DAGSTUHL REPORTS

Volume 14, Issue 2, February 2024

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
July, 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.14.2.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 24061

Are Knowledge Graphs Ready for the Real World? Challenges and Perspective

David Chaves-Fraga*[1], Oscar Corcho*[2], Anastasia Dimou*[3], Maria-Esther Vidal*[4], Ana Iglesias-Molina†[5], and Dylan Van Assche†[6]

1 **Universidade de Santiago de Compostela, ES.** `david.chaves@usc.es`
2 **Universidad Politécnica de Madrid, ES.** `ocorcho@fi.upm.es`
3 **KU Leuven, BE.** `anastasia.dimou@kuleuven.be`
4 **TIB – Hannover, DE.** `vidal@l3s.de`
5 **Universidad Politécnica de Madrid, ES.** `ana.iglesiasm@upm.es`
6 **IDLab - Ghent University - imec, BE.** `Dylan.VanAssche@ugent.be`

Abstract

This report documents the program and results of the Dagstuhl Seminar 24061 "Are Knowledge Graphs Ready for the Real World? Challenges and Perspectives". The seminar focused on gaining a better understanding of the open challenges required for the development of Knowledge Graph ecosystems. The seminar focused on four different topics: access control and privacy in decentralized knowledge graphs, knowledge graph construction lifecycle, software methods for improving KG implementation, and a new wave of knowledge engineers and their expected skills. By focusing on these relevant research topics, the seminar aimed to reflect on KGs from a more fundamental computer science perspective. It brought together interdisciplinary researchers from academia and industry to discuss foundations, concepts, and implementations that will pave the way for the next generation of KGs ready for real-world use.

Seminar February 4–9, 2024 – https://www.dagstuhl.de/24061

2012 ACM Subject Classification Information systems → Resource Description Framework (RDF); Information systems → Data exchange; Information systems → Graph-based database models; Information systems; Information systems → Data model extensions; Information systems → Query languages; Information systems → Data structures; Information systems → Information integration; Information systems → Semantic web description languages; Information systems → Web applications; Social and professional topics → Computing education; Security and privacy → Security services; Theory of computation → Semantics and reasoning; Information systems → Information retrieval

Keywords and phrases access control and privacy, federated query processing, intelligent knowledge graph management, programming paradigms for knowledge graphs, semantic data integration

Digital Object Identifier 10.4230/DagRep.14.2.1

* Editor / Organizer
† Editorial Assistant / Collector

Are Knowledge Graphs Ready for the Real World? Challenges and Perspective, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 1–70

Editors: David Chaves-Fraga, Oscar Corcho, Anastasia Dimou, and Maria-Esther Vidal

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

1 Executive Summary

David Chaves-Fraga (Universidade de Santiago de Compostela, ES, david.chaves@usc.es)
Oscar Corcho (Universidad Politécnica de Madrid, ES, oscar.corcho@upm.es)
Anastasia Dimou (KU Leuven, BE, anastasia.dimou@kuleuven.be)
Maria-Esther Vidal (TIB – Hannover, DE, vidal@l3s.de)

Graphs and knowledge bases have been around for many decades, and research results have had a tremendous impact on areas such as mathematics, artificial intelligence, and databases. However, although the term has been coined by the scientific community, technological developments and astronomical data growth have made knowledge graph (KG) management a fundamental topic in various areas of computer science today. The scientific and industrial communities have responded to the emerging field of knowledge management. As a result, formal frameworks for defining and representing KGs, as well as methods for creating, exploring, and analyzing KGs, have flourished to make KGs a reality. However, despite the tangible results, sustainability is still compromised by the lack of transparent and accountable management of CCs. The real-world application of KGs requires programming paradigms for KG management, transparent data integration and quality assessment techniques, and methods for maintaining access control and privacy. In addition to technological advances, societal adjustments can have a tremendous impact on the management of KGs. The seminar addressed these socio-technical challenges with a mix of invited talks, lightning talks, and small group workshops as follows:

The Incremental Creation of Knowledge Graphs. Creating a Knowledge Graph (KG) involves several open research challenges, such as data extraction, data quality, data integration, and data security. It also requires attention to architectural aspects such as scalability and interoperability. A working group was formed to discuss and focus on two main topics: the definition of a general pipeline for KG construction and its relationship to data quality. The main outcome is a standard formalization of the KG construction lifecycle and its associated components. This definition is accompanied by quality measures and provenance tracking of all steps.

Support of Knowledge Graph Implementation. Software engineering and programming languages have created approaches and techniques that support complex tasks during software development such as software dependencies, error identification, testing, syntactic validation, software lifecycle, etc. We look into these proposals to determine a set of requirements in software lifecycle management for knowledge graphs. They will improve and facilitate the implementation of knowledge graphs in industrial and complex environments, taking into account the relationships and dependencies between all the artifacts used (ontologies, shapes, mappings, tests, etc.) as well as their evolution and versioning. To achieve this goal, we believe that it is necessary to have a better understanding and general overview of how knowledge graphs are implemented. Therefore, a workshop on this topic has been proposed at ISWC2024[1]. After its celebration, the next step will be to create a community around this topic with researchers and industry stakeholders to standardize and implement the identified challenges/requirements.

[1] https://w3id.org/soflim4kg

Access Control in Decentralized Knowledge Graphs. Exploring access control in decentralized Knowledge Graphs has been a relatively underexplored area. Specifically, mechanisms for restricting access to knowledge to safeguard confidential information and personal data, as well as establishing consent models for the processing of personal data, have not received substantial attention within the realm of Knowledge Graph management. Additionally, ensuring compliance with usage policies has been inadequately addressed, particularly in the context of decentralized Knowledge Graphs. During the seminar, a dedicated group convened to deliberate on approaches for managing Knowledge Graphs across a federation of decentralized instances.

A New Generation of Knowledge Engineers. Improving the utilization and management of knowledge graphs requires educating a diverse audience about both the social and technical aspects of knowledge work. To address this need, a dedicated working group was established. This group conducted an analysis to identify existing educational resources and gaps in knowledge, exploring how consensus could be fostered among various stakeholders in the field. Moreover, the group investigated the specific educational requirements tailored to different audiences, including professional students, undergraduates, and postgraduates. By thoroughly examining these aspects, the working group aimed to formulate strategies for enhancing education and understanding in the domain of knowledge graph utilization and management.

2 Table of Contents

Breakout Groups

3 Invited Talks

3.1 Semantic Data Integration

Maurizio Lenzerini (Sapienza University of Rome, IT)

Data integration is considered as one of the old problems in data management. The definition assumed in the talk is the one characterizing data integration as the problem of providing a unified and transparent view to a collection of data stored in multiple, autonomous, and heterogeneous data sources. The unified view is achieved through a global schema, and is realized either through a materialized database (warehousing, or exchange), or through a virtualization mechanism based on querying (virtual integration). Formal approaches to data integration started in the 90's (see for example the PODS 1995 by A. Y. Levy and colleagues). Since then, research both in academia and in industry has addressed a huge variety of aspects of the general problem. In this talk the focus was on the idea of using semantics for making data integration more effective. But what does it mean to pose semantics at the center of the scene? The main element is that the unified view is not simply a data schema harmonizing the schemas of the data sources, but it is a structure formally representing the domain of interest for the users of the system. The unified view to information consumers is not merely a data structure accommodating the various data at the sources, but a semantically rich description of the relevant concepts in the domain of interest, as well as the relationships between such concepts. The distinguishing feature of the whole approach is that users of the system will be freed from all the details of how to use the data sources, as they will express their needs (e.g., a query) in the terms of the concepts, the relations, and the processes described in the domain model. Since ontologies are suitable formalisms for domain modeling, Ontology-based Data Integration (OBDI) is a promising direction for realizing the above idea. The founding principle of OBDI is to apply suitable techniques from the area of Knowledge Representation (KR) and Reasoning for a new way to achieve data governance and integration, based on the principle of managing heterogeneous data through the lens of an ontology. OBDI resorts to a three-level architecture, constituted by the ontology, the data sources, and the mapping between the two:

- The data layer is constituted by the existing data sources that are part of the organization.
- The ontology is a declarative and explicit representation of the domain of interest for the organization, specified by means of a formal and high level description of all relevant aspects.
- The mapping is a set of declarative assertions specifying how the available sources in the data layer and the computational resources used in the organization relate to the ontology.

The above components are used to provide novel, sophisticated services in governing and integrating data. In particular, the ontology is the pivotal element for documenting the IT resources of the organization, expressing their semantics at an abstract, conceptual level. By working with a specification mechanism that is close to the conceptual view of the domain, the designer is facilitated in the initial specification, evolution and maintenance of IT resources, and the final user interacts with the data in a more natural way than in the case of a unified view expressed in terms of a data schema. The further step that is becoming more and more popular is the idea of rendering the ontology in terms of a Knowledge Graph, thus exposing the meaning of data and the structure of the domain knowledge to users/applications in a very flexible and effective way. Now, the idea of using Semantic Networks for information integration is an old one, and since Knowledge Graphs have a lot in common with Semantic Networks, the talk presented a brief history regarding KR formalisms, from Semantic Networks to Knowledge Graphs, not only discussing the similarities, but also pointing out several crucial differences, and delving into the details of three of them.

The first difference is the central role played by formal semantics when Knowledge Graphs are used to express ontologies. Indeed, it is generally accepted to formalize ontologies in logic, and in any logic the expressions built according to the appropriate syntax rules are assigned formal semantics. In turn, this has the important consequences of providing a solid formalization of the notion of inference (in particular, deductive inference) in the context of Knowledge Graphs, something that was missing in the early Semantic Networks. Such formalization opens up the possibility of defining precise methods for reasoning about the Knowledge Graph and the mappings, and for reformulating the needs (e.g., queries) expressed over the Knowledge Graph in terms of appropriate calls to services provided by the data source managers.

The second difference is the attention to the cost of reasoning. In order to translate the services expressed over the ontology into correct and efficient computations over the data sources, techniques typical of the two areas of Knowledge Representation and Automated Reasoning are crucial. Unfortunately, many reasoning problems in the context of expressive formalisms are intractable or even undecidable. It is therefore crucial to find a trade-off between expressive powers of axioms formalizing Knowledge Graphs and the computational complexity of reasoning over them. Moreover, OBDI introduces new challenges to these areas. Indeed, while Knowledge Representation techniques are often confined to scenarios where the complexity resides in the rules governing the domain, in OBDI one often faces the problem of a huge amount of data in the data layer, and this poses completely new requirements for the reasoning tasks that the system should be able to carry out. For example, the notion of data complexity, by which one measures the computational complexity on the basis of the size of the data layer only, is of paramount importance in OBDI.

Finally the third difference is that the architecture of OBDI suggests considering new types of reasoning problems that may be useful in real world scenarios and are addressed in recent papers. In the talk, three of them were discussed in more detail. One is the need for reasoning about metamodeling features, one is the challenge of semantically characterizing in terms of the Knowledge Graphs the services expressed over the data source, and the third is the problem of assessing the quality of data by using the domain model expressed in the Knowledge Graph. The first issue arises from the fact that Knowledge Graphs naturally mix intensional and extensional knowledge while the great majority of research works on ontology (e.g., the ones based on OWL and Description Logics) keep them separated. The second issue calls for new types of reasoning tasks in the context of OBDI. A recent proposal in this direction defines the notion of abstraction: given a process P expressed at the level of the sources, the goal is to compute the query over the Knowledge Graphs, called abstraction, that captures P for every configuration of the data layer. The third issue refers to the possibility of using the domain model represented in the Knowledge Graph as a yardstick for measuring the quality of data at the sources along various dimensions, such as, for example, consistency, accuracy, and completeness.

3.2 Access Control, Policies and Constraints

Sabrina Kirrane (Wirtschaftsuniversität Wien, AT)

Knowledge graphs hold immense potential for the public and the private sector alike, both in terms of internal enterprise data management and data sharing across enterprises. In both cases there is a need for restricting or constraining access to knowledge in order to protect confidential information and personal data. There are already several standards that can be used to specify licenses (the Open Digital Rights Language (ODRL)), legal requirements (LegalRuleML), and constraints (the Shapes Constraint Language (SHACL)) that are particularly suitable for knowledge graphs. ODRL is a W3C recommendation, which is composed of two complementary specifications (i) the ODRL Information Model; and (ii) the ODRL Vocabulary and Expressions. Although the language is most suitable for modeling intellectual property rights, over the years it has also been used to model contracts, regulatory requirements, access policies, and privacy policies. A more suitable choice for modeling legal texts (e.g., legislation, regulations, contracts) is LegalRuleMl, which is an OASIS specification that extends the modular rule markup language (RuleML) with the ability to encode legal norms. From a constraints perspective, the W3C SHACL specification has proven its effectiveness for validating Resource Description (RDF) based graph databases. More recently, researchers have been exploring the use of SHACL for encoding access policies.

Additionally researchers have demonstrated the potential of the W3C Web Ontology Language (OWL) general purpose modeling language in order to model consent for personal data processing and the corresponding legal requirements stipulated in the General Data Protection Regulation (GDPR). Another noteworthy initiative is the work conducted by the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) in relation to modeling privacy and data protection vocabularies primarily derived from GDPR. Also, extensive access control research has resulted in proposals that can be used to specify access policies at different levels of granularity: named graphs, views, triples, triple patterns, and quad patterns. Many of these works either adopt the OASIS eXtensible Access Control Markup Language (XACML) standard enforcement components or propose derivatives that were inspired by the XACML standard.

In recent years the focus has shifted from access control to usage control, which strives to constrain what happens with data and knowledge after access has been granted. There are a number of different usage control frameworks that could potentially be applied to knowledge graphs, however in a decentralized setting once data or knowledge is shared, different copies and derivations of the same data or knowledge can be easily shared with others. Although it is extremely difficult to constrain ad hoc sharing there are a number of desideratum: data producers (individuals and organizations) need to be able to attach usage policies to data and knowledge; there is a need for techniques that can ensure continuous adherence to these policies; and legal fall back measures need to be supported via appropriate technical governance mechanisms. At present there are different approaches: (i) assume that data consumers want to demonstrate compliance and show how policy languages and system logs can together be used to provide transparency or evidences of compliance; or (ii) use a combination of hardware and software, such as Intel's Software Guard Extensions (SGX) trusted execution environment, to ensure compliance.

Looking more broadly at the topic of access control, policies, and quality, there have been a number of journal special issues soliciting state of the art research in terms of knowledge graph governance, validation, and quality assurance. From a data and algorithmic governance perspective a recent call for papers[2] was very broad aiming to attract research ranging from privacy and data protection to trust, authenticity, integrity, and bias detection and mitigation. However, the special issue attracted works that aimed to support companies in complying with the GDPR, with contributions in relation to: consent approaches and best practices; ontologies for representing personal data processing information flows; privacy-preserving data analytics; and differential privacy for knowledge graphs. From a validation and a quality assurance perspective a recent call for papers[3] solicited theoretical and practical methods and techniques for assessing and validating the quality of knowledge graphs with either quality or validation appearing in the majority of the listed topics. Interesting the special issue attracted a broad range of contributions on: the quality of linked open data sources; techniques for learning SHACL shapes from knowledge graphs; a single unified graph data model that embraces both RDF and property graphs; assertion and alignment correction for large knowledge graphs; exploring how subgraphs change over time; and facilitating composition and reusability.

Since the GDPR, the legal landscape in Europe has evolved considerably with several important new acts appearing under the digital strategy for Europe umbrella, such as, the Copyright in the Digital Single Market Directive[4], the EU Data Governance Act[5], the EU Data Act[6], and the proposed EU Artificial Intelligence Act[7]. These acts will not only have a major influence on both the public and private sector, but will also be a major driver for research in relation to knowledge graph governance, validation, and quality assurance.

Additionally there are a number of open challenges and opportunities when it comes to privacy, policies, and quality. We need to put more emphasis on tech transfer and develop best practices for software engineers and architects. Performance, scalability, and usability need to be assessed in practical real world settings. There is no standard general purpose policy language capable of representing various policies, norms, and preferences. Machine-readable policies must faithfully represent human policies and norms. Technical usage control is difficult, which means we often need to rely on legal agreements. We need to get into the practice of defining attacker models for privacy and security use case scenarios.

3.3 Programming Languages

Martin Giese (University of Oslo, NO)

The talk addressed the relationship between knowledge graphs (KGs) and programming languages. Two central observations are:

1. There is more than one interesting connection: while it is interesting to ask for a programming language or paradigm particularly suitable for generating, transforming, or accessing KGs, programs can also be seen as specifications of processes, i.e. how

[2] `https://www.semantic-web-journal.net/blog/call-papers-special-issue-semantic-technologies-data-and-algorithmic-governance`
[3] `https://www.semantic-web-journal.net/blog/call-papers-special-issue-knowledge-graphs-validation-and-quality`
[4] `https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32019L0790`
[5] `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868`
[6] `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN`
[7] `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206`

situations change over time. Desirable language features for these two purposes will be quite different. E.g., while efficiency and scalable software engineering are important in the first case, the ability to verify properties deductively or by model checking is more important in the second case.

2. One of the main goals of semantic technologies is to treat as much as possible of the "business logic" through knowledge representation, and to implement generic software that is driven by arbitrary ontologies, constraints, etc., rather than implementing software that is strongly tied to an application. Therefore, before asking which programming language is best suited for knowledge graph work, one should always ask first if a generic tool would be more appropriate.

The talk continued by discussing a selection of approaches from the literature, that address the combination.

The most obvious approach is maybe to use a logic programming language like Prolog: it delivers a very tight integration between programming and knowledge representation, since Prolog rules and facts can in fact be seen as executable representations of domain knowledge. Logic programming is extremely well researched, and the relation to databases (in the form of Datalog) is also well known. There are several drawbacks however: If there is no separation between the "external knowledge" to be manipulated and the program itself, then representing change, i.e. the evolution of knowledge is not straightforward – unless one implements knowledge bases as a data structure in Prolog, losing much of the original appeal. Similarly, reasoning is built-in, but only on Prolog's terms. If something else is needed, one needs to work around what Prolog provides, rather than using it.

Another interesting approach is that of programming languages that explicitly have a knowledge base as part of the program state and that let programs query and modify this knowledge base. We reviewed the 2015 work of Zarieß and Claßen that builds on Levesque and Reiter's GOLOG from 1997. The language allows querying the knowledge base using epistemic operators; i.e. queries are explicitly about all known instances of some concepts, it is possible to "sense" parts of the knowledge base, after which corresponding parts of the state become known, and it possible to modify the knowledge base by means of "effects" which change the KB in such a way that a fact becomes true. The authors show how to verify properties expressed in CTL (Computation Tree Logic) for such programs. Although the language presented is a toy language, with comparatively complicated semantics, it has many interesting properties. A related approach was presented by Calvanese, de Giacomo, Lenzerini, and Rosati in 2011. Where Zarieß and Claßen's proposal uses a semantic that modifies epistemic interpretations, i.e. semantic entities, the work of Calvanese et al modifies the ABox, i.e. the syntax of the KB, using the theory of Knowledge Base revision when changes lead to inconsistency. While these languages certainly show a direction for programming with knowledge, it is not clear how appropriate the approach would be for the practical manipulation of knowledge graphs. Similarly to the criticism of logic programming, the built-in reasoning and update of knowledge bases may or may not be what is required, and a more direct access to the representations and not only the knowledge is useful in many cases.

The remainder of the talk addressed attempts at reconciling object oriented (OO) programming and modeling with semantic approaches. While an OO program can deal with knowledge graphs as data structures, as is done in frameworks like Jena or RDF4J, the wish here is to include some more of the semantics into the programs. We reviewed some of the most important issues that make a straightforward mapping between OO classes and OWL classes, UML associations and object properties, etc., problematic:

1. OO models and programs describe data, in a closed world way. Comparing an OWL statement that every person has a father to a similar UML model or an OO programming asserting that the hasFather field of every person is non-null, we see that the OO semantics is closer to the closed-world view of databases or SHACL than to OWL.
2. The OO view is almost universally that resources have a single type that they receive upon creation, that will not change, and that dictates their behavior. This is quite different from the semantic technology view where resources can be inferred to have arbitrarily many types.
3. Associations in UML and fields in programs are placeholders for data values. There is no direct correspondence between UML associations and OWL/RDFS range and domain statements.

We briefly reviewed the work on Type-safe programming for the Web by Martin Leinberger and others. The main idea here is to extend the Java type system with types corresponding to DL concepts. This work circumvents the described projects by essentially keeping the OO and DL class hierarchies separate.

Finally, we gave an overview of our own work on the SMOL language (smolang.org), developed in the frame of the PeTWIN project (petwin.org) to enable the semantic orchestration of digital twins. The core idea of SMOL is the semantic lifting of program states. In its simplest form, this means that a direct mapping of OO program states, including all objects and fields, to RDF is defined. During the program execution, this RDF graph can be queried from the program, giving the programmer a means of semantic reflection. This becomes more interesting if an ontology and more specific mappings are added that describe the application domain and its relationship to the program data structures. In this way a coherent semantic view on program states can be maintained, and used from within the program, while maintaining the necessary and useful "semantic gap" between the object oriented and semantic views. The approach has been used in a case study connecting an elaborate formal ontology of structural geology to SMOL code for the qualitative simulation of geological processes. On the theoretical side, we have been able to establish type safety properties, using theorem proving to reestablish type safety despite the untyped nature of SPARQL.

To summarize, one should avoid putting domain knowledge into programs if possible. If it is necessary, the choice of language will depend on what one wants to achieve – Execution? Modeling? Reasoning? – and how much actual programming will be needed. If "real" programming is required, then a proper, e.g. object oriented language may be required, but the semantic gap between the OO and semantic perspectives must be kept in mind.

3.4 Industry Perspectives

Souripriya Das (Oracle Corp. – Nashua, US)

The content presented in this talk consisted of three parts: 1) the first part focused on identifying some of the major factors that affect adoption of a technology in the industry; 2) the second part briefly outlined how Oracle's graph technology products try to take these factors into account; 3) the third part identified, and suggested possible approaches to addressing, some new requirements that may be critical for wider adoption of graph technologies in the industry.

At the outset of the first part, we explained a sequence of phases that a technology typically goes through: 1) innovators and researchers, from academics and industry, invent novel concepts, techniques and algorithms that try to create new or improved solutions to emerging or pre-existing problems; 2) experts from the relevant domains then work on those novel ideas to create standards that refine and augment those ideas to ensure their implementability in practice; 3) availability of those standards enables vendors to create products that implement them and relevant tools; 4) customers then use those products to create solutions for the end users; 5) the end users make use of those solutions in their business use cases.

Next, we outlined some of the factors that we have seen as playing a major role in the industrial adoption of new technologies. The first and foremost important factor is that the proposed novel concepts and techniques must provide substantial tangible benefits for important existing or anticipated business use cases. Substantial effort has to be expended to ensure this is reliably established so that the proposal receives positive attention from the industry. The second factor is that the technology must be powerful yet intuitive enough that it can be grasped easily by the developers and the developers in turn can explain its potential benefits to the business leaders in their companies. Being on the positive side for this second factor enables experts from the industry to obtain approval to participate in the standards work for the technology. Involvement of industry experts with prior experience in dealing with use of related technologies helps in keeping the standards concise yet sufficient for most of the anticipated major requirements in practice. Several factors come in next as the vendors create products that implement the standards. These factors have to do with going beyond implementing just the functionality. These include scalability and efficiency in handling large amounts of data and large numbers of concurrent users, access control, security, and availability. Additionally, a set of tools relevant for the technology must be made available or be present in the near-term roadmap.

As shortcomings of a new technology for existing or newly-arising use cases get identified, it becomes important to enhance the technology. We explained why critical factors relevant for technology upgrade include not only the simplicity of the revisions to avoid a "major reboot" of the existing mental model but also ensuring full backward compatibility so that every pre-existing thing – already-loaded data, formatted input data, queries, tools – must continue to work without requiring any changes. Invalidation of pre-existing queries, for example, becomes a major expense for an enterprise because large numbers of queries may need to be rewritten.

Next, we shared the good news that over the last 20 years, graph technologies in Oracle went from being relatively unknown inside Oracle to something that shows up as a top technology in presentations by Oracle Executives: graph is now listed as a major data model in the Oracle Converged Database along with relational and JSON. We also indicated that our internal groups have started creating and using ontologies with a goal to provide cloud services for integrating industrial data from a large number of customers in areas such as energy and water utilities.

The second part of the talk briefly outlined how Oracle products for graph technologies – RDF and property graph – took into account some of the major factors for industry adoption outlined in the first part. Graph technologies in Oracle Database provide comprehensive support for standards including W3C's RDF, RDFS, OWL, SPARQL, RDB2RDF, and OGC's GeoSPARQL. Besides scalable and efficient support for concurrent use of RDF, Oracle also provides extensive support for data sharing and access control, including fine-grained (triple-level) access control using its OLS (Oracle Label Security) technology. Furthermore, given the importance of tools for industry adoption, Oracle's graph technology offering includes native tools like RDF Server and Query UI and Graph Studio, plugin for Oracle REST Data Services (ORDS), and adapters for popular open source tools Eclipse RDF4J, Apache Jena, and Protege. Oracle's support for SPARQL in SQL (using the SEM_MATCH table function) and the SQL/PGQ standard enables integrated access to graph and relational data.

The third part of the talk included topics on identifying and exploring possible ways of addressing new requirements that may be critical for industrial adoption. First, we identified the following as some of the possible reasons why SQL/PGQ became part of SQL standard but SPARQL in SQL did not: the complexity of IRIs in RDF (vs. scalar values in property graph), extreme flexibility of RDF (vs. strictly-typed properties, no direct support for multi-valued properties, and only natively-typed values in property graphs), no support for edge-properties in RDF, and path-query support in SPARQL being limited only to reachability.

We outlined the latest thoughts in the W3C RDF Working Group on extending RDF to include support for named occurrences of RDF triples that would not only address the absence of edge-property support but go beyond to allow statements about statements. An example would be representing "A knows B, according to C" as 1) an occurrence of the A knows B triple, named as :e, and 2) then adding the triple :e :accordingTo :C.

Finally, we moved to a discussion on a potential new requirement – neighborhood-aware path traversal – that we had given a presentation on at the 2022 Knowledge Graph Conference (KGC). The requirement involves the ability for a query to return information about paths that match a given iterative pattern and satisfies path constraints such as "no two adjacent vertices in a path can have the same color". SPARQL not only cannot return path information, it cannot even compare properties of two adjacent vertices on paths of length > 1. This is due to support being limited to reachability only and no access to properties of interior elements. Although property graph query languages can handle constraints on two adjacent vertices, they fail when constraints involve three or more adjacent vertices (or two or more adjacent edges). We presented a possible extension to SPARQL to accommodate this requirement. Given the iterative pattern ?x :knows+ ?y, let us assume that $x and $y represent the subject and object of each individual edge (triple) in the matching paths. A new PATH clause can then be used to express constraints on the path. For example, the constraint that "no two adjacent vertices can have the same color" may be expressed as: PATH (FILTER($x.color != $y.color)). To extend the constraint to three adjacent vertices (instead of two), we can extend the PATH clause above to include an additional check of the form FOR EACH <segment-pattern> FILTER(...) to apply to each path segment pattern of length two: PATH (FILTER($x.color != $y.color) FOR EACH ?s ?p ?o . ?o ?p2 ?o2 FILTER(?s.color != ?o2.color)).

Next, we introduced the following additional clauses for use under a PATH clause: AGGREGATE clause for allowing path ("horizontal") aggregation over the properties of the vertices in a path, HAVING clause to filter over some of the path aggregates, and TOP <k> clause to pick the best k paths determined by the path aggregates.

Example: Consider an RDF graph where each vertex has (single) values for properties :name, :color, and :worth. We want to retrieve the highest max(worth) :knows path from the vertex named "A" to the vertex named "D". The path must satisfy the following constraints: 1) no two adjacent vertices can have the same color and 2) max(worth) for the path must be less than 100. Using the proposed extensions, a SPARQL query for such a neighborhood-aware, aggregate-sensitive path traversal can be written as follows:

```
SELECT ?nameList ?maxWorth
{ ?x :name "A" . ?y :name "D" .
  ( ?x :knows+ ?y PATH (
     FILTER ($x.color != $y.color)
     AGGREGATE ( (GROUP_CONCAT($y.name) as ?nameList) (MAX(?worth) as ?maxWorth) )
     HAVING (?maxWorth < 100)
     TOP 1 (ORDER BY desc(?maxWorth))
  ))
}
}
```

We argued that support for neighborhood-aware path traversal in SPARQL and in property graph query languages is a critical requirement. The SQL/PGQ folks are already working on addressing this requirement and we hope that the SPARQL community too will address this in the near future.

4 Lightning Talks

4.1 From linked data for regulatory reporting towards an interoperable data layer for the rail sector: The ERA KG

Marina Aguado (European Union Agency for Railways, FR)

The European Union Agency for Railways is an institution of the European Union mandated to contribute to the implementation of the one Single European Railway Area and is system authority for a set of registers and datastores referred to in several Directives. The goal of these datastores is to facilitate the railway data interoperability. In 2019, the European Union Agency for Railway initiated its journey towards data centricity supported but the ERA rail ontology and the ERA KG. It was decided at Management Board level that linked data and linked data and semantic web technologies and standardized semantic artifacts will be the basis for any software development in the Agency[8]. Since then we have managed to make the ERA ontology legally binding and it also appears as an instrument in the European Mobility Data Space and in particular in the Regulatory Reporting Data Space[9,10].

The two first registers in the ERA KG were the European Register of Infrastructure RINF and the ERATV register of railway vehicle types. This dataset includes various aspects of the railway infrastructure such as track parameters, signalling systems, speeds, load capacities, station facilities and interoperability characteristics. This information is essential for planning European cross-border railway services as well as for ensuring route compatibility of rolling stock across different European countries' railway networks. By harmonizing and sharing infrastructure data, ERA KG helps to reduce barriers to cross border railway operations and contributes to the development of a more integrated and efficient European railway system. Nowadays, the registers describes more than 270K track segments, 50K stations and more

[8] https://www.era.europa.eu/content/decision-n%C2%B0250-management-board-european-union-agency-railways-adopting-roadmap-data-and_en
[9] https://transport.ec.europa.eu/system/files/2023-11/COM_2023_751.pdf
[10] https://ec.europa.eu/newsroom/dae/redirection/document/101623

than 50K geo referenced objects from 27 countries[11,12]. The ERA KG hosts more than 36 million triples more than 31k lines of mappings and more than 100 SHACL shapes. The Agency acts as a ERA as neutral vocabulary provider and identity provider for the data exchange in the EU Common European Mobility data space facilitating data interoperability in the Transport Sector.

The major challenges experienced in our roadmap can be grouped in social challenges (the normal reluctancy to changes, adoption of new technologies) and technical challenges. In relation to these last group, it is worthy to mention that we have experienced quite a technical challenging scenario in regards the availability of standard solutions for access control policies between different vendors, and also from the low performance of shacl validation engines. Another challenge is the heavy burden of backwards compatibility requirements and legacy backwards compatibility. Another technical challenge is the need to be able to go back in time and versioning of the ERA KG. In this regard, the Agency is exploring the Linked Data Event Stream technology.

4.2 Knowledge Graphs for the Circular Economy

Eva Blomqvist (Linköping University, SE)

The current societal transformation that is needed to reach a sustainable society for future generations will require organizations and citizens to collaborate in completely new ways. It will also require participation from all, not only large corporations or technically skilled people. Further, a key to such transformation is access to data, and data and information that accompanies physical objects, which can be accessed and used many years after the object was first produced or introduced to society. Certainly, trust, security, confidentiality, and ethics play a big role in such collaborations and data sharing scenarios. A specific scenario, and a key enabler for sustainability, is the Circular Economy (CE). The aim of the CE is to minimize waste, by maintaining the value of products, as resources, and applying various strategies for reusing, remanufacturing, refurbishing, or recycling materials and products, instead of letting them become waste. However, due to lack of information about a certain product or material, today much of the resources in our society cannot be retained, but become waste that is either incinerated or left as landfill.

Knowledge Graphs (KG) are already used for many applications, and increasingly so also in distributed scenarios. However, large scale scenarios, such as data sharing and integration for the CE, are still quite unexplored. Nevertheless, such scenarios pose very challenging, yet interesting, research questions for the future of KG research. Some of the challenges in data sharing for the CE include:

- The global scale of actors that need to share data – a very high number of actors, across industry domains and across borders.
- The variety of data that needs to be shared – data needs to be interoperable across industry domains.

[11] `https://data-interop.era.europa.eu/`
[12] `https://data.europa.eu/en/publications/datastories/linking-data-route-compatibility-check`

- The means of interaction needed for discovery, retrieval, integration, analysis, and understanding of the data – the data needs to be accessible and actionable.
- The time scale of data sharing, which can vary from days to many decades, e.g. compare the life cycle of a paper cup for coffee, with the lifetime of a glass window put into a building.
- Confidentiality of business data versus the need to share data to facilitate CE strategies, as well as trust in the accuracy and quality of the data, and security of the actual sharing.

Ongoing projects are pushing the limits, and testing current technologies, to explore to what extent they can meet these challenges. For instance, ontologies as a means of facilitating semantic interoperability in the CE are being explored, but there are many challenges in terms of modularity and flexibility of CE ontology networks that need further research. Similarly on the data sharing, and distributed KGs side, attempts are made to explore the use of Solid pods for CE data sharing, but limitations also of these technologies need to be overcome. We have a lot of research ahead of us, to really overcome the barriers for data sharing in the CE!

4.3 Personal Data Protection in Knowledge Graphs (and KR in general)

Piero Andrea Bonatti (University of Naples, IT)

Knowledge graphs are increasingly used to encode personal, possibly sensitive information. As a consequence, in order to fulfill application requirements and comply with personal data protection regulations, it is necessary to design and implement access control and anonymization mechanisms for restricting access to knowledge graphs.

Currently, standard KG implementations do not support such mechanisms. Moreover, most of the theoretical work on this topic does not really protect knowledge due to an improper choice of the goal of the protection mechanisms (i.e. the adopted *confidentiality* or *anonymization criterion*). This results in a false sense of security [1].

The main challenges in this area comprise:

- Raising the awareness about the limitations of the current approaches. Problems to be solved include (i) improving methodology, by requiring papers to specify explicit and nontrivial attacker models, and (ii) removing misunderstandings and misconceptions related to fundamental concepts such as indistinguishability and k-anonymity.
- Knowledge graphs make it easy to link different pieces of information and knowledge related to a same person, that are stored in a variety of distributed knowledge sources – KG have been expressly designed for this purpose. From a privacy perspective, such linkage is one of the most dangerous operations, as it results in the creation of rich personal profiles. This makes the protection of KG a particularly difficult task, that may potentially involve the processing of a wide set of distributed sources.
- Anonymizing KG is more difficult than anonymizing a single relational table. Generalizing the standard definitions of anonymity for relational DB is a nontrivial task. From a computational perspective, even *checking* that a graph is anonymous is computationally difficult.

We are currently working on a notion of k-anonymity for KG, using different notions of *suppressors* (the functions that anonymize KG), that range from uniform substitutions that turn some constants into blank nodes, to non-uniform substitutions that may create multiple copies of nodes and triples. We are studying the complexity of anonymity checking, of checking whether anonymizations exist, and the problem of finding optimal anonymizations (that hide a minimal amount of information). We have already identified a few tractable cases; in general, however, the complexity of the above problems ranges from GI (the complexity of the graph isomorphism problem) and NP.

Future work includes the investigation of: more kinds of suppressors (capable of deleting nodes and weakening triples); l-diversity and t-closeness; differential privacy for KG.

References

1 Bonatti, P. A false sense of security. *Artificial Intelligence.* **310** pp. 103741 (2022)

4.4 Private Knowledge Graph Construction

Carlos Buil-Aranda (TU Federico Santa María – Valparaíso, CL)

Knowledge Graph construction faces a challenge for providing private access to the data. Data stored in these graphs may come from private sources in which data is sensitive to certain groups of consumers but not to others. How to define what data can be shared with others, and provide means to do so is a challenge. It is a challenge since it is difficult to define what data can be shared when data is incomplete, and guaranteeing privacy constraints over unknown data is hard. I believe that we need means for defining what data can be shared privately, which is part of building knowledge graphs.

4.5 Challenges in Procedural Knowledge governance in Industry 5.0

Irene Celino (CEFRIEL – Milan, IT)

Procedural Knowledge (PK) is **knowing-how** to perform some task, to do a specific job; it is usually opposed to descriptive or declarative knowledge, i.e. *knowing-what* in terms of facts and notions. This type of knowledge is specifically relevant in industrial environments. The governance of PK shows different challenges: PK is **hard to explicitly articulate**, because it derives from *experience* and often it is *not documented*; PK is **hard to explain and describe**, because it may include *tacit* and/or *commonsense* knowledge; PK is **hard to access and retrieve**, because it is spread in many different sources.

I believe that solutions based on Knowledge Graphs can provide value to reuse and correctly execute procedures: once PK is digitised and opportunely structured in a KG, applications can be built to access the procedural KG and provide support or guidance to the people that need to understand and carry out such procedures. But how do we create and populate Procedural Knowledge Graphs? This a sort of *"cold start" problem* and can be addressed at two levels: (1) **PK extraction from unstructured data** and (2) **PK**

capture from humans. In my lightning talk, I discussed both "cold start" problems, highlighting on the one hand the challenges for a "fit-for-purpose" PK extraction, even with the latest and best performing AI solutions, and on the other hand the need to adopt approaches with the **direct involvement of people**, to take into account human and social phenomena.

I concluded by mentioning my current work on the design and development of proper tools to capture PK in the European project PERKS[13], which aims at providing support for the holistic management of PK lifecycle in Industry 5.0, with solutions based on Knowledge Graphs but also other AI technologies. PERKS goal is to provide industry workers with tools that prove to be: (1) **fit-for-purpose**, i.e. not "perfect", but acceptable quality for practical use, (2) **easy-to-use** and not time-consuming, and (3) mostly automated but with **human-in-the-loop**, i.e. employing advanced AI technologies, but leaving the final word to people.

4.6 Are Knowledge Graphs ready for the World... Wide Web?

Pierre-Antoine Champin (INRIA – Sophia Antipolis, FR)

Judging from the news and the scientific literature in the past decade, and how "Knowledge Graphs" are increasingly mentioned and advertised, it seems that Knowledge Graph are not only ready for the real world, but are well installed in it. The question remains, however, to know if they are ready for the web? This is also a relevant question, given the way the web has radically changed the world we live in, especially since the COVID pandemic.

The web is a decentralized, interconnected and interoperable data space. However many knowledge graphs are centralized, and developed and used behind closed doors. The technologies for building "Web Knowledge Graphs" (RDF, SPARQL, OWL...) are available and have been long studied, but their adoption did not always meet expectations.

What can we do to make these technologies ready for the real world – or to make the real world ready for them? The active RDF-star W3C Working Work is striving to improve interoperability between RDF and the popular Label Property Graph databases; the JSON-LD W3C Working Group is developing "symbiotic syntaxes" to allow RDF interpretations to co-exist with "traditional" interpretations of widespreads data format (JSON, YAML, CBOR). Web Knowledge Graphs technologies should not strive to replace other technologies, but to build bridges between them.

[13] http://www.perks-project.eu/

4.7 On the Need for Project Management for Knowledge Graph Construction and Usage Projects

Oscar Corcho (Universidad Politécnica de Madrid, ES) and David Chaves-Fraga (Universidade de Santiago de Compostela, ES)

Knowledge graph construction projects today require creating a diverse set of artefacts (OWL ontologies, SHACL shapes, declarative mappings, SPARQL queries, etc.). Most of these projects may look now like an art. However, they should become a proper engineering activity, where all artefacts are well controlled and maintained, all processes are well understood and systematised, and, in general, we can be sure that they can be easily maintained and replicated. Let's work on this and normalise how these projects are done in the future.

4.8 Grounding KGs in Natural Language

Christophe Debruyne (University of Liège, BE)

I commence my talk by arguing that the first question of this Dagstuhl Seminar should be rephrased. First, we should discuss "languages" instead of "programming languages" as the latter is too narrow. Secondly, I believe that "using knowledge graphs" should be replaced with "engaging with knowledge graphs" as the former is unilateral, and the latter implies a bilateral interaction between agents (human and computer-based) and knowledge graphs. The question thus becomes: *"What are the key requirements for language paradigms for modeling, representing, storing, engaging with, and managing KGs in the real world?"* One of the requirements is to include humans by adopting their language.

Knowledge graphs are designed for machines, not for humans. Humans engage with each other using natural language, evidenced by the popularity of generative AI to engage with information. Humans must be kept in the loop of constructing, maintaining, and using knowledge graphs, but we cannot expect them to become "KG-literate." My call to arms is to critically reflect on the role of humans in a KG "ecosystem," as reducing them to "users" would be a disservice to them. There have been initiatives in the past where people used controlled natural languages such as NIAM [1] and RIDL* [2]) for knowledge engineering and querying. These initiatives can be applied to knowledge graphs, so our community should consider learning from the past. De Leenheer et al. [3], for instance, adopted these principles for a knowledge engineering method where knowledge is declared and used separately (a principle called double articulation), which allows for knowledge to be used in different interrelated contexts, much like humans perceive things from different angles depending on their activity or task. This talk briefly mentions these principles to open the room for discussion.

References

1 Nijssen, G. & Halpin, T. Conceptual schema and relational database design – a fact oriented approach. (Prentice Hall,1989)
2 Troyer, O. RIDL*: A Tool for the Computer-Assisted Engineering of Large Databases in the Presence of Integrity Constraints. *Proceedings Of The 1989 ACM SIGMOD International Conference On Management Of Data, Portland, Oregon, USA, May 31 – June 2, 1989.* pp. 418-429 (1989), https://doi.org/10.1145/67544.66965

3 Leenheer, P., Moor, A. & Meersman, R. Context Dependency Management in Ontology Engineering: A Formal Approach. *J. Data Semant.*. **8** pp. 26-56 (2007), `https://doi.org/10.1007/978-3-540-70664-9%5C_2`

4.9 KG4SE & SE4KG: Exploring the Intersection of Knowledge Graph and Software Engineering Research

Coen De Roover (VU – Brussels, BE)

Treating source code as data is common in software engineering research. Moreover, many software engineering success stories bear resemblance to knowledge graphs. For instance, logic rules over a database of program facts have proven popular for implementing program analyses [3] as well as program queries [8, 2] that identify code of interest (e.g., design patterns [5, 6], defective code to be repaired [4], meta-level code making assumptions about base-level code [7], etc ...) in a project. The same goes for graph query languages over various graph-based representations. In my talk, I will showcase the latter through a graph-based representation of the control and data flow within Ansible Infrastructure-as-Code scripts [10], which enables detecting design and security smells [11] through straightforward graph queries. Another notable success story in software engineering is the creation and sharing of datasets through mining software repositories. Examples include datasets of Helm Kubernetes charts [9], of build and test results [1], of StackOverflow posts [12], ... The goal of my talk is to raise the question "What if true knowledge graphs were used in all these success stories (KG4SE)?", and also "What software engineering needs exist among knowledge graph engineers (SE4KG)?". This with the aim of stimulating discussion and fostering new insights at the intersection of software engineering and knowledge graph research.

References

1 Beller, M., Gousios, G. & Zaidman, A. TravisTorrent: Synthesizing Travis CI and GitHub for Full-Stack Research on Continuous Integration. *Proceedings Of The 14th International Conference On Mining Software Repositories (MSR 2017)*. pp. 447-450 (2017,5)
2 De Roover, C., Noguera, C., Kellens, A. & Jonckers, V. The SOUL Tool Suite for Querying Programs in Symbiosis with Eclipse. *Proceedings Of The 9th International Conference On The Principles And Practice Of Programming In Java (PPPJ11)*. (2011)
3 Bravenboer, M. & Smaragdakis, Y. Strictly declarative specification of sophisticated points-to analyses. *Proceedings Of The 24th ACM SIGPLAN Conference On Object-Oriented Programming, Systems, Languages, And Applications (OOPSLA09*. **44**, 243-262 (2009)
4 De Roover, C. & Inoue, K. The Ekeko/X Program Transformation Tool. *Proceedings Of 14th IEEE International Working Conference On Source Code Analysis And Manipulation, Tool Demo Track (SCAM14)*. (2014)
5 De Roover, C. A Logic Meta-Programming Foundation for Example-Driven Pattern Detection in Object-Oriented Programs. *Proceedings Of The 27th IEEE International Conference On Software Maintenance (ICSM11)*. (2011)
6 Molderez, T. & De Roover, C. Search-Based Generalization and Refinement of Code Templates. *Proceedings Of The 8th International Symposium On Search-Based Software Engineering (SSBSE16)*. (2016)

7 Fabry, J., De Roover, C. & Jonckers, V. Aspectual Source Code Analysis with GASR. *Proceedings Of The 13th International Working Conference On Source Code Analysis And Manipulation (SCAM13).* (2013)

8 Hajiyev, E., Verbaere, M. & De Moor, O. CodeQuest: Scalable Source Code Queries with Datalog. *Proceedings Of The 20th European Conference On Object-Oriented Programming (ECOOP06).* **4067** pp. 2-27 (2006)

9 Zerouali, A., Opdebeeck, R. & De Roover, C. Helm Charts for Kubernetes Applications: Evolution, Outdatedness and Security Risks. *Proceedings Of The 20th International Conference On Mining Software Repositories (MSR 2023).* pp. 523-533 (2023)

10 Opdebeeck, R., Zerouali, A. & De Roover, C. Smelly Variables in Ansible Infrastructure Code: Detection, Prevalence, and Lifetime. *Proceedings Of The 19th International Conference On Mining Software Repositories (MSR 2022).* pp. 61-72 (2022)

11 Opdebeeck, R., Zerouali, A. & De Roover, C. Control and Data Flow in Security Smell Detection for Infrastructure as Code: Is It Worth the Effort?. *Proceedings Of The 2023 IEEE/ACM 20th International Conference On Mining Software Repositories (MSR 2023).* pp. 534-545 (2023)

12 Baltes, S., Dumani, L., Treude, C. & Diehl, S. SOTorrent: reconstructing and analyzing the evolution of stack overflow posts. *Proceedings Of The 15th International Conference On Mining Software Repositories (MSR 2018).* (2018)

4.10 Challenges on Knowledge Graph Data Management Quality

Anastasia Dimou (KU Leuven, BE)

Data quality is not a new research topic that only affects the knowledge graph data management, but it is extensibly studied across various domains beyond knowledge graphs [1]. However, the survey on knowledge graphs quality [2] revealed that the dimensions of data quality for knowledge graphs extend far beyond data quality, while well-known aspects of data quality take new dimensions when they are examined within the context of knowledge graphs. Different vocabularies were proposed to describe the quality of knowledge graphs, such as DAQ [3], DQV [4], DQM [5], Similarly, different quality assessment frameworks were proposed to assess the quality of knowledge graphs, such as Sieve [6], RDFUnit [8] and Luzzu [7]. These vocabularies and frameworks made the need for a validation framework tailored for knowledge graphs evident.

Following this need, two prominent shapes languages were proposed: the W3C-recommended Shapes Constraint Language (SHACL) [9] and the Shapes Expressions Languaege (ShEx) [10]. Shapes are defined mostly manually by domain experts, however, as it is a time-consuming task, many approaches were proposed to extract shapes. Shapes are extracted from *RDF graphs*, e.g., QSE [11], SHACLGEN [12], ABSTAT [13], ShapeInduction [14], SHACLearner [15], or ShapeDesigner [16]. Shapes are also extracted from other artefacts that contribute to the construction of RDF graphs, such as ontologies [18], data schemas [19, 20] and mapping rules [21], or a combination of them [17].

Although considerable research has delved into defining shapes, developing efficient validation frameworks, providing explanations for the violations and repairing the knowledge graphs [22] or their artefacts, e.g., data, ontologies and mapping rules, after validation [25, 26, 24, 23] have received considerable less attention. There remains ample opportunity for further enhancement of systems and research.

References

1 Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. 2009. Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41, 3, Article 16 (2009). https://doi.org/10.1145/1541880.1541883

2 Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, Sören Auer: Quality assessment for Linked Data: A Survey. Semantic Web 7(1): 63-93 (2016)

3 Jeremy Debattista, Christoph Lange, and Sören Auer. daQ, an Ontology for Dataset Quality Information. LDOW 2014, CEUR, vol. 1184.

4 Riccardo Albertoni, & Antoine Isaac. Introducing the Data Quality Vocabulary (DQV). *Semantic Web*. **12**, 81-97 (2021,1), https://doi.org/10.3233/SW-200382

5 Fürber, C. & Hepp, M. Towards a vocabulary for data quality management in semantic web architectures. *Proceedings Of The 1st International Workshop On Linked Web Data Management*. pp. 1-8 (2011), https://doi.org/10.1145/1966901.1966903

6 Mendes, P., Mühleisen, H. & Bizer, C. Sieve: linked data quality assessment and fusion. *Proceedings Of The 2012 Joint EDBT/ICDT Workshops*. pp. 116-123 (2012), https://doi.org/10.1145/2320765.2320803

7 Debattista, J., Auer, S. & Lange, C. Luzzu—A Methodology and Framework for Linked Data Quality Assessment. *J. Data And Information Quality*. **8** (2016,10), https://doi.org/10.1145/2992786

8 Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R. & Zaveri, A. Test-driven evaluation of linked data quality. *Proceedings Of The 23rd International Conference On World Wide Web*. pp. 747-758 (2014), https://doi.org/10.1145/2566486.2568002

9 Knublauch, H., Allemang, D. & Steyskal, S. SHACL Advanced Features. (World Wide Web Consortium (W3C),2017,6), https://www.w3.org/TR/shacl-af/

10 Prud'hommeaux, E. Shape Expressions 1.0 Primer. (World Wide Web Consortium (W3C),2014,6), https://www.w3.org/Submission/2014/SUBM-shex-primer-20140602/

11 Rabbani, K., Lissandrini, M. & Hose, K. Extraction of Validating Shapes from Very Large Knowledge Graphs. *Proceedings Of The VLDB Endowment*. **16**, 1023-1032 (2023), https://doi.org/10.14778/3579075.3579078

12 Arndt, N. SHACLGEN. (GitHub), https://github.com/AKSW/shaclgen, Accessed on 20.09.2023

13 Spahiu, B., Maurino, A. & Palmonari, M. Towards Improving the Quality of Knowledge Graphs with Data-driven Ontology Patterns and SHACL. *Workshop On Ontology Design Patterns (WOP) At ISWC (Best Workshop Papers)*. **2195** pp. 52-66 (2018)

14 Mihindukulasooriya, N., Rashid, M., Rizzo, G., Garcia-Castro, R., Corcho, O. & Torchiano, M. RDF Shape Induction using Knowledge Base Profiling. *Proceedings Of The 33rd ACM/SIGAPP Symposium On Applied Computing*. (2017)

15 Ghiasnezhad Omran, P., Taylor, K., Rodriguez Mendez, S., Haller, A. & Others Towards SHACL learning from knowledge graphs. *Proceedings of The ISWC 2020 Demos And Industry Tracks: From Novel Ideas To Industrial Practice Co-located With 19th International Semantic Web Conference (ISWC 2020)*. 2721 pp. 94-99 (2020)

16 Boneva, I., Dusart, J., Fernández Alvarez, D., & Gayo, J. Shape Designer for ShEx and SHACL Constraints. *Proceedings Of The ISWC 2019 Satellite Tracks (Poster & Demonstrations, Industry, And Outrageous Ideas)*. 2456 pp. 269-272 (2019,10), https://hal.archives-ouvertes.fr/hal-02268667

17 Duan, X., Chaves-Fraga, D., Derom, O. & Dimou, A. SCOOP all the Constraints' Flavours for your Knowledge Graph. *Proceedings Of The 21th Extended Semantic Web Conference (ESWC)*. (2024)

18 Cimmino, A., Fernandez-Izquierdo, A. & Garcia-Castro, R. Astrea: Automatic Generation of SHACL Shapes from Ontologies. *European Semantic Web Conference*. pp. 497-513 (2020)

19 Duan, X., Chaves-Fraga, D. & Dimou, A. XSD2SHACL: Capturing RDF Constraints from XML Schema. *Proceedings Of The 12th Knowledge Capture Conference 2023*. pp. 214-222 (2023), `https://doi.org/10.1145/3587259.3627565`

20 Garcia-Gonzalez, H. & Labra-Gayo, J. XMLSchema2ShEx: Converting XML validation to RDF validation. *Semantic Web*. **11** (2020)

21 Delva, T., Smedt, B., Oo, S., Assche, D., Lieber, S. & Dimou, A. RML2SHACL: RDF Generation Taking Shape. *Proceedings Of The 11th On Knowledge Capture Conference*. pp. 153-160 (2021,12)

22 Ahmetaj, S., David, R., Polleres, A. & Šimkus, M. Repairing SHACL Constraint Violations Using Answer Set Programming. *The Semantic Web – ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23 – 27, 2022, Proceedings*. pp. 375-391 (2022), `https://doi.org/10.1007/978-3-031-19433-7_22`

23 Randles, A., Junior, A. & O'Sullivan, D. A Framework for Assessing and Refining the Quality of R2RML mappings. *22nd International Conference On Information Integration And Web-based Applications & Services (iiWAS '20)*. (2021), `https://doi.org/10.1145/3428757.3429089`

24 Paulheim, H. Data-Driven Joint Debugging of the DBpedia Mappings and Ontology. *The Semantic Web*. (2017), `https://doi.org/10.1007/978-3-319-58068-5_25`

25 Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., Hellmann, S. & Walle, R. Assessing and Refining Mappings to RDF to Improve Dataset Quality. *SEMANTIC WEB – ISWC*. (2015), `https://doi.org/10.1007/978-3-319-25010-6_8`

26 Heyvaert, P., Meester, B., Dimou, A. & Verborgh, R. Rule-driven inconsistency resolution for knowledge graph generation rules. *Semantic Web*. **10**, 1071-1086 (2019), `https://doi.org/10.3233/SW-190358`

4.11 Semantic Knowledge Graphs as a Foundation for Advanced AI Applications

Michel Dumontier (Maastricht University, NL)

Following on the FAIR data principles, the past decade has been characterised by improving access to data. The unforeseen success of large-language models (LLMs) have created new opportunities for natural-language based interaction with documents and structured datasets. However, LLMs are only clever token prediction systems and (currently) lack advanced reasoning capabilities that make them prone to making incorrect inferences to even well known answers.

Knowledge Graphs (KGs) offer a foundational knowledge representation for truth maintenance that should be a critical part of future AI solutions. Semantic KGs extend the capabilities of heterogeneous or informal knowledge graphs with automated inference, so as to check the consistency of encoded knowledge and for inferring additional knowledge beyond the facts in the data graph portion of it. However, effective collaboration on shared schemas and vocabularies continue to pose significant hurdles due to a lack of socio-technological infrastructure that links expertise in formal knowledge representation with everyday users and contributors. Future work should explore new paradigms that leverage advancements in generative AI technologies with semantic knowledge graphs while keeping humans in the loop to maximise discovery and minimise false knowledge generation.

Neurosymbolic methods offer the tantalising possibility to advance the state of the art by seamlessly integrating language understanding with zero-shot prediction and logical reasoning capabilities. Hybrid question-answering systems, leveraging LLMs, FAIR (Findable, Accessible, Interoperable, Reusable) KGs, and web services, represent the frontier in creating more intelligent, reliable, and accessible information systems. The convergence of major systems is part of an evolving landscape of data management that will shape the future of knowledge management and (AI-based) knowledge discovery.

4.12 "Real World" Considered Harmful

George Fletcher (TU Eindhoven, NL)

I study data and knowledge systems. Traditionally, this means working out formal foundations and, in tandem, designing and engineering systems, i.e., theory building and empirical investigation. In recent years there has been an increasing interest in the human context of data and knowledge systems, what can be called a humanistic turn in the discipline. This centering of people in our work calls for us to revisit the language we use when we talk about data and knowledge. I highlight two examples. First, we must deprecate the concept of "the user" as the catch-all stand-in for people. Instead we should talk first of "people" or "humans", and only after this talk about particular roles which people can play, e.g., citizens, customers, refugees, business analysts,... There is much more to the world (and of equal, if not greater, importance for effective data and knowledge systems) than just users. Second, we must deprecate the concept of "the real world". It is clear that one person's or community's interests are not necessarily always those of another person or community. And, much of the work of data and knowledge is about bridging (and sometimes integrating) these heterogeneous interests. To speak of "the real world" implies the privilege of one perspective over another, and obscures or even effaces these practical difficulties, hindering our work. Furthermore, the language of "real world" is pejorative, i.e., is actually a negative term. After all, the opposite of "real" is "fake", and naming the interests of others as fake only inhibits collaboration, thriving, and success. We should instead be using positive unifying language. We should be speaking instead of perspectives of interest: practices, communities, applications, application domains, and so forth. In other words, moving from talking of "the real world" to talking of "something someone or some community is interested in". We can capture this in a slogan: Real World Considered Harmful. The future of data and knowledge work is people and our always complicated and conflicting and converging and diverging and evolving interests. Our scientific terminology must evolve to keep up with this reality.

4.13 Reliable Knowledge Graphs Need Reliable Processes

Paul Groth (University of Amsterdam, NL)

In 1929, Frank Ramsey defined knowledge as a "set of beliefs that are true, certain and obtained by a reliable process." I would argue that this notion of reliable process is central to making sure that the knowledge captured in knowledge graphs is indeed high quality.

This is particularly important as the complexity of the systems that are used to construct and maintain knowledge graphs only increases. Moreover, the integration of large language models into knowledge graph systems means that they can evolve even more rapidly from both a content as well as functional perspective. Given that knowledge graphs are complex socio-technical systems, to obtain reliable knowledge graphs, we need a renewed focus on updating and codifying the principles and practices of their engineering. This includes aligning software engineering and knowledge engineering practice with a focus on developer user experience; improved knowledge engineering with a focus on better support for manual curation by a range of people and support for multiple modalities; and improved engineering with large language models that fully embraces the capabilities of these models. Lastly, we need to improve our engineering approaches to effectively support the social processes that undergird knowledge graphs, for example by supporting "getting people on the same page" and embracing lessons learned from community curated knowledge graphs such as Wikidata and Wikipathways. By focusing on the processes we can make sure that the term knowledge in "knowledge graphs" is not a misnomer.

4.14 Challenges for Knowledge Graphs (KGs) in an Era of Machines that Learn

Claudio Gutierrez (University of Chile – Santiago de Chile, CL)

We are living a revolution in machine development whose essence is the passage from machines that obey instructions to machines that learn. This is specially relevant for computing, and probably more relevant for areas that were in the "higher levels" of the hierarchy from the human to the machine (from human ideas, written requirements, (graphical) models, programming languages, machine languages, to machines). KGs can be viewed as a sort of integration of areas such as databases, knowledge bases, knowledge representation, systems architecture.

With the advent of machine learning, neural networks, and especially with LLMs today, the whole field needs to be rethinking. Large language models (for natural language, visual languages, videos, etc.) are reshaping completely the traditional hierarchy, in terms of tasks, people, disciplines and goals. We are entering a world in which the communication and interaction among humans and machines is becoming more and more integrated.

Today we need to think about the whole cycle of (digital) knowledge in terms of machines that learn (and thus deal with knowledge) as part of the environment. I would like to raise awareness of the relevance of rethinking what we were traditionally doing, and help reflect on how this new scenario is reshaping our field. I advance three points to contribute to this reflection.

1. **A change of view of the notion of KGs**. Before the "AI" wave, KGs were considered essentially artifacts (i.e. objects), and treated as such. After (namely, today), *KGs are better thought of as a methodology*, that is, a system of methods and tools that uses graphs and semantic machinery, but whose main value is a characteristic form of organizing the digital world of knowledge.
2. **An interpretation of the sources of KGs**. The Semantic Web's original idea (which in a great degree precludes KGs topics) was *to deploy a universal infrastructure where*

agents (machines) can act. In the early 2000's this meant building languages (logical languages, taxonomies, ontologies, etc.) so that machines can understand each other (T. Berners-Lee) These ideas plus the "natural" space (the Web) to deploy them, triggered semi-structured languages, tree-like languages, and finally graph-like languages. In parallel, graph databases were being developed to address these (and other) concerns in a world of big data. In 2010, the notion of KGs embraced most of these ideas as a natural way of thinking and modeling a digital world that was increasingly populated by artifacts (machines, humans: the nodes) that exchanged information and knowledge (the arcs) .

3. **Surfing a tsunami**. In parallel (IMHO, slowly considered by our community), the neural network boom was exploding. This tsunami displaced most of the other concerns of the last decades. We still need to understand the internals of this phenomenon. I believe we should *be open to reengineering the original objectives of the area*: autonomous digital agents supporting as stewards, services, negotiators, and generators of new ideas in a space of digital information and knowledge. I dare to suggest two lines of work in this direction: (1) Develop infrastructure and languages for the standards of explainability that will allow us to realize the idea of agents negotiating and exchanging all over the world. (2) Develop the graph infrastructure for assembling LLMs and AI agents (graphs as ...).

4.15 Knowledge Graphs in the Real World with metaphactory

Peter Haase (Metaphacts GmbH – Walldorf, DE)

In this presentation I shared the metaphacts perspective on "knowledge graphs in the real world", based on the past ten years of experience with our enterprise knowledge graph platform metaphactory. I covered the things that are well established and proven to work as well as open challenges and areas of research.

Our approach to developing knowledge graph applications follows an agile and iterative process that allows to create value very rapidly. Business users and domain experts are involved throughout the process, which starts with pinpointing the information needs of the users. Domain experts together with ontology engineers then collaboratively define a knowledge model. This is supported by the semantic modeling environment of metaphactory. The knowledge model is validated with business users in an iterative manner. The model is then used to connect and integrate relevant data sources. Using the application building components of metaphactory, the user experience is built by translating end-user information needs into intuitive, model-driven interfaces.

Based on the experience from dozens of knowledge graph projects I summed up the state-of-the-art in the "real world":

- Semantic knowledge modeling is successfully used in enterprises. While the modeling is still largely done by expert users, new tooling like metaphactory has significantly lowered the entry barrier for non-experts.
- Data integration using ontologies is well established and becomes more and more model-driven, with declarative pipelines for knowledge graph construction. It still is largely based on materialization (as opposed to virtualization).
- Low-code, model-driven user interfaces provide a sustainable approach to application development.

I concluded with stating open challenges:

- For a large-scale adoption of semantic knowledge modeling in enterprises, we need tooling that considers aspects of governance in large organizations.
- There is large potential in hybrid approaches for data integration, combining materialization and virtualization, but scalable federated query processing is still a challenge.
- Enterprise knowledge graphs are the way to go to move from an application-centric to a data-centric enterprise, but more work needs to be done to support a multitude of diverse applications and to properly consider security and access control.

4.16 Improving UX/DX in Querying Federations of KGs

Olaf Hartig (Linköping University, SE)

One of my areas of research is on the virtual integration of knowledge graphs (KGs), with a focus on querying federations of KGs. While there has been research on this topic in the past, a question to ask in the spirit of the theme of this Dagstuhl Seminar is whether the approaches and systems developed in this context are ready for the real world? My answer to this question is: no! In the context of a COST Action on Distributed KGs, we recently organized a three-days hackathon on query federation over KGs to which we invited i) people who had published on the topic and had built some query federation engine and ii) use case providers who prepared use cases in which they thought these engines can be applied. This event was great fun, but the outcomes were quite sobering. For most of the use case providers, a major challenge was how to even write the queries that make sense, and this was not because of a lack of understanding of the query language but because of a lack of visibility into how the data in the federation members is shaped and how exactly it is connected across the federation. Once they had some queries, and they had managed to set up an engine that they wanted to play with, the next major issue was to understand what was going on when something went wrong with the queries. So, overall the user experience was quite disappointing for the use case providers in this hackathon. A challenge that I want to work on is to improve the user experience of developers who are considering to employ query federation in practice.

4.17 Can Large Language Models improve the Usability of Knowledge Graphs?

Aidan Hogan (University of Chile – Santiago de Chile, CL)

Large Language Models have left many wondering what the future is for Knowledge Graphs. Will Knowledge Graphs still be needed in the future? Can they be combined with Large Language Models in some useful way? In this lightning talk, I will put forward the proposal that these are both complementary technologies, whereby Knowledge Graphs can be used to improve the reliability and factuality of Large Language Models, with Large Language

Models can be used to potentially address one of the fundamental issues of Knowledge Graphs: a lack of usability. To illustrate this idea, I will show examples of questions that ChatGPT currently fails to answer, and show how such questions can be answered on the Wikidata Knowledge Graph. However, answering questions on Wikidata requires expert knowledge of languages such as SPARQL, where, in the creation of such SPARQL queries, Large Language Models can certainly be of use. This provides a concrete but simple idea of how Large Language Models and Knowledge Graphs could complement each other in future.

4.18 No Intelligence without Knowledge

Katja Hose (TU Wien, AT)

While large amounts of data are being generated and collected, we are still struggling to efficiently exploit, interpret, and extract meaningful insights from such heterogeneous data. Because of their ability to capture factual knowledge, knowledge graphs are applied in versatile settings to organize, integrate, share, and access information. In this sense, also Machine Learning approaches, such as LLMs, benefit from the access to factual knowledge provided by knowledge graphs. In doing so, they can help mitigate hallucinations by providing access to verifiable knowledge, reliable facts, patterns, and enable a deeper understanding of the underlying domains. However, any approach can only be as good as the data it is built (or trained) upon ("garbage in, garbage out"). Hence, the quality of knowledge graphs needs to be ensured, e.g., by using SHACL constraints, access to a knowledge graphs needs to be efficient, evolving knowledge needs to be captured and made accessible, provenance needs to be available, etc. It will be exciting to witness the advances in this exciting field in the next years.

4.19 Lowering the Barriers for Declarative Knowledge Graph Construction Adoption

Ana Iglesias-Molina (Universidad Politécnica de Madrid, ES)

Declarative KG construction has undergone a great process of improvement and optimization in the last decade. Currently, a wide ecosystem of resources, languages and compliant tools, is available with an active community around maintaining and enhancing them. Despite this progress, there is still room for improvement, not only to make them as versatile as ad-hoc approaches, but especially in boosting their usability, since there is still some resistance to their adoption. One of the main reasons behind it is that these technologies pose a learning curve for users to use them, as it is not a familiar environment. Hence, there is room for improving the user involvement, providing understandable outputs and keeping the process seamless without adding any overhead. The rise of LLMs and their success when interacting with users opens the door to multiple possibilities for addressing this issue, since traditional approaches (user friendly interfaces and serializations) have had limited success. Not only can we help facilitate the process for KG construction, but also in several additional tasks and steps involved in the knowledge graph life cycle.

4.20 Knowledge Graph Adoption: Unveiling User Perspective and Challenges

Samaneh Jozashoori (Metaphacts GmbH – Walldorf, DE)

Knowledge graphs are already part of the "reality" of several enterprises. One of the challenges that I've observed in these enterprises is onboarding "knowledge stewards" to adopt knowledge graphs as the solution. To be precise, we (at metaphacts) define knowledge stewards as those whose role is to curate ontologies, vocabularies, and instance data. Like any change adoption in an enterprise, we can consider the following primary elements as the reasons for resistance and discuss the potential contributions of the enterprise and the scientific community in addressing them:

1. Lack of urgency or vision: If the enterprise did not have a data integration solution previously, knowledge stewards may lack the urgency and necessity to adopt knowledge graphs. On the other hand, if there has been a previous failed solution for data integration problems, knowledge stewards will be skeptical that a knowledge graph can be "the" solution.
2. Lack of motivation for adoption: Adopting a simple, unfamiliar technology is more difficult than using a complex, familiar system. So the more familiar knowledge stewards become with knowledge graphs and the technology surrounding them, the easier the acceptance process becomes.
3. Insufficient confidence: One of the main concerns knowledge stewards express is the quality of their models. For many new knowledge stewards, especially the majority who are just starting out, grasping the fundamentals of modeling doesn't always conduct confidence. The question remains, how can the scientific community contribute to overcoming these obstacles.

4.21 Semantic Reflection

Eduard Kamburjan (University of Oslo, NO)

While knowledge graphs and ontologies are eminently useful to represent formal knowledge about a system's individuals and universals, programming languages are designed to describe a system's evolution. To address the dichotomy, we use a mapping that lifts the program states of an object oriented programming language into a knowledge graph, including the running program's objects, fields, and call stack. The resulting graph is exposed as a semantic reflection [1] layer within the programming language that can be accessed from the very same program that is lifted, allowing programmers to leverage knowledge of the application domain in their programs. We formalize semantic lifting and reflection for a core programming language, SMOL, to explain the operational aspects of the language, and consider type correctness and virtualisation for program queries through the semantic reflection layer. We illustrate the approach by a case study of geological modeling [2]. The language implementation is open source and available online under `http://www.smolang.org`.

References

1 Kamburjan, E., Klungre, V., Schlatte, R., Johnsen, E. & Giese, M. Programming and Debugging with Semantically Lifted States. *ESWC*. **12731** pp. 126-142 (2021)

2 Qu, Y., Kamburjan, E., Torabi, A. & Giese, M. Semantically triggered qualitative simulation of a geological process. *Applied Computing And Geosciences*. **21** pp. 100152 (2024)

4.22 Exploiting Semantics for Integrating Data on Critical Minerals

Craig A. Knoblock (USC – Marina del Rey, US)

Critical minerals like lithium, cobalt, and nickel are essential for transitioning to a green economy but are in short supply, necessitating a search for more domestic sources. We are working on building a knowledge graph to create grade and tonnage models of various commodities, leveraging artificial intelligence and machine learning to automate the laborious manual process of assembling data from diverse sources. The technical challenges in this endeavor are substantial, including locating source information, accurately extracting information from text documents using Large Language Models (LLMs), automatically modeling tables, precisely linking entities, and performing efficient spatial/temporal queries. These challenges stem from the need to handle diverse data types and formats, ensure data accuracy and consistency, and integrate information from disparate sources to build comprehensive and accurate models. The project aims to rapidly build and maintain high-quality models, enabling timely updates as new information becomes available and supporting the identification of critical mineral resources.

4.23 Knowledge-enhanced Representation Learning to Accelerate Scientific Discovery

Vanessa López (IBM Research – Dublin, IE)

Foundation models mark a significant advancement in AI, not just for natural language processing, but they also have a great potential to unlock scientific discoveries. This potential extends particularly to domains such as drug discovery, where these models can assist experts in, for example, identifying suitable drug small molecules from a large pool of candidates that may bind to protein targets causing disease. While scientific data is highly multimodal, many models have largely remained unimodal. Key challenges in representation learning include effectively utilizing multimodal information and achieving multimodal fusion. In this context, Research on Multimodal Knowledge Graphs (MKG) is finding increasing applications beyond language modeling and computer vision into the biomedical domain. KGs are often used to understand the underlying complexity of the underlying data and combine rich factual knowledge from heterogeneous sources. In a MKG, entities and attributes may convey information about their modality, with typical examples including text, protein sequences, SMILES, images, 3D structures, numerical and categorical values. As such, MKG can capture correspondences between multimodal entities and attributes through labeled relations.

Recent approaches such as OtterKnowledge [1] use Graph Neural Networks (GNNs) for learning representations from MKGs. Otter Knowledge leverages existent encoders (single modal foundation models) to compute initial embeddings for each modality, and learns how to transform or fuse different modalities based on the rich neighborhood information for each entity. During inference, these knowledge-enhanced pre-trained representations are applied to downstream tasks, such as predicting binding affinity between protein and molecules. Essentially, this system aligns the representation spaces of an arbitrary number of unimodal representation learning models through a multi-task learning regime. The key for the multi-task learning involves building a MKG describing each of the entities (e.g., proteins, drugs, or diseases), how they interact with each other, and what their multimodal properties are (e.g., protein sequence, structure, functional annotations as gene ontology terms, or descriptions).

There are many opportunities and challenges to advance life science discovery by democratizing vast human knowledge accumulated in human-curated multimodal sources, and incorporating that knowledge into AI-enriched multimodal models. Knowledge-graphs can serve as a powerful tool to integrate a broader range of heterogeneous data and modalities. In turn, knowledge-enhanced multimodal representations may improve foundation models for predictive downstream tasks and hypothesis generation in discovery domains, addressing the question of whether approaches like this can lead to success in real-life applications where single-modal methods fail to learn something new about the natural world.

References

1 Hoang, T., Sbodio, M., Galindo, M., Zayats, M., Fernandez-Diaz, R., Valls, V., Picco, G., Berrospi, C. & Lopez, V. Knowledge Enhanced Representation Learning for Drug Discovery. *Proceedings Of The AAAI Conference On Artificial Intelligence.* **38**, 10544-10552 (2024)

4.24 Using LLMs to Augment KG Construction

Paco Nathan (Derwen – Sebastopol, US)

The goal of this project is to improve semi-automated KG construction from large collections of unstructured text sources, while leveraging feedback from domain experts and maintaining quality checks for the aggregated results. We explore hybrid applications which leverage LLMs to improve natural language processing (NLP) pipeline components, which are also complemented by other deep learning models, graph queries, semantic inference, and related APIs. In contrast to "black-box" methods of using chat agents to generate KG data, we focus on how LLMs can be used to augment specific, well-defined tasks, while maintaining quality checks? To this end, we consider where can the data (training datasets, benchmarks, evals) be reworked to improve performance on tasks, among the research projects evaluated here? Also, it would be intractable in terms of time and funding to rewrite code and then re-evaluate models for the many research projects which are within the scope of this work. Therefore reproducibility of published results – based on open source code, models, evals, etc. – becomes crucial for determining whether other projects may be adapted for production use. We propose a rubric used for this evaluation process.

4.25 Queries over Evolving Knowledge Graphs

Edelmira Pasarella (UPC Barcelona Tech, ES)

Nowadays data are in motion, continuously changing, (possibly) unbounded, which means data sources are constantly evolving. This breaks the paradigm –from the data persistence point of view – of always having dynamic but stable data sources. This, together with the increasing number of real-time applications for making critical decisions based on a data stream, raises the need for re-thinking the data and the query models for these new requirements. We tackled the problem of querying evolving knowledge graphs, i.e., graphs having *volatile relations*. These kinds of relations are created according to a data stream by feeding the knowledge graph at a time interval. Once this time interval is over, these relations cease to be valid and must be removed. The evaluated queries over data streams are known in the literature as continuous queries.

Our approach for querying evolving knowledge graphs is twofold. First, we use a data model based on a knowledge graph data source and decompose its relations into stable and volatile relations. Volatile relations induce subgraphs that exist while these relations are valid and, hence the knowledge graph remains constantly evolving. Secondly, we evaluate continuous queries following a stream processing technique –the dynamic pipeline computational model. The dynamic pipeline approach allows for emitting answers as query patterns are identified in the knowledge graph.

Stream based applications are direct beneficiaries of our proposal because they can query knowledge graphs and get answers from "fresh" data as they are produced and avoid the computational overhead of discarding non-valid data.

4.26 From Traditional Data Quality to Knowledge Graph Quality

Anisa Rula (University of Brescia, IT)

In recent years, there has been a notable increase in the creation of large-scale and interconnected knowledge graphs within both academic and industrial domains. However, the diverse quality of these knowledge graphs presents significant challenges for researchers and practitioners alike. First, this challenge is exacerbated by the transition from conventional data quality assessment paradigms to the unique characteristics of knowledge graph quality evaluation, owing to the structural differences inherent in the new data model. It is essential to critically examine how quality metrics have evolved from traditional relational databases to knowledge graphs. This involves identifying which metrics persist, adapt, or emerge anew in the context of knowledge graphs.

Second challenge regards, the absence of a standardized benchmark for evaluating knowledge graph quality. Therefore, it is necessary to contextualize quality within a scenario which underscores the importance of developing tailored assessment methods that consider the specific objectives and characteristics of individual knowledge graphs. A key consideration in quality assessment is the concept of linkable data, where the quality of individual datasets significantly influences the effectiveness of integration efforts. The richer the semantic content of datasets, the better the quality of the data integration, suggesting that quality assessment could be seen as an optimization problem.

Third, most of these approaches are focused on the quality assessment of datasets and not on the quality assessment of mappings used to transform tabular data to Knowledge Graphs. To deal with low-quality data, one may need to revise the procedures of generating those knowledge graphs. As such, the quality of knowledge graphs is not only contingent on the input and output data but also on the quality of the mappings.

Last but not less important, is the challenge of quality interpretation. In this context, leveraging advanced technologies such as Large Language Models becomes crucial for interpreting quality results and extracting meaningful insights from complex knowledge graph structures.

4.27 The Future of Knowledge is Social

Juan F. Sequeda (data.world – Austin, US)

My position for this seminar is three-fold:

1. **Data and Knowledge management is a SOCIO-technical phenomena**. We have been focusing mainly on the technical side for the past 30 years and doing the same for different results; Einstein's definition of insanity. The data and knowledge community has ignored the socio side. That's the paradigm shift we need.
2. **Generative AI / Large Language Models are now an important motivation to invest in Knowledge, thus also, a motivation for Knowledge Graphs**. Enterprises (i.e. the "real world") are realizing that without investing adequately in knowledge they can't use AI applications in production because it lacks accuracy and explainability. We must seize this moment. In our research we have found evidence that enterprise question answering on enterprise SQL databases becomes 3X more accurate if it's over a Knowledge Graph!
3. **We must update/upgrade Knowledge Engineering Methodologies**. These methodologies have existed for 30+ years, but we must adapt them to the status of today, including also using LLMs as copilots to the knowledge engineers. The methodologies must focus on people and process and of course technology that uses AI/LLM. In our current research, we are taking every step in the methodology and analyzing how we can automate it with LLMs.

Please ask yourself the following questions:

- When was the last time you spoke to a user?
- What does accuracy mean?
- Explainable to whom?
- How do you define "real world" and how do you define success?
- How do we educate the current and next generation of data professionals so they understand the value of investing in Knowledge?

4.28 Towards Resource-efficient and Hybrid KG Construction Approaches

Dylan Van Assche (Ghent University, BE)

Knowledge Graph construction approaches integrate heterogeneous data sources from files, databases, Web APIs, and many more. Either by materializing the Knowledge Graph or providing virtualized access to it. Each of these approaches have their set of trade-offs regarding resource consumption and execution time. However, there is no combined hybrid approach which leverages the best of both materialization and virtualization. In this talk, we open the discussion of combining materialization and virtualization approaches for Knowledge Graph construction. How can we combine materialization and virtualization in a hybrid fashion to make Knowledge Graph construction more efficient in terms of resources e.g. CPU time, memory usage, storage usage, and execution time? This way, we open the path towards a more resource-efficient Knowledge Graph construction approach which may also be suitable for smaller embedded devices e.g., smartphones, laptops, and many more.

4.29 MisLED: Linked Enterprise Data Left without Research Attention

Ivo Velitchkov (Brussels, BE)

Semantic technologies fail the 4U test[14]. They work and are useful but are barely used[15]. The focus of the research community on open data has led to little adoption in enterprises. Even the Linked Open Data (LOD) publishers don't use Linked Data themselves. Because of that, they perceive LOD mainly as cost. In addition, not using daily applications with an RDF backend excludes them from influencing the evolution of semantic web technologies, thus producing a vicious cycle. At the same time, the contemporary corporate landscape is shaped by the application-centric mindset, leading to a high cost of data integration, high-cost change and accumulation of technical debt. The standards for creating semantic knowledge graphs are well-suited to solve this problem by supporting its alternative, data-centric architectures[16]. Yet, the required technologies for a data-centric digital transformation via Linked Enterprise Data (LED), transforming data, applications, and access control, respectively, are developed disproportionately. The knowledge construction has advanced, but RML is not yet on the W3C standards track. There are RDF libraries, but transforming enterprise applications to work with RDF backend remains a challenge. The worst is the situation with access control. While there are some approaches, there is no standard declarative way to transform the access control configuration from the current applications to the enterprise knowledge graph.

The momentum currently created by the data-centric movement presents one more opportunity for the research community to shift its focus and contribute to increasing the adoption of LED-based enterprise knowledge graphs.

[14] https://www.linkandth.ink/p/4u2p
[15] https://www.strategicstructures.com/?p=2193
[16] http://datacentricmanifesto.org

4.30 Data and Knowledge Management in Knowledge Graphs

Maria-Esther Vidal (Leibniz University of Hannover & TIB-Leibniz Information Centre for Science and Technology, DE)

Knowledge graphs (KGs) represent the convergence of data and knowledge using a graph data model. Although coined by the research community for several decades, KGs are currently playing an increasingly relevant role in scientific and industrial domains. In particular, the richness of data in encyclopedic KGs such as DBpedia [1], Wikidata [2], ORKG [3] or domain-specific KGs (e.g. Bio2RDF [4]) demonstrates the feasibility of data integration following Linked Data principles.

The scientific and industrial communities have responded to the emerging field of KG management [5]. As a result, formal frameworks for defining and representing KGs and methods for creating, exploring, and analyzing KGs have flourished to make KGs a reality. However, despite the noticeable results, KG management is cumbersome, thus preventing full adoption and commonality [6].

This talk presented the challenges of knowledge management and data integration. First, we reviewed the state of the art, put the challenges faced in these real-world applications into perspective, and discussed the limitations of current approaches. Specifically, we discussed the need for programming paradigms for KM management, transparent data integration and quality assessment techniques, and scalable and interpretable approaches to knowledge exploration. We concluded with a road-map for making KGs usable in the real world and supporting the needs of the different users who play a role in KGs.

References

1 Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P., Hellmann, S., Morsey, M., Kleef, P., Auer, S. & Bizer, C. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web.* **6**, 167-195 (2015), `https://doi.org/10.3233/SW-140134`

2 Vrandecic, D. & Krötzsch, M. Wikidata: a free collaborative knowledgebase. *Commun. ACM.* **57**, 78-85 (2014), `https://doi.org/10.1145/2629489`

3 Brack, A., Hoppe, A., Stocker, M., Auer, S. & Ewerth, R. Analysing the requirements for an Open Research Knowledge Graph: use cases, quality requirements, and construction strategies. *Int. J. Digit. Libr..* **23**, 33-55 (2022), `https://doi.org/10.1007/s00799-021-00306-x`

4 Belleau, F., Nolin, M., Tourigny, N., Rigault, P. & Morissette, J. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Informatics.* **41**, 706-716 (2008), `https://doi.org/10.1016/j.jbi.2008.03.004`

5 Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A. & Taylor, J. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM.* **62**, 36-43 (2019), `https://doi.org/10.1145/3331166`

6 Li, H., Appleby, G., Brumar, C., Chang, R. & Suh, A. Knowledge Graphs in Practice: Characterizing their Users, Challenges, and Visualization Opportunities. *IEEE Transactions On Visualization And Computer Graphics.* **30**, 584-594 (2024)

5 Breakout Groups

5.1 Access and Usage Control for Federations of Knowledge Graphs

Piero Andrea Bonatti (University of Naples, IT)
Pierre-Antoine Champin (INRIA – Sophia Antipolis, FR)
Anastasia Dimou (KU Leuven, BE)
Peter Haase (Metaphacts GmbH – Walldorf, DE)
Olaf Hartig (Linköping University, SE)
Katja Hose (TU Wien, AT)
Sabrina Kirrane (Wirtschaftsuniversität Wien, AT)
Edelmira Pasarella (UPC Barcelona Tech, ES)

5.1.1 Introduction

The European Union (EU) has introduced a number of regulations and directives that have a major impact on access to and usage of personal, corporate, and governmental data. The overarching goal of these legal acts is to provide legal certainty with respect to the provision and consumption of data, as well as clarifying usage rights and obligations. Several of these legal acts are particularly relevant when it comes to Knowledge Graphs (KGs):

- The General Data Protection Regulation (GDPR)[17] provides citizens with control over the processing and sharing of data that concerns them. The regulation stipulates the legal bases under which personal data may be processed and requires data controllers and processors to provide transparency with respect to this processing. Given that KGs are particularly suitable for encoding data in an unambiguous manner, there have been several initiatives (SPECIAL[18], TRAPEZE[19], DPVCG[20]) that have employed KGs to assist with personal data processing, risk assessment, and compliance checking. Additionally, data controllers and processors (located in Europe or processing personal data about European citizens) who encode personal data in KGs are bound by the GDPR.
- The Open Data Directive[21] strives to make public sector data (e.g. geospatial, environmental, meteorological, statistical, mobility) available with the dual goal of ensuring transparency and facilitating re-use and exploitation of public sector data by governments, businesses, and individuals. KGs are at the core of many open data initiatives. For instance, the open source software of the CKAN open data portal[22], which is used by several governments, has been extended[23] such that it is possible to expose and consume metadata from other catalogs using Resource Description Framework (RDF) documents.
- The primary objective of the copyright and related rights in the Digital Single Market directive[24] is to harmonize copyright legislation across Europe, thus providing clarity for copyright holders (with respect to literary, dramatic, musical, and artistic works) and the

[17] https://eur-lex.europa.eu/eli/reg/2016/679/oj
[18] https://specialprivacy.ercim.eu/
[19] https://trapeze-project.eu/
[20] https://www.w3.org/community/dpvcg/
[21] https://eur-lex.europa.eu/eli/dir/2019/1024/oj
[22] https://ckan.org/
[23] https://github.com/ckan/ckanext-dcat
[24] https://eur-lex.europa.eu/eli/dir/2019/790/oj

users of copyrighted material across borders and in digital environments. Here, again, the KG community has demonstrated how KGs can be used in order to encode common licenses or facilitate custom license generation (DALICC[25]). Additionally, given that the default license in many EU member states is all rights reserved, KGs themselves should come with clear licensing terms and conditions.
- The Data Governance Act[26] stimulates the reuse of public data that is not considered open data, for instance data relating to citizens or businesses that could be used with the consent of the data subject or owner. The regulation also foresees the setup of data intermediaries that will act as trustworthy entities that facilitate data sharing and pooling. KGs are particularly suitable for sharing and reuse as entities and relations are specified in a manner that facilitates integration and extension.
- The Data Act[27] builds upon the Data Governance Act by providing legal certainty to both data producers and data consumers with respect to who can use what data for which purposes. The regulation also aims to ensure a fair playing field in terms of data access with the help of model data sharing contracts and measures to prevent vendor lock-in by data processing services. Here again there is a need to encode usage rights using clear unambiguous formalisms that facilitates compliance checking.
- The proposed Artificial Intelligence Act[28] aims to unlock the potential of artificial intelligence by putting in place a risk-based assessment process that will facilitate secure, trustworthy, and ethical artificial intelligence. The term artificial intelligence includes both symbolic and sub-symbolic processing. Given that the act imposes regulatory burdens if AI systems could potentially infringe upon fundamental rights and safety, it spans many domains (e.g., critical infrastructure, education, employment, public and private services, law enforcement). Thus, many vendors of products and services built on top of knowledge graphs will need to adhere to legal requirements and obligations stipulated in the AI Act.

The various acts aim to stimulate the digital economy by making data available for re-use, while at the same time ensuring that data subjects, data owners, and data publishers can control how data is used. Consequently, both access and usage control are at the core of these acts. Access control is concerned with permitting or prohibiting access to the data or systems based on some credentials or attributes. In turn, usage control is concerned with ensuring that data is used in accordance with specified terms and conditions (e.g., privacy preferences, intellectual property rights) after access has been granted.

In this vision of a digital economy, the value of data increases when data from multiple sources can be combined and integrated. During the seminar we discussed virtual integration settings in which such a combination of multiple KGs can be achieved by means of federated query processing. Federated query processing focuses on scenarios in which multiple data sources operate as fully autonomous members in a federation architecture in which a mediator component provides the functionality to answer global queries over the virtual union of the data of all federation members. While the technical aspects of federations of KGs have already been studied, the impact of access and usage control on these federated architectures has been mostly unexplored.

[25] https://www.dalicc.net/
[26] https://eur-lex.europa.eu/eli/reg/2022/868/oj
[27] https://eur-lex.europa.eu/eli/reg/2023/2854/oj
[28] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

5.1.2 State of the Art

Access control for RDF-based KGs has been extensively studied and consequently there have been many different proposals. A comprehensive survey by Kirrane et al. [1] categorizes existing proposals for access control specification, enforcement, and administration. The predominant specification approaches define access control policies based on triple patterns or views. Enforcement is facilitated either by rewriting SPARQL queries such that access to certain patterns or views is permitted / prohibited or alternatively by creating a filtered dataset according to the policies that only contains permitted data. Although there has been less emphasis on access control policy administration, researchers have considered the delegation of access rights, the consistency and safety of access policies, as well as usability and understandability. Many of the existing works have been inspired by traditional access control approaches that were originally applied to relational databases. For instance, the authorization framework proposed by Jajodia et al. [3] has been adapted to demonstrate how graph patterns, propagation rules, conflict resolution policies, and integrity constraints together can be used as a blueprint for access control specification and enforcement [4]. Additionally, there has been some work that extends static access control with dynamic context awareness [2].

There has also been a body of work focusing on providing access to knowledge, while at the same time protecting personal data by removing both direct and indirect personal identifiers. For instance, Radulovic et al. [5] and Heitmann et al. [6] demonstrate how k-anonymity can be applied to RDF-based KGs. Whereas Grau and Kostylev [8] propose a privacy model that encompasses notions of safe and optimal anonymization. Safety is used to provide protection against linking attacks, while optimality strives to release as much knowledge as possible without negatively affecting safety. When it comes to the execution of statistical queries over KGs, Silva et al. [7] demonstrate how differential privacy can be applied to RDF-based KGs and propose a mechanism for computing relevant parameters for the differential privacy mechanism. More recently, Buil-Aranda et al. [9] investigate how new differential privacy techniques that are particularly suitable for various SQL joins can be applied to SPARQL counting queries.

Continuing on the topic of personal data protection, although the GDPR outlines several different legal bases that provide legal grounds for the collection and processing of personal data, consent is particularly important as it gives data subjects control over their personal data and provides data controllers and processors with access to data they would otherwise not be permitted to collect and process. In this context, there have been several papers that demonstrate how OWL-based policies can be used to encode consent, legal requirements, and business policies, and demonstrate how KG-based tools can be used for automated compliance checking [10, 12]. The language was subsequently extended to support sticky policies [13]. An extensive overview of existing KG-based approaches for representing and managing content is provided by Kurteva et al. [11].

Existing work focusing on safeguarding intellectual property rights promotes reuse by proposing machine readable licenses and facilitates legal compliance by ensuring compliance. Both Rodríguez Doncel et al. [14] and Pellegrini et al. [16] demonstrate how the W3C Open Digital Rights Language (ODRL) can be used to encode licenses and provide access to a catalog of machine readable licenses. These licenses can be used to facilitate license identification and derivation, as well as conflict identification and resolution [16, 15]. In the case of the latter, Moreau and Serrano-Alvarado [17] demonstrate how federated SPARQL queries executed over several RDF-based KGs, can be relaxed such that the query result complies with licenses associated with the respective KGs.

In terms of research on queries over federations of KGs, the main focus of existing work has been on developing efficient query processing techniques [20, 21, 22, 19]. The majority of the state-of-the-art approaches in this context assume that the data of the federation members can be accessed and used without any restrictions. Yet, a few approaches to enforce policies during the execution of queries over a federation of KGs exist: The SAFE approach by Khan et al. [23] takes into account access control policies during the source selection phase of query planning, where these access control policies focus on authorizing specific users to access and retrieve designated subgraphs of the knowledge graph of a federation member. In contrast, the BOUNCER approach by Endris et al. [24] generates query execution plans by taking into account general (not user-specific) policies that permit or prohibit particular types of operations for designated attributes in the data of a federation member, where these operations are the retrieval of the respective attribute value and joins on the attribute values. Hence, the BOUNCER approach is more fine-grained in terms of the operations considered in the policies, but it does not distinguish between different users as done by the SAFE approach. Moreover, none of these approaches considers policies as described by standard policy languages for usage control such as ODRL.

5.1.3 Current Challenges

The challenges and the open research questions in this area can be roughly divided into four groups:

1. Lack of standard access control mechanisms in available implementations of KG systems;
2. Awareness/methodological issues related to security and privacy in the knowledge representation area at large;
3. Challenges in generalizing security and privacy approaches to KGs;
4. Challenges specific to federations of KGs.

Point 1 is essential to set up more sophisticated protection mechanisms for inference and usage control. It may be addressed by referring to the rich arsenal of policy models and enforcement mechanisms that can be found in computer security – in this respect we do not see any hard open research questions. However, individual initiatives may jeopardize the interoperability of KGs and cause vendor lock-in effects. Thus, the main challenge related to point 1 consists of avoiding such negative side effects, e.g. by standardizing both the security-aware protocols for accessing KGs and the related policy models. Authentication is not an issue, as several well-engineered solutions exist, covering distributed and – possibly – federated settings (for example Kerberos, Shibboleth, Cassandra, plus several identity management systems).

Point 2 can only be addressed by disseminating security and privacy culture in the KG community. It is essential to require that the papers that introduce inference control and anonymization methods provide explicit attack models, equipped with explicit hypotheses about the attacker. It is also very important to remove the frequent misconception that preventing the logical inference of secrets suffices to protect confidential data. Focusing confidentiality criteria on the indistinguishability of knowledge sources, as opposed to models of one source, is also of paramount importance: indistinguishability should mean that every source (or KG, in our context) that contains a secret should have the same observable behavior (in terms of accessible axioms and query results) as another source that contains no secrets, while some authors interpret this criterion as: every source that has a model (i.e., a logical interpretation) that satisfies a secret, should also have a model that satisfies no secrets. The latter interpretation of indistinguishability does not suffice to protect confidentiality from attackers that know the adopted anonymization/access control algorithms.

Point 3 constitutes the main group of technical challenges related to security and privacy. The notions of k-anonymity, l-diversity, and t-closeness need to be suitably generalized due to the greater complexity of knowledge bases with respect to relational databases (to the best of our knowledge no such generalizations have been introduced so far). For example, ontology axioms capture relationships that are not expressible by standard database integrity constraints. Unpublished results show that even in the simple case of KGs without ontologies, generalizing k-anonymity is a nontrivial task. For example, it does not suffice to require that each individual should be matchable to at least k distinct blanks in the anonymized KG, because this condition still permits re-identification in some cases.

Besides the above difficulties, the algorithms related to anonymization need to solve hard problems, similar (sometimes polynomially equivalent) to graph matching, or harder (NP-complete), depending on the task (e.g. anonymity checking, optimal anonymization, etc). Some preliminary, unpublished results (presented at this seminar) show that complexity depends also on the choice of anonymization technique: the anonymization procedure may consist of uniform substitutions of IRIs and literals with blanks, or it may map different occurrences of a same constant to different blanks (so as to "remove equalities"); furthermore, multiple copies of a same axiom may be created to obstacle re-identification, or – following an opposite approach – some axioms may be removed to make individuals indistinguishable. The complexity of a same anonymization task may range from trivial to NP-complete, depending on the above features of the anonymization procedure. Also, the cost function (which measures the information loss caused by anonymization) may influence the worst case complexity.

The instantiation of differential privacy in the KG area is even less developed. All we have to apply differential privacy in a symbolic setting is the so-called exponential mechanism that assumes a utility function for measuring the effect of the pseudo-random answer distortions required by differential privacy. In practice, using such a mechanism, the queries to a KG contain random errors, and the exponential mechanism guarantees that such errors have a limited effect on utility, as specified by the utility function. In practice, there is no methodology for defining the utility function. The KG "lies" randomly, and there is no theoretical support to estimating the possible negative effects of such lies, which is one of the reasons why differential privacy is not yet exploited in KR. Bringing differential privacy to its full potential in this area is an interesting research direction.

Point 4 focuses on security and privacy challenges specific to the federation setting. From an access control point of view, each federation member can have its own local access control policy and enforcement mechanism. The federation could also have a global access control policy. Since KGs have both a semantic dimension (their ontologies) and a data dimension, it is expected that access control models used for defining local policies in a federation of KGs take into consideration the semantic definitions of KGs in each case. Therefore, if necessary, local access control enforcement mechanisms would infer from the corresponding ontologies the underlying knowledge about subjects, actions, and resources to enforce access control policies (see [18]). Notwithstanding, from the point of view of the federation, this semantic dimension of KGs raises several access control challenges mainly related to (but not limited to) the semantic interoperability [25] of local ontologies. Some of them are the following:

- To define the model for unifying or establishing a common semantic framework for interconnecting the different domains and ontologies of the federation of KGs members, i.e., a model that guarantees the governance from the semantic point of view. This implies to "homogenize" to some extent the semantic dimension of the whole federation. This model should guarantee the interoperability among federation members and the federation itself. For example, let us assume for query evaluation purposes, there is an authorization request for accessing a resource that requires to infer knowledge about the entities involved. Most likely, these entities will be defined in different ways in local ontologies. Thus, in this situation, it is necessary to deal with some problems raised due to the general context of the federation, starting from just a naming problem to more complex logical issues such as, for instance, inconsistency, soundness, and incompleteness of definitions.

- To prevent possible information leaks. In the context of trust management systems, this problem can take place by means of a probing attack [26]. A probing attack consists in (systematically) sending authorization requests and registering the system's response, the attacker – who may be someone with low-ranking authorization permission for some type of information – gains knowledge about access control policies. In the scenario of a federation of KGs, there is a risk of vulnerability to some kind of probing attack that jeopardizes the privacy and security of the entities in the protection state of one of the KGs. Information leakage could occur because there could be sensitive information inferred from a local ontology. Then, through the grant/deny decision of the corresponding local access control application mechanism, it could be made accessible to the federation and thus, possibly, also to the querying user. That is, the result of query evaluation could provide users with some insights or full details about the inferred sensitive information.
- To integrate the global access control policy and the local policies of the members of a federation of KGs in such a way that the accomplishment of both, semantic and data conformance of the whole federation are guaranteed. In this regard there are two main challenges:
  - **(a)** To prevent inconsistency. Inconsistency arises in a trust management system when it contains contradictory authorization rules and thus, there are authorization requests that can be both granted and denied by different rules. In the case of a federation of KGs, inconsistency could be caused by the different ways in which the ontologies link to KGs members and the federation itself to formalize concepts and their relationships.
  - **(b)** To deal with the incompleteness of knowledge. This is, to establish a proper semantic approach for the global access control enforcement mechanism, the closed world assumption, the open world assumption, or some specific adaptation as, for instance, Local Closed-World Assumption [27]. This is necessary because if, to grant/deny an authorization request, the global access control enforcement mechanism depends on inferred knowledge of some of the KG members and one or more of these KG are not able to deduce that knowledge, there must be a logical support to justify the response of the global access control enforcement mechanism.
- To define a mechanism capable of creating a global explanation of the answer to a query by combining local explanations without revealing sensitive information obtained from (locally) inferred knowledge.

5.1.4 Requirements to Move Forward

As a first step to address the aforementioned challenges and to develop access control approaches that work in a federated setting, we need to specify the requirements to be satisfied by such approaches. During our discussions in the seminar, we observed that such requirements cannot be defined in a general manner but, instead, are use case specific. In particular, we identified the following three classes of use cases for creating and querying federations of knowledge graphs, and discussed the requirements for each of these classes.

Class 1 – Single Organization:

This class covers use cases in which all federation members are under the control of a single organization with a desire to decentralize management and maintenance of different portions of their organization-wide knowledge graph. Such federated architectures become increasingly important in enterprise architectures where knowledge graphs are used as an instrument for data integration following a virtualization approach. The Enterprise Knowledge Graph (EKG) acts as a central means for unified and integrated access across distributed sources, i.e. the EKG can be queried transparently through the federation layer, while the data resides and is managed autonomously in multiple separate databases. These federation architectures are advantageous where, for reasons of data security, data privacy, enterprise governance or scalability, a centralized approach to data integration is not feasible or desirable. Still, in an enterprise, a federation approach will naturally compete and be compared with centralized approaches, which raises a number of requirements:

- Data virtualization and federation can create significant additional performance overhead compared to centralized integration. Query optimization is thus of significant importance.
- Even though we are considering integration in a single-organization environment, data must remain properly segmented, secured, and isolated between different departments, units, and sub-organizations.
- In centralized alternatives, in particular in traditional application-centric architectures, access control patterns are well established. Enterprise Knowledge Graphs are a new form of data-centric architecture, in which access control needs to be defined on the federated knowledge graph itself, while still respecting the access control defined by the federation members.
- Authentication and identity management across the entire federation landscape, i.e., from the applications via the Enterprise Knowledge Graph down to the individual federation members, need to be managed in a coherent manner.

While the above mentioned requirements pose special challenges, there are also a number of favorable characteristics that can be utilized in reducing complexity compared to other environments:

- While the data sources (federation members) are autonomous and under the control of multiple sub-organisations, ultimately the entire federated system is under the control of a single organization, i.e., the autonomy of the members is counterbalanced with central coordination in the organization.
- We can make certain assumptions about the characteristics of the environment, e.g., performance of the individual subsystems, network latency, reliability of nodes, service levels, etc.
- We have extended options to gather and publish statistics and metadata of the federation members required for query optimization.
- Access control in such federated settings can lead to more performant query processing when properly utilized during query optimization. For example, source selection can immediately prune federation members from the query plan if access to the data of the member is forbidden during query execution.

In summary, there are special requirements in enterprise environments that make the problem of federation data access both more challenging and more manageable.

Class 2 – Collaborating Organizations:

The use cases in this class are characterized by the need for different organizations (or perhaps also different units within the same organization) to share and integrate some of their respective data within a collaboration or consortium that has been created to achieve a common goal (e.g., in a Data Space). Typical examples of such federations are federations that consist of several *private* SPARQL endpoints where each of them provides access to an *enterprise* knowledge graph of one of the involved organizations. This would require the query engine for the federation to authenticate itself at such a federation member on behalf of its user or application.

In these cases, the federation engine might not be aware of the complete set of policies of its federation members, e.g., what all the data of a federation member is and who is allowed to access which data and for which purpose. The federation engine may only be aware of the metadata about a federation member. This federation member can then be accessed by the federation engine and may be aware of what the corresponding usage policies are (or not). While pursuing a common goal, the different federation members may not be in a position to use the exact same access control mechanisms, e.g. a federation member may belong to multiple federation systems. This may result in some level of heterogeneity of such mechanisms across the federation. Hence, for such use cases, the query engine of the federation may be needed to interact with a large variety of access control mechanisms.

Once a federation member becomes part of a federation, the federation member may (or not) inform the federation engine about which of its data the federation engine may access and what can be done with this data. The federation engine may request access to more data on demand, i.e., the federation engine may ask a federation member whether it has more data than the available that the federation engine can access. Each federation member may choose whether it makes its usage policies available to the federation engine in advance, or whether it allows the federation engine to request the usage policies on demand; another alternative is that the federation engine gets informed about the usage policies once it retrieves certain data.

Thus, there may be cases in which a federation engine has the option to optimize its query execution plans based on information about data access and usage policies of the federation members, but there may also be cases in which such an option does not exist, or perhaps only for some of the federation members.

5.1.5 Class 3 – Autonomous Existing Sources:

These use cases require the execution of queries over a combination of several existing data sources that independent parties have put up for general use. Hence, any federation that emerges by such a combination of data sources consists of federation members that have not been created explicitly for participating in that particular federation, and that may not even be aware of their participation. Typical examples of such federations are federations that consist of several public SPARQL endpoints that provide access to knowledge graphs on the Web. So far, such public SPARQL endpoints have been made available without explicit access control restrictions. Yet, as is customary for many public REST APIs, we foresee cases in which providers of public data sources for a federation of knowledge graphs enforce access quota (e.g., a specific number of requests per day), potentially combined with subscription plans for increased quota. Or there may be offers for prioritized (faster) processing for paying users. Such a practice would require the query engine for the federation to authenticate itself at such a federation member on behalf of its user or application, and to take potential access limitations into account when creating query execution plans. Given that the data sources in this class of use cases are provided without any coordination, the access control mechanisms employed by any one of the federation members can be assumed to be completely separate from potential access control at other members of the federation, and there may be a high heterogeneity in terms of such mechanisms across the federation. Hence, a query engine for such use cases may be required to be able to interact with a large variety of access control mechanisms. Besides access control, data sources that are publicly accessible may have usage control policies for their data, which must be enforced by the federation engine. As an example, if the data obtained from some of the federation members is associated with a license, then the engine has to be able to determine whether these licenses allow the engine to combine the data and, if so, to determine what the license for the query result produced from the combined data should be.

Our Vision

During the seminar we discussed several options on how to enable general access control for federated data management. Our vision to study this question in a systematic manner is to develop an abstract framework that can be instantiated with respect to different use cases as well as the diverse technical setups we are encountering in federated setups. In particular, we discussed potential instantiations for the classes of use cases discussed above.

As main components of such a framework, we identified (i) federation members, which are providing access to data under certain policies that they enforce, (ii) a federation engine, which coordinates access to the federation members by performing global policy enforcement and optimizing the execution of user queries potentially involving multiple federation members, and (iii) clients, which send the user queries to the federation engine along with authentication tokens, which are passed down to the federation members via the federation engine.

During the discussion we identified several challenges for designing such an architecture. One of them is that it has to fulfill multiple potentially conflicting requirements, e.g., while additional metadata at the federation engine can be helpful to optimize queries and increase efficiency, this might be in conflict with privacy control. Hence, particular design choices for such a system are to identify the level of detail at which metadata can be shared and to determine which site – centralized (federation engine) and/or decentralized (participants) – needs to perform authentication and enforce policies.

Another challenge is that users – sending requests via the clients – would require some kind of explanation for their executed queries, i.e., what policies have been enforced and maybe also some quantification about additional query results (if allowed by the respective policies) that might have potentially been missed due to policy restrictions.

References

1. Kirrane, S., Mileo, A. & Decker, S. Access control and the resource description framework: A survey. *Semantic Web*. **8**, 311-352 (2017)
2. Costabello, L. & Gandon, F. Context-Aware Presentation of Linked Data on Mobile. *Information Retrieval And Management: Concepts, Methodologies, Tools, And Applications*. pp. 1940-1971 (2018)
3. Jajodia, S., Samarati, P., Sapino, M. & Subrahmanian, V. Flexible support for multiple access control policies. *ACM Transactions On Database Systems (TODS)*. **26**, 214-260 (2001)
4. Kirrane, S., Abdelrahman, A., Mileo, A. & Decker, S. Secure manipulation of Linked Data. *The Semantic Web – ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I 12*. pp. 248-263 (2013)

5 Radulovic, F., García-Castro, R. & Gómez-Pérez, A. Towards the Anonymisation of RDF Data. *The 27th International Conference On Software Engineering And Knowledge Engineering, SEKE 2015, Wyndham Pittsburgh University Center, Pittsburgh, PA, USA, July 6-8, 2015*. (2015)

6 Heitmann, B., Hermsen, F. & Decker, S. k-RDF-Neighbourhood Anonymity: Combining Structural and Attribute-based Anonymisation for Linked Data. *PrivOn@ ISWC*. **1951** (2017)

7 Silva, R., Leal, B., Brito, F., Vidal, V. & Machado, J. A differentially private approach for querying RDF data of social networks. *Proceedings Of The 21st International Database Engineering & Applications Symposium*. pp. 74-81 (2017)

8 Grau, B. & Kostylev, E. Logical Foundations of Privacy-Preserving Publishing of Linked Data. *Proceedings Of The Thirtieth AAAI Conference On Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*. pp. 943-949 (2016)

9 Buil-Aranda, C., Lobo, J. & Olmedo, F. Differential privacy and SPARQL. *Semantic Web.*, 1-29 (2021)

10 Bonatti, P., Kirrane, S., Petrova, I. & Sauro, L. Machine understandable policies and GDPR compliance checking. *KI-Künstliche Intelligenz*. **34** pp. 303-315 (2020)

11 Kurteva, A., Chhetri, T., Pandit, H. & Fensel, A. Consent through the lens of semantics: State of the art survey and best practices. *Semantic Web.*, 1-27 (2021)

12 Bonatti, P., Sauro, L. & Langens, J. Representing consent and policies for compliance. *2021 IEEE European Symposium On Security And Privacy Workshops (EuroS&PW)*. pp. 283-291 (2021)

13 Bonatti, P. & Sauro, L. Sticky Policies in OWL2: Extending PL with Fixpoints and Transitive Closure. *Proceedings Of The International Conference On Principles Of Knowledge Representation And Reasoning*. pp. 73-80 (2022)

14 Rodríguez-Doncel, V., Villata, S. & Gómez-Pérez, A. A dataset of RDF licenses. *Legal Knowledge And Information Systems – JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014*. **271** pp. 187-188 (2014)

15 Governatori, G., Lam, H., Rotolo, A., Villata, S., Atemezing, G. & Gandon, F. Checking Licenses Compatibility between Vocabularies and Data.. *COLD*. **1264** (2014)

16 Pellegrini, T., Havur, G., Steyskal, S., Panasiuk, O., Fensel, A., Mireles, V., Thurner, T., Polleres, A., Kirrane, S. & Schönhofer, A. DALICC: a license management framework for digital assets. *Proceedings Of The Internationales Rechtsinformatik Symposion (IRIS)*. **10** (2019)

17 Moreau, B. & Serrano-Alvarado, P. Ensuring license compliance in linked data with query relaxation. *Transactions On Large-Scale Data-and Knowledge-Centered Systems XLIX: Special Issue On Data Management – Principles, Technologies And Applications*. pp. 97-129 (2021)

18 Amini, M. & Jalili, R. Multi-level authorisation model and framework for distributed semantic-aware environments. *IET Information Security*. **4**, 301-321 (2010)

19 Gu, Z., Corcoglioniti, F., Lanti, D., Mosca, A., Xiao, G., Xiong, J. & Calvanese, D. A systematic overview of data federation systems. *Semantic Web*. **15**, 107-165 (2024)

20 Schwarte, A., Haase, P., Hose, K., Schenkel, R. & Schmidt, M. FedX: Optimization Techniques for Federated Query Processing on Linked Data. *ISWC (1)*. **7031** pp. 601-616 (2011)

21 Saleem, M., Khan, Y., Hasnain, A., Ermilov, I. & Ngomo, A. A fine-grained evaluation of SPARQL endpoint federation systems. *Semantic Web*. **7**, 493-518 (2016)

22 Acosta, M., Hartig, O. & Sequeda, J. Federated RDF Query Processing. *Encyclopedia Of Big Data Technologies*. (2019)

23 Khan, Y., Saleem, M., Mehdi, M., Hogan, A., Mehmood, Q., Rebholz-Schuhmann, D. & Sahay, R. SAFE: SPARQL Federation over RDF Data Cubes with Access Control. *J. Biomed. Semant.*. **8** (2017)
24 Endris, K., Almhithawi, Z., Lytra, I., Vidal, M. & Auer, S. BOUNCER: Privacy-Aware Query Processing over Federations of RDF Datasets. *Proceedings Of The 29th International Conference On Database And Expert Systems Applications (DEXA)*. (2018)
25 Slater, L., Gkoutos, G. & Hoehndorf, R. Towards semantic interoperability: finding and repairing hidden contradictions in biomedical ontologies. *BMC Medical Informatics And Decision Making*. **20** pp. 1-13 (2020)
26 Gurevich, Y. & Neeman, I. DKAL: Distributed-knowledge authorization language. *2008 21st IEEE Computer Security Foundations Symposium*. pp. 149-162 (2008)
27 Bertino, E., Provetti, A. & Salvetti, F. Local Closed-World Assumptions for reasoning about Semantic Web data.. *APPIA-GULP-PRODE*. pp. 314-323 (2003)

5.2 Quality-aware Knowledge Graph Construction

Cinzia Cappiello (Polytechnic University of Milan, IT)
Maria Esther Vidal (TIB – Hannover, DE)
Sandra Geisler (RWTH Aachen, DE)
Ana Iglesias-Molina (Universidad Politécnica de Madrid, ES)
Dylan Van Assche (IDLab - Ghent University - imec, BE)
David Chaves-Fraga (Universidade de Santiago de Compostela, ES)
Anastasia Dimou (KU Leuven, BE)
Irene Celino (CEFRIEL – Milan, IT)
Anisa Rula (University of Brescia, IT)
Maurizio Lenzerini (Sapienza University of Rome, IT)

5.2.1 Introduction

A Knowledge Graph (KG) is a structured graph-based representation of knowledge that captures relationships between entities in a domain [22]. It is constructed by starting from knowledge bases, which gather information from heterogeneous sources such as free text, web pages, databases, and media content.

Knowledge graphs often incorporate concepts from knowledge representation and ontology modeling to formalize domain-specific knowledge [22]. They are used in various domains, including search engines, recommendation systems, question-answering systems, semantic web applications, knowledge management systems, and more. They serve as powerful tools for organizing, representing, and leveraging knowledge to support advanced analytics, decision-making, and information retrieval tasks.

For example, operators usually manage their knowledge in a distributed manner in industrial settings, especially in manufacturing. The relevant information they need to support their daily work is often spread across many different and heterogeneous sources that are difficult to access and retrieve. For example, shopfloor employees, willing to solve a problem on the production line, face the challenge of looking for answers in different documents (e.g., manuals, procedures, technical diagrams) and in different systems (e.g., machinery inventories, ticketing systems, logs, retrofitting/sensor data), when not even in (un)official communication channels (e.g., chat, emails).

Often, relevant knowledge is not documented or digitized because it comes from experience, remains tacit in people's heads, and is transferred only orally. In this context, organizing industrial knowledge with knowledge graphs would overcome the challenges of access, heterogeneity, and documentation. Industrial knowledge graphs would make it easy to get answers to compelling questions in different areas, relate concepts with each other, and define the context for that knowledge [21].

However, constructing a Knowledge Graph (KG) involves various challenges and issues, including data extraction, data quality, data integration, and data security issues as well as architectural aspects such as scalability and interoperability [19, 22, 20]. The Dagstuhl Seminar allowed us to discuss and focus on two main topics: the definition of a general pipeline for the KG construction and data quality. The following sections will summarize discussions and findings.

5.2.2 State of the Art

Knowledge is usually distributed among several heterogeneous sources. The paper by Lenzerini [5] provides the theoretical background to solve this problem from a data perspective. Ontologies play an important role in this process [6], as they provide a unified view of a specific domain that is used to integrate the input data [26]. Therefore, the construction of a knowledge graph is defined as a data integration system [7, 18].

The construction can be performed in either a materialized or a virtual way. In the first one, the input data is transformed according to the ontology terms, while in the second, data remain in the original format and queries following the ontology are transformed into queries over the input source using the mapping rules. There exist multiple techniques to create and maintain knowledge graphs, which could either be (semi-)automatic methods, e.g., based on NLP, or manual approaches. However, only some approaches consider or define the life cycle of a knowledge graph, i.e., how a knowledge graph is generated, how it evolves, and how its quality is assessed during the evolution. For example, [26] present a KG life cycle with six steps: design and requirements scoping, data ingestion, data enrichment, storage, consumption, and maintenance. Additionally, they present a platform how to implement these steps. Similarly, [27] and [28] describe four phases of a lifecycle, comprising KG creation, hosting, curation, and deployment. [27] also present with a corresponding framework, Helio, which realizes the requirements they elicitated in their work. However, still a general formalization of the KG life cycle is missing, which considers also the various actors and roles, requirements, and constraints in the different steps of a life cycle.

Concerning previous approaches, data quality for a KG life cycle does not necessarily cover the broader spectrum of the quality of the KG ecosystem. The quality of the KG ecosystem includes the quality of the data feeding into the KG, the quality of the data transformation process for constructing the KG, and the quality of the KG data itself. Managing data quality in a knowledge graph involves various processes, including quality assessment in terms of error detection and quality improvement in knowledge graph completion. The following discusses the different aspects.

- **Quality of data feeding the KG:** *Quality of data sources* refers to the quality of the raw data obtained from various sources before it is integrated into the KG. It involves aspects such as accuracy, completeness, consistency, timeliness, and relevancy of the raw data. *Data cleaning* involves identifying and correcting errors, inconsistencies, and inaccuracies in the raw data before transforming it into KG format. Data cleaning techniques such as deduplication, outlier detection, and data validation are employed in this phase [8, 9].

- **Quality of the data transformation for the KG Construction:** *Quality of data transformation* assesses how accurately data from the source is mapped to entities and relationships in the KG with respect to the applied ontologies [11, 12, 13, 10] or how complete the mappings are if all the relevant source data is properly mapped into the KG. Only a few approaches are proposed to assess the quality of mappings [11, 13, 10].
- **Quality of the Knowledge Graph Data:** Quality assessment of the final Knowledge Graphs may include: (i) *Semantic Consistency* refers to the coherence and logical consistency of the relationships and entities represented within the KG; (ii) *Accuracy and Completeness*, the KG should accurately reflect the real-world domain it represents, and it should contain comprehensive coverage of relevant entities and relationships within that domain; (iii) *Timeliness*, the KG should be regularly updated to reflect changes and updates in the underlying data sources. Timeliness ensures that the KG remains relevant and up-to-date over time [17]
  RDF validators assess whether RDF data complies with specified restrictions, typically using standards like SHACL [14]. These tools analyze excerpts of RDF data along with given restrictions to generate validation reports. On the other hand, RDF quality tools focus on evaluating and quantifying the quality of a KG and the factors affecting it [15].
  KG improvement is crucial for constructing a Knowledge Graph (KG). It involves identifying and rectifying errors, inconsistencies, and inaccuracies. In particular, the contributions are focused on improving RDF links, which aims to connect local RDF resources with those in different Knowledge Graphs (KGs) [16].
- **Quality of Ontologies** The quality of ontologies is a prerequisite to organize the knowledge in a knowledge graph in a sensible way. There exist a multitude of quality metrics for ontologies, which refer to the different structural components of the ontology, such as the hierarchy of the concepts, data and object properties, or instances [25]. The metrics can be (semi-)automatically be checked using acknowledged tools, such as the OOPS! pitfall scanner [24]. Such tools can be employed as services to be integrated into a life cycle step of validation.

5.2.3 Current Challenges and Open Research Questions

The construction of a knowledge graph requires to:

- Identify the tasks within a generalized workflow for the KG construction process.
- Characterize the inputs, policies (e.g., governance, guidelines, etc.) to be followed, and expected output.
- Guarantee traceability and documentation of decisions made at each step so they can be explained later.
- Test for quality at each step of the process: verify, explain, trace.

Data quality has an important role in the KG construction since KGs are frequently messy and, hence, normally have data quality problems. The main causes for poor quality are various. In fact, KGs' flexibility often implies a change in data quality requirements. KGs could inherit data quality issues from the underlying data sources they are created from. Data quality issues may also occur because of the construction pipeline itself (e.g., errors in mappings and lack of updates when ontologies are updated). Finally, the employed ontologies can be subject to quality issues.

Common data quality issues are related to missing entities and/or properties, inaccuracies, and redundancies. Such issues might affect all parts of the KG, including schema and attributes, which means that validation is a challenge. Data quality assessment and improvement

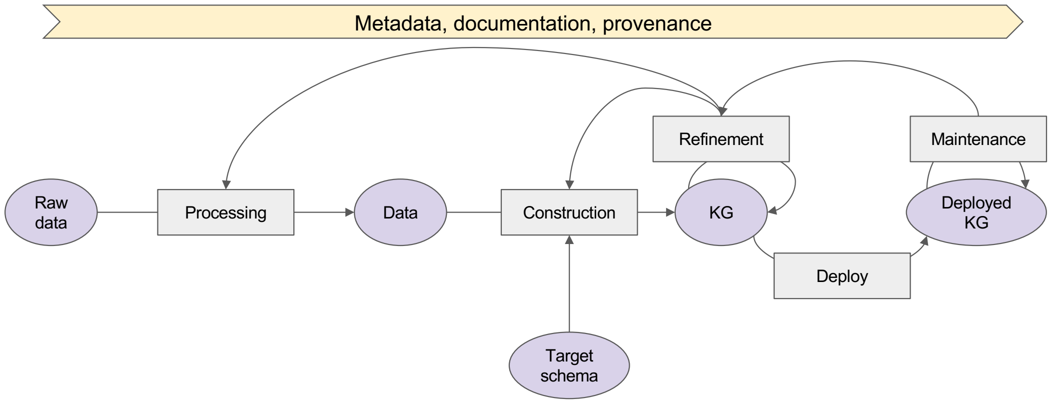


Figure 1 KG lifecycle.

are difficult in this context. Not all aspects are quantifiable, and non-functional quality aspects such as trustworthiness are difficult to assess. On the other hand, the volume and variety of values contained in the KG hampered data cleaning.

In the discussion, a general agreement was found on recognizing a lack of a comprehensive and holistic methodology for creating a KG quality life cycle. Quality management can be seen as an orthogonal process. Quality should be taken into account in the entire KG lifecycle, from the creation phase to the maintenance processes (e.g., KG refinement/link prediction). It is also important to consider users' and curation feedback processes in all the stages to propagate quality repairs back to the source (or the underlying data collection processes themselves); moreover, the perspective of different users can bring different quality requirements (e.g., the KG is "good enough" for a specific user task according to some qualitative aspects, opposed to quantitative metrics like accuracy or precision).

For this reason, it is necessary to represent the quality requirements in each life cycle aspect. The specification of quality requirements is not trivial. In fact, KG quality is a function of potential use. Hence, defining a specification of quality requirements for those uses is challenging. Directions of potential interest for expressing such a specification include design patterns and graph normal forms. LLMs might be explored to see if they can be useful in this task.

For the data quality improvement, the Pay as you go approach by [23] can be adopted:

- How can we progressively improve KG quality?
- How do we allocate technical and social efforts for human-driven quality assessment?
- How can we design systems that make iterative improvements?
- How do we account for the quality of the maintenance processes over time?

5.2.4 Vision and Possible Approaches

The discussion leads to possible approaches. Figure 1 shows the result of the first discussions to represent a general workflow.

Data quality assessment can be performed on consistency, completeness, and accuracy, considering input data, mappings, and ontologies.

Drawing from principles in data integration [2], semantic data management [3], and data ecosystems [1], the concept of *Knowledge Graph Ecosystems*(KGE) is presented as an abstraction that specifies a knowledge graph based on six components. They include:

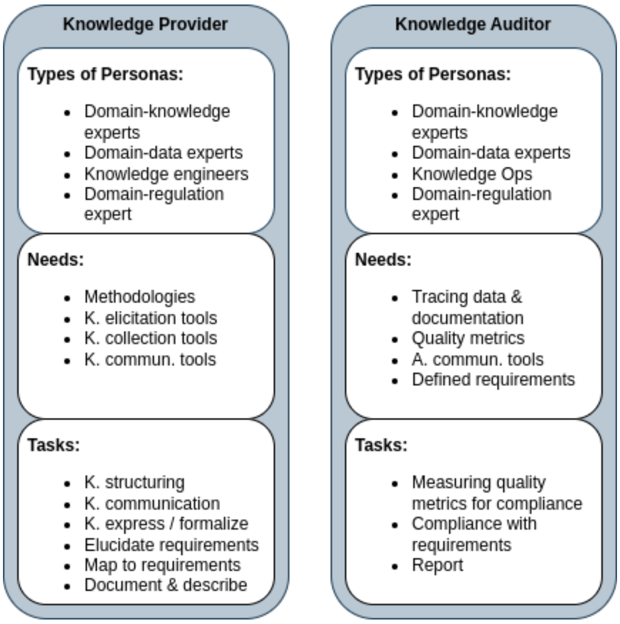


Figure 2 KG new roles.

i) the data sources;
ii) a unified schema corresponding to an ontology expressed by a formal description of the domain of interest;
iii) the mappings between the unified schema and the data sources;
iv) domain constraints;
v) a knowledge graph (logically or physically materialized) populated with data from the data sources respecting the mappings, the ontology, and the domain constraints; and
vi) a log, that records the evolution of the ecosystem.

We argue that the abstraction of a knowledge graph ecosystem allows for capturing the main characteristics of the various components and operations that enable the execution of data governance tasks towards the creation and maintenance of a knowledge graph, as well as the traceability of its evolution.

During the seminar, such an abstraction was formalized also considering the people involved. The next subsection shows the achieved results.

5.2.4.1 Roles in the KGE lifecycle

Different actors are involved in this process. A previous contribution [4], identifies three major stakeholders: KG Builders, KG Analysts, and KG Consumers. We adapt these stakeholders to the KGEs so they are not restricted only to KGs (i.e., Knowledge Builders instead of KG Builders) and identify two additional stakeholders: Knowledge Providers and Knowledge Auditors (see Figure 2). We describe how these stakeholders intervene in KGEs, defining the roles (s) they can play, tasks, and needs. Each stakeholder may play more than one role in the different KG lifecycle steps.

Knowledge Providers bring expertise into the KGE. They do not only provide input on the KGE subject matter (i.e. as domain-knowledge experts), but also on the data, required regulations, and knowledge engineering aspects. Together, they define the needs and tasks of what the KGE must serve, to (i) specify the requirements for the Knowledge Builders; (ii) comply with the Knowledge Auditors requirements; and (iii) ensure that the needs of the Knowledge Consumers and Knowledge Analysts are met. Knowledge Providers then require the means for seamless communication with the rest of the stakeholders, e.g., communication and visualization tools, and for collecting and sharing their knowledge as input for the KGE.

Knowledge Builders are responsible for integrating the knowledge from the Knowledge Providers and building the KGE resources. This stakeholder group comprises experts in KGE-related technologies, such as knowledge engineers, application developers, and KnowledgeOps. Their output must be up to the coverage and quality standards of the Knowledge Auditor, and be appropriate for its use by Knowledge Consumers and Knowledge Analysts. Therefore, Knowledge Builders must report, document, and provide provenance traces for all the resources produced and knowledge processing performed.

Knowledge Auditors review and evaluate the KGE in terms of quality and compliance with the requirements. This task is mainly performed by domain-knowledge, domain-data, domain-regulation experts, and the KnowledgeOps. They define the metrics for evaluation depending on the use case, regulations, and corresponding requirements. Their efforts serve to check and improve the KGE's quality, comply with regulations, and ensure that it is valid for consumption to perform the tasks that it is built for.

Knowledge Analysts directly interact with the KGE to generate insights. These stakeholders are usually data scientists, ML/AI experts, or even app developers. They are not necessarily knowledge engineering experts as the Knowledge Builders, but possess the skills to interact, extract information and support discovery with them in the KGE. Their work output is then shared and reported for Knowledge Consumers to use and Knowledge Auditors to verify that their needs are fulfilled.

Knowledge Consumers are the end-users of the KGE. They do not usually interact directly with the KGE, so they do not require technical skills and tend to use user-friendly interfaces. They need the documentation, reports, and interfaces to consume the KGE and communicate whether the KGE meets their requirements.

5.2.4.2 A formal definition of the KGE

Such a section formalizes the KGE and the related lifecycle.
A *life cycle step* serves as the fundamental unit of operations executed within a knowledge graph ecosystem. It encompasses a service responsible for implementing operations over the ecosystem. Additionally, contextual information of the knowledge graph ecosystem defines the actors and their roles, as detailed in Section 5.2.4.1. These roles include requirements to be satisfied, constraints to be validated, and the needs of actors expressed as both, requirements and constraints, during their involvement in executing the life cycle over a knowledge graph ecosystem. The life cycle steps are combined to form life cycles; the steps can be executed concurrently or sequentially based on execution dependencies between life cycle steps specified in the life cycle. In the following, we formally define the sketched concepts of Knowledge Graph Ecosystems, life cycles, and the steps.

Knowledge Graph Ecosystems

A *knowledge graph ecosystem (KGE)* is defined by a tuple $KGE = (D, O, M, DC, KG, L)$ where:

- D: represents a set of data sources, each defined by its schema ($\theta(ds)$) and its corresponding instances ($\alpha(ds)$). The schema ($\theta(ds)$) specifies the structure and attributes of the data source, while the instances ($\alpha(ds)$) constitute the actual data organized according to the schema.
- O: denotes an ontology expressed as a logical theory, capturing the conceptual framework and relationships within the domain of interest. This provides a structured vocabulary for describing entities and their relationships.
- M: signifies a set of mapping assertions that establish relationships between elements of the ontology (O) and the data sources (D). These mappings provide a declarative definition of the attributes ($\theta(ds)$) of each data source ds in terms of the concepts defined in the ontology (O).
- DC: represents a collection of domain-specific constraints, expressed in a formal language. These constraints ensure the consistency, accuracy, and quality of the data for all the components of the ecosystem.
- KG: corresponds to the knowledge graph itself, which can be empty or the rendering of the ontology O with individuals generated by data collected from the data sources described by D based on mapping assertions in M;
- L: An ordered list of log entries, each containing a timestamp, a before-image, an after-image, and a description of the execution of life cycle step over the ecosystem. This logging mechanism facilitates tracking and auditing of changes and activities within the KGE.

▶ **Example 1.** In a healthcare setting, a knowledge graph ecosystem ($HKGE$) can be employed to integrate and manage diverse healthcare data sources, such as electronic health records (EHRs), medical literature databases, clinical trial repositories, and medical imaging archives. A $HKGE$ aims to facilitate various tasks, including patient diagnosis, treatment planning, research, and healthcare analytics. The components of a HKGE can be summarized as follows:

- **Data Sources (D)**: For example, this can be Electronic Health Records (EHRs), Medical Literature Databases, Clinical Trial Repositories, or Medical Imaging Archives.
- **Ontology (O)**: Represents a conceptual model of healthcare domain entities, relationships, and attributes.
- **Mapping Assertions (M)**: Specify how data from each source (D) aligns with the ontology (O).
- **Domain-specific Integrity Constraints (DC)**: Define rules to ensure data consistency, accuracy, and quality.
- **Knowledge Graph (KG)**: Represents the integrated view of healthcare data structured according to the ontology (O).
- **Log Entries (L)**: Maintain a record of lifecycle events, including data integration, ontology updates, and analysis results.

KG Life Cycles

A knowledge graph ecosystem KGE is subject of *life cycles*, consisting of a series of ordered *life cycle steps* and potential sub-life cycles. These life cycles orchestrate the manipulation and evolution of KGE components, guiding their creation, validation, curation, maintenance,

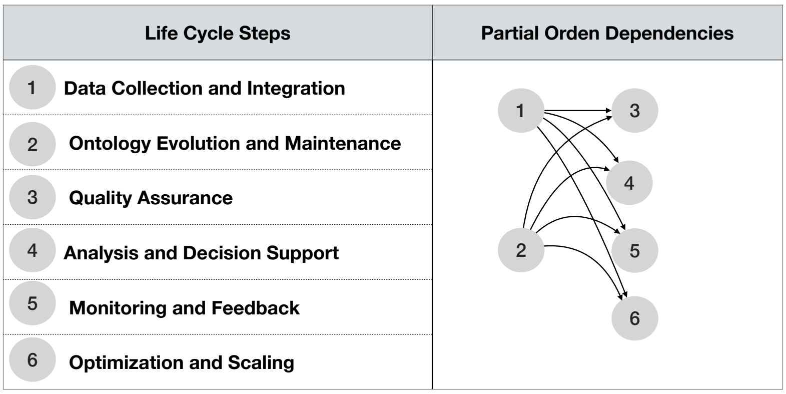


Figure 3 Life cycle composed of six life cycle steps.

traversal, and analysis. Each life cycle step operates within a defined partial order, ensuring systematic execution and progression throughout the life cycle of KGE.
A life cycle LC is a partial order (LCS,R), where LCS is a set of life cycle steps and R is a precedence relation between the elements of LCS. The relation R satisfies reflexivity, antisymmetry, and transitivity. LC=(LCS,R) is inductively defined as follows:

Base Case: LC=($\{lcs\}$,$\{(lcs, lcs)\}$) for a is life cycle step lcs.

Inductive Case: Consider a life cycle LC'=(LCS',R') and lcs denotes a life cycle step which does not belong to LCS'. In this case, a new life cycle LC=(LCS,R) can be defined as follows:

- LCS corresponds to the union of LCS' and $\{lcs\}$.
- The precedence relation R is defined by extending R' with pairs (lcs, lcs') or (lcs', lcs) for each $lcs' \in R'$ such that lcs' precedes lcs or lcs precedes lcs', respectively.

▶ **Example 2.** A life cycle over HKGE involves a series of steps, each contributing to the overall management and evolution of the ecosystem. These steps are interrelated and follow a partial order, ensuring systematic progression throughout the lifecycle of HKGE.

1. **Data Collection and Integration**:
   - Extract data from diverse healthcare sources (D), including Electronic Health Records (EHRs), Medical Literature Databases, Clinical Trial Repositories, and medical images.
   - Integrate the collected data into the knowledge graph (KG) based on mapping assertions (M).
2. **Ontology Evolution and Maintenance**:
   - Update the healthcare ontology (O) to accommodate new concepts, terminology, and domain-specific knowledge.
   - Ensure consistency between the ontology (O) and the integrated data in KG.
3. **Quality Assurance**:
   - Validate healthcare data against domain-specific constraints (DC) to ensure data quality, accuracy, and compliance with standards.
   - Perform data cleaning, deduplication, and normalization to improve data quality and consistency.
4. **Analysis and Decision Support**:
   - Perform healthcare analytics using the knowledge graph (KG) to identify patient cohorts, predict outcomes, and recommend treatments.

  - Support clinical decision-making by providing insights derived from integrated healthcare data and knowledge.
5. **Monitoring and Feedback**:
  - Monitor performance and usage of the HKGE components.
  - Collect feedback from healthcare professionals and researchers to improve HKGE effectiveness and usability.
6. **Optimization and Scaling**:
  - Optimize the performance of the HKGE components to handle large-scale data and complex analytics tasks.
  - Scale the system to integrate growing data volumes and user demands while maintaining efficiency and reliability.

Figure 3 summarizes the life cycle and the dependencies between the six life cycle steps. As indicated in the figure, 1 and 2 should be executed before steps 3, 4, 5, and 6. Thus, the partial order between the life cycle steps enables the management and evolving of the different components of HKGE.

Life Cycle Steps

A *life cycle step* is defined as $lcs = (S, \langle P, Ro, C, Re, N \rangle)$, where $< Ro, P, C, Re, N >$ comprises the contextual information that guides the execution of lcs over a knowledge graph ecosystem KGE.

- S: represents a service related to knowledge graph operations (e.g., creation, quality assessment, updates, or querying) to be executed within a KGE. Services are defined as functions, where inputs include KGE, and outputs may result in modifications or analysis of KGE.
- P: represents a set of actors who fulfill roles in executing the life cycle step. Actors may include individuals contributing to or overseeing the execution of the services. 5.2.4.1 presents the actors that can participate in a life cycle.
- Ro: denotes a set of roles responsible for executing the services within KGE. As defined in 5.2.4.1, a role could be a knowledge engineer or a data scientist.
- C: refers to a set of constraints expressed using a logical formalism. These constraints may encompass data quality standards, compliance requirements, or other conditions that must be satisfied during the execution of the life cycle step.
- Re: Signifies a set of requirements expressed in a logical formalism, outlining desired outcomes or conditions that the execution of the life cycle step aims to achieve.
- N: represents a set of needs, where each need specifies requirements, constraints, roles, and actors involved in executing the life cycle step. Each need is characterized by a quadruple consisting of a set of requirements and constraints stated by an actor while playing a role.

A life cycle $LC = (LCS, R)$ is applied to a knowledge graph ecosystem KGE to generate a new knowledge graph ecosystem KGE'. This process involves executing each life cycle step in LCS while adhering to the dependencies specified in R. We denote the result of executing LC over KGE as $\sigma(LC, KGE)$.

Consider knowledge graph ecosystems $KGE = (D, O, M, DC, KG, L)$ and $KGE' = (D', O', M', DC', KG', L')$, along with a life cycle step $lcs = (S, \langle P, Ro, C, Re, N \rangle)$. The execution of lcs over KGE, denoted as $\lambda(lcs, KGE)$, results in KGE' by applying the service S to KGE, while ensuring compliance with the specified needs in N, especially validating the constraints in C.

Given a knowledge graph ecosystem $KGE = (D, O, M, DC, KG, L)$ and a life cycle $LC = (LCS, R)$, the execution of LC over KGE (denoted as $\sigma(LC, KGE)$) is inductively defined based on the life cycle steps of LC and the precedence relation specified by R.

Base Case: The life cycle comprises a single life cycle step $lcs = (S, \langle Ro, P, C, Re, N\rangle)$, i.e., $LC = (\{lcs\}, \{(lcs, lcs)\})$. Executing LC over KGE yields $\sigma(lcs, KGE) = (D', O', M', DC',$
$KG', L')$, where D', O', M', and KG' result from applying the service S to D, O, M, and KG, according to the needs in N and the validation of constraints in C. The log L' includes all entries in L along with input received by S and the produced output, annotated with the timestamp of execution. Additionally, L' records the results of validating constraints in C during the execution of S on KGE, represented using a specified language.

Inductive Case: Let $KGE' = (D', O', M', DC', KG', L')$ be a knowledge graph ecosystem resulting from executing a life cycle $LC' = (LCS', R')$ over $KGE = (D, O, M, DC, KG, L)$, i.e., $KGE' = \sigma(LC', KGE)$. Consider a life cycle $LC = (LCS, R)$ created by adding a life cycle step lcs to LCS', i.e., $LC = LCS' \cup \{lcs\}$, and updating R to include dependencies in R' and between lcs and other life cycle steps in LCS'. Suppose $lcs = (S, \langle P, Ro, C, Re, N\rangle)$. Executing LC over KGE' results in $KGE'' = (D'', O'', M'', DC'', KG'', L'')$, where D'', O'', M'', DC'', and KG'' are obtained by applying service S to D', O', M', DC', and KG' according to the needs in N and the validation of constraints in C. The log L'' includes all entries in L', input received by S, and the output annotated with the timestamp of execution.

▶ **Example 3.** Consider the life cycle step **Quality Assurance** and suppose HKGE$_{1,2}$=($D_{1,2}$,$O_{1,2}$,$M_{1,2}$,$KG_{1,2}$,DC,$L_{1,2}$) is the result of applying the steps (1) and (2) from Figure 3 over the HKGE. **Quality Assurance** comprises a service S_3 for ensuring that $KG_{1,2}$ meets the contextual information $\langle P_3, Ro_3, C_3, Re_3, N_3\rangle$ defined as follows:

- P_3: actors include **Knowledge Builder**, **Knowledge Analyst**, **Knowledge Provider**, and **Knowledge Auditor**.
- Ro_3: the roles within the **Knowledge Builder** actor include *knowledge engineer*, *knowledge ops*, and *app developer*. The **Knowledge Analyst** actor encompasses the role of *researcher*. The **Knowledge Provider** and **Knowledge Auditor** actors consist of the roles *domain data expert* and *domain knowledge expert*.
- C_3: constraints representing the correctness of the data collected and integrated by step (1), and their compliance with clinical guidelines and treatment protocols.
- Re_3: requirements stating rules that indicate how entities, which violate the constraints, must be curated, i.e., how entities that invalidate the constraint will be treated or curated.
- N_3: needs include requirements for data curation when constraints regarding clinical guidelines and medical protocols are violated, requested by **Knowledge Providers** in their roles as domain data and knowledge experts. **Knowledge Auditors**, also acting as domain experts, require that changes resulting from executing S_3 are traced in the log $L_{1,2}$. Finally, **Knowledge Builders** in their roles of knowledge engineers, ops, and app developers demand that S_3 is efficiently executed and capable of scaling to handle the large volume of data integrated into $KG_{1,2}$.

The execution of the **Quality Assurance** life cycle step results in the creation of a new knowledge graph ecosystem ($D_{1,2}$, $O_{1,2}$, $M_{1,2}$, KG_3, DC, L_3). Here, KG_3 represents the outcome of the curation process performed by S_3, adhering to the contextual information. L_3 encompasses all traces detailing the changes made during the curation process.

References

1 Geisler, S., Vidal, M., Cappiello, C., Lóscio, B., Gal, A., Jarke, M., Lenzerini, M., Missier, P., Otto, B., Paja, E., Pernici, B. & Rehof, J. Knowledge-Driven Data Ecosystems Toward Data Transparency. *ACM J. Data Inf. Qual.*. **14**, 3:1-3:12 (2022)
2 Lenzerini, M. Data Integration: A Theoretical Perspective. *Proceedings Of The Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium On Principles Of Database Systems, June 3-5, Madison, Wisconsin, USA*. pp. 233-246 (2002)
3 Lenzerini, M. Ontology-Based Data Management. *Proceedings Of The 6th Alberto Mendelzon International Workshop On Foundations Of Data Management, Ouro Preto, Brazil, June 27-30, 2012*. **866** pp. 12-15 (2012)
4 Li, H., Appleby, G., Brumar, C., Chang, R. & Suh, A. Knowledge Graphs in Practice: Characterizing their Users, Challenges, and Visualization Opportunities. *IEEE Transactions On Visualization And Computer Graphics*. **30**, 584-594 (2024)
5 Lenzerini, M. Data integration: A theoretical perspective. *Proceedings Of The Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium On Principles Of Database Systems*. pp. 233-246 (2002)
6 Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M. & Rosati, R. Linking data to ontologies. *Journal On Data Semantics X*. pp. 133-173 (2008)
7 Iglesias, E., Jozashoori, S., Chaves-Fraga, D., Collarana, D. & Vidal, M. SDM-RDFizer: An RML interpreter for the efficient creation of RDF knowledge graphs. *Proceedings Of The 29th ACM International Conference On Information & Knowledge Management*. pp. 3039-3046 (2020)
8 Batini, C. & Scannapieco, M. Data and Information Quality – Dimensions, Principles and Techniques. (Springer,2016), https://doi.org/10.1007/978-3-319-24106-7
9 Wang, R. & Strong, D. Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.*. **12**, 5-33 (1996,3), https://doi.org/10.1080/07421222.1996.11518099
10 Heyvaert, P., Meester, B., Dimou, A. & Verborgh, R. Rule-driven inconsistency resolution for knowledge graph generation rules. *Semantic Web*. **10**, 1071-1086 (2019), https://doi.org/10.3233/SW-190358
11 Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., Hellmann, S. & Walle, R. Assessing and Refining Mappings to RDF to Improve Dataset Quality. *SEMANTIC WEB – ISWC*. (2015), https://doi.org/10.1007/978-3-319-25010-6_8
12 Paulheim, H. Data-Driven Joint Debugging of the DBpedia Mappings and Ontology. *The Semantic Web*. (2017), https://doi.org/10.1007/978-3-319-58068-5_25
13 Randles, A., Junior, A. & O'Sullivan, D. A Framework for Assessing and Refining the Quality of R2RML mappings. *22nd International Conference On Information Integration And Web-based Applications & Services (iiWAS '20)*. (2021), https://doi.org/10.1145/3428757.3429089
14 Knublauch, H. & Kontokostas, D. Shapes constraint language (SHACL). (W3C,2017,7), https://www.w3.org/TR/shacl/
15 Issa, S., Adekunle, O., Hamdi, F., Cherfi, S., Dumontier, M. & Zaveri, A. Knowledge graph completeness: A systematic literature review. *IEEE Access*. **9** pp. 31322-31339 (2021), https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9344615
16 Nentwig, M., Hartung, M., Ngomo, A. & Rahm, E. A survey of current Link Discovery frameworks. *Semantic Web*. **8**, 419-436 (2017), https://doi.org/10.3233/SW-150210
17 Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. & Auer, S. Quality assessment for Linked Data: A Survey. *Semantic Web*. **7**, 63-93 (2016), https://doi.org/10.3233/SW-150175

18 De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A. & Rosati, R. Using Ontologies for Semantic Data Integration. *A Comprehensive Guide Through The Italian Database Research Over The Last 25 Years*. pp. 187-202 (2018), https://doi.org/10.1007/978-3-319-61893-7_11

19 Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A. & Taylor, J. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue*. **17**, 48-75 (2019,4), https://doi.org/10.1145/3329781.3332266

20 Peng, C., Xia, F., Naseriparsa, M. & Osborne, F. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*. **56**, 13071-13102 (2023)

21 Buchgeher, G., Gabauer, D., Martinez-Gil, J. & Ehrlinger, L. Knowledge Graphs in Manufacturing and Production: A Systematic Literature Review. *IEEE Access*. **9** pp. 55537-55554 (2021)

22 Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G., Gutierrez, C., Kirrane, S., Gayo, J., Navigli, R., Neumaier, S., Ngomo, A., Polleres, A., Rashid, S., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S. & Zimmermann, A. Knowledge Graphs. *ACM Comput. Surv.*. **54** (2021,7), https://doi.org/10.1145/3447772

23 Sequeda, J., Briggs, W., Miranker, D. & Heideman, W. A pay-as-you-go methodology to design and build enterprise knowledge graphs from relational databases. *The Semantic Web – ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26 – 30, 2019, Proceedings, Part II 18*. pp. 526-545 (2019)

24 Poveda-Villalón, M., Gómez-Pérez, A. & Suárez-Figueroa, M. Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal On Semantic Web And Information Systems (IJSWIS)*. **10**, 7-34 (2014)

25 Mc Gurk, S., Abela, C. & Debattista, J. Towards ontology quality assessment. *Proc. Of The 4th Workshop On Linked Data Quality (LDQ2017), Co-located With The 14th Extended Semantic Web Conference (ESWC)*. pp. 94-106 (2017)

26 Yip, H. & Sheth, A. The EMPWR Platform: Data and Knowledge-Driven Processes for the Knowledge Graph Lifecycle. *IEEE Internet Computing*. **28**, 61-69 (2024)

27 Cimmino, A. & García-Castro, R. Helio: a framework for implementing the life cycle of knowledge graphs. *Semantic Web.*, 1-27 (2022)

28 Simsek, U., Angele, K., Kärle, E., Opdenplatz, J., Sommer, D., Umbrich, J. & Fensel, D. Knowledge graph lifecycle: Building and maintaining knowledge graphs. *Second International Workshop On Knowledge Graph Construction*. (2021)

5.3 Knowledge Engineering and Education

Juan F. Sequeda (data.world – Austin, US)
Aidan Hogan (University of Chile – Santiago de Chile, CL)
Axel Polleres (Wirtschaftsuniversität Wien, AT)
Christophe Debruyne (University of Liège, BE)
Eva Blomqvist (Linköping University, SE)
Samaneh Jozashoori (Metaphacts GmbH – Walldorf, DE)
Paul Groth (University of Amsterdam, NL)
Claudio Gutierrez (University of Chile – Santiago de Chile, CL)
George Fletcher (TU Eindhoven, NL)
Ivo Velitchkov (Brussels, BE)

5.3.1 Introduction

The success of Knowledge Graphs in the real world is dependent on educating a wide audience on the social and technical knowledge work that needs to be done, and most importantly motivating the need for it. The focus of this group throughout the week was on discussing:

- Knowledge work: what are the known and unexplored themes
- Educational material: who is the audience, what material exists and what is needed?

In order to define what we mean by knowledge work, the group started by sharing the themes of what we consider to be part of knowledge work through our experience. Afterwards, we brainstormed themes that we believe should be part of knowledge work but for which we lack expertise. Furthermore, the knowledge work is independent of any specific technology such as knowledge graphs.

5.3.2 State of the Art

We present the known themes of knowledge work categorized as social and technical.

Social

The common social themes for knowledge work are the following:

Building Consensus. Strive to build consensus towards shared goals by bridging and integrating how different stakeholders think about a domain by embracing differences.

Intermediating between Actors. Serve as a bridge between different stakeholders, in order to acquire knowledge and find a way to build consensus.

Methodologies. Leverage methodologies and frameworks in order to acquire knowledge and create ontologies in a systematic manner using competency questions, etc.

People Skills. Communicate with stakeholders of different profiles and backgrounds, which may involve interviews, storytelling, etc.

Technical

The common technical themes for knowledge work are the following:

Metadata Management. Create and manage metadata about data assets to search and find relevant assets; define provenance in order to support upstream and downstream applications.

Knowledge Maintenance. Understand how knowledge evolves, in terms of quality, provenance and reuse.

Tools. Master schema, ontology and taxonomy editors (e.g. ER, UML, OWL), mapping tools, and visualization tools.

Computational modalities. Understand the expressive power of knowledge representation languages.

Knowledge Construction. Build knowledge graphs from unstructured (text), semi-structured (csv, json) and structured data (SQL) following different methodologies.

5.3.3 Current Challenges/Open Research Questions

The current challenges are the unknown themes of knowledge work, which we also categorized by social and technical.

Social

Motivation. Even though knowledge engineering has been a field for many decades, it is still not prominent in enterprises and academic curricula. One can argue that the field has not been able to convince a broad audience of the value. There is a need to understand how to explain the value and what are the consequences of not doing the knowledge work appropriately. Another argument can focus on the maturity and adaptability of the field, prompting the demand for a practical guideline for knowledge engineering. Another aspect to consider is what is the right metaphor for knowledge to help motivate? For example, in data the common metaphors are pipeline and architecture.

Politics of Knowledge. Knowledge of an organization exists within tools and the people who use them. Knowledge may be used as a means to retain power and its value as an organizational asset/capital is often underestimated. Governing and managing knowledge dynamics and sharing becomes of utmost importance: including knowledge as a "first class citizen" in organizational processes means taking care of personal and group behavior, people motivation, change management, continuous stakeholder engagement. If we are to make knowledge a first class citizen, this means that a paradigm shift is required, thus change is imminent. Thus the areas of change management and stakeholder management become a core, especially in organizations where roles are cross-functional and dynamic. Within this core, understanding behaviors and incentives (personal, professional, etc.) becomes key.

Emerging Methodologies. Traditional knowledge engineering methodologies have been designed for a world where there is small data, less complex organizations and technology. We must learn how to adapt and extend methodologies to the scenarios of today. For example, how do we key an eye on blind spots, how do we incorporate concepts of product management, how to deal with our own view vs authoritative, how do we acquire knowledge in a cross-cultural/multilingual environment.

Knowledge Variety. What are the types of knowledge that we are not capturing that we should be capturing? For example, how should we represent knowledge of not only what exists (e.g. descriptive knowledge) but of how things are done (e.g. procedural knowledge)? How do we capture the unknown knowledge (i.e. ignorance)? How do we support people in making explicit the knowledge that comes from experience or intuition?

Epistemic Challenges. There are many aspects that are not right or wrong, but rather "it depends". For example, there is the art and science of modeling. What are the guidelines where we can draw this line? Further challenges are to determine when and how conditions are met for a knowledge engineering system to be deemed trustworthy by users, bearing in mind that distrust can also be epistemically productive in knowledge production. What are the conditions for appropriate distribution of responsibility and meaningful human contribution?

Bridge Across Different Disciplines Knowledge work has been well-studied outside of Computer Science in disciplines such as Library and Information Science, Management Information Systems (MIS) among others. It is critical to bridge these disciplines and define a common nomenclature.

Technical

Advances in Computational Linguistics Advances such as Large Language Models (LLMs) and Generative, AI have the potential to impact knowledge engineering. This is an open area of research where we need to understand where and how an LLM can be used, and even if it is the right thing to do? Where do they fit? Where do they not fit? Are there special types of LLMs that need to be trained? How should knowledge engineering tools be designed with LLMs?

Migration from legacy to modern Legacy tools that contain and manage knowledge, such as wikis, may be prevalent in an organization. How do we bridge and migrate legacy tools to modern tools? When is it necessary, beneficial or redundant to do so?

Nuances in Technology Knowledge work involves people, process and technology. Therefore there may be a lot of nuances on what technology should be used for certain people and processes.

5.3.4 Requirements to move forward

The key requirement we focused on to move forward is the need for educational material. This is a broad requirement; thus the first step is to identify who would be the audience for the education material. We centered on the following dimensions:

University
- Lower bachelors (mandatory)
- Upper bachelors (elective)
- Graduate (advanced)

Enterprise
- People with IT/CS background
- Decision makers / middle management
- Executives / upper management
- Thought leaders

Discipline/Background
- Computer Science
- Data Science
- Information and Library Sciences

- Business Information Systems
- Information Technology
- Digital Economy
- General

Personas/Roles
- Executives
- Information architect
- Enterprise architect
- Data Engineers
- Analytics Engineers
- Data Stewards
- Knowledge Stewards
- Knowledge Engineers
- HR personnel
- Manager
- IT maintenance staff
- Citizens / communities
- Librarians
- Archivist

Expected Background
- Novice
- Intermediate
- Expert

5.3.5 Vision/Possible approaches

We envision having core content defined that can be expressed for the aforementioned audiences in the form of books, manuscripts, etc. We do not envision a single text book given the variety of audience. Additionally, there are plentiful courses on knowledge engineering in universities across the world and this core content can serve as a way of defining a potential standardized course.

This core content is the following, centered around the concept of Knowledge in Action:
- Motivation/Why should we care?
- What is Data?
- What is Knowledge?
- What are the relationships between Data and Knowledge?
- What is the path from Data to Knowledge (and back)?
- How do people (and other agents) collaboratively create, gather, maintain and leverage Knowledge and Data?
- How do we connect Knowledge?
- How do we find/access/share Knowledge?
- Politics of Knowledge
- Verticals / Applications
- Tools

As an exercise, our group came up with several book pitches that resemble the opportunities for context on knowledge work. We believe that this will inspire not only ourselves but a wider community to join forces in order to make knowledge work a first class citizen and work jointly on respective, modern textbooks and educational materials.

Book Pitch 1

By Aidan Hogan, Axel Polleres, Christophe Debruyne, and Eva Blomqvist

How do we make data and knowledge easier to use and process, both for humans and technology? Recognizing that most of the required expertise to answer this question is scattered across many disciplines and seminal works, this book aims at providing a holistic perspective for unleashing knowledge through data. While on the one hand, in organizations we collect and maintain digital traces of various aspects of what we "know", on the other hand many aspects of knowledge are inherently hard to capture. Along these lines, the present book covers the whole spectrum from data to knowledge:

- Details a novel conceptual framework for data-centric knowledge management along with the principles and processes for putting it into practice
- Synthesizes key concepts, methods and techniques from several disciplines spanning the spectrum from data to knowledge as relevant for the emerging ubiquity of A.I.
- Specifies methodologies, languages and tools to modernize how enterprises and organizations capitalize on knowledge
- Addresses the contemporary challenges faced by decision makers, engineers, and knowledge workers.
- Illustrates aspects of data-centric knowledge management in practical scenarios.

Book Pitch 2

By Samaneh Jozashoori and Juan Sequeda

Embracing the Unknown: Igniting business transformation through Knowledge Engineering

In an age where organizations are drenched in data and seeking to embrace A.I., there may seem to be little to no concern about managing this wealth of information. "Embracing the Unknown" aims to shift your perspective, emphasizing the monumental potential and value that lies not just in data, but more importantly, in knowledge.

This book is invaluable to the following readers:

- If you find yourself skeptical about investing in knowledge work like domain modeling, we delve into its depths and why it's essential to investigate the potentials.
- Those who have recognized the presence of knowledge-related challenges but struggle to present these implications convincingly to the broader parts of the organization
- Those who are looking for principled solutions to knowledge challenges, regardless of technology.

Our book helps debunk some common misconceptions about organizational struggles – highlighting that the root cause often lies not with technology, but rather with people and processes.

Most importantly, "Embracing the Unknown" delineates why investing in knowledge work is the cornerstone to succeeding in a highly competitive environment. We showcase how ignoring this vital component equates to unseen financial losses, not being able to navigate successfully in constantly changing environments, and equip you with the approaches not just to confront but to resolve many of today's data and knowledge challenges.

We invite you to navigate this enlightening journey with us to transform your organization landscapes through the potent power of Knowledge Engineering.

Book Pitch 3

By Eva Blomqvist

In today's fast-changing society managing our knowledge is a challenge. But what is knowledge and how does it relate to data? And what is the connection to technology, tools and systems for managing knowledge? This book is the reference for both decision makers and engineers to understand knowledge and data, in the context of modeling and engineering knowledge. We start from the "why?" and end in concrete techniques and methods for putting knowledge into action. The book contains useful examples and concrete guidelines for knowledge practices. After reading the book you will have a cross-disciplinary understanding of knowledge and data, as well as concrete understanding of modern tools to start your own knowledge engineering processes.

Book Pitch 4

By Paul Groth, Claudio Gutierrez, George Fletcher

This book provides a new foundational text for knowledge work, designed for all Computer Science and Data Science students. Just as statistics is a cornerstone of the field, the book situates knowledge work as an indispensable component of all forward-looking curricula. The textbook offers a unified view of knowledge work that is currently lacking in the discipline.

What sets the book apart is the integration of both the social and technical aspects of knowledge into a single framework. Topics are cohesively integrated, drawing from the variety of disciplines and traditions which have studied knowledge work in isolation. The book puts forward a fresh perspective on knowledge work, one that sees it as a collaborative effort between humans and machines, focusing on consensus-making. It embraces the full potential of modern and emerging AI systems, and embodies a contemporary, multi-disciplinary approach to knowledge work with a focus on action and practice. Unlike traditional approaches, we prioritize empirical and qualitative methodologies, ensuring that students are equipped with the perspectives, tools, and insights needed to tackle real-world data and knowledge challenges head-on.

Book Pitch 5

By Ivo Velitchkov

How do we deal with data and knowledge in a data-dominated world populated with knowledge workers? Can we rely on mainstream data management? Or do we expect that AI will take care of it? There is a need for a new kind of knowledge literacy to reduce data waste, bring knowledge center-stage and put it into action. That is what this book does. It goes below and beyond trends and technologies to bring a fundamental understanding of how knowledge needs to be managed within and between organizations. In a world where technologies and applications capture all the attention, this book reverts the focus. Now, knowledge is what matters, and technologies and applications are there to serve it.

5.4 Project Management for Knowledge Graph Construction and Usage

Oscar Corcho (Universidad Politécnica de Madrid – Madrid, ES))
Coen De Roover (VU – Brussels, BE)
Paco Nathan (Derwen – Sebastopol, US)
Souripriya Das (Oracle Corp. – Nashua, US)
Eduard Kamburjan (University of Oslo, NO)
Martin Giese (University of Oslo, NO)
Marina Aguado (European Union Agency for Railways, FR)
David Chaves-Fraga (Universidade de Santiago de Compostela, ES)
Carlos Buil-Aranda (TU Federico Santa María – Valparaíso, CL)
Wouter Beek (Triply BV, NL)

5.4.1 Introduction

The Knowledge Graph construction and usage process has so far not relied on a mechanised process and tool support, which would provide support for the creation and maintenance of all the artefacts used during the process (ontologies, thesauri, declarative mappings, RDF dumps, SHACL files, SPARQL queries, etc.). The concept of dependencies among artefacts has not been addressed yet in a systematic manner.

This working group was formed based on this assumption, understanding the need for addressing the evolution and co-evolution of semantic artefacts and/or program code during the knowledge graph lifecycle.

The following requirements/needs and pportunities served as a starting point for the group and were refined during the time slots dedicated during the seminar:

- Need for a holistic view of all the artefacts used during KG construction and use.
- Need for algorithms and tools that generate dependency maps across all the (semantic) artefacts that are generated during the KG construction and use process: ontologies, mappings, SHACL shapes, SPARQL queries, programming code, etc. This analysis should also identify which kind of messages/problems need to be informed to KG engineers and software developers.
- Need for refactoring support for the project (e.g., simple or complex changes in the ontology highlight potential changes in other artefacts). This will require a deeper analysis of diffs in semantic artefacts.
- Opportunity to create a project specification file that aggregates all these dependency maps and configuration items required to support this process.
- Opportunity for the generation of template-based instantiations of KG projects (create a new KG project and setup the folder structure and all the configuration files).
- Opportunity to create Language Server Protocol (LSP) support, so as to have plugins for tools like Visual Studio Code.
- Opportunity to provide continuous integration support, including the generation of tests.
- Opportunity to ensure the separation of concerns across the different actors (subject-matter experts, ontology engineers, KG engineers, software engineers)

5.4.2 Activities performed during the seminar

During the reserved slots for this working group in the seminar, several activities were done, which led into the discussion items, conclusions and action points that are presented in the following sections. This section provides a summary of those activities:

First, we started by comparing the data processing pipelines and architectures used for the construction of two knowledge graphs in two real applications, one coming from industry (and in this case private and not disclosable outside the context of the working group) and another one coming from a public organisation (namely, the knowledge graph being built for the European Agency for Railways (ERA). From a methodological point of view, we know that using only two use cases is not enough to derive any strong conclusions, but the exercise was useful in order to understand, collectively, the main similarities and differences between the data processing pipelines, and the different types of artefacts used for them.

The analysis allowed generating an initial set of dependencies among artefacts. These were classified between coarse-grained dependencies (e.g., a set of SHACL shapes are derived from the OWL file of an ontology, a file with mappings depends on the ontology and on the source dataset) and fine-grained or vocabulary dependencies (e.g., when a data property is changed into an object property in the ontology, specific mappings, SHACL shapes, SPARQL queries, etc., need to be revised). This led to the proposal of some ideas on how to handle all these dependencies, as it is done in traditional software engineering.

Additional work was done in relation to identifying some key differences between traditional application development and KG-based application development. For instance, one of the conclusions that was clear was that in KG-based applications there is much business logic that is done in the SPARQL queries that are evaluated against the knowledge graph instead of in the code itself, as it happens with traditional software engineering.

5.4.3 Next steps

Several action points were identified as next steps for the members of the group, and are currently ongoing at the time of writing this report.

Creation of template project specification files for KG-based application projects

One of the needs that was identified was related to the fact that there are no template project specification files for typical KG-based application projects. These types of files are common in other software engineering projects, and useful to establish a template of all the artefacts used in the project and the structure of the project itsefl.

These files can be made available using Gradle, Conda or giter8, among others.

One of the actions identified for this group was the creation of such template specification files for these types of project, covering at least the coarse-grained dependencies among artefacts used. This project specification file may be used later by those tools that will calculate fine-grained dependencies

Identification of fine-grained dependencies

Based on the literature available on ontology changes, it was proposed to have an action point related to the identification of the types of changes in the different artefacts used during knowledge graph creation and knowledge-graph application development which may have a cascading impact into other artefacts.

As a result of this activity, a map of dependencies among artefacts used in these projects can be proposed. Normally, this may normally start with the change in an ontology (e.g., add/delete/rename class C, add/delete subclass C1, add/delete subproperty P1, add/delete/remove P, add deprecated class C), although it may be also derived from changes in other artefacts.

Once identified, one potential way to detect these dependencies may be by using SPARQL queries across different artefacts, which are normally implemented in RDF, making this possible. For instance, the following query shows the classes ?x that have some mapping ?y in some mapping file ?g:

```
SELECT DISTINCT ?x ?g ?y WHERE {
GRAPH <.../ont> {?x a owl:Class}
GRAPH ?g {
?y a rr:TriplesMap .
?y rr:subjectMap ?s .
?s rr:class ?x}}
```

The OPTIONAL clause may be used for detecting coverage problems (classes defined in the ontology that are not covered in mappings, or viceversa).

Organising a workshop

One of the action points derived from the working group was the organisation of a workshop for the following International Semantic Web Conference (ISWC2024), with the objective of bringing together practitioners in knowledge graph creation, and collecting experiences from them. The workshop has been accepted for ISWC2024, with the title Software Lifecycle Management for Knowledge Graphs Workshop (SofLiM4KG). Next we provide the current text for the call for contributions:

Knowledge graphs result from a complex construction process utilizing numerous tools and data sources, generated in elaborate pipelines utilizing a wide variety of semantic technologies, such as RML, OTTR, SHACL... All the involved artifacts, ontologies, mapping scripts, graph shapes, etc., are interdependent and changes in one of them require adjustment in others.

In current practice, managing the dependencies and artifacts is a manual process using ad hoc approaches. Despite the numerous work on KG construction, there is a focus on the technical aspects of the single steps and little attention has been paid to the practical aspects of (a) organizing and managing knowledge graphs projects in terms of change management, dependencies between semantic artifacts, as well as DevOps for knowledge graphs, and (b) automating building and deploying of the resulting knowledge graph and adjacent artifacts.

The Software Lifecycle Management for KG workshop aims to collect experiences in successful and abandoned knowledge graph projects from this perspective to (a) carve out the specifics in knowledge graph engineering that pose challenges beyond software engineering practices, (b) to establish best practices and anti-patterns for the community, and (c) build the foundations for the systematic investigation of the connection to software engineering, as well as qualitative and quantitative studies in project management of knowledge graphs.

The topics of the workshop can be summarised as:

- Best practices and experience reports from managing knowledge graph
- Reports of problems and hindrances in abandoned knowledge graph projects
- Tools for mechanizing building of knowledge graphs
- Tools for mechanizing maintenance of knowledge graphs

- Applications of DevOps or agile patterns in development
- Connections to software engineering practices, such as build tools
- Dependency management among semantic artifacts used in knowledge graph construction and maintenance
- Methods of testing applications based on knowledge graphs

6 Emerging Discussions

6.1 Knowledge Management and Knowledge Graphs

Axel Polleres (WU Vienna, Austria, axel.polleres@wu.ac.at)
Irene Celino (Cefriel, Italy, irene.celino@cefriel.com)

In discussing the relationship between Knowledge Graphs and the "real world", according to the seminar title, most of the participants and working groups naturally ended up discussing about Knowledge *Engineering* practices, since Knowledge Graphs – intended as digital artifacts – are usually the product of socio-technical practices aimed at the conceptualisation, representation, and exchange of *explicit* knowledge in a digital form.

On the other hand, activities to create, update and manage Knowledge Graphs could be explored and investigated also through the lense of Knowledge Management, the discipline rooted in organizational sciences, which aims to capture the processes of knowledge creation and transfer, especially in the endeavour to transform *tacit* knowledge into explicit knowledge.

Therefore, we think that some of the challenges that the Knowledge Graphs community is currently facing – including those discussed in the present seminar – could benefit from a Knowledge Management perspective. Some open questions that we should aim to answer are:

- What different kinds of knowledge [1] are already included or could be represented in Knowledge Graphs (e.g., declarative, relational, procedural, conditional, etc.)?
- Can Knowledge Graph creation (and Knowledge Engineering processes) be framed in the well-known SECI model of knowledge transformation [2]? How to make tacit knowledge explicit in and by means of Knowledge Graphs?
- How can Knowledge Graphs be successfully employed as a means to provide people and organizations with knowledge intended as the "capacity to act" [3]? I.e., how can Knowledge Graphs have a concrete and measurable impact in making knowledge *actionable* also from an organizational point of view?

We invite interested colleagues to work and collaborate on these questions, with an aim at (a) providing a survey of existing literature on different perspectives on types of Knowledge and tasks in Knowledge Management from a holistic perspectives, and (b) systematic evaluations of tools and use cases in the field of Knowledge Engineering, mapping out to which of these tasks and perspectives our technologies contribute.

References

1 Alavi, M. & Leidner, D. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*. pp. 107-136 (2001)
2 Nonaka, I., Toyama, R. & Konno, N. SECI, Ba and leadership: a unified model of dynamic knowledge creation. *Long Range Planning*. **33**, 5-34 (2000)
3 Sveiby, K. A knowledge-based theory of the firm to guide in strategy formulation. *Journal Of Intellectual Capital*. **2**, 344-358 (2001)

7 Conclusions

In conclusion, this seminar stands as a pivotal gathering that convened researchers and industry partners from diverse backgrounds. Together, we explored the complexities, challenges, and advancements inherent in managing and leveraging knowledge graphs within real-world contexts. Spanning from technical considerations to social dimensions, we identified essential requirements, imperatives, and actionable strategies necessary to foster the development of a new generation of knowledge graph ecosystems.

Given the advent of generative AI and its demonstrated benefits when integrated with intricate data structures such as knowledge graphs, ensuring readiness across all facets of knowledge graph implementation is paramount. The convergence of knowledge graphs with emerging technologies presents novel avenues for advancing knowledge representation, reasoning, and applications. Our discussions underscored the significance of robust quality assessment mechanisms and stressed the importance of integrating human expertise and feedback loops throughout the knowledge graph lifecycle. From an educational standpoint, it is imperative for experts to disseminate their knowledge through educational programs tailored to different levels of learning and professional training. However, standardizing competencies across all levels is essential to ensure a uniform understanding of fundamental concepts among potential knowledge graph practitioners.

Participants

- Marina Aguado
European Union Agency for Railways – Valenciennes, FR
- Wouter Beek
Triply B.V. – Bussum, NL
- Eva Blomqvist
Linköping University, SE
- Piero Andrea Bonatti
University of Naples, IT
- Carlos Buil-Aranda
Thuban Technology Services – Madrid, ES
- Cinzia Cappiello
Polytechnic University of Milan, IT
- Irene Celino
CEFRIEL – Milan, IT
- Pierre-Antoine Champin
INRIA – Sophia Antipolis, FR
- David Chaves-Fraga
Universidade de Santiago de Compostela, ES
- Oscar Corcho
Universidad Politécnica de Madrid, ES
- Souripriya Das
Oracle Corp. – Nashua, US
- Coen De Roover
VU – Brussels, BE
- Christophe Debruyne
University of Liège, BE
- Anastasia Dimou
KU Leuven, BE
- Michel Dumontier
Maastricht University, NL
- George Fletcher
TU Eindhoven, NL
- Sandra Geisler
RWTH Aachen, DE
- Martin Giese
University of Oslo, NO
- Paul Groth
University of Amsterdam, NL
- Claudio Gutierrez
University of Chile – Santiago de Chile, CL
- Peter Haase
Metaphacts GmbH – Walldorf, DE
- Olaf Hartig
Linköping University, SE
- Aidan Hogan
University of Chile – Santiago de Chile, CL
- Katja Hose
TU Wien, AT
- Ana Iglesias-Molina
Universidad Politécnica de Madrid, ES
- Samaneh Jozashoori
Metaphacts GmbH – Walldorf, DE
- Eduard Kamburjan
University of Oslo, NO
- Sabrina Kirrane
Wirtschaftsuniversität Wien, AT
- Craig A. Knoblock
USC – Marinael Rey, US
- Maurizio Lenzerini
Sapienza University of Rome, IT
- Vanessa López
IBM Research – Dublin, IE
- Paco Nathan
Derwen – Sebastopol, US
- Edelmira Pasarella
UPC Barcelona Tech, ES
- Axel Polleres
Wirtschaftsuniversität Wien, AT
- Anisa Rula
University of Brescia, IT
- Juan F. Sequeda
data.world – Austin, US
- Dylan Van Assche
IDLab - Ghent University - imec, BE
- Ivo Velitchkov
Brussels, BE
- Maria-Esther Vidal
TIB – Hannover, DE

Report from Dagstuhl Seminar 24062

Beyond-Planar Graphs: Models, Structures and Geometric Representations

Vida Dujmović*[1], Seok-Hee Hong*[2], Michael Kaufmann*[3], János Pach*[4], and Henry Förster†[5]

1 **University of Ottawa, CA.** `vida@cs.mcgill.ca`
2 **The University of Sydney, AU.** `seokhee.hong@sydney.edu.au`
3 **Universität Tübingen, DE.** `michael.kaufmann@uni-tuebingen.de`
4 **Alfréd Rényi Institute – Budapest, HU & EPFL – Lausanne, CH.** `pach@renyi.hu`
5 **Universität Tübingen, DE.** `henry.foerster@uni-tuebingen.de`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24062 "Beyond-Planar Graphs: Models, Structures and Geometric Representations". The seminar investigated beyond-planar graphs, in particular, their combinatorial and topological structures, computational complexity and algorithmics for recognition, geometric representations, and their applications to real-world network visualization. Compared to the previous two editions of the seminar, we focus more on aspects of combinatorics and geometry.

The program consists of four invited talks on beyond planar graphs, open problem session, problem solving sessions and progress report sessions. Specific open problems include questions regarding the combinatorial structures and topology (e.g., k^+-real face graphs, beyond upward planar graphs, sparse universal geometric graphs, local-crossing-critical graphs), the geometric representations (e.g., constrained outer string graphs, rerouting curves on surface), and applications.

The details of the invited talks and progress reports from each working groups are included in this report.

Seminar February, 4–9, 2024 – https://www.dagstuhl.de/24062

2012 ACM Subject Classification Mathematics of computing → Graph theory; Theory of computation → Design and analysis of algorithms

Keywords and phrases Combinatorial geometry, Graph algorithm, Graph drawing, Graph theory, Network visualization

Digital Object Identifier 10.4230/DagRep.14.2.71

1 Executive Summary

Vida Dujmović (University of Ottawa, CA, vida@cs.mcgill.ca)
Seok-Hee Hong (The University of Sydney, AU, seokhee.hong@sydney.edu.au)
Michael Kaufmann (Universität Tübingen, DE, michael.kaufmann@uni-tuebingen.de)
János Pach (Alfréd Rényi Institute – Budapest, HU & EPFL – Lausanne, CH, pach@renyi.hu)

Many big data sets in various application domains have complex relationships, which can be modeled as *graphs*, consisting of entities and relationships between them. Consequently, graphs are extensively studied in both mathematics and computer science. In particular,

* Editor / Organizer
† Editorial Assistant / Collector

planar graphs, which can be drawn without edge crossings in the plane, form a distinguished role in graph theory and graph algorithms. Many structural properties of planar graphs are investigated in terms of excluded minors, low density, and small separators, leading to efficient planar graph algorithms. Consequently, fundamental algorithms for planar graphs have been discovered.

However, most real-world graphs, such as social networks and biological networks, are *nonplanar*. For example, the scale-free networks, which are used to model web graphs, social networks, and biological networks, are globally sparse nonplanar graphs with locally dense clusters and low diameters. To understand such real-world networks, we must solve fundamental mathematical and algorithmic research questions on *beyond-planar graphs*, which generalize the notion of planar graphs regarding topological constraints or forbidden edge crossing patterns.

This Dagstuhl Seminar investigated beyond-planar graphs, in particular, their combinatorial and topological structures (i.e., density, thickness, crossing pattern, chromatic number, queue number, and stack number), computational complexity and algorithmics for recognition, geometric representations (i.e., straight-line drawing, polyline drawing, intersection graphs), and their applications to real-world network visualization.

Compared to the previous two editions of the seminar, we focus more on aspects of combinatorics and geometry. Therefore, we included one new organizer and more participants from the corresponding fields. Thirty-two participants accepted the invitation to participate and arrived on Sunday afternoon.

On Monday morning, the program started with an introduction of all participants, followed by four invited talks to provide fundamental background knowledge on related research fields. We organized an open problems session on Monday afternoon and formed new working groups for research collaboration.

Many new problems related to combinatorics and geometry of beyond-planar graphs have been proposed. Specific open problems include questions regarding the combinatorial structures and topology (e.g., k^+-real face graphs, beyond upward planar graphs, sparse universal geometric graphs, local-crossing-critical graphs), the geometric representations (e.g., constrained outer string graphs, rerouting curves on the surface), and applications.

Two progress report sessions were organized on Tuesday and Thursday afternoons to report progress and plans for future publications and follow-up meetings among researchers. From the participants' feedback, the seminar has initiated new research collaboration and led to new research ideas and directions.

Taking this opportunity, we thank Schloss Dagstuhl for providing an environment for fruitful research collaboration.

2 Table of Contents

3 Overview of Talks

3.1 Crossing numbers of crossing-critical graphs

Géza Tóth (Alfréd Rényi Institute of Mathematics – Budapest, HU, geza@renyi.hu)

Joint work of Géza Tóth, János Barát

A graph G is k-crossing-critical if $\mathrm{cr}(G) \geq k$, but for any edge e of G, $\mathrm{cr}(G-e) < k$. In 1993 Richter and Thomassen conjectured that for any k-crossing-critical graph G, $\mathrm{cr}(G) \leq k + c\sqrt{k}$ and proved that $\mathrm{cr}(G) \leq 5k/2 + 16$. We improve it to $\mathrm{cr}(G) \leq 2k + 6\sqrt{k} + 47$.

3.2 The Density Formula for Beyond-Planar Graph Classes

Torsten Ueckerdt (Karlsruhe Institute of Technology, DE, torsten.ueckerdt@kit.edu)

Joint work of Michael Kaufmann, Boris Klemz, Kristin Knorr, Meghana M. Reddy, Felix Schröder, Torsten Ueckerdt
Main reference Michael Kaufmann, Boris Klemz, Kristin Knorr, Meghana M. Reddy, Felix Schröder, Torsten Ueckerdt: "The Density Formula: One Lemma to Bound Them All", CoRR, Vol. abs/2311.06193, 2023.
URL https://doi.org/10.48550/ARXIV.2311.06193

We introduce the Density Formula for drawings of graphs on the sphere, which can be used to derive tight upper bounds for the density (maximum number of edges for given number of vertices) of several beyond-planar graph classes, such as 1- and 2-planar graphs, fan-planar graphs, k-bend RAC graphs, and quasiplanar graphs. While in some cases we even obtain the first tight upper bounds, the real strength of the Density Formula is its simplicity and versatility. In this talk, I showcase the Density Formula with a few examples and mention a few open problems that seem worth investigating next.

3.3 Connected Dominating Sets in Triangulations

Pat Morin (Carleton University – Ottawa, CA, morin@scs.carleton.ca)

Joint work of Prosenjit Bose, Vida Dujmovic, Hussein Houdrouge, Pat Morin, Saeed Odak
Main reference Prosenjit Bose, Vida Dujmovic, Hussein Houdrouge, Pat Morin, Saeed Odak: "Connected Dominating Sets in Triangulations", CoRR, Vol. abs/2312.03399, 2023.
URL https://doi.org/10.48550/ARXIV.2312.03399

We show that every n-vertex triangulation has a connected dominating set of size at most $10n/21$. Equivalently, every n vertex triangulation has a spanning tree with at least $11n/21$ leaves. Prior to the current work, the best known bounds were $n/2$, which follows from work of Albertson, Berman, Hutchinson, and Thomassen (J. Graph Theory **14**(2):247–258). One immediate consequence of this result is an improved bound for the SEFENOMAP graph drawing problem of Angelini, Evans, Frati, and Gudmundsson (J. Graph Theory **82**(1):45–64). As a second application, we show that for every set P of $\lceil 11n/21 \rceil$ points in $\mathbb{R}^2$ every n-vertex planar graph has a one-bend non-crossing drawing in which some set of $11n/21$

vertices is drawn on the points of P. The main result extends to n-vertex triangulations of genus-g surfaces, and implies that these have connected dominating sets of size at most $10n/21 + O(\sqrt{gn})$.

3.4 Beyond-planar Euclidean spanners

Csaba D. Tóth (California State University – Northridge, US, csaba.toth@csun.edu)

Main reference Csaba D. Tóth: "Minimum weight Euclidean $(1+\varepsilon)$-spanners", European Journal of Combinatorics, Vol. 118, p. 103927, 2024.
URL https://doi.org/10.1016/j.ejc.2024.103927

For a set P of n points in the plane and a parameter $t \geq 1$, a t-spanner is a geometric graph G such that for all pairs $u, v \in P$, the shortest path distance in G (with Euclidean edge weights) approximates the Euclidean distance between u and v up to a factor of at most t; the parameter t is the *stretch* of H. For example, the Delaunay triangulation is 1.998-spanner, but in general plane graphs on P cannot achieve a stretch less than $\pi/2$. If edge crossings are allowed, the stretch can be arbitrarily close to 1: For every $\varepsilon > 0$ there are $(1+\varepsilon)$-spanners with $O(\varepsilon^{-1}n)$ edges and $\tilde{O}(\varepsilon^{-2}) \cdot MST(P)$ weight. These bounds are the best possible, and such spanners also have separators of size $\varepsilon^{-O(1)}\sqrt{n})$. However, it remains an open problem to quantify, in terms of $\varepsilon > 0$, how much $(1+\varepsilon)$-spanners are beyond planar graphs.

4 Working Groups

4.1 Constrained Outerstring Graphs

Therese Biedl (University of Waterloo, CA, biedl@uwaterloo.ca)
Sabine Cornelsen (University of Konstanz, DE, sabine.cornelsen@uni-konstanz.de)
Jan Kratochvíl (Charles University Prague, CZ, honza@kam.mff.cuni.cz)
Ignaz Rutter (Universität Passau, DE, ignaz.rutter@uni-passau.de)

In an *outer-string representation* [6] (implicitly defined and first results obtained in [9]) of a graph each vertex is drawn as a simple curve $\partial(v)$ within a simple closed region D such that one end point of $\partial(v)$, called the anchor of v, is on the simple closed curve bounding D. The curves $\partial(v)$ and $\partial(w)$ of two vertices v and w intersect if and only if v and w are adjacent. It is NP-hard to decide whether a graph has an outer-string representation [8]. Unfortunately, outer-string representations sometimes need exponentially many crossings [1]. So it is interesting to investigate which graphs allow an outer-string representation with a restricted number of crossings. In an *outer-1-string representation*, it is additionally required that the curves $\partial(v)$ and $\partial(w)$ of two vertices v and w intersect at most once. This is similar to the intersection graph of pseudosegments [4], however, with the additional constraint that the anchors still have to be on the boundary of a simple closed region containing all pseudosegments. Representing chordal graphs as intersections of pseudosegments was considered in [3].

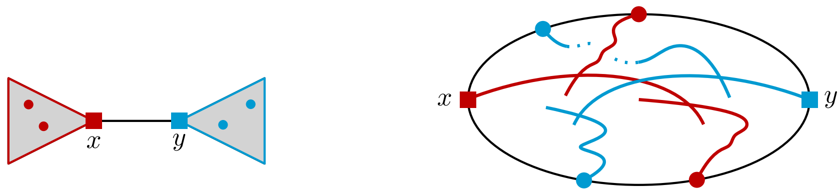


Figure 1 Triply interleaved bridge.

In [2], the order of crossings along a string was constrained. We focus on the constrained version where the cyclic order of the anchors is fixed, i.e., an instance of constrained outer-(1)-string representation consists of a graph and a cyclic order of the vertices. In addition to general outer-string and outer-1-string representations, we also consider *L-shaped* [7, 5] and *U-shaped representations* in which the anchors are on a horizontal line and the vertices are 1- or 2-bend orthogonal polylines below that line. I.e., in particular, we also allow ⅃s. See Figures 3b and 3c. In the constrained version, the *linear* order of the anchors is fixed.

▶ **Theorem 1** ([9]). *The complement of a simple cycle with at least four vertices does not have a constrained outer-string representation, i.e., if the cyclic order of the vertices is* $v_1, \dots, v_n$ *then the graph with edge set* $E = \{\{v_i, v_j\}; |i-j| \notin \{1, n-1\}\}$ *does not have a constrained outer-string representation.*

4.1.1 Summary of Results

We say that two sets V_1 and V_2 of vertices are *interleaved* if in the (cyclic) order no two vertices of V_1 nor two vertices of V_2 are consecutive. Observe that the complement of a 4-cycle consists of two *interleaved independent edges.*

▶ **Theorem 2.** *A chordal graph with a fixed cyclic order of the vertices admits a constrained outer-string representation if and only if it contains no two interleaved independent edges.*

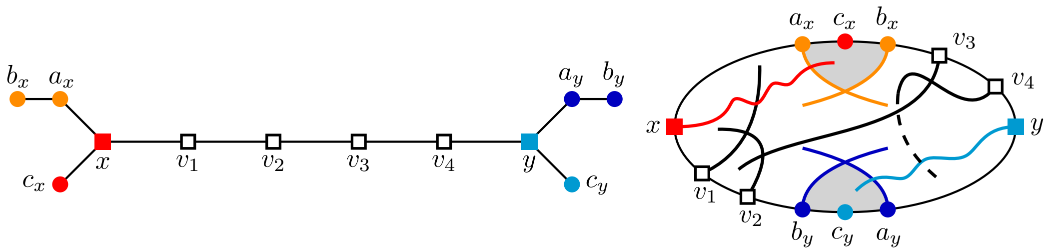
The following instances do not admit a constrained outer-1-string representations: (a) a *triply interleaved bridge*, i.e., a bridge e of the graph G, such that the two connected components of $G - e$ containing the end vertices of e each contain a set X and Y of three vertices such that X and Y are interleaved. See Figure 1. (b) An *X-obstruction*; see Figure 2.

▶ **Theorem 3.** *A tree with a fixed cyclic order of the vertices admits a constrained outer-1-string representation if and only if it contains no two interleaved independent edges, no triply interleaved bridge, nor an X-obstruction. Moreover, for trees there is a certifying polynomial-time recognition algorithm, which either outputs a constrained outer-1-string representation or an obstruction.*

An *extended complement of a 5-cycle* is either the complement of a 5-cycle or a subpath $w_1 v' v u u' w_2$ of a cycle whose anchors are in the order $w_1 u v' u' v w_2$. See Figure 3a.

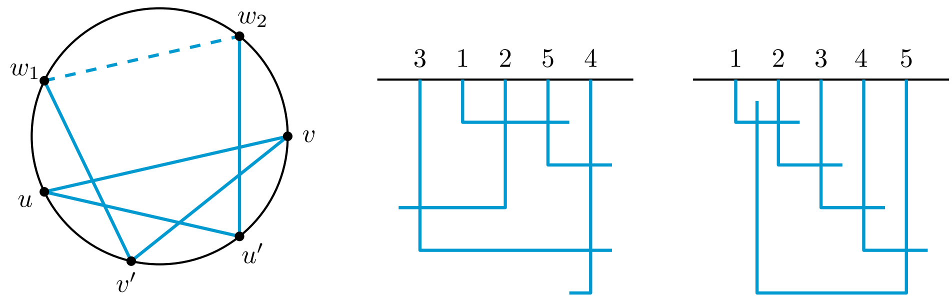
▶ **Theorem 4.** *Let* $G = (V, E)$ *be a simple cycle and let* $\prec$ *be a cyclic order of* V*. Then the following are equivalent.*

1. $(G, \prec)$ *has a constrained outer-1-string representation*
2. *For every edge* $\{u, v\}$ *of* G *one of the following sequences* uv, $uu'v'v$, $uu'v$, *or* $uv'v$, *or their reverse is a subsequence of* $\prec$, *where* u' *and* v' *are the neighbors of* u *and* v *other then* v *and* u, *respectively.*



(a) graph **(b)** forbidden cyclic ordering

Figure 2 An *X-obstruction* contains the vertices i, a_i, b_i, c_i, d_i and the edges ia_i, a_ib_i, ic_i for $i = x, y$ as well as an x-y path $x = v_0, v_1, \ldots, v_{\ell-1}, v_\ell = y$ of arbitrary length, including zero. For any $k = -1, \ldots, \ell$, the set $\{v_0, \ldots, v_k, a_x, b_x, c_x\}$ of vertices appears consecutive (not necessarily in this order) in the cyclic order $\prec$ and for $i = x, y$ the pairs $\{a_i, b_i\}$ and $\{i, c_i\}$ are interleaved.



(a) extended compl. of 5-cycle **(b)** L-shaped **(c)** U-shaped

Figure 3 An obstruction and two representations of a 5-cycle.

3. $(G, \prec)$ *does not contain two interleaving independent edges nor an extended complement of a 5-cycle.*

Observe that a path has a constrained L-shaped outer-1-string representation if there are no two independent edges that are interleaved. Every simple cycle with a fixed linear order of the vertices that admits a constrained outer-1-string representation also admits a constrained U-shaped outer-1-string representation.

▶ **Theorem 5.** *It can be tested in polynomial time whether a graph with a given ordering of the vertices admits a constrained L-shaped outer-1-string representation.*

4.1.2 Open Problems

What is the complexity of testing whether a graph has an outer-1-string, a constrained outer-1-string, or a constarined outer-string representation? What if the instances are restricted to graphs with bounded treewidth?

References

1 Therese Biedl, Ahmad Biniaz, and Martin Derka. On the size of outer-string representations. In *16th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT*, volume 101 of *LIPIcs*, pages 10:1–10:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPICS.SWAT.2018.10.

2 Therese Biedl and Martin Derka. Order-preserving 1-string representations of planar graphs. In *SOFSEM 2017: Theory and Practice of Computer Science*, volume 10139 of *Lecture Notes in Computer Science*, pages 283–294. Springer, 2017. doi:10.1007/978-3-319-51963-0_22.

3 Cornelia Dangelmayr, Stefan Felsner, and William T. Trotter. Intersection graphs of pseudosegments: Chordal graphs. *J. Graph Algorithms Appl.*, 14(2):199–220, 2010. doi:10.7155/JGAA.00204.
4 Gideon Ehrlich, Shimon Even, and Robert Endre Tarjan. Intersection graphs of curves in the plane. *J. Comb. Theory, Ser. B*, 21(1):8–20, 1976. doi:10.1016/0095-8956(76)90022-8.
5 Vít Jelínek and Martin Töpfer. On grounded L-graphs and their relatives. *Electron. J. Comb.*, 26(3):P3.17, 2019. doi:10.37236/8096.
6 Jan Kratochvíl. String graphs. In *Graphs and Other Combinatorial Topics, Proceedings Third Czechoslovak Symposium on Graph Theory, Prague*, volume 59 of *Teubner Texte zur Mathematik*, pages 168–172. Teubner, Berlin, 1982.
7 Sean McGuinness. On bounding the chromatic number of L-graphs. *Discret. Math.*, 154(1-3):179–187, 1996. doi:10.1016/0012-365X(95)00316-O.
8 Alexandre Rok and Bartosz Walczak. Outerstring graphs are χ-bounded. *SIAM J. Discret. Math.*, 33(4):2181–2199, 2019. doi:10.1137/17M1157374.
9 F. W. Sinden. Topology of thin film RC circuits. *Bell System Tech. J.*, 45:1639–1662, 1966. doi:10.1002/j.1538-7305.1966.tb01713.x.

4.2 Universal Geometric Graphs

Vida Dujmović (University of Ottawa, CA, vida.dujmovic@uottawa.ca)
Fabrizio Frati (Roma Tre University, IT, fabrizio.frati@uniroma3.it)
Michael Hoffmann (ETH Zürich, CH, hoffmann@inf.ethz.ch)
Pat Morin (Carleton University, CA, morin@scs.carleton.ca)

4.2.1 Problem statement

A *drawing* of a graph is a mapping of each vertex to a point in the plane and of each edge to a Jordan arc between its endvertices. A drawing is *straight-line* if each edge is represented by a straight-line segment and it is *planar* if no two edges intersect, except at common endvertices. A *planar graph* is a graph that admits a planar drawing. An *embedding* of a graph is a planar straight-line drawing of it. Every planar graph admits an embedding [15, 23].

A *geometric graph* is a graph whose vertices are points and whose edges are straight-line segments. A geometric graph is *planar* if it defines an embedding of the underlying (abstract) graph. A geometric graph is *universal* for a family $\mathcal{F}$ of planar graphs if it contains an embedding of every graph in $\mathcal{F}$. That is, for every graph $G \in \mathcal{F}$, there exists a subgraph of the universal graph which is isomorphic to G and is planar.

The question we study is the following.

▶ Problem 1. Let $f(n)$ be the minimum number of edges of any geometric graph that is universal for the family of the n-vertex planar graphs. What is the asymptotic growth of $f(n)$?

4.2.2 Related results

Universality has long been studied from a graph-theoretic perspective, starting from a paper by Rado in the 1960s [21]. An (abstract) graph is *universal* for a family $\mathcal{F}$ of graphs if it contains every graph in $\mathcal{F}$ as a subgraph. Clearly, the complete graph K_n with n vertices is universal for any family $\mathcal{H}$ of n-vertex graphs. Henceforth, research has been conducted on determining upper and lower bounds on the number of edges that universal graphs for notable families of n-vertex sparse graphs must have. Babai et al. [4] proved that a universal graph with $O(m^2 \log\log m/\log m)$ edges exists for the family of all graphs with m edges, whereas any such a universal graph has $\Omega(m^2/\log^2 m)$ edges. Alon et al. [1, 2] proved that there exists a universal graph with $O(n^{2-2/k})$ edges for the n-vertex graphs with maximum degree k; such a bound is tight in the worst case.

A special attention has been devoted to planar graphs and their subclasses. It has long been known [4] that there exists a graph with $O(n^{3/2})$ edges that is universal for the family of the n-vertex planar graphs. This bound was recently improved to $n \cdot 2^{O(\sqrt{\log n \cdot \log\log n})}$ by Esperet et al. [14]. For bounded-degree planar graphs, there exists an (optimal) $O(n)$ bound, due to Capalbo [9]. Böttcher et al. [7, 8] proved that every n-vertex graph with minimum degree $\Omega(n)$ is universal for the n-vertex planar graphs of bounded degree. Chung and Graham [11, 12] constructed a universal graph with $O(n \log n)$ edges for the n-vertex trees. This bound is the best possible, apart from constant factors.

Universal geometric graphs were first defined and studied by Frati, Hoffmann, and Tóth [17]. They strengthened Chung and Graham result [11, 12] by proving that there exists an n-vertex geometric graph with $O(n \log n)$ edges that is universal for the n-vertex trees. They also proved that every n-vertex convex geometric graph that is universal for the n-vertex outerplanar graphs has $\Omega_h(n^{2-1/h})$ edges, for every positive integer h, which almost matches the trivial $O(n^2)$ upper bound given by a convex complete geometric graph.

The study of universal geometric graphs has a strong relationship with the study of universal point sets. A set $\mathcal{P}$ of points is *universal* for a family $\mathcal{F}$ of planar graphs if every graph in $\mathcal{F}$ has an embedding in which the vertex set is mapped to a subset of $\mathcal{P}$. The question is then, for a family $\mathcal{F}$ of n-vertex planar graphs, what is the asymptotic growth of the function representing the minimum number of points of a universal point set for the graphs in $\mathcal{F}$. Answering such a question for the family of all n-vertex planar graphs is perhaps the most famous graph drawing open problem. It is has been known for a long time that there exists a universal point set for the n-vertex planar graphs with $O(n^2)$ points [13], see also [5], while the currently best known lower bound is only linear, namely $(1.293 - o(1))n$ [22]; see also [10, 20]. Universal point sets with sub-quadratic size are known for the 2-outerplanar graphs and the simply nested graphs [3], and for the n-vertex stacked triangulations [18]. Linear-size universal point sets are known for the n-vertex outerplanar graphs [6, 19], as well as for the cubic planar graphs and the bipartite planar graphs [16].

Consider a point set $\mathcal{P}$ which is universal for a family $\mathcal{F}$ of planar graphs. Then the complete geometric graph with vertex set $\mathcal{P}$ is universal for $\mathcal{F}$. This connection, together with the existence of a quadratic-size universal point set for the n-vertex planar graphs, gives us an $O(n^4)$ upper bound on the number of edges of a universal geometric graph for the n-vertex planar graphs, which is the best known upper bound we are aware of for Problem 1. On the other hand, the best known lower bound is only $\Omega(n \log n)$, which comes from the described graph-theoretic setting [11, 12].

4.2.3 Our research

Our research aimed at finding an upper bound better than $O(n^4)$ for Problem 1. We now explain the strategy we pursued in order to achieve such a goal.

As already mentioned, de Fraysseix, Pach, and Pollack proved the existence of a universal point set $\mathcal{P}$ with $O(n^2)$ points (in fact, a $2n \times n$ section of the integer lattice) for the n-vertex planar graphs [13]. The embedding of any n-vertex planar graph G on $\mathcal{P}$ can be constructed incrementally as follows. First, one can assume without loss of generality that G is a *maximal plane graph.* Indeed, maximality can be guaranteed by an initial edge-augmentation. Furthermore, a maximal planar graph has a unique combinatorial embedding (this is the circular order of the incident edges in an embedding); this, together with a choice of the outer face, enhances G to a maximal plane graph. Second, every maximal plane graph G with $n \geq 3$ vertices and with outer face (u, v, z) admits a *canonical ordering*. This is a labeling of the vertices $v_1 = u, v_2 = v, v_3, \ldots, v_{n-1}, v_n = z$ meeting the following requirements for every $k = 3, \ldots, n$:

- The plane subgraph $G_k \subseteq G$ induced by $v_1, v_2, \ldots, v_k$ is 2-connected; let C_k be the cycle bounding its outer face;
- v_k is in the outer face of G_{k-1}, and its neighbors in G_{k-1} form an (at least 2-element) subinterval of the path $C_{k-1} - (u, v)$.

A *canonical drawing* of G can be constructed from a canonical ordering of G in $n-2$ steps. At step 1, a planar straight-line drawing Γ_3 of G_3 is constructed with v_1 at $(0, 0)$, with v_2 at $(2, 0)$, and with v_3 at $(1, 1)$. Auxiliary sets $M_3(v_1) := \{v_1, v_2, v_3\}$, $M_3(v_3) := \{v_2, v_3\}$, and $M_3(v_2) := \{v_2\}$ are also defined. For $k = 4, \ldots, n$, at step $k - 2$, a planar straight-line drawing Γ_k of G_k is constructed from Γ_{k-1}, as follows. Let $w_1 = u, w_2, \ldots, w_r = v$ be the clockwise order of the vertices along the outer face of G_{k-1}, where $w_p, w_{p+1}, \ldots, w_q$ are the neighbors of v_k in G_{k-1}, for some $1 \leq p < q \leq r$. Then Γ_k is constructed from Γ_{k-1} by "shifting" the vertices in $M_{k-1}(w_{p+1})$ by one unit to the right, by shifting the vertices in $M_{k-1}(w_q)$ by one additional unit to the right, and by placing v_k at the intersection point of the line through w_p with slope $+1$ and of the line through w_q with slope -1. Step $k - 2$ is completed by defining the sets:

- $M_k(w_i) = M_{k-1}(w_i) \cup \{v_k\}$, for $i = 1, \ldots, p$;
- $M_k(v_k) = M_{k-1}(w_{p+1}) \cup \{v_k\}$; and
- $M_k(w_i) = M_{k-1}(w_i)$, for $i = q, \ldots, r$.

Note that the above described construction maintains the x-monotonicity of the boundary of the drawing at every step. The shifting of the vertices in the sets $M_{k-1}(w_{p+1})$ and $M_{k-1}(w_q)$ makes room for drawing the edges incident to the newly inserted vertex v_k in a planar way.

The starting observation of our approach is that a canonical ordering of G can be used in a much simpler way to obtain a planar straight-line drawing of G, entirely avoiding the shifting phase and the definition of the sets $M_{\cdot}(\cdot)$. Indeed, because of the x-monotonicity of the boundary of the drawing, one can simply place v_k at a "sufficiently high" point in the interior of the x-interval spanned by its neighbors $w_p, w_{p+1}, \ldots, w_q$. This ensures planarity and maintains the x-monotonicity of the boundary of the drawing. We call *generalized canonical drawing* a drawing constructed in this way.

Now, consider an $n \times n$ *stretched grid*. This is a point set obtained from an $n \times n$ section of the integer lattice by translating grid rows upwards, in such a way that each point is above the line through any two points in lower rows that are not vertically aligned. Stretched grids were used in [18]. It can be proved that every n-vertex maximal plane graph G has a generalized canonical drawing in which the vertex set is mapped to a subset of any $n \times n$ stretched grid $\mathcal{S}$; thus, $\mathcal{S}$ is a universal point set for the n-vertex planar graphs. This can be proved as follows. First, compute a canonical ordering $v_1, v_2, \ldots, v_n$ of G. Second, define a partial order Y of the vertices of G iteratively, so that each vertex v_k follows all its neighbors $w_p, w_{p+1}, \ldots, w_q$ in G_{k-1}. Third, define a partial order X of the vertices of G iteratively, so that each vertex v_k follows its first neighbor w_p and precedes its last neighbor w_q in G_{k-1}. It is easy to see that any assignment of the vertices of G to the points of $\mathcal{S}$ such that:

- if a vertex u precedes a vertex v in Y, then u is assigned to a lower row than v; and
- if a vertex u precedes a vertex v in X, then u is assigned to a column to the left of the one of v

results in a generalized canonical drawing of G whose vertex set lies at $\mathcal{S}$.

Our intuition is that a universal geometric graph $\mathcal{G}$ with $o(n^4)$ edges that is universal for the n-vertex planar graphs can be constructed so that its vertex set is an $n \times n$ stretched grid $\mathcal{S}$, possibly slightly perturbed so that each row defines a convex point set. Our approach for defining $\mathcal{G}$ consists of connecting the points on each column of $\mathcal{S}$ to all the points on a number of adjacent columns which depends on the index of the column. More specifically, consider the sequence π_i which is inductively defined as follows: (i) $\pi_0 := 1$; (ii) $\pi_i := \pi_{i-1} \circ 2^i \circ \pi_{i-1}$. For example, $\pi_3 = 1, 2, 1, 4, 1, 2, 1, 8, 1, 2, 1, 4, 1, 2, 1$. Let i be sufficiently large so that π_i has at least n elements. For $j = 1, \dots, n$, assign the j-th element of π_i to the j-th column of $\mathcal{S}$. Then the points on the j-th column of $\mathcal{S}$ are connected to all the points on a number of adjacent columns which is equal to the element of π_i assigned to the column times some integer constant $c > 0$. Since the sum of the elements assigned to the columns of $\mathcal{S}$ is in $O(n \log n)$, the number of edges of the resulting geometric graph $\mathcal{G}$ is in $O(n^3 \log n)$. Whether $\mathcal{G}$ is actually a universal geometric graph for the n-vertex planar graphs however remains to be proved.

References

1. Noga Alon and Michael R. Capalbo. Optimal universal graphs with deterministic embedding. In *Proc. 19th ACM-SIAM Sympos. Discrete Algorithms (SODA)*, pages 373–378, 2008.
2. Noga Alon, Michael R. Capalbo, Yoshiharu Kohayakawa, Vojtech Rödl, Andrzej Rucinski, and Endre Szemerédi. Universality and tolerance. In *Proc. 41st IEEE Sympos. Foundations of Computer Science (FOCS)*, pages 14–21, 2000.
3. Patrizio Angelini, Till Bruckdorfer, Giuseppe Di Battista, Michael Kaufmann, Tamara Mchedlidze, Vincenzo Roselli, and Claudio Squarcella. Small universal point sets for k-outerplanar graphs. *Discret. Comput. Geom.*, 60(2):430–470, 2018.
4. László Babai, Fan R. K. Chung, Paul Erdős, Ronald L. Graham, and Joel H. Spencer. On graphs which contain all sparse graphs. *Annals Discrete Math.*, 12:21–26, 1982.
5. Michael J. Bannister, Zhanpeng Cheng, William E. Devanny, and David Eppstein. Superpatterns and universal point sets. *J. Graph Algorithms Appl.*, 18(2):177–209, 2014.
6. Prosenjit Bose. On embedding an outer-planar graph in a point set. *Comput. Geom.*, 23(3):303–312, 2002.
7. Julia Böttcher, Klaas P. Pruessmann, Anusch Taraz, and Andreas Würfl. Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs. *European Journal of Combinatorics*, 31(5):1217–1227, 2010.
8. Julia Böttcher, Mathias Schacht, and Anusch Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen*, 343(1):175–205, 2009.
9. Michael R. Capalbo. Small universal graphs for bounded-degree planar graphs. *Combinatorica*, 22(3):345–359, 2002.
10. Jean Cardinal, Michael Hoffmann, and Vincent Kusters. On universal point sets for planar graphs. *J. Graph Algorithms Appl.*, 19(1):529–547, 2015.
11. Fan R. K. Chung and Ronald L. Graham. On graphs which contain all small trees. *J. Combin. Theory Ser. B*, 24(1):14–23, 1978.
12. Fan R. K. Chung and Ronald L. Graham. On universal graphs for spanning trees. *J. London Math. Soc.*, 27(2):203–211, 1983.
13. Hubert de Fraysseix, János Pach, and Richard Pollack. How to draw a planar graph on a grid. *Combinatorica*, 10(1):41–51, 1990.
14. Louis Esperet, Gwenaël Joret, and Pat Morin. Sparse universal graphs for planarity. *CoRR*, abs/2010.05779, 2020.
15. I. Fáry. On straight line represention of planar graphs. *Acta Scientiarum Mathematicarum*, 11:229–233, 1948.

16 Stefan Felsner, Hendrik Schrezenmaier, Felix Schröder, and Raphael Steiner. Linear size universal point sets for classes of planar graphs. In Erin W. Chambers and Joachim Gudmundsson, editors, *39th International Symposium on Computational Geometry, SoCG 2023, June 12-15, 2023, Dallas, Texas, USA*, volume 258 of *LIPIcs*, pages 31:1–31:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
17 Fabrizio Frati, Michael Hoffmann, and Csaba D. Tóth. Universal geometric graphs. *Comb. Probab. Comput.*, 32(5):742–761, 2023.
18 Radoslav Fulek and Csaba D. Tóth. Universal point sets for planar three-trees. *J. Discrete Algorithms*, 30:101–112, 2015.
19 Peter Gritzmann, Bojan Mohar, János Pach, and Richard Pollack. Embedding a planar triangulation with vertices at specified points. *Amer. Math. Monthly*, 98(2):165–166, 1991.
20 Maciej Kurowski. A 1.235 lower bound on the number of points needed to draw all n-vertex planar graphs. *Inf. Process. Lett.*, 92(2):95–98, 2004.
21 Richard Rado. Universal graphs and universal functions. *Acta Arithmetica*, 9(4):331–340, 1964.
22 Manfred Scheucher, Hendrik Schrezenmaier, and Raphael Steiner. A note on universal point sets for planar graphs. *J. Graph Algorithms Appl.*, 24(3):247–267, 2020.
23 K. Wagner. Bemerkungen zum vierfarbenproblem. *Jahresbericht. German. Math.-Verein*, 2:26–32, 1936.

4.3 Recognizing k^+-real Face Graphs

Michael A. Bekos (University of Ioannina, GR, bekos@uoi.gr)
Giuseppe Di Battista (Università Roma Tre, IT, giuseppe.dibattista@uniroma3.it)
Emilio Di Giacomo (University of Perugia, IT, emilio.digiacomo@unipg.it)
Walter Didimo (University of Perugia, IT, walter.didimo@unipg.it)
Michael Kaufmann (Universität Tübingen, DE, michael.kaufmann@uni-tuebingen.de)
Fabrizio Montecchiani (University of Perugia, IT, fabrizio.montecchiani@unipg.it)

Abstract. A nonplanar drawing Γ of a graph G divides the plane into topologically connected regions, called *faces* (or *cells*). The boundary of each face is formed by vertices/crossings and edges. Given a positive integer k, we say that Γ is a *k^+-real face drawing* of G if the boundary of each face of Γ contains at least k vertices of G. The study of k^+-real face drawings started in a paper by Binucci et al. (WG 2023), where edge density bounds and results about the relationship with other beyond-planar graph classes are given. In this seminar we have investigated the complexity of recognizing k^+-real face graphs, i.e., graphs that admit a k^+-real face drawing. We have studied both the general unconstrained scenario and the 2-layer scenario in which the graph is bipartite, the vertices of the two partite sets are placed on two distinct horizontal layers, and the edges are drawn as straight segments (or equivalently as vertical monotone curves).

4.3.1 Introduction

The study of k^+-real face drawings of (nonplanar) graphs started in a recent paper by Binucci et al. [1]. In a k^+-real face drawing, the boundary of each face contains at least k vertices of the graph, where $k \geq 1$ is a given integer. In particular, for any positive integer k, a k^+-real face drawing forbids faces formed only by crossing points and edges. From the practical side, the interest in k^+-real face graphs is motivated by the intuition that faces mostly consisting of crossing points make the graph layout less readable. From the theoretical side, k^+-real face drawings can be regarded as a generalization of planar drawings whose face sizes are above a desired threshold [2, 3, 4].

Basic Notations and Terminology. Let G be a graph. We assume that G is simple, that is, it contains neither multiple (i.e., parallel) edges nor self-loops. We also assume, without loss of generality, that G is connected, as otherwise we can just consider each connected component of G independently. We denote by $V(G)$ and $E(G)$ the set of vertices and the set of edges of G, respectively. A *drawing* Γ of G is a geometric representation of G that maps each vertex $v \in V(G)$ to a distinct point of the plane and each edge $(u, v) \in E(G)$ to a simple Jordan arc between the points corresponding to u and v. We always assume that Γ is a *simple* drawing, that is: (i) adjacent edges do not intersect, except at their common endpoint; (ii) two independent (i.e., non-adjacent) edges intersect in at most one of their interior points, called a *crossing point*; and (iii) no three edges intersect at a common crossing point.

A *vertex* of Γ is either a point corresponding to a vertex of G, called a *real-vertex*, or a point corresponding to a crossing point, called a *crossing-vertex*. Since the drawing is simple, a crossing-vertex has always degree four. We denote by $V(\Gamma)$ the set of vertices of Γ. An *edge* of Γ is a curve connecting two vertices of Γ; an edge of Γ whose endpoints are both real-vertices coincides with an edge of G; otherwise it is just a proper portion of an edge of G. We denote by $E(\Gamma)$ the set of edges of Γ. Drawing Γ subdivides the plane into topologically connected regions, called *faces* (or *cells*). The boundary of each face consists of a circular sequence of vertices and edges of Γ. The set of faces of Γ is denoted by $F(\Gamma)$. Exactly one face in $F(\Gamma)$ corresponds to an infinite region of the plane, called the *external face* (or *outer face*) of Γ; the other faces are the *internal faces* of Γ. When the boundary of a face f of Γ contains a vertex v (or an edge e), we also say that f contains v (or e).

Given an integer $k \geq 1$, a *k^+-real face drawing* of a graph G is such that each face contains at least k real-vertices. If G admits such a drawing, then we call G a *k^+-real face graph*. If G is bipartite, then a *2-layer k^+-real face drawing* of G is a k^+-real face drawing Γ of G such that the vertices of the two parts of its vertex partition are drawn on two distinct horizontal lines, called *layers*, and each edge is a straight-line segment. If G admits such a drawing, then we call G a *2-layer k^+-real face graph*.

4.3.2 Contribution

During the seminar we investigated the complexity of recognizing k^+-real face graphs, that is, the complexity of testing whether, given a graph G and a positive integer k, there exists a k^+-real face drawing of G. We studied both the general (unconstrained) scenario and the 2-layer drawing scenario. A summary of the main contributions is given below.

- In the general case, we are able to show that recognizing k^+-real face graphs for values of $k \in \{1, 2\}$ is NP-complete. For the hardness proof we exploit a reduction from the well-known 3-Partition problem. Note that, for $k \geq 3$, *optimal* k^+-real face graphs (i.e., k^+-real face graphs with the maximum possible edge density) are always planar graphs with all faces of degree k (see [1]). Hence, recognizing optimal k^+-real face graphs when $k \geq 3$ is equivalent to testing whether the graph admits a planar embedding where all faces have size at least k, a problem studied in [5].
- We proved tights upper bounds on the maximum number of edges in a 2-layer k^+-real face graph, for every value of k. These types of results can help in the design of recognition

algorithms. Specifically, we established that 1^+-real face and 2^+-real face graphs with n vertices have at most $2n-4$ and $1.5n-2$ edges, respectively. Also, for $k \geq 3$, optimal 2-layer k^+-real face graphs are caterpillar graphs, and therefore have $n-1$ edges.

- We believe that it is possible to efficiently recognize 2-layer 2^+-real face graphs. In particular, during the seminar we designed a testing algorithm that seems to work in linear time in the size of the graph. We plan to give a formal description and a proof of correctness of this algorithm in a near future article.
- For 2-layer 1^+-real face graphs, we characterized the structure of optimal graphs (i.e., 2-layer 1^+-real face graphs with exactly $2n-4$ edges) and of biconncted graphs. These characterizations should lead to efficient recognition algorithms. Recognizing 2-layer 1^+-real face graphs that are not biconnected seems to be more difficult; we are still working on establishing whether a polynomial-time algorithm exists in this case, even if the graph is a tree.

References

1 Carla Binucci, Giuseppe Di Battista, Walter Didimo, Seok-Hee Hong, Michael Kaufmann, Giuseppe Liotta, Pat Morin, and Alessandra Tappini. Nonplanar Graph Drawings with k Vertices per Face. *Graph-Theoretic Concepts in Computer Science – 49th International Workshop, WG 2023*, volume 14093 of LNCS, pages 86–100, Springer, 2023.
2 Patrick Ali, Peter Dankelmann, and Simon Mukwembi. The radius of k-connected planar graphs with bounded faces *Discrete Mathematics*, 312(24): 3636–3642, 2012.
3 Brandon Du Preez. Plane graphs with large faces and small diamater. *Asutralasian Journal of Combinatorics*, 80(3): 401–418, 2021.
4 Yongxin Lan, Yongtang Shi, and Zi-Xia Song. Extremal h-free planar graphs. *Electronics Journal of Combinatorics*, 26(2):2, 2019.
5 Giordano Da Lozzo, Vít Jelínek, Jan Kratochvíl, and Ignaz Rutter. Planar embeddings with small and uniform faces. *Algorithms and Computation – 25th International Symposium, ISAAC 2014*, volume 8889 of LNCS, pages 633–645, Springer, 2014.

4.4 Local-crossing-critical graphs and covering complete geometric graphs

János Pach (Rényi Institute – Budapest, HU & EPFL – Lausanne, CH, pach@renyi.hu)
Géza Tóth (Rényi Institute, Budapest, HU, geza@renyi.hu)
Pavel Valtr (Charles University, Prague, CZ, valtr@kam.mff.cuni.cz)

4.4.1 Local-crossing-critical graphs

The crossing number of a graph G, $\mathrm{cr}(G)$ is the minimum number of edge crossings of G over all its drawings on the plane. G is called k-crossing-critical if $\mathrm{cr}(G) \geq k$, but for any edge e of G, $\mathrm{cr}(G \setminus e) < k$. Richter and Thomassen [6] proved that the crossing number of a k-crossing-critical graph cannot be arbitrarily large, if G is a k-crossing-critical graph, then $\mathrm{cr}(G) \leq 5k/2 + 16$. It was improved by Barát and Tóth [2], [4] to $\mathrm{cr}(G) \leq 2k + 6\sqrt{k} + 47$. It is conjectured that for any such graph G we have $\mathrm{cr}(G) \leq k + c\sqrt{k}$.

We worked on the following related problem. The local crossing number of a graph G, $\mathrm{lcr}(G)$ is the minimum number l with the property that G can be drawn in the plane with at most l crossings on each edge. In other words, $\mathrm{lcr}(G)$ is the minimum number l such that G is l-planar. A graph G is k-local-crossing-critical if $\mathrm{lcr}(G) \geq k$, but for any edge e of G, $\mathrm{lcr}(G \setminus e) < k$.

Is there a function $f(k)$ with the property that for any k-local-crossing-critical graph G, $\text{lcr}(G) \leq f(k)$?

1-local-crossing critical graphs are easy to describe, removing any edge we get a planar graph, but the graph itself is not planar. It follows from Kuratowski's theorem, that these graphs are the topological K_5 and $K_{3,3}$ graphs, therefore, $f(1) = 1$.

Observation. *If $f(2)$ exists, then $f(k)$ exists for all k, and $f(k) \leq (k-1)f(2)$.*

Proof. Suppose that we know that $f(2) = f$ exists. That is, if G is a graph with the property that for any edge e of G, the graph $G - e$ is 1-planar, then G is f-planar.

Let $k > 2$ and suppose that G is a k-local-crossing-critical graph. Replace each edge of G by a path of length $k-1$ (that is, ($k-1$ edges, $k-2$ subdividing vertices), let H be the resulting graph. Remove an edge e from H. It follows from the assumption on G that that $H - e$ an be drawn such that each path that replaces an edge of G contains at most $k-1$ crossings. But then the subdividing vertices can be arranged so that there is at most one crossing on each edge. Therefore, H is 2-local-crossing-critical. Consequently, H is f-planar. Consider an f-planar drawing of H. In the corresponding drawing of G, there are at most $f(k-1)$ crossings on each edge. This finishes the proof.

We are left with the case $k = 2$: Is there an $f > 0$ so that the following satement holds? Suppose that G is a graph with the property that for any edge e of G, the graph $G - e$ is 1-planar. Is there a number f then G is f-planar.

We tried to use the ideas of Richter and Thomassen and other related papers on crossing-critical graphs, but there were some unexpected and very exciting difficulties.

4.4.2 Covering complete geometric graphs with plane trees and forests

Definition. A geometric graph is a graph drawn in the plane with possibly crossing straight-line edges. A plane star-forest is a geometric graph in which each component is a star (a tree with exactly one non-leaf vertex) and no two edges the graph cross. A complete convex geometric graph is a geometric graph whose vertex set is a set of points in the plane in strictly convex position, where every pair of vertices are connected by an edge.

Answering a question of Dujmović and Wood [3] Pach, Saghafian, and Schnider [5] proved that the edge set of a complete *convex* geometric graph on n vertices cannot be covered by fewer than $n-1$ plane star-forests. This bound is tight. They made the following

Conjecture. No complete geometric graph can be covered with less than $3n/4$ plane star forests.

This was proved to be *false* [1]

Theorem. (Antić, Gliǐć, Milivojčević): *There are infinitely many even values of n, for which there exists a complete geometric graph with n vertices whose edges set can be covered by $n/2+1$ plane star-forests.*

We studied the analogous problem where instead of star-forests, we are allowed to use any plane trees. It appears to be true that there exists a constant $c > 0$ such that the edge set of every complete geometric graph can be covered by $(1-c)n$ plane trees. We verified this conjecture in some special cases.

References

1 Todor Antić, Jelena Glišić, and Milan Milivojčević. Star-forest decompositions of certain geometric graphs. In *40th European Workshop on Computational Geometry (EuroCG24)*, 2024.

2 János Barát and Géza Tóth. Improvement on the crossing number of crossing-critical graphs. In *International Symposium on Graph Drawing and Network Visualization*, pages 372–381. Springer, 2020.
3 Vida Dujmovic and David R Wood. Graph treewidth and geometric thickness parameters. *Discrete & Computational Geometry*, 37:641–670, 2007.
4 Barát János and Tóth Géza. Improvement on the crossing number of crossing-critical graphs. *Discrete & Computational Geometry*, 67(2):595–604, 2022.
5 János Pach, Morteza Saghafian, and Patrick Schnider. Decomposition of geometric graphs into star-forests. In *International Symposium on Graph Drawing and Network Visualization*, pages 339–346. Springer, 2023.
6 R Bruce Richter and Carsten Thomassen. Minimal graphs with crossing number at least k. *Journal of Combinatorial Theory, Series B*, 58(2):217–224, 1993.

4.5 Rerouting Curves on Surfaces

Stefan Felsner (Technische Universität Berlin, DE, felsner@math.tu-berlin.de)
Henry Förster (Universität Tübingen, DE, henry.foerster@uni-tuebingen.de)
Stephen Kobourov (University of Arizona, US, kobourov@cs.arizona.edu)
Anna Lubiw (University of Waterloo, CA, alubiw@uwaterloo.ca)
Yoshio Okamoto (The University of Electro-Communications, JP, okamotoy@uec.ac.jp)
János Pach (Rényi Institute – Budapest, HU & EPFL – Lausanne, CH, pach@renyi.hu)
Csaba D. Tóth (California State University Northridge, US, csaba.toth@csun.edu)
Géza Tóth (Rényi Institute of Mathematics, HU, geza@renyi.hu)
Torsten Ueckerdt (Karlsruhe Institute of Technology, DE, torsten.ueckerdt@kit.edu)
Pavel Valtr (Charles University Prague, CZ, valtr@@kam.mff.cuni.cz)

4.5.1 The Problem

We study the problem of reconfiguring graph embeddings on an orientable surface, where the vertices are fixed and each reconfiguration step redraws one edge curve. Consider a set of points $\mathcal{S}$ on an orientable surface Σ, and two embeddings $\mathcal{P}$ and $\mathcal{Q}$ of the same graph G on vertices $\mathcal{S}$. Here an *embedding* means that each edge is drawn as a curve, which we call an *edge curve*, on the surface and no two edge curves intersect except at a common endpoint. Note that the edge curves of $\mathcal{P}$ may cross the edge curves of $\mathcal{Q}$. We assume that the correspondence between edge curves of $\mathcal{P}$ and $\mathcal{Q}$ is given. A *reconfiguration step* or *move* replaces one edge curve γ of an embedded graph G by a new curve γ' to obtain a new embedding of G – in other words, γ' may not cross any of the other edge curves of the embedded graph, though we allow γ and γ' to intersect. The question we address is whether $\mathcal{P}$ can be reconfigured to $\mathcal{Q}$ via a sequence of moves.

The special case where the graph is a matching consisting of two disjoint edges was considered by Ito, Iwamasa, Kakimura, Kobayashi, Maezawa, Nozaki, Okamoto, and Ozeki [8]. In this restricted situation, they showed that reconfiguration is not always possible in the plane (see Fig. 4), but is always possible on a surface Σ of genus $g \geq 1$. (Note that their paper is primarily about reconfiguration in a more discrete setting where $\mathcal{P}$ and $\mathcal{Q}$ consist of disjoint paths in a fixed graph.)

Our main result is that if the graph G is a matching and the surface Σ is a torus, then reconfiguration is always possible. This immediately extends to any orientable surface of genus $g \geq 1$ and nonorientable surface of genus $g \geq 2$. The only open case remains the projective plane. We extend the result to the case where G is a tree. The result does not extend to general embeddings of a graph on the torus, as we show by an example. However, we conjecture that reconfiguration is possible if we restrict to plane embeddings, and we prove this for the special case of series-parallel graphs.

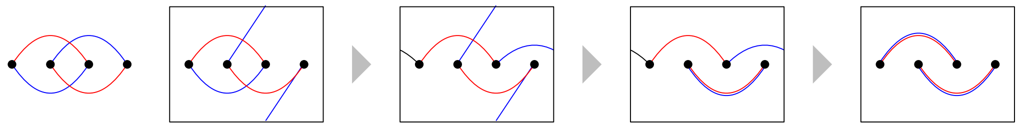


Figure 4 (a) Two embeddings of a matching of two edges (red and blue) that cannot be reconfigured on the plane. (b) Reconfiguration of the two embeddings on the torus using 4 steps.

4.5.2 Related Work

The problem of morphing graph drawings on a torus [1, 7] is different in that the vertices are allowed to move but the edges must remain straight segments on the flat torus. The problem of tightening or untangling curves on a surface [2, 3, 4, 6] is also different in that they consider drawings with possible crossings (i.e., immersions rather than embeddings), and deform the edge curves continuously via so-called homotopy moves (local moves that modify the topology of the immersion). We also point the interested reader to Colin de Verdière's survey [5] on graphs on surfaces.

4.5.3 Rerouting of Matchings on the Torus

We have a set $\mathcal{P}$ of n non-crossing blue paths on the torus that form a matching of $2n$ points, and we have a set $\mathcal{Q}$ of n non-crossing red paths that form the same matching of the points. Our algorithm consists of the following three steps:

1. Draw the torus as a flat torus with all the red paths inside (i.e., none of them cross the torus boundary).
2. Re-draw the blue paths so that none of them cross the torus boundary.
3. Use the top/bottom boundary of the flat torus which now forms a clean handle (i.e., a closed non-separating curve not crossed by any red or blue path) to solve the problem.

We remark that for the third step, it would be enough to re-draw the blue paths so that none of them cross the top/bottom boundary. Curiously, however, our proof for the second step will establish the stronger property of clearing the entire boundary.

4.5.3.1 Draw the torus as a flat torus with all the red paths inside

In this step, we begin with an arbitrary projection of the red paths on the flat torus. We now seek a closed non-separating curve σ that avoids all red paths. Note that σ necessarily exists as the red paths form a non-crossing matching. We use σ as the new horizontal boundary of the flat torus. The argument can be repeated to obtain a new vertical boundary of the flat torus; see Fig. 5.

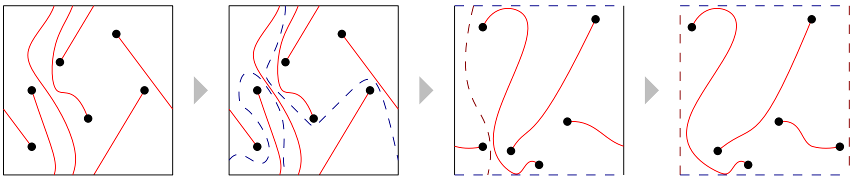


Figure 5 Illustration for Step 1.

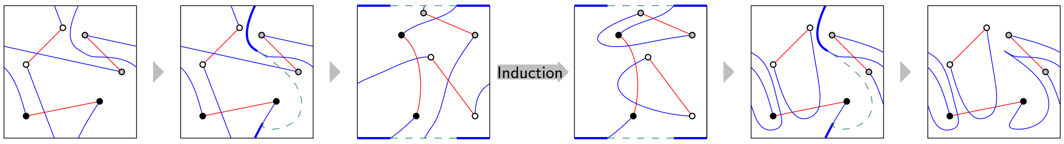


Figure 6 Illustration for Step 2.

4.5.3.2 Re-draw the blue paths so that none of them cross the torus boundary

In this step, we focus on the blue paths. The high-level idea (also see Fig. 6) is to pick one of the blue paths $p \in \mathcal{P}$ that crosses the flat torus boundary and reduce the number of times that it crosses the flat torus boundary. To this end, we find a *shortcut* γ such that

- γ lies in the torus boundary,
- the endpoints of γ lie in p, and there are no other intersections between γ and p,
- let p' be the piece of p that makes a cycle with γ,
- rerouting p' to γ reduces the number of times the path crosses the torus boundary,
- the cycle $p' \cup \gamma$ is a non-separating cycle on the torus.

The proof that γ exists requires a careful argument. Our goal is to reroute p' to γ but note that γ may cross other blue paths. Thus we must first clear all the crossings where other blue curves cross γ. This is done by induction on an appropriate (different) flat torus. After that we can reroute p along γ which reduces the number of times that p crosses the flat torus boundary. Observe that this comes at the expense of possibly increasing the number of times that other blue curves cross the torus boundary.

4.5.3.3 Use the clean handle to solve the problem

Once the clean handle is established, we can solve the problem similarly to the example shown in Fig. 4. To this end, we redraw one blue path $p \in \mathcal{P}$ at a time and resolve one of the crossings of its corresponding red path $q \in \mathcal{Q}$ which is closest to one of its endpoints in each step; see Fig. 7. We can use one boundary (say the vertical) to reroute p from its first crossing. Then, we reroute the crossing path c such that it avoids the crossing using the other boundary (say the horizontal). Now we can redraw p so to follow the trajectory of q until q's second crossing. Finally, we redraw c so that it does not cross the boundary of the flat torus, avoiding p.

4.5.3.4 Extension to forests

Finally, we remark that our result can be generalized to the case where $\mathcal{P}$ and $\mathcal{Q}$ are toroidal embeddings of a forest. While our strategy remains the same, this requires a slightly more careful analysis.

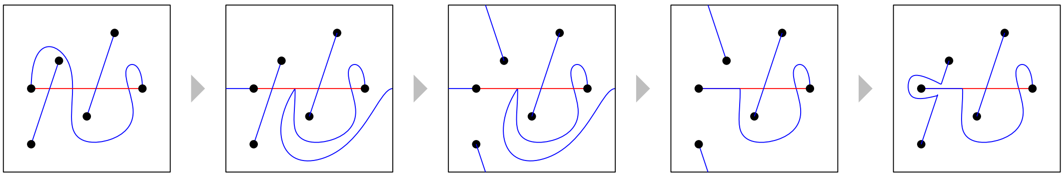


Figure 7 Illustration for Step 3.

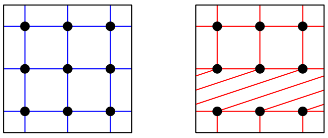


Figure 8 Two toric embeddings $\mathcal{P}$ (blue) and $\mathcal{Q}$ (red) that cannot be reconfigured into each other using a sequence of moves.

4.5.4 Non-Reroutable Toric Graph Embeddings

Following our previous positive result, one may wonder if it is always possible to reconfigure a given toric embedding $\mathcal{P}$ with a sequence of moves into another given toric embedding $\mathcal{Q}$. Unfortunately, this is not always possible as the example in Fig. 8 demonstrates.

For this example, it can be easily verified that a single curve of $\mathcal{P}$ can only be replaced by a topologically equivalent curve, i.e., it is impossible to change the embedding by replacing a single edge per move. Moreover, observe that both embeddings correspond to a quadrangulation of the torus where the embedding $\mathcal{Q}$ differs from $\mathcal{P}$ by a twist of the torus. This observation implies that we can generalize this result easily to a surface Σ of higher genus, i.e., one can use a suitably rigid tessellation of Σ for $\mathcal{P}$ and then perform a twist along a non-separating curve to obtain another embedding $\mathcal{Q}$ into which $\mathcal{P}$ cannot be reconfigured.

4.5.5 Rerouting of Plane Graphs on the Torus

While we showed that not all toric graphs can be reconfigured on the torus one may still wonder what happens if we restrict the embeddings $\mathcal{P}$ and $\mathcal{Q}$ of the graph to be plane. We remark that one may ask the same question for a surface Σ of higher genus g by requiring embeddings $\mathcal{P}$ and $\mathcal{Q}$ to be embeddable on a surface of genus $g-1$.

In particular, we can show that a plane embedding $\mathcal{P}$ can be reconfigured into another plane embedding $\mathcal{Q}$ on the torus if the input graph G is *series-parallel*. To this end, recall that the family of series-parallel graphs can be defined recursively as follows:

1. The graph consisting of a single edge st is a series-parallel graph with poles s and t.
2. Given two series-parallel graphs G_1 with poles s_1 and t_1 and G_2 with poles s_2 and t_2, the series composition obtained by identifying t_1 and s_2 is a series-parallel graph with poles s_1 and t_2.
3. Given two series-parallel graphs G_1 with poles s_1 and t_1 and G_2 with poles s_2 and t_2, the parallel composition obtained by identifying s_1 and s_2 as well as t_1 and t_2 is a series-parallel graph with poles $s_1 = s_2$ and $t_1 = t_2$.

Notably, all plane embeddings of series-parallel graphs differ only in the order in which parallel subgraphs are sorted at their common poles. We schematically show in Fig. 9 how two consecutive parallel components can be resorted at their common poles. Transforming $\mathcal{P}$ into $\mathcal{Q}$ then reduces to a sequence of such reorderings.

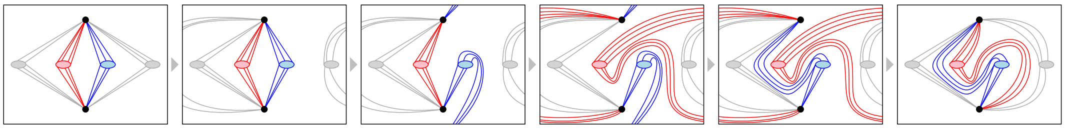


Figure 9 Reordering of two parallel components (red and blue) in a plane embedding on the torus.

4.5.6 Next Steps

As a follow-up to the above results found at the Dagstuhl Seminar, we intend to work on the following aspects:

1. Most importantly, we want to formalize our approaches further and provide reasonable bounds on their run times.
2. Our result on plane embeddings on series-parallel may be generalizable to plane embeddings of general planar graphs on the torus. The missing link is the analysis of triconnected planar graph whose embedding we have to be able to mirror.
3. Finally, we want to consider additional types of surfaces. With respect to our results on matchings, we want to attempt to achieve a similar result on the projective plane. Moreover, our results on plane embeddings motivates to study embeddings embeddable on surfaces of genus $g-1$ on a surface of genus g.

References

1. Erin Wolf Chambers, Jeff Erickson, Patrick Lin, and Salman Parsa. How to morph graphs on the torus. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2759–2778, 2021.
2. Hsien-Chih Chang and Arnaud de Mesmay. Tightening curves on surfaces monotonically with applications. *ACM Trans. Algorithms*, 18(4):36:1–36:32, 2022.
3. Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Discret. Comput. Geom.*, 58(4):889–920, 2017.
4. Hsien-Chih Chang, Jeff Erickson, David Letscher, Arnaud de Mesmay, Saul Schleimer, Eric Sedgwick, Dylan Thurston, and Stephan Tillmann. Tightening curves on surfaces via local moves. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 121–135, 2018.
5. Éric Colin de Verdière. Computational topology of graphs on surfaces. In Csaba D. Tóth, Joseph O'Rourke, and Jacob E. Goodman, editors, *Handbook of Discrete and Computational Geometry*, chapter 23, pages 605–636. Chapman and Hall/CRC, 2017.
6. Éric Colin de Verdière, Vincent Despré, and Loïc Dubois. Untangling graphs on surfaces. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4909–4941, 2024.
7. Jeff Erickson and Patrick Lin. Planar and toroidal morphs made easier. *Journal of Graph Algorithms and Applications*, 27:95–118, 2023.
8. Takehiro Ito, Yuni Iwamasa, Naonori Kakimura, Yusuke Kobayashi, Shun-ichi Maezawa, Yuta Nozaki, Yoshio Okamoto, and Kenta Ozeki. Rerouting planar curves and disjoint paths. In *50th International Colloquium on Automata, Languages, and Programming, ICALP 2023, July 10-14, 2023, Paderborn, Germany*, volume 261 of *LIPIcs*, pages 81:1–81:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.

4.6 Upward Drawings Beyond Planarity

Patrizio Angelini (John Cabot University, Rome, IT, pangelini@johncabot.edu)
Therese Biedl (University of Waterloo, Canada, biedl@uwaterloo.ca)
Markus Chimani (Universität Osnabrück, DE, markus.chimani@uni-osnabrueck.de)
Sabine Cornelsen (University of Konstanz, DE, sabine.cornelsen@uni-konstanz.de)
Giordano Da Lozzo (Roma Tre University, IT, giordano.dalozzo@uniroma3.it)
Seok-Hee Hong (University of Sydney, AU, seokhee.hong@sydney.edu.au)
Giuseppe Liotta (University of Perugia, IT, giuseppe.liotta@unipg.it)
Maurizio Patrignani (Roma Tre University, IT, maurizio.patrignani@uniroma3.it)
Sergey Pupyrev (Meta, Menlo Park, USA, spupyrev@gmail.com)
Ignaz Rutter (University of Passau, DE, rutter@fim.uni-passau.de)
Alexander Wolff (Universität Würzburg, DE, alexander.wolff@uni-wuerzburg.de)

4.6.1 Summary of Results

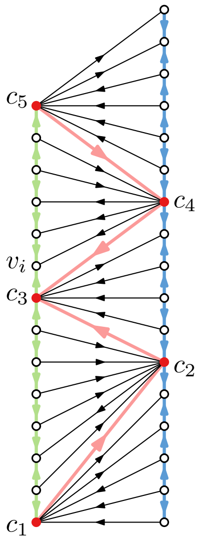
It is known that not every directed acyclic graph whose underlying undirected graph is planar admits an upward planar drawing. We are interested in pushing the notion of upward drawings beyond planarity. We investigate the "price of upwardness" for drawing planar directed acyclic graphs upwards – in terms of the maximum number of crossings per edge. More formally, we say that the drawing of a directed graph is *upward k-planar* if each edge is a y-monotone curve that is crossed at most k times by other edges. Our aim is to give good bounds on this parameter k for classes of planar directed acyclic graphs. For example, it is easy to see that every tree, no matter how its edges are directed, admits a planar upward drawing. On the other hand, Papakostas [2] showed that there is a directed acyclic 8-vertex outerpath that does not admit a planar upward drawing. (An outerpath is an outerplanar graph whose weak dual is a path.)

We have studied the problem both from a combinatorial and an algorithmic perspective. While this is still work in progress, we briefly summarize our results below. Let a *fan* be an outerpath in which there is a vertex, the *apex* of the fan, that is adjacent to all other vertices. We first show that every directed fan has an upward 2-planar drawing with specific properties (see Theorem 1). We then use this to show that every outerpath has an upward 2-planar drawing (see Theorem 2).

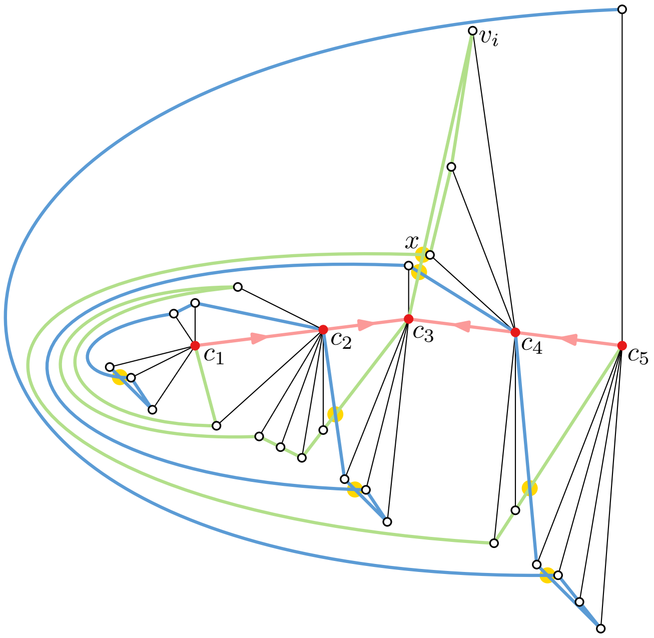
The edges incident to the apex are the *inner edges* of the fan. The other edges are the *outer edges* of the fan. Observe that the outer edges of a fan induce a path.

▶ **Lemma 1.** *Let c be the apex of a directed acyclic fan G, and let $P = \langle v_1, v_2, \ldots, v_{n-1} \rangle$ be the path of the remaining vertices in G. Let $P_1, P_2, \ldots, P_k$ be an ordered partition of P into maximal subpaths such that, for every $i \in \{1, 2, \ldots, k\}$, the edges between P_i and c are either all directed towards c or are all directed away from c. Then there is an upward 2-planar drawing of G with the following properties.*

1. *No inner edge is crossed.*
2. *Vertex v_1 has x-coordinate 1, the apex c and the vertex v_{n-1} have x-coordinate $n-1$, and the x-coordinates of $v_2, v_3, \ldots, v_{n-2}$ are distinct values in the set $\{2, 3, \ldots, n-2\}$.*
3. *For all edges all x-coordinates of the curves are at most $n-1$. All inner edges and all edges of the subpaths $P_1, \ldots, P_k$ are in the vertical strip between 1 and $n-1$.*
4. *The edge between P_1 and P_2 is crossed at most once if P_1 is a directed path.*



(a) a directed outerpath G'



(b) upward drawing of G' with at most two crossings per edge

Figure 10 Example input and output of our algorithm for drawing outerpaths upward (edge crossings are highlighted in yellow).

We use Theorem 1 to prove the following.

▶ **Theorem 2.** *Every directed acyclic outerpath admits a upward 2-planar drawing.*

Proof. We assume that the given outerpath is maximal. If the outerpath has interior faces that are not triangles, we triangulate them using additional edges, which we direct such that they do not induce directed cycles. After drawing the resulting maximal outerpath, we remove the additional edges.

Let G' be such a graph; see Figure 10a. Let $c_1, c_2, \ldots, c_k$ be the vertices of degree at least 4 in G' (marked red in Figure 10). These vertices form a path (light red in Figure 10); let them be numbered along this path, which we call the *backbone* of G'. We draw the backbone in an x-monotone fashion, with very small slopes, going up and down as needed; see Figure 10b. For $i \in \{1, 2, \ldots, k-1\}$, we set $x(c_{i+1})$ to $x(c_i)$ plus the number of inner edges incident to c_{i+1}. For $i \in \{1, 2, \ldots, k\}$, we place the vertices incident to backbone vertex c_i using the algorithm for drawing a fan as detailed in the proof of Theorem 1. The vertices above (below) c_i are placed above (below) the backbone. If $i < k$, then the last vertex in the fan of c_i is connected to c_{i+1} and c_i is connected to the first vertex v_i in the fan of c_{i+1}. These two edges may cross each other. If the edge $c_i v_i$ goes, say, up but the following outer edges go down until a vertex v_k below c_{i+1} is reached, then the edge e_i between c_i and v_i may be crossed a second time by the edge e between v_{k-1} and v_k – as the crossing labeled x on the edge $c_3 v_3$ in Figure 10b – but, due to our invariant for drawing fans, e had been crossed only once within its fan. Also, the edge e_i cannot have a third crossing. Thus, in total no edge is crossed three times. ◀

Theorem 2 naturally raises the question about whether we can extend the proof to any graph having pathwidth 2. This is not the case, as we can prove the following.

▶ **Lemma 3.** *For every $k \geq 1$, there exists a directed acyclic graph with pathwidth 2 and $O(k)$ vertices that does not admit an upward k-planar drawing.*

Another research direction motivated by Theorem 2 is whether the result about outerpaths can be extended to any outerplanar graph. Also this question has a negative answer. Namely, we can prove the following.

▶ **Lemma 4.** *For every $k \geq 1$, there exists an outerplanar directed acyclic graph that does not admit an upward k-planar drawing.*

We have also studied the complexity of testing upward k-planarity of directed acyclic graphs. An st-graph is a directed acyclic graph with only one source and only one sink. Every *planar* st-graph with the source and the sink on the same face is upward planar, that is, it admits an upward drawing where no edge is crossed [1]. Leaving the domain of planar st-graphs, we can prove the following.

▶ **Theorem 5.** *Testing upward 1-planarity is NP-complete even for st-graphs both with and without a fixed rotation system.*

On the positive side, we are working on proving the following recognition result concerning outer upward 1-planar graphs, that is, graphs that admit an upward 1-planar drawing where all vertices lie on the outer face.

▶ **Theorem 6.** *Outer upward 1-planarity can be tested in polynomial time for single-source graphs.*

4.6.2 Open Problems

The research activity in Dagstuhl has also identified a list of related problems that can be the subject of future studies. Among them are the following questions.

1. Is there a directed outerpath that does not admit an upward 1-planar drawing?
2. Consider the class $\mathcal{O}_\Delta$ of outerplanar graphs (or even 2-trees) of maximum degree Δ. Is there a function f such that every graph in $\mathcal{O}_\Delta$ admits an upward $f(\Delta)$-planar drawing?
3. For which families of biconnected directed acyclic graphs is testing upward 1-planarity tractable?

References

1 Giuseppe Di Battista and Roberto Tamassia. Algorithms for plane representations of acyclic digraphs. *Theoret. Comput. Sci.*, 61:175–198, 1988. doi:10.1016/0304-3975(88)90123-5.
2 Achilleas Papakostas. Upward planarity testing of outerplanar DAGs. In Roberto Tamassia and Ioanni G. Tollis, editors, *Proc. Int. Sympos. Graph Drawing (GD)*, volume 894 of *LNCS*, pages 298–306. Springer, 1994. doi:10.1007/3-540-58950-3_385.

Participants

- Patrizio Angelini
John Cabot University – Rome, IT
- Michael A. Bekos
University of Ioannina, GR
- Therese Biedl
University of Waterloo, CA
- Markus Chimani
Universität Osnabrück, DE
- Sabine Cornelsen
Universität Konstanz, DE
- Giordano Da Lozzo
University of Rome III, IT
- Giuseppe Di Battista
University of Rome III, IT
- Emilio Di Giacomo
University of Perugia, IT
- Walter Didimo
University of Perugia, IT
- Vida Dujmovic
University of Ottawa, CA
- Stefan Felsner
TU Berlin, DE
- Henry Förster
Universität Tübingen, DE
- Fabrizio Frati
University of Rome III, IT
- Michael Hoffmann
ETH Zürich, CH
- Seok-Hee Hong
The University of Sydney, AU
- Michael Kaufmann
Universität Tübingen, DE
- Stephen G. Kobourov
University of Arizona – Tucson, US
- Jan Kratochvil
Charles University – Prague, CZ
- Giuseppe Liotta
University of Perugia, IT
- Anna Lubiw
University of Waterloo, CA
- Fabrizio Montecchiani
University of Perugia, IT
- Pat Morin
Carleton University – Ottawa, CA
- Yoshio Okamoto
The University of Electro-Communications – Tokyo, JP
- János Pach
Alfréd Rényi Institute – Budapest, HU & EPFL – Lausanne, CH
- Maurizio Patrignani
University of Rome III, IT
- Sergey Pupyrev
Facebook – Menlo Park, US
- Ignaz Rutter
Universität Passau, DE
- Csaba Tóth
California State University – Northridge, US
- Géza Tóth
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Torsten Ueckerdt
KIT – Karlsruher Institut für Technologie, DE
- Pavel Valtr
Charles University – Prague, CZ
- Alexander Wolff
Universität Würzburg, DE

Report from Dagstuhl Seminar 24071

Safety Assurance for Autonomous Mobility

Jyotirmoy Deshmukh[*1], Bettina Könighofer[*2], Dejan Ničković[*3], and Filip Cano[†4]

1 **USC – Los Angeles, US.** jyotirmoy.deshmukh@usc.edu
2 **TU Graz, AT.** bettina.koenighofer@iaik.tugraz.at
3 **AIT – Austrian Institute of Technology – Wien, AT.** dejan.nickovic@ait.ac.at
4 **TU Graz, AT.** filip.cano@iaik.tugraz.at

Abstract

This report documents the program and the outcomes of the Dagstuhl Seminar "Safety Assurance for Autonomous Mobility" (24071). The seminar brought together an interdisciplinary group of researchers and practitioners from the fields of formal methods, cyber-physical systems, and artificial intelligence, with a common interest in autonomous mobility. Through a series of talks, working groups, and open problem discussions, participants explored the challenges and opportunities associated with ensuring the safety of autonomous systems in various domains, including industrial automation, automotive, railways, and aerospace. Key topics addressed included the need for industrial-grade autonomous products to operate reliably in safety-critical environments, highlighting the lack of standardized procedures for obtaining safety certifications for AI-based systems. Recent advancements in the verification and validation (V&V) of autonomous mobility systems were presented, focusing on requirements verification, testing, certification, and correct-by-design approaches. Overall, the seminar provided a comprehensive overview of the current state and future directions in safe autonomous mobility, emphasizing the need for interdisciplinary collaboration and innovation to address the complex challenges in this rapidly evolving field.

Seminar February 11–16, 2024 – https://www.dagstuhl.de/24071

2012 ACM Subject Classification Applied computing → Transportation; Computer systems organization → Embedded and cyber-physical systems; Hardware → Robustness; Software and its engineering → Software verification and validation

Keywords and phrases aerospace, automotive, autonomy, formal methods, railway

Digital Object Identifier 10.4230/DagRep.14.2.95

1 Executive Summary

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT)
Jyotirmoy Deshmukh (USC – Los Angeles, US)
Bettina Könighofer (TU Graz, AT)
Filip Cano (TU Graz, AT)

As autonomous mobility systems gain traction worldwide, ensuring their safety, robustness, and dependability has become a paramount concern for their implementation at scale. The Dagstuhl Seminar on "Safety Assurance for Autonomous Mobility" gathered experts from academia, and industry to address the critical challenges and opportunities posed by the rapid

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

growth of autonomous technologies across various mobility domains, including automotive, aerospace, robotics, and railways. This seminar provided a much-needed platform for researchers and practitioners to exchange insights, collaborate on emerging ideas, and build a shared understanding of safety assurance in this rapidly evolving field.

Seminar Context and Structure

The seminar convened a diverse and multidisciplinary group of participants, each bringing specialized expertise in formal methods, software verification, embedded systems, and transportation safety. We believe that autonomous mobility is a field where progress cannot be just limited to academic research. The ideas and methods developed in theoretical settings build more robust applications, and the challenges faced in industrial settings guide theoretical research towards productive solutions. To reflect this drive, the group of participants was chosen to strike a balance between academia and industry, with many participants having experience in both domains. The seminar was structured around discussions in small working groups, different each day. Each group had a topic or problem to tackle, and the key challenges and state of the art solutions were shared at the end of each day to all participants of the seminar. One of the key objectives of this seminar was to bring together ideas and researchers from academic and industrial backgrounds. To this end, a sessions on Wednesday were focused on the current state of the practice being used in industrial applications, with each industrial partner sharing knowledge about their respective application fields.

Key Themes and Discussions

- Participants explored various formal methods and verification techniques designed to enhance the reliability of autonomous systems. Discussions highlighted the need to advance state-of-the-art formal verification approaches to accommodate the complexity of modern autonomous systems.
- The seminar also emphasized the importance of building resilient systems capable of functioning reliably in dynamic environments. Discussions tackled challenges related to the various modules (e.g. perception, motion planning, etc.) in autonomous mobile cyber-physical systems, considering how to incorporate robustness into the design phase and beyond.
- The specific challenges unique to each transportation sector were discussed, emphasizing tailored strategies for addressing safety assurance in automotive, aerospace, and railway systems. The cross-sectoral dialogue shed light on shared challenges and provided new perspectives that will inform future efforts.

Outcomes and Future Directions

The seminar generated a consensus on the urgent need for more research and collaboration across sectors. Participants emphasized the importance of combining expertise from different domains to address the interdisciplinary nature of safety assurance in autonomous mobility.

Moreover, the discussions underscored the potential for ongoing interdisciplinary seminars and follow-up workshops that would ensure continuous engagement among stakeholders. These future events would also provide venues for updating each other on progress, refining safety standards, and accelerating technological advancements in this field.

Overall, the seminar succeeded in creating a collaborative environment that not only identified existing challenges but also laid the groundwork for innovative solutions in safety assurance for autonomous mobility.

2 Table of Contents

3 Overview of Talks

3.1 Industrial-grade Autonomous Products – Challenge and Applied Approaches for Safety

Christof Budnik (Siemens – Princeton, US)

Many businesses recognize the vast potential of artificial intelligence (AI) to enhance efficiency, productivity, and quality in autonomous industrial production. While numerous prototypes have emerged, only a limited number of products exhibit the capability to operate reliably in safety-critical environments. Notably, the absence of a secure person locator in manufacturing poses a significant challenge, hindering worker collaboration with autonomously operating machines in open environments. The elevated risk of AI failures causing harm to workers underscores the urgent need for solutions to mitigate such risks and avoid adverse consequences for companies. In this context, gaining a competitive edge in verifying and measuring AI safety becomes imperative for quality-focused enterprises. This imperative is further underscored by the establishment of standards and industrial regulations. The presented discussion outlined prevailing challenges and barriers associated with implementing safe AI for autonomous robots in the manufacturing floor. Specifically, it provided insights into how autonomous systems are further shaping the future in mobility and smart cities, presenting challenges that resonate within industry practices. One prominent challenge discussed pertains to the absence of standardized procedures for obtaining safety certifications for AI-based systems. Addressing this issue, the presentation introduced three distinct approaches: the development of a comprehensive test infrastructure supporting the entire DevOps lifecycle, leveraging AI to generate test cases, and ensuring end-to-end assurance of autonomous machines from an architectural perspective. By exploring these strategies, businesses can actively navigate the complexities of AI safety certification and enhance their ability to deploy reliable and secure autonomous systems in industrial settings.

3.2 Industrial Challenges in Assuring Autonomy

Mauricio Castillo-Effen (Lockheed Systems – Arlington, US)

In this talk, I introduced three ideas:
- The necessity and feasibility of agility and adaptability,
- The importance of context, and
- The challenges in modeling system evolution.

I demonstrated that current Systems Engineering practice has shortcomings because it assumes systems must be developed with all foreseeable conditions in mind. This is impractical and significantly burdens design assurance, limiting timely deployment opportunities and hindering learning from operations. This issue can be mitigated by focusing on targeted conditions, reusing assurance artifacts, and enabling continuous improvement, specifically by gradually expanding the operational domain under the protection of runtime assurance and safe learning mechanisms.

In the second part of the talk, Operational Design Domains (ODDs) were presented as a framework for modeling context. Seven research challenges were proposed to advance our understanding of ODD similarity, coverage, and scenarios as ODD samples.

The final section revealed a conceptual diagram illustrating the evolution of systems through the interplay of four categories of artifacts: specifications, design, implementation, and assurance. Systems evolve over time and through a spectrum of variants. It is necessary to develop formalisms and algorithmic reasoning to address system evolution effectively, focusing on assurance to create safe configuration baselines and variants offering guarantees.

3.3 Challenges in Autonomous Vehicle Development

Patricia Derler (Zoox Inc. – Foster City, US)

Automotive autonomous vehicles (AVs) development faces many challenges stemming from (a) the need to express the desired behavior, (b) the need to (formally) capture this desired behavior, and (c) the need to verify whether the AV exhibited the desired behavior. This talk discusses some of the challenges surrounding specification of requirements from traffic rules as written in the rules of the road (e.g. how should one formalize the rule in the Austrian road traffic act stating that one "is not allowed to drive so fast that he dirty other road users or things on the road" – § 20 in [1], the assumptions that one needs to make about other road users (should one always assume that the pedestrian on the side walk could jump in front of the ego vehicle as it is passing?), the timing and latency requirements (e.g. what is the latency requirement from first sensing a previously occluded road participant to braking?), and the lack of good, useful, formal models at various levels of abstractions that lend themselves to formal verification.

References

1 Austria, Straßenverkehrsordnung 1960, BGBl. Nr. 159/1960. (Austrian Road Traffic Act of 1960)

3.4 Future Mobility: Challenges and Recent Developments in V&V

Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US)

Joint work of Bardh Hoxha, Danil Prokhorov, Dejan Ničković, Georgios Fainekos, Hideki Okamoto, Hoang-Dung Tran, Jacob Anderson, Jyotirmoy Deshmukh, Navid Hashemi, Sungwoo Choi, Tomoya Yamaguchi, Xiaodong Yang

The talk delves into the emerging challenges and recent advancements in the Verification and Validation (V&V) of autonomous mobility systems. Focused on enhancing safety assurance, the discussion encompasses a broad range of cyber-physical systems, including smart cities, medical devices, and autonomous driving systems (ADS). The research emphasizes the critical role of requirements verification, validation, testing, certification, and correct-by-design approaches. The talk provides a number of methods for safety verification of machine learning-enabled CPS, safe planning and control of heterogeneous multi-agent systems, and human-robot interaction. The team aims to provide formal guarantees on system functionality and performance under uncertainty. The presentation showcases tools and case studies, highlighting the importance of rigorous V&V processes in realizing the future of autonomous mobility safely and efficiently.

References

1 Jacob Anderson, Georgios Fainekos, Bardh Hoxha, Hideki Okamoto, Danil Prokhorov, *Pattern Matching for Perception Streams*, Runtime Verification (RV), 2023
2 Tomoya Yamaguchi, Bardh Hoxha, Dejan Ničković, *RTAMT – Runtime Robustness Monitors with Application to CPS and Robotics*, International Journal on Software Tools for Technology Transfer (STTT), 2023
3 Hoang-Dung Tran, Sungwoo Choi, Hideki Okamoto, Bardh Hoxha, Georgios Fainekos, Danil Prokhorov, *Quantitative Verification for Neural Networks using ProbStars*, Hybrid Systems: Computation and Control (HSCC), 2023
4 Hoang-Dung Tran, Sungwoo Choi, Xiaodong Yang, Tomoya Yamaguchi, Bardh Hoxha, Danil Prokhorov, *Verification of Recurrent Neural Networks with Star Reachability*, Hybrid Systems: Computation and Control (HSCC), 2023
5 Navid Hashemi, Bardh Hoxha, Tomoya Yamaguchi, Danil Prokhorov, Georgios Fainekos, Jyotirmoy Deshmukh, *A Neurosymbolic Approach to the Verification of Temporal Logic Properties of Learning enabled Control Systems*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS), 2023

3.5 Clearsy: Safety Critical Systems And Forthcoming Autonomy In the Railways

Thierry Lecomte (CLEARSY – Aix-en-Provence, FR)

Joint work of Jan Peleska, Anne E. Haxthausen, Thierry Lecomte
Main reference Jan Peleska, Anne E. Haxthausen, Thierry Lecomte: "Standardisation Considerations for Autonomous Train Control", in Proc. of the Leveraging Applications of Formal Methods, Verification and Validation. Practice – 11th International Symposium, ISoLA 2022, Rhodes, Greece, October 22-30, 2022, Proceedings, Part IV, Lecture Notes in Computer Science, Vol. 13704, pp. 286–307, Springer, 2022.
URL https://doi.org/10.1007/978-3-031-19762-8_22

This presentation demonstrates how safety critical systems are designed, developed, and certified in the railways. Safety is about failing systems. Failing parts to consider are exposed: wrong specification, wrong program, wrong binary, wrong execution, bad hardware (hardware caonatains other functions/interface that described in the datasheet), failing hardware (entropy leading to dysfunctional gates, drifting clock, etc.), wrong environment specification, and wrong exploitation procedure. Failure, either systematic or random, have an impact on the behaviour of the system. Safety is about keeping the probability of catastrophic failure below a treshold. Safety demonstration has to convince an independent expert that the feared events are not going to happen more frequently than expected. CLEARSY is using formal methods and related tools at different levels (software, data, system) to complete this safety demonstration, in accordance with the safety standards. The introduction of ML-based technologies to enable autonomous driving is raising safety issues. AI is at the moment not recommended to develop the most critical function in the railways, the main argument geing the lack of explanability of the ML black box function. UIC (http://uic.org) has launched a project titled "New methods for safety demonstration" that is aimed at helping human certifiers to envisage the certification of functions developed with AI or using IoT/Cybersecurity. Several projects of autonomous trains have been initiated in France, some are related to existing lines (freight, passengers, high-speed) while others are targeting low traffic, regional lines with specific/adapted infrastructure. These projects are expected to contribute to the standards, based on the results obtained.

3.6 Introduction to Conformal Prediction

Lars Lindemann (USC – Los Angeles, US)

Learning-enabled systems promise to enable many future technologies such as autonomous driving, intelligent transportation, and robotics. Accelerated by the computational advances in machine learning and AI, there has been tremendous success in the design of learning-enabled systems. At the same time, however, new fundamental challenges arise regarding the safety and reliability of these increasingly complex systems that operate in unknown dynamic environments. In this tutorial, I will provide new insights and discuss exciting opportunities to address these challenges by using conformal prediction (CP), a statistical tool for uncertainty quantification. I will advocate for the use of CP in systems theory due to its simplicity, generality, and efficiency as opposed to existing model-based techniques that are either conservative or have scalability issues.

References

1 Shafer, Glenn, and Vladimir Vovk. *A tutorial on conformal prediction*, Journal of Machine Learning Research (JMLR) 2008.
2 Angelopoulos, Anastasios N., and Stephen Bates. *Conformal prediction: A gentle introduction*, Foundations and Trends in Machine Learning 2023.
3 Romano, Yaniv, Evan Patterson, and Emmanuel Candes. *Conformalized quantile regression*, Conference on Neural Information Processing Systems (NeurIPS) 2019.
4 Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel Candes, Aaditya Ramdas. *Conformal prediction under covariate shift*, Conference on Neural Information Processing Systems (NeurIPS) 2019.
5 Lars Lindemann, Matthew Cleaveland, Gihyun Shim, George J. Pappas. *Safe planning in dynamic environments using conformal prediction*, IEEE Robotics and Automation Letters (RA-L) 2023.
6 Matthew Cleaveland, Insup Lee, George J. Pappas, Lars Lindemann. *Conformal Prediction for Time Series using Linear Complementarity Programming*, AAAI Conference on Artificial Intelligence (AAAI) 2024.
7 Lars Lindemann, Xin Qin, Jyotirmoy V. Deshmukh, George J. Pappas. *Conformal prediction for STL runtime verification*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) 2023.
8 Yiqi Zhao, Bardh Hoxha, Georgios Fainekos, Jyotirmoy V. Deshmukh, Lars Lindemann. *Robust Conformal Prediction for STL Runtime Verification under Distribution Shift*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) 2024.
9 Yiqi Zhao, Xinyi Yu, Jyotirmoy V. Deshmukh, Lars Lindemann. *Conformal Predictive Programming for Chance Constrained Optimization* arXiv 2024.

3.7 Denso: Critical Scenario Identification

Selma Music (Denso Automotive – Eching, DE)

Proving the safe operation of autonomous driving is one of the biggest challenges in the automotive domain. Conventional test methods are not sufficient due to the large amount of test kilometers that are needed to argue about system safety guarantees. Therefore, scenario-based testing is a promising and widely studied method to address this challenge in simulation. In this talk, I will discuss critical scenario identification (CSI), one of the ways to perform scenario-based testing. We will refer to requirements from the safety standard ISO 21448 (SOTIF) and focus on discovering scenario "unknown unknowns" using CSI methods. We will mainly focus on intelligent testing methods within CSI, using optimization-based testing, and we will discuss open challenges, e.g., the selection of appropriate input parameters and the appropriate design of the scenario. The talk will contain illustrative examples and results to demonstrate the effectiveness of the CSI approach for the safety assurance of autonomous driving.

3.8 AVL framework for close loop testing

Darko Stern (AVL – Graz, AT)

Benefits to automated vehicles are greater road safety, cost savings, more productivity and reduced use of fuels, thus contributing to a green future. End users are ready to invest in autonomous vehicles, however, 64% demand increased safety for autonomous vehicles. Customers expect that autonomous systems cause no dangerous situations and do not influence the traffic flow, however, most systems currently on the market cannot handle complex situations. So, how to make sure that the automated vehicle behaves correctly in EVERY situation? In my talk, I presented a framework for close-loop testing and verification of automated vehicles that are under development in the AVL List. I presented ways how to avoid expensive full factorial testing, by using the Active DoE approach, followed by a comprehensive overview of the co-simulation environment for its execution. The reliability of the test depends on the fidelity of the 3D environment, which needs to support close-loop testing and the possibility of creating rare events. I finished my talk with the progress AVL made in modelling sensors and pedestrian behaviour.

3.9 Safety Assurance of Automated Driving Systems – Selected Concepts, Standards, and Approaches

Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

The talk covered several topics in the area of safety assurance for automated driving systems. Based on a model to reason about the interaction of required, specified and implemented behaviors as well as the definition of the Operational Design Domain, the purposes and scopes of two most prevalent safety standards for road vehicles have been discussed. Furthermore, the Pegasus-VVM framework for scenario-based safety assurance as well as a specific method for open context domain analysis have been introduced.

4 Working groups

4.1 What Do We Need to Provide Safety Assurance?

Houssam Abbas (Oregon State University – Corvallis, US), Ezio Bartocci (TU Wien, AT), Chih-Hong Cheng (Universität Hildesheim, DE), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Stefan Pranger (TU Graz, AT), and Alessandro Zanardi (ETH Zürich, CH)

This working group discussed what artifacts are need to build a safety assurance case for an autonomous vehicle or system of autonomous vehicles. We started by considering what things we think are currently working or should be part of an eventual reliable solution. Answers included:

1) Natural language specifications (unstructured natural language vs constrained natural language)
2) Tools that ensure "horizontal" traceability from natural language requirement to formal requirements to test cases to results (with intermediate steps in the middle)
3) Syntactic generators of scenarios: while still in infancy in this domain, they are showing usefulness within industry. Declarative/program-like specifications

We did an overview of types of specs (functional, performance, security, architectural design, ODD etc.) and some attendees suggested that rather than adhere strictly to the use of unambiguous specifications, perhaps we should allow multiple interpretations with some margins of flexibility to account for genuine uncertainty (borne out of the designers' uncertainty, as opposed to wrong interpretations.) Another things that works really well is the existence of a near-universal system description language (like Verilog for hardware) which enables industry-wide tool development and agreement on pre-competitive technologies.

In answer to "Where do you see the biggest gap?" the attendees offered:

1) No ways to efficiently capture the different interpretations of certain requirements, or which ones are relevant.

2) How to account efficiently for new factors in the assurance case? (E.g. we discover a new factor that affects a given outcome). Measures of coverage?
3) How to capture expert (implicit) knowledge and intuition?
4) How to create vertical traceability, from application-specific assurance outcomes (e.g. no running over pedestrians) to assurance requirements one level below, and one level below that, etc.

4.2 Methods for Safety Assurance

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Christof Budnik (Siemens – Princeton, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Marie Farrell (University of Manchester, GB), Rong Gu (Mälardalen University – Västerås, SE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Lars Lindemann (USC – Los Angeles, US), Selma Music (Denso Automotive – Eching, DE), Necmiye Ozay (University of Michigan – Ann Arbor, US), Giulia Pedrielli (Arizona State University – Tempe, US), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Darko Stern (AVL – Graz, AT)

The discussion was centered around the prerequisites needed in to provide safety assurance to autonomous mobile systems in terms of requirements, specifications, assumptions, descriptions of the Operational Design Domain (ODD), architecture models, digital twins and simulators. This rather broad list of topics was narrowed down during the discussion and the participants identified the most critical aspects with respect to the safety assurance.

The first observation was that the requirements are still predominantly given in the form of natural language with possibly some guidelines regarding the structure of the text. In the past, there have been works on constrained natural language templates for specifications, but they had limited success of practical adoption. On the other hand, formal specifications are needed to reduce ambiguities, facilitate exchange of requirements between teems, and automate certain design and verification activities. The process of formalizing requirements is not straightforward as certain natural language statements can sometimes have (even on purpose) multiple interpretations, and it is hard to capture different interpretations with formal specification. Formal specifications typically require providing one precise interpretation of the natural language sentence. Another problem is that there are no efficient ways to capture implicit knowledge and intuitions with formal specifications, which are often part of any design. Formal specifications can take many different forms, from declarative (e.g. temporal logics) to operational (e.g. program-like specifications). Specifications can also address many different aspects of the design such as its functional properties, performance, security, architecture design, ODDs, etc. Finally, the participants identified the need for tools that ensure horizontal traceability throughout the design cycle (from natural language requirements to formal specifications to test cases with everything in between). More specifically, there is the question of how to create vertical traceability, from application-specific assurance outcomes (e.g. no running over pedestrians) to assurance requirements one level below, down to the requirements of individual components. In order to facilitate the safety assurance reasoning, the participants identified a framework for reusability of artefacts as part of a potential solution. Such a framework would permit to keep track of specifications, historical evidence on prior arguments, scenarios and domain models, raw datasets and assurance patterns.

4.3 How To Design for Assurance

Bettina Könighofer (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Martin Fränzle (Universität Oldenburg, DE), Mahsa Ghasemi (Purdue University – West Lafayette, US), Radu Grosu (TU Wien, AT), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Sayan Mitra (University of Illinois – Urbana Champaign, US), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

A pivotal area of discussion in our group was how to present evidence for safety assurance. We discussed the importance of constructing and showcasing worst-case scenarios from each contract. However, we quickly realized that these scenarios might contradict each other. This led us to the consensus that analyzing combined worst-case behaviors, instead of individual ones, would be more effective, especially to tackle dependence problems and ensure the system's safety through invariance properties.

When it came to integrating specifications, contracts, and tests, we saw much potential in mining contracts from tests and generating tests to cover scenarios not addressed by the contracts. This approach ensures a comprehensive coverage of potential failures and strengthens the safety argument.

The group recognized the challenges in verifying large codebases and model-based design. We acknowledged that tools are not scalable enough yet but concluded that focusing verification on critical properties and having good layered architectural designs could be a workaround. Similarly, for model-based testing, the difficulty lies in obtaining models, so we suggested advancing techniques like abstract interpretation, which automatically generates models from code.

We also tackled the issue of information overflow and how to pinpoint the relevant properties that need our attention. Integrating different kinds of evidence into a consistent assurance argument emerged as a significant challenge. Here, we saw a need for heterogeneous contracts that can seamlessly combine probabilistic and Boolean elements, despite the current lack of extensive research in this area.

Specific attention went to the verification of perception systems. We discussed managing various uncertainties by defining precise contracts that set tolerated error ranges and ensure consistency across data sequences. This method aims to mitigate risks associated with perception errors.

The synthesis of systems brought up concerns over explainability. We all agreed that guided synthesis, imposing constraints on architectural design, could make the synthesis process quicker and the systems more understandable.

Lastly, we discussed existing solutions that we think have high potential. Layered and error-aware architectures are great because they allow for error tracing and include mechanisms for recovery. They are built to make safe decisions based on the current knowledge and have plans in place for potential deviations. Contracts that mediate trust between different layers of an architecture are also part of the solution.

In sum, our discussions underscored a multifaceted approach to designing safety into autonomous mobility products. We believe that a combination of worst-case analysis, contract-based design, targeted verification, and sophisticated handling of uncertainties will pave the way for safer autonomous systems.

4.4 Application Domain: Automotive – Challenges

Patricia Derler (Zoox Inc. – Foster City, US), Ezio Bartocci (TU Wien, AT), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Darko Stern (AVL – Graz, AT), Alessandro Zanardi (ETH Zürich, CH), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

This working group session discussed important and hard problems in the area of automotive autonomous mobility. The problems were roughly divided into the areas of

(a) requirements and specification engineering,
(b) the development of automotive autonomous systems, and
(c) verification and validation (V&V) activities.

For (a), participants highlighted the difficulty in formalizing requirements, especially given the diverse range of requirement types such as collision avoidance, comfort, adherence to traffic and social rules, among others. Moreover, concerns arose regarding the verification of requirements' completeness across different levels, spanning from the vehicle to component levels. Complex Operational Design Domains (ODDs) further compounded these challenges, prompting discussions on how to formalize, refine, and extend them effectively while balancing usefulness and safety. Defining safety requirements proved intricate, with passive safety measures like stopping deemed insufficient for freeways. Additionally, challenges in formulating requirements that help gain trust in autonomous systems were discussed. Such requirements necessitate interpreting social behaviors and developing robust models for the environment, AV, as well as components and sub-components, reflecting the multifaceted nature of autonomous automotive systems.

As for problems surrounding the development (b), the complexities of handling and certifying machine learning (ML) components were front and center. Debates arose regarding the autonomy levels and the suitability of supervised versus unsupervised approaches. There were concerns regarding the ambiguous nature of certification processes and the need to address performance issues, including the adequacy of data for training and the challenges of generalizing it effectively. Architecture flaws, particularly the lack of redundancy, emerged as a focal point, emphasizing the necessity for robust system designs. Operational Design Domains posed significant hurdles, raising questions about developing and detecting specific ODDs and the system's behavior beyond these domains. The workshop also shed light on the myriad tooling and engineering issues inherent in designing and deploying autonomous systems, underscoring the multifaceted nature of the development process.

Key problems identified in (c) are the balance between simulation and testing, questioning what could effectively be simulated versus what necessitated physical testing for robust validation. Concerns arose regarding the fidelity of simulation environments and determining the appropriate level of simulation, whether at the holistic AV and environment level or at the component level. Closed-loop testing presented difficulties, along with debates surrounding the utility of abstractions in models and simulations, particularly in mapping them to real-world scenarios. Coverage of Operational Design Domains in simulation and testing remained a significant challenge, as did the processing and extraction of information from data to uncover special cases effectively.

4.5 Application Domain: Automotive – Solutions

Patricia Derler (Zoox Inc. – Foster City, US), Ezio Bartocci (TU Wien, AT), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Darko Stern (AVL – Graz, AT), Alessandro Zanardi (ETH Zürich, CH), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

Following the working group discussions on problems for autonomous automotive systems, this session focused on solution approaches and existing solutions. International safety standards were recognized as pivotal, although some standards posed additional challenges or lacked actionable directives, underscoring the importance of concerted efforts in this area. Advances in simulation, exemplified by initiatives like VISTA 2.0 [1], offered promising avenues for enhancing capabilities in multimodal sensing and policy learning for autonomous vehicles. Bridging natural language and formal methods emerged as a key strategy, with solutions integrating language and logic for formalizing Operational Design Domains (ODDs) and requirements to provide clearer and more actionable specifications. Moreover, leveraging probabilistic modeling and verification techniques showed promise in enhancing the robustness of autonomous systems, particularly in interpreting social behaviors and demonstrating intent. Digital twin frameworks were identified as valuable tools for gaining insights into real-world scenarios and validating autonomous systems. Open source reference implementations were also highlighted for fostering collaboration, transparency, and innovation within the autonomous automotive community. Specific solutions discussed included improving AI/ML explainability through redundancy, independence assumptions, and model-based approaches like neuro-symbolic ML/AI. The importance of explainable AI was emphasized, stressing the need for comprehensive understanding and enabling replayable errors for effective analysis and improvement. Safety frameworks, such as the Safety Case by Waymo [2], were recognized for providing structured methodologies to ensure the safety and reliability of autonomous systems. Formalization efforts, such as Responsibility-Sensitive Safety (RSS) [5], the Pegasus Project [3], and IEEE 2846 [4], were deemed essential for establishing clear guidelines and standards for system behavior and domain coverage. Communication protocols like Collaborative Awareness Messages [6] and Collective Perception Messages [7] were seen as critical for facilitating effective communication between autonomous vehicles and infrastructure, thereby enhancing overall system efficiency and safety.

References

1 Alexander Amini, Tsun-Hsuan Wang, Igor Gilitschenski, Wilko Schwarting, Zhijian Liu, Song Han, Sertac Karaman, Daniela Rus. *Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles*, 2022 International Conference on Robotics and Automation (ICRA), 2022.

2 The Waymo Team, *A Blueprint for AV Safety: Waymo's Toolkit For Building a Credible Safety Case*, 2023.

3 Hermann Winner, Karsten Lemmer, Thomas Form, Jens Mazzega. *PEGASUS—First steps for the safe introduction of automated driving.* Road Vehicle Automation 5. Springer International Publishing, 2019.

4 IEEE, *IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems.* 2022.

5 Shalev-Shwartz, Shai, Shaked Shammah, and Amnon Shashua. *On a formal model of safe and scalable self-driving cars.* arXiv preprint arXiv:1708.06374 (2017).

6 F. Raviglione, S. Zocca, A. Minetto, M. Malinverno, C. Casetti, C. F. Chiasserini, and F. Dovis, *From collaborative awareness to collaborative information enhancement in vehicular networks,* Vehicular Communications, vol. 36, p. 100497, 2022.

7 Mohammad Raashid Ansari, Jean-Philippe Monteuuis, Jonathan Petit, Cong Chen. *V2x misbehavior and collective perception service: Considerations for standardization*, IEEE Conference on Standards for Communications and Networking (CSCN) 2021.

4.6 Application Domain: Robotics – Challenges

Stefan Pranger (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), and Bettina Könighofer (TU Graz, AT)

This working group discussed important problems in the field of autonomous robotics. The topics discussed can roughly be divided into problems that arise during design time and problems faced after deployment of an autonomous actor.

The discussion in this working group started off with problems that occur during design time. The first hard problems that have been discussed concerned the design and definition of concrete requirements and specifications, as well as assumptions that can be taken at design time. Clear definitions of safety and performance are often highly dependent on the specific application area. The group discussed several examples from industry, an example to highlight the intricacy of the problem are autonomous delivery robots in a japanese hospital. The robots are used to pick up and deliver e.g. blood samples and are therefore allowed to use elevators. Since the hospital is in an area prone to earthquakes it is of special interest that robots do not hinder the fast evacuation of patients and staff in case of an earthquake. The group put a special emphasis on the arising complexity w.r.t. the different application areas of autonomous robots. In contrast to the problems faced in autonomous mobility, an area in which the legislative body has already compiled a manifest that autonomous systems need to adhere to. The group discussed a second issue that arises during the design time with respect to simulation software, especially the robot operating system (ROS). The issue raised regarding ROS is its dependency on the scheduling of the underlying operating system.

The simulation might not be faithful as it cannot ensure real-time execution due to this dependency. This issue has been discussed in the seminar group, we refer the reader to the abstract regarding discussed solutions.

The discussion continued with topics related to the deployment of autonomous robots. The discussed problems heavily relate to the problem mentioned above regarding the different application areas, especially under the presence of humans. The problem mentioned here is twofold: It is an open problem, how a robot should behave to be the least harmful towards humans in safety critical situations, and it is a formidable problem to incorporate arbitrary human behaviour, such that autonomous robots can behave appropriately. Additionally, the problem of reliable perception has been labeled as very interesting for industry partners.

Lastly, the group wanted to include security as a hard problem. The topic has initially been mentioned, but has not been discussed.

4.7 Application Domain: Robotics – Solutions

Stefan Pranger (TU Graz, AT), Filip Cano (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Bettina Könighofer (TU Graz, AT), and Selma Music (Denso Automotive – Eching, DE)

The group followed up on the discussion about hard problems for autonomous robotics by discussing potential solution methods. The complexity of generalizing specifications regarding safety and utility in the domain of autonomous robotics has been identified as a key problem. This stems from the wide range of application areas. Therefore, the need for formal definitions of Operational Design Domains (ODDs) per applications has been identified as key driver during the discussion. The group highlighted the different aspects that are to be taken into account, specifically with regards to human interaction. In order to be able to allow safe behavior there is the need for dynamic modelling of human behaviour. The group mentioned data-driven approaches as a solution. Apart from a model of potential human behaviour, specifications of the types of intended interaction with an autonomous robot needs to be part of an ODD.

The group continued by discussing methods that enhance the explainability through counterfactual analysis.

Regarding open problems related to the faithfulness of simulations, the group discussed how parts of the ROS can be formally verified, and the capabilities of ROS to allow real-time simulations.

4.8 Application Domain: Railway – Challenges

Chih-Hong Cheng (Universität Hildesheim, DE), Christof Budnik (Siemens – Princeton, US), Martin Fränzle (Universität Oldenburg, DE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

The development of machine learning has created great promises, not only for autonomous driving but also for other domains. Within the railway domain, projects such as safe.trAIn aim to uncover the potential of ML-based perception systems for increasing the Grade of Automation (GoA) via enabling driverless regional trains.

When aligning railway applications with urban autonomous driving, one can find some challenges sitting on the common ground. The first is the positive contribution of ML in safety arguments. The second is the introduction of security attacks. The third is the offering of guarantees to cover sufficiently the environmental conditions. The fourth is the ability to explain why ML fails by generalizing a single prediction error.

Nevertheless, there are also some railway-specific challenges. The first challenge comes with the problem where trains are operated on high-speed and have a long stopping distance. Thus, precisely characterizing acceptable behavior by considering aspects such as safety margin is a concern. The long stopping distance also implies that the decision making requires (1) a different sensor suite that allows look ahead way further than autonomous driving, and (2) considering high uncertainties associated with prediction due to detected objects being very far. The final challenge is related to the use of synthetic data. In contrast to autonomous driving where obtaining data is relatively straightforward, for railway applications such as defect inspection or rail track covered by muds, they occur very rarely, implying a must to use synthetic data. Arguing whether the degree of simulation fidelity is "enough" remains a challenging issue.

4.9 Application Domain: Railway – Solutions

Christof Budnik (Siemens – Princeton, US), Chih-Hong Cheng (Universität Hildesheim, DE), Martin Fränzle (Universität Oldenburg, DE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

In the intricate world of autonomous train safety, a blend of traditional and innovative solutions have been discussed which can ensure a secure railway environment.

Traditional safety measures like guard-rails, isolation protocols, and radio signals act as the backbone, guiding trains safely on their tracks and preventing collisions. These time-tested methods establish a reliable foundation for safe railway operations. Safety is further fortified through sensor diverse redundancy, where an array of safety-focused sensors act as vigilant guardians. Redundancy ensures that even if one sensor encounters an issue, others are ready to step in, contributing to fail-safe detection capabilities. This principle of sensor redundancy has been extensively studied and discussed in sensor fusion for autonomous car engineering utilizing different sensor technologies and their detection ranges. The introduction of telescope cameras for instance adds a futuristic dimension to safety measures, allowing us to “see the future” of train travel. This innovation poses interesting challenges, particularly in the context of Railway Vehicles (RV), and has been a subject of exploration in the field of railway safety research. In alignment with the philosophy of designing for verifiability verification processes and approaches can be streamlined such as when tracks are only designed as straight-line tracks. This design approach simplifies safety checks and addresses railway infrastructure design. Furthermore, if full specifiability can be reached then there is no need for complex Deep Neural Networks (DNNs).

Looking towards promising and innovative solutions online runtime monitoring has been identified as a beacon for ensuring safety in real-time operations. The emphasis on making these monitors as safe as possible is a foundational principle, underscored in safety studies on autonomous systems. This real-time vigilance serves as a safeguard against potential risks, ensuring the continuous safety of autonomous trains throughout their journeys. However, the advent of the autonomous train era brings forth new challenges, notably the domain’s distribution shift. Placing monitors in the safety loop requires thoughtful consideration, as an excessive presence might overwhelm the system. This delicate balance between effective monitoring and system efficiency is a topic to be further explored on autonomous system architectures. Striking the right equilibrium becomes pivotal in managing the distribution shift effectively without compromising safety.

Another aspect discussed is the adapting infrastructure as a forward-thinking strategy, heralding a promising future for autonomous trains. Rather than solely replacing the driver, reshaping the environment in which trains operate is a paradigm shift discussed on autonomous transportation infrastructure. This approach ensures a holistic enhancement of safety measures while leveraging existing infrastructure investments. The characteristic of the rarity of failures in autonomous train systems sparked the idea of injecting artificial failures to address the synthetic data gap. This proactive approach to failure simulation aims to bridge the gap in rare failure occurrences, ensuring comprehensive safety testing. Rigorous envelope protection, with an emphasis on minimum braking capabilities, establishes another robust standard for safety, ensuring autonomous trains possess the necessary safeguards to handle unforeseen circumstances. Finally, the integration of neuro-symbolic approaches, incorporating temporal stability, consistency across modalities, and physics-assisted techniques, is seen as cutting-edge methodology to filter potential safety issues in autonomous train operations. Neuro-symbolic approaches leverage a combination of symbolic reasoning and neural networks, offering a sophisticated methodology to enhance safety in autonomous train operations.

4.10 Application Domain: Aerospace – Challenges & Solutions

Lars Lindemann (USC – Los Angeles, US), Houssam Abbas (Oregon State University – Corvallis, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Marie Farrell (University of Manchester, GB), Taylor T. Johnson (Vanderbilt University – Nashville, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Necmiye Ozay (University of Michigan – Ann Arbor, US), and Giulia Pedrielli (Arizona State University – Tempe, US)

For aerospace system design, the discussion initially centered around the critical balance between safety and utility, the quantification of risk, and the pursuit of more modular and principled design methodologies. The discussion further centered around the multifaceted challenge of reconciling the inherent trade-offs between ensuring safety and achieving optimal performance and utility. Central to these discussions is the notion of risk, traditionally quantified as the product of severity and probability. This classical framework, however, prompts a reevaluation in the context of both offline design and runtime adaption of aerospace systems. The discourse underscored the necessity for nuanced conceptions of risk (such as the conditional value of risk) that are responsive to the dynamic operational environments and the uncertainties that pervade these phases. The discussion also focused on the complexities of flight certification, with risk-friendly and risk-averse approaches, and challenges due to the sim2real gap. Specifically, models in aerospace may be highly nontrivial and not correct due to lack of data. Another point of discussion was on academic strategic repositioning towards emphasizing the economic benefits of safety verification ("Safety does not sell") in accelerating system development and deployment. Marie brought up another point of discussion, that of the regulatory landscapes of NASA, ESA, and private entities like SpaceX, the latter being in absence of regulatory frameworks. In the end, the conversation touched upon educational and research methodologies, discussing the value of integrating courses on assurance cases and system safety into the aerospace (or more broadly any engineering) curriculum. This is to equip future engineers with the skills necessary to navigate the complex interplay of safety, risk, and performance in aerospace design and operation and to prepare them best for real world challenges. Lastly, the discussion centered on the development of case studies tailored to the aerospace and space sectors that link academia and industry more closely (further bridging the gap between theoretical safety tools and their application in real-world scenarios).

4.11 Techniques for Safety Assurance: : Low-level Control

Taylor T. Johnson (Vanderbilt University – Nashville, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Patricia Derler (Zoox Inc. – Foster City, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Lars Lindemann (USC – Los Angeles, US), Selma Music (Denso Automotive – Eching, DE), and Necmiye Ozay (University of Michigan – Ann Arbor, US)

In the working group dedicated to low-level control, we discussed the integration of various autonomous mobility components, such as perception, planning, and low-level control, and the types of guarantees that can be provided for safe execution. A key focus was on the architectures used in system design, including the distinction between functional and physical architectures and how structure is defined through interfaces. We explored the use of languages and standards such as AADL, SysML, and Autosar for specifying system architecture from various perspectives, including functionality, safety, behavior, timing, and enabling different types of analyses. Contract-based verification was highlighted as a critical method for ensuring system reliability. Challenges included integrating systems composed of many components, particularly when dealing with legacy systems and proprietary models, especially interfaces to low-level control through application programmer interfaces (APIs) that often may necessitate crossing cross-layer boundaries between high-level and low-level control, similar to cross-layer optimizations in network protocols. Overall, the discussion covered the need for a more flexible and adaptive approach to system design, one that can accommodate changing requirements and negotiations between stakeholders. Various techniques for software development and verification were discussed, including structure, composition, certification, schedulability analysis, modeling, and verification.

The discussion also touched upon the aerospace industry's adoption of architectures defined by the customer, similar to Autosar, underlining the importance of system architecture in the design and verification process. The challenge of low-level control not being fully accessible to the control designer was discussed, emphasizing the gap created by hidden system models and the necessity for system identification, and the use of lookup tables (LUTs) to manage nonlinearities, as well as approximations of large LUTs with other methods, such as via neural networks. More broadly, there was discussion on verification of these types of systems, ranging from the classical designs in low-level control like LUTs, to the usage of artificial intelligence (AI) and machine learning surrogates for these and broader tasks, where robustness of these components are critical.

We delved into the complexities of managing systems with time-varying parameters, such as battery degradation and the effects of aging, and discussed the potential for systems to self-identify and monitor changes. The conversation also covered the concept of conducting verification not only a priori but online, through the generation of conditional evidence and dynamic assurances, so that overall system-level and component-level specifications are monitored for assurance during operation.

The ability to refine high-level architectures through methods such as optimization modulo theories was considered crucial for overcoming system integration challenges, particularly the issue of abstraction leading to loss of detail and mismatches at the low level. We debated the importance of establishing low-level metrics that aid in reasoning about system-level integration and verification, including considerations for common mode failures, logical sensor fusion, and the balance between safety and availability.

Risk modulation and data-driven approaches were discussed for measuring risk either fleet-wise or at the individual vehicle level, with considerations on scoring individual artifacts, Safety Integrity Levels (SIL), and blame analysis. Questions were raised about the systematic approach to decision-making, such as selecting driving routes based on safety assessments.

Finally, the session touched upon the role of falsification in continuous integration and continuous delivery (CI/CD) processes, underscoring the evolving nature of safety and verification in the context of autonomous mobility. The discussion illuminated the multifaceted challenges and innovative strategies involved in ensuring the safe execution of autonomous robots, cars, planes, and trains, with a particular emphasis on the crucial role of low-level control in the broader context of autonomous system safety, all of which will require collaboration in teams of varied expertise to address.

4.12 Techniques for Safety Assurance: : Perception

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Christof Budnik (Siemens – Princeton, US), Martin Fränzle (Universität Oldenburg, DE), Rong Gu (Mälardalen University – Västerås, SE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Stefan Pranger (TU Graz, AT), Andoni Rodríguez (IMDEA Software Institute – Madrid, ES), Darko Stern (AVL – Graz, AT), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

The main topic of the discussion was the perception module and its role in the safety of smart mobility applications. The perception module is a complex system that typically relies on multiple sensors (e.g. camera, lidar and radar in the automotive domain), which are fused together, sometimes even using additional information extracted from other sources such as maps. The type and configuration of sensors used for the perception is domain-specific. For instance, the railways domain relies more on the infrastructure sensors rather than the train sensors, due to the large breaking time for trains. Modern perception systems must also have predictive capabilities (e.g. predict the next action of the detected pedestrian). Finally, perception has to deal with multiple sources of uncertainty. The participants identified three sources of uncertainty: state uncertainty, classification of uncertainty and existential uncertainty. Perception contracts were proposed as a potential solution towards dealing with uncertainty and building safe perception with error bounds on the state estimation.

The participants noted that the problem of testing perception is quite different from operational perception. Testing perception is notoriously hard and opens many challenging problems – from the synthetic generation of realistic inputs to sensors to the integration of perception in closed-loop testing with the ability to catch rare scenarios. There is also the problem of storage when testing perception. For instance, recording one hour of raw sensor data in a car requires several TBs of storage. There is hence a need to pre-process and filter this data on the edge, before sending it to the cloud. When looking at the perception module in isolation, it is not clear what is the right set of properties to verify. The participants noted that testing perceptions shall be done on a system level, since the important part is the functional impact of a misperception on the overall system. Perception contracts were proposed as a potential solution for building safe perception with error bounds on the state estimation.

4.13 Techniques for Safety Assurance: : Machine Learning in Planning

Alessandro Zanardi (ETH Zürich, CH), Ezio Bartocci (TU Wien, AT), Filip Cano (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Marie Farrell (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Bettina Könighofer (TU Graz, AT), Assaf Marron (Weizmann Institute – Rehovot, IL), and Doron A. Peled (Bar-Ilan University – Ramat Gan, IL)

The discussion concentrated primarily on the crucial role of Machine Learning (ML) in effective planning, emphasizing its indispensable nature in trajectory planning. ML's ability to consider the conduct of a multitude of actors in scenarios such as switching lanes in traffic jams is indeed invaluable. One of the core discussion points was the significance of re-evaluating our predictions about the environment. This point underscores the necessity for dynamically adapting to changes that were previously unforeseen.

In terms of alternatives and security, the group outlined the importance of having a Plan B to fall back on when assumptions are not met while planning. They also presented the idea of having logical safeguards in place, alongside a Simplex architecture, to ensure that the planning process is robust and able to handle unexpected changes or challenges. A substantial portion of the discussion was devoted to explainable planning, highlighting the need for ML and symbolic planning to work in parallel. The symbolic planner, they detailed, would be capable of elucidating decisions when inquired, an essential feature for understanding and improving the plan's operation.

The group also shed light on how planning often involves juggling various objectives that might be conflicting. Addressing this challenge, the concept of minimal-violation planning was introduced, which emphasized having different property levels, treating safety and performance at distinct levels.

The discussion then directed toward handling risks and uncertainties, recognizing the crucial role of conformal prediction in this context. Motion Planning, as deterministic as it may appear, also contains a slew of uncertainties, often arising due to perception nuance. Therefore, developing strategies to evaluate and handle these risks plays a part in successful planning. The involvement of symbolic and sub-symbolic elements in end-to-end modular learning, which encompasses perception, planning, and control, was clarified. The group also elaborated on the utilization of Belief-Desire-Intention (BDI) agents to increase explainability and transparency.

Practical aspects like the costs to monitor, the timing, and the demands towards optimization engines were explored, with a particular focus on time budgets. The meeting acknowledged that there are numerous different solutions, and the challenge lies in how to combine them effectively. Emphasis was laid upon verification aspects and how to propagate results between modules—an essential part of improving and refining the planning process.

Lastly, the discussion proposed possible solutions, advocating for the use of all the ML resources, including black-box models, yet enclosing them within a safety net. This safety net could comprise Hoare logic-based controllers, safety filters, as well as rulebooks.

4.14 Techniques for Safety Assurance: Large Language Models

Marie Farrell (University of Manchester, GB), Christof Budnik (Siemens – Princeton, US), Filip Cano (TU Graz, AT), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Rayna Dimitrova (CISPA – Saarbrücken, DE), Martin Fränzle (Universität Oldenburg, DE), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Taylor T. Johnson (Vanderbilt University – Nashville, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Necmiye Ozay (University of Michigan – Ann Arbor, US), Andoni Rodríguez (IMDEA Software Institute – Madrid, ES), and Darko Stern (AVL – Graz, AT)

The discussion centered on the use of LLMs in the development of safety-critical autonomous mobile vehicles. During this discussion, we used ChatGPT to guide our discussion. We asked questions including: What are the hot topics regarding the use of LLMs in the development of safe autonomous systems (for mobility), what are specific applications of LLMs in the development of safety-critical autonomous mobility systems, how can LLMS be used in the development process of safety-critical autonomous mobility systems, what makes the use of LLMs in this domain unreliable and untrustworthy, what role can LLMs have in the development process of planning for autonomous mobility systems, and what are the dangers of using LLMs in this domain?

It is clear that LLMs bring both benefit and potentially create undesirable outcomes when used in the development of safety-critical autonomous mobility systems. Some of the positive ways that LLMs can be used are as follows. LLMs can improve the natural-language understanding of autonomous vehicles, potentially improving the safety of the interactions that the system has with the user. For example, responding accurately to verbal commands. They can also be used to provide clear instructions to passengers in an emergency. LLMs can be used to recognise and interpret traffic signs, signals and road markings. They can potentially even be used in adaptive navigation by assisting in dynamic route planning. One particular strength of LLMs in this domain is to provide fast incident reporting and documentation which can be used in insurance claims, compliance, and post-incident analysis. During the development process itself, LLMs can be used to generate requirement specifications and associated documentation. Other uses include the provision of a natural language interface for design and configuration. This would facilitate intuitive communication with system designers and engineers during the design phase. LLMs can be used to summarise regulatory compliance and standards documentation to help the developers to demonstrate compliance. Along this vein, LLMs can be used to carry out risk/hazard analyses by processing relevant pre-existing documentation. They can also provide useful ways of generating user manuals and training materials as well as improve communication with stakeholders. Each of these positive perspectives can contribute to the overall safety and reliability of these systems.

However, this all comes with some major caveats. LLMs can struggle to gain a deep understanding of context in real-world, dynamic situations and current documentation that LLMs are trained on is likely incomplete. Since LLMs are sensitive to ambiguities in natural language, misinterpretation of cues or context can lead to incorrect, potentially dangerous situations. The age-old adage of “garbage in, garbage out” holds true and LLMs are likely to inherit biases in training data that could inspire inequitable decision-making which would raise ethical concerns. They are highly dependent on the quality and diversity of their training data but these models may not accurately reflect real-world conditions, leading to unsafe responses. LLMs are vulnerable to adversarial attacks that could be exploited by malicious agents and compromise the safety of the system. The environments that mobile autonomous systems operate within are rapidly changing and LLMs may struggle to adapt to dynamic scenarios such as road closures. Although LLMs are great at producing explanations, they do not possess a deep understanding of the operation at hand, merely they produce reasonable explanations based on data they have seen. However, understanding natural-language and its nuances is not trivial and it is possible that LLMs may misinterpret instructions that are given by the user. Though LLMs produce explanations, the LLM itself is not transparent and does not explain how it made the decisions that it made, this lack of transparency and explainability can make it difficult to trust LLMs and the output that they generate.

Participants

- Houssam Abbas
Oregon State University – Corvallis, US
- Ezio Bartocci
TU Wien, AT
- Christof Budnik
Siemens – Princeton, US
- Filip Cano
TU Graz, AT
- Mauricio Castillo-Effen
Lockheed Systems – Arlington, US
- Chih-Hong Cheng
Universität Hildesheim, DE
- Patricia Derler
Zoox Inc. – Foster City, US
- Jyotirmoy Deshmukh
USC – Los Angeles, US
- Rayna Dimitrova
CISPA – Saarbrücken, DE
- Marie Farrell
University of Manchester, GB
- Michael Fisher
University of Manchester, GB
- Martin Fränzle
Universität Oldenburg, DE
- Mahsa Ghasemi
Purdue University – West Lafayette, US
- Radu Grosu
TU Wien, AT
- Rong Gu
Mälardalen University – Västerås, SE
- Ichiro Hasuo
National Institute of Informatics – Tokyo, JP
- Bardh Hoxha
Toyota Research Institute North America- Ann Arbor, US
- Taylor T. Johnson
Vanderbilt University – Nashville, US
- Panagiotis Katsaros
Aristotle University of Thessaloniki, GR
- Bettina Könighofer
TU Graz, AT
- Thierry Lecomte
CLEARSY – Aix-en-Provence, FR
- Lars Lindemann
USC – Los Angeles, US
- Assaf Marron
Weizmann Institute – Rehovot, IL
- Sayan Mitra
University of Illinois – Urbana Champaign, US
- Selma Music
Denso Automotive – Eching, DE
- Dejan Nickovic
AIT – Austrian Institute of Technology – Wien, AT
- Necmiye Ozay
University of Michigan – Ann Arbor, US
- Giulia Pedrielli
Arizona State University – Tempe, US
- Doron A. Peled
Bar-Ilan University – Ramat Gan, IL
- Stefan Pranger
TU Graz, AT
- Andoni Rodríguez
IMDEA Software Institute – Madrid, ES
- Darko Stern
AVL – Graz, AT
- Alessandro Zanardi
ETH Zürich, CH
- Dirk Ziegenbein
Robert Bosch GmbH – Renningen, DE

Report from Dagstuhl Seminar 24072

Triangulations in Geometry and Topology

Maike Buchin*[1], Jean Cardinal*[2], Arnaud de Mesmay*[3], Jonathan Spreer*[4], and Alex He†[5]

1 **Ruhr-Universität Bochum, DE.** maike.buchin@rub.de
2 **ULB – Brussels, BE.** jean.cardinal@ulb.be
3 **CNRS, Gustave Eiffel University – Marne-la-Vallée, FR.** arnaud.de-mesmay@univ-eiffel.fr
4 **University of Sydney, AU.** jonathan.spreer@sydney.edu.au
5 **The University of Queensland – Brisbane, AU.** a.he@uq.net.au

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar "Triangulations in Geometry and Topology" (24072). The seminar was held from February 12 to February 16, 2024, gathered 31 participants, and started with four introductory talks and an open problem session. Then the participants spread into small groups to work on open problems on diverse topics including reconfiguration of geometric shapes, geodesics on triangulated surfaces, distances in flip graphs, geometric cycles and algorithms in 3-manifold topology.

Seminar February 11–16, 2024 – https://www.dagstuhl.de/24072

2012 ACM Subject Classification Human-centered computing → Graph drawings; Mathematics of computing → Geometric topology; Theory of computation → Computational geometry

Keywords and phrases computational geometry, geometric topology, triangulations

Digital Object Identifier 10.4230/DagRep.14.2.120

1 Executive Summary

Maike Buchin
Jean Cardinal
Arnaud de Mesmay
Jonathan Spreer

This seminar was a followup to the Dagstuhl Seminars "Applications of Topology to the Analysis of 1-Dimensional Objects" (17072), "Computation in Low-Dimensional Geometry and Topology" (19352), and "Computation and Reconfiguration in Low-Dimensional Topological Spaces" (22062). The common idea behind all of these seminars is to bring together researchers from different communities (such as computational geometry, graph drawing, or geometric topology) with a shared interest in low-dimensional objects (e.g., curves, embedded graphs, knots, or surfaces). The goal of this approach is to foster collaborative work and synergies: The mathematical study of low-dimensional objects has a rich and old history, but research into their algorithmic and combinatorial properties and the underlying computational questions is still young. This makes for a vibrant research situation, giving strong opportunities for interdisciplinary work involving our multiple communities.

* Editor / Organizer
† Editorial Assistant / Collector

The focus of this Dagstuhl Seminar was placed on **triangulations**: partitions of the plane into triangles, or, more generally, of a space into simplices, which are required to meet face-to-face. Triangulations are typically constrained to use a given set of points as vertices and are fundamental tools in many applications such as computer graphics or geographic information systems. Alternatively, a triangulation can be defined on a topological space as a simplicial complex together with a homeomorphism from this simplicial complex to the space. These triangulations play an important role in the study of metrics on surfaces and their moduli space. The multiple facets of triangulations make them an essential and ubiquitous object of study in the combinatorial, algorithmic, geometric and topological properties of low-dimensional spaces, and thus they constitute a fertile ground for collaborations.

The seminar started with a quick introduction of all the participants, and four keynote talks on different aspects of triangulations.

- *Lionel Pournin* gave an overview of flip graphs and their combinatorial and algorithmic properties.
- *Saul Schleimer* surveyed the use of triangulations in 3-manifold topology and the numerous computational challenges that arise from them.
- *Linda Kleist* presented two snapshots on triangulations, first on the computational complexity of linearly embedding simplicial complexes, and then on some hamiltonicity properties of polytopes arising from flipping triangulations.
- *Mikkel Abrahamsen* explained new hardness proofs for algorithmic problems related to packing, covering and partitioning simple polygons with unit squares.

We refer to the abstracts later in this report for more details on these contributions.

These keynote introductory talks were followed with an extended open problem session where we gathered a large collection of open problems. Some of these were circulated in advance of the meeting, many of them were new. The remainder of the week was spent working in small groups actively trying to make progress on the most popular open problems. We made extensive use of the tool "Coauthor," designed by Erik Demaine (MIT). This allowed for a very efficient recording of the progress made in the different groups. Regular progress reports allowed participants to easily switch between groups during the week, or to start new groups, leading to a very dynamic working environment.

We now quickly survey the different problems that have been worked on during this very productive week:

- **Computational complexity of problems in 3-manifold topology.** This group discussed computational complexity for important algorithmic problems from 3-manifold topology. Examples of problems include showing that a knot is ribbon, and testing 0-efficiency in a 3-manifold triangulation (lead: Eric Sedgwick).
- **Veering triangulations and the flip graph.** This group studied the effect on flip distances of surface triangulations if a veering structure in the associated layered triangulation is (lead: Saul Schleimer).
- **Flip distances.** This group investigated distances in the flip graphs of triangulated convex n-gons or annuli, both theoretically and algorithmically (lead: Jonathan Spreer).
- **Hardness for simple polygons.** The group considered NP-hard problems on polygons with holes, and how to show that these are also NP-hard on simple polygons. (lead: Lena Schlipf)
- **Catching balls on Curves.** This group considered the problem of characterising curves for which there exist a strategy to catch a ball from any initial configuration. (lead: Maarten Löffler)

- **Computing geodesic paths using edge flips.** This group investigated the FlipOut algorithm defined by Crane and Sharp to compute geodesics on intrinsic geometric triangulations and its possible variants. (lead: Hsien-Chih Chang)
- **Rendering a knot without self-intersections.** This group focused on using topological data analysis techniques to produce instructive 3D models of link diagrams (lead: Clément Maria)
- **Shortest cycle separating k objects from $n-k$.** This group investigated the complexity of, given a collection of n objects, computing the shortest cycle separating k objects from $n-k$ objects, and how it behaves with respect to the parameter k. This was explored in the setting of the plane with obstacles and then planar graphs with obstacles, with an eye towards separating k handles in graphs embedded on surfaces. (lead: Éric Colin de Verdière)

In summary, this Dagstuhl Seminar provided a very fruitful research environment, allowing participants from very different backgrounds to work together on important open problems. Survey feedback from the participants highlighted how much the emphasis on intensive work in small groups was appreciated. In several of the working groups, significant progress was made, and the results are currently prepared to be submitted for publication. As in the previous meetings, the excellent quality of the Dagstuhl infrastructure and the impeccable support of the Dagstuhl staff provided a seamless experience for all the participants. We are hopeful that this successful experience will lead to follow-up Dagstuhl Seminars on related topics.

2 Table of Contents

3 Overview of Talks

3.1 Packing, Covering and Partitioning Simple Polygons with Unit Squares

Mikkel Abrahamsen (University of Copenhagen, DK)

Main reference Mikkel Abrahamsen, Jack Stade: "Hardness of Packing, Covering and Partitioning Simple Polygons with Unit Squares", CoRR, Vol. abs/2404.09835, 2024.
URL https://doi.org/10.48550/ARXIV.2404.09835
Main reference Mikkel Abrahamsen, Nichlas Langhoff Rasmussen: "Partitioning a Polygon Into Small Pieces", CoRR, Vol. abs/2211.01359, 2022.
URL https://doi.org/10.48550/ARXIV.2211.01359

We show that packing axis-aligned unit squares into a simple polygon P is NP-hard, even when P is an orthogonal and orthogonally convex polygon with half-integer coordinates. It has been known since the early 80s that packing unit squares into a polygon with holes is NP-hard [Fowler, Paterson, Tanimoto, Inf. Process. Lett., 1981], but the version without holes was conjectured to be polynomial-time solvable more than two decades ago [Baur and Fekete, Algorithmica, 2001].

Our reduction relies on a new way of reducing from Planar-3SAT. Interestingly, our geometric realization of a planar formula is non-planar. Vertices become rows and edges become columns, with crossings being allowed. The planarity ensures that all endpoints of rows and columns are incident to the outer face of the resulting drawing. We can then construct a polygon following the outer face that realizes all the logic of the formula geometrically, without the need of any holes.

This new reduction technique proves to be general enough to also show hardness of two natural covering and partitioning problems, even when the input polygon is simple. We say that a polygon Q is *small* if Q is contained in a unit square. We prove that it is NP-hard to find a minimum number of small polygons whose union is P (covering) and to find a minimum number of pairwise interior-disjoint small polygons whose union is P (partitioning), when P is an orthogonal simple polygon with half-integer coordinates. This is the first partitioning problem known to be NP-hard for polygons without holes, with the usual objective of minimizing the number of pieces.

We also describe a 13-approximation algorithm for the partitioning problem, and $O(1)$-approximation algorithms for several related problems.

3.2 2 Snapshots – Geometric Embedding of Complexes & Facet-Hamiltonian Cycles in Generalized Associahedra

Linda Kleist (TU Braunschweig, DE)

Joint work of Part 1: Mikkel Abrahamsen and Tillmann Miltzow; Part 2: Jean Cardinal, Stefan Felsner, Robert Lauff
Main reference Mikkel Abrahamsen, Linda Kleist, Tillmann Miltzow: "Geometric Embeddability of Complexes Is ∃ℝ-Complete", in Proc. of the 39th International Symposium on Computational Geometry, SoCG 2023, June 12-15, 2023, Dallas, Texas, USA, LIPIcs, Vol. 258, pp. 1:1–1:19, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
URL https://doi.org/10.4230/LIPICS.SOCG.2023.1
Main reference Jean Cardinal, Stefan Felsner, Linda Kleist, Robert Lauff. Facet-Hamiltonian Cycles in Permutohedra and Generalized Associahedra. (in preparation)

In this talk, we will consider two snapshots of triangulations in geometry and topology.

In the first part, we consider the decision problem of determining whether a given (abstract simplicial) k-complex has a geometric embedding in $\mathbb{R}^d$. In particular, we consider the case $d = 3$ and $k = 2$, i.e., a geometric embedding corresponds to a set of triangles in 3D, and show that this problem is complete for the Existential Theory of the Reals (ETR). As a matter of fact the result can be extended to all $d \geq 3$ and $k \in \{d-1, d\}$. We note that ETR-hardness implies NP-hardness and that our work constitutes the first hardness result for the algorithmic problem of geometric embedding of (abstract simplicial) complexes. This complements recent breakthroughs for the computational complexity of piece-wise linear embeddability. This part is based on joint work with Mikkel Abrahamsen and Tillmann Miltzow.

In the second part, we investigate facet-Hamiltonian cycles of polytopes, i.e., cycles in the skeleton such that every facet of the polytope is visited exactly once. These cycles can be understood as shortest tours that guard the facets of a polytope. Among other polytopes, we consider associahedra of type A where the 1-skeletons correspond to the flip graphs of triangulations of points in convex position. The skeletons of their generalizations, known as associahedra of type B/C and D, also allow for models involving triangulations which help to construct facet-Hamiltonian cycles. This part is based on ongoing joint work with Jean Cardinal, Stefan Felsner, and Robert Lauff.

3.3 Flips and surface triangulations

Lionel Pournin (Université Paris 13 – Villetaneuse, FR)

Joint work of Lionel Pournin, Hugo Parlier, Zili Wang

A triangulation of a topological surface Σ is a set of pairwise non-crossing arcs in Σ that decompose Σ into triangles and whose endpoints belong to a prescribed finite set of vertices. The flip-graph $\mathcal{F}(\Sigma)$ of Σ is the graph whose vertices are these triangulations and whose edges connect two triangulations when they differ by a single arc. This graph is always connected and, except for a few surfaces, it is infinite. By the Švarc–Milnor lemma, $\mathcal{F}(\Sigma)$ provides a combinatorial model for the mapping class group of Σ. More precisely, $\mathcal{F}(\Sigma)$ is quasi-isometric to the Cayley graphs of the mapping class group of Σ. When Σ is a Euclidean polygon whose vertex set is the vertex set of the triangulations, $\mathcal{F}(\Sigma)$ is finite and already interesting: it is the edge-graph of a polytope – the associahedron – and computing distances in this graph is a major open problem in computer science. This talk reviews the results obtained over the last few decades on the geometry of $\mathcal{F}(\Sigma)$.

In a first part, the case when Σ is a Euclidean polygon is reviewed. The focus is on diameter estimates for $\mathcal{F}(\Sigma)$ and on the computational problem of determining the distance of two given triangulations T_1 and T_2 of Σ. A popular heuristic for this problem, based on the number of arcs crossings between T_1 and T_2 is presented and discussed. The higher dimensional case and associated open problems are also mentioned.

In a second part, the general case when Σ is an arbitrary topological surface is reviewed. In that case, quotienting $\mathcal{F}(\Sigma)$ by the homeomorphisms of Σ results in a finite graph $\mathcal{MF}(\Sigma)$ whose diameter allows to estimate the constants of the quasi-isometry between $\mathcal{F}(\Sigma)$ and the mapping class group of Σ. The known results on the diameter of $\mathcal{MF}(\Sigma)$ [1, 2, 3] are presented and a number of open problems are given.

The third part focuses on the 3-dimensional model for the paths in $\mathcal{F}(\Sigma)$ due to Daniel Sleator, Robert Tarjan, and Wiliam Thurston in the case when Σ is a polygon and the vertex set of the triangulations coincides with that of Σ. In their seminal 1988 article [5], these authors prove that, if the path is a geodesic, then this model is a simplicial complex. The proof [4] that this simplicial complex is flag is presented. Several consequences are discussed as for instance, a proof that the arc crossings based flip distance estimation heuristic is, at best, a 1.5-approximation algorithm.

References

1. Hugo Parlier and Lionel Pournin, *Flip-graph moduli spaces of filling surfaces*, Journal of the European Mathematical Society **19** (2017), no. 9, 2697–2737.
2. Hugo Parlier and Lionel Pournin, *Modular flip-graphs of one holed surfaces*, European Journal of Combinatorics **67** (2018), 158–173.
3. Hugo Parlier and Lionel Pournin, *Once punctured disks, non-convex polygons, and pointihedra*, Annals of Combinatorics **22** (2018), 619–640.
4. Lionel Pournin and Zili Wang, *Strong convexity in flip-graphs*, arXiv:2106.08012 (2021).
5. Daniel Sleator, Robert Tarjan and William Thurston, *Rotation distance, triangulations, and hyperbolic geometry*, Journal of the American Mathematical Society **1** (1988), no. 3, 647–681.

3.4 The homeomorphism problem

Saul Schleimer (University of Warwick – Coventry, GB)

We discuss the homeomorphism problem for manifolds, mostly in dimension three, from an algorithmic point of view. We give a few definitions, a bit of history, and a selection of open problems.

This is an edited, and slightly expanded, version of my talk. I have assumed a certain amount of background in computational complexity. I have not attempted to trace the open problems back to their original posers.

3.4.1 Manifolds

A basic object of study in topology is the *n–manifold* (with boundary): a (second countable and Hausdorff) topological space M which is locally modelled on $\mathbb{R}^{n-1} \times \mathbb{R}_{\geq 0}$. The *boundary* of M, denoted ∂M, is the subset of points of M which do not have charts in $\mathbb{R}^n$. We call a manifold M *closed* if M is compact and ∂M is empty.

As a few simple, but relevant, initial examples we have the n–*sphere*, the n–*ball*, and the n–*torus*:

$$S^n = \{x \in \mathbb{R}^{n+1} \mid |x| = 1\}$$
$$B^n = \{x \in \mathbb{R}^n \mid |x| \leq 1\}$$
$$T^n = S^1 \times S^1 \times \ldots \times S^1 \text{ (n copies)}$$

We write $M \cong N$ if the manifolds M and N are homeomorphic. By *invariance of domain* if $M \cong N$ then M and N necessarily have the same dimension. Here, then, is the foundational *homeomorphism problem*:

Homeo(n)

Instance: A pair of n–manifolds M and N.

Question: Is $M \cong N$?

Note that the homeomorphism problem immediately reduces to the connected case.

Suppose that N is a connected n–manifold. We may simplify the homeomorphism problem to the *recognition problem*:

Recog(N)

Instance: An n–manifold M.

Question: Is $M \cong N$?

The recognition problem should, "morally speaking", be easier than the homeomorphism problem. This is because we may use properties particular to N when constructing our algorithm.

3.4.2 PL-triangulations

To give a connected manifold M (say, as input to a decision problem) we require a uniform and finitary way to describe it. Note that there are uncountably many homeomorphism types of non-compact n–manifolds (for $n \geq 2$); thus, they do not admit finitary descriptions. Accordingly, we henceforth restrict our attention to compact and connected manifolds.

Freedman and Zuddas (2018) prove that compact n–manifolds (in the categories TOP, PL, and DIFF) admit uniform and finitary descriptions. The most difficult case occurs in dimension four. To avoid these subtleties, we will restrict our attention to manifolds presented via *PL-triangulations*. We begin as follows.

▶ **Definition 1.** The *standard n-simplex* is defined by

$$\Delta^n = \left\{x \in \mathbb{R}^{n+1} \mid x_i \geq 0 \text{ and } \textstyle\sum_i x_i = 1\right\}$$

A *model simplex* is simply a copy of Δ^n. A *face-pairing* is an affine isomorphism between two (co-dimension one) faces of two (possibly equal) model simplices. A *triangulation* $\mathcal{T}$ is a finite collection of model simplices together with finitely many face-pairings. The *realisation* of $\mathcal{T}$, denoted $|\mathcal{T}|$, is obtained by taking the disjoint union of the model simplices and quotienting by the face-pairings.

Suppose that M is an n-manifold. Suppose that $\mathcal{T}$ is a triangulation. We say that $\mathcal{T}$ is a *triangulation of* M if $M \cong |\mathcal{T}|$.

Next we give the (necessarily recursive) definition of a *PL-triangulation*.

▶ **Definition 2.** Suppose that M is an n-manifold. Suppose that $\mathcal{T}$ is a triangulation of M. We say that $\mathcal{T}$ is a *PL-triangulation of M* if the vertex links in $\mathcal{T}$ are PL-triangulations of S^{n-1}.

Two PL-triangulations of M are *PL-homeomorphic* if they have isomorphic subdivisions.

All manifolds in dimensions three and below have triangulations, all such triangulations are PL, and any two such PL-triangulations on a single manifold are PL-homeomorphic (Radó 1925 and Moise 1953). On the other hand, work of Freedman, Casson, and Donaldson (and also Kirby-Siebenmann) gives

- four-manifolds that have no triangulation at all and
- four-manifolds that have many distinct PL-structures.

Note, however, that a triangulation of a four-manifold is necessarily a PL-triangulation. [1] In dimension five and above the situation is even more complicated.

To avoid these subtle points, we will essentially restrict our attention to the PL category.

3.4.3 Away from dimension three

Recall our standing hypotheses: the n–manifold M is assumed to be connected, compact, and given via a PL-triangulation.

Thus the only zero-manifold allowed is a single point; this has a unique triangulation and so Homeo(0) has a constant-time solution. The only one-manifold allowed is the circle; so Homeo(1) amounts to recognising when a finite graph is a cycle. This can be done in polynomial time.

The two-dimensional case is more difficult. By the *classification of surfaces* a pair of compact, connected two-manifolds are homeomorphic if and only if they

- have the same Euler characteristic,
- have the same number of boundary components, and
- are either both orientable or both non-orientable.

This gives a polynomial-time solution to Homeo(2).

In higher dimensions (again working in the PL category) there are many negative results.

- For $n \geq 4$ the problem Homeo(n) is not decidable (Markov 1958).
- For $n \geq 5$ the problem Recog(S^n) is not decidable (Novikov).
- Thus, in dimensions six and higher, there is no algorithm to decide if a triangulation is PL.

In dimension four there are many well-known, and widely open, algorithmic questions.

Recog(S^4)

Instance: A four–manifold M, given by a PL-triangulation.

Question: Is M PL-homeomorphic to S^4?

This relates to, but would not be answered by, the last remaining part of the Poincaré conjecture: namely, is there a unique smooth (and thus PL) structure on the four-sphere?

Bounding

Instance: A PL three-sphere M, as a submanifold of S^4.

Question: Is M PL-isotopic to a great three-sphere?

This is solved by the trivial algorithm (printing Yes) if the Schoenflies problem has an affirmative answer in dimension four.

[1] This relies on the resolution of the three-dimension Poincaré conjecture (Hamilton, Perelman).

UnknotRecog(4)

Instance: A PL two-sphere M, as a submanifold of S^4.

Question: Is M PL-isotopic to a great two-sphere?

Note that the fundamental groups arising from two-knots in the four-sphere are not well understood.

PromoteHomeomorphism

Instance: A pair of PL four-manifolds M and N which are TOP-homeomorphic.

Question: Is M PL-homeomorphic to N?

This is closely related to the problem of classifying the PL-structures on a given closed TOP four-manifold.

3.4.4 Homeo(3)

It is a "folklore result" (Kuperberg, 2019) that the geometrisation theorem (Perelman, 2002-3) implies that Homeo(3) is decidable. Kuperberg sharpens this result by proving that there is an elementary recursive algorithm (for closed, connected, oriented three-manifolds). Improving this upper bound seems possible.

▶ **Question 3.** Does Homeo(3) lie in NP?

We suggest approaches to this problem below.

▶ **Question 4.** Does Homeo(3) lie in coNP?

This second question seems much more doubtful. For example, there is a polynomial-time reduction of GraphIso (graph isomorphism) to Homeo(3) (Lackenby 2022). While GraphIso is now known to be quasi-polynomial time (Babai 2016), it is not yet known to lie in coNP.

In the other direction, Burton asks for lower bounds.

▶ **Question 5.** Is Homeo(3) NP-hard?

The three questions above are in tension, due to the standard conjectures in computational complexity. There is evidence that Homeo(3) is difficult: for example knot genus in arbitrary manifolds is NP-complete (Agol, Hass, Thurston 2006) and various diagrammatic problems for links are NP-complete or NP-hard (Lackenby 2017; Koenig, Tsvietkova 2021). There are other candidate problems which may be NP–complete.

HeegaardGenus

Instance: A three-manifold M and an integer g.

Question: Does M have Heegaard genus at most g?

This is known to be decidable (Li 2011) and NP–hard (Bachman, Derby-Talbot, Sedgwick 2017). It seems that showing this lies in NP requires a delicate control of the *normal tori* in a triangulation.

Embed(S^3)

Instance: A three-manifold M.

Question: Does M embed in S^3?

Again, this is decidable (Matoušek, Sedgwick, Tancer, Wagner 2018) and NP–hard (de Mesmay, Rieck, Sedgwick, Tancer 2020).

3.4.5 Triangulations for their own sake

Suppose that $\mathcal{T}$ is a triangulation; let $c(\mathcal{T})$ be the number of tetrahedra of $\mathcal{T}$. For any compact connected three-manifold M we define its *complexity* as follows.

$$c(M) = \min\{c(\mathcal{T}) \mid M \cong |\mathcal{T}|\}$$

The complexity of M is morally similar to, but finer than, the Heegaard genus of M.

Complexity

Instance: A three-manifold M and an integer n.

Question: Is $c(M) \leq n$?

This problem is decidable, because Homeo(3) is decidable. However, we do not have efficient algorithms to solve this question. We do not even know how, in general, to estimate $c(M)$ within any fixed factor $K \geq 1$. The exact computation of $c(M)$ is only known for a finite number of examples (coming from various censuses) and some special families (Jaco, Rubinstein, Spreer, Tillmann). Estimating $c(M)$ to within a constant factor is known in some special families (Lackenby, Purcell, Jackson). In another other direction we have a question of Gromov:

▶ **Question 6.** How many triangulations does S^3 have with at most n tetrahedra?

The growth is at least exponential and at most exp-poly. We end this section with a question which feels, paradoxically, related to the above.

▶ **Question 7.** How many finite-volume hyperbolic, closed, connected three-manifolds M are there with $c(M) \leq n$?

As a closely related problem, we should ask for a count of finite-volume cusped hyperbolic three-manifolds with complexity at most n. It may be possible to relate the counts, in the closed and cusped cases, by understanding how the complexity $c(M)$ changes under Dehn filling.

3.4.6 The graph of triangulations

Suppose that Δ is the four-simplex. We divide $\partial\Delta$ combinatorially into a pair of triangulated three-balls, say B and B'. Let k and k' be the number of tetrahedra in B and B' respectively. Note that $k + k' = 5$.

Suppose that $\mathcal{T}$ is a triangulation of a three-manifold M. Suppose that a copy of B appears as a subtriangulation of $\mathcal{T}$. Then we may produce a triangulation $\mathcal{T}'$, again of M, by replacing the copy of B in $\mathcal{T}$ by a copy of B'. This move, from $\mathcal{T}$ to $\mathcal{T}'$, is called a k-k' *bistellar flip*.

Suppose that M is a compact connected three-manifold. We now define $\mathrm{GT}(M)$ to be the *graph of triangulations* of M: vertices are (isomorphism classes of) triangulations of M and edges are bistellar flips. Note that $\mathrm{GT}(M)$ is connected (Pachner 1991). Navigation in the graph $\mathrm{GT}(M)$ is important both in theory and in practice.

▶ **Question 8.** Suppose that M is a compact connected three-manifold. Let d_M be the distance function in $\mathrm{GT}(M)$. Is $d_M(\mathcal{T}, \mathcal{T}')$ is bounded above by a some polynomial f_M in $c(\mathcal{T})$ and $c(\mathcal{T}')$?

If we restrict ourselves to geometric triangulations of hyperbolic manifolds M (with a lower bound on injectivity radius) then f_M is at worst cubic (Kalelkar, Phanse 2021). Supposing that such polynomials f_M exist, we may ask the following.

▶ **Question 9.** Does f_M depend on M? Are there manifolds M where f_M is linear?

▶ **Question 10.** Suppose that M is a compact connected three-manifold. Let h_M be the *height* function in $\mathrm{GT}(M)$; that is, $h_M(\mathcal{T}, \mathcal{T}')$ is the smallest possible maximal complexity appearing in a path (in $\mathrm{GT}(M)$) from $\mathcal{T}$ to $\mathcal{T}'$. Is there a polynomial upper bound on the height?

One problem related to the above is as follows: fix a manifold M and a number k and then try to find a random triangulation $\mathcal{T}$ of M with $k = c(\mathcal{T})$. Another is as follows: fix a number k and try to find a random manifold M with $c(M) \leq k$.

3.4.7 Recognition

Suppose that $\mathcal{C}$ is a class of three-manifolds.

Recog($\mathcal{C}$)

Instance: A triangulation $\mathcal{T}$ of a three–manifold.

Question: Does $|\mathcal{T}|$ belong to the class $\mathcal{C}$?

When $\mathcal{C}$ is the class of a single manifold M then this reduces to Recog(M). For example, recognising the three-sphere lies in the class NP (Schleimer 2011). More generally, recognising elliptic manifolds lies in the class NP (Lackenby, Schleimer 2022). The same holds for torus bundles over the circle (folklore) and Seifert fibred spaces with boundary (Jackson 2023).

▶ **Question 11.** Is the problem of recognising small Seifert fibred spaces in NP?

This problem is decidable (Rubinstein 2004, Li 2006). One possible NP-certificate might start by finding a fibre circle with length at most linear (in the complexity of the given triangulation).

▶ **Question 12.** Is the problem of recognising closed hyperbolic three-manifolds in NP?

This problem is decidable (Casson, Manning 2002). One possible NP-certificate might start by finding a systole with at most linear length (in the complexity of the given triangulation).

▶ **Question 13.** Suppose that M is a surface bundle over the circle, given via a triangulation $\mathcal{T}$. Suppose that $H_1(M, \mathbb{Z}) = \mathbb{Z}$. Is there a uniform description, with complexity at most polynomial in $c(\mathcal{T})$, of the monodromy of the surface bundle structure?

This is a delicate question; the genus of the fibre may be exponentially large in the complexity of the given triangulation.

Acknowledgements

I thank Arnaud de Mesmay for his useful comments.

4 Working groups

4.1 Computing Geodesic Paths using Edge Flips

Hsien-Chih Chang (Dartmouth College – Hanover, US), Florestan Brunck (University of Copenhagen, DK), Arnaud de Mesmay (CNRS, Gustave Eiffel University – Marne-la-Vallée, FR), Tim Ophelders (Utrecht University, NL), Rodrigo I. Silveira (UPC Barcelona Tech, ES), Stephan Tillmann (University of Sydney, AU), Zili Wang (Sun Yat-Sen University – Shenzen, CN), and Erin Moriarty Wolf Chambers (St. Louis University, US)

Let T be a triangulation embedded on a surface $\mathcal{M}$, which comes with a metric ℓ equipped where every triangle is flat. The pair $(\mathcal{M}, \ell)$ is called an *intrinsic triangulation*. We view T as a Δ-complex in the sense of Hatcher, as well as an embedded (multi-)graph on $\mathcal{M}$ with possibly loops and multiedges. A *geodesic* on an intrinsic triangulation $(\mathcal{M}, \ell)$ is a locally shortest path (in the sense that small perturbation cannot make the curve shorter). One way to characterize the geodesics on an intrinsic triangulation is that the path consists of straight segments on the triangles and never passes through a positive-curvature vertex; when a subpath $[abc]$ passes through a negative-curvature vertex b, the two angles on the two sides of the consecutive segments ab and bc are both at least degree π (half a circle).

Computing geodesics is an important question due to its applications in many areas including path planning [1], surface partitioning [2] and terrain navigation [3]. In this report we focus on an algorithm that computes geodesics quickly and accurately in practice using the idea of "edge flip", which is the replacement of one diagonal in a convex quadrilateral by the other. Consider the FLIPOUT algorithm, designed by Nick Sharp and Keenan Crane [4]: The input is a triangulation T, and a *flexible*[2] joint $[abc]$ in a given walk W on T, where $\angle abc < \pi$ and segments ab and bc are distinct. The output is a shorter edge path connecting a to c in an updated triangulation T.

```
FlipOut(T, [abc]):
  while any β_i < π:
      j ← min i such that β_i < π
      FlipEdge(T, bn_j)

  return (T, [a, n_1, ..., n_{k−1}, c])
```

FLIPOUT iteratively modifies the triangulation T and the walk W on the triangulation in a way that shortens the walk. Sharp and Crane proved that the algorithm terminates in finitely many steps if the input walk W is a path with two fixed endpoints. They also provided a modification of the FLIPOUT algorithm so that it can be applied to free closed loops on the surface. However they could not prove termination of the algorithm with this modification.

▶ **Question 1.** Does the FLIPOUT algorithm terminate when the input walk is a free closed loop?

We can also ask the question about efficiency.

[2] See [4] for definition. Intuitively a joint is flexible if we can safely shortens it using edge flips without causing the curve to self-intersect.

▶ **Question 2.** How many edge flips does the FlipOut perform to turn the input walk into a geodesic?

Termination. Our group mostly focused on investigating the first question. The technical challenge can be summarized as follows. When applied to a loop, the FlipOut algorithm can get stuck in the specific case where it reduced the input curve to a single edge of the intrinsic triangulation (which is thus a loop) while still not being geodesic because there is an angle less than π at the unique vertex of that edge. In this situation, the flip specified by the algorithm might be impossible. Crane and Sharp propose a work-around in this case: they apply an *edge-move*, which has the effect of slightly pushing the loop and replacing it with another loop with more edges. While they provide a proof that this makes progress locally (in the sense that the FlipOut algorithm will not immediately revert back to the offending loop), this edge-move hinders the proofs of termination of the algorithm. Indeed, in the case of walks, one of the main ingredients in the proof that the FlipOut algorithm terminates is that the length of the walk is monotonically decreasing throughout the process. In contrast, here, the edge move might increase the length of the curve, thereby rendering the monotonicity argument invalid.

In the case where the curve to be reduced is separating, it seems that strong constraints on the local geometry of the situation when the reduced curve consists of a single edge of the triangulation can help and provide a positive answer to the termination question. The rationale is as follows. Let us denote by γ the curve that we are reducing. Throughout the algorithm, either γ never gets reduced to a single edge, in which case, the proof of termination for walks applies. Otherwise, let us denote by γ_1 the first occurrence in time where it becomes a single edge. Then, applying an edge-move pushes γ on one side of γ_1, without loss of generality let us choose an orientation on γ_1 so that this push is to its right. Then Crane and Sharp prove that the following move will once again push to the right. At this stage, the curve γ is disjoint from γ_1. And the same proof shows that no subsequent move in the algorithm will ever move the curve so that it hits back γ_1 from the right. But by the assumption that γ is separating, so is γ_1, and thus this means that γ will forever stay disjoint from γ_1. So γ_1 acts as some sort of barrier for the FlipOut algorithm, removing a portion of the intrinsic triangulation from consideration. Since the triangulation is finite, this is a definite measure of progress, which we can use to prove termination.

Let us observe that on a topological sphere, all the closed loops are separating, so this observation is already of wide applicability. The case of surface of genus larger than one could also be amenable to such an analysis, since even though it could a priori happen that the curve γ comes back to hit one of the barrier curves γ_i from the left, it will do so without being homotopic to it. The case of the torus however remains impervious to this analysis, as it cannot rule out a curve γ infinitely looping around the torus, always going to the right, jumping from one barrier curve to the next. A significant effort was spent investigating other reasons why the algorithm terminates in that case, as well as in designing variants and alternatives to the edge-move for which termination can be proved, and we are hopeful that these efforts will coalesce into definite progress in follow-up work after the seminar.

Efficiency. The question of the complexity of the algorithm seems delicate, and we began exploring various approaches to it. Of course, without a proof of termination for surfaces, it seems necessary to consider additional restrictions on the input triangulation. Here, a case familiar to computational geometers which is of wide utility, could be the one of polyhedral terrains, so we began by investigating this setting and attempting to give a polynomial bound. However, the property of being a terrain is not preserved by the FlipOut algorithm, which makes this hypothesis quite at odds with the philosophy of this algorithm. It seems necessary to consider more general triangulated but simply connected settings to make further progress.

References

1 Ferenc A. Jolesz, William E. Lorensen, Hiroshi Shinmoto, Hideki Atsumi, Shin Nakajima, Peter Kavanaugh, Pairash Saiviroonporn, Steven E. Seltzer, Stuart G. Silverman, Margaret Phillips, and Ron Kikinis. Interactive virtual endoscopy. *American Journal of Roentgenology*, 196(5):1229–1235, 1997.
2 Sagi Katz and Ayellet Tal. Hierarchical mesh decomposition using fuzzy clustering and cuts. *ACM Transactions on Graphics*, 22(3):954–961, 2003.
3 Lanthier Mark, Anil Maheshwari, and J.-R. Sack. Approximating shortest paths on weighted polyhedral surfaces. *Algorithmica*, 30(4):527–562, 2001.
4 Nicholas Sharp and Keenan Crane. You can find geodesic paths in triangle meshes by just flipping edges. *ACM Trans. Graph.*, 39(6), Nov 2020. doi:10.1145/3414685.3417839.

4.2 Cycles Enclosing Objects in the Plane

Éric Colin de Verdière (Gustave Eiffel University – Marne-la-Vallée, FR), Therese Biedl (University of Waterloo, CA), Fabrizio Frati (University of Rome III, IT), Anna Lubiw (University of Waterloo, CA), Günter Rote (FU Berlin, DE), and Alexander Wolff (Universität Würzburg, DE)

4.2.1 Introduction

We investigate the problem of enclosing a subset of objects in the plane using a cycle of minimum total length. More specifically, given n disjoint polygons in the plane, of total complexity N, and a number k, the problem is to find a minimum-perimeter weakly simple polygon containing exactly k of the polygons. We also consider a variant of the problem where the k polygons to be enclosed are specified as part of the input (see Figure 1). We are interested in the case where k is small with respect to n, and aim for efficient algorithms in this fixed parameter setting.

We also consider a different, less geometric, setting where we are given as input a plane graph G of total complexity N, a set of n marked faces, an integer k, and we have to compute a shortest weakly simple closed walk in the graph that encloses exactly k of the n marked faces. As above, we also consider the variant where the k faces to be enclosed are given as part of the input. The graph setting is more general than the polygon setting, via the following polynomial time reduction, which is based on the observation that the desired cycle walks on visibility edges: given a set of disjoint polygons, construct a plane graph by finding all visibility segments in the free space and adding vertices at their intersections to create a plane graph.

Other variants include the problem of enclosing *at least* k objects (rather than *exactly* k), and – in the geometric setting – dealing with the presence of polygonal obstacles that may freely be included in, or excluded from, (but not cut by) the desired cycle.

These enclosure problems look very natural but seem rather unexplored, and are related to many well-studied problems, in particular Steiner tree and subset TSP. Our goal is to nail down the computational complexity of the problem and its many variations. We first survey related previous work and simple observations, and then present our results and other promising directions.

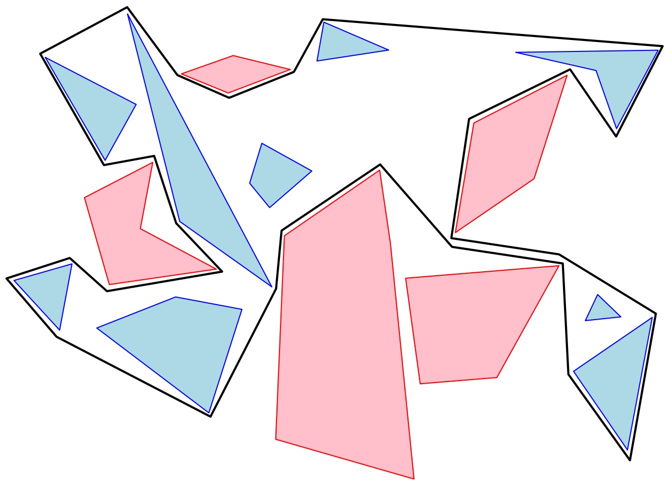


Figure 1 An instance of our geometric problem in which one looks for a minimum perimeter weakly simple polygon containing the blue polygons but none of the red ones. Note that the polygon "runs along itself" naturally, in other words it is weakly simple but not simple.

4.2.2 Related Previous Work and Simple Observations

Both in the graph setting and in the geometric setting, we must allow the cycle or the boundary of the solution polygon to use the same edge more than once. Such a cycle, which may touch itself but must not cross, is called weakly simple. Figure 1 shows a symbolic drawing where the two copies of an edge are visually separated.

Point objects. A natural special case of our problem is when each polygon is actually a point. Given a set of n points in the plane and an integer k, a minimum-perimeter polygon enclosing exactly k points is convex, and can thus be found in polynomial time using a result by Eppstein et al. [10, Corollary 5.3(3)]. Here the run-time depends only on n, i.e., k need not be restricted to a constant.

On the other hand, if we are given the k points to enclose, then the problem is NP-hard if k is not restricted, by a result by Eades and Rappaport [8]. However, this NP-hardness proof breaks down if k is small, say $k = O(1)$.

Separation by fences. In a recent work, Abrahamsen et al. [1] studied a variation of the problem in which one is given a set of disjoint *colored* polygons, and the goal is to compute a graph H embedded in the *free space* (the plane minus the interior of the polygons) such that the polygons inside any face of H all have the same color. The desired solution is a graph, which is not necessarily a weakly simple cycle, and may even be disconnected. This gives the problem a very different character. The problem is related to minimum cut (in the case of two colors) and multicut (in the general case). In contrast, our problem does not seem to have a direct relation to cuts, although it does relate to homology tools used for cut problems.

Computational topology. Enclosure problems have a topological flavor, so computational topology tools may help. Shortest homotopic cycles can be computed in polynomial time, both in the geometric setting [2, 9] and the graph setting [6, 5]. Thus, our problem boils down to finding the correct homotopy class of the solution polygon.

Homology is a coarser relation than homotopy that is more closely related to our problem. In particular, when the k objects to be enclosed are specified, our problem boils down to computing a shortest cycle in a given homology class (over $\mathbb{Z}_2$, in the plane minus the objects to be enclosed). The best known algorithm for this purpose [4] implies a $2^n \cdot N^{O(1)}$-time algorithm, and the question is whether we can do better for small k.

Finally, the problem of computing a shortest weakly simple cycle that encloses a given set of k objects, and possibly more, is solvable in polynomial time (even for unrestricted k), using techniques for shortest k-essential cycles [11, Section 5].

Steiner tree. The graph version of the enclosing problem has a flavor similar to the Steiner tree problem. Marx et al. [12] proved that, in an unweighted planar graph of size N, Steiner tree with k terminals cannot be solved in $2^{o(k)} \cdot N^{O(1)}$ time, assuming the Exponential Time Hypothesis (ETH). This immediately implies a hardness result for our problem if we force the cycle to behave like a Steiner tree.

▶ **Lemma 1.** The following problem cannot be solved in $2^{o(k)} \cdot N^{O(1)}$ time assuming ETH: Given an unweighted connected planar graph G of size N, a set F of k marked faces, compute a shortest weakly simple cycle enclosing F but no other face of G.

4.2.3 Exact Algorithm for the Geometric Setting

We have promising strategies to solve the following problem.

▶ **Problem 2.** Given as input a set P of n disjoint *convex* polygons in the plane, of total complexity N, and a subset $Q \subseteq P$ of k polygons, compute a shortest weakly simple polygon in $\mathbb{R}^2 \setminus P$ that encloses each polygon in Q but no polygon in $P \setminus Q$.

Note that if we only want to enclose k polygons, without specifying the subset Q, we may enumerate each of the $n^{O(k)}$ subsets Q of P of size k and apply the above result to each such subset Q in turn.

We mimic the dynamic programming algorithm of Dreyfus and Wagner [7] that computes Steiner trees in graphs. We aim for an algorithm with running time $3^k \cdot N^{O(1)}$. The basic idea is to compute, for each subset Q' of Q and each line segment uv, where u and v are input vertices, a shortest cycle $c(Q', uv)$ having uv on its boundary and such that a polygon has odd winding number with respect to the cycle if and only if the polygon lies in Q'. No edge of $c(Q', uv)$ may cross a polygon of P except that uv may cross polygons of Q. We expect to be able to compute the values of $c(Q', uv)$ by increasing size of Q' by merging smaller solutions in the same spirit as the Dreyfus–Wagner approach. At the root of the recursion, we untangle the cycle to make it weakly simple without making it longer, and prove that it encloses precisely Q by the condition on the winding numbers.

Possible improvements include removing the convexity assumption.

4.2.4 An Approximation Algorithm for the Graph Setting

▶ **Problem 3.** Given as input a connected plane graph G with N vertices, edges, and faces, a positive weight $w(e)$ for each edge e, a subset P of n marked faces and a subset $F \subseteq P$ of k faces, compute a minimum-weight weakly simple cycle in G that encloses the faces of F but no other face in P.

▶ **Proposition 4.** There is an approximation algorithm for Problem 3 that runs in time $N^{O(k)}$ and computes a $(k + 2 - \frac{2}{k})$-approximation.

Lemma 1 implies that solving the problem exactly cannot be done in $2^{o(k)} \cdot N^{O(1)}$ time assuming ETH.

Proof sketch. We can assume that $k < n$, for otherwise it suffices to compute a weakly simple cycle homotopic to the outer face, which can be done in polynomial time. Let $X = P \setminus F$ be the set of faces that we want to exclude, and let $F = \{f_1, \dots, f_k\}$.

Let $(r_1, \dots, r_k) \in \{0, \dots, N\}^k$. Below, we describe a procedure that will return a feasible solution to the problem (if it returns anything at all). If, for every $i \in \{1, \dots, k\}$, r_i equals the distance (in terms of the number of edges of a shortest path) from each face f_i to c_{OPT}, the procedure will return a solution, and it will be a $(k + 2 - 2/k)$-approximation of the optimal solution. Our overall algorithm returns the shortest solution found by the procedure, over all choices of $(r_1, \dots, r_k) \in \{0, \dots, N\}^k$.

The procedure for a given k-tuple $(r_1, \dots, r_k)$ is the following.

1. For each $i = 1, \dots, k$, we remove all vertices at distance less than r_i from f_i (where the "distance" is the number of edges), enlarging f_i to a larger face f'_i. If some f'_i contains some face in X, we exit without returning anything. Let G' be the resulting graph.
2. We describe a subroutine that computes a set Γ of cycles in G' satisfying, at each step of the subroutine, the following invariants:
   - the cycles in Γ are weakly simple and weakly pairwise disjoint;
   - they are pairwise non-nested;
   - no cycle in Γ encloses a face in X;
   - (*) for each cycle γ in Γ, there is a face $f \in F$ such that γ is, in $\mathbb{R}^2 \setminus (X \cup f')$, a shortest cycle homotopic to the boundary of f' (in G').

   At the end of the subroutine, each face of F is enclosed by a (unique) cycle in Γ.
   Here is the subroutine. Initially, Γ is empty. While some face f of F is enclosed by no cycle in Γ, we do the following:
   a. Let γ be the shortest cycle in G' among those that are weakly disjoint from Γ and homotopic, in $\mathbb{R}^2 \setminus \{X \cup f'\}$, to the boundary of f'. We can compute γ in near-linear time as follows: Remove the interior of each cycle in Γ, and then compute a shortest cycle homotopic, in $\mathbb{R}^2 \setminus \{X \cup f'\}$, to the boundary of f', using, e.g., [3, Lemma 4.1]; γ will automatically be weakly simple.
   b. Add γ to Γ, and then remove from Γ all cycles enclosed by γ, if any.

   The process clearly terminates after at most k steps. The invariants are clearly satisfied, except the last one (*), which we can prove separately. (It suffices to prove that some shortest cycle homotopic, in $\mathbb{R}^2 \setminus \{X \cup f'\}$, to the boundary of f' is weakly simple and weakly disjoint from Γ, and we omit the proof here.)
3. We compute a minimum spanning tree $\hat{T}$ in an auxiliary graph $\hat{G}$, which is a complete graph with a node for every cycle in Γ. The edge weights are the shortest distances between the corresponding cycles in G'. Every edge of $\hat{T}$ corresponds to a path in G'. When we overlay these paths they form a subgraph $\bar{T}$ that connects the cycles in Γ. We remove multiple edges from $\bar{T}$ and (conceivably) edges that create cycles in $\bar{T}$, until we end up with a forest T that still connects the cycles in Γ. (T is a *Steiner tree* in the graph where each cycle in Γ and its interior is contracted into a single vertex.)
4. We combine T and the cycles in Γ to obtain a weakly simple cycle of length $2w(T) + w(\Gamma)$ enclosing the faces of F but no face of X; see Figure 2.

The running time of the overall algorithm is clearly $N^{O(k)}$.

It remains to bound the length of the output. Let c_{OPT} be a shortest weakly simple cycle enclosing F but no face of X. We consider the cycle c obtained in the previous procedure in the case where r_i is the graph distance from c_{OPT} to f_i. It suffices to prove that

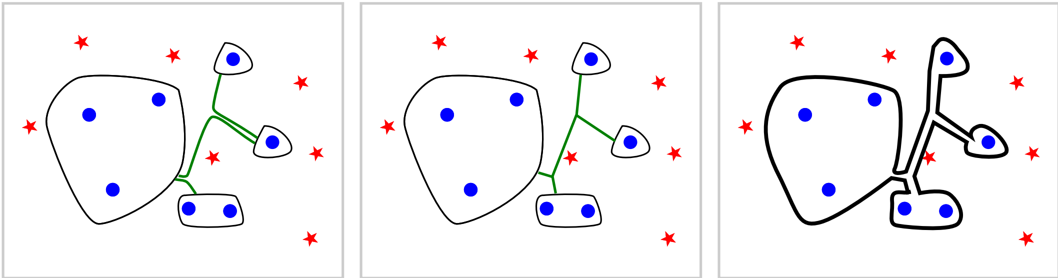


Figure 2 Illustration for the proof of Proposition 4. Faces in F are denoted by blue dots, faces in $P \setminus F$ are denoted by red stars. Left: The cycles in Γ are connected using a minimum spanning tree in an auxiliary graph. Middle: The spanning tree is turned into a Steiner tree T. Right: We assemble the cycles in Γ and the Steiner tree T to form a cycle enclosing F.

$w(c) \leq (k + 2 - \frac{2}{k})w(c_{\text{OPT}})$. We have $w(c) = 2w(T) + w(\Gamma)$, where T and Γ are obtained in the iteration computing c. Moreover, by the choice of the r_i's:

- Each of the at most k cycles γ in Γ is no longer than c_{OPT}. Indeed, by Property (*), γ is a shortest cycle homotopic to the boundary of f' in $\mathbb{R}^2 \setminus (X \cup f')$, for some $f \in F$, and c_{OPT} is homotopic to the boundary of f' in $\mathbb{R}^2 \setminus (X \cup f')$, for each $f \in F$.
- $w(T) \leq w(c_{\text{OPT}})\frac{k-1}{k}$, because c_{OPT} is connected and intersects the boundary of each cycle in Γ (perhaps unless Γ is a single cycle, in which case the claim is still true). Let $p_1, \ldots, p_{k'}$ be intersection points between c_{OPT} and the $k' \leq k$ cycles of Γ, in the order in which they are visited by c_{OPT}. Then the sections of c_{OPT} from p_1 to p_2, from p_2 to p_3, etc, from $p_{k'-1}$ to $p_{k'}$, whose total length is bounded by $w(c_{\text{OPT}})$, yield an upper bound on the weight of the minimum spanning tree $\hat{T}$ in Step 3, and hence $w(T) \leq w(c_{\text{OPT}})$. By cyclically shifting the sequence $p_1, \ldots, p_{k'}$ such that the omitted section from $p_{k'}$ for p_1 is the longest among the k' segments, the bound is improved by the factor $\frac{k'-1}{k'} \leq \frac{k-1}{k}$. ◄

References

1 M. Abrahamsen, P. Giannopoulos, M. Löffler, and G. Rote. Geometry multicut: shortest fences for separating groups of objects in the plane. *Discrete Comput. Geom.*, 64:575–607, 2020. doi:10.1007/s00454-020-00232-w.

2 S. Bespamyatnikh. Computing homotopic shortest paths in the plane. *J. Algorithms*, 49(2):284–303, 2003. doi:10.1016/S0196-6774(03)00090-7.

3 S. Cabello, M. DeVos, J. Erickson, and B. Mohar. Finding one tight cycle. *ACM Trans. Algorithms*, 6(4):Article 61, 2010. doi:10.1145/1824777.1824781.

4 E. W. Chambers, J. Erickson, K. Fox, and A. Nayyeri. Minimum cuts in surface graphs. *SIAM J. Comput.*, 52(1):156–195, 2023. doi:https://doi.org/10.1137/19M1291820.

5 É. Colin de Verdière and J. Erickson. Tightening nonsimple paths and cycles on surfaces. *SIAM J. Comput.*, 39(8):3784–3813, 2010. URL: http://monge.univ-mlv.fr/~colinde/pub/04octagons.pdf, doi:10.1137/090761653.

6 É. Colin de Verdière and F. Lazarus. Optimal pants decompositions and shortest homotopic cycles on an orientable surface. *J. ACM*, 54(4):Article 18, 2007. URL: https://monge.univ-mlv.fr/~colinde/pub/02pants.pdf, doi:10.1145/1255443.1255446.

7 S. E. Dreyfus and R. A. Wagner. The Steiner problem in graphs. *Networks*, 1(3):195–207, 1971. doi:10.1002/net.3230010302.

8 P. Eades and D. Rappaport. The complexity of computing minimum separating polygons. *Patt. Recog. Lett.*, 14(9):715–718, 1993. doi:10.1016/0167-8655(93)90140-9.

9 A. Efrat, S. G. Kobourov, and A. Lubiw. Computing homotopic shortest paths efficiently. *Comput. Geom.*, 35:162–172, 2006. doi:10.1016/j.comgeo.2006.03.003.

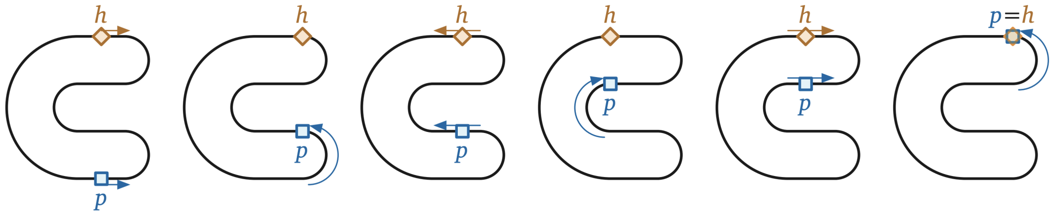


Figure 3 Catching a iron ball with a magnet on a curve.

10 D. Eppstein, M. Overmars, G. Rote, and G. Woeginger. Finding minimum area k-gons. *Discrete Comp. Geom.*, 7:45–58, 1992. doi:10.1007/BF02187823.

11 J. Erickson and P. Worah. Computing the shortest essential cycle. *Discrete Comput. Geom.*, 44(4):912–930, 2010. doi:10.1007/s00454-010-9241-8.

12 D. Marx, M. Pilipczuk, and M. Pilipczuk. On subexponential parameterized algorithms for Steiner tree and directed subset TSP on planar graphs. In *Proc. 59th Ann. IEEE Symp. Foundat. Comput. Sci. (FOCS)*, pages 474–484, 2018. doi:10.1109/FOCS.2018.00052.

4.3 Catching Balls (or Puppies) On Curves of Rotation Number 1

Maarten Löffler (Utrecht University, NL), Florestan Brunck (University of Copenhagen, DK), and Rodrigo I. Silveira (UPC Barcelona Tech, ES)

Main reference Mikkel Abrahamsen, Jeff Erickson, Irina Kostitsyna, Maarten Löffler, Tillmann Miltzow, Jérôme Urhausen, Jordi L. Vermeulen, Giovanni Viglietta: "Chasing Puppies: Mobile Beacon Routing on Closed Curves", in Proc. of the 37th International Symposium on Computational Geometry, SoCG 2021, June 7-11, 2021, Buffalo, NY, USA (Virtual Conference), LIPIcs, Vol. 189, pp. 5:1–5:19, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.
URL https://doi.org/10.4230/LIPICS.SOCG.2021.5

4.3.1 Problem Statement

We are given a smooth closed curve in the plane, described by a function $\gamma : S^1 \to \mathbb{R}^2$, and two points a and b on the curve. Here a is the *attractor* (think of a magnet) and b is a *ball* (think of an iron or nickel ball). Then b always moves closer to a if possible while staying on the curve. We can directly control the position a, but we have no direct control over b. We assume that b moves infinitely faster than a. Our goal is to move a along the curve in such a way that a meets b (see Figure 3). Note that a bit of care is required to properly define this problem and make sure γ is generic enough for the conversation to hold. For more details and a careful description of the setting, see [1, 3.1].

It is known that for every *simple* curve γ (that is, without self-intersections), there always exists strategy to catch the ball from every initial configuration [1].

It is also known that there exist curves for which this is not possible; the simplest such curve is the "doubled circle", and in fact any curve that is "doubled" or "multipled" is an example of such a curve, since a and b will always stay close to each other on different "copies" of the curve. Refer to Figure 4 for some examples; for proper definitions and proofs see [1].

A natural question is then to see whether one can characterise those curves for which there does exist a strategy to catch the ball from every initial configuration. Perhaps the first and simplest invariant of a curve is the *rotation number* (also called the *index* or *Whitney*

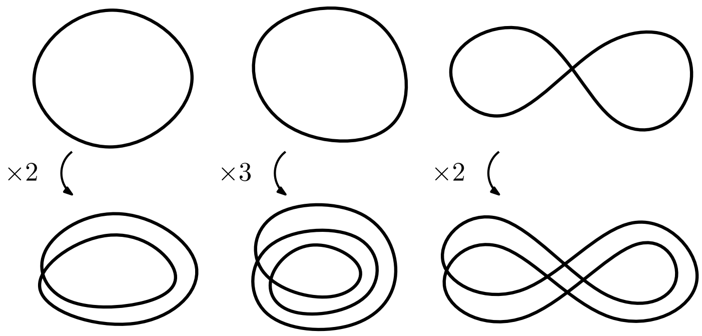


Figure 4 Any curve can by "multiplied" by taking multiple copies of it at (almost) the same location, cutting it at an arbitrary point, and switching endpoints before reconnecting.

index). This is defined as the number of revolutions a tangent vector completes as it traverses the curve once. It is related to the number of left-to-right and the number of right-to-left crossings of the curve. [3]

A question posed in [1] is:

▶ **Question 1.** Is it true that, for every curve of rotation number 1, there always exists a way to catch the ball?

Experimental evidence suggests the answer should be yes. You can play around with an implementation here to try for yourself: `https://github.com/viglietta/Chasing-Puppies`

4.3.2 Progress

Rotation number

We answer Question 1 negatively: there do exist curves of rotation number 1 for which there are starting configurations from which we can never catch the ball.

▶ **Theorem 2.** There exists a curve γ of rotation number 1 and a pair of points (a, b) on γ such that, no matter how a moves on γ, a will never reach b.

Proof (sketch). The idea is to first construct a family of curves $c_1, c_2, \ldots, c_7$ that satisfy some properties:

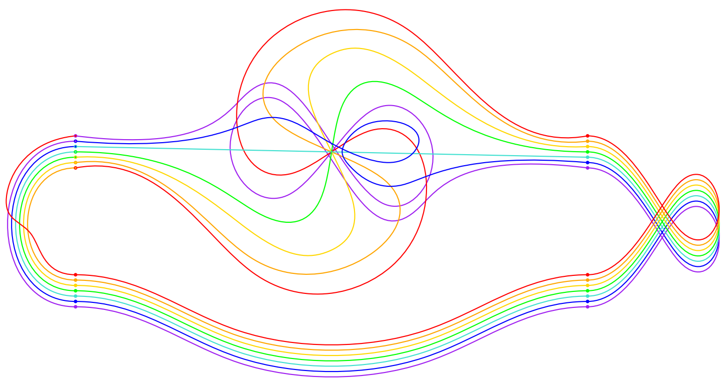
1. All c_i start somewhere at x-coordinate -1 with a tangent vector $(1, 0)$, and end somewhere at x-coordinate 1 with a tangent vector $(1, 0)$.
2. The rest of the curves are contained in the strip $-1 < x < 1$.
3. All except 1 of the curves make a net 0 rotations, while the final curve makes one full rotation.
4. When a is on curve c_i, and b is on curve c_{i+1} (indices wrap around), b will locally follow a as we move from left to right on the curve; in particular, when a is on the left end then so is b and when a is on the right end then so is b.

Then, we connect these 7 curves into a single closed curve by looping the curves around, as in Figure 5. To make sure that the total rotation number is 1, we must also add an additional counter-clockwise loop. ◀

While our counterexample answers the question as posed, the curve used in the proof has 108 self-intersections and is arguable not very close to being simple.

Instead, we can also try to use a more fine-grained classification of curves.

A curve of rotation number 1 (with 108 crossings).



When a is on **purple**, b follows on **blue**.
When a is on **blue**, b follows on **turquoise**.
When a is on **turquoise**, b follows on **green**.
When a is on **green**, b follows on **yellow**.
When a is on **yellow**, b follows on **orange**.
When a is on **orange**, b follows on **red**.
When a is on **red**, b follows on **purple**.

Figure 5 A single curve of rotation number 1, for which there exist configurations from which the ball can never be caught; for example, start with a on the the lowest point of the curve (on the purple strand) and b just above it (on the blue strand).

Curve classifications

Working towards the goal of characterising which curves admit winning strategies for all starting configurations, it is natural to wonder which characteristics and invariants of a curve might have repercussions on this problem. While the rotation number is a fundamental invariant of plane curves, it is also quite coarse: it merely characterises the connected components in the space of immersions (see [4]). It is thus perhaps not so surprising that two curves sharing that invariant would behave very differently regarding our problem. In fact, the classification of curves based on the rotation number is up to homotopy in the class of closed curves, but not up to diffeomorphisms of the whole plane and therefore ignores all about how the curves are situated in the plane. See Figure 6 for an example of two very different curves with the same rotation number.

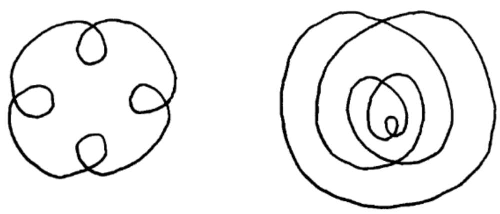


Figure 6 Two curves with the same rotation number but very different geometries.

Various finer classifications and invariants have been introduced to address this. Most notably, Arnold's three invariants classify generic curves in much finer details ([2]; see Figure 7). Together with his use of Gauss diagrams (see [2, p28]), Arnold's classification for a fixed number of crossings now differentiates many curves previously bundled in the same class and might help provide finer insights into the behaviour of each class of curve with respect to our problem.

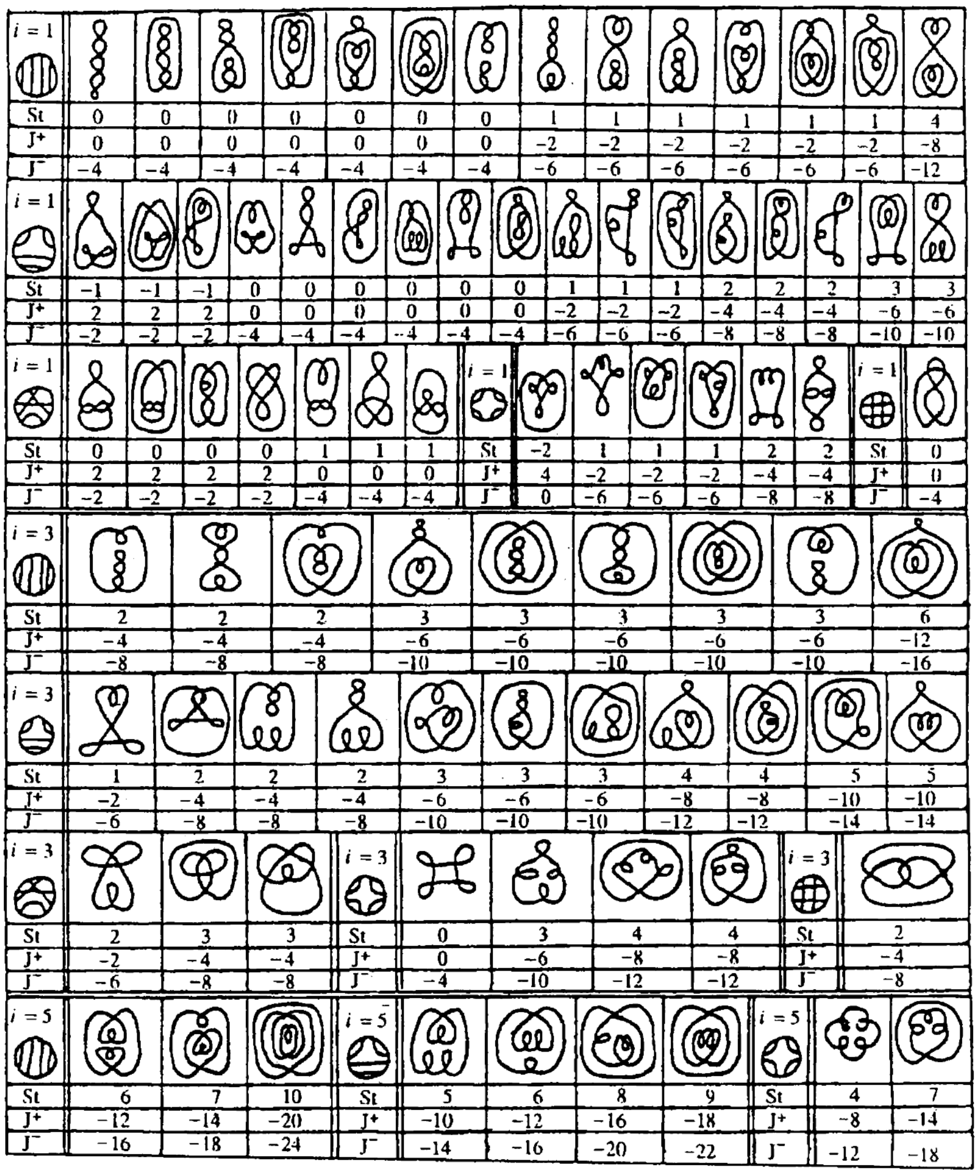


Figure 7 A classification of curves with crossing number 4 based on Arnold's three invariants, J^+, J^-, St, the index and the Gauss diagrams; image taken from [2].

Regardless of our preferred choice of curve classification, we can study the behaviour of the problem at hand on the different members of each class.

A priori, there are three possibilities for each class:

1. for every curve in the class, and for every initial configuration on the curve, there exists a strategy to catch the ball; or
2. there is at least one curve in the class such that for every initial configuration there is a strategy to catch the ball, and at least one curve in the class for which there exists a starting position from which we cannot catch the ball; or
3. for every curve in the class, there is a starting position from which we cannot catch the ball.

However, it is easy to see that the third option cannot occur.

▶ **Lemma 3.** In every class, there is at least one curve such that for every starting position we can catch the ball.

Proof (sketch). Take any (geometric) curve in the class. Now find a point on the convex hull, and make a small opening. Attach a huge loop to this opening (radius at least twice the diameter of the original curve).

We claim that on this adapted curve, we can catch the ball by simply keeping walking in one direction. The reason is that, while we are in the original curve, the ball will never go "around" the big loop, since the farthest point on the loop is always too far away. So, we might walk around the big loop once while the ball is somewhere in the original curve, but by the time we enter the big loop for the second time, we must have encountered the ball along the way. ◀

Because of this observation, there are only two possible states for each class: "always yes" or "depends on the geometry".

To conclude this report, we believe it would be important to understand this classification in more detail; as a small first step towards this goal we present a table with the known behaviours of a few classes of curves (grouped in a coarser 2-parameter table with respect to index and crossing number) in Figure 8.

References

1 Mikkel Abrahamsen, Jeff Erickson, Irina Kostitsyna, Maarten Löffler, Tillmann Miltzow, Jérôme Urhausen, Jordi Vermeulen, and Giovanni Viglietta. *Chasing Puppies: Mobile Beacon Routing on Closed Curves*. Journal of Computational Geometry, vol 13(2):115–150, 2022.

2 Vladimir I. Arnold. *Plane curves, their invariants, Perestroikas, and classifications*. `http://www.pdmi.ras.ru/~arnsem/Arnold/Arnold-PC93.pdf`, 1994.

3 Damián Wesenberg. *A new formula for rotation number*. arXiv:2010.01422, 2020.

4 Hassler Whitney. *On regular closed curves in the plane*. Compositio Mathematica, vol 4:276–284, 1937.

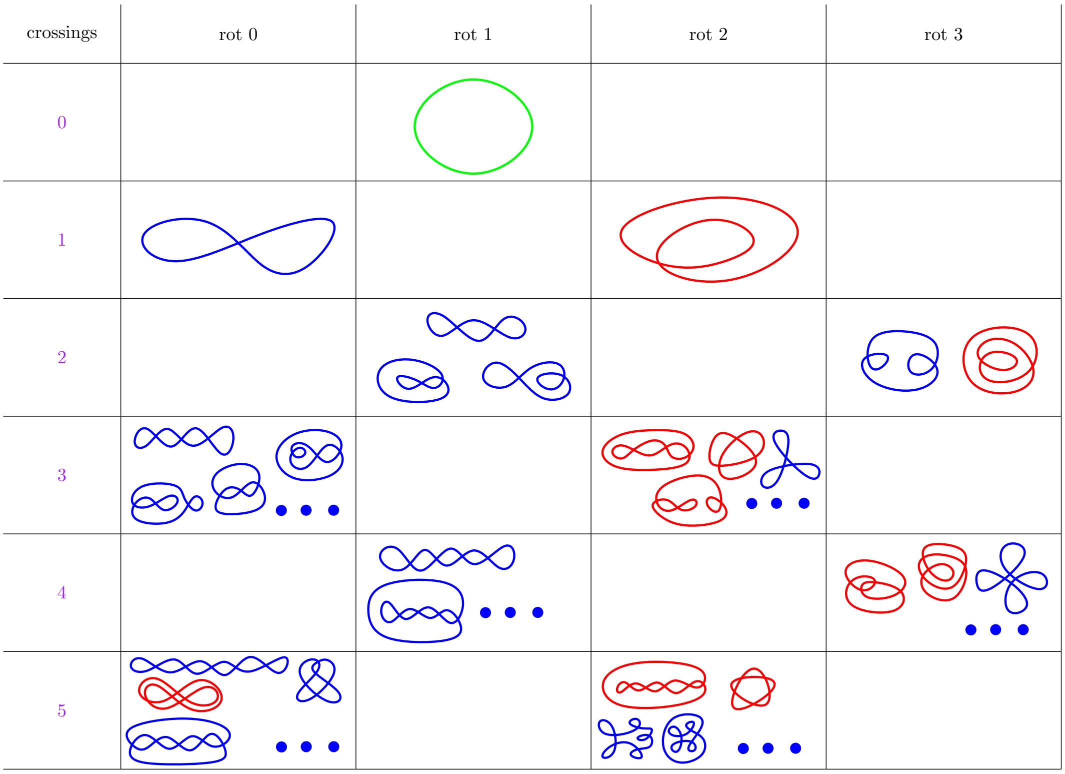


Figure 8 What we know. For green curve classes, the answer is always yes; for red curve classes, there are known counter-examples. For blue classes, we do not know. Blue dots mean that we did not draw all curves in the cell. Note that the curve in Figure 5 would also be red in this table, and would sit at coordinates (1, 108).

4.4 A practical implementation of the link thickening problem

Clément Maria (INRIA – Sophia Antipolis, FR) and Benjamin Burton (The University of Queensland – Brisbane, AU)

We give a practical solution to the *link thickening* problem: Given a collection of embedded closed curves L in $\mathbb{R}^3$, compute the maximal value $r > 0$ such that the r-thickening of L and L have the same homotopy type.

By r-thickening, we mean the neighborhood of L in $\mathbb{R}^3$, made of all points at distance at most r from L. We call r^* the supremum of r for which the r-thickening of L and L have same homotopy type.

The main goal is to implement an experimentally fast rendering algorithm to visualize knots and links in the `Regina` [3] library. Our main constraint is that our solution to the link thickening problem should be close to real time in practice.

4.4.1 Persistent homology approach

Persistent homology [4] studies the evolution of the homology of the level-sets

$$\left(f^{-1}((-\infty; r])\right)_{r\in\mathbb{R}}$$

of a continuous function $f\colon X \to \mathbb{R}$ on a topological space X. A most studied situation is when the space X is the Euclidean space $\mathbb{R}^d$, and the function f is the Euclidean distance function to some compact subset of $\mathbb{R}^d$. Available implementations, such as in the `Gudhi` library [5], essentially work with finite point set approximations of compacts.

In our framework, the ambient space is the 3-dimensional Euclidean space $\mathbb{R}^3$, the compact subset is the collection of closed polygonal curves L forming the link, and an r-*thickening* of L is the level set $f^{-1}((-\infty; r])$ of the function f assigning to any point $x \in \mathbb{R}^3$ the distance from x to L.

We propose a heuristic solution to the *link thickening* problem with good practical performance, based on persistent homology. For implementation reasons, it is easier to work with neighborhoods of point clouds, instead of polygonal curves. We start by subsampling the polygonal curves of L with a collection of regularly spaced points $P \subset L$. Let ϵ be the *density* of P in L, *i.e.*, the maximal distance between L and P:

$$\varepsilon := \sup_{x\in L} \inf_{p\in P} \mathrm{d}(x, p).$$

Heuristically, we assume $0 < \varepsilon \ll r^*$ by subsampling sufficiently.

Call $\mathrm{d}_P : \mathbb{R}^3 \to \mathbb{R}$ the distance function to the point cloud. Let n be the cardinality of P, and k the number of connected components of the link L.

In practice, the topology of the level sets $\mathrm{d}_P^{-1}((-\infty; r])$, consisting of the union of balls of radius r centered at the points in P, follows several phases for $r > 0$:

Phase 1 for $0 < r < \varepsilon$, the level set $\mathrm{d}_P^{-1}((-\infty; r])$ is a collection of strictly more than k connected components,

Phase 2 for $\varepsilon < r < r^*$, the level set $\mathrm{d}_P^{-1}((-\infty; r])$ has the homotopy type of L,

Phase 3 for $r^* < r$, the level set $\mathrm{d}_P^{-1}((-\infty; r])$ does not have the homotopy type of L.

Intuitively, the phase $0 < r < \varepsilon$ consists of n disjoint radius r-balls centered at the points in P when r is close to 0, that eventually merge into a regular neighborhood of the link L, once $\varepsilon < r < r^*$.

For $\varepsilon < r < r^*$, and under valid density assumptions [1], the level set $\mathrm{d}_P^{-1}((-\infty; r])$, or equivalently the union of balls of radius r centered at the points of P, is guaranteed to have the same homotopy type as the link L, *i.e.*,, a collection of k circles.

When passing the critical value $r = r^*$, the topology of $\mathrm{d}_P^{-1}((-\infty; r])$ changes by definition, which happens in three possible ways: generically,

Case 1 either one component K of the link L fills up, in which case a solid torus neighborhood of K turns into a 3-ball,

Case 2 or a component K of the link L self intersects, in which case a solid torus neighborhood of K turns into a genus 2 handlebody,

Case 3 or two distinct components K_1, K_2 of the link L connect to form a genus 2 handlebody.

All three cases are captured by the homology of $\mathrm{d}_P^{-1}((-\infty; r])$. Indeed, in case 1, the dimension of the first homology group $H_1(\mathrm{d}_P^{-1}((-\infty; r]), \mathbb{Z}_2)$ is reduced by 1. In case 2, the dimension of $H_1(\mathrm{d}_P^{-1}((-\infty; r]), \mathbb{Z}_2)$ increases by 1. Finally, in case 3, the number of connected components, which is equal to the dimension of the homology group $H_0(\mathrm{d}_P^{-1}((-\infty; r]), \mathbb{Z}_2)$, decreases by 1.

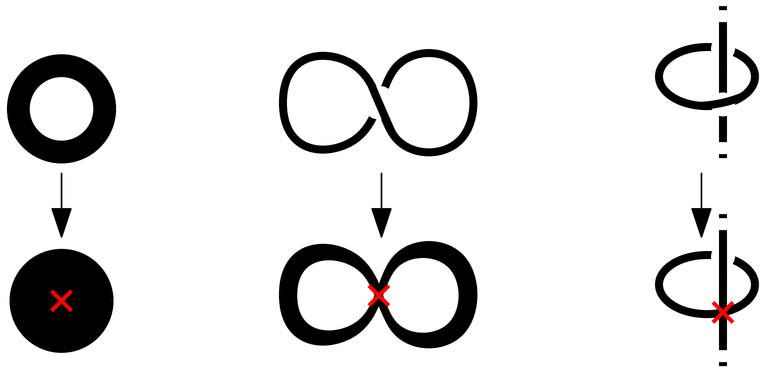


Figure 9 Three possible changes of homotopy type at $r = r^*$; the change of topology happens at the red crosses. From left to right: a solid torus fills up into a 3-ball, a solid torus self intersects into a genus 2 handlebody, two distinct solid tori intersect to form a genus 2 handlebody.

All cases are readily available from the persistent diagrams of the level sets of the function d_P^{-1}.

4.4.2 Implementation with `Gudhi`

The *Cech complex* based on P and of threshold r is a simplicial complex which has the same homotopy type as the level set $\mathrm{d}_P^{-1}((-\infty; r])$. However, it is challenging and relatively slow to implement its construction, in comparison to the *Rips complex* [4]. Rips complexes can be interpreted as an approximation of the Cech complex, and are generally preferred in software implementation. When their construction is extremely fast [2], they result into noisier persistence diagrams. However, in our context where P is a noiseless subsample of a link in $\mathbb{R}^3$, we do not observe experimentally significant differences in the persistence diagrams of the two filtered complexes.

We interface the `Gudhi` library with `Regina` to implement the algorithm above. For efficiency reasons, we stop the computation when the change of homology at $r = r^*$ is detected. First experiments show a close to real time performance on simple knots and links. See Figure 10 for an example of rendering.

References

1. Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and Topological Inference*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2018.
2. Jean-Daniel Boissonnat and Clément Maria. The simplex tree: An efficient data structure for general simplicial complexes. *Algorithmica*, 70(3):406–427, 2014.
3. Benjamin A. Burton, Ryan Budney, William Pettersson, et al. Regina: Software for low-dimensional topology. `http://regina-normal.github.io/`, 1999–2023.
4. Herbert Edelsbrunner and John Harer. *Computational Topology – an Introduction*. American Mathematical Society, 2010.
5. Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The Gudhi library: Simplicial complexes and persistent homology. In Hoon Hong and Chee Yap, editors, *Mathematical Software - ICMS 2014 – 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings*, volume 8592 of *Lecture Notes in Computer Science*, pages 167–174. Springer, 2014.

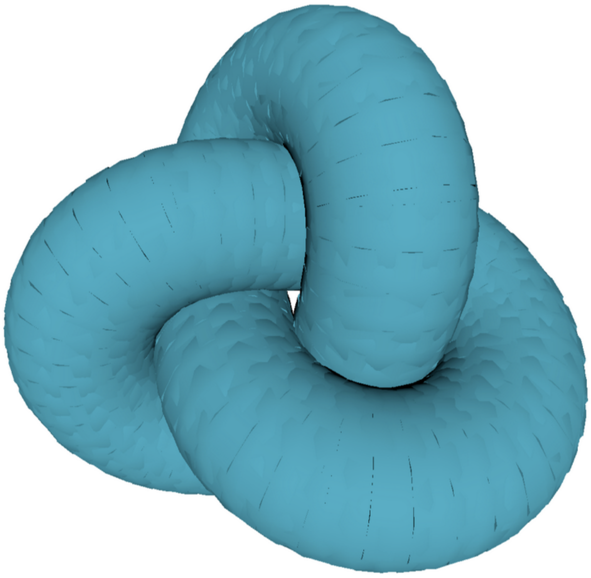


Figure 10 Trefoil knot, sampled, and thickened up to r^*.

4.5 Complexity of Reconfiguration Problems

Lena Schlipf (Universität Tübingen, DE), Mikkel Abrahamsen (University of Copenhagen, DK), Kevin Buchin (TU Dortmund, DE), Maike Buchin (Ruhr-Universität Bochum, DE), Linda Kleist (TU Braunschweig, DE), Maarten Löffler (Utrecht University, NL), and André Schulz (FernUniversität in Hagen, DE)

Numerous NP-hardness proofs for polygons use a reduction from (a version of) PLANAR 3SAT. An instance ϕ of PLANAR 3SAT (or a version thereof) is given and a polygon P is build on top of ϕ, but then the bounded faces of ϕ turn (most of the times) into holes in the polygon. Thus, most proofs work only for polygons with holes. However, there are quite some problems that seem to be hard even for simple polygons. But for proving their hardness new ideas are needed. New insights and techniques for such hardness proofs for simple polygons were recently presented by Abrahamsen and Stade [1]. They show that packing unit squares in a simple polygon is NP-hard by a reduction from MONOTONE PLANAR 3SAT. This working group focuses on how to extend their result to other similar problems.

The first problem that we investigate is motion planning for unit square robots. Given a set of unit squares (robots) in a workspace and a set of target positions, the goal is to move the robots such that (i) each target position is occupied by some robot (*unlabeled* version), or (ii) each target position is occupied by a specific robot (*labeled* version). If the workspace is a domain with obstacles, the problem is known to be PSPACE-hard for both variants, labeled and unlabeled [3, 7]. Building on the result of [1], we could show the following.

▶ **Theorem 1.** Motion planning for labeled unit square robots in a simple polygon is NP-hard.

Proof-Sketch. The proof follows quite easily from the NP-hardness result of packing unit squares in a simple polygon [1]. The basic idea is to add a small tunnel to the polygon of a packing instance. The robots are then initially placed in the tunnel on the target positions

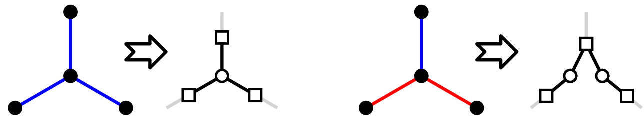


Figure 11 Mapping an NCL-Instance to an embedding of a planar 3-SAT formula. Variable nodes are drawn as squares, Clause nodes are drawn as disks.

but need to be reordered. The robots can only be reordered if all of them move into the original polygon and then move back into the tunnel in the required reordered manner. Hence, the reordering is only possible if the robots can be packed into the original polygon. ◀

We also investigate PSPACE-hardness. Many SAT reconfiguration problems are known to be PSPACE-complete: Given a SAT formula and two satisfying assignments, the question is whether it is possible to find a sequence of flips that takes us from one assignment to the other where a flip allows to change the value of exactly one variable. This is equivalent to the problem of deciding whether the two satisfying assignments are connected in the flip graph of the given SAT formula. We denote this problem by SAT-Reconfiguration; and analogously, we denote the problem for other SAT variants.

The proof for Theorem 1 uses a reduction from monotone planar 3SAT, thus, using the reconfiguration problem of the same version of 3SAT seems like a good choice for proving PSPACE-hardness of the problem of Theorem 1. It was not known whether the reconfiguration problem of monotone planar 3SAT is PSPACE-complete, hence, we first show the following result.

▶ **Theorem 2.** Deciding whether two satisfying assignments are connected in the flip graph of a monotone planar 3SAT formula is PSPACE-complete.

Proof-Sketch. The proof is a reduction from Nondeterministic Constrained Logic (NCL) [5, 6]. An NCL instance is given by a directed planar graph G and an edge e_r of G. The edges are colored blue or red, where every vertex is either incident to three blue edges or to one blue and two red edges. A feasible orientation must respect an indegree of at least two at every vertex, where blue edges count with multiplicity 2. An NCL-instance is positive, if and only if there is a sequence of edge-reversals that finally reverses e_r and any intermediate orientation is feasible.

We first describe a mapping from G to a planar 3-SAT formula ϕ with embedding by substituting every vertex with incident half edges as shown in Figure 11. For every oriented edge $e = uv$ in G we introduce one variable x_e in the formula. This variable appears as x_e at the clause nodes induced by v and as $\neg x_e$ at the clause nodes induced by u. The truth assignment of the formula decodes, which edges are reversed (false variables are the reversed edges). The reversal of an edge in G is represented by flipping a variable.

It remains to show that ϕ can be turned into a monotone planar 3-SAT formula; for an illustration, consider Figure 12. In a first step, we subdivide the edges in G and replace every edge $e = uv$ by edges uw and wv. We then substitute u and v as before, but the variable node at u is called x_{e-} and the (now distinct) variable node at v is now called x_{e+}. The new vertex w will be associated with a clause node $(\neg x_{e+} \vee \neg x_{e-})$. In the clauses defined by the original nodes of G all variables are nonnegated.

We can now derive a truth assignment from the orientation of G as follows: If the edge $e = uv$ is not reversed in G, we set x_{e-} = false and x_{e+} = true. If the edge $e = uv$ is reversed in G we set x_{e-} = true and x_{e+} = false. In order to get an orientation from a satisfying truth

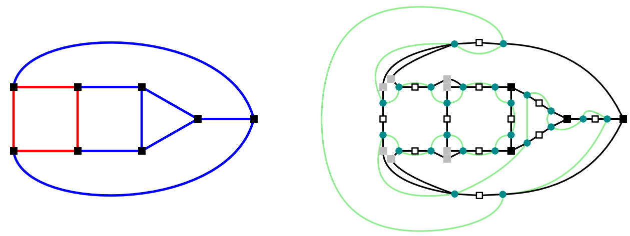


Figure 12 Mapping between a graph G of an NCL instance (left) and an embedded monotone 3-SAT formula ϕ (right).

Table 1 Overview of computational complexity for the satisfying and reconfiguration problems of important SAT variants.

	SATISFYING	RECONFIGURATION
3SAT	NP-complete	PSPACE-complete
PLANAR 3SAT	NP-complete	PSPACE-complete
MON. PLANAR NOT-ALL-EQUAL 3SAT	P	PSPACE-complete
MON. PLANAR 3SAT	NP-complete	PSPACE-complete
POSITIVE 1-IN-3SAT	NP-complete	P
POSITIVE NOT-ALL-EQUAL 3SAT	NP-complete	PSPACE-complete

assignment we use the same mapping but we could also have x_{e+} = false and x_{e-} = false as a satisfying assignment for ϕ. In this case the edge e would not contribute to the indegree constraint at neither u nor v in G. Thus, we can orient e arbitrarily, say in the original orientation.

Our construction yields faces (for the embedding ϕ) of even degree. Hence the dual graph is Eulerian and supports an Euler tour. This tour can be selected such that it can be shortcutted to a cycle that visits all variable nodes without crossing any edge of the embedding of ϕ (see Figure 12(b)). ◀

We believe that using Theorem 2 and exploiting the ideas of Abrahamsen and Stade [1] leads to a PSPACE-hardness proof of the problem of Theorem 1.

In Table 1, we give a summary of some variants of 3SAT and the complexity of their satisfying and reconfiguration problems. Interestingly, most satisfying and reconfiguration problems are hard, but there are some surprising exceptions.

Additionally, we consider the motion planning problem for unit disk robots. Recently, it was proven that the unlabeled motion planning problem for disk robots with two different radii in a domain with obstacles is PSPACE-hard [2]. We turn our attention to unit disk robots and have some ideas of showing that the problem remains PSPACE-hard.

▶ **Theorem 3.** Reconfiguration of unit disk robots in an arc-gon/polygon with holes is PSPACE-complete.

Proof-Sketch. We reduce from Constrained Logic. For this we need gadgets to represent two types of vertices (AND and OR gadgets), and edge gadgets that connect them.

Our construction closely follows Brocken et al. [2]; that is, vertices are represented by rooms and edges are represented by tunnels that are filled with disks. The main difference is that we require a gadget that ensures an edge can only have one orientation at a time.

Our new edge gadget is illustrated in Figure 13. Note that the tunnel is shaped (i.e. has small dents) such that the middle three discs cannot move too far out. We claim that:

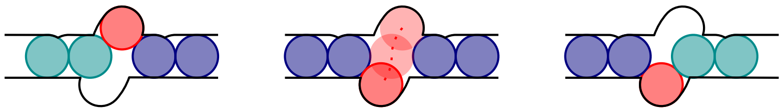


Figure 13 A sequence of moves that imitates the reorientation of an edge.

▶ **Claim.** An edge gadget can have all disks "pushed towards the right", or all disks "pushed towards the left", or it can have the left half pushed towards the left and the right half pushed towards the right; this is required as an intermediate state when we reconfigure from one state to the other (this corresponds to reorienting the edge in the constraint logic graph). But it is not possible to have all disks in the middle.

The AND and OR gadgets are similar to the ones from Brocken et al. [2]. ◀

References

1 Mikkel Abrahamsen and Jack Stade. Hardness of Packing, Covering and Partitioning Simple Polygons with Unit Squares. In preparation.

2 Thomas Brocken, G. Wessel van der Heijden, Irina Kostitsyna, Lloyd E. Lo-Wong, and Remco J. A. Surtel. Multi-Robot Motion Planning of k-Colored Discs Is PSPACE-Hard. 10th International Conference on Fun with Algorithms, FUN: 15:1–15:16

3 Josh Brunner, Lily Chung, Erik D. Demaine, Dylan H. Hendrickson, Adam Hesterberg, Adam Suhl, and Avi Zeff. 1 X 1 Rush Hour with Fixed Blocks Is PSPACE-Complete. 10th International Conference on Fun with Algorithms, FUN: 7:1–7:14, 2021.

4 Parikshit Gopalan, Phokion G. Kolaitis, Elitza Maneva, and Christos H. Papadimitriou. The Connectivity of Boolean Satisfiability: Computational and Structural Dichotomies. SIAM Journal of Computing 38(6):2330–2355 (2009).

5 Robert A. Hearn and Erik D. Demaine. The nondeterministic constraint logic model of computation: Reductions and applications. Automata, Languages and Programming, 29th International Colloquium, ICALP 2002: 401–413.

6 Robert A. Hearn and Erik D. Demaine. Games, puzzles, and computation. CRC Press, 2009.

7 Kiril Solovey and Dan Halperin. On the hardness of unlabeled multi-robot motion planning. Int. J. Robotics Res. 35(14): 1750–1759 (2016).

4.6 Complexity of Testing 0-Efficiency

Eric Sedgwick (DePaul University – Chicago, US), Benjamin Burton (The University of Queensland – Brisbane, AU), Alex He (The University of Queensland – Brisbane, AU), Clément Maria (INRIA – Sophia Antipolis, FR), Tim Ophelders (Utrecht University, NL), Saul Schleimer (University of Warwick – Coventry, GB), Jonathan Spreer (University of Sydney, AU), and Stephan Tillmann (University of Sydney, AU)

This working group was formed to discuss the complexity of testing the 0-efficiency of a 3-manifold triangulation. This problem appears to be quite challenging, and the group ended up discussing, and making progress on a number of related problems.

Each of these problems concern *normal surface theory*, a frequently used technique for 3-manifold decision problems. In normal surface theory, a 3-manifold is given via a triangulation, and conclusions about the manifold can be drawn by understanding the *normal surfaces*, those properly embedded surfaces that intersect each tetrahedron in triangles and/or quadrilaterals.

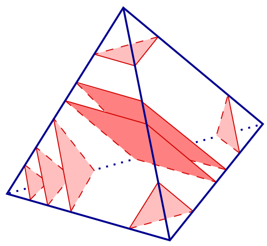


Figure 14 A normal surface meets each tetrahedron in triangles and/or quadrilaterals.

4.6.1 0-Efficiency, Non-vertex linking normal surfaces

A connected normal surface is *vertex linking* if it contains only triangles. In a closed 3-manifold, a vertex linking surface will be sphere that is the boundary of a regular neighborhood of some vertex of the triangulation. Each vertex of a triangulation has a vertex linking sphere surrounding it, are there any other normal spheres? A 3-manifold triangulation is *0-efficient* if there are none, i.e., all of its normal spheres are vertex linking. Almost all 3-manifolds have a 0-efficient triangulation and working with one yields significant computational advantages [5, 2].

The working group was formed to consider the computational complexity of the following problem:

▶ **Problem 1.** *Given a 3-manifold triangulation, is it 0-efficient?*

Alternatively, we can consider the complementary problem:

▶ **Problem 2.** *Given a 3-manifold triangulation, does it contain a non-vertex linking 3-sphere?*

The work of Hass, Lagarias, and Pippenger [4] shows that if there is a non-vertex linking normal sphere, then there is one with an (exponentially) bounded number of triangles and quadrilaterals. This determines a coordinate vector with linear bit-length which is sufficient to certify that the underlying normal surface is a non-vertex linking sphere. It follows that Problem 2 is in **NP** (and Problem 1 is in **co-NP**).

The working group endeavored to show that Problem 2 is also **NP-hard**, primarily by demonstrating a reduction from a problem such as **3-SAT**. One possible model is the work of Agol, Hass and Thurston [1], who used a reduction from **3-SAT** to show that the problem **3-Manifold Knot Genus** is **NP-hard**. Unfortunately, it seems quite challenging to adapt their approach to Problem 2. In particular, their reduction yields surfaces with large genus, whereas Problem 2 would require a reduction that always produced a sphere.

The group also considered various analogues and/or variations on Problems 1 and 2, see the sections below. Towards the end of the group's meetings, a proposal was made to weaken Problem 2 to allow surfaces of higher genus:

▶ **Problem 3.** *Given a 3-manifold triangulation, does it contain a normal surface that is not subcomplex linking?*

A *subcomplex linking* surface is a normal surface that links a subcomplex of the triangulation. Vertex linking spheres are the simplest case, and always exist, but it is also possible that a normal surface links an edge or even a larger subcomplex of the triangulation.

It seems possible that Problem 3 is **NP-hard**. This appears easier to approach than showing that Problem 2 is **NP-hard**.

4.6.2 Composite Knot Recognition

The working group also considered the computational complexity of deciding whether a given knot is composite. In particular, we believe that the following problem is in **NP**:

▶ **Problem 4.** *Given a knot $K \subset S^3$, is K composite, i.e., can K be expressed as a sum of non-trivial knots $K = K_1 \# K_2$?*

Typically a knot is specified by providing a *knot diagram*, a projection of the knot into the plane, with its over and under crossings indicated. A few observations are helpful:

1. A composite knot is characterized by the existence of an *decomposing annulus*, a meridional essential annulus that is properly embedded in the *knot exterior*, $X = S^3 - \eta(K)$, where $\eta(K)$ is an open tubular neighborhood of K. The knot exterior is a compact manifold with a single torus boundary component. Cutting the exterior X along the decomposing annulus, cuts the knot exterior into the exteriors of the knots K_1 and K_2.
2. Given a diagram of a knot drawn in the plane with c crossings, it is straightforward to produce a triangulation of the knot exterior with at most $O(c)$ tetrahedra [4].

Using these observations we can recognize a composite knot by decomposing along this annulus. In particular, we need to

1. Produce the decomposing annulus A, and
2. Show that when X is cut along A, each of the resulting components is a non-trivial knot exterior.

We may assume that the knot exterior X is given by a 0-efficient triangulation. Then, some decomposing annulus A is realized as a fundamental normal surface. In particular, this means that its coordinates (number of triangles and quadrilaterals) are bounded exponentially and the bit size of the coordinate vector is linear in the number tetrahedra. The coordinate vector of this annulus will serve as the **NP** certificate. Indeed, using the coordinate vector, we can verify that it represents an incompressible annulus.

It remains to show that when X is cut along A, the two knot exteriors produced are **non-trivial** (equivalently, that the annulus A is **not** peripheral in X). For this we can use Lackenby's result that the unknotting problem is in **co-NP** [8]. Unfortunately, naively cutting X along A will not be sufficient as the resulting triangulation is only exponentially bounded. Three alternatives were proposed to mitigate this complication:

1. Cut in a smarter way so that the resulting triangulation (or handle structure) has polynomial (linear?) complexity in terms of the original.
2. Crush X along the annulus A instead of cutting along it [5, 2]. This is equivalent to Dehn filling the meridian (producing S^3) and then crushing along a capped off sphere. This reduces the number of tetrahedra, but we would need to prove that knowledge of the summands can be tracked through the crushing procedure.
3. Work with a spun normal annulus in an ideal triangulation. We would need to show that a peripheral annulus would not occur as a spun normal annulus.

Each of these approaches holds promise and would require working out some significant details. Regardless, this result appears to be the most attainable considered by the group.

4.6.3 Ribbon Knot Recognition

The working group also discussed the complexity of ribbon knot recognition with the aim of showing that this problem is in **NP** for small knots:

▶ **Problem 5.** *Given a small knot $K \subset S^3$, is K a ribbon knot?*

A *small* knot is a knot with no closed essential surfaces in its exterior. Every knot $K \subset S^3$ bounds an immersed disk. A knot K is a *ribbon knot*, if it bounds a *ribbon disk* an immersed disk D with only ribbon singularities. A *ribbon singularity* is an arc of self intersection whose pre-image in the disk consist of two arcs, one arc in the interior of the disk (a slit) and one arc with both endpoints in the boundary of the disk (the ribbon passing through the slit).

Given a ribbon knot, choose a ribbon disk that minimizes the number of ribbon singularities. Each of the, say n, ribbon singularities can be desingularized, in two distinct ways, by a cut and paste of the two sheets that meet the singularity. Resolving all n singularities produces a Seifert surface for the knot with Euler characteristic $1 - 2n$.

▶ **Question 6.** *Among the 2^n choices for desingularization, is there always one that produces an incompressible surface?*

While there are clear cases where bad choices produce a compressible surface, an affirmative answer yields an approach to ribbon knot recognition.

Suppose that there is a desingularization producing an incompressible surface F. When the knot K is small, F will be realized by a fundamental normal surface. (If F is a normal sum, then one of the summands is a closed essential surface contradicting that K is small [7, 6]).

The goal would then be to exhibit an **NP** certificate consisting of the fundamental normal surface F along with a collection of n disks in its complement that demonstrate how to resingularize the surface to produce a ribbon disk for K.

Currently, resolving Question 6 is the main obstacle to this approach. A first step could be gathering evidence using Burton's *Regina* [3].

References

1 Ian Agol, Joel Hass, and William Thurston. 3-manifold knot genus is NP-complete. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 761–766. ACM, New York, 2002.

2 Benjamin Burton. A new approach to crushing 3-manifold triangulations. *Discrete Comput. Geom.*, 51(4):864–892, 2014.

3 Benjamin A. Burton. Introducing Regina, The 3-Manifold Topology Software. *Experimental Mathematics*, 13(3):267 – 272, 2004.
4 Joel Hass, Jeffrey C. Lagarias, and Nicholas Pippenger. The computational complexity of knot and link problems. *Journal of the ACM (JACM)*, 46(2):185–211, 1999.
5 William Jaco and J. Hyam Rubinstein. 0-efficient triangulations of 3-manifolds. *J. Differential Geom.*, 65(1):61–168, 2003.
6 William Jaco and Eric Sedgwick. Decision problems in the space of Dehn fillings. *Topology*, 42(4):845–906, 2003.
7 William Jaco and Jeffrey L. Tollefson. Algorithms for the complete decomposition of a closed 3-manifold. *Illinois Journal of Mathematics*, 39(3):358 – 406, 1995.
8 Marc Lackenby. The efficient certification of knottedness and Thurston norm. *Advances in Mathematics*, 387:107796, 2021.

4.7 Two problems on flip distances

Jonathan Spreer (University of Sydney, AU), Jean Cardinal (ULB – Brussels, BE), Francis Lazarus (CNRS – Grenoble, FR), Lionel Pournin (Université Paris 13 – Villetaneuse, FR), Günter Rote (FU Berlin, DE), Saul Schleimer (University of Warwick – Coventry, GB), and Birgit Vogtenhuber (TU Graz, AT)

These are the notes of the working group looking into flip distances. They contain two initial findings. The first one is a family of pairs of triangulations of the convex $2k$-gon, where k is at least 5, that produces flip paths of worst possible length between them for a natural heuristic. The second one consists of upper and lower bounds on the flip distances in the modular flip graph of the annulus with arbitrary numbers of marked points on its boundary components.

4.7.1 Introduction

Flip graphs – graphs whose vertices are the triangulations of a geometric or topological space and whose edges correspond to a local operation between them called a flip – appear naturally in computational geometry [15, 20, 24] and in low-dimensional topology [2, 9, 17]. When the triangulated space is a convex polygon, that graph is also relevant to algebraic combinatorics since in that case, it is precisely the edge graph of the associahedron [14], a polytope that has been generalized in a number of ways [5, 10, 11, 21].

The geometry of flip graphs has attracted considerable attention because it is intimately connected to the complexity of balancing binary trees [27], provides a dissimilarity measure between these trees [7], serves as a combinatorial model for the mapping class group of a surface [9, 17], or can be used as a tool to optimize triangulations [1, 3, 4]. The diameter of the associahedron is known exactly [23], but the complexity of computing the distance between two of its vertices is still an open problem [15, 20].

This note has two parts. In Section 4.7.2, we discuss two natural heuristics to find short paths between triangulations of the convex n-gon and show how they are linked. The heuristics use what is called shadow vertex paths and crossing-monotone flip sequences. We then present a family of pairs of triangulations for which one of these heuristics performs worst possible, up to a constant factor, and discuss implications for the other heuristics. In

Section 4.7.3 we prove upper and lower bounds on flip distances in the modular flip graph of triangulations of the annulus with n_1 and n_2 marked points on its boundary components, $n = n_1 + n_2$. More specifically, we prove that the diameter of this flip graph must lie between $\frac{9}{4}n$ and $\frac{10}{4}n$, see Lemmas 9 and 10.

4.7.2 Negative results for flip path heuristics

In this section, we consider the flip graph $\mathcal{F}(P)$ of a Euclidean convex n-gon P. A triangulation of P is a set of pairwise non-crossing arcs between the vertices of P that decompose P into triangles. In this context, an arc is just a straight line segment. The vertices of $\mathcal{F}(P)$ are the triangulations of P and its edges link and pair of triangulations that differ by exactly one arc. It is well-known that $\mathcal{F}(P)$ is the edge graph of the $(n-3)$-dimensional associahedron [14].

We consider two natural heuristics for computing short flip sequences between two given triangulations $\mathcal{T}_1$ and $\mathcal{T}_2$ of a convex n-gon:

1. *Shadow vertex paths* are obtained by projecting the associahedron onto a plane. The two directions defining the projection (the shadow) are chosen such that vertices corresponding to $\mathcal{T}_1$ and $\mathcal{T}_2$ are both extremal. Then, the projection is used to describe a sequence of flips turning $\mathcal{T}_1$ into $\mathcal{T}_2$. The resulting flip path depends on the chosen realization of the associahedron and on the choice of directions for $\mathcal{T}_1$ and $\mathcal{T}_2$ in their respective normal cones. A natural choice of realization of the associahedron is the realization as a secondary polytope for a regular n-gon.
2. Superimpose $\mathcal{T}_1$ and $\mathcal{T}_2$. A *crossing-monotone path* is a sequence of flips, strictly reducing the number of crossing pairs of arcs between $\mathcal{T}_2$ and the intermediate triangulations obtained from $\mathcal{T}_1$ at each step.

We show how these two heuristics are linked, obtain negative results for the first heuristic, and discuss the implications on the second. More precisely, we present pairs of triangulations, where the path constructed by the first heuristic is of length $O(n^2)$. This path realizes the worst case asymptotics and is much longer than the actual shortest path of length $O(n)$.

4.7.2.1 Crossing-monotone paths

It is known that crossing-monotone paths exist for all pairs of triangulations of a set of points in $\mathbb{R}^2$ [12]. The same holds for topological surfaces [2, 9, 16]. In this section we give a short proof of this fact in the special case of a convex n-gon. Given are two triangulations $\mathcal{T}_1$ and $\mathcal{T}_2$ of the convex n-gon. We start with $\mathcal{T}_1$ and flip edges until we reach the target triangulation $\mathcal{T}_2$. We refer to $\mathcal{T}_1$ and all intermediate triangulations as the *red triangulation* and to $\mathcal{T}_2$ as the *blue triangulation*.

We start by superimposing $\mathcal{T}_1$ and $\mathcal{T}_2$. We enumerate the set of red edges with the maximum number of blue intersections. Each of these edges splits the n-gon into two pieces. We call such an edge *outermost*, if one of the two pieces does not contain any red edge with maximum number of intersections.

▶ **Lemma 1.** Flipping an outermost red edge decreases the number of intersections between the red and blue triangulations by at least one.

Proof. Let e be an outermost red edge with maximum number of blue intersections, and let B and D be its endpoints. Let t_1 and t_2 be the two red triangles containing e with vertices B, C, D, and A, B, D respectively. Without loss of generality, let C be the vertex on the outermost side. Denote the number of blue edge segments separating one of the four corners A, B, C, and D of the quadrilateral $t_1 \cup t_2$ from the other three by a, b, c, and d, respectively. The quadrilateral may also have x transversing edge segments in one of two directions. Without loss of generality, let these intersect the edges running between B and C, and A and D, respectively. See Figure 15 for an illustration.

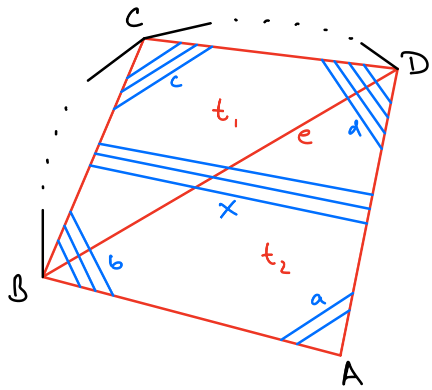


Figure 15 An outermost red edge e with the maximum number of blue intersections.

We then have for the number of intersections of the 5 edges of t_1 and t_2: e is intersected $x + b + d = m$ times. The edge between B and D is intersected $x + b + c < m$ times (note that outermost edges must have strictly fewer than m intersections by construction). It follows that we have $c < d$. Similarly, the edge between A and D is intersected $x + a + d \leq m$ times and hence $a \leq b$. It follows that flipping e creates an edge which is intersected $x + a + c < x + b + d = m$ times and the overall number of intersections between the red and blue triangulations has strictly decreased. ◀

From Lemma 1 the existence of crossing-monotone paths immediately follows: iteratively flip an outermost red edge with maximum number of crossings, until the red and blue triangulations coincide. Since $\mathcal{T}_1$ and $\mathcal{T}_2$ initially have $O(n^2)$ crossings, this is also an upper bound for the length of such a path.

4.7.2.2 Shadow vertex paths

Secondary polytopes and regular triangulations. Regular triangulations of point sets in $\mathbb{R}^d$ are projections of the lower convex hulls of lifted point sets, in which every point is assigned an arbitrary height as $(d + 1)$th coordinate. It is well-known that regular triangulations are one-to-one with vertices of the *secondary polytope* [11]. The latter is realized as the convex hull of points whose coordinates are the combined volume of the simplices incident to each point in the triangulation. The edges of the secondary polytopes are one-to-one with bistellar flips. Note that, in the general case, flips are no longer always edge flips and a more general definition is needed for them (see for instance [24, Definition 4]). In particular, flips can remove or introduce a vertex in a triangulation, provided that this vertex is not a vertex of the convex hull of the considered set of points. For a comprehensive introduction to flips and flip graphs in arbitrary dimensions, see [8].

The method. A shadow vertex paths on the secondary polytope is obtained by linearly interpolating between height functions of the two triangulations corresponding to the two endpoints. A height function can be interpreted as a direction in space, such that the lower convex hull defines the triangulation associated with the optimal vertex of the secondary polytope in this direction. Recall that a path in the edge graph of a polytope P is *non-revisiting* when the vertices of that path that belong to any given facet of P are consecutive. In other words, such a path never re-enters a facet of P after it leaves it.

▶ **Lemma 2.** Shadow vertex paths on the secondary polytope of a set of points in general position are non-revisiting. This implies that a shadow vertex path on the secondary polytope of a set of n points in convex position in the plane always has length at most $(n-2)(n-3)/2$.

Proof. This follows from the known fact that, when a face of a triangulation disappears along a shadow vertex path, this face cannot reappear later in that path [24, Theorem 8]. Facets of secondary polytopes are one-to-one with subdivisions that can only be coarsened by the trivial subdivision [8]. Consider such a subdivision S that corresponds to a facet F of the secondary polytope. When the point set that gives rise to the secondary polytope is in general position, a face of S that is not full-dimensional must be a simplex. The shadow vertex path leaves F precisely when one of these simplices disappears. As they cannot reappear later along the path, F cannot be revisited.

For the secondary polytope P of n points in convex position in the plane, the facets correspond to the $n(n-3)/2$ diagonals of the polygon. Since a vertex of P is incident to $n-3$ facets and the shadow vertex paths on P are non-revisiting, such a path has length at most

$$\frac{n(n-3)}{2} - (n-3) = \frac{(n-2)(n-3)}{2}$$

as desired. ◀

Implications and scope of the method. We recall that a polytope is Hirsch when the diameter of its edge graph is at most the number of its facets minus its dimension [13, 26].

▶ **Theorem 3.** Secondary polytopes of point sets in general position are Hirsch.

Proof. The existence of a non-revisiting path between every pair of vertices implies Hirsch's bound on the diameter [13, 26]. ◀

Recall that a polytope is *non-revisiting* if all the geodesics paths between its vertices are non-revisiting (this is called the *non-leaving face property* in [6]).

▶ **Lemma 4.** There exist secondary polytopes of d-dimensional point sets, for some d, that are not non-revisiting.

Proof. This is obtained with the so-called *mother of all examples* [8]: point sets consisting of two homothetic simplices one of whose is contained in the interior of the other. Their secondary polytope is known as the *stellohedron* [22], and is known to have revisiting geodesics [6]. ◀

This shows that for higher-dimensional regular triangulations, there exist geodesics that are *not* shadow vertex paths.

▶ **Lemma 5.** There exist secondary polytopes and geodesics on those, that are not shadow vertex paths. Hence there does not exist any height function whose interpolation yields a shortest flip sequence.

4.7.2.3 A family of examples producing long paths

In this section we present a family of pairs of triangulations of the $n = 2k$-gon $(\mathcal{E}_k^1, \mathcal{E}_k^2)_{k\geq 5}$ for which the methods described in Sections 4.7.2.1 and 4.7.2.2 produce paths of asymptotically worst possible length $\Theta(n^2)$.

Figure 16 shows four members of our family of examples, each consisting of a start and a target triangulation. They consist of triangulations of a $n = 2k$-gon, obtained by gluing fan-triangulations of two $(k+1)$-gons in a point-symmetric way. Start and end triangulations are identical, except for a rotation of one $2k$th of a full twist.

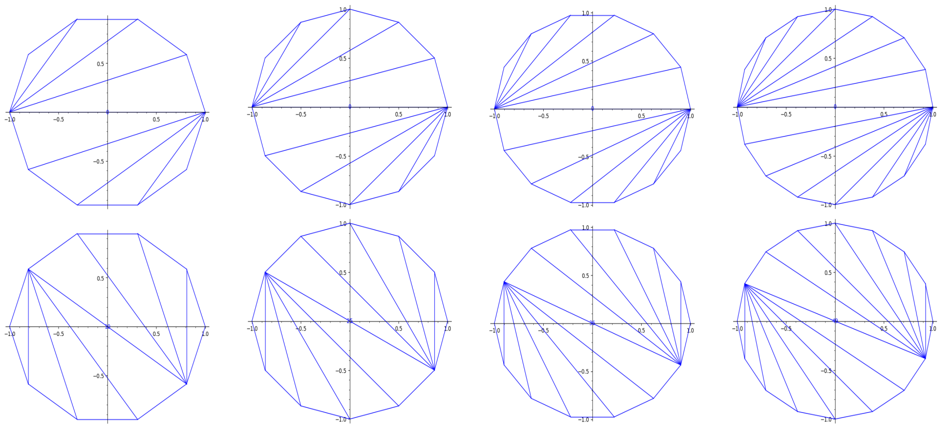


Figure 16 Our family of examples for $k = 5, 6, 7, 8$. Top row: start triangulations $\mathcal{E}_k^1$. Bottom row: target triangulations $\mathcal{E}_k^2$.

▶ **Lemma 6.** There exist shadow vertex paths transforming triangulations of the $n = 2k$-gon $\mathcal{E}_k^1$ into $\mathcal{E}_k^2$ in $(n/2-1)^2$ steps.

In particular, there exist shadow vertex paths of length $\Omega(n^2)$ on the (secondary polytope realization of the) associahedron, hence the upper bound $n(n-3)/2$ from Lemma 2 is tight, up to a factor of 2.

Sketch of the proof. It is enough to provide a pair of height functions for every pair $(\mathcal{E}_k^1, \mathcal{E}_k^2)$ of triangulations with the desired properties.

We pick vertices of the $2k$-gon with coordinates $(\cos(j\pi/k), \sin(j\pi/k))$, $0 \leq j < 2k$, and label them from 0 to $2k-1$ accordingly. For $0 \leq j < k$, vertices j and $k+j$ start with heights $-j^2$ and end with heights $2kj - j^2$, linearly interpolating between these to values to flip from $\mathcal{E}_k^1$ to $\mathcal{E}_k^2$. The heights can be perturbed by small amounts to ensure general position throughout the procedure.

We claim without proof that the shadow vertex paths obtained in this way have the desired property. ◀

▶ Remark. Gradually adding small, random perturbations to the height functions proposed in the proof of Lemma 6 allows us to find other flip sequences using the shadow vertex path method. For instance, a guided search based on iteratively performing small random permutations, finds flip sequences between $\mathcal{E}_5^1$ and $\mathcal{E}_5^2$ of length 10 (a geodesic) and 17.

▶ **Lemma 7.** There exist crossing-monotone sequences of flips between triangulations of a convex n-gon that are of quadratic size (and hence worst possible).

Proof. The flip sequence from the proof of Lemma 6 is crossing-monotone. ◀

The crossings based heuristics described at the beginning of the section is implemented in practice by selecting at each step a flip that makes the number of crossings decrease the most. In that case, it is known that the heuristics is, at best a 1.5-approximation algorithm [25]. However, computational evidence suggests that it is accurate most of the time [7]. This implementation works very well on the examples that we describe here and we ask the following.

▶ **Question 8.** Is the *greedy* crossing heuristic, in which we flip an arc that decreases the number of crossings the most, an approximation algorithm?

4.7.3 The modular flip graph of the annulus

Consider an orientable surface Σ equipped with a finite set of points such that each boundary component of Σ contains at least one of these points. By a triangulation of Σ, we mean an inclusion-wise maximal set of pairwise non-crossing arcs on Σ between two of these prescribed points. Note that the arcs are considered up to isotopy, so in particular, when we say that two arcs are non-crossing, this means that two representatives in the corresponding isotopy classes are non-crossing. The flip graph $\mathcal{F}(\Sigma)$ of Σ is the graph whose vertices are the triangulations of Σ and whose edges correspond to flipping arcs.

The modular flip graph $\mathcal{MF}(\Sigma)$ of Σ is obtained by quotienting $\mathcal{F}(\Sigma)$ by the homeomorphisms of Σ that fix the vertices of the triangulations pointwise [17]. It is known that when Σ is an annulus with n points on one boundary and a single point on the other, the diameter of the modular flip graph of Σ is exactly $\lfloor 5n/2 \rfloor - 2$ [17, Theorem 1.4]. In this section, we consider the slightly more general case when Σ is an annulus with n_1 points on one boundary and n_2 on the other, with $n = n_1 + n_2$. We aim at bounding the worst-case diameter of the modular flip graph of Σ as a function of n.

4.7.3.1 Upper bounds

In this section we prove the following statement.

▶ **Lemma 9.** Let Σ be an annulus with a total of n marked points on its two boundary. The diameter of $\mathcal{MF}(\Sigma)$ is $\lfloor 5(n-1)/2 \rfloor - 2$.

Proof. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two arbitrary triangulations of the annulus Σ. Let v be a vertex on the inner boundary of Σ such that in $\mathcal{T}_2$, v is adjacent to at least one vertex w on the outer boundary of Σ. (Note that each triangulation of Σ has at least one such vertex v). Let n_1 and n_2, $n_1, n_2 \geq 1$ and $n_1 + n_2 = n$, be the number of vertices on the outer and on the inner boundary component of Σ, respectively.

We start by flipping $\mathcal{T}_1$ to a common target. For this we iteratively increase the vertex degree of v in $\mathcal{T}_1$ until it is adjacent to all other vertices. Note that, as long as v is not yet adjacent to all other vertices, it is always possible to find a flip that strictly increases the degree of v: every edge that is not a boundary edge is flippable, and the vertex link of v contains a non-boundary edge if and only if v is not yet adjacent to all other vertices. Since there are n non-boundary edges, and v may initially not be adjacent to a single one, this can always be done in at most n flips.

The result is a triangulation $\mathcal{T}_1'$ with – necessarily – the following properties:

- there exists a unique vertex u on the other boundary with two edges running between v and u, enclosing this inner boundary and forming a heart shape
- every other vertex on the outer boundary has exactly one edge running between it and v
- there is a loop edge running from v to v and encircling the inner boundary

- there exists a unique vertex x on the inner boundary with two edges running between v and x
- every other vertex on the inner boundary has exactly one edge running between it and v

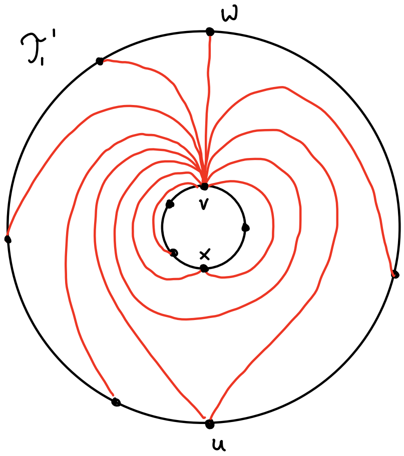


Figure 17 The triangulation $\mathcal{T}_1'$ of the annulus with $n_1 = 6$ vertices on the outer, and $n_2 = 5$ vertices on the inner boundary components.

Note that the vertices u, v, and x in the above properties uniquely determine $\mathcal{T}_1'$. See Figure 17 for an illustration.

For $\mathcal{T}_2$, we cut Σ along the edge vw to obtain a "rectangle" (cycle with $n+2$ vertices), whose corners are two copies of each of v and w. Two opposite sides of the rectangle contain the outer and inner boundary points of Σ, respectively, and the other two sides are the two copies of the edge vw.

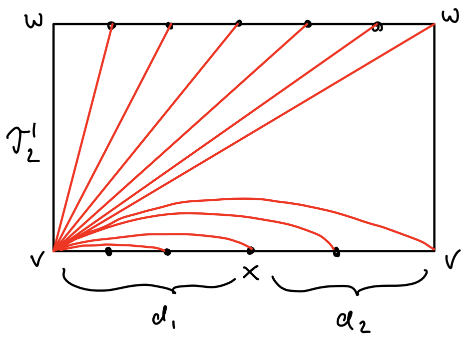


Figure 18 The triangulation $\mathcal{T}_2'$, again with $n_1 = 6$ vertices on the outer, and $n_2 = 5$ vertices on the inner boundary components. Note that x is only one flip away from being twice adjacent to v.

Note that the edges of $\mathcal{T}_2$ form a triangulation of this rectangle and that flips between triangulations in the rectangle correspond to flips between triangulations of the annulus Σ. Using at most $n-1$ flips, we can flip in this rectangle from $\mathcal{T}_2$ to a fan at any of the two copies of v. Gluing the resulting triangulation along edge vw yields a triangulation $\mathcal{T}_2'$ of Σ where v has maximum degree,

Note that we can arbitrarily choose, which copy of v we use as the center of the fan. We make this choice in the following way: Consider the number d_1 of vertices of Σ that lie between one copy v and x along the inner boundary part of the rectangle, and the number

d_2 of vertices of Σ that lie between the other copy of v and x along the inner boundary part of the rectangle. For the centre of the fan, we choose the copy of v for which this number is greater. Note that since $d_1 + d_2 = n_2 - 2$, we have $\min\{d_1, d_2\} \leq \lfloor n_2 - 2/2 \rfloor = \lfloor n_2/2 \rfloor - 1$.

We can now flip $\mathcal{T}_1'$ to $\mathcal{T}_2'$ in the following way: Along the outer boundary of Σ, we keep on flipping one of the two edges between v and (initially) u until there are two edges running between v and w. Since we can choose the direction around the outer circle in which we travel, this can be done in at most $\lfloor n_1/2 \rfloor$ flips. Next we turn to the inner boundary of Σ: here, similarly as before, we can flip x to be twice adjacent to v in at most $\lfloor n_2/2 \rfloor - 1$ flips. This is possible because of how we chose the centre of the fan. After these additional at most $\lfloor n_1/2 \rfloor + \lfloor n_2/2 \rfloor - 1$ flips, the triangulations now coincide.

Altogether, the flip sequence described above is of length $\lfloor 5n/2 \rfloor - 2$. ◀

4.7.3.2 Lower bounds

The tools to prove lower bounds on the diameter of flip-graphs of surfaces can be found in [17, 18, 19, 23].

The best introduction to the lower bound techniques would be to read [17] first, though the case of the annulus with n points on a boundary and one point on the other (matching lower and upper bounds for the diameter of the flip-graph) is treated in [17]. Note also that a general upper bound of 4n+K (K is a constant) on the diameter of the flip-graph is given in [17] in the case when there's a boundary containing n points and the rest of the surface is arbitrary but fixed. These techniques were introduced in [23] in the special case of convex polygons.

▶ **Lemma 10.** There exist triangulations $\mathcal{T}_1$ and $\mathcal{T}_2$ of the annulus with $n = n_1 + n_2$ points on their boundaries that are $2.25n + O(1)$ flips apart.

Proof. Start with the diameter realising pair of triangulations of the case $n_1 = n - 1$, $n_2 = 1$. This requires a flip sequence of length at least $\approx 5n_1/2$ [17].

Next we think of the inner boundary to have more than one point. Since there is a loop edge encircling the inner boundary (see [17] for details), the space bounded by the inner boundary together with the loop edge is an n_2-gon of diameter $\approx 2n_2$. Since, according to [27], a geodesic between triangulations having a common edge, preserves this edge, the loop edge of these examples remains intact in a shortest path from one to the other. It follows that these triangulations require at least $5n_1/2 + 2n_2 + O(1)$ flips to be connected. Since inner and outer boundary are symmetric, this bound is lowest for $n_1 \approx n_2$ and hence we obtain a lower bound of $2.25n + O(1)$. ◀

References

1 Eduardo Altmann and Jonathan Spreer. Sampling triangulations of manifolds using Monte Carlo methods. `https://arxiv.org/abs/2310.07372`, 2023. 29 pages, 6 figures.

2 Mark C. Bell. Simplifying triangulations. *Discrete & Computational Geometry*, 66(1):1–11, 2021.

3 Anders Björner and Frank H. Lutz. Simplicial manifolds, bistellar flips and a 16-vertex triangulation of the Poincaré homology 3-sphere. *Experiment. Math.*, 9(2):275–289, 2000.

4 Benjamin A. Burton. The Pachner graph and the simplification of 3-sphere triangulations. In *SoCG '11: Proceedings of the twenty-seventh annual symposium on Computational geometry*, pages 153–162. ACM, 2011.

5 Michael P. Carr and Satyan L. Devadoss. Coxeter complexes and graph-associahedra. *Topology and its Applications*, 153:2155–2168, 2006.

6 Cesar Ceballos and Vincent Pilaud. The diameter of type D associahedra and the non-leaving-face property. *European Journal of Combinatorics*, 51:109–124, 2016.
7 Sean Cleary and Roland Maio. Edge conflicts do not determine geodesics in the associahedron. *SIAM Journal on Discrete Mathematics*, 32(2):1003–1015, 2018.
8 Jesús A. de Loera, Jörg Rambau, and Francisco Santos. *Triangulations: structures for algorithms and applications*, volume 25 of *Algorithms and Computation in Mathematics*. Springer, 2010.
9 Valentina Disarlo and Hugo Parlier. The geometry of flip graphs and mapping class groups. *Transactions of the American Mathematical Society*, 372(6):3809–3844, 2019.
10 Sergey Fomin and Andrei Zelevinsky. Y-systems and generalized associahedra. *Annals of Mathematics*, 158:977–1018, 2003.
11 Israel M. Gel'fand, Andrei V. Zelevinsky, and Mikhail M. Kapranov. Newton polytopes of principal A-determinants. *Soviet Mathematics. Doklady*, 40:278–281, 1990.
12 Sabine Hanke, Thomas Ottmann, and Sven Schuierer. The edge-flipping distance of triangulations. *Journal of Universal Computer Science*, 2(8):570–579, 1997.
13 Victor Klee and David W. Walkup. The d-step conjecture for polyhedra of dimension $d < 6$. *Acta Mathematica*, 117:53–78, 1967.
14 Carl W. Lee. The associahedron and triangulations of the n-gon. *European Journal of Combinatorics*, 10(6):551–560, 1989.
15 Anna Lubiw and Vinayak Pathak. Flip distance between two triangulations of a point set is NP-complete. *Computational Geometry*, 49:17–23, 2015.
16 Lee Mosher. Mapping class groups are automatic. *Annals of Mathematics*, 142(2):303–384, 1995.
17 Hugo Parlier and Lionel Pournin. Flip-graph moduli spaces of filling surfaces. *Journal of the European Mathematical Society*, 19(9):2697–2737, 2017.
18 Hugo Parlier and Lionel Pournin. Modular flip-graphs of one holed surfaces. *European Journal of Combinatorics*, 67:158–173, 2018.
19 Hugo Parlier and Lionel Pournin. Once punctured disks, non-convex polygons, and pointihedra. *Annals of Combinatorics*, 22:619–640, 2018.
20 Alexander Pilz. Flip distance between triangulations of a planar point set is APX-hard. *Computational Geometry*, 47:589–604, 2014.
21 Alexander Postnikov. Permutohedra, associahedra, and beyond. *International Mathematics Research Notices*, 2009(6):1026–1106, 2009.
22 Alexander Postnikov, Victor Reiner, and Lauren Williams. Faces of generalized permutohedra. *Documenta Mathematica*, 13:207–273, 2008.
23 Lionel Pournin. The diameter of associahedra. *Advances in Mathematics*, 259:13–42, 2014.
24 Lionel Pournin and Thomas M. Liebling. Constrained paths in the flip-graph of regular triangulations. *Computational Geometry*, 37(2):134–140, 2007.
25 Lionel Pournin and Zili Wang. Strong convexity in flip-graphs. *arXiv:2106.08012*, 2021.
26 Francisco Santos. A counterexample to the Hirsch conjecture. *Annals of Mathematics*, 176:383–412, 2012.
27 Daniel Sleator, Robert Tarjan, and William Thurston. Rotation distance, triangulations, and hyperbolic geometry. *Journal of the American Mathematical Society*, 1:647–681, 1988.

Participants

- Mikkel Abrahamsen
University of Copenhagen, DK
- Therese Biedl
University of Waterloo, CA
- Florestan Brunck
IST Austria –
HKlosterneuburg, AT
- Kevin Buchin
TU Dortmund, DE
- Maike Buchin
Ruhr-Universität Bochum, DE
- Benjamin Burton
The University of Queensland –
Brisbane, AU
- Jean Cardinal
ULB – Brussels, BE
- Hsien-Chih Chang
Dartmouth College –
Hanover, US
- éric Colin de Verdière
Gustave Eiffel University –
Marne-la-Vallée, FR
- Arnaud de Mesmay
CNRS, Gustave Eiffel University
– Marne-la-Vallée, FR
- Fabrizio Frati
University of Rome III, IT
- Alex He
The University of Queensland –
Brisbane, AU
- Linda Kleist
TU Braunschweig, DE
- Francis Lazarus
CNRS – Grenoble, FR
- Maarten Löffler
Utrecht University, NL
- Anna Lubiw
University of Waterloo, CA
- Clément Maria
INRIA – Sophia Antipolis, FR
- Tim Ophelders
Utrecht University, NL
- Lionel Pournin
Université Paris 13 –
Villetaneuse, FR
- Günter Rote
FU Berlin, DE
- Saul Schleimer
University of Warwick –
Coventry, GB
- Lena Schlipf
Universität Tübingen, DE
- André Schulz
FernUniversität in Hagen, DE
- Eric Sedgwick
DePaul University – Chicago, US
- Rodrigo I. Silveira
UPC Barcelona Tech, ES
- Jonathan Spreer
University of Sydney, AU
- Stephan Tillmann
University of Sydney, AU
- Birgit Vogtenhuber
TU Graz, AT
- Zili Wang
Sun Yat-Sen University –
Shenzen, CN
- Erin Moriarty Wolf Chambers
St. Louis University, US
- Alexander Wolff
Universität Würzburg, DE

Report from Dagstuhl Seminar 24081

Computational Approaches to Strategy and Tactics in Sports

Ulf Brefeld*[1], Jesse Davis*[2], Laura de Jong*[3], and Stephanie Kovalchik*[4]

1 **Leuphana Universität Lüneburg, DE.** `ulf.brefeld@leuphana.de`
2 **KU Leuven, BE.** `jesse.davis@kuleuven.be`
3 **Deakin University – Melbourne, AU.** `mail@lmsdejong.nl`
4 **Zelus Analytics – Austin, US.** `skovalchik@zelusanalytics.com`

Abstract

One of the most challenging and interesting aspects in sports are *Strategy* and *Tactics*. In this interdisciplinary Dagstuhl Seminar, we aimed to develop a computational understanding of these concepts in an interdisciplinary setting with researchers and practitioners from Machine Learning, Statistics, and Sports. The seminar was organized around the themes "Discovery", "Evaluation", and "Communication" that were introduced with tutorial and overview style talks about the key concepts to facilitate a common ground among researchers with different backgrounds. These were augmented by more in-depth presentations on specific problems or techniques. Besides several topical discussions in larger groups, there were two panel discussions dealing with differences between individual and team sports and bringing computational analytics into practice, respectively.

Seminar February 18–23, 2024 – https://www.dagstuhl.de/24081
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning
Keywords and phrases AI, machine learning, sports, team, athletes, strategy, tactics
Digital Object Identifier 10.4230/DagRep.14.2.164

1 Executive Summary

Ulf Brefeld
Jesse Davis
Laura de Jong
Stephanie Kovalchik

The rapid growth in spatio-temporal data in sport over the past decade has generated numerous methodological developments from the statistical and machine learning communities. The richness of modern sports data is enabling sports researchers to analyze every action and decision during a competitive event in increasing detail. Two central topics that have emerged from this new phase of methodological research in sport are data-driven approaches to *strategy* & *tactics*. In a nutshell, *strategy* & *tactics* allow weaker teams or athletes to win over stronger ones. Therefore, they are one of the most interesting and challenging aspects in sports.

* Editor / Organizer

Computational Approaches to Strategy and Tactics in Sports, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 164–181
Editors: Ulf Brefeld, Jesse Davis, Laura de Jong, and Stephanie Kovalchik
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Although both terms describe similar aspects and are even sometimes used interchangeably, they range on different time scales. A *strategy* serves as an overarching umbrella to reach long-term goals. Hence, strategic decisions involve long-term training plans, signing players and coaches, as well as deciding on team formation, pacing, equipment, rotations, or playing philosophy. On a shorter time scale a match/race strategy is the plan made by coaches and athletes before the start of the match or race.

Tactics, on the other hand, is rather short-term. Tactics are the execution and adaptations to the planned strategy to have an edge over the opponent during the match or race. Tactics are therefore often broken down into building blocks or patterns that can be easily communicated to athletes. Note that communication is of utmost importance as tactics are invented by the coaching staff while their implementation is the task of the players/athletes. A tactical pattern may thus involve only subgroups of athletes, or subsections of a race and assign concrete tasks for predefined, context-sensitive situations.

The goal of this Dagstuhl Seminar was to bring together a diverse set of researchers from both academia and industry working on these topics. The seminar drew from people with various backgrounds in terms of area of specialization (Artificial Intelligence, Operations Research, Sport Science, Statistics), role (Academic, Data Provider, Federation, Sports Club) and sport (Australia Rules Football, Baseball, Basketball, Darts, Ice Hockey, Soccer, Speed Skating, Tennis, Wheelchair Rugby).

The seminar was structured around three themes:

Discovery The goal of this theme was to discuss different methods that can automatically identify tactical and strategic patterns from spatio-temporal data. Examples were given for problems such as detecting formations, identifying commonly occurring sequences of actions (e.g., passing sequences), discovering player movement trajectories, and deciding where players should aim a tennis serve.

Evaluation This theme focused on the challenges and pitfalls associated with trying to evaluate the finding of computational approaches to identifying strategies and tactics. This theme focused on highlighting a number of methodological issues and describing ways to assess the validity of discoveries. There were a number of interesting examples given about how causal analysis could be used to evaluate the efficacy of certain tactics. Finally, the potential and risks for using large language models in sports were also discussed.

Communication This theme tackled the problem of how to communicate the findings from tactical studies to an interdisciplinary audience. The emphasis was on how to marry finding from the research literature to things that could be translated into practice. A key point that was made is that it is crucial to think about what types of information will be useful and actionable for practitioners.

The first three days of the seminar focused on one theme, which was introduced with a longer tutorial and then shorter presentations. The final full day of the seminar was open to all topics under the themes and there was a greater focus on presentations from early-career researchers in attendance. The seminar also featured two panels and (small) group discussions about five different topics.

Results

During the seminar, we identified and agreed upon the following action points aimed at trying to continue integrating the various different communities (Sports Science, Operations Research, Statistics, Artificial Intelligence) working on computational approaches to tactics in sports:

1. We will collect a list of venues where computational approaches to tactics in sports are often published. We will host this on the web: `https://dtai.cs.kuleuven.be/sports/venues/`
2. We will explore setting up a slack or discord channel to facilitate more continuous interaction and the ability to quickly get answers to questions. Joris Bekkers and Jan Van Haaren will take the lead on this point.
3. We have setup a document that contains the biographies, contact details, and topics of interest for all seminar participants that are willing to share their information. That will help people stay in touch.
4. We will strive to setup some basic tutorials that illustrate how to implement standard, concepts that reoccur across sports. For example, many team sports have variants of plus-minus, expected possession value metrics, and expected statistics such as expected goals (soccer, ice hockey) or expected rush yard gained (American Football).
5. We will continue to promote the mailing list for disseminating computational sports-related information (job ads, conference call for papers, etc.) and we will use this list to distribute the report on the seminar to reinforce our thanks to the attendees and excitement about the seminar's outcomes: ml-ai-4sports@googlegroups.com

2 Table of Contents

3 Overview of Talks

3.1 The Secrets of Competition: Using AVATARS to Better Understand Exercise Regulation and Competition

Florentina Hettinga (Northumbria University, GB)

Competition between athletes is central to sport. Athletes need to determine how and when to invest their limited energy resources to optimise and self-regulate their pace depending on their physiological and biomechanical capacity as well as environmental factors, such as the presence and behaviour of another athlete. Remote technology can be used to explore mechanisms involved in exercise regulation and competition, for example via the use of avatars, which can be a graphical representation of an athlete's own performance or that of another athlete. I will overview a series of studies using avatar scenario's to better understand decision-making and pacing in sport and competition.

3.2 Using a Markov Chain Model to Identify Optimal Football Match Tactics

Benjamin Holmes (University of Liverpool, GB)

In this paper, we develop a Markov chain representation of football. States consist of different locations on the pitch and set-pieces, and different actions, such as passes, shots, and tackles, move the chain between these states. Novel variables which describe the abilities of the opposing teams in different aspects of football over different zones of the pitch drive the transition probabilities. By simulating a match numerous times using different scenarios, we can identify the optimal choice of tactics: who should play and in what formation, or what style of football to implement. A case-study using a recent match between Everton and Chelsea demonstrates the usefulness of the model, as well as the detailed predictions one can obtain.

3.3 Unveiling Tactical Advantage in Baseball: Insights from Kinematic Analysis and Predictive Modelling

Mamiko Kato (Toyo University, JP)

A remarkable development of data measurement and collection technology in baseball has provided us with opportunities to gather comprehensive performance data. Our research team conducted detailed analyses of batter and fielder performances in the official professional baseball games. Specifically, the kinematic characteristics of batted balls were analysed collectively to evaluate how fast, how high, and in which direction the batters should aim to hit the ball to increase the chance of making base hits and home runs. Additionally, we introduced a novel method for fielder performance evaluation, addressing the challenge of comparing abilities fairly due to asymmetrical distributions of batted ball characteristics across the baseball field. A machine learning algorithm was used to predict the probability of flyout from the kinematic characteristics of fly balls and compared the probability score for a systematically constructed set of fly balls, uniformly distributed across the field. We discovered that hitting towards the same side of the field increased the chance of a base hit and argued that the advantage was attributable to the large variance in both the direction and magnitude of deflection seen in the trajectories of the pulled fly balls. Based on these studies, using classical statistical methods as well as simulation with predictive modelling with vast amounts of batted ball data will help us conduct fair and rigorous comparisons of players' performance and provide target values with a given aim in baseball. The future development of this methodology in baseball and its potential applications in other sports will be discussed to provide findings and implications in practical situations effectively.

3.4 Expected Thread Models – Overview and Ideas of Validation

Matthias Kempe (University of Groningen, NL)

In this talk, I will give a short intro one a new line of research in 1 vs 1 actions. To study them and to quantify action in football in general, one needs a valid success measure. Expected Thread is one of the suggested measures. However, models that circulate right now might have different short comings and are not validated properly. I like to give an overview on the different models in the literature and propose some standards and guidelines to validate them.

3.5 SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data

Hyunsung Kim (Seoul National University, KR)

In fluid team sports such as soccer and basketball, analyzing team formation is one of the most intuitive ways to understand tactics from domain participants' point of view. However, existing approaches either assume that the team formation is consistent throughout a match or assign formations frame-by-frame, which disagree with real situations. To tackle this issue, we propose a change-point detection framework named SoccerCPD that distinguishes tactically intended formation and role changes from temporary changes in soccer matches. We first assign roles to players frame-by-frame and perform two-step change-point detections: (1) formation change-point detection based on the sequence of role-adjacency matrices and (2) role change-point detection based on the sequence of role permutations. The evaluation of SoccerCPD using the ground truth annotated by domain experts shows that our method accurately detects the points of tactical changes and estimates the formation and role assignment per segment. Lastly, we introduce practical use-cases that domain participants can easily interpret and utilize.

3.6 What Is Supportive for Coaching Practice and What Is Not (So Much)?

Martin Lames (TU München, DE)

I try to classify computational approaches to strategy and tactics according to their usefulness for sports practice. Basic research in informatics based on sports data, Messi-and-Ronaldo-look-good-in-my-data or I-found-a-new-performance-indicator typed studies may be criticized in this respect. In the narrow sense of support for coaches and teams (not in the wider sense of sports analytics), practical hints for training should be the ultimate aim. The analysis process in sports practice is presented and more concrete requirements for support are derived.

3.7 Match Analysis in German Beachvolleyball – An Ecosystem for Data Collection, Data Analysis, Communication of Results and Training

Daniel Link (TU München, DE)

This talk introduces the game observation concept used by the German national teams in beach volleyball. First, the talk discusses the type of statements in match analysis from a conceptual perspective and explores questions that relevant for beach volleyball coaches. The methods section outlines the logical work steps of match analysis, including: i) querying game scenes according to a classifier; ii) the quantitative preliminary analysis supported by descriptive statistics and graphical reports; and iii) the main qualitative analysis based on video-recordings. It also shows how results are communicated to the German athletes by using a custom made presentation tool and how they use the data for anticipation training.

3.8 Scaling Soccer Data and Analysis using Multimodal Language Modeling

Patrick Lucey (Stats Perform – Chicago, US)

The integration of data and artificial intelligence (AI) has significantly enhanced performance measurement in sports, particularly in soccer. Event data enables the assessment of on-ball performance, encompassing metrics such as xG (expected goals), Possession Value, and win probability. Complementary tracking data captures off-the-ball actions, informing fitness assessments, team tactics, defensive strategies, and passing options. Despite advancements, the scalability of these measurements remains constrained by the lack of "complete" tracking data. For 25 years, tracking data in soccer has relied on in-venue systems, established since 1998. However, the necessity for being in-venue for collection has limited it wide use. Recent years have witnessed efforts to utilize tracking data from broadcast video as a scalable alternative. Yet, challenges persist, with occlusions caused by players out of view (or complete segments of play being missed completely) resulting in incomplete data and therefore incomplete downstream analysis. In this presentation, I will explore the application of Multimodal LLM approaches to address this limitation. By leveraging these methodologies, I will discuss how complete tracking data can be extracted from broadcast video, ensuring accuracy and completeness akin to in-venue systems. This advancement holds promise for reliable and trustworthy analysis and strategic decision-making in soccer and beyond.

3.9 Team Tactical Performance in Small-sided Games in Football

Sigrid Olthof (John Moores University – Liverpool, GB)

The purpose of this presentation is three-fold: i) measuring team tactical performance, ii) using small-sided games, and iii) highlighting applications to the football practice. Team tactical performance in football (soccer) refers to the cooperation of players within a team and the competition between teams. Positional data from tracking technology is used for team tactical performance metrics representing the positioning and dispersion of players on the pitch. Knowledge of match performance allows for developing and improving performance through training. Small-sided games (SSGs) are training formats representing the match (phases) and therefore a popular training drill. Usually, changing the pitch size and number of players in SSGs affects the individual performance (physical and technical) and team tactical performance compared to match performance, compromising the usefulness and representativeness of SSGs. In this presentation, I shared insights of optimising SSGs by using a similar relative pitch area (total pitch surface / number of players) as the match. This leads to similar team tactical performance in SSGs and the match. These insights can be applied to the football practice with examples of new designs of grassroots competitions, coach dashboards, and a range of games for training programs. Involving coaches in this process is crucial for successful applications.

3.10 Presenting Multiagent Challenges in Team Sports Analytics

David Radke (Chicago Blackhawks, US)

This talk will present several challenges and opportunities within the area of team sports analytics and key research areas within multiagent systems (MAS). We specifically consider invasion games, where players invade the opposing team's territory and can interact anywhere on a playing surface (ice hockey or soccer). We discuss how MAS is well-equipped to study invasion games and will benefit both MAS and sports analytics fields. We highlight topics along two axes: short-term strategy (coaching) and long-term team planning (management).

3.11 Unlocking Insights: Interpretable Models in Soccer Analytics

Pegah Ramihian (Twelve Football – Stockholm, SE)

In the fast-evolving landscape of soccer analytics, leveraging the power of deep learning has become imperative for gaining a competitive edge. However, the opacity of these models often leaves coaches and players in the dark, hindering the practical application of insights. We need to delve into the significance of using interpretable models in soccer analytics, shedding light on the "why" behind every "what."

3.12 Masked Autoencoder for Multiagent Trajectories

Yannick Rudolph (Leuphana Universität Lüneburg, DE)

Automatically labeling trajectories of multiple agents is key to behavioral analyses but usually requires a large amount of manual annotations. This also applies to the domain of team sport analyses. In this paper, we specifically show how pretraining transformer models improve the classification performance on tracking data from professional soccer. For this purpose, we propose a novel self-supervised masked autoencoder for multiagent trajectories to effectively learn from only a few labeled sequences. Our approach employs a masking scheme on the level of individual agent trajectories and makes novel use of a factorized transformer architecture for multiagent trajectory data. As a result, our model allows for a reconstruction of masked trajectory segments while being permutation equivariant with respect to the agent trajectories. In contrast to related work, our approach is conceptually much simpler, does not require handcrafted features and naturally allows for permutation invariance in downstream tasks.

3.13 Practical Implications and Solutions to Foster Innovation in Sports Data Analytics

Martin Rumo (OYM AG – Cham, CH)

The transformation into a data-driven culture presents significant challenges for sports organizations, notably the issue of data silos that impede the free flow of information necessary for innovation in sports data analytics. This talk aims to tackle these barriers by offering practical solutions and discussing their implications. Firstly, we will introduce a data-centric approach to break down internal data silos within sports clubs, ensuring optimized access to data for analytics. Secondly, for industry-wide data silos, we will present the blockchain-based protocol OCEAN, illustrated through a case study from Swiss ice hockey, to privacy conserving data sharing across different clubs. Furthermore, the talk will address the necessity of seamless integration of academic research and solutions into the sports industry's existing technological frameworks. Strategies to facilitate this integration, enhancing the flow of innovative ideas from universities to the field, will be explored. Additionally, the growing complexity of sports information systems is identified as a challenge to innovation; we propose the adoption of microservices architecture as a scalable and flexible solution to this problem. Conclusively, the presentation will outline actionable steps that organizations can take to overcome these innovation barriers, paving the way for a more integrated, efficient, and innovative future in sports data analytics.

3.14 An Experiment to Investigate the Spatial Component of Serving Strategy in Tennis

Nathan Sandholtz (Brigham Young University, US)

We conducted an experiment with the Brigham Young University Men's Tennis Team to investigate the spatial element of serving strategies in tennis. Serve data, including precise spatial coordinates, were collected for 12 players, with known targets for each serve. Leveraging this data, we estimate player-specific optimal aim locations, accounting for factors such as first vs second serve, speed, and the distribution of their serves around the intended targets, termed "execution error". Our experiment also provides insights on the interplay between conscious beliefs and on-court performance. Our preliminary results reveal apparent differences between players' subconscious behavior and their explicit articulation of optimal aiming locations.

3.15 Towards full automation and scalability when collecting spatiotemporal data in tennis

Joshua Smith (Concordia University – Montreal, CA)

Data analytics in tennis is a fast growing topic of interest, often related to fan engagement at tournaments. Higher ranked players are also able to benefit from the statistics and data collected by these tournaments. However, the collection of spatiotemporal (and even event) data has historically been complicated and expensive, and so developing players at all levels still do not have access to these types of resources. Recent advances in computing power and AI algorithms have allowed for cheaper and more efficient data collection for individual tennis matches. This promises to expose more players to the idea of analytics, with the hope that it can level the playing field. But there are still challenges to consider when building for even more scalability and versatility. This talk will explore some of the issues that arise when trying to fully automate the data collection process of a tennis match. This includes many facets of the approach from court detection, player identification, bounce detection etc. and includes the edge cases that one needs to consider when aiming for fully automatic data collection.

3.16 Tactical Problems in Football using Tracking Data and Causal Methods

Tim Swartz (Simon Fraser University – Burnaby, CA)

A frequent impediment to applied causal analysis is the identification and quantification of confounding variables. With the advent of tracking data in sport, there is often a realistic chance of dealing with the confounding variable problem. In this presentation, we consider three questions involving soccer tactics that are each approached using causal methods (a) what is the benefit of crossing the ball? (b) what is the benefit of playing with pace? and (c) what is the benefit associated with throw-in decisions? The problems each have a common structure, and we provide a template for approaching such problems. The differences between the problems lie in the nature of the independent variables, the dependent variables and the confounding variables.

3.17 Evaluating Sports Analytics Models

Jan Van Haaren (Club Brugge & KU Leuven, BE)

Joint work of Jan Van Haaren, Maaike Van Roy, Pieter Robberechts, Jesse Davis

There has been an explosion of data collected about sports. Because such data is extremely rich and complicated, machine learning is increasingly being used to extract actionable insights from it. Typically, machine learning is used to build models and indicators that capture the skills, capabilities, and tendencies of athletes and teams. Such indicators and models are in turn used to inform decision-making at professional clubs. Unfortunately, how to evaluate the use of machine learning in the context of sports remains extremely challenging. On the one hand, it is necessary to evaluate the developed indicators themselves, where one is confronted by a lack of labels and small sample sizes. On the other hand, it is necessary to evaluate the models themselves, which is complicated by the noisy and non-stationary nature of sports data. The goal of this presentation is three-fold. First, we detail some aspects about how analytics are used within a club environment. Second, we discuss pitfalls and best practices for evaluating models learned from sports data. Third, we overview various ways to validate developed indicators, which requires assessing if they can provide value to the workflow of practitioners.

3.18 A Markov Framework for Learning and Reasoning About Strategies in Professional Soccer

Maaike Van Roy (KU Leuven, BE)

Strategy-optimization is a fundamental element of dynamic and complex team sports such as soccer, American football, and basketball. As the amount of data that is collected from matches in these sports has increased, so has the demand for data-driven decision-making support. If alternative strategies need to be balanced, a data-driven approach can uncover insights that are not available from qualitative analysis. This could tremendously aid teams in their match preparations. In this talk, I present a novel Markov model-based framework for soccer that allows reasoning about the specific strategies teams use in order to gain insights into the efficiency of each strategy. The framework consists of two components: (1) a learning component, which entails modeling a team's offensive behavior by learning a Markov decision process (MDP) from event data that is collected from the team's matches, and (2) a reasoning component, which involves a novel application of probabilistic model checking to reason about the efficacy of the learned strategies of each team. I will illustrate the framework on one use case, namely that it can be used to optimise a team's defensive strategies when playing against a particular team.

3.19 Discovering Tactics from Team Sports Data

Albrecht Zimmermann (Caen University, FR)

Team sports tactics are about who (which player and/or player role) does what (in terms of legal actions, which includes moving around) where on the pitch, court or whatever the playing field is called, potentially modified by when (in the shot clock, game clock etc).I used my presentation to quickly touch on those different aspects (excluding time) by summarizing a number of works from the literature. The lines are not always clearly drawn: some papers mix what athletes do with where it happens or with who does it. The "what" part is arguably the most complex one and the papers I touched on all use what I've called the "vocabulary" of possible actions/movements to describe what teams and individual athletes do. Some do so rather explicitly, using topic models originally developed for text document analysis, others learn or predefine patterns that occur in certain situations or for certain teams. Some learn the vocabulary from the available data, others predefine it and "only" learn the "phrases" that are being formed. When it comes to the "who", finally, network-based modeling proves to be very powerful and allows to explore "what if" scenarios that would occur if one put different line-ups in the field.

3.20 Shape Descriptors Applied to Tactical Analysis in Football

Felipe Arruda Moura (State University of Londrina, BR)

The idea of the presentation is to introduce and discuss the use of shape descriptors related to team behaviour during football matches. In the literature, the surface area is usually represented by the area of a polygon obtained from the position of the teammates, and represents the total space covered by a given team on the pitch. Although the absolute values of the surface area represent good parameters for the characterization of teams' organization during a match, one can argue that different distributions of players on the pitch can provide equal surface areas or, for similar organization shapes, it is possible that the teams present different areas. Thus, the shape description of the polygon provides more in-depth information on the complexity of organization during the matches. Some shape descriptors will be presented and discussed about their validity. Also, Multiscale Fractal Dimension will be presented as an alternative for shape description. Finally, some applied results associating shape description with performance indicators will be presented.

4 Panel discussions

We organized two panels during the seminar. The topic of the first panel was "Team versus Individual Sports" and the panelists were Gabriel Anzer (RB Leipzig), Tim Chan (University of Toronto), Florentina Hettinga (University of Northumbria), and Stephanie Kovalchik (Zelus Analytics). The topics discussed during the panel included:

- What do tactics encompass in team and individual sports?
- In team sports, is there tension around salaries in terms of an athlete doing something that is for the good of the team (e.g., playing multiple different positions, playing out of position) that may hurt them individually?
- Is there a trend of team sports reaching out to experts in individual sports to get help with a specific skill or problem?
- At what point is a sport individual and what point is it team? Does this distinction matter?
- Is the type of analytics different between team and individual sports?
- What impact does gamification have in sport? Is this a positive or negative impact? Could playing a game translate to becoming a professional in some sports? How close does a game need to be to mimic real sport?
- Major League Baseball has a revenue of around $10 Billion and its teams employ around 500 analysts. For professional tennis, the annual revenue is around $1 Billion: Why are there not 50 tennis analysts?

The topic of the second panel was "Putting Analytics into Practice" and the panelists were Joris Bekkers (Freelance Data Analyst), Max Goldsmith (Royal Belgian Football Association), Sigrid Olthof (Liverpool John Moores University), and Darren O'Shaughnessy (St. Kilda Football Club). The topics discussed during the panel included:

- How do you handle communicating with domain experts in a practical setting? Do you explicitly undertake initiatives to raise data literacy in the sports organizations?
- How do you manage expectations, e.g., avoid over-promising?
- How do you deal with probabilistic outputs and uncertainty and communicating this to experts? Or how do you overcome the fact that people are bad at thinking statistically, understanding sample sizes, etc.?
- Working at a club can be very tenuous and uncertain as there can be significant turnover based on how a team performs. Strategically, how can you set yourself for the next job, particularly when much of the work you do is protected?
- Earlier on, there was often a mismatch between the types of problems considered by researchers and what the practitioners actually need? Does this gap still exist, or are there more researchers tackling problems that are directly applicable to practitioners?
- How has data analytics influenced coach/practitioners behaviour in the teams you've worked with?

5 Discussion Topics

5.1 Tactics vs Strategy

There was no final agreement on the definition of tactics and strategy after this break-out session and we also realised that the vocabulary is used differently across the world. For example, in North America the word tactics is barely used while in Europe the word tactics is more common. Most participants agreed that tactics have a shorter time-line than strategies. The outcome of the groups is as follows.

Group 1 defined tactics as a set of unconstrained actions and reactions that are anchored in a respective strategy and take the environment into account. In contrast, strategy is planned out and ranges on a longer term. This group also questioned whether it is possible to quantify the quality of a strategy.

Group 2 viewed strategy also as a long term, pre-planned idea of how a team aims to play. Tactics are decisions made on the field, where it makes a difference who takes the decision. While the strategy is decided by coaches, the tactics is rather a joint effort by players and coaches. The difference is also dependent on the actual sport and there may be fluent transitions but all strategies are pre-planned (like load management and player rotation) while this does not hold for all tactics.

Group 3 focused on athletics and skill action plans as tactical knowledge. The group also agreed that tactical decisions are very short term and considered a continuum that ranges from one player, via groups of players, and the team to a game and finally to the entire season and the philosophy of the coach. There are certainly action plans with various time scales, however, strategy can be viewed as long term tactics.

Group 4 also aimed to draw a line between tactics and strategy. The long term strategy was placed at a business level, however discussions went about where to draw the line for in-game decisions? The group agreed that strategy is decided on before the game while tactics are adjustments of the strategy during the game.

5.2 Common Data Format

Event data or play-by-play data is a common type of data that is collected about many different team sports such as soccer, ice hockey, and rugby among others. This type of data records specific semantically meaningful events that occur during a match. For example, relevant events in soccer include passes, tackles, and shots. Each event is annotated with information such as the players involved and the location where the event took place. Such data forms the backbone of many different analysis tasks.

However, event data can be challenging to work with because it is collected by multiple different providers in each sport and each provider often records the data in a different format. This makes working with similar data from different providers tedious for practitioners because each data source usually has to be converted or mapped into a unified format. An even bigger challenge is that each provider may use different definitions for certain events, particularly since some events are inherently subjective.

Less critical, but also challenging in the domain of soccer is the comparison of different tracking data sources. While the format of the data-set itself is somehow restricted to x/y/z-coordinates of either the center of mass or different body-points, different vendors share the data also in different formats, and use different preprocessing algorithms.

Gabriel Anzer, Pascal Bauer and Joshua Smith are involved in a project called the **Common Data Format** (initiated by FIFA and the DFB) that will attempt to address this challenge for soccer. They presented the key ideas underlying the format, which strives for a touch-based model that incorporates both event and positional data. Moreover, they are committed to providing clear definitions of events that companies can use to collect data in the proposed format.

Most participants had experienced some of the aforementioned issues. From an academic and club perspective, there is a clear need for such a unification as academia and clubs cannot reproduce or repeat their experiments on other data formats and comparability of approaches and results is an issue. However, from a provider perspective there is an appreciation that changing definitions is difficult from an operational perspective and creates legacy issues for data that was collected differently. Moreover, the data format also has to adhere to the needs of other user, such as the broadcasting media, which makes reaching an agreement on a joint format difficult.

5.3 Communication and Visualization

Communication between analysts and coaches can be difficult as there is a gap between the terms used by analysts and coaches, which can make it challenging for analysts to prepare what the coaches want. Martin Rumo calls this the "semantic gap". To help bridge this gap, he advocates for a co-creation approach between the analyst and user/coach.

More generally, several challenges exist that hinder the adoption of a common terminology. First, Jan Van Haaren noted that sometimes no term exists and people simply operate by describing the concept. In these cases one may need to invent a term. Second, terminology is often specific to a club and may stay consistent even with changes to personnel. Third, cultural factors affect terminology. For example, players come from different places and players from South American may use different terminology than those from Europe. Sigrid Olthof noted that this an issue in academic work too as papers from the computational literature may use, e.g., artificial intelligence specific terms instead of ones used in sports science. Fourth, misalignment of goals can lead to different terminology. For example, people may build their career around specific terminology and hence are not incentivized to change. One initiative in football (soccer) to help mitigate this problem is FIFA's common football language that will be used in coach & analyst courses.[1] As more people follow these training courses, it will help introduce a more standard vocabulary.

Beyond this, several pieces of advice that were offered included:

- It can be useful to provide a clip library to illustrate good / bad behavior.
- When making visualizations, avoid using colors associated with a rival.
- When discussing probabilities, it can be useful to explicitly state the chances of each open. For example, instead of just saying "There is an 80% chance that we reach the playoffs" also add "and there is a 20% chance that we do not reach the playoffs".
- The Pysport website contains a list of visualization packages: `https://opensource.pysport.org/?categories=Visualization`
- The following article summarizes some good advice about visualizations: `https://knowablemagazine.org/content/article/mind/2019/science-data-visualization`

[1] `https://www.fifatrainingcentre.com/en/resources-tools/football-language/`

5.4 Longitudinal Data

Currently most analysis is done on data from a few matches, or a few seasons. However, very interesting questions can be posed when looking at data that has been collected over many years. For example women's soccer has gone through a relatively quick development and is considered to have a very different style than the men's game. With longitudinal analysis it could be investigated if the women's game is developing to get closer to the men's game over time and if this development for example is going faster than how the men's game developed over the last 20 years. Other questions that could be answered with longitudinal analysis are around variability across seasons, changes in coaching styles and finally talent development. Having longitudinal data from developmental players starting when they join the academy at a very young age until they are retired can answers questions about talent development programs but also about how to maximize a players value/time at the club.

5.5 Context

It may be important when doing analysis to normalize data or exclude outliers because it does not help finding large trends in the data. However, if data on the context was collected, such as players injured or weather data new insights might be created from these outliers. Context also matters on a match level: For example, using event data alone allows track whereabouts of the ball, but not whether the ball possessing player is facing one, two, or even more opponents. The context of how many defenders the player faces influences where he/she will pass the ball to.

Participants

- Gabriel Anzer
Hertha BSC – Berlin, DE
- Felipe Arruda Moura
State University of Londrina, BR
- Pascal Bauer
Frankfurt am Main, DE
- Joris Bekkers
Breda, NL
- Luke Bornn
Zelus Analytics –
QSacramento, US
- Timothy Chan
University of Toronto, CA
- Jesse Davis
KU Leuven, BE
- Laura de Jong
Deakin University –
Melbourne, AU
- Uwe Dick
Sportec Solutions –
Unterföhring, DE
- Max Goldsmith
RBFA – Tubize, BE
- Florentina Hettinga
University of Northumbria –
Newcastle, GB
- Benjamin Holmes
University of Liverpool, GB
- Mamiko Kato
Toyo University – Tokyo, JP
- Matthias Kempe
University of Groningen, NL
- Hyunsung Kim
Seoul National University, KR
- Stephanie Kovalchik
Zelus Analytics – Austin, US
- Martin Lames
TU München, DE
- Daniel Link
TU München, DE
- Jim Little
University of British Columbia –
Vancouver, CA
- Patrick Lucey
Stats Perform – Chicago, US
- Jakub Michalczyk
Sportec Solutions –
Unterföhring, DE
- Darren O'Shaughnessy
St Kilda Football Club –
Moorabbin, AU
- Sigrid Olthof
John Moores University –
Liverpool, GB
- David Radke
Chicago Blackhawks, US
- Pegah Rahimian
Twelve Football – Stockholm, SE
- Yannick Rudolph
Leuphana Universität
Lüneburg, DE
- Martin Rumo
OYM AG – Cham, CH
- Nathan Sandholtz
Brigham Young University, US
- Raimund Seidel
Universität des Saarlandes –
Saarbrücken, DE
- Joshua Smith
Concordia University –
Montreal, CA
- Tim Swartz
Simon Fraser University –
Burnaby, CA
- Jan Van Haaren
FC Bruges, BE
- Maaike Van Roy
KU Leuven, BE
- Christoph Weber
DBS – Frechen-Buschbell, DE
- Hendrik Weber
DFL – Frankfurt, DE
- Albrecht Zimmermann
Caen University, FR

Report from Dagstuhl Seminar 24082

AI for Social Good

Claudia Clopath*[1], Ruben De Winne*[2], Mohammad Emtiyaz Khan*[3], and Jacopo Margutti*[4]

1 **Imperial College London, GB.** c.clopath@imperial.ac.uk
2 **Oxfam Novib – The Hague, NL.** ruben_de_winne@hotmail.com
3 **RIKEN – Tokyo, JP.** emtiyaz.khan@riken.jp
4 **510 / Netherlands Red Cross – The Hague, NL.** jmargutti@redcross.nl

Abstract

Progress in the field of Artificial intelligence (AI) and machine learning (ML) has not slowed down in recent years. Long-standing challenges like Go have fallen and the technology has entered daily use via the vision, speech or translation capabilities in billions of smartphones. The pace of research progress shows no signs of slowing down, and demand for talent is unprecedented. AI for Social Good in general is trying to ensure that the social good does not become an afterthought, but that society benefits as a whole. In this Dagstuhl Seminar, which can be considered a follow-up edition of Dagstuhl Seminars 19082 and 22091 with the same title, we brought together AI and machine learning researchers with non-governmental organisations (NGOs), as they already pursue a social good goal, have rich domain knowledge, and vast networks with (non-)governmental actors in developing countries. Such collaborations benefit both sides: on the one hand, the new techniques can help with prediction, data analysis, modelling, or decision making. On the other hand, the NGOs' domains contain many non-standard conditions, like missing data, side-effects, or multiple competing objectives, all of which are fascinating research challenges in themselves. And of course, publication impact is substantially enhanced when a method has real-world impact. In this seminar, researchers and practitioners from diverse areas of machine learning joined stakeholders from a range of NGOs to spend a week together. We first pursued an improved understanding of each side's challenges and established a common language, via presentations and discussion groups. Building on this foundation, we organised a hackathon around some existing technical questions within the NGOs to scope the applicability of AI methods and seed collaborations. Finally, we discussed topics that cut across the AI for social good field, such as how to properly evaluate AI models that are used for good.

Seminar February 18–23, 2024 – https://www.dagstuhl.de/24082
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Applied computing → Computers in other domains; Computing methodologies → Machine learning
Keywords and phrases artificial intelligence, interdisciplinary, machine learning, non-governmental organizations, social good
Digital Object Identifier 10.4230/DagRep.14.2.182

1 Executive Summary

Ruben De Winne (Oxfam Novib – The Hague, NL)

AI and ML have made impressive progress in the last few years. Long-standing challenges like Go have fallen and the technology has entered daily use via the vision, speech or translation capabilities in billions of smartphones, and more recently via general uptake of software

* Editor / Organizer

AI for Social Good, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 182–190
Editors: Claudia Clopath, Ruben De Winne, Mohammad Emtiyaz Khan, and Jacopo Margutti
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

applications built on large language models. The pace of research progress shows no signs of slowing down, and demand for talent is unprecedented. But as part of a wider AI for Social Good trend, this seminar wanted to contribute to ensuring that the social good does not become an afterthought in the rapid AI and ML evolution, but that society benefits as a whole.

The five-day seminar brought together AI and ML researchers from various universities with representatives from NGOs pursuing various social good goals, such as providing legal aid, providing humanitarian assistance, advocating for gender justice, denouncing growing levels of inequality, and defeating poverty. On these topics, NGOs have rich domain knowledge, just like they have vast networks with (non-)governmental actors in developing countries. Mostly, NGOs have their finger on the pulse of the challenges that the world & especially its most vulnerable inhabitants are facing today, and will be facing tomorrow.

The objective of the seminar was to look at these challenges through an AI and ML lens, to explore if and how these technologies could help NGOs to address these challenges. The motivation was also that collaborations between AI and ML researchers and NGOs could benefit both sides: on the one hand, the new techniques can help with prediction, data analysis, modelling, or decision making. On the other hand, the NGOs' domains contain many non-standard conditions, like missing data, side-effects, or multiple competing objectives, all of which are fascinating research challenges in themselves. And of course, publication impact is substantially enhanced when a method has real-world impact.

The seminar facilitated the exploration of possible collaborations between AI and ML researchers and NGOs through a two-pronged approach. This approach combined high-level talks & discussions on the one hand with a hands-on hackathon on the other hand. High-level talks & discussions focused first on the central concepts and theories in AI and ML and in the NGOs' development work, before diving into specific issues such as generalizability, data pipelines, and explainability. These talks and discussions allowed all participants – in a very short timeframe – to reach a sufficient level of understanding of each other's work. This understanding was the basis to then start investigating jointly through a hackathon how AI and ML could help addressing the real-world challenges presented by the NGOs. At the start of the hackathon, an open marketplace-like setting allowed AI and ML researchers and NGOs to find the best match between technological supply and demand. When teams of researchers and NGOs were established, their initial objective was not to start coding, but to define objectives, assess scope and feasibility.

The intense exchanges during the hackathon allowed NGOs with a lower AI/ML maturity increased to increase understanding of the capabilities of AI/ML and define actions to effectively start working with AI/ML. NGOs that already had a more advanced understanding and use of AI/ML technology prior to the seminar, could take their AI maturity to the next level by trying out new ML approaches, designing and testing tailored ML models, or simply exploring new partnerships. Key to this success of the hackathon – and the seminar at large – was the presence of AI/ML experts whose respective fields of expertise could seamlessly be matched with the various needs of the various NGOs. This excellent group composition also facilitated a productive discussion about topics that cut across the AI for social good field, such as how to properly evaluate AI models that are used for good.

2 Table of Contents

3 Overview of Talks

3.1 Potential Generic Tools for AI for Social Good: Multi-objective Optimization & AutoML

Frank Hutter (Universität Freiburg, DE)

Professor Hutter presented the potential of AutoML for Multi-objective Optimization in the context of AI for Social Good. Setting the scene, professor Hutter used the example problem of algorithmic fairness in face recognition. Facial recognition systems are widely acknowledged to be prone to bias, particularly along sociodemographic dimensions such as gender and race. Given their pervasive use in sensitive applications like law enforcement for suspect identification and missing person tracking, there is a pressing need to address this bias. Professor Hutter explored the challenge of improving the fairness of face recognition algorithms while maintaining high accuracy rates. A key strategy discussed is multi-objective optimization, which involves balancing competing objectives such as minimizing errors while reducing bias. This approach mirrors similar tradeoffs found in other domains like food production and language generation algorithms such as GPT. By leveraging multi-objective AutoML (Automated Machine Learning), it becomes possible to develop AI systems that are not only performant but also fair, calibrated, energy-efficient, and robust. The advantages of AutoML are highlighted, including its ability to streamline ML application development, ensure reproducibility and transparency, and potentially surpass human performance on various tasks. Also, especially relevant in the AI for Social Good context, is the potential of unblocking applications because AutoML can remove the requirement for a (scarce, highly-paid) human ML expert in the inner loop. However, it's cautioned that while tools for single-objective tasks and tabular data are mature, multi-objective AutoML tooling is still somewhat in its early stages. Moreover, the presentation emphasized the importance of considering the broader socio-technical context in which AI systems operate. While technical solutions like AutoML offer promise, they must be integrated thoughtfully within the larger socio-technical system to truly address issues of fairness and bias in face recognition and other AI applications.

3.2 Method seeds

Mohammad Emtiyaz Khan (RIKEN – Tokyo, JP)

Mainly for the NGO participants who lacked an in-depth comprehension of what Machine Learning entails, professor Khan gave a talk on the main ML concepts and methods. He suggested his definition of ML ("making a computer intelligent without explicitly programming it"), noted a number of historical and recent success stories as well as notorious failure cases, and made participants aware of some key challenges (e.g. existing methods require a large amount of 'good-quality' data). He also explained main divisions within the ML realm and the most important methods ((un)supervised learning, reinforcement learning, . . .).

3.3 AI for Data, Models, Decisions

Subhransu Maji (University of Massachusetts – Amherst, US)

Professor Maji started his presentation by explaining that billions of birds migrate every year, mostly in the cover of darkness. But these birds are visible on RARD networks in the continental US. Subhransu explained that using Machine Learning, we can learn how migration has changed over the last 25+ years. A team of ecologists and computer scientists worked together to analyze this bird migration data at scale. Challenges and unique opportunities that this collaboration had were also discussed. Subhransu concluded with three pieces of advice to the participants:

- Don't throw away noisy data! You might be able to correct for noise.
- Don't throw away info on who labeled data!
- Don't throw away intermediate things, might be useful at some point for training.

Professor Maji also highlighted the iNaturalist application of AI for biodiversity mapping. He asked the question what if AI is not reliable? Many applications require a total count – e.g., how many birds migrate in a year, or how many damaged buildings are in city / county / state. Estimates can be biased (sometimes by a lot!).

3.4 A brief history of NLP

Virginia Patridge (University of Massachusetts – Amherst, US)

To complement professor Khan's presentation, Virginia Partridge zoomed in on the topic of Natural Language Processing (NLP). She explained the difference between generative and predictive NLP tasks, gave an overview of the history of NLP, and walked participants through an overview of NLP models.

3.5 Machine Learning for Peace: Digital Tools for Civic Actors

Jeremy Springman (University of Pennsylvania – Philadelphia, US)

Dr. Springman presented Digital Tools for Civic Actors from UPenn's Machine Learning for Peace lab. In general, the lab's approach for using data to contribute to crisis response is the following:

1. Awareness: data on what's happening very recently
   - Mass scraping online news + ML to track events
   - Interactive data dashboard
2. Planning: predictive analytics for strategic decisions
   - Forecasting political events
   - Civic Space Early Warning System

The lab uses online news from 300+ sources in 35 languages as input data. To ensure data quality, the focus is on reputable local sources. Overall, there is much better coverage than extant archivers/aggregators (GDELT, Wayback, Lexis Nexis, etc.). Dr. Springman concluded with a few concrete examples of detecting and forecasting civic events, such as arrests in Uganda or civic activism in Angola. In the future, the lab wants to be capable of forecasting external event data (Travel Advisories) and of extracting new information from text.

4 Working groups

4.1 Sustainability and ownership of AI models

Ruben De Winne (Oxfam Novib – The Hague, NL)

1. Common failures: A typical failure scenario involves enthusiasm during deployment, followed by eventual abandonment due to unforeseen issues and lack of ongoing usage.
2. Knowledge loss: High turnover rates, typically every 2-3 years, contribute to knowledge loss within organizations, affecting the sustainability of AI projects.
3. Maintenance challenges: While building prototypes may be exciting for students/researchers, finding skilled software engineers for basic maintenance is difficult, highlighting the importance of technical expertise.
4. NGO ownership: NGOs possessing the right profiles and resources can successfully own and maintain AI projects.
5. Business models: Offering digital products as services to other organizations can be a sustainable business model for AI applications.
6. Avoiding dependencies: Careful selection of services is crucial to avoid dependencies on technical partners who may package solutions in a way that limits maintenance without their continued support.
7. Partnerships for trust and resources: Partnerships with academia and other organizations can enhance trust, provide resources, and facilitate peer review of methods.
8. Communities of practice: Engaging in communities such as NetHope can provide opportunities for brainstorming and accessing resources.
9. University partnerships: Collaborating with universities, as seen with KoboToolbox, can yield successful outcomes in AI project sustainability.
10. Focus on reusable tools: Reusable tools are easier to manage and sustain as their impact can be easily demonstrated, facilitating resource allocation and support.

4.2 Evaluation of AI models

Daphne Ezer (University of York, GB) and Ruben De Winne (Oxfam Novib – The Hague, NL)

In the session on evaluating AI models, several key points were discussed:

- NGO Concerns: NGOs are concerned about ensuring AI does no harm.
- Metrics for Chatbots: There's ambiguity regarding evaluation metrics for Generative AI technologies like chatbots. Questions arise about whether chatbots are effectively answering questions and how well they're performing overall.
- Building Benchmarks: Developing benchmarks for chatbots involves defining basic proxy metrics such as factuality, toxicity, and relevance. However, the ultimate measure of success lies in user actions, like registration.
- A/B Testing: A/B testing involves comparing different versions of the chatbot to measure user engagement, typically assessed by how long users interact with it.
- Multi-Objective Optimization: To monitor progress towards goals, it's essential to define and measure multiple metrics specific to the use case.
- Avoiding Bias: Defining metrics before designing methods or models helps avoid confirmation bias.
- Iterative Improvement: Expect multiple iterations of testing and improvement to refine AI models.
- Beware of Overfitting: Using publicly available benchmarks might lead to overestimating model performance if the model has been trained on those benchmarks.
- Monitoring Performance: Implement alerts to detect significant changes in prediction distributions, especially for categorical outputs.
- Readiness for Production: Models are deemed ready for production based on predefined metrics, followed by extensive qualitative checks.
- Domain Knowledge: Evaluation now requires more domain knowledge than traditional feature engineering.
- Categorizing Errors: Errors should be categorized based on their impact, distinguishing between manageable and catastrophic mistakes.
- Focus on Worst Outcomes: Pay attention to extreme cases rather than just averages when evaluating model performance.
- Task-Specific Evaluation: Evaluation frameworks need to be tailored to specific tasks, making generalizable frameworks challenging to create.

5 Panel discussions

5.1 Feedback for the organizers of a follow-up AI for Social Good seminar at Dagstuhl

Claudia Clopath (Imperial College London, GB)

- Don't change too much because it works really good.
- Keep the pre-call between AI participants and NGO participants.
- Ensure that everyone is at the same starting point, e.g. by being clear on problem, sharing reading material on what ML is or Nature Comms paper, have a checklist for the prep call (e.g. re: data, problem, ...).
- Develop an MOOC or similar to make sure that everyone is at the same level.
- Be clear beforehand what the expectations are, so that people can prepare properly (explicitly that actually bringing data to the seminar could be helpful).
- Good to have a general introduction (Data science lifecycle, demystifying AI) session to start the week.
- Develop glossary in advance, maybe turn it into a fun quiz in the beginning.
- The case studies may be presented with some slides instead that only in an oral form.

Participants

- Asma Atamna
Ruhr-Universität Bochum, DE
- Annabelle Behnke
Deutsches Rotes Kreuz e.V. – Berlin, DE
- Siu Lun Chau
CISPA – Saarbrücken, DE
- Claudia Clopath
Imperial College London, GB
- Jorn Dallinga
WWF – Zeist, NL
- Ruben De Winne
Oxfam Novib – The Hague, NL
- Michael Dhatemwa
Oxfam Novib – The Hague, NL
- Daphne Ezer
University of York, GB
- Frank Hutter
Universität Freiburg, DE
- Roberto Interdonato
CIRAD – Montpellier, FR
- Mohammad Emtiyaz Khan
RIKEN – Tokyo, JP
- Isabell Klipper
Deutsches Rotes Kreuz e.V. – Berlin, DE
- Parvathy Krishnan
Analytics for a Better World – Amsterdam, NL
- Derek Loots
Médecins Sans Frontières – Amsterdam, NL
- Subhransu Maji
University of Massachusetts – Amherst, US
- Jacopo Margutti
510 / Netherlands Red Cross – The Hague, NL
- Marieke Meeske
Tilburg University, NL & Oxfam Novib – Den Haag, NL
- Krikamol Muandet
CISPA – Saarbrücken, DE
- N. N.
Internews – Washington, US
- Virginia Partridge
University of Massachusetts – Amherst, US
- Julia Proskurnia
Google – Zürich, CH
- Lennart Purucker
Universität Freiburg, DE
- Jake Robertson
Universität Freiburg, DE
- Andrés Roure Cuzzoni
Propel – Den Haag, NL
- Tom Schaul
Google DeepMind – London, GB
- Jeremy Springman
University of Pennsylvania – Philadelphia, US
- Maïna Vergonjanne
Droits Quotidiens Legal Tech – Montpellier, FR

Report from Dagstuhl Seminar 24091

Reflections on Pandemic Visualization

Daniel Archambault*[1], Fintan McGee*[2], Nicolás Reinoso-Schiller†[3], Tatiana von Landesberger*[4], and Simone Scheithauer*[5]

1 **Newcastle University, GB.** daniel.archambault@newcastle.ac.uk
2 **Luxembourg Inst. of Science & Technology, LU.** fintan.mcgee@list.lu
3 **Institute for Hospital Hygiene and Infectious Diseases, Universitätsmedizin Göttingen, DE.** nicolas.reinoso-schiller@med.uni-goettingen.de
4 **Universität Köln, DE.** tanat1979@gmail.com
5 **Institute for Hospital Hygiene and Infectious Diseases, Universitätsmedizin Göttingen, DE.** simone.scheithauer@med.uni-goettingen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar "Reflections on Pandemic Visualization" (24091). The fight against COVID-19 has highlighted the crucial role of data visualization and analytics, prompting significant innovations and collaborations. This Dagstuhl Seminar brought together experts from various fields to reflect on the lessons learned. The aim is to document and disseminate these insights, enhancing preparedness for future global health crises.

Seminar February 25 – March 1, 2024 – https://www.dagstuhl.de/24091

2012 ACM Subject Classification Applied computing → Health care information systems; Applied computing → Health informatics; Human-centered computing → Human computer interaction (HCI); Human-centered computing → Visualization application domains; Human-centered computing → Visualization techniques

Keywords and phrases Epidemiology, Pandemic, Preparedness, Visualisation

Digital Object Identifier 10.4230/DagRep.14.2.191

1 Executive Summary

Daniel Archambault (Newcastle University, GB)
Fintan McGee (Luxembourg Inst. of Science & Technology, LU)
Simone Scheithauer (Universitätsmedinzin Göttingen, DE)
Tatiana von Landesberger (Universität Köln, DE)

During the recent SARS-CoV-2 pandemic, visualizations were omnipresent, playing a central role in communicating with the public, drawing from multiple data sources, and serving diverse goals. In a short period, public health messages spread globally. In this Dagstuhl Seminar, we brought together 37 experts in visualization, mathematics, modeling, public health, infectious diseases, and psychology from Europe, Asia, Australia, and North America to summarize and discuss their personal insights gained over the three years of the pandemic.

Due to the heterogeneity of expertise and different tasks performed by each individual during the pandemic, and despite intensive collaboration and contact between experts in both fields over the past three years, it was felt necessary to establish some common ground

* Editor / Organizer
† Editorial Assistant / Collector

for the work and concepts of both medical and visualization experts. The first two days of the seminar included short talks on public health, infectious diseases, modeling, big data, and an overview of visualizations that participants used or appreciated during the SARS-CoV-2 pandemic. These introductory talks were specially designed to foster discussions and personal exchange, allowing significant time for in-depth conversations. After engaging discussions during and after the sessions, several key topics emerged across the medical and visualization fields that require deeper reflection. The organizers then clustered these relevant topics into five overarching areas of interest: the use of dashboards during the pandemic, communication to the public, preparedness, data visualization methodology in emergency responses and users tasks and medium.

Over the last two days of the seminar, breakout sessions of six participants were created to work on these key issues. The fruitful discussions in the breakout sessions had as their first output a presentation by each group summarizing their discussions. These discussions will be extended into a second output in the form of several publications that will appear in a Computer Graphics and Applications issue later this year which are now in preparation. Important connections were made between experts in healthcare-related disciplines and visualization specialists, addressing a significant need. As a result, one of the organizers was invited to participate in a symposium on automated surveillance of bloodstream infections in Germany, which was promptly accepted. We look forward to hosting more interdisciplinary seminars, which are extremely rewarding and valuable.

2 Table of Contents

3 Overview of Talks

3.1 Reflections on Pandemic Visualisation: An Introduction

Daniel Archambault (Newcastle University, GB)

Welcome to Dagstuhl! I hope you enjoy your seminar. This talk is an introduction to the seminar and its participants. We hope to facilitate time for writing retrospective articles in the area of pandemic visualisation to better respond to public health crises such as COVID-19. I hope you have a productive stay!

3.2 Aspects of pandemic preparedness

Johannes Dreesman (Niedersächsisches Landesgesundheitsamt – Hannover, DE)

The SARS-CoV-2-Pandemic brought many challenges to societies and governments. These challenges are highlighted from a public health agencies perspective. The lecture is structured according to the German pandemic preparedness plan, which adresses many aspects of the experiences made during the pandemics. Concerning surveillance, the mandatory surveillance system had to fulfill many tasks, which was not possible to completely fulfill with one system. In the course of the pandemic several more surveillance systems were established. Dashboard presentations are affected by the "rural district fallacy", leading to incident rates in districts with small populations having very high volatility and producing the most eyecatching results with little public health meaning. Concerning pandemic measures we have learned that non pharmaceutical interventions and particularly contact reducing measures have to be carefully evaluated to allow for withdrawal if they are not effective. Another challenge was the priorization of the vaccine provision and the monitoring of vaccine coverage and adverse effects. It requires well designed additional surveillance systems to fulfill these requirements. Visualisation methodology is key to support all these requirements coming up during a pandemic or comparable types of emergencies.

3.3 Visualization for Pandemics and Public Health: Review and Opportunities David Ebert

David S. Ebert (University of Oklahoma – Norman, US)

In this talk, I discussed the key roles in developing visualizations for public health, focusing on the data, who is using it, when they are using it, and what they are trying to do. This was highlighted with examples for each case. I then reflected on useful visualizations, their features, and how they were used. Finally, discussion on the challenges and opportunities of new sources of data and challenges in transforming this data into actionable information through visual analytic interfaces led to lively discussions.

3.4 Compartmental Modelling and the Pandemic Response in Wales

Biagio Lucini (Swansea University, GB)

The COVID-19 pandemic has resulted in huge strains on various aspects of our life. In Wales, the need to understand, adapt, and respond to the evolving situation has generated unprecedented challenges for the devolved health policies. As a first urgent response, the Technical Advisory Cell was created, which identified modelling as a high priority. This request led to the formation of the Swansea Modelling Team, a multidisciplinary team of Epidemiologists, Mathematicians, Biologists, Computer Scientists, and Research Software Engineers. Through numerical simulations that produce likely scenarios under evolving conditions, this modelling effort has been the main forward-looking input that has informed government policies and containment measures. In this talk, I will review how the team formed and went on to produce the earliest set of scenarios. Then, I will provide an overview of the underlying mathematical and computational methods and discuss the key results and findings. Finally, I will reflect on lessons learned and give an overview of the challenges that we are likely to encounter in a potential future pandemic.

3.5 COVID-19

Mathias Pletz (Universitätsklinikum Jena, DE)

This talk gave an overview about 5 medical topics related to the pandemics evolution of the virus, non-pharmaceutical interventions, differences between COVID-19 and influenza, medical management and vaccination. It gave insight about my experience with the implementation of mandatory masking in our hospital and the city of Jena – the first German city to implement masking. I explained the different medical strategies related to the time period of infection: early →antivirals, late state → steroids. Furthermore, it explained pros and cons of the vaccine and the impact of the selective pressure by immunity on the emergence of escape variants. Some key insights included that human behaviour and related cultural differences are hard to measure, but had a huge impact on viral spread and the death toll for different countries. In this regard, the impact of visualizations for public decision makers and the right communications strategy, e.g. how to communicate scientific insecurities to lay persons, was discussed.

3.6 Probably not famous last words...

Simone Scheithauer (Universitätsmedizin Göttingen, DE)

Now it's time for the probably not-so-famous last words. I have the honor of moderating the conclusion of our joint Dagstuhl Seminar. I have divided my contribution into three parts. I would like to start with a narrative that ran through our discussions and, I believe, through all the groups. It also seems to me to symbolize the pandemic: Time. Time as a resource, time as a symbol for working under pressure. But time is also an essential parameter that can influence the pandemic when it comes to when decisions are made or not. We had the quote that "Speed trumps perfection." This quote describes a long-known fact that I researched 10 years ago and used for an editorial during the pandemic. The quote is about doing the right thing instead of doing things right. It was written by Peter Drucker in 1963. It fundamentally captures the confusion between effectiveness and efficiency–between doing the right things and doing things right. I think we have struck a good balance between the two. My second part is an assignment that I, as co-organizer, would of course like to fulfill–I may and would like to remind us all to deliver our reports and, later on, manuscripts. I think all the Dagstuhl reports are already in. Nico can say something about that in a moment. We all received valuable information about the publications from Daniel yesterday and today. If you have any questions, please do not hesitate to ask us by email or other means. I am convinced that together we will succeed in delivering a good collection on our common theme. Now I come to the conclusion, which also revolves around the topic of time. Unfortunately, our time together here has come to an end, and I truly regret that. I had no idea what I was getting myself into and was a little apprehensive, even though I have already organized two congresses as president. But this is something completely different and unique. I have been able to meet many great people and learn a lot. I would like to say thank you for that. I would also like to thank Fintan, Daniel, and Tatiana for giving me the opportunity. And a big thank you to Tom, Max, and Nicolas, without whom we would not have managed the organization. And of course, my thanks go to you, because Dagstuhl can only be as good as the group is. From my point of view, this is fundamentally true, and I hope that only our time together ends here and that we stay in contact while working on the manuscripts and beyond. Thank you for everything, have a safe journey home, and please stay in touch!

3.7 Visualisation examples in the pandemic: The Good, the Bad and the Effective

Max Sondag (Universität Köln, DE)

Throughout the pandemic, a whole host of different visualisations were used. All participants of the seminar collected and submitted a collection of these that drew their interest or attention, with the goal of reflection on it as a whole and highlighting what made some special. Within the talk, we will identify what potential was there, where problems existed, and to find patterns in how and when they were used. The session focuses on facilitating and encouraging an open-ended discussion from the different perspectives of the participants. This will allow us to collect opinions and topics for the working groups.

Based on the results of this discussion, as well as the presentations and previous discussions, we will then identify the salient topics within pandemic visualisation that are worthwhile to reflect upon. After filtering, these topics will form the basis of our working groups in which we will report, reflect and publish upon in the days afterwards. Using the Tatiana-method of group selection, we then divide the participants into groups, reconvening at regular moments to realign and obtain perspectives for the group as a whole.

3.8 Pandemics: A case in point for Big Data

Antje Wulff (Universität Oldenburg, DE)

With the onset of the COVID-19 pandemic, Big Data was recognized as a "potentially powerful weapon" in the fight against the virus. Over time, researchers have realized that it is indeed possible to use data science and visualizations to help – however, we also have realized that it is not as easy as we might have thought in the beginning: Big Data is not available "out-of-the-box". In this talk, the characteristics and challenges of Big Data in pandemics are discussed by going through the Big Data V's and the Big Data pipeline. In particular, routine hospital data is discussed: it is heterogeneous, non-standardized and proprietary formatted, making it initially unsuitable for secondary uses such as pandemic visualizations. Standardized and (inter)national data platform approaches are needed to effectively use routine data. Current German and international initiatives cover this topic but, in the end, it remains highly intensive, interdisciplinary and resource-intense work. Based on such sharing platforms, innovative algorithms and visualizations can be developed and implemented in valuable decision-support tools. Finally, this talk leads to the question and discussion of what a "readiness" Big Data platform for pandemic visualization might look like, and what are the implications of these aspects for visualizations.

4 Working groups

4.1 Dashboard Group

Alessio Arleo (TU Eindhoven, NL), Rita Borgo (King's College London, GB), Jörn Kohlhammer (Fraunhofer IGD – Darmstadt, DE), Roy Ruddle (University of Leeds, GB), Holger Scharlach (Niedersächsisches Landesgesundheitsamt – Hannover, DE), and Xiaoru Yuan (Peking University, CN)

The Dashboard group united varied expertise from the visualization and medical informatics communities. The scope of this group is to reflect on the practical implications following the extensive use of dashboards during Covid-19 pandemic. During that time, dashboards have been used for different purposes and were designed for a wide variety of audiences. For example, they have been developed to inform the public about the spread of the pandemic in terms of cases and deaths since the very early stage of the pandemic. During the same time, they were also used by experts, e.g., for decision making and disaster mitigation. In the course of the pandemic more and more data sources were added e.g. patients in intensive care units, wastewater treatment data or genome comparison (variant identification). Besides dashboards internal daily or weekly reports were set up to inform decision makers.

For visualization researchers, this has been a massive challenge – and the collaboration with medical experts represented a great opportunity, but it wasn't short of misunderstandings and labored development. Overall, it is possible to conclude how dashboards played a pivotal role in informing people about the pandemic at every level: however, among several success stories, a few "monsters" spawned as well.

Within this context, the group investigated the impact of dashboard visualizations during the pandemic from an holistic perspective. The purpose of this group's work is not to draw a comprehensive landscape of the research on the dashboards developed and published during and shortly after the pandemic. Instead, our objective is to comment, dissect, and elaborate on the experience of experts, and how they worked with visualizations actually used on the front line.

Taking advantage of the unique setting of this Dagstuhl Seminar, we conducted semi-structured interviews with five fellow participants who worked in four distinct roles during the Covid pandemic – as clinicians, epidemiologists, health authorities, and communicating with journalists and the general public. They shared with us the types of dashboard they used, the aims and audience of their work, details of the dashboards and the data, issues and successes that occurred, the time required to learn how to use the dashboards, and how their general experience has fed through to their work today. We recorded their interviews and spent the remaining time discussing and clustering the commonalities and divergences of their respective stories.

The seminar atmosphere at Schloss Dagstuhl offered a rare and invaluable opportunity for its participants to not only engage in intellectual exchange but also foster an environment where social interaction and collaboration go hand in hand. This has had a direct impact on the quality and richness of material that was collected during interviews. Interview setting was both technical and colloquial, with both parties at ease in prompting questions as well as sharing experiences.

4.2 Data visualization methodology in emergency responses

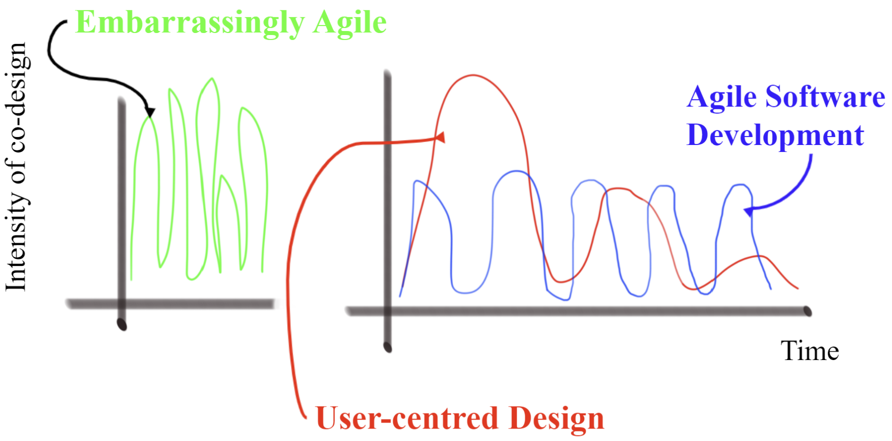
Barbora Kozlíková (Masaryk University – Brno, CZ), Daniel Archambault (Newcastle University, GB), Johannes Dreesman (Niedersächsisches Landesgesundheitsamt – Hannover, DE), Andreas Kerren (Linköping University, SE), Biagio Lucini (Swansea University, GB), Huamin Qu (HKUST – Hong Kong, HK), and Cagatay Turkay (University of Warwick – Coventry, GB)

Speed trumps perfection. Perfection is the enemy of the good when it comes to emergency responses – Dr Mike Ryan, March 14, 2020.

The main topic and goal of this group was to discuss the specifics of responding to an emergency situation (inspired by the COVID-19 pandemic case) by designing dedicated visual representations in a very limited time and with limited resources. In the discussions, we decided to focus solely on the target group of experts in epidemiology and infectious diseases in general, rather than on the general public, as these target audiences require very different solutions.

When designing appropriate visual representations in a standard visualization research project, we are always posing questions about the most suitable methodology for given data, tasks, and users. However, in an emergency response, when there is no time to select the most appropriate methodology from the existing ones, the questions we are posing are rather similar to "Has this methodology any chance to work reasonably well?".

Our discussions were inspired and based on three existing publications that focused on design study methodologies. In the paper by Sedlmair et al. [1], the authors discuss when the



Based on this observation, we built our proposed methodology on the concept of time *t* available for development, entitled *Agile(t) - visualization methodology as a function of time t*.

design study methodology is appropriate and propose a nine-stage framework that describes a linear but iterative process of designing visual representations. In a follow-up paper [3], a "lite" version of the previous methodology was proposed, where the first precondition stages collapsed into one and the core and analysis stages were partially overlapping. Publication [2] emerged as a reaction to the COVID-19 pandemic, where the authors designed a visual analytics approach for contact tracing policy simulations during an emergency response. Our aim within the seminar and also in the planned upcoming publication is to retrospectively evaluate the existing methodologies, summarize the lessons learned, and discuss the impact of time pressure on the methodology.

From experience, the development in an emergency situation follows the agile methodology. In the very first stage, it is crucial to build initial trust between the experts in infectious diseases and in visualization. To do so, the initial sketches and early prototypes should be developed as soon as possible. Inspired by diagrams in [4], we concluded that the development under time pressure resembles a very chaotic and almost randomly distributed intensity of co-design (meaning the intensity of collaborative efforts between the experts for infectious diseases and the visualization designers). We call this phase the embarrassingly agile one (green line in the graph below). The second graph shows a situation when already more time for development is available. Here, we can still see differences in the patterns – when still having limited time resources, the process resembles agile software development (blue line), whereas the red line captures the user-centered design approach, when there is enough time for discussions the designs and the created prototypes with the experts for infectious diseases. In this case, the iterations are converging to the final deployed solution much faster.

Based on this observation, we built our proposed methodology on the concept of time t available for development, entitled Agile(t) – visualization methodology as a function of time t.

Based on the time available, we are moving along the following "timeline", with a fully agile approach on the left side and a thorough methodological approach on the right side.

Embarrassingly Agile <——————————> Munzner Methodology [1]

As our methodology focuses on emergency response situations, we are focusing our interest on the left part of the above interval. To identify the presence and content of individual phases of the process described in Munzner's methodology [1], we looked at these phases from the perspective of available time. As already mentioned, if we have limited time for creating visualizations as an emergency response, we have to sacrifice several stages of the user-centered pipeline. We identified that only the most crucial stages – discover, design, and implement – are the core ones that are preserved in all situations. However, when the time available is very short (almost close to 0), we argue that even these stages are merged into one. This corresponds to the situation when the visualization designers with very limited input from the experts are trying to choose the most appropriate representation from the existing solutions. When time increases, the implementation stage starts to separate from the discover and design stages, as we start to utilize libraries and build more customized solutions. If we have even more time available, also the discover and design stages start to separate and we can incorporate more and more techniques that help to better analyze the problems and choose the most appropriate solution. However, the interconnection between the stages is still crucial, as it was proposed in the original Munzner's methodology.

Furthermore, we were discussing how to shorten time while preserving the quality of the solution. Here we identified that, for example, extracting guidelines for visual mapping that would guide the visualization developers to map certain tasks to appropriate visual encodings, would significantly contribute to this effort.

We also compiled a questionnaire asking about the experience with data, design, and development of visual representations and tools within the COVID-19 pandemic and distributed it among the seminar participants. The outcomes will be summarized as lessons learned in the upcoming publication.

References

1 Michael Sedlmair, Miriah D. Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2431–2440, 2012.

2 Max Sondag, Cagatay Turkay, Kai Xu, Louise Matthews, Sibylle Mohr, and Daniel Archambault. Visual analytics of contact tracing policy simulations during an emergency response. *Comput. Graph. Forum*, 41(3):29–41, 2022.

3 Uzma Haque Syeda, Prasanth Murali, Lisa Roe, Becca Berkey, and Michelle A. Borkin. Design study "lite" methodology: Expediting design studies and enabling the synergy of visualization pedagogy and social good. In Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik, editors, *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, pages 1–13. ACM, 2020.

4 Jo Wood, Alexander Kachkaev, and Jason Dykes. Design exposition with literate visualization. *IEEE Trans. Vis. Comput. Graph.*, 25(1):759–768, 2019.

4.3 "Getting ready": a visualisation perspective on pandemics

Fintan McGee (Luxembourg Inst. of Science & Technology, LU), Muna Abu Sin (RKI – Berlin, DE), Min Chen (University of Oxford, GB), David S. Ebert (University of Oklahoma – Norman, US), Kazuo Misue (University of Tsukuba, JP), Panagiotis Ritsos (Bangor University, GB), Tatiana von Landesberger (Universität Köln, DE), and Antje Wulff (Universität Oldenburg, DE)

Scope

The World Health Organisation defines pandemic preparedness as: "continuous process of planning, exercising, revising and translating into action..." [1]. In the context of visualisation this can be expressed in two main pillars: to devise and deploy visualisation technologies that are available during a pandemic, that can be deployed in quick-response mechanisms. to devise and deploy visualisation technologies for supporting operations and activities before and towards a future pandemic, to facilitate preparedness for when it arrives.

During the COVID-19 pandemic, we experienced a misbalance between perfection and speed [2]. As emergency situations were unfamiliar, mindsets were still prone to deliver perfect visualisation solutions rather than deploying a quick start. The time until the next pandemic hits can be used to prepare for a quicker response but also to expose shortcomings in advance, bringing us one step closer to perfection.

Our considerations of challenges include considerations on the difference in responses to the pandemic between low and high income countries. The approach followed in this report is to have an international outlook, as the synergies between populations and areas are impossible to ignore. Moreover, advances and lessons learned in different settings can inform the response and infrastructure at local, national and global level.

Challenges and opportunities In our working group, we identified the following challenges and opportunities for visualisation and related technologies towards improving our preparedness for the next pandemic:

Parameter optimisation for improved pandemic modelling, which can be improved by using visualisation approaches. The lack of standards-based data infrastructures and access that enables interoperability between different systems and services. Building awareness of the capabilities of visualisation in the expert domain remains an issue that affects the utilisation, adoption and development of visualisation tools. The need for in-depth requirements engineering including participatory visualisation methods and communication between stakeholders and developers. There are diverse methods for evaluating visualisation approaches with non-standardised indicators not quickly or automatically applicable. The integration of future technologies, ranging from mobile communications and tracking to Augmented Reality, personalised sensors, and pervasive / ubiquitous displays, which offer new opportunities for visualisation and public information. Ensuring the developed tools, approaches and solutions are transferable to other emergency situations, beyond pandemics.

For each challenge, we discussed the following three questions: What are the details of the challenge? Based on our experience and existing research in the current pandemic, why is it challenging? What approach could be proposed?

References

1 World Health Organisation. (2011) Pandemic Preparedness. [Online]. Available https://www.who.int/europe/news-room/fact-sheets/item/pandemic-preparedness. Accessed: 29/02/24.

2 J. Dykes et al., "Visualization for epidemiological modelling: challenges, solutions, reflections and recommendations," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 380, no. 2233, Oct. 2022.

4.4 Reflections on Pandemic Visualizations: Communication, Behaviour and Reactions

Nicolas Reinoso-Schiller (Universitätsmedinzin Göttingen, DE), Tatiana Losev (Simon Fraser University – Surrey, CA), Nicolas Medoc (Luxembourg Inst. of Science & Technology, LU), Klaus Mueller (Stony Brook University, US), Max Sondag (Universität Köln, DE), Yong Wang (SMU – Singapore, SG), Michael Wybrow (Monash University – Clayton, AU), and Hajo Zeeb (BIPS – Bremen, DE)

The main topic and goal of this group was to reflect on how pandemic (COVID-19) visualizations were communicated to the general public, especially with respect to how the communication happened, how they influenced behaviour and the reaction of the public. The group consists of a diverse team comprising epidemiologists, psychologists, data analysts, and experts in data visualization collaborates on infection prevention and control initiatives both within hospitals and in the public sphere. Their responsibilities extend to creating informative visualizations for public awareness campaigns, providing consultation to government officials, and delivering lectures at universities. This reflection's intended audience consists mainly of other visualization experts, but also government officials and anyone who wants to better understand what might be the key point of the pandemic reflection.

To explore the topic and the different perspectives from the various group members, we opted for an individualized reflection before discussing these reflections with the entire group to find larger and broader patterns. We used the following set of initial questions as a starting point for reflection, broadly categorized in experience, and future outlooks.

Experience:

- What is the relationship between your work and public behaviour / response during the pandemic?
- My thoughts, impressions, experiences?
- My experiences from my work?
- Who would I choose to work with for the next pandemic?
- What did you find to be an effective visualization? Example?
- How do you think visualization impacted misinformation?
- Future outlooks
- What would you change in communication to the public for next time?
- What would I change for the next pandemic if there were no constraints?

We discussed the individual reflections on these topics centrally over the next two days. From this, several core topics (Personalisation, Communication pathways, Intentions and actuality, Targeting) emerged, which we briefly highlight here:

(1) Personalisation. Visualizations during the pandemic were often aimed at abstract information and abstract target audiences. For example: How fast is the diseases spreading? What is the hospital bed occupation? How many people are currently infected? Which groups are getting infected most? However, a missed opportunity was to develop personalized visualizations taken the perspective of the individual. One could for example take in the group one belongs to and tailor the visualization to highlight their personal risks and COVID situation in their surroundings.

(2) Communication pathways. During the pandemic, we all noticed that we had to skip some steps in the standard methodologies for the purpose of speed. In particular, users were included only late in the design of the visualization. We discussed several ways that the group members encountered this, as well as potential solutions to establish communication pathways early and play to the strengths of the communities we are trying to reach.

(3) Intentions and actuality. Visualizations were used differently during the pandemic than how they were intended to be used by the designers. We discussed various situations where intentions did not match the actualities. While it is not always an issue if visualizations were differently, sometimes this can have severe consequences such as when taking visualizations designed for experts out of context and presenting it to the public to misinform.

(4) Targeting. Many visualizations were made and presented without specific target audiences in mind beyond the "general public". However, there is a very large breadth of audiences within the general public, and many visualizations could have benefitted from targeting more specific groups. We identified that specific communities within the general public have different needs for visualization information, as well as potential risks of stigmatization that can occur when developing for such a targeted audience.

We discussed these 4 core topics in depth during the group working sessions at Dagstuhl, and they will form the basis for envisioned People and Practices' article for Computer Graphics and Applications.

4.5 Empowering Communities: Tailored Pandemic Data Visualization for Varied Tasks and Users

Nikita Srivastava (Universitätsmedinzin Göttingen, DE), Tom Baumgartl (Universität Köln, DE), Mohammad Ghoniem (Luxembourg Inst. of Science & Technology, LU), Georgeta Elisabeta Marai (University of Illinois – Chicago, US), Silvia Miksch (TU Wien, AT), Sibylle Mohr (University of Glasgow, GB), and Simone Scheithauer (Universitätsmedinzin Göttingen, DE)

Data visualization concepts and methodologies were intensively leveraged during the COVID-19 pandemic. However, a systematic investigation into the needs of the particular users, their data, and their tasks, as well as the visual media, is still missing. We review our design experience working across six countries and over interdisciplinary COVID-19 pandemic projects.

We describe the challenges we met in these projects, characterize the user communities served by these projects, the goals, and tasks we supported, the data types and visual media we worked with. Furthermore, we instantiate these characterizations in a series of case studies covering the known purposes of visualization: exploratory analysis, confirmatory analysis, and presentation for communication during the pandemic. Finally, we describe the Visual Analytics lessons we learned, considering future pandemics.

Participants

- Muna Abu Sin
RKI – Berlin, DE
- Daniel Archambault
Newcastle University, GB
- Alessio Arleo
TU Eindhoven, NL
- Tom Baumgartl
Universität Köln, DE
- Rita Borgo
King's College London, GB
- Min Chen
University of Oxford, GB
- Johannes Dreesman
Niedersächsisches Landesgesundheitsamt – Hannover, DE
- David S. Ebert
University of Oklahoma – Norman, US
- Mohammad Ghoniem
Luxembourg Inst. of Science & Technology, LU
- Andreas Kerren
Linköping University, SE
- Jörn Kohlhammer
Fraunhofer IGD – Darmstadt, DE
- Barbora Kozlíková
Masaryk University – Brno, CZ
- Tatiana Losev
Simon Fraser University – Surrey, CA
- Biagio Lucini
Swansea University, GB
- Georgeta Elisabeta Marai
University of Illinois – Chicago, US
- Fintan McGee
Luxembourg Inst. of Science & Technology, LU
- Nicolas Medoc
Luxembourg Inst. of Science & Technology, LU
- Silvia Miksch
TU Wien, AT
- Kazuo Misue
University of Tsukuba, JP
- Sibylle Mohr
University of Glasgow, GB
- Klaus Mueller
Stony Brook University, US
- Mathias Pletz
Universitätsklinikum Jena, DE
- Huamin Qu
HKUST – Hong Kong, HK
- Nicolas Reinoso-Schiller
Universitätsmedinzin Göttingen, DE
- Panagiotis Ritsos
Bangor University, GB
- Roy Ruddle
University of Leeds, GB
- Holger Scharlach
Niedersächsisches Landesgesundheitsamt – Hannover, DE
- Simone Scheithauer
Universitätsmedinzin Göttingen, DE
- Max Sondag
Universität Köln, DE
- Nikita Srivastava
Universitätsmedinzin Göttingen, DE
- Cagatay Turkay
University of Warwick – Coventry, GB
- Tatiana von Landesberger
Universität Köln, DE
- Yong Wang
SMU – Singapore, SG
- Antje Wulff
Universität Oldenburg, DE
- Michael Wybrow
Monash University – Clayton, AU
- Xiaoru Yuan
Peking University, CN
- Hajo Zeeb
BIPS – Bremen, DE

Report from Dagstuhl Seminar 24092

Applied and Combinatorial Topology

Paweł Dłotko*[1], Dmitry Feichtner-Kozlov*[2], Anastasios Stefanou*[3], Yusu Wang*[4], and Jan F Senge†[5]

1 **Polish Academy of Science, PL.** pdlotko@impan.pl
2 **Universität Bremen, DE.** dfk@uni-bremen.de
3 **Universität Bremen, DE.** stefanou@uni-bremen.de
4 **University of California, San Diego – La Jolla, US.** yusuwang@ucsd.edu
5 **Universität Bremen, DE.** janfsenge@uni-bremen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24092 "Applied and Combinatorial Topology".

The last twenty years of rapid development of Topological Data Analysis (TDA) have shown the need to analyze the shape of data to better understand the data. Since an explosion of new ideas in 2000' including those of Persistent Homology and Mapper Algorithms, the community rushed to solve detailed theoretical questions related to the existing invariants. However, topology and geometry still have much to offer to the data science community. New tools and techniques are within reach, waiting to be brought over the fence to enrich our understanding and potential to analyze data. At the same time, the fields of Discrete Morse Theory (DMT) and Combinatorial Topology (CT) are developed in parallel with no strong connection to data-intensive TDA or to other statistical pipelines (e.g. machine learning).

This Dagstuhl Seminar brought together a number of experts in Discrete Morse Theory, Combinatorial Topology, Topological Data Analysis, and Statistics to (i) enhance the existing interactions between these fields on the one hand, and (ii) discuss the possibilities of adopting new invariants from algebra, geometry, and topology; in particular inspired by continuous and discrete Morse theory and combinatorial topology; to analyze and better understand the notion of shape of the data.

The different talks in the seminar included both introductory talks as well as current research expositions and proved fruitful for the open problem and break-out sessions. The topics that were discussed included

1. algorithmic aspects for efficient computation as well as Morse theoretic approximations
2. topological information gain of multiparameter persistence
3. understanding the magnitude function and its relation to graph problems.

Seminar February 25 – March 1, 2024 – https://www.dagstuhl.de/24092

2012 ACM Subject Classification Information systems → Data structures; Theory of computation → Computational geometry; Mathematics of computing → Discrete mathematics

Keywords and phrases Applied Topology, Topological Data Analysis, Discrete Morse Theory, Combinatorial Topology, Statistics

Digital Object Identifier 10.4230/DagRep.14.2.206

* Editor / Organizer
† Editorial Assistant / Collector

Applied and Combinatorial Topology, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 206–239
Editors: Paweł Dłotko, Dmitry Feichtner-Kozlov, Anastasios Stefanou, and Yusu Wang
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Paweł Dłotko (Polish Academy of Science – Warsaw, PL)
Dmitry Feichtner-Kozlov (Universität Bremen, DE)
Anastasios Stefanou (Universität Bremen, DE)
Yusu Wang (University of California, San Diego – La Jolla, US)

The Dagstuhl Seminar titled "Applied and Combinatorial Topology" brought together researchers in mathematics and computer science to engage in active discussions and exchange of ideas on theoretical, computational, and practical aspects of applied and combinatorial topology. The seminar has led to further the connections between the Discrete Morse Theory, Computational Topology, and Statistics communities and identification of open problems that can be addressed together.

Context

Applied Topology is a new and rapidly increasing research field within applied mathematics. Its main focus is to utilize topological methods to solve applied problems. The common emphasis of the methods in Applied Topology is on the computational aspect. Application areas include: data analysis, computational biology, network analysis, graph visualization and reconstruction, feature selection, and many more.

Goals

The Dagstuhl Seminar 24092 (February 25–March 1, 2024) brought together three research communities, namely researchers in discrete Morse theory (DMT), computational topology, and statistics. The aim was to facilitate collaborations that could strengthen the existing interactions of the fields (e.g. Reeb graphs, Mappers and discrete vector fields) and collaborations that may lead to the development of new descriptors of data (e.g. persistence invariants, magnitude functions, etc.), which in turn have the potential to be inputted into statistical methodologies and to provide their efficient implementations.

Topics

We chose three research topics for which the respective communities will benefit from a knowledge exchange and mutual discussion.

Discrete Morse theory. The research field of discrete Morse theory, developed by R. Forman, is a discrete counterpart of a continuous Morse theory. It has recently found many practical applications both within mathematics (e.g. configuration spaces, and homology computation) as well as outside mathematics, such as computer science (e.g. denoising and mesh compression). The target of discrete Morse theory is to construct a discrete vector field that either simplifies the data at hand, without losing its important features, or to introduce discrete dynamics on the data. The resulting dynamics can be further analyzed to extract certain interesting sets, for instance invariant sets. In both cases, the aim is to simplify the data or to find the important regions of the data. One example of use of Morse theory in data analysis is the Reeb graph of a Morse function. A discrete adaptation of Reeb graphs which uses ideas of partial clustering is known as Mapper and has seen a great deal of success in data analysis. Another example is an adaptation of discrete Morse theory to computations of persistent homology in topological data analysis: the machinery of discrete Morse theory can be used to help reduce the complexity of the evolving topology in the filtrations of datasets. Moreover, it was recently shown that discrete Morse theory can be utilized for simplification and complexity reduction also in the multiparameter persistence setting. Furthermore, discrete Morse theory has been studied in conjunction with persistent homology theory and has found interesting applications, such as reconstruction of grayscale digital images and reconstruction of graphs, see for instance 2D road reconstruction and 3D neuron reconstruction. Another powerful application of Discrete Morse theory is in distributed computing. In both cases, the discrete Morse theory is used to simplify data at hand, and recover their invariants. We believe that using this machinery, one can do even more. We would like to take the paradigm of discrete Morse theory further and directly try to recover certain invariants from the constructed discrete vector fields.

Computational Topology. The research field of combinatorial topology originated from the study of topological invariants derived from combinatorial decompositions of spaces (cf. simplicial approximation theorem), known as simplicial complexes. One of the main examples of such invariants are the Betti numbers. Combinatorial topology was later named algebraic topology due to the switch of focus of the field on its algebraic aspects (as homology groups), which is attributed to Emmy Noether. In the research area of computational topology (also known as topological data analysis), we are interested in studying a single parameter filtration of complexes associated with a data set (viewed as finite metric space) such as the Vietoris-Rips filtration, or multiparameter filtrations of complexes associated to datasets, such as the function-Rips bifiltration and the multicover bifiltration. Those structures can be simplified with the tools of discrete Morse theory. A homology functor is then being applied to those filtrations resulting in a single or multiparameter persistence module. Single-parameter persistence modules are visualized by their persistence diagrams. A well-known invariant of multiparameter persistence modules is the rank invariant which captures important persistence information about multifiltrations of datasets. Recently, there have been some refinements of the rank invariant and also a generalization of the notion of persistence diagram (induced by the rank invariant). Developing algorithms for the efficient computation of multiparameter persistence modules and their rank invariants, is one of the big challenges of computational topology and topological data analysis (TDA).

Statistics in Topological Data Analysis. Persistence invariants such as the persistence diagram are equipped with a family of metrics, e.g. the ℓ^p-Wasserstein distances and the bottleneck distance. To make these signatures applicable, one must interface them with standard statistical methods. This has already been done e.g. when developing statistics on persistent diagrams, or other signatures such as persistence landscapes. However, much remains unknown in the case of limits of persistence diagrams when the number of points goes to infinity. A good example of a successful synergy between statistics and combinatorial topology is a process of vectorization of persistence diagrams. This process allowed the community to build multiple applications of persistent homology into many branches of science and engineering. We believe that, if new invariants originated from discrete Morse theory and combinatorial topology are introduced, such as the recently introduced Mapper graph of datasets, a work needs to be done, to incorporate them into existing statistical pipelines, hypothesis testing methods and similar. Moreover, a vectorization method for Mapper graphs needs to be established and their limit behavior (when e.g. the number of points goes to infinity) need to be studied. Also an application in standard statistics will be further explored; It is widely known that one should not rely on summary statistics, but always attempt to visualize the data. However, oftentimes the data are very high dimensional. In this case, Mapper type algorithms may serve as a surrogate of a scatter plot in visualization by providing a graph–based summary of the data. Our aim will be to explore this connection and look for ways of inputting Mappers into standard statistical pipelines, e.g. including concepts of averages and central limit theorems. We will also explore the connections between Mapper and other combinatorial topology concepts via, for instance, discrete Morse theory.

Participants, Schedule, and Organization

The attendees were strongly encouraged to prepare talks that will include open problems and new research directions. The program for the week consisted of talks of different lengths, open problem sessions, breakout sessions, and summary sessions with the participants. On Monday, we started with an 1-hour session where the participants introduced themselves, and then we had 6 introductory talks, two on Discrete Morse theory, two on computational topology and two on Statistics in TDA. Then, we had an open problem session where participants identified certain open problems and directions for research for the breakout sessions.

Participants chose one or more from the following proposed topics for breakout sessions:

1. Can we compute representatives of generators of persistent homology in less than cubic time? (proposed by Tamal Dey)
2. Optimal Discrete Morse function given a partial matching (proposed by Yusu Wang)
3. Topological information, i.e. "how much topological information remains when going from one to two dimensional filtrations (or from Reeb graphs to Reeb spaces)" (proposed by Bei Wang Phillips)
4. Manifold reconstruction guarantees (proposed by Ulrich Bauer)
5. Algorithmic questions on (multiparameter) persistence (proposed by Fabian Lenzen)
6. Can TDA detect planted cliques? (proposed by Bastian Rieck)
7. Monotonicity of magnitude functions of Euclidean metric spaces (proposed by Sara Kalisnik)
8. General applied topology (proposed by Dmitry Feichtner-Kozlov).

Tuesday to Thursday in the morning we had the lecture talks and we organized breakout sessions on Tuesday and Thursday afternoon. We reserved three rooms for the breakout sessions that ran in parallel, the main seminar room for topics (1)–(5), another room for topics (6)–(7), and a small room for topic (8). On Wednesday afternoon we organized some groups for hiking near Schloss Dagstuhl. Representatives from the working groups summarized the discussions during their breakout sessions and presented it to all participants on Thursday evening and Friday morning.

Results and Reflection

The seminar successfully facilitated a rich exchange of ideas and expertise among participants. The varied program, including talks, open problem sessions, breakout discussions, and outdoor activities, created an environment conducive to collaborative exploration. Attendees expressed satisfaction with the content and structure of the seminar, indicating a strong interest in future editions. During the breakout sessions, it was encouraging to note that some participants reported preliminary results related to the open problems presented. These early findings sparked lively discussions and provided valuable insights into potential directions for further research. The seminar served as a platform not only for sharing existing knowledge but also for generating new ideas and approaches.

2 Table of Contents

3 Overview of Talks

3.1 Discrete Morse theory and persistent homology of geometric complexes

Ulrich Bauer (TU München, DE)

I will discuss the interplay between geometry and topology, and between Morse theory and persistent homology, in the setting of geometric complexes. This concerns constructions like Rips, Čech, Delaunay, and Wrap complexes, which are fundamental construction in topological data analysis. The tandem of Morse theory and homology shows the topological equivalence of several of these constructions, helps in speeding up their computation by a huge factor (in the software Ripser), reveals thresholds at which homology necessarily vanishes (with links to a classical result by Rips and Gromov), and relates optimal representative cycles for persistent homology to the industry-tested Wrap reconstruction algorithm.

3.2 (Discrete) Morse Theory and Inverse Problems

Julian Brüggemann (Universität Bonn, DE)

Morse theory and its discrete version are well established toolboxes in pure topology. They both serve a similar purpose: use the combinatorics of the real numbers via well-behaved real-valued functions to compute topological invariants of geometric objects. In some instances, certain collections of topological invariants allow for a complete classification of the given class of spaces, which in turn might allow for a reconstruction of the original objects from the computed collection of invariants, most time up to some suitable notion of equivalence. In this talk, I will give a brief overview over smooth and discrete Morse theory and mention some classification results in topology as well as solutions to inverse problems in TDA.

3.3 A Statistical Perspective on Multiparameter Persistent Homology

Mathieu Carrière (Centre Inria d'Université Côte d'Azur – Sophia Antipolis, FR)

Multiparameter persistent homology is a generalization of persistent homology that allows for more than a single filtration function. Such constructions arise naturally when considering data with outliers or variations in density, time-varying data, or functional data. Even though its algebraic roots are substantially more complicated, several new invariants have been proposed recently. In this talk, I will go over such invariants, as well as their stability, vectorizations and implementations in statistical machine learning.

3.4 Computational Topology for Zigzag Persistence

Tamal K. Dey (Purdue University – West Lafayette, US)

In topological data analysis, zigzag persistence has become an important component because it enhances the applicability of persistence theory by allowing both insertions and deletions of simplices in a simplicial filtration. Such filtrations occur in applications where a space or a function on it changes over time. For example, in network analysis, new connections appear and existing connections disappear over time. The standard persistence algorithm for non-zigzag filtrations does not work for the zigzag case. After laying out the background and earlier work on computations of zigzag persistence, we present a new algorithm FastZigzag for computing zigzag persistence from an input filtration. We follow it with the discussion of the well known vineyard problem in the zigzag case. We present a recent efficient algorithm for computing the zigzag vineyard. Akin to the non-zigzag case, the special but important case of graphs allow certain optimizations that make the computations of zigzag barcode and their vineyards more efficient. We go over some of these developments. Finally, we indicate some of the applications of zigzag persistence, in particular to data analysis in TDA with multiparameter persistence.

3.5 Hypergraph Barcodes: a way to Link two Different Notions of Hypergraph Homology

Robert Green (University at Albany, US)

Hypergraphs are a natural data structure to consider when studying networks with multiway connections. One approach to characterizing the features of these networks involves defining a form of hypergraph homology and then leveraging these homological traits to delineate the hypergraphs. There are many different ways however to define hypergraph homology and different approaches yield different types of features. In this talk I will present two different approaches to this problem and then connect them by presenting a persistence module they both live inside of.

3.6 Merge Tree for Periodic Data

Teresa Heiss (Institute of Science and Technology Austria, AT)

Periodic data is abundant in material science, for example the atoms of a crystalline material repeat periodically. Additionally, periodic boundary conditions are used in many further applications, for example in cosmology, to remove boundary effects. It is unclear how to deal with the periodicity of the data when computing topological descriptors, like the merge tree or persistent homology. A classical approach is to compute the respective descriptor simply on the torus. However, this does not give the information needed for many applications and is in some sense even unstable under noise. Therefore, we suggest decorating the periodic merge tree gained from the torus with additional information, describing for each connected component how many components of the infinite periodic covering space map to it. The resulting periodic merge tree carries the desired information and fulfills all the desired properties, in particular: stability and efficient computability.

3.7 When Do Two Distributions Yield the Same Expected Euler Characteristic Curve in the Thermodynamic Limit

Niklas Hellmer (Polish Academy of Science, PL)

Joint work of Tobias Fleckenstein, Niklas Hellmer
Main reference Tobias Fleckenstein, Niklas Hellmer: "When Do Two Distributions Yield the Same Expected Euler Characteristic Curve in the Thermodynamic Limit?", CoRR, Vol. abs/2401.04580, 2024.
URL https://arxiv.org/abs/2401.04580

Given a probability distribution F on $\mathbb{R}^d$ with density f, consider a sample X_n of n points sampled from F i.i.d.. We study the Euler characteristic curve (ECC) of the union of balls $\bigcup_{x \in X_n} \overline{B}_{r_n}(x)$ in the thermodynamic limit. That is, as $n \to \infty$, we let $r_n \to 0$ such that nr_n^d approaches a finite, non-zero limit. It turns out that two distributions yield the same expected ECC in this setting if and only if they have the same excess mass. Whether this condition is also necessary for the distributions of the ECCs to coincide in the limit remains an open question.

3.8 Topological descriptors for efficient analysis of electronic structures

Ingrid Hotz (Linköping University, SE)

In this talk, I will present some of our ongoing work on visual analysis and comparison of electronic structures in molecules or crystals based on topological analysis. The application context is to develop novel materials with some desired properties by simulating material configurations from a large number of possible candidates. Our aim is to characterize such materials based on their electronic structure, represented by their electron density fields. Therefore, we have experimented with various multiscale descriptors that support quantitative and visual comparative analysis to help scientists better understand the differences in their structures through visual exploration guided by automatic analyzes such as outlier detection, clustering, and similar structure searches.

3.9 Magnitude, Alpha Magnitude and Applications

Sara Kalisnik (ETH Zürich – Zürich, CH)

Joint work of Sara Kalisnik, Miguel O'Malley, Nina Otter

Magnitude is an isometric invariant for metric spaces that was introduced by Leinster around 2010, and is currently the object of intense research, since it has been shown to encode many known invariants of metric spaces. In recent work, Govc and Hepworth introduced persistent magnitude, a numerical invariant of a filtered simplicial complex associated to a metric space. Inspired by Govc and Hepworth's definition, we introduced alpha magnitude. Alpha magnitude presents computational advantages over both magnitude as well as Rips magnitude, and is thus an easily computable new measure for the estimation of fractal dimensions of real-world data sets.

3.10 Graphcodes

Michael Kerber (TU Graz, AT)

Joint work of Michael Kerber, Florian Russold

We introduce graphcodes, a novel multi-scale summary of the topological properties of a data set that is based on the well-established theory of persistent homology. Graphcodes handle data sets that are filtered along two real-valued scale parameters. Such multi-parameter topological summaries are usually based on complicated theoretical foundations and difficult to compute; in contrast, graphcodes yield an informative and interpretable summary and can be computed as efficient as one-parameter summaries. Moreover, a graphcode is simply an embedded graph and can therefore be readily integrated in machine learning pipelines using graph neural networks. We describe such a pipeline and demonstrate that on data sets with rich topological features, graphcodes achieve better classification accuracy than state-of-the-art approaches.

3.11 The discriminating power of the generalized rank invariant

Woojin Kim (Duke University – Durham, US & KAIST – Daejeon, KR)

Joint work of Woojin Kim, Nathaniel Clause, Facundo Mémoli
Main reference Nate Clause, Woojin Kim, Facundo Mémoli: "The discriminating power of the generalized rank invariant", CoRR, Vol. abs/2207.11591, 2022.
URL https://doi.org/10.48550/ARXIV.2207.11591

In topological data analysis, the rank invariant is one of the best known invariants of persistence modules over posets. The rank invariant of a persistence module M over a given poset P is defined as the map that sends each comparable pair $p \leq q$ in P to the rank of the linear map $M(p \leq q)$. The recently introduced notion of generalized rank invariant acquires more discriminating power than the rank invariant at the expense of enlarging the domain of rank invariant to a collection I of intervals of P that contains all segments of P. In this talk, we discuss the tension that exists between computational efficiency and the discriminating power of the generalized rank invariant, depending on its domain I. The Möbius inversion formula will assume a significant role in clarifying the discriminating power, even in cases where the domain I is not locally finite. Along the way, we show that the possibility of encoding the generalized rank invariant of M over a non-locally-finite I into a multiset of signed intervals of P depends on how "tame" M is. Such a multiset, if it exists, is obtained via Möbius inversion of the generalized rank invariant over a suitable locally finite subset of I.

3.12 Barcodes for the topological analysis of gradient-like vector fields

Claudia Landi (University of Modena and Reggio Emilia, IT)

Joint work of Clemens Bannwart, Claudia Landi
Main reference Clemens Bannwart, Claudia Landi: "Barcodes for the topological analysis of gradient-like vector fields", CoRR, Vol. abs/2401.08466, 2024.
URL https://doi.org/10.48550/ARXIV.2401.08466

Intending to introduce a method for the topological analysis of fields, we present a pipeline that takes as an input a weighted and based chain complex, produces a tame epimorphic parametrized chain complex, and encodes it as a barcode of tagged intervals. We show how to apply this pipeline to the weighted and based chain complex of a gradient-like Morse-Smale vector field on a compact Riemannian manifold in both the smooth and discrete settings. Interestingly for computations, it turns out that there is an isometry between tame epimorphic parametrized chain complexes endowed with the interleaving distance and barcodes of tagged intervals endowed with the bottleneck distance. Concerning stability, we show that the map taking a generic enough gradient-like vector field to its barcode of tagged intervals is continuous. Finally, we prove that the barcode of any such vector field can be approximated by the barcode of a combinatorial version of it with arbitrary precision.

3.13 Challenges in two- and multi-parameter persistent cohomology

Fabian Lenzen (TU Berlin, DE)

In the last years, research in persistent homology has started to focus on multi-parameter persistent homology, which studies the homology of a space filtered by multiple parameters independently. For example, this can be used to overcome the notorious susceptibility of persistent homology to outliers, to deal with data sats of inhomogeneous density, or to study filtration types that rely on more than one parameter.

Computing multi-parameter persistent homology is challenging, both algebraically and algorithmically. In particular, current software is orders of magnitudes slower than common software for one-parameter persistence.

We will discover why persistent cohomology – a key ingredient in the efficiency of one-parameter persistence software – is inherently more difficult in multi-parameter persistence, how this is dealt with in the software package 2pac, and what problems still remain.

3.14 Models of Subdivision Bifiltrations

Michael Lesnick (University at Albany – New York, US)

Joint work of Michael Lesnick, Kenneth McCabe

We study the size of Sheehy's subdivision bifiltrations, up to homotopy. We focus in particular on the subdivision-Rips bifiltration $\mathcal{SR}$, the only density-sensitive bifiltration on metric spaces known to satisfy a strong robustness property. Given a simplicial filtration $\mathcal{F}$ with a total of m maximal simplices across all indices, we introduce a simplicial model for its subdivision bifiltration $\mathcal{SF}$ whose k-skeleton has size $O(m^{k+1})$. We also show that the 0-skeleton of any simplicial model of $\mathcal{SF}$ has size at least m. We give several applications: For arbitrary metric spaces, we introduce a $\sqrt{2}$-approximation to $\mathcal{SR}$ with poly-size skeleta, improving on the previous best approximation bound of $\sqrt{3}$. Moreover, we show that the approximation factor of $\sqrt{2}$ is tight; in particular, there exists no exact model of $\mathcal{SR}$ with poly-size skeleta. On the other hand, we show that for data in a fixed-dimensional Euclidean space with the ℓ_p-metric, there exists an exact model of $\mathcal{SR}$ with poly-size skeleta for $p \in \{1, \infty\}$, as well as a $(1+\epsilon)$-approximation to $\mathcal{SR}$ with poly-size skeleta for any $p \in (1, \infty)$ and fixed $\epsilon > 0$.

3.15 Large Simple d-Cycles in Simplicial Complexes

Roy Meshulam (Technion – Haifa, IL)

Joint work of Roy Meshulam, Ilan Newman, Yuri Rabinovich

Let $G = (V, E)$ be a finite simple graph. A classical result of Erdos and Gallai asserts that if $|E| > \frac{k(|V|-1)}{2}$, then G contains a simple cycle of length $> k$. We study the analogous question for higher dimensional simplicial complexes. A set $\{\sigma_1, \ldots, \sigma_k\}$ of d-dimensional simplices in a simplicial complex X is a *simple d-cycle over a field F* if $\{\partial\sigma_1, \ldots, \partial\sigma_k\}$ is a minimal linearly dependent set in the space of d-chains $C_d(X; F)$. Let $f_i(X)$ denote the number of i-dimensional simplices in X. It is shown that any d-dimensional X contains a simple d-cycle of size

$$k \geq \sqrt{\frac{2f_d(X)}{(d+1)f_{d-1}(X)}} - 1.$$

3.16 Bounding the Interleaving Distance for Mapper Graphs with a Loss Function

Elizabeth Munch (Michigan State University, US)

Data consisting of a graph with a function to $\mathbb{R}^d$ arise in many data applications, encompassing structures such as Reeb graphs, geometric graphs, and knot embeddings. As such, the ability to compare and cluster such objects is required in a data analysis pipeline, leading to a need for distances or metrics between them. In this work, we study the interleaving distance on discretizations of these objects, $\mathbb{R}^d$-mapper graphs, where functor representations of the data can be compared by finding pairs of natural transformations between them. However, in many cases, computation of the interleaving distance is NP-hard. For this reason, we take inspiration from the work of Robinson to find quality measures for families of maps that do not rise to the level of a natural transformation, called assignments. We then endow the functor images with the extra structure of a metric space and define a loss function which measures how far an assignment is from making the required diagrams of an interleaving commute. Finally we show that the computation of the loss function is polynomial. We believe this idea is both powerful and translatable, with the potential to be used for approximation and bounds on interleavings in a broad array of contexts.

3.17 Topologically Attributed Graphs

Tom Needham (Florida State University – Tallahassee, US)

Joint work of Tom Needham, Justin Curry, Washington Mio, Osman Berat Okutan, Florian Russold

I will describe recent work with Curry, Mio, Okutan and Russold which fuses graphical and persistence invariants of datasets. The basic idea is to attribute the nodes of a Reeb or Mapper graph of a dataset with persistence diagrams, which encode localized, higher-dimensional homological features of the data. These enriched graphical summaries can be used, for example, as inputs to a graph neural network for shape classification tasks. I will also discuss the (fairly subtle) theoretical stability properties of these invariants.

3.18 Directed paths and duality

Martin Raussen (Aalborg University, DK)

An important class of Higher Dimensional Automata (HDA) in concurrency theory arises from semaphore protocols or PV-programs originally described by Dijkstra. In order to understand their behaviour, one must analyse the *space* of *all* schedules (directed paths) between (any) start and end state. How can one translate the orders of lock and unlock commands into a recipe describing this space?

By definition, the space of allowed directed paths is an intersection (limit) of elementary spaces – each having the homotopy type of a sphere – in the infinite-dimensional space of all directed paths. There is a homotopy equivalence embedding the (allowed) paths as a *configuration space* into a *finite-dimensional* sphere. The complement of this configuration space in that sphere is a union (colimit) of elementary spaces. Its topology can therefore be described as the homotopy colimit of certain spaces for which we have a "low-dimensional" description arising directly from the PV-encoding. In favourable cases, this homotopy colimit can be described explicitly. Alexander duality allows then to determine the homology of the complement, and hence of the space of all allowed directed paths.

3.19 From Coarse to Fine and Back Again: Geometry and Topology in Machine Learning

Bastian Grossenbacher-Rieck (Helmholtz Zentrum München, DE)

A large driver contributing to the success of deep learning models is their ability to synthesise task-specific features from data. For a long time, the predominant belief was that 'given enough data, all features can be learned.' However, it turns out that certain tasks require imbuing models with inductive biases such as invariances that cannot be readily gleaned from the data! This is particularly true for data sets that model real-world phenomena, creating a crucial need for different approaches. This talk will present novel advances in harnessing multi-scale geometrical and topological characteristics of data. I will particularly focus on how geometry and topology can improve (un)supervised representation learning tasks. Underscoring the generality of a hybrid geometrical-topological perspective, I will furthermore showcase applications from a diverse set of data domains, including point clouds, graphs, and higher-order combinatorial complexes.

3.20 Overview of Discrete Morse Theory

Nick Scoville (Ursinus College – Collegeville, US)
Leonard Wienke (Universität Bremen, DE)

This Overview of Discrete Morse Theory is two-fold.

In the first part, we give an introduction to the basic concepts of Discrete Morse Theory. In particular, we discuss the equivalence of simplicial collapses, acyclic matchings, and poset maps with small fibers. We then define the Morse complex that computes simplicial homology and consider examples.

In the second part, we discuss open problems as well as newer directions of research. We will look at open problems in both random Discrete Morse Theory and the complex of discrete Morse functions. We will then survey several variations of Discrete Morse Theory, inluding stratified and Bestvina-Brady, which may prove useful in simplifying a complex.

3.21 Combinatorial Topological Models for Phylogenetic Reconstruction Networks

Jan F Senge (Universität Bremen, DE & Polish Academy of Science, PL)

Joint work of Paweł Dłotko, Jan Felix Senge, Anastasios Stefanou
Main reference Paweł Dłotko, Jan Felix Senge, Anastasios Stefanou: "Combinatorial Topological Models for Phylogenetic Networks and the Mergegram Invariant", CoRR, Vol, abs/2305.04860, 2023.
URL https://arxiv.org/abs/2305.04860

Phylogenetic networks are vital for understanding complex evolutionary processes, where traditional tree-like structures fall short. The application of topological data analysis (TDA) has emerged as a powerful approach for exploring such networks, revealing underlying geometric and topological structures. This talk focuses on a lattice theoretical approach of representing such networks and relating them to TDA. We will discuss the applications of TDA techniques in analyzing phylogenetic networks, aiming to uncover hidden patterns and gain deeper insights into their evolutionary dynamics. Additionally, we introduce the facegram, a simplicial lattice model that generalizes the dendrogram model for phylogenetic trees, which enables an alternative way to visualize filtrations of complexes, and show some more recent applications of these ideas and connections.

3.22 Reeb Graphs and Their Variants: Theory and Applications

Bei Wang Phillips (University of Utah – Salt Lake City, US)

A Reeb graph is a graphical representation of a scalar function on a topological space that encodes the topology of the level sets. Reeb graphs and their variants are popular tools in topological data analysis and visualization. As an overview talk for TDA+statistics, I will review theoretical advances in studying Reeb graphs and their variants, as well as their applications in data mining and machine learning.

From a theoretical perspective, the questions surrounding Reeb graphs are as follows.

- **Comparative analysis**: What is a reasonable distance or similarity measure between a pair of Reeb graphs? Desirable properties of a distance/measure is for it to be a metric or pseudometric, discriminative, and easy to compute (both in terms of computational complexity and practical implementation). See Yan et al. (2021) [35] and Bollen et al. (2022) [9] for surveys.
- **Stability**: How stable is a Reeb graph w.r.t. simplification or perturbation of the underlying function?
- **Information content**: What information is encoded by the Reeb graph? How much information can we recover about the original data from the Reeb graph by solving an inverse problem?

The questions surrounding mapper graphs (discrete approximations of Reeb graphs) include comparative analysis, information content, and additionally:

- **Stability**: What is the structural stability of the mapper with respect to perturbations of its function, domain and cover?
- **Convergence**: What is an appropriate metric under which the mapper converges to the Reeb graph as the number of sampled points goes to infinity and the granularity of the cover goes to zero?
- **Parameter tuning**: How to effectively and automatically tune the parameters that best capture the topology of the underlying data?

Reeb graphs have many variants, including:

- Mapper construction/mapper graph: Singh et al. (2007) [33].
- α-Reeb graph that considers the cover of range space with open intervals of length at most α, see Chazal and Sun (2014) [17].
- Multiscale mapper considers a hierarchical family of covers and the maps between them, see Dey et al. (2016) [18].
- Multinerve mapper computes the multinerve of a connected cover, see Carriere and Oudot (2018) [15].

- Joint Contour Net (JCN) works with a piecewise linear (PL) mapping over a simplicial mesh with multiple real-valued functions, see Carr and Duke (2013) [14] and Geng et al. (2014) [22].
- Extended Reeb graph uses cover elements from a partition of the domain without overlaps, see Barral and Biasotti (2014) [4].
- Enhanced mapper graph considers additionally the inverse map of intersections among cover elements, see Brown et al. (2021) [10].
- Ball mapper that may not require a filter function, see Dłotko (2019) [20].

Reeb graphs and their variants, in particular, mapper graphs, have seen many applications in topological data analysis and visualization. They have been used for shape skeletonization (Pascucci et al., 2007) [29] and symmetry detection (Thomas and Natarajan 2011) [34]. They have recently been utilized to study artificial neuron activations in deep learning, see the works by Purvine et al. (2023) [30], and Rathore et al. (2021, 2023) [31, 32]. For applications in visualization, see Yan et al. (2021) [35] for a survey.

3.23 Persistent cup modules

Ling Zhou (Duke University – Durham, US)

Joint work of Facundo Mémoli, Anastasios Stefanou, Ling Zhou

One-dimensional persistent homology is arguably the most important and heavily used computational tool in topological data analysis. Additional information can be extracted from datasets by studying multi-dimensional persistence modules and by utilizing cohomological ideas, e.g. the cohomological cup product. In this work, given a single parameter filtration, we investigate a certain 2-dimensional persistence module structure associated with persistent cohomology, where one parameter is the cup-length and the other is the filtration parameter. This new persistence structure, called the persistent cup module, is induced by the cohomological cup product and adapted to the persistence setting. Furthermore, we show that this persistence structure is stable. By fixing the cup-length parameter, we obtain a 1-dimensional persistence module and again show it is stable in the interleaving distance sense, and study their associated generalized persistence diagrams.

4 Open Problems

4.1 Problem 1: Can we compute representatives of bars of persistent homology in less than cubic time?

Tamal K. Dey (Purdue University – West Lafayette, US)

Matrix reduction algorithm for persistent homology: $R = DV$ reduced boundary.

		σ	
τ_1		1	
τ_2		1	
τ_3		1	

In this example of a reduced matrix, σ is a negative simplex, because it kills the cycle $(\tau_1+\tau_2+\tau_3)$. The representative of a bar can be read off from the column of the corresponding negative simplex: In this case it would be $(\tau_1 + \tau_2 + \tau_3)$. Similarly, the representatives of infinite bars can be read off from the columns of the matrix V. The reduced matrix R and the matrix V, and hence the representatives, can be computed in $\mathcal{O}(m^3)$ time. On the other hand, the persistence pairing (in other words, the pivot positions of the reduced matrix) can be computed even in matrix multiplication time $\mathcal{O}(m^\omega)$, with $\omega < 2.373$, for classical persistent homology and even for zigzag persistent homology. The question is whether this holds only for the computation of the persistence pairs, or also for the computations of their representatives.

Questions:

1. Compute homological representative cycles for all bars in $\mathcal{O}(m^{<3})$ time where m is the size of the input filtration?
2. Zigzag in $\mathcal{O}(m^{<4})$ or $\mathcal{O}(m^3)$ or even $\mathcal{O}(m^{<3})$ time?

Reference to "Zigzag Persistent Homology in Matrix Multiplication Time", Milosavljević, Nikola and Morozov, Dmitriy and Skraba, Primoz, Proceedings of the twenty-seventh Annual Symposium on Computational Geometry, 216–225, 2011. `https://www.mrzv.org/publications/zzph-mmt/socg11/`

During this Dagstuhl Seminar, this question has been investigated by the working group on algorithms.

Tamal's comment: The above problem is not open anymore. It turns out that the algorithm in the paper mentioned above can be utilized to compute a representative for standard persistence in $O(m^\omega)$ time. The authors of the paper exchanged notes through email which have convinced us that indeed the algorithm can be adapted to compute the representatives in the stated time. For zigzag, it is not clear if the algorithm can be adapted straightforwardly to do the same. However, the proposer of the problem with his students has gotten an algorithm which can compute the representatives in $O(m^3)$ time. Interested people may contact the proposer to have a preprint. The paper is currently under some revision which is planned to be submitted for publication in a near future. The question of computing the representatives for zigzag in less than cubic time remains open. It is not entirely clear if it can be done at all because there is no bound better than cubic known for the output size in this case.

4.2 Problem 2: Optimal Discrete Morse function given a partial matching

Yusu Wang (University of California, San Diego-La Jolla, US)

Base DM (Discrete Morse) Graph Reconstruction algorithm (see e.g., [19])
Input: Triangulation K of domain $I \subset \mathbb{R}^d$, function $f : K \to \mathbb{R}$, threshold δ

- Step 1: persistence computation
- Step 2: persistence-guided Morse simplification.
  - Spanning forest construction based on persistence output
  - retrieval of 1-stable manifold based on spanning forest for critical edges with persistence larger than δ

The above algorithm is already known. However, in practical applications, we often need to have additional constraints as input to the above graph reconstruction algorithm. For example, in neural bundle reconstruction from 3D images, we may want to add constraints that there are desired "flow directions" at certain locations. These flow directions could have been computed locally based on computer vision-based approaches. Hence, the high level problem we aim to solve is a discrete Morse based graph reconstruction with additional constraints. As a first step, we encode the "flow direction" constraints simply as a set of discrete gradient vectors. More precise, see the following description:

Input:

- a simplicial complex K with vertices $V = V(K)$
- a function $f : V \to \mathbb{R}$
- collection of pairs $P = \{(\sigma, \tau) | \sigma \subset \tau, |\tau| = |\sigma| + 1\}$, indicating desirable "gradient vectors"

Goal:

1. What would be a good way to define an "optimal" discrete gradient vector field (or a (generalized) discrete Morse function) that combine both types of input
2. In particular, what theoretical properties can we provide (including properties of the graph reconstructed using the earlier algorithms)

Idea:
convex polytope solution (together with Linear / Quadratic Programming); space of morse functions and hyperplanes

During this Dagstuhl Seminar, this question has been investigated by the working group on algorithms.

4.3 Problem 3: Algorithmic questions on (multiparameter) persistence

Fabian Lenzen (TU Berlin, DE)

Let M be a matrix where rows carry labels $\ell_i \in \mathbb{Z}^2$, fixed once and for all. If V is invertible, we let

$$\ell(MV_j) = \bigvee_{(MV)_{ij} \neq 0} \ell_i, \quad \text{where} \quad (x, y) \vee (x', y') = (\max\{x, x'\}, \max\{y, y'\}).$$

We can prove[1] that there exists V that minimizes $\ell(MV_j)$ simultaneously for all j. The order of the columns does not matter.

Algorithm: Let P and Q be permutations such that PM and QM have rows in lexicographic and colexicographic order w.r.t. ℓ.

1. Compute V' such that PMV' is reduced
2. Order columns of PMV' by pivot
3. Compute *upper triangular* V'' such that $QMV'V''$ is reduced.

With $V = V'V''$, the ordering of the columns in step 2 ensures that afterwards, both PMV and QMV are reduced. See the paper for a proof that this minimizes all $\ell(MV_j)$.

Problem: If M is a coboundary matrix of a two-parameter function-Rips filtration, the algorithm produces tremendous fill-in in step 3: Out of millions of columns, there are often 1–10 columns that for which 50–90% of the entries are non-zero.

Question: Can we devise an algorithm that (in most cases) does not produce fill-in?

Motivation: Fill-in in only a few columns is currently a central bottleneck in computing Vietoris–Rips persistent cohomology. The above task can be seen as a two-parameter generalization of the clearing-idea. Experiments show that fill-in is an even worse problem in two-parameter persistence. Solving this problem would be a major step towards efficient algorithms for multi-parameter persistence.

Besides, Uli also mentioned that matrix fill-in is a central bottleneck in computing image persistence.

4.4 Problem 4: Manifold reconstruction guarantees

Ulrich Bauer (TU München, DE)

Manifold reconstruction guarantees for the method described in the talk (exhaustive reduction method for most persistent top dimensional cycle). Note that the exhaustive reduction algorithm gives the lexicographically minimal cycle representative.

[1] See `https://doi.org/10.4230/LIPIcs.SoCG.2023.15` for a proof and for motivation.

- As a motivation, we considered a visualization of the exhaustive reduction method (eliminate all entries that can be eliminated, not only the lowest ones).
- The reduction starts with one d-simplex (in $\mathbb{R}^d$) and its boundary, and you add more until you reach the reduced cycle.

Under which sampling conditions can we guarantee that this lexicographically minimal cycle reconstructs the manifold? And does it always make sense to take the most persistent top-dimensional cycle? If the manifold has multiple components, we would need one cycle for each component.

As a first step, we would need to investigate: What methods do people use to prove such results?

4.5 Problem 5: Topological information gain

Bei Wang Phillips (University of Utah – Salt Lake City, US)

We are interested in quantifying the information content of multiparameter topological descriptors, in particular, multiparameter persistence modules [13], Reeb spaces [21], and mappers [33].

Let X be a topological space equipped with a pair of functions, $f, g : X \to \mathbb{R}$. We first utilize the function f (resp., g) to obtain a 1-parameter topological descriptor, denoted as $P(X, f)$ (resp., $P(X, g)$). We then use both f and g to construct a 2-parameter topological descriptor, denoted as $P(X, f, g)$. For example, if $P(X, f)$ is a Reeb graph, then $P(X, f, g)$ is a Reeb space. Alternatively, $P(X, f)$ and $P(X, f, g)$ are 1- and 2-parameter persistence modules, respectively. We ask the following question: how do we quantify the information gain from a 1-dimensional topological descriptor $P(X, f)$ to a 2-dimensional topological descriptor $P(X, f, g)$?

This problem first appeared in the work of Zhou et al. [37], where the authors investigated a method for stitching a pair of 1-parameter mappers together into a 2-parameter mapper, quantified and visualized topological notions of information gains during such a process. While the work in [37] provides some initial thoughts on this problem, including graph entropy, fiber-wise homology and Euler characteristics, it leaves much to be desired.

During this Dagstuhl Seminar, this question has been investigated by a working group studying topological information gain.

4.6 Problem 6: Can TDA detect planted cliques?

Bastian Rieck (Helmholtz Zentrum München, DE)

In *complexity theory*, a classical problem is the so-called *planted clique problem*. Given natural numbers n and k, the planted clique problem uses the following procedure:

1. Create an Erdős–Rényi graph G on n vertices with edge creation probability $p = 0.5$.
2. With probability 0.5, select a subset of k vertices in G and turn them into a k-clique.
3. Return G.

Given G, the goal is now to *detect* whether a clique has been planted or not. The planted clique problem is most interesting – and hard – for a specific range of k, viz.

$$2\log_2(n) \ll k \ll \sqrt{n}. \tag{1}$$

My **open question** is whether topological descriptors can detect planted cliques in the distributional sense, for instance via *persistence landscapes*, calculated from an appropriately-selected filtration. (See the working groups section below for the preliminary results).

4.7 Problem 7: Monotonicity of magnitude functions of Euclidean metric spaces

Sara Kalisnik (ETH Zürich – Zürich, CH)

Sara Kalisnik posed the open question: Are magnitude functions of Euclidean metric spaces monotonic? (see working groups section below for the details and preliminary results).

5 Working Groups

5.1 Working Group: Algorithms

Ulrich Bauer (TU München, DE)
Tamal K. Dey (Purdue University – West Lafayette, US)
Teresa Heiss (Institute of Science and Technology Austria, AT)
Fabian Lenzen (TU Berlin, DE)
Nick Scoville (Ursinus College – Collegeville, US)
Bei Wang Phillips (University of Utah – Salt Lake City, US)
Yusu Wang (University of California, San Diego-La Jolla, US)

5.1.1 Question 1

Compute the generators for bars in persistent homology in $< O(n^3)$ time, where n is the number of filtration steps, and (by assuming every filtration step adds exactly simplex) also equals the number of simplices. We also ask the same question for Zigzag persistence. Here, we assume that the filtration starts and ends with the empty complex, and every filtration step either adds or removes exactly one simplex. Hence, the number of simplices used overall is at most $\frac{n}{2}$.

As a first step, we contacted Primoz Skraba and Dmitriy Morozov about their paper [28] which introduces the algorithm to compute persistence barcodes (both for classical persistent homology and zigzag persistent homology) in matrix multiplication time. We asked them if they were aware if additionally to the barcode itself, the generators can be computed in matrix multiplicaiton time. They were confident that for classical persistent homology, the generators could be read off of a matrix called Z in their paper on this algorithm. After further communication, it turned out it seems non-trivial, but indeed possible. However, that approach does not seem possible/easy to extend to zigzag persistence, where even the output might be cubic. One possible future direction would be convincing ourselves in detail that it is indeed possible to read off the generators of classical persistent homology from said matrix Z. Another possible future direction of research would be finding an explicit example of a zigzag generator output that is cubic. This would of course disprove the possibility of less than cubic running time for zigzag generators.

Another question we tried to answer is the following: Given a matrix time algorithm to compute generators in the classical persistent homology setting (which seems to exist, see above), can we compute zigzag generators in matrix multiplication time? As we are worried that the zigzag generator output could be cubic, we were brainstorming for alternative outputs that would be less than cubic, and from which it would be easy to compute the (possibly cubic) wanted output. In the specific setting of level-set zigzag persistent homology, an idea for this is somehow pulling back the generator of ordinary persistence through the Mayer–Vietoris pyramid [12] and even further (for this we need the relative interlevel set cohomology [5], which is an extension of the Mayer Vietoris pyramid that also includes maps between homological degrees) to get to a generator in a node of the relative interlevel set cohomology from which all the maps to every step of the zigzag module are forward pointing. If that pull-back operation can be done in matrix multiplication time, or even quadratic time, which seems plausible, there is hope for some kind of zigzag generator output, computable in matrix multiplication time, even if that output is not precisely the output needed in many applications.

5.1.2 Question 2

Given a simplicial complex K along with an acyclic matching P on K and a specified set of values f on the vertices, find an optimal discrete Morse function g that is compatible with P and that is as close to f as possible.
We decided that given a function f on the vertices of K and a partial matching P of K (without assuming any relation between f and P), we wish to find a monotone function g on K with the hard constraint of g being (weakly) compatible with P (i.e. g has to map matched simplices to the same value) that minimizes $\|g-f\|_2^2$. This is a quadratic optimization over a convex polytope (because the constraint of being a monotone function compatible with P is determined by equalities and inequalities), and we can solve it via standard optimization procedures. Note that the result is only a monotone function (not necessarily discrete Morse), but we can easily perturb it into a discrete Morse function g_ϵ satisfying $\|g-g_\epsilon\|<\epsilon$ for any $\epsilon>0$. The gradient of the Morse function could either be P or another matching containing P as a subset. This construction is entirely elementary, namely adding a discrete Morse function h that is compatible with P and has a small norm to g, in order to break the tie between simplices that get mapped to the same function value under g. Note that in some applications it might not even be necessary to pass to the discrete Morse function g_ϵ, because a monotone function g might be enough. In those applications it might be more useful to use the approach in optimal simplification of discrete functions [6] (see also Ulrich Bauer's PhD Thesis) to find the maximal pairing that contains the pairing P and is still compatible with g.

Note that it is possible to turn the hard constraint of being compatible with P into a soft constraint. In this case, the monotonicity constraints are the only hard constraints and the optimization function is $\|g-f\|_2^2+\sum_{(a,b)\in P} w_{(a,b)}(g(a)-g(b))^2$ for appropriate weights $(w_{(a,b)})_{(a,b)\in P}$.

An alternative construction was discussed, namely, a modified version of the King–Knudson–Mramor algorithm. We observed that this algorithm is actually very closely related to the approach described above.

5.2 Working Group: Topological Information Gain

Julian Brüggemann (Universität Bonn, DE)
Mathieu Carrière (Centre Inria d'Université Côte d'Azur – Sophia Antipolis, FR)
Paweł Dłotko (Polish Academy of Science, PL)
Teresa Heiss (Institute of Science and Technology Austria, AT)
Ingrid Hotz (Linköping University, SE)
Michael Kerber (TU Graz, AT)
Woojin Kim (Duke University – Durham, US & KAIST – Daejeon, KR)
Claudia Landi (University of Modena and Reggio Emilia, IT)
Bei Wang Phillips (University of Utah – Salt Lake City, US)
Yusu Wang (University of California, San Diego-La Jolla, US)

5.2.1 Introduction

There are a number of multiparameter topological descriptors, such as multiparameter persistence modules [13], Reeb spaces [21], and mappers [33]. Before utilizing these descriptors in topological data analysis, a key question is: when does a multiparameter topological descriptor provide quantifiable benefits in addition to a set of single parameter topological descriptors derived from the same data?

Simply put, we are interested in studying a topological notion of information gain from a 1-parameter topological descriptor to a 2-parameter one.

Let X be a topological space equipped with a pair of functions $f_1, f_2 : X \to \mathbb{R}$. X may be a d-dimensional manifold or a point cloud in $\mathbb{R}^d$. We construct a 1-parameter topological descriptor from f_1, denoted as $P(X, f_1)$. We then construct a 2-parameter topological descriptor using both f_1 and f_2, denoted as $P(X, f_1, f_2)$. If $P(X, f)$ is a Reeb graph, then $P(X, f_1, f_2)$ is a Reeb space. If $P(X, f_1)$ is a 1-parameter persistence module, then $P(X, f_1, f_2)$ is a 2-parameter one. We pose the following question: how do we quantify the information gain from a 1-dimensional topological descriptor $P(X, f_1)$ to a 2-dimensional one $P(X, f_1, f_2)$?

During the Dagstuhl Seminar, we have discussed many potential approaches to address this question, some of which are summarized below.

5.2.2 A Lifting Approach

We first consider a simple 2-parameter setting: given f_1 and f_2, we want to find a single function g, such that $P(X, g)$ optimally describes $P(X, f_1, f_2)$. We consider lifting g to a 2-parameter space by replacing $P(X, g)$ with $P(X, g, g)$, and compare the information content between $P(X, f_1, f_2)$ and $P(X, g, g)$.

More abstractly, let $\mathcal{P}_1$ denote the set of all topological descriptors of the form $P(X, f)$ (e.g., 1-parameter persistence modules) and $\mathcal{P}_2$ the set of all topological descriptors of the form $P(X, f_1, f_2)$ (e.g., 2-parameter persistence modules). Let $L : \mathcal{P}_1 \to \mathcal{P}_2$ denote a lifting operator, for instance, $L : P(X,g) \mapsto P(X,g,g)$. Let $d : \mathcal{P}_2 \times \mathcal{P}_2 \to [0,\infty)$ be a distance measure on $\mathcal{P}_2$. Then we are asking for a function g such that

$$d(P(X,f_1,f_2),L(P(X,g)))$$

is minimized. Depending on the constraints imposed on g, the search space of g may be a linear combination of f_1 and f_2, or simply the set $\{f_1,f_2\}$.

Depending on the choices of P, L, and d, we might obtain a different set of desirable properties of the optimizer for g. As an example, consider the situation where g is constrained to be a *convex* combination of f_1 and f_2, and P is the persistence module. Let $d := d_I$ denote the interleaving distance between 2-parameter persistence modules. Let the lifting of $P(X,g)$ to be the persistence module induced by the sublevel set filtration of g in both directions, that is, $L(P(X,g)) := P(X,g,g)$. We are looking for a parameter $\alpha \in [0,1]$ such that for $g_\alpha := \alpha f_1 + (1-\alpha) f_2$, the interleaving distance between $P(X,f_1,f_2)$ and $P(X,g_\alpha,g_\alpha)$ is minimized.

Here is an algorithmic sketch, inspired by [7, 23]. We aim to find some α' such that, for a fixed $\varepsilon > 0$, the interleaving distance between $P(X,f_1,f_2)$ and $P(X,g_{\alpha'},g_{\alpha'})$ is at most ε larger than the optimum. To that end, we need two ingredients. First, we need to be able to evaluate the interleaving distance $d_I\left(P(X,f_1,f_2),P(X,g_{\alpha_0},g_{\alpha_0})\right)$ for any $\alpha_0 \in [0,1]$. Second, we need a bound for the *variance*, that is, the difference between $d_I\left(P(X,f_1,f_2),(X,g_{\alpha_0},g_{\alpha_0})\right)$ and $d_I\left(P(X,f_1,f_2),(X,g_{\alpha_1},g_{\alpha_1})\right)$ when $|\alpha_0-\alpha_1|$ is at most δ. Such a bound may be obtained through the stability of the interleaving distance (e.g., [16]) and the details have to be worked out. In any case, if we then sample the unit interval finely enough such that the variance in every subinterval is at most ε, we can evaluate the interleaving distance at the sample points and take the minimum as our solution.

We can further optimize the above strategy with an adaptive subdivision process. We keep splitting the unit interval into subintervals, evaluate at the midpoint and compute the variance of the interval. We also remember the smaller interleaving distance Δ_{min} we have seen so far. If for a subinterval, the interleaving distance at the midpoint is Δ_0 and the variance is σ and $\Delta_0 - \sigma \geq \Delta_{min}$ holds, we can avoid further subdivision in that subinterval.

The interleaving distance is NP-hard to compute and to approximate [8]. However, we could in practice replace it with any stable distance and the above algorithm should transfer without any changes. The same algorithm will also work in principle for more than two input functions, by requiring to search in a high-dimensional simplex. However, it will also require computing distances between persistence modules with more than two parameters, which is likely unpractical with current technology.

A potential criticism of the above algorithm is that it has a rather strong assumption that $P(X,f_1,f_2)$ should be reasonably similar to $P(X,g,g)$ for some g. It seems reasonable to imagine the situation that $P(X,f_1,f_2)$ is similar to $P(X,g_\alpha,ag_\alpha+b)$ for some $a,b \in \mathbb{R}$. Therefore, a possible direction is to explore the distance between these two topological descriptors, by optimizing parameters α, a and b.

Finally, the lifting approach may be generalized to the setting where we want to find a single function g to optimally describe the topological descriptor of a set of function $f_1,\ldots,f_p : X \to \mathbb{R}$, for $p \geq 2$. For example, let $p = 4$. We can use the algorithm above to find some g_{12} such that the distance between $P(X,f_1,f_2)$ and $P(X,g_{12},g_{12})$ is minimized. Likewise, we can find g_{34} such that the distance between $P(X,f_3,f_4)$ and $P(X,g_{34},g_{34})$ is minimized. As the last step, we can find a function g such that the distance between $P(X,g_{12},g_{34})$ and $P(X,g,g)$ is minimized. In other words, we construct a binary tree, called a *tournament tree*, and always combine two functions into another one that is used for optimization higher up in the tree. It remains speculative whether the final outcome is a good proxy for the global optimum of the set of functions. It is also not clear how stable the outcome is under a permutation of the input functions, which would yield a different tournament tree.

5.2.3 A Matrix Sketching Approach

Now, we consider the multiparameter setting using a matrix sketching approach. We start with a topological space X together with a set of functions $f_1, \ldots, f_p : X \to \mathbb{R}$. We assume an operator P is given, which returns for each pair (X, f_i) $(1 \le i \le p)$ a topological descriptor $P(X, f_i)$. Motivated by dimensionality reduction, in particular, principal component analysis (PCA), the goal is to compute a smaller set of functions $g_1, \ldots, g_q : X \to \mathbb{R}$ (with $q \ll p$) such that the set of descriptors $P(X, g_1), \ldots, P(X, g_q)$ optimally describes the set of descriptors $P(X, f_1), \ldots, P(X, f_p)$.

We consider several types of constraints on the functions g_j $(1 \le j \le q)$:

- We can require that $\{g_1, \ldots, g_q\} \subset \{f_1, \ldots, f_p\}$.
- We can require that each g_j is an linear combination among the f_i.

Li et al. [27] partially addressed the above problem by solving a related one: given a large set $\mathcal{T}$ of merge trees, find a much smaller set of basis trees $\mathcal{S}$ such that each tree in $\mathcal{T}$ can be approximately reconstructed from a linear combination of trees in $\mathcal{S}$.

First, we recall the standard PCA. We are given a dataset of p points with ℓ features, represented as a $\ell \times p$ matrix A (with row-wise zero empirical mean). Pick a parameter q, PCA finds a q-dimensional subspace $\mathbb{H}$ of $\mathbb{R}^\ell$ that minimizes the average squared distance between the points and their projections onto $\mathbb{H}$. Algebraically, PCA tries to approximate the data matrix A with a matrix $\hat{A}$, that is, $A \approx \hat{A} = BY$, such that $||A - \hat{A}||_F^2 = ||A - BY||_F^2$ is minimized. B is a $\ell \times q$ matrix whose columns form an orthonormal basis for $\mathbb{H}$, whereas Y is a $q \times p$ coefficient matrix. By construction, each column vector in A is now approximated by a linear combination of basis vectors in B with coefficients specified in Y. Lin et al. [27] applied PCA and column subset selection (CSS, another matrix sketching technique) to a set of vectorized merge trees, and obtains a set of basis vectors that could be converted back into merge trees.

In our current setting, let V denote a vectorization method that assigns to a topological descriptor $P(X, f_i)$ a vector in some Euclidean space $\mathbb{R}^\ell$. For shorter notation, we write $V_i := V(P(X, f_i))$. There is a rich literature of such vectorization methods in the context of persistence diagrams, including persistence landscapes [11], persistence images [1], and persistence kernels (e.g. [36]), etc. It is also possible to vectorize merge trees (and similarly contour trees) using techniques from optimal transport [27].

Having fixed a set of vectors $V_1, \ldots, V_p$, we can proceed by applying a standard PCA on the $l \times p$ matrix spanned by the vectors and compute the first q principal components. Each principal direction can be written as a linear combination of the V_i. Let us assume for now that there is a canonical or good way to choose the weights α_i, and let $a_k = \alpha_1 V_1 + \ldots + \alpha_p V_p$ denote the k-th principal component (for $1 \le k \le q$). We can simply set

$$g_k := \alpha_1 f_1 + \ldots + \alpha_p f_p,$$

meaning that we pull back the PCA solution to the function space we started with. With a chosen q, this procedure would hopefully yield the functions $g_1, \ldots, g_q$ as desired. The key question is whether such a pull back produces meaningful solutions.

5.2.4 Multi-Parameter Persistent Entropy

Another option for comparing $P(X, f_1, f_2)$ and $P(X, g, g)$ is to develop a multi-parameter version of the *persistent entropy*.

Persistent entropy was originally defined as a statistic for persistence diagrams [2]. It reduces a persistence diagram to a single number, thus discarding a lot of information. However, it is useful for comparing distributions of diagrams with usual statistical tests. For a single-parameter persistence diagram consisting of a set $\mathcal{I}$ of finite bars of the form $I := [I_b, I_d) \subset \mathbb{R}$, the persistent entropy is defined as:

$$E(\mathrm{dgm}_{\mathcal{I}}) := -\sum_{I\in\mathcal{I}} \frac{I_d - I_b}{L} \cdot \log\left(\frac{I_d - I_b}{L}\right),$$

where L is the total persistence (excluding the infinite bars), $L := \sum_{I\in\mathcal{I}}(I_d - I_b)$. It would be interesting to investigate the extension of this statistic to the multiparameter setting and its theoretical properties, following [2, 3].

A natural extension may be obtained using the Möbius inversions. Fix a family $\mathcal{I}$ of intervals in $\mathbb{R}^p$ (p is the number of parameters), we generate a signed barcode by computing the Möbius inversion of the generalized rank invariant over $\mathcal{I}$ [24]. In practice, this barcode is simply defined as a set of "birth" and "death" intervals (denoted by $\mathcal{I}^+$ or $\mathcal{I}^-$, respectively), which can be seen as a discrete measure with positive and negative signs:

$$\mathrm{dgm}_{\mathcal{I}} = \sum_{I^+\in\mathcal{I}^+} \delta_{I^+} - \sum_{I^-\in\mathcal{I}^-} \delta_{I^-}.$$

Intuitively, this signed barcode "agrees" with the generalized rank invariant (GRI) on $\mathcal{I}$:

$$\forall I \in \mathcal{I},\ \mathrm{GRI}(I) = \sum_{I\subseteq J, J\in\mathcal{I}} \mathrm{dgm}_{\mathcal{I}}(J).$$

Then the question becomes how to adapt the persistent entropy to these signed barcodes, ideally in a way such that $E(P(X, f)) = \tilde{E}(P(X, f, f))$, where $\tilde{E}$ is the extended persistent entropy. A possibility could be to use, for an interval I, the longest diagonal that is included in I:

$$\tilde{E}(\mathrm{dgm}_{\mathcal{I}}) = -\sum_{I\in\mathrm{supp}(\mathrm{dgm}_{\mathcal{I}})} \frac{\ell_I}{L} \cdot \log\left(\frac{\ell_I}{L}\right),$$

where $\ell_I = \sup\{\delta > 0 : x, x + \delta \cdot [1, \ldots, 1] \subseteq I\}$ and $L = \sum_{I\in\mathrm{supp}(\mathrm{dgm}_{\mathcal{I}})} \ell_I$.

5.3 Working Group: Can Topological Descriptors Detect Planted Cliques?

Niklas Hellmer (Polish Academy of Science, PL)
Bastian Rieck (Helmholtz Zentrum München, DE)
Jan Felix Senge (Polish Academy of Science, PL & Universität Bremen, DE)

In *complexity theory*, a classical problem is the so-called *planted clique problem*. Given natural numbers n and k, the planted clique problem uses the following procedure:

1. Create an Erdős–Rényi graph G on n vertices with edge creation probability $p = 0.5$.
2. With probability 0.5, select a subset of k vertices in G and turn them into a k-clique.
3. Return G.

Given G, the goal is now to *detect* whether a clique has been planted or not. The planted clique problem is most interesting – and hard – for a specific range of k, viz.

$$2\log_2(n) \ll k \ll \sqrt{n}. \tag{2}$$

My **open question** is whether topological descriptors can detect planted cliques in the distributional sense, for instance via *persistence landscapes*, calculated from an appropriately-selected filtration. (See the working groups section below for the preliminary results).

Preliminary results

Using the magnitude of metric spaces, we found that some regimes, i.e. choices of k and n, afford a detection that is substantially better than random chance; see table 1 for an overview. Moving forward, we will work on deriving more theoretical bounds.

Table 1 Detecting *planted cliques* with metric space magnitude. We used graphs with $n = 600$ vertices and generated $N = 500$ graphs. The table reports the accuracies for detecting a planted clique as a function of k, the clique size.

k	Accuracy
19	58.40 ± 4.88
20	62.40 ± 4.67
21	61.00 ± 4.34
22	65.00 ± 2.97
23	62.80 ± 0.98
24	60.20 ± 4.07

5.4 Working Group: Monotonicity of Magnitude functions of Euclidean metric spaces

Sara Kalisnik (ETH Zürich – Zürich, CH)
Claudia Landi (University of Modena and Reggio Emilia, IT)
Elizabeth Munch (Michigan State University, US)
Tom Needham (Florida State University – Tallahassee, US)
Anastasios Stefanou (Universität Bremen, DE)
Ling Zhou (Duke University – Durham, US)
Paweł Dłotko (Polish Academy of Science, PL)

This group worked on a basic problem in the theory of metric space magnitude. Magnitude was introduced by Leinster and its fundamental properties are worked out in [26, 25]. The notation and initial exposition below follow [25].

5.4.1 Definitions and problem statement

Let $X = (X, d)$ be a finite metric space. We fix an order on the points of X and denote them as $x_1, \ldots, x_n$. The distance matrix of X, with respect to this ordering, is denoted

$$D = [d(x_i, x_j)]_{ij} = [D_{ij}]_{ij}.$$

From the distance matrix, we derive a *similarity matrix* $Z = Z_X$ with entries $Z_{ij} = e^{-D_{ij}}$. Assuming that the similarity matrix is invertible, the *magnitude* of X, denoted $|X|$ is given by

$$|X| = \sum_{i,j=1}^{n} (Z^{-1})_{ij}.$$

Speaking informally, the magnitude $|X|$ captures the effective number of points in X. For example, if X is a two point metric space whose points are at distance t, then $|X| = \frac{2}{1+e^{-t}}$, so that $|X| \to 1$ as $t \to 0$ and $|X| \to 2$ as $t \to \infty$.

The example above suggests studying the magnitude of rescalings of the metric space. For $t > 0$, let tX denote the metric space with underlying set X and with metric $t \cdot d$; that is, the metric function is uniformly rescaled by a factor of t. The *magnitude function of* X is the function $M = M_X$ defined by

$$\begin{aligned} M : \mathbb{R}_{>0} &\to \mathbb{R} \\ t &\mapsto |tX|. \end{aligned}$$

For example, if X is the two point metric space whose points are at unit distance from one another, then the example from the previous paragraph implies that

$$M_X(t) = \frac{2}{1 + e^{-t}}.$$

Observe that the magnitude function is only well-defined for metric spaces X such that the similarity matrix Z_{tX} is invertible for all $t > 0$. It can be shown by example that invertibility of $Z_{t_0 X}$ for some particular value of t_0 is not enough to imply invertibility at all t.

The behavior of the magnitude function for several explicit examples, as well as numerical evidence, suggests the following:

Main Question: Is M *monotone increasing* when we assume that distance d comes from an embedding of X *into a Euclidean space*?

5.4.2 Progress

We were able to answer the main question in the affirmative in a few very specific cases. The general question remains open. In this subsection, we present our partial results.

Two point space

For a two point space, with given distance d between the two points, we have $D = \begin{pmatrix} 0 & d \\ d & 0 \end{pmatrix}$. Below, we use tZ to denote the similarity matrix Z_{tX}, to simplify notation. Then

$$tZ = \begin{pmatrix} 1 & e^{-td} \\ e^{-td} & 1 \end{pmatrix} \qquad \text{and} \qquad (tZ)^{-1} = \begin{pmatrix} 1 & -e^{-td} \\ -e^{-td} & 1 \end{pmatrix},$$

so that

$$|tX| = \frac{2}{1-e^{-2td}}(1+e^{-td}) = 2(1+e^{-td})^{-1},$$

and

$$\frac{d}{dt}|tX| = 2de^{-dt}(e^{-dt}+1)^{-2}.$$

Since $\frac{d}{dt}|tX|$ is positive (as $d > 0$ and $e^{-dt} > 0$), the magnitude function of X is monotonically increasing.

Three point equilateral space

Next, consider the equilateral triangle space $X = \{x_1, x_2, x_3\}$ with $d(x_i, x_j) = 1$ for $i \neq j$. Note that we will later show that magnitude is monotone increasing for a general equilateral space (i.e., on an arbitrary number of points), but we begin with this hands-on example. For the three-point space tX, the similarity matrix is given by

$$tZ = \begin{pmatrix} 1 & e^{-t} & e^{-t} \\ e^{-t} & 1 & e^{-t} \\ e^{-t} & e^{-t} & 1 \end{pmatrix}.$$

The inverse of tZ is:

$$(tZ)^{-1} = \frac{e^t}{e^t + e^2t - 2}\begin{pmatrix} e^t+1 & -1 & -1 \\ -1 & e^t+1 & -1 \\ -1 & -1 & e^t+1 \end{pmatrix}.$$

To get the magnitude function, we sum all the entires, which results in the function

$$M(t) = \frac{3e^t(e^t-1)}{e^t+e^{2t}-2}.$$

The derivative of this function is

$$M'(t) = \frac{6e^t}{(e^t+2)^2}.$$

Thus $M'(t) \geq 0$, so $M(t)$ is monotone increasing – see Figure 1.

Finite equilateral space

Before proceeding, let us make some general observations about the magnitude function. Let X be a metric space such that tZ is invertible for all $t > 0$ and let $\underline{u} = (1, \cdots, 1)^T$. Then

$$|tX| = \sum_{i,j}((tZ)^{-1})_{i,j} = \underline{u}^T(tZ)^{-1}\underline{u}.$$

We want to show that $|tX|$ is non-decreasing by showing that $\frac{d}{dt}|tX| = |tX|'$ is non-negative. Note that

$$\begin{aligned} |tX|' &= \left(\underline{u}^T(tZ)^{-1}\underline{u}\right)' \\ &= \underline{u}^T\left((tZ)^{-1}\right)'\underline{u} \\ &= -\underline{u}^T(tZ)^{-1}(tZ)'(tZ)^{-1}\underline{u}. \end{aligned}$$

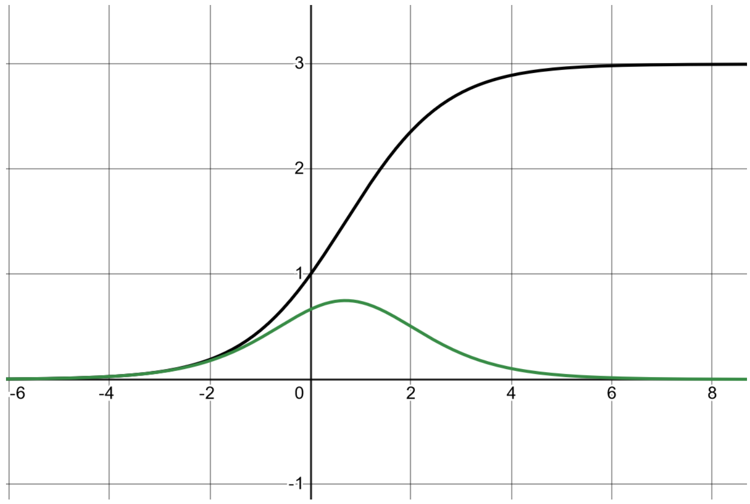


Figure 1 The black curve is $M(t)$ for the three point equilateral space and the green curve is $M'(t)$. Note that $\lim_{t\to 0} M(t) = 1$ and $\lim_{t\to\infty} M(t) = 3$ as we expect based on the informal meaning of magnitude described above.

Now we specialize and generalize the three-point equilateral space considered above. Let X denote the metric space on n points such that $d(x_i, x_j) = a$ for all $i \neq j$. In this case, tZ and $(tZ)'$ both have $\underline{u}$ as an eigenvector, and the associated eigenvalue is the row sum (all rows have the same sum) of the corresponding matrix. For any matrix A, let $R_k(A) = \sum_i A_{ki}$ be the sum of the k-th row of A. Then,

$$(tZ)\underline{u} = R_1(tZ)\underline{u} = \left(1 + (n-1)e^{-at}\right)\underline{u}$$

implies

$$(tZ)^{-1}\underline{u} = \frac{1}{(1 + (n-1)e^{-at})}\underline{u}$$

and

$$(tZ)'\underline{u} = R_1((tZ)')\underline{u} = -(n-1)ae^{-at}\underline{u}.$$

Thus, noting that $(tZ)^{-1}$ is symmetric and hence $\underline{u}^T(tZ)^{-1} = ((tZ)^{-1}\underline{u})^T$, we have

$$\begin{aligned}|tX|' &= -\underline{u}^T(tZ)^{-1}(tZ)'(tZ)^{-1}\underline{u} \\ &= -\frac{1}{(1+(n-1)e^{-at})}(-(n-1)ae^{-at})\frac{1}{(1+(n-1)e^{-at})}\underline{u}^T\underline{u} \\ &= \frac{(n-1)ae^{-at}}{(1+(n-1)e^{-at})^2}n > 0.\end{aligned}$$

This proves that for any finite equilateral space, its magnitude is monotonically increasing.

The obstruction in extending this method to non-equilateral cases is due to the fact that $\underline{u}$ is not necessarily an eigenvector of tZ when considering an arbitrary finite metric space.

Numerical Experiments

We performed some simple numerical experiments whose results agree with our assertion that the main question should be answered in the affirmative. Fixing a dimension k and number of points n, we sampled n points from $\mathbb{R}^k$ to create a finite Euclidean metric

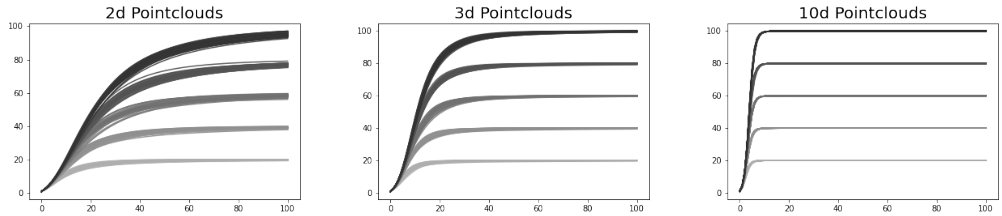


Figure 2 Results of our numerical experiments. The horizontal axis is the t-axis, and the functions plotted are magnitude functions $t \mapsto |tX|$ for various Euclidean pointclouds X. See the text for details.

space X, then computed the magnitude function $M(t) = |tX|$ (at densely and uniformly sampled values of $t \in (0, 100]$). We performed this experiment for $k \in \{2, 3, 10\}$ and $n \in \{20, 40, 60, 80, 100\}$, and for each choice of parameters, we repeated the experiment 50 times. Plots of the resulting magnitude functions are shown in Figure 2. Observe that in each case, the magnitude functions are monotonically increasing.

References

1. Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.
2. Nieves Atienza, Rocio Gonzalez-Diaz, and Matteo Rucco. Persistent entropy for separating topological features from noise in Vietoris-Rips complexes. *Journal of Intelligent Information Systems*, 52(3):637–655, 2019.
3. Nieves Atienza, Rocio Gonzalez-Díaz, and Manuel Soriano-Trigueros. On the stability of persistent entropy and new summary functions for topological data analysis. *Pattern Recognition*, 107, 2020.
4. Vincent Barral and Silvia Biasotti. 3D shape retrieval and classification using multiple kernel learning on extended reeb graphs. *The Visual Computer*, 30(11):1247–1259, 2014.
5. Ulrich Bauer, Magnus Bakke Botnan, and Benedikt Fluhr. Structure and interleavings of relative interlevel set cohomology. *arXiv preprint arXiv:2108.09298*, 2021.
6. Ulrich Bauer, Carsten Lange, and Max Wardetzky. Optimal Topological Simplification of Discrete Functions on Surfaces. *Discrete & Computational Geometry*, 47(2):347–377, 2012.
7. Silvia Biasotti, Andrea Cerri, Patrizio Frosini, and Daniela Giorgi. A new algorithm for computing the 2-dimensional matching distance between size functions. *Pattern Recognition Letters*, 32:1735–1746, 10 2011.
8. Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is np-hard. *Foundations of Computational Mathematics*, 20:1237–1271, 2018.
9. Brian Bollen, Erin Chambers, Joshua A. Levine, and Elizabeth Munch. Reeb graph metrics from the ground up. *arXiv preprint arXiv:2110.05631*, 2022.
10. Adam Brown, Omer Bobrowski, Elizabeth Munch, and Bei Wang. Probabilistic convergence and stability of random mapper graphs. *Journal of Applied and Computational Topology*, 5(1):99–140, 2021.
11. Peter Bubenik. The persistence landscape and some of its properties. In Nils A. Baas, Gunnar E. Carlsson, Gereon Quick, Markus Szymik, and Marius Thaule, editors, *Topological Data Analysis. Abel Symposia*, volume 15, pages 97–117. Springer, Cham, 2020.

12 Gunnar Carlsson, Vin de Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the 25th Annual Symposium on Computational Geometry*, pages 247–256, New York, NY, USA, 2009.
13 Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009.
14 Hamish Carr and David Duke. Joint Contour Nets. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1100–1113, 2013.
15 Mathieu Carriére and Steve Oudot. Structure and stability of the one-dimensional mapper. *Foundations of Computational Mathematics*, 18(6):1333–1396, 2018.
16 Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, Cham, 2016.
17 Frédéric Chazal, Ruqi Huang, and Jian Sun. Gromov-Hausdorff approximation of filamentary structures using Reeb-type graphs. *Discrete & Computational Geometry*, 53(3):621–649, 2015.
18 Tamal K. Dey, Facundo Mémoli, and Yusu Wang. Mutiscale mapper: A framework for topological summarization of data and maps. *Proceedings of the 27th annual ACM-SIAM symposium on Discrete algorithms*, pages 997–1013, 2016.
19 Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Graph reconstruction by discrete morse theory. In *Proc. 34th Intl. Sympos. Comput. Geom. (SoCG)*, pages 31:1–31:15, 2018.
20 Paweł Dłotko. Ball mapper: a shape summary for topological data analysis. *arXiv: 1901.07410*, 2019.
21 Herbert Edelsbrunner, John Harer, and Amit K Patel. Reeb spaces of piecewise linear mappings. In *Proceedings of the 24th Annual Symposium on Computational Geometry*, pages 242–250, College Park, MD, USA, 2008.
22 Zhao Geng, David J Duke, Hamish A Carr, and Amit Chattopadhyay. Visual analysis of hurricane data using joint contour net. In *TPCG*, pages 33–40, 2014.
23 Michael Kerber and Arnur Nigmetov. Efficient approximation of the matching distance for 2-parameter persistence. In Sergio Cabello and Danny Z. Chen, editors, *Proceedings of the 36th International Symposium on Computational Geometry*, volume 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
24 Woojin Kim and Facundo Mémoli. Generalized persistence diagrams for persistence modules over posets. *Journal of Applied and Computational Topology*, 5(4):533–581, 2021.
25 Tom Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013.
26 Tom Leinster and Simon Willerton. On the asymptotic magnitude of subsets of euclidean space. *Geometriae Dedicata*, 164:287–310, 2013.
27 Mingzhe Li, Sourabh Palande, Lin Yan, and Bei Wang. Sketching merge trees for scientific visualization. In *IEEE Workshop on Topological Data Analysis and Visualization (TopoInVis) at IEEE VIS*, pages 61–71, 2023.
28 Nikola Milosavljević, Dmitriy Morozov, and Primoz Skraba. Zigzag persistent homology in matrix multiplication time. In *Proceedings of the 27th Annual Symposium on Computational Geometry*, pages 216–225, 2011.
29 Valerio Pascucci, Giorgio Scorzelli, Peer-Timo Bremer, and Ajith Mascarenhas. Robust on-line computation of reeb graphs: Simplicity and speed. *ACM Trans. Graph.*, 26(3):58–es, jul 2007.
30 Emilie Purvine, Davis Brown, Brett Jefferson, Cliff Joslyn, Brenda Praggastis, Archit Rathore, Madelyn Shapiro, Bei Wang, and Youjia Zhou. Experimental observations of the topology of convolutional neural network activations. *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 2023.

31 Archit Rathore, Nithin Chalapathi, Sourabh Palande, and Bei Wang. TopoAct: Visually exploring the shape of activations in deep learning. *Computer Graphics Forum (CGF)*, 40(1):382–397, 2021.
32 Archit Rathore, Yichu Zhou, Vivek Srikumar, and Bei Wang. TopoBERT: Exploring the topology of fine-tuned word representations. *Information Visualization*, 2023.
33 Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Eurographics Symposium on Point-Based Graphics*, pages 91–100, 2007.
34 Dilip Mathew Thomas and Vijay Natarajan. Symmetry in scalar field topology. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2035–2044, 2011.
35 Lin Yan, Talha Bin Masood, Raghavendra Sridharamurthy, Farhan Rasheed, Vijay Natarajan, Ingrid Hotz, and Bei Wang. Scalar field comparison with topological descriptors: Properties and applications for scientific visualization. *Computer Graphics Forum (CGF)*, 40(3):599–633, 2021.
36 Qi Zhao and Yusu Wang. Learning metrics for persistence-based summaries and applications for graph classification. In *Advances in Neural Information Processing Systems*, pages 9859–9870, 2019.
37 Youjia Zhou, Nathaniel Saul, Ilkin Safarli, Bala Krishnamoorthy, and Bei Wang. Stitch fix for mapper and topological gains. In Ellen Gasparovic, Vanessa Robins, and Kate Turner, editors, *Research in Computational Topology 2*, volume 30 of *Association for Women in Mathematics Series*, pages 265–294. Springer International Publishing, 2021.

Participants

- Ulrich Bauer
TU München, DE
- Julian Brüggeman
Universität Bonn, DE
- Mathieu Carrière
Centre Inria d'Université Côte d'Azur – Sophia Antipolis, FR
- Tamal K. Dey
Purdue University – West Lafayette, US
- Paweł Dłotko
Polish Academy of Science, PL
- Dmitry Feichtner-Kozlov
Universität Bremen, DE
- Eric Goubault
Ecole Polytechnique – Palaiseau, FR
- Robert Green
University at Albany, US
- Teresa Heiss
Institute of Science and Technology Austria, AT
- Niklas Hellmer
Polish Academy of Science, PL
- Sara Kališnik Hintz
ETH Zürich, CH
- Michael Kerber
TU Graz, AT
- Woojin Kim
Duke University – Durham, US & KAIST – Daejeon, KR
- Ingrid Hotz
Linköping University, SE
- Claudia Landi
University of Modena, IT
- Fabian Lenzen
TU Berlin, DE
- Michael Lesnick
University at Albany, US
- Roy Meshulam
Technion – Haifa, IL
- Elizabeth Munch
Michigan State University, US
- Tom Needham
Florida State University – Tallahassee, US
- Sergio Rajsbaum
National Autonomous University of Mexico, MX
- Martin Raussen
Aalborg University, DK
- Bastian Rieck
Helmholtz Zentrum München, DE
- Nicholas Scoville
Ursinus College – Collegeville, US
- Jan Felix Senge
Polish Academy of Science, PL & Universität Bremen, DE
- Anastasios Stefanou
Universität Bremen, DE
- Yusu Wang
University of California, San Diego – La Jolla, US
- Bei Wang Phillips
University of Utah – Salt Lake City, US
- Leonard Wienke
Universität Bremen, DE
- Ling Zhou
Duke University – Durham, US