples [4]. In neuroscience, calcium imaging allows to record time-series activity of thousands of neurons. Many scientific fields that traditionally did not have to deal with high-dimensional data analysis now face similar challenges; for example, a digital library can yield a dataset with tens of millions of samples and hundreds, if not millions, of features [5].

Such datasets require adequate computational methods for data analysis, including unsupervised data exploration. In fact, exploratory statistical analysis has become an essential tool in many scientific disciplines, allowing researchers to compactly visualise, represent and make sense of their data. It became commonplace to explore low-dimensional embeddings of the data, generated by methods like t-SNE [6] or UMAP [7]. Such visualisation has proven to be a valuable tool for exploring the data, performing quality control, and generating scientific hypotheses (Figure 1).

Similar algorithms are also applied in artificial intelligence research to visualise massive datasets used to train state-of-the-art artificial intelligence models, such as image-based and text-based generative models. This allows researchers to discover biases and gaps in the data, to highlight model limitations, and ultimately to develop better models (Figure 2). A concise overview of the model's training data can also be helpful for societal oversight and public communication.

Neighbour embedding methods like t-SNE and UMAP create a low-dimensional map of the data based on the k-nearest-neighbour graph. As a result, they are often unable to reproduce large-scale global structure of the data [12], creating potentially misleading

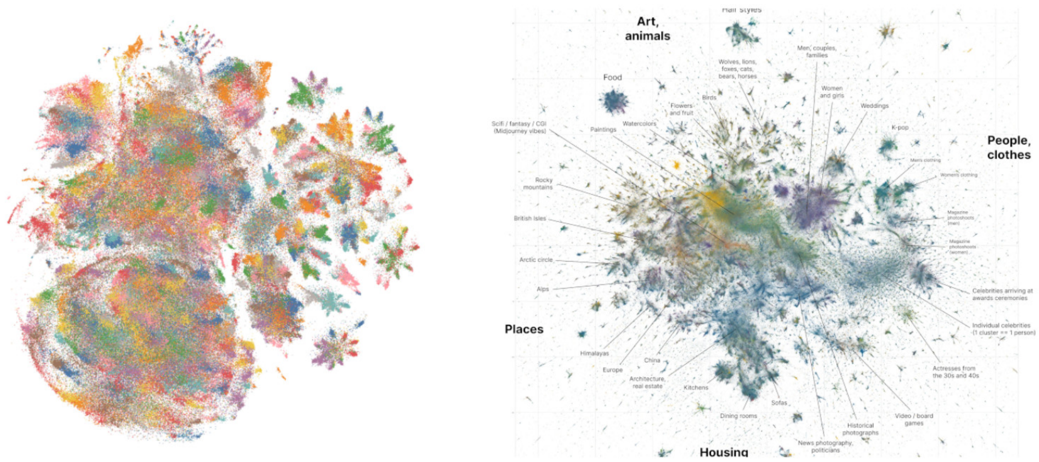


Figure 2 Example applications in artificial intelligence. Left: GPT4All-J training data [11], n = 800K. Right: image captions from LAION-Aesthetics dataset (figure by Dadid McClure), n = 12M.

visualizations [13]. Acquisition of increasingly high-dimensional data across scientific fields has sparked widespread interest in employing dimensionality reduction and visualisation methods. However, there is a gap between method developers who propose and implement these algorithms, and domain experts who aim to use them. The purpose of this seminar was to bring together machine learning researchers, theoreticians, and practitioners, to address current gaps in theoretical guarantees and evaluation measures for state-of-the-art approaches, highlight practical challenges in applying these techniques in different domains, brainstorm the solutions, and set up new collaborations to tackle open problems in this vibrant field.

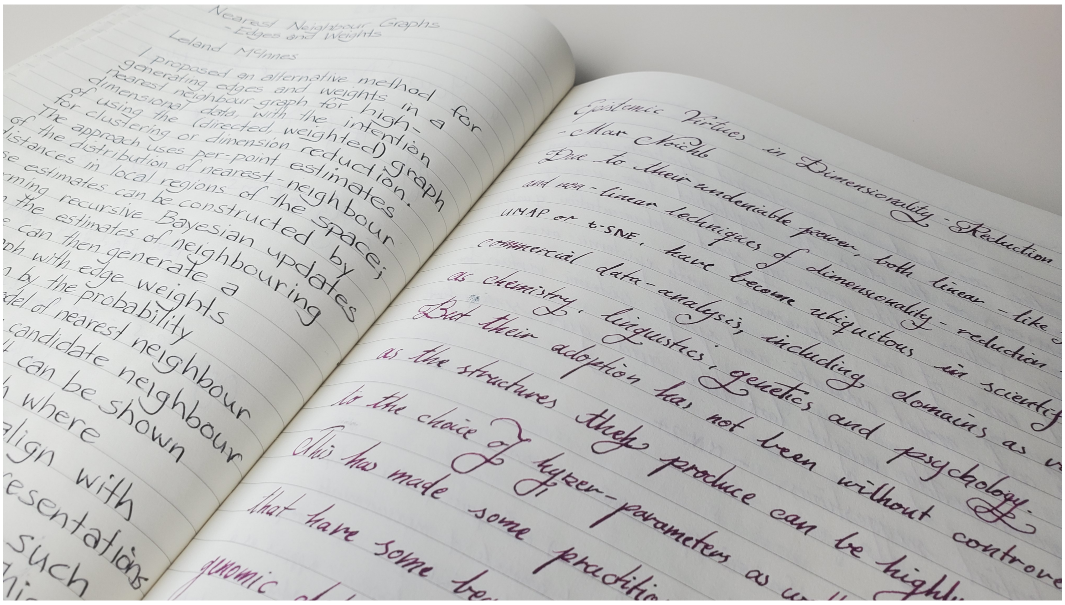
Seminar topics

The overarching purpose of this Dagstuhl Seminar was to brainstorm open problems and challenges in the field of low-dimensional embeddings, as seen by (i) practitioners; (ii) theoreticians and mathematicians; and (iii) machine learning researchers — leading to new collaborations to tackle these problems. The seminar focused on the following open questions, grouped into four areas.

Low-dimensional embeddings in actual practice

Single-cell biology, working with large quantities of high-dimensional data and interested in exploratory research, became a field heavily relying on low-dimensional embeddings. But embeddings of texts [5], of genomes [14], of graph nodes [15], of chemical structures [16], etc., are also rapidly gaining popularity. Seminar participants discussed and brainstormed which fields in the coming years are likely to generate data amenable for embedding methods, and compared challenges raised by each of these application fields.

Neighbour embeddings have a number of well-known limitations [12]: for example, they can strongly distort the global structure of the data and are unable to represent high-dimensional topological features of the data. These artefacts can lead practitioners to incorrect scientific conclusions or to chasing unfounded hypotheses. We extensively discussed



Nearest Neighbour Graphs
- Edges and Weights
Leland McInnes
I proposed an alternative method for
generating edges and weights in a
nearest neighbour graph for high-
dimensional data, with the intention
of using the (directed, weighted) graph
for clustering or dimension reduction.
The approach uses per-point estimates
of the distribution of nearest-neighbour
distances in local regions of the space;
se estimates can be constructed by
rming recursive Bayesian updates
n the estimates of neighbouring
e can then generate a
ph with edge weights
n by the probability
del of nearest neighbour
candidate neighbour
t can be shown
n where
align with
esentations
such
hi

Epistemic Virtues in Dimensionality-Reduction
- Max Noichl
Due to their undeniable power, both linear - like
and non-linear techniques of dimensionality-reduction
UMAP or t-SNE, have become ubiquitous in scientif
commercial data-analysis, including domains as
as chemistry, linguistics, genetics and psychology.
But their adoption has not been without controve
as the structures they produce can be highl
to the choice of hyper-parameters as w
This has made some practitio
that have some be
genomic d

Figure 3 Handwritten abstracts from our seminar.

(i) which limitations can be addressed by the new generation of algorithms; (ii) how to diagnose misleading aspects of any given embedding; and (iii) what evaluation metrics are necessary and sufficient for comparing different visualisation techniques.

Moreover, two-dimensional embeddings have been recently criticised as being dangerously misleading [13]. At the same time, they are widely used across many disciplines and can be helpful in actual scientific practice, if used with care [12]. In several talks and multiple discussions, seminar participants talked about specific examples of how and where the embeddings are useful, and which best practices can help to avoid them being misleading.

Common themes across state-of-the-art algorithms and relevant trade-offs

One common theme in multiple talks and discussions was trade-offs between various embedding algorithms.

First, methods like t-SNE or UMAP are typically used to produce 2D or 3D embeddings, while spectral methods like Laplacian eigenmaps [17] produce low-dimensional embeddings that are often used with more embedding dimensions. This is less suitable for visualisation but may be better suited for downstream data analysis. Several seminar participants reported successfully applying UMAP to intermediate dimensionality too, with particular benefits for downstream density-based clustering (using HDBSCAN algorithm).

Second, all neighbour embedding algorithms operate on the kNN graph of the data but use different loss functions and different attractive/repulsive forces to arrive at the final layout. This yields various trade-offs in the quality of global/local structure preservation [18].

Third, neighbour embedding algorithms are typically run on a kNN graph constructed using pairwise Euclidean distances, but in principle any other metric can be used as well. Specifically, metric design can be useful for incorporating domain knowledge and statistical priors on the data [19, 20]. We discussed what kinds of data can profit from using non-Euclidean distance metrics, or from kNN graph post-processing, such as diffusion-based smoothing.

Fourth, more generally, neighbour embeddings are known to be related to the self-supervised learning approach known as contrastive learning [21]. However, despite substantial progress in each of these two fields, they stayed largely disconnected from each other. Seminar participants argued that both contrastive learning and neighbour embedding research can benefit from each other's state-of-the-art approaches, and in particular can be combined to develop new algorithms for visualising textual and/or graph-based data.

Fifth, while neighbour embeddings only aim to preserve nearest neighbours, methods based on MDS aim to preserve all pairwise distances including the large ones. In Isomap [22] and PHATE [23], pairwise distances are obtained as graph distances on the kNN graph. Isomap uses short path distance, while PHATE uses diffusion-based distance called potential distance. LDLE [24] uses bottom-up manifold learning to align low-distortion local embeddings to a global embedding. We discussed to what extent these approaches can capture both the local and the global structure of the data, and what the advantages and the disadvantages of aiming to preserve global aspects of the data are.

Interactive embeddings

Another extensively discussed topic was interactive visualizations of 2D embeddings (in particular see abstracts by Benjamin M. Schmidt and B. P. F. Lelieveldt). While most often low-dimensional embeddings are depicted as static images, they can be powerful tools for *interactive* data explorers. NomicAI has been developing software for in-browser interactive explorers, while the group of B. P. F. Lelieveldt has been working on stand-alone software for interactive explorerd of RNA-sequencing data.

Perspective paper

During the seminar, participants decided to work together on a perspective paper, provisionally titled like the seminar: "Low-dimensional embeddings of high-dimensional data". During the seminar, we organized several brainstorming sessions on what should the paper cover and how the material should be organized. The writing is currently underway and we hope to be able to release the work some time in summer 2024.

References

1 Kobak, Dmitry and Berens, Philipp *The art of using t-SNE for single-cell transcriptomics*, Nature Communications, 10(1): 1–14, 2019.

2 Belkina, Anna C and Ciccolella, Christopher O and Anno, Rina and others *Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets*, Nature Communications, 10(1): 5415, 2019.

3 Diaz-Papkovich, Alex and Anderson-Trocmé, Luke and Ben-Eghan, Chief and Gravel, Simon *UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts*, PLoS genetics, 15(11): e1008432, 2019.

4 Kollmorgen, Sepp and Hahnloser, Richard HR and Mante, Valerio *Nearest neighbours reveal fast and slow components of motor learning*, Nature, 577(7791): 526–530, 2020.

5 Schmidt, Benjamin *Stable random projection: Lightweight, general-purpose dimensionality reduction for digitized libraries*, Journal of Cultural Analytics, 3(1), 2018.

6 van der Maaten, Laurens and Hinton, Geoffrey *Visualizing data using t-SNE*, Journal of Machine Learning Research, 9(11), 2008.

7 McInnes, Leland and Healy, John and Melville, James *UMAP: Uniform manifold approximation and projection for dimension reduction*, arXiv preprint arXiv:1802.03426, 2018.

8 Yao, Zizhen and Van Velthoven, Cindy TJ and Nguyen, Thuc Nghi and others *A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation*, Cell, 184(12): 3222–3241, 2021.

9 Kanton, Sabina and Boyle, Michael James and He, Zhisong and others *Organoid single-cell genomic atlas uncovers human-specific features of brain development*, Nature, 2019.

10 Triana, Sergio and Vonficht, Dominik and Jopp-Saile, Lea and others *Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states*, Nature Immunology, 22(12): 1577–1589, 2021.

11 Anand, Yuvanesh and Nussbaum, Zach and Treat, Adam and other *GPT4All: An Ecosystem of Open Source Compressed Language Models*, arXiv preprint arXiv:2311.04931, 2023.

12 Wattenberg, Martin and Viégas, Fernanda and Johnson, Ian *How to Use t-SNE Effectively*, Distill, 2016, 10.23915/distill.00002.

13 Chari, Tara and Pachter, Lior *The specious art of single-cell genomics*, PLoS Computational Biology, 19(8): e1011288, 2023.

14 Diaz-Papkovich, Alex and Anderson-Trocmé, Luke and Gravel, Simon *A review of UMAP in population genetics*, Journal of Human Genetics, 66(1): 85–91, 2021.

15 Hu, Yifan *Efficient, high-quality force-directed graph drawing*, Mathematica journal, 10(1): 37–71, 2005.

16 Probst, Daniel and Reymond, Jean-Louis *Visualization of very large high-dimensional data sets as minimum spanning trees*, Journal of Cheminformatics, 12(1): 12, 2020.

17 Belkin, Mikhail and Niyogi, Partha *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Computation, 15(6): 1373–1396, 2003.

18 Böhm, Jan Niklas and Berens, Philipp and Kobak, Dmitry *Attraction-Repulsion Spectrum in Neighbor Embeddings*, Journal of Machine Learning Research, 23(95), 2022.

19 Talmon, Ronen and Coifman, Ronald R *Empirical intrinsic geometry for nonlinear modeling and time series filtering*, Proceedings of the National Academy of Sciences, 110(31): 12535–12540, 2013.

20 Mishne, Gal and Talmon, Ronen and Meir, Ron and others *Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery*, IEEE Journal of Selected Topics in Signal Processing, 10(7): 1238–1253, 2016.

21 Damrich, Sebastian and Böhm, Niklas and Hamprecht, Fred A and Kobak, Dmitry *From t-SNE to UMAP with contrastive learning*, In *The Eleventh International Conference on Learning Representations*, 2023.

22 Tenenbaum, Joshua B and Silva, Vin de and Langford, John C *A global geometric framework for nonlinear dimensionality reduction*, Science, 290(5500): 2319–2323, 2000.

23 Moon, Kevin R and Van Dijk, David and Wang, Zheng and others *Visualizing structure and transitions in high-dimensional biological data*, Nature Biotechnology, 37(12): 1482–1492, 2019.

24 Kohli, Dhruv and Cloninger, Alexander and Mishne, Gal *LDLE: Low distortion local eigenmaps*, Journal of machine learning research, 22(282): 1–64, 2021.

2 Table of Contents

3 Overview of Talks

3.1 RNA Velocity Embeddings in Curved Spaces

Michael Bleher (Universität Heidelberg, DE)

RNA velocity data provides a snapshot of cell states and their current rate of change. It promises insights into the behaviour of individual cells and the dynamics governing cell division and differentiation processes. To explore single cell RNA-sequence data one often relies on low-dimensional visualizations, e.g. tSNE or UMAP. A priori it is not obvious how RNA velocities carry over to such representations and current methods have several drawbacks.

It was recently suggested that one should fix a biologically motivated, low-dimensional manifold and infer RNA velocities strictly in terms of an embedding of the data in that manifold. I expand on that idea and argue that low-dimensional representations of position-velocity pairs should utilize the Sasakian geometry on the tangent bundle of curved target spaces. Moreover, I propose that non-linear neighbour embeddings into low- or middle-dimensional symmetric spaces provide a geometric representation of the principal dynamical components in the data. This geometrization provides interesting future directions regarding the analysis of the dynamical processes captured in single cell data.

3.2 Dimensionality Reduction for Scientific Machine Learning – First Steps towards Task-driven Mechanistic Model Reduction

Kerstin Bunte (University of Groningen, NL)

Joint work of Kerstin Bunte, Peter Tiňo, Elisa Oostwal, Janis Norden, Michael Chappell
Main reference Yuan Shen, Peter Tino, Krasimira Tsaneva-Atanasova: "Classification framework for partially observed dynamical systems", Phys. Rev. E, Vol. 95, p. 043303, American Physical Society, 2017.
URL https://doi.org/10.1103/PhysRevE.95.043303
Main reference Kerstin Bunte, David J Smith, Michael J Chappell, Zaki Hassan-Smith, Jeremy W Tomlinson, Wiebke Arlt, Peter Tiňo: "Learning pharmacokinetic models for in vivo glucocorticoid activation". J Theor Biol. 2018 Oct 14;455:222-231. Epub 2018 Jul 23. PMID: 30048717.
URL https://doi.org/10.1016/j.jtbi.2018.07.025

Nowadays, most successful machine learning (ML) techniques for the analysis of complex interdisciplinary data use significant amounts of measurements as input to a statistical system. The domain expert knowledge is often only used in data preprocessing. The subsequently trained technique appears as a "black box", which is difficult to interpret and rarely allows insight into the underlying natural process. Especially in critical domains such as medicine and engineering, the analysis of dynamic data in the form of sequences and time series is often difficult. Due to natural or cost limitations and ethical considerations data is often irregularly and sparsely sampled and the underlying dynamic system is complex. Therefore, domain experts currently enter a time-consuming and laborious cycle of mechanistic model construction and simulation, often without direct use of the experimental data or the task at hand. We now combine the predictive power of ML and the explanatory power of mechanistic models. Therefore we perform learning in the space of dynamic models that represent the complex underlying natural processes, with potentially very few measurements. We use principles of dimensionality reduction, such as subspace learning, to determine relevant areas in the parameter space of the underlying model as a first step to achieve task-driven model reduction.

3.3 Tree-based Dimensionality Reduction and Clustering

Miguel Á. Carreira-Perpiñán (University of California – Merced, US)

URL http://faculty.ucmerced.edu/mcarreira-perpinan/research/TAO.html

I describe recent work about tree-structured dimensionality reduction, with applications to interpretability, fast training and inference, and scalability to large datasets. This relies on learning optimal sparse oblique decision trees, which have hyperplane splits using few features (rather than the traditional single-feature splits). I make connections to methods ranging from PCA to autoencoders to t-SNE, and extensions to clustering and other topics.

3.4 Mapping the Embedding Multiverse

Corinna Coupette (MPI für Informatik – Saarbrücken, DE)

Joint work of Jeremy Wayland, Corinna Coupette, Bastian Rieck
Main reference Jeremy Wayland, Corinna Coupette, Bastian Rieck: "Mapping the Multiverse of Latent Representations", CoRR, Vol. abs/2402.01514, 2024.
URL https://doi.org/10.48550/ARXIV.2402.01514

Echoing recent calls to counter reliability and robustness concerns in machine learning via multiverse analysis, we present PRESTO, a principled framework for mapping the multiverse of machine-learning models that rely on latent representations. Although such models enjoy widespread adoption, the variability in their embeddings remains poorly understood, resulting in unnecessary complexity and untrustworthy representations. Our framework uses persistent homology to characterize the latent spaces arising from different combinations of diverse machine-learning methods, (hyper)parameter configurations, and datasets, allowing us to measure their pairwise (dis)similarity and statistically reason about their distributions. As we demonstrate both theoretically and empirically, our pipeline preserves desirable properties of collections of latent representations, and it can be leveraged to perform sensitivity analysis, detect anomalous embeddings, or efficiently and effectively navigate hyperparameter search spaces.

3.5 Detecting the Topology of High-dimensional Data with Spectral Methods

Sebastian Damrich (Universität Tübingen, DE)

Joint work of Sebastian Damrich, Philipp Berens, Dmitry Kobak
Main reference Sebastian Damrich, Philipp Berens, Dmitry Kobak: "Persistent homology for high-dimensional data based on spectral methods", CoRR, Vol. abs/2311.03087, 2023.
URL https://doi.org/10.48550/ARXIV.2311.03087

Persistent homology is a popular computational tool for finding the global shape (topology) of point clouds, such as the presence of loops or voids. However, many real-world datasets with low intrinsic dimensionality reside in an ambient space of much higher dimensionality. We show that in this case traditional persistent homology becomes very sensitive to noise and fails to detect the correct topology. The same holds true for existing refinements of persistent homology. As a remedy, we find that spectral distances, such as diffusion distance and effective resistance, allow persistent homology to detect the correct topology even in the presence of high-dimensional noise. Finally, we apply these methods to high-dimensional single-cell RNA-sequencing data.

3.6 Neighbor Embedding Algorithms: Missing Data, Fast Multiscale Approaches, and Interpretability

Cyril de Bodt (University of Louvain, BE)

Joint work of Cyril de Bodt, Dounia Mulders, Michel Verleysen, Pierre Lambert, Edouard Couplet
Main reference Cyril de Bodt, Dounia Mulders, Michel Verleysen, John Aldo Lee: "Fast Multiscale Neighbor Embedding", IEEE Transactions on Neural Networks and Learning Systems, Vol. 33(4), pp. 1546–1560, 2022.
URL https://doi.org/10.1109/TNNLS.2020.3042807

Dimensionality reduction (DR) aims at computing relevant low-dimensional (LD) representations of high-dimensional (HD) data sets, mainly for exploratory visualization. Different paradigms have emerged to formalize mappings from HD to LD coordinates, e.g., through the reproduction of distances or neighborhoods. In the data visualization context, neighbor embedding (NE) algorithms, such as stochastic neighbor embedding (SNE) and variants (t-SNE, UMAP, etc.), reach outstanding DR performance compared to older techniques.

After quickly introducing the field of dimensionality reduction for data visualization and NE algorithms in particular, this talk will summarize three lines of projects recently explored in our lab:

- The visualization of databases with missing entries [1];
- The acceleration of multiscale NE schemes, which aim at better preserving both local and global HD structures in LD embeddings [2];
- The interpretability of NE algorithms, through the design of both post-hoc techniques and natively interpretable methods [3, 4].

References

1 de Bodt, Cyril and Mulders, Dounia and Verleysen, Michel and Lee, John Aldo *Nonlinear dimensionality reduction with missing data using parametric multiple imputations*, IEEE Transactions on Neural Networks and Learning Systems, 30(4): 1166–1179, 2018. `https://ieeexplore.ieee.org/abstract/document/8447227`

2 De Bodt, Cyril and Mulders, Dounia and Verleysen, Michel and Lee, John Aldo *Fast multiscale neighbor embedding*, IEEE Transactions on Neural Networks and Learning Systems, 33(4): 1546–1560, 2020. `https://ieeexplore.ieee.org/abstract/document/9308987`

3 Lambert, Pierre and Marion, Rebecca and Albert, Julien and others *Globally local and fast explanations of t-SNE-like nonlinear embeddings*, 2022. `https://dial.uclouvain.be/pr/boreal/object/boreal:265533`

4 Couplet, Edouard and Lambert, Pierre and Verleysen, Michel and others *Natively Interpretable t-SNE*, 2023. `https://dial.uclouvain.be/pr/boreal/object/boreal:279549`

3.7 What is a Population? Insights from Topological Analysis of Biobank Data

Alex Diaz-Papkovich (Brown University – Providence, US)

Joint work of Alex Diaz-Papkovich, Shadi Zabad, Chief Ben-Eghan, Luke Anderson-Trocmé, Georgette Femerling, Vikram Nathan, Jenisha Patel, Simon Gravel
Main reference Alex Diaz-Papkovich, Shadi Zabad, Chief Ben-Eghan, Luke Anderson-Trocmé, Georgette Femerling, Vikram Nathan, Jenisha Patel, Simon Gravel: "Topological stratification of continuous genetic variation in large biobanks", bioRxiv, Cold Spring Harbor Laboratory, 2023.
URL https://doi.org/10.1101/2023.07.06.548007

Population genetics methods necessarily rely on some definition of a population for analysis. Many methods exist, and most either model a discrete number of populations and their mixtures or define an archetype of a population and fit data to that. Alternatively, we can use density clustering after having processed the data with UMAP specifically parametrized for clustering. Using this approach, we can visualize and study biobank data from a genetic perspective, allowing us to better understand the complexity of the gene-geography-environment relationship, explore potential analyses, and ultimately learn much more about the data upon which so many analyses are based.

3.8 Compound-SNE for Comparative Alignment of Multiple t-SNEs & Eco-velo for RNA-velocity estimation

Laleh Haghverdi (Max-Delbrück-Centrum – Berlin, DE)

Main reference Colin G. Cess, Laleh Haghverdi: "Compound-SNE: Comparative alignment of t-SNEs for multiple single-cell omics data visualisation", bioRxiv, Cold Spring Harbor Laboratory, 2024.
URL https://doi.org/10.1101/2024.02.29.582536
Main reference Valérie Marot-Lassauzaie, Brigitte Joanne Bouman, Fearghal Declan Donaghy, Yasmin Demerdash, Marieke Alida Gertruda Essers, Laleh Haghverdi: "Towards reliable quantification of cell state velocities", PLoS Comput. Biol., Vol. 18(9), p. 1010031, 2022.
URL https://doi.org/10.1371/JOURNAL.PCBI.1010031

One of the first steps in single-cell omics data analysis is visualization, which allows researchers to see how well-separated cell-types are from each other. In order to improve visual comparisons between large numbers of samples, we introduce Compound-SNE, which performs what we term a soft alignment of samples in embedding space. We show that Compound-SNE is able to align cell-types in embedding space across samples and data modalities, while preserving local embedding structures from when samples are embedded independently. I also talked about application of the Nostrum projection method for visualisation of RNA-velocities, as well as our cost-efficient Eco-velo approach, which skips the current unreliable gene-by gene parameter fitting approaches for velocity estimation.

3.9 Using Embeddings in the Social Sciences: Examples and Open Problems

Ágnes Horvát (Northwestern University – Evanston, US)

Joint work of Ágnes Horvát, Daniel Romero, Hao Peng, Stasa Milojevic, Orsolya Vasarhelyi, Igor Zakhlebin, Sohyeon Hwang, Katherine O'Toole, Henry Dambanemuya, Brian Uzzi
Main reference Hao Peng, Daniel M. Romero, Emoke-Ágnes Horvát: "Dynamics of Cross-Platform Attention to Retracted Papers: Pervasiveness, Audience Skepticism, and Timing of Retractions", CoRR, Vol. abs/2110.07798, 2021.
URL https://arxiv.org/abs/2110.07798

In recent years, there has been an explosion of interest in quantitative methods that rely on low-dimensional embeddings for pattern extraction and visualization. Social scientists increasingly recognize that these techniques open up new methodological opportunities. This brief talk presented examples from our work relying on digital trace data to understand online science communication [1, 4, 3, 2], musical creativity [5], and capital allocation [6], highlighting the challenges where social science applications need further methodological development.

References

1 Peng, H., Romero, D. & Horvát, E. Dynamics of cross-platform attention to retracted papers. *Proceedings Of The National Academy Of Sciences.* **119**, e2119086119 (2022)

2 Hwang, S., Horvát, E. & Romero, D. Information Retention in the Multi-Platform Sharing of Science. *Proceedings Of The Seventeenth International AAAI Conference On Web And Social Media, ICWSM 2023, June 5-8, 2023, Limassol, Cyprus.* pp. 375-386 (2023), https://doi.org/10.1609/icwsm.v17i1.22153

3 Vásárhelyi, O., Zakhlebin, I., Milojević, S. & Horvát, E. Gender inequities in the online dissemination of scholars' work. *Proceedings Of The National Academy Of Sciences.* **118** (2021)

4 Zakhlebin, I. & Horvát, E. Diffusion of Scientific Articles across Online Platforms. *Proc. Int. AAAI Conf. Web. Soc. Media.* **14**, 762-773 (2020,5), https://ojs.aaai.org/index.php/ICWSM/article/view/7341

5 O'Toole, K. & Horvát, E. Extending human creativity with AI. *Journal Of Creativity.* **34**, 100080 (2024), https://www.sciencedirect.com/science/article/pii/S2713374524000062

6 Horvát, E., Dambanemuya, H., Uparna, J. & Uzzi, B. Hidden Indicators of Collective Intelligence in Crowdfunding. *Proceedings Of The ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 – 4 May 2023.* pp. 3806-3815 (2023), https://doi.org/10.1145/3543507.3583414

3.10 Neighbour Embeddings Meet Contrastive Learning

Dmitry Kobak (Universität Tübingen, DE)

Main reference Jan Niklas Böhm, Philipp Berens, Dmitry Kobak: "Attraction-Repulsion Spectrum in Neighbor Embeddings", Journal of Machine Learning Research, Vol. 23(95), pp. 1–32, 2022.
URL http://jmlr.org/papers/v23/21-0055.html

In recent years, neighbor embedding methods like t-SNE and UMAP have become widely used across several application fields, in particular in single-cell biology. Given this attention, it is very important to understand possibilities, shortcomings, and trade-offs of neighbor embedding methods. In this talk, I present our recent work on the attraction-repulsion spectrum of neighbor embeddings and the involved trade-offs [1, 2]. I also explain how neighbor embeddings are related to contrastive learning, a popular framework for self-supervised learning of image data. This leads to our recent work on contrastive visualizations of image datasets (t-SimCNE) [3].

References

1 Böhm, Jan Niklas and Berens, Philipp and Kobak, Dmitry *Attraction-Repulsion Spectrum in Neighbor Embeddings*, Journal of Machine Learning Research, 23(95), 2022.

2 Damrich, Sebastian and Böhm, Niklas and Hamprecht, Fred A and Kobak, Dmitry *From t-SNE to UMAP with contrastive learning*, In *The Eleventh International Conference on Learning Representations*, 2023.

3 Böhm, Niklas and Berens, Philipp and Kobak, Dmitry *Unsupervised visualization of image datasets using contrastive learning*, In *The Eleventh International Conference on Learning Representations*, 2023.

3.11 Tear and Repulsion Enabled Registration of Point Clouds for Manifold Learning

Dhruv Kohli (University of California – San Diego, US)

Joint work of Dhruv Kohli, Gal Mishne, Alex Cloninger

We present a framework for aligning the local views of a possibly closed/non-orientable data manifold to produce an embedding in its intrinsic dimension through tearing. Through a spectral coloring scheme, we render the embeddings of the points across the tear with matching colors, enabling a visual recovery of the topology of the data manifold. The embedding is further equipped with a tear-aware metric that enables computation of shortest paths while accounting for the tear. To measure the quality of an embedding, we propose two Lipschitz-type notions of global distortion—a stronger and a weaker one—along with their pointwise counterparts for a finer assessment of the embedding. Subsequently, we bound them using the distortion of the local views and the alignment error between them. We show that our theoretical result on strong distortion leads to a new perspective on the need for a repulsion term in manifold learning objectives. As a result, we enhance our alignment approach by incorporating repulsion. Finally, we compare various strategies for the tear and repulsion enabled alignment, with regard to their speed of convergence and the quality of the embeddings produced.

3.12 Heat Diffusion Distances, Manifold Embeddings and Geodesics

Smita Krishnaswamy (Yale University – New Haven, US)

Here we explore the connection between heat diffusion on data and recovery of manifold or more intrinsic distances in data for low dimensional embeddings and dimensionality reduction. The main approach here is to view the data as a graph over which random walks or heat diffusion are conducted to discover distances "through" the data between points and then embed them in low dimensions. We introduce the idea of random walk based distance, which is a feature of diffusion maps and our PHATE method. With the latter using an M-divergence between data (discrete) diffusion probabilities. Next we introduce the heat kernel which involves exponential powers of the graph laplacian, which can be used to discover a more generalized multiscale distance and preservation options which weigh near and far distances under different schema to create a continuum between neighbor preservation embeddings (like SNE) and global embeddings like PHATE. Finally we showed how to use these distances to regularize autoencoders whose latent spaces can then be used for population flows and discovery of dynamics from static snapshot data via our Neural FIM and Mioflow frameworks.

3.13 Unsupervised Dimensionality Reduction: Multi-Scale Methods & Quality Assessment

John Aldo Lee (UC Louvain-la-Neuve, BE) and Cyril de Bodt (University of Louvain, BE)

Joint work of John A. Lee, Cyril de Bodt, Pierre Lambert, Edouard Couplet, Michel Verleysen
Main reference John A. Lee, Diego H. Peluffo-Ordóñez, Michel Verleysen: "Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure", Neurocomputing, Vol. 169, pp. 246–261, 2015.
URL https://doi.org/10.1016/j.neucom.2014.12.095

Since 2008, methods of neighbor embedding (NE) have gained much popularity and have outperformed mostly all other paradigms of dimensionality reduction DR. A method like t-SNE yields results that clearly outperform stress-based multidimensional scaling (MDS) for instance. However, NE is known to be a local method, preserving small neighborhoods, whereas MDS is more of a global method, keeping the global data arrangement. This work is interested in developing NE methods that are local and global, as well as quality criteria to evaluate them. Multi-scale NE can be achieved by using entropic affinities by browsing a range of neighborhood sizes (a.k.a. perplexities in NE) like powers of 2 up to about N/2. Then entropic affinities are averaged to get multi-scale affinities that can be matched with information-theoretic divergences.

In order to evaluate these methods, quality criteria have been developed, based on neighborhood rank preservation. As those criteria depend on the neighborhood size K, curves of neighborhood agreement with respect to K can be drawn. Rescaling the criterion to account for random embedding and having a log axis for abscissa K visually emphasizes local neighborhoods; the area under the curve yields a scalar score for each compared embedding. DR quality assessment can then be considered in a (DR – DR QA – user) loop for iterative exploratory data analysis. Some examples are discussed and a software interface for exploratory data analysis are presented. A complementary topic is a new method of MDS, working with a low-cost stochastic optmization, coined SQuaD-MDS (stochastic quartet descent). Like other flavors of MDS, SQuaD-MDS is more of a global method. However, it can be combined with accelerated local methods of NE to address their main shortcoming of overlooking the global structure.

A companion talk is given by Cyril de Bodt with recent projects and papers along that line (NE with missing data, fast multi-scale NE, interpretable NE).

3.14 Interactive Visual Analytics and Hypothesis Generation with Non-linear Similarity Embeddings

B.P.F. Lelieveldt (Leiden University Medical Center, NL)

Joint work of B.P.F. Lelieveldt, Nicola Pezzotti, Thomas Hoellt, Anna Vilanova, Thomas Kroes, Julian Thijssen, Jeroen Eggermont, Baldur van Lew, Soumyadeep Basu, Alexander Vieth, Chang Li
Main reference Nicola Pezzotti, Thomas Höllt, Boudewijn P. F. Lelieveldt, Elmar Eisemann, Anna Vilanova: "Hierarchical Stochastic Neighbor Embedding", Comput. Graph. Forum, Vol. 35(3), pp. 21–30, 2016.
URL https://doi.org/10.1111/CGF.12878

Non-linear similarity embedding techniques such tSNE and UMAP have rapidly gained traction for exploratory data analysis and visualization. They have demonstrated their utility for hypothesis generation, and following from that, the formulation of highly targeted experimental setups for verification of these visualization-inspired hypotheses. Key enabling factor for this hypothesis generation is the development of high-performance tools to interact with embeddings that enable on-the-fly drill-ins, re-embedding and complementary views on the data: a visualization paradigm known as visual analytics. This presentation discussed a number of methods to enable and integrate interactivity, as well as embedding dynamics and quality control cues into the visual exploration of high-dimensional data. Departing from the scalable embedding technique Hierarchical Stochastic Neighbor Embedding (HSNE), methods such as progressive visualization of attraction force reduction during embedding, dual sample-feature views, magic lenses for localized alterations in attraction force, elastically-coupled multi-view embeddings, and strategies for "focus and context" drill-in options for multi-million datapoint datasets were discussed. Application examples were focused on life-sciences (single-cell and spatial transcriptomics) and hyperspectral image analysis (satellite imagery and paintings).

3.15 Nearest Neighbour Graphs – Edges and Weights

Leland McInnes (Tutte Institute for Mathematics&Computing – Ottawa, CA)

I proposed an alternative method for generating weights in a nearest neighbour graph, with the intention of using the (directed, weighted) graph for clustering or dimension reduction. The approach uses per point estimates of the distribution of nearest neighbour distances in local regions of the data space; these estimates can be constructed by performing recursive Bayesian updates of estimates based on the estimates of neighbouring points. One can then generate a neighbour graph with edge weights (of affinities) given by the probability (under the points model of nearest neighbour distances) that a given candidate neighbour is a nearest neighbour. It can be shown that results in a graph where edge weights more closely align with distances in low-dimensional representations given by neighbour graph methods such as t-SNE, TriMAP, MDE and UMAP. This provides a potential approach for performing clustering directly on high dimensional data that is competitive with approaches such as UMAP+HDBCSAN.

3.16 The Case for Intermediate-dimensional Embeddings – Looking Deep into the Spectrum of the Graph Laplacian

Gal Mishne (University of California, San Diego – La Jolla, US)

Joint work of Gal Mishne, Xiuyuan Cheng, Raphy Coifman, Hadas Benisty, Dhruv Kohli, Alex Cloninger, Devika Narain, Bas Nieuwenhuis
Main reference Dhruv Kohli, Alexander Cloninger, Gal Mishne: "LDLE: Low Distortion Local Eigenmaps", J. Mach. Learn. Res., Vol. 22, pp. 282:1–282:64, 2021.
URL http://jmlr.org/papers/v22/21-0131.html
Main reference Xiuyuan Cheng, Gal Mishne: "Spectral Embedding Norm: Looking Deep into the Spectrum of the Graph Laplacian", SIAM J. Imaging Sci., Vol. 13(2), pp. 1015–1048, 2020.
URL https://doi.org/10.1137/18M1283160
Main reference Gal Mishne, Ronald R. Coifman, Maria Lavzin, Jackie Schiller: "Automated cellular structure extraction in biological images with applications to calcium imaging data", bioRxiv, Cold Spring Harbor Laboratory, 2018.
URL https://doi.org/10.1101/313981

In this talk, I introduce new unsupervised geometric approaches for extracting structure from large-scale high-dimensional data. The traditional viewpoint of spectral approaches to clustering and manifold learning is to construct a data-driven graph on the data-points and use the top eigenvectors of the graph Laplacian matrix to embed the data. However, in recent work, we have shown the benefit of looking deep within the spectrum of the graph-Laplacian to identify subsets of eigenvectors that characterize the data locally. First, I will present a new robust measure, the Spectral Embedding Norm, to separate clusters from background, and demonstrate its application to both outlier detection and image segmentation. Based on this measure we developed a greedy method for extracting overlapping clusters from a dominant background compound, which we demonstrate on calcium imaging data at different spatial scales (e.g., cellular, widefield). Finally, I will present Low Distortion Local Eigenmaps (LDLE), a "bottom-up" manifold learning technique that constructs a set of low distortion local views of a dataset in lower dimension and registers them to obtain a global embedding. In contrast to existing data visualization techniques, LDLE is more geometric and can embed manifolds without boundary as well as non-orientable manifolds into their intrinsic dimension.

3.17 Probabilistic Embedding Models

Ian Nabney (University of Bristol, GB)

This talk discussed briefly the importance of user involvement in method development with an example from model evaluation: how does a non-expert user know whether further work is needed on a specific model?

The main aim of the talk was to describe how latent variable models can be used for dimensionality reduction and the characteristics of the statistical probability analysis viewpoint. Principal Component Analysis was defined as a probabilistic model and it was shown how it can be generalised to a density model for the data (latent variable model exemplified by Generative Topographic Mapping – GTM). The value of this is the use of a single coherent framework: probabilities (noise and statistical viewpoint not an afterthought but inherent in the model), latent variables, inference, EM algorithm, Bayes. We then discussed how GTM can be extended to deal with missing values, discrete and mixed data types, time-dependent data, hierarchies, and feature selection. Illustrations from real applications were provided throughout.

3.18 On the Epistemic Virtues of Dimensionality Reduction

Maximilian Noichl (Utrecht University, NL)

In the present contribution, we focus on novel techniques of dimensionality-reduction. These methods can be useful both as independent analyses in there own right, as preprocessing steps for further analysis, e. g. clustering, and as visualisation techniques that translate data into two or three dimensions. Because of their undeniable power, both linear variants, like the older PCA, as well as somewhat novel non-linear variants, like t-SNE or UMAP, have become ubiquitous in scientific and commercial data analysis, including domains as varied as chemistry, linguistics, genetics and psychology. Importantly, they are also used to inspect the features learned by neural networks and to visualise their learning-process. But their adoption has not been without controversy, as the structures they produce can be highly sensitive to the choice of hyper-parameters as well as random initialisation. This has made some practitioners cautious in their interpretation and communication of their results, especially regarding settings that have some bearing on social questions. UMAP or t-SNE-visualizations of human genomic data can for example give the impression of clear separation of human groups that is not warranted by the data, a visual feature that has led them to be widely shared in racist internet-communities. In our contribution, we investigate the emergence of epistemic virtues, a notion we borrow from Lorraine Daston's and Peter Galison's work on the virtue of objectivity, surrounding these techniques. We base our analysis on published articles, open-sourced code, tutorials, as well as a computational analysis of social media content, and interviews with key-actors in the domain. Based on our analysis we suggest a first account of the epistemic virtues which in our view ought to surround their practical usage. We suggest that virtues like accessibility, interactivity, explorability can supersede virtues like mechanisation and process-determinacy im some cases. We further highlight how deeply non-epistemic values of software-implementation, like speed and ease of use interweave with epistemic one, and make some suggestions for how the the maintainers of open source packages can improve the environment in which end-users find themselves to contribute to a responsible and scientifically profitable practice.

3.19 VERA: Generating Visual Explanations of Two-Dimensional Embeddings via Region Annotation

Pavlin G. Poličar (University of Ljubljana, SI)

Joint work of Pavlin G. Poličar, Blaž Zupan
Main reference Pavlin G. Poličar, Blaž Zupan: "VERA: Generating Visual Explanations of Two-Dimensional Embeddings via Region Annotation", 2024.
URL https://arxiv.org/abs/2406.04808

Two-dimensional embeddings obtained from dimensionality reduction techniques, such as MDS, t-SNE, and UMAP, are widely used across various disciplines to visualize high-dimensional data. These visualizations provide a valuable tool for exploratory data analysis, allowing researchers to visually identify clusters, outliers, and other interesting patterns in the data. However, interpreting the resulting visualizations can be challenging, as it often requires additional manual inspection to understand the differences between data points in different regions of the embedding space. To address this issue, we propose Visual Explanations via Region Annotation (VERA), an automatic embedding-annotation approach that generates visual explanations for any two-dimensional embedding. VERA produces informative explanations that characterize distinct regions in the embedding space, allowing users to gain an overview of the embedding landscape at a glance. Unlike most existing approaches, which typically require some degree of manual user intervention, VERA produces static explanations, automatically identifying and selecting the most informative visual explanations to show to the user. We illustrate the usage of VERA on a real-world data set and validate the utility of our approach with a comparative user study. Our results demonstrate that the explanations generated by VERA are as useful as fully-fledged interactive tools on typical exploratory data analysis tasks but require significantly less time and effort from the user.

3.20 No Metric to Rule Them All: Gauging the Graphicality of Graph Data

Bastian Rieck (Helmholtz Zentrum München, DE)

Graphs are ubiquitous and constitute the primary data type in many application domains. Modern graph learning algorithms, like *graph neural networks*, permit dealing with graph data in such contexts. Recent research, however, shows that these algorithms are *biased* in the sense that they use the graph structure for tasks even when it unnecessary or detrimental for task performance. Thus, there is a crucial need for understanding to what extent the structure of a graph and its attributes are related. We address this by *lifting* the problem to a comparison of metric spaces defined by either the attributes or the structure of a graph. This defines a new measure that we refer to as *graphicality*. We demonstrate its utility via a suite of experiments while also proving its stability properties.

3.21 Distances and Trees

Enrique Fita Sanmartin (Universität Heidelberg, DE)

Joint work of Sebastian Damrich, Christoph Schnörr, Fred A. Hamprecht
Main reference Enrique Fita Sanmartín, Sebastian Damrich, Fred Hamprecht: "The Algebraic Path Problem for Graph Metrics", in Proc. of the 39th International Conference on Machine Learning, Proceedings of Machine Learning Research, Vol. 162, pp. 19178–19204, PMLR, 2022.
URL https://proceedings.mlr.press/v162/sanmarti-n22a.html
Main reference Enrique Fita Sanmartín, Christoph Schnörr, Fred A. Hamprecht: "The Central Spanning Tree Problem", CoRR, Vol. abs/2404.06447, 2024.
URL https://doi.org/10.48550/ARXIV.2404.06447

In the first part of the talk, we present the "log-norm" family of distances, a novel family of metrics on graphs that interpolates between the shortest path, minimax and commute cost distances. The log-norm family is based on the "algebraic path problem" framework, a generalization of the shortest path problem. In the second part, we introduce a family of robust spanning trees embedded in Euclidean space, named central spanning tree (CST), whose geometric structure is resilient against perturbations such as noise. The family of trees is defined through a parameterized NP-hard minimization problem over the edge lengths, with specific instances including the minimum spanning tree or the Euclidean Steiner tree. The minimization problem weighs the length of the edges by their tree edge-centralities, which are regulated by a parameter α. Two variants of the problem are explored: one permitting the inclusion of Steiner points (referred to as branched central spanning tree or BCST), and another that does not. The effect of α on tree robustness is empirically analyzed, and a heuristic for approximating the optimal solution is proposed.

3.22 Scalable Interaction in Browser-based Embedding Visualizations

Benjamin M. Schmidt (Nomic AI – New York, US)

Joint work of Brandon Duderstadt, Andriy Mulyar, Robert Lesser, Wilson Jr. Marcilio, Vincent Giardina, Aaron Miller, Richard Guo, Benjamin M. Schmidt
URL https://atlas.nomic.ai

Practices of two-dimensional embedding representations that have emerged from the cultural heritage community offer useful models for advancing human-computer interaction techniques in dimensionality reduction. Domain experts in the fields often have extremely little investment in programming but can easily understand and read individual points given a sufficiently advanced interface. In this talk I describe the tactics used in Deepscatter, a typescript/WebGL library, that is able to progressively serve, render, and interactively animate billion-point scatterplots over the web using Apache Arrow and other technologies by storing data in a progressively-loaded quadtree format designed to allow mutations and editing through asynchronous transformations. I also describe the language of data interaction we have developed in the Nomic AI Atlas product for easing the tasks of large-scale filtering, selection, tagging, and search on data represented upstream as embeddings; the creation of selections of data and interactive repositioning of points is an important component of interaction that allows improving models and avoiding the misreadings that are easy when relying on only a single, static view that makes interrogating individual points difficult or impossible.

3.23 Using Higher-Order De Bruijn Graphs to Learn Causality-Aware Representations of Temporal Graphs

Ingo Scholtes (Universität Würzburg, DE)

Joint work of Ingo Scholtes, Lisi Qarkaxhija, Vincenzo Perri, Franziska Heeg
Main reference Lisi Qarkaxhija, Vincenzo Perri, Ingo Scholtes: "De Bruijn Goes Neural: Causality-Aware Graph Neural Networks for Time Series Data on Dynamic Graphs", in Proc. of the First Learning on Graphs Conference, Proceedings of Machine Learning Research, Vol. 198, pp. 51:1–51:21, PMLR, 2022.
URL https://proceedings.mlr.press/v198/qarkaxhija22a.html

Graph Neural Networks (GNNs) have become a cornerstone for the application of deep learning to data on complex networks. However, we increasingly have access to time-resolved data that not only capture which nodes are connected to each other, but also when and in which temporal order those connections occur. A number of works have shown how the timing and ordering of links shapes the causal topology of networked systems, i.e. which nodes can possibly influence each other via so-called time-respecting paths that account for the arrow of time [5]. Moreover, higher-order graph models have been developed that allow us to model patterns in the resulting causal topology [4, 3]. Building on these works, we introduce De Bruijn Graph Neural Networks (DBGNNs), a novel time-aware graph neural network architecture for time-resolved data on dynamic graphs. Our approach accounts for temporal-topological patterns that unfold via causal walks, i.e. temporally ordered sequences of links by which nodes can influence each other over time. This enables us to learn patterns in the causal topology of time series data on complex networks, which facilitates to address learning tasks in temporal graphs.

In my talk, I will show how we can use higher-order De Bruijn graph models of time-respecting paths to learn low-dimensional Euclidean representations that capture both temporal and topological patterns in data on temporal graphs. Building on a generalization of graph Laplacians to higher-order De Bruijn graph models [5], I will show how we can use a Laplacian embedding to detect temporal-topological cluster patterns in temporal graphs. I further demonstrate a neural representation learning technique that is based on the De Bruijn Graph Neural Network (DBGNN) architecture [2]. Apart from facilitating node classification it has recently been used to predict temporal node centralities in temporal graphs [1].

References

1 Franziska Heeg, Ingo Scholtes: *Using Causality-Aware Graph Neural Networks to Predict Temporal Centralities in Dynamic Graphs*. CoRR abs/2310.15865 (2023)

2 Lisi Qarkahija, Vincenzo Perri, Ingo Scholtes. *De Bruijn goes Neural: Causality-Aware Graph Neural Networks for Time Series Data on Dynamic Graphs*. Learning on Graphs Conference, LoG 2022, 9-12 December 2022, Virtual Event, Proceedings of Machine Learning Research, Vol. 198 (2022)

3 R Lambiotte, M Rosvall, I Scholtes: *From networks to optimal higher-order models of complex systems*, Nature Physics, Vol. 15, pp. 313-320 (2019)

4 Ingo Scholtes: *When is a Network a Network?: Multi-Order Graphical Model Selection in Pathways and Temporal Networks*. KDD 2017: 1037-1046 (2017)

5 Ingo Scholtes, Nicolas Wider, Rene Pfitzner, Antonios Garas, Claudio Tessone, Frank Schweitzer: *Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks*, Nature Communications 5 (2014)

3.24 Guided Data Exploration with (Semi-)Supervised Manifold Learning

Guy Wolf (University of Montreal, CA & MILA – Montreal, CA)

Modern challenges in exploratory data analysis, especially in biomedical applications involving single cell data, give rise to representation learning techniques that aim to capture intrinsic data geometry (e.g., patterns and structures), while separating it from data distribution that is typically biased by data availability and collection artifacts, thus allowing discovery of rare subpopulations and sparse transitions between meta-stable states. A common approach in this area, which I discuss in this talk, is the construction of a data-driven diffusion geometry that both captures intrinsic structure in data and provides a generalization of Fourier harmonics on it, combining tools and perspectives from a range of fields including manifold learning, graph signal processing, and harmonic analysis. However, most methods following this paradigm rely on unsupervised learning, under the assumption that the target phenomena of interest will form the dominant emergent patterns in the data, uncovered by the extracted representation. While this is the case in certain controlled experiment conditions, such property cannot be guaranteed in many observational services settings. As an alternative, here we discuss semi-supervised approaches that leverage annotations and meta information that often accompanies collected data, in order to guide the data geometry to accentuate task-informed structures in the learned representation. This approach is demonstrated in data exploration tasks including visualization and multimodal data fusion.

Participants

- Michael Bleher
Universität Heidelberg, DE
- Kerstin Bunte
University of Groningen, NL
- Corinna Coupette
MPI für Informatik – Saarbrücken, DE
- Sebastian Damrich
Universität Tübingen, DE
- Cyril de Bodt
University of Louvain, BE
- Alex Diaz-Papkovich
Brown University – Providence, US
- Laleh Haghverdi
Max-Delbrück-Centrum – Berlin, DE
- Fred Hamprecht
Universität Heidelberg, DE
- Ágnes Horvát
Northwestern University – Evanston, US
- Dmitry Kobak
Universität Tübingen, DE
- Dhruv Kohli
University of California – San Diego, US
- Smita Krishnaswamy
Yale University – New Haven, US
- John Aldo Lee
UC Louvain-la-Neuve, BE
- B.P.F. Lelieveldt
Leiden University Medical Center, NL
- Leland McInnes
Tutte Institute for Mathematics & Computing – Ottawa, CA
- Gal Mishne
University of California, San Diego – La Jolla, US
- Ian Nabney
University of Bristol, GB
- Maximilian Noichl
Utrecht University, NL
- Pavlin Poličar
University of Ljubljana, SI
- Bastian Rieck
Helmholtz Zentrum München, DE
- Enrique Fita Sanmartin
Universität Heidelberg, DE
- Benjamin M. Schmidt
Nomic AI – New York, US
- Ingo Scholtes
Universität Würzburg, DE
- Guy Wolf
University of Montreal, CA & MILA – Montreal, CA

Remote Participants

- Miguel Á. Carreira-Perpiñán
University of California – Merced, US