Higher-Order Graph Models: From Theoretical Foundations to Machine Learning

Edited by

Tina Eliassi-Rad[1], Vito Latora[2], Martin Rosvall[3], and Ingo Scholtes [*][4]

1   **Northeastern University – Boston, US,** `tina@eliassi.org`
2   **Queen Mary University of London, GB,** `v.latora@qmul.ac.uk`
3   **University of Umeå, SE,** `martin.rosvall@umu.se`
4   **Julius-Maximilians-Universität Würzburg, DE & Universität Zürich, CH,**
    `ingo.scholtes@uni-wuerzburg.de`

─── **Abstract** ───

Graph and network models are essential for data science applications in computer science, social sciences, and life sciences. They help to detect patterns in data on dyadic relations between pairs of genes, humans, or documents, and have improved our understanding of complex networks across disciplines. While the advantages of graph models of relational data are undisputed, we often have access to data with multiple types of higher-order relations not captured by simple graphs. Such data arise in social systems with non-dyadic or group-based interactions, multi-modal transportation networks with multiple connection types, or time series containing specific sequences of nodes traversed on paths. The complex relational structure of such data questions the validity of graph-based data mining and modelling, and jeopardises interdisciplinary applications of network analysis and machine learning.

To address this challenge, researchers in topological data analysis, network science, machine learning, and physics recently started to generalise network analysis to higher-order graph models that capture more than dyadic relations. These higher-order models differ from standard network analysis in assumptions, applications, and mathematical formalisms. As a result, the emerging field lacks a shared terminology, common challenges, benchmark data and metrics to facilitate fair comparisons. By bringing together researchers from different disciplines, Dagstuhl Seminar 21352 "Higher-Order Graph Models: From Theoretical Foundations to Machine Learning" aimed at the development of a common language and a shared understanding of key challenges in the field that foster progress in data analytics and machine learning for data with complex relational structure. This report documents the program and the outcomes of this seminar.

─────────────

1 Executive Summary

Ingo Scholtes (Julius-Maximilians-Universität Würzburg, DE & Universität Zürich, CH)
Tina Eliassi-Rad (Northeastern University – Boston, US)
Vito Latora (Queen Mary University of London, GB)
*Martin Rosvall (University of Umeå, SE)*

The network science and graph mining community has created a rich portfolio of data analysis and visualisation techniques that have become a cornerstone for knowledge extraction from relational data on complex systems. Most of those techniques build on simple graph abstractions, where nodes represent a system's elements, and links represent *dyadic interactions, relations, or dependencies* between those elements. This mathematical formalism has proven useful for reasoning, e.g., about the centrality of nodes, the evolution and control of dynamical processes, and the community or cluster structure in complex systems, given that we have access to *relational data* [17]. However, the graph abstractions used in those methods typically do not account for **higher-order relations** between nodes that are present in many real complex systems. Important examples for such data include:

- relational data that is inherently non-dyadic, such as (unordered) sets of authors co-authoring scientific articles, protein triplets in a cell that simultaneously interact with each other, or actors in social systems engaging in group collaborations,
- time-stamped data on social networks with chronologically ordered sequences of (dyadic) interactions, where specific sequences of nodes interact via *causal paths*
- sequential data on networked systems, such as user click streams, mobility trajectories, financial transaction sequences, citation paths, or directed acyclic graphs that give rise to a chronologically or topologically ordered sequences of nodes traversed by processes
- data on networked systems with multiple types or layers of links that cannot be reduced to a simple graph model

Over the past years, researchers have shown that the presence of such higher-order interactions can fundamentally alter our understanding of complex systems. They can change our notion of the importance of nodes captured by centrality measures, affect the detection of cluster and community structures in graphs, and influence dynamical processes like diffusion or epidemic spreading, as well as associated control strategies in non-trivial ways [24, 28, 33, 4, 5, 10, 35]. To further develop graph-based representations of data and broaden their potential application in pattern recognition, data analysis, and machine learning, over the past few years researchers have developed a rich portfolio of **higher-order network models and representations** that capture more than just dyadic dependencies in complex systems. The organisers of this seminar have recently summarised current research and open challenges in this area in three independent overview and perspective articles [1, 30, 15]. An incomplete list of approaches explored over the past few years include:

- hypergraphs, where each *hyperedge* can connect an arbitrary number of nodes [11]
- simplicial network models, where *simplices* represent d-dimensional group interactions [12, 10]
- d-dimensional De Bruijn graphs, where edges capture *ordered* sets of d dyadic interactions [28, 16]
- memory networks, where memory nodes capture Non-Markovian properties in time series data [24]
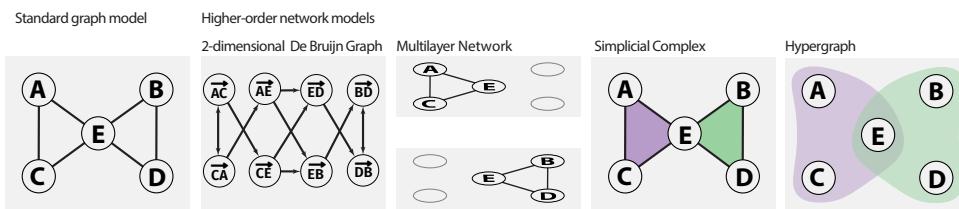
■ **Figure 1** Illustration of standard graph model (left) and four modelling approaches capturing different types of higher-order interactions proposed in topological data analysis, network science, and computer science. Figure adapted from [15].

- higher-, variable-, and multi-order Markov models for temporal networks [33, 20, 4]
- multi-layer and multiplex networks with multiple types of links between nodes [13]
- applications of categorical sequence mining techniques to model patterns in sequences of node sets [7]

In Figure 1, we illustrate some of the higher-order graph models listed above. All these modelling approaches address the same fundamental limitation of graph models when studying complex systems: **we cannot understand a system's structure and dynamics by decomposing direct and indirect interactions between elements into a set of dyadic relations with a single type**. However, the similarities and differences between these different approaches are still not fully understood.

At a critical time for the community, this Dagstuhl seminar intended to improve our understanding of the strengths, weaknesses, commonalities, and differences of these different approaches along with their resulting computational and epistemological challenges. The seminar aimed to create a common foundation for developing graph mining and machine learning techniques that use recent advances in the study of higher-order graph models by gathering key researchers from different communities, including machine learning, information retrieval and data mining, complex systems theory, theoretical physics, network science, computational social science, and mathematics. The participants included senior and junior researchers focusing on four related and intersecting topics: (i) Topological and Graph-Theoretic Foundations, (ii) Higher-Order Models for Dynamical Processes, (iii) Higher-Order Pattern Recognition and Machine Learning, (iv) Computational Aspects in Higher-Order Graph Analysis and Graph Mining.

The organisers used the four topics to structure the seminar program and derive the participants' initial assignment to possible working groups. After an initial round of brief opening statements, participants introduced themselves and stated their specific interests for the seminar during five-minute lightning talks. During a match-making session taking place in the afternoon of day one, all interests expressed by the participants were consolidated into a set of working groups, addressing the following six areas: (i) Visualisation and Interpretability of Higher-Order Graph Models, (ii) Learning and Model Selection, (iii) Unification of Different Higher-Order Modelling Frameworks, (iv) Benchmark Data and Evaluation Practices, (v) Applications of Higher-Order Graph Models, and (vi) Societal Impact, Robustness, and Fairness. In the remaining time of the seminar, participants worked on those issues in the groups. This report includes summaries of the opening statements, the results of the working groups, and a summary of a panel discussion taking place on the evening of day two.

References

  1 Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 2020.

2 Austin R. Benson. Tools for higher-order network analysis. *CoRR*, abs/1802.06820, 2018.

3 Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, 2018.

4 Austin R. Benson, David F. Gleich, and Jure Leskovec. Tensor spectral clustering for partitioning higher-order network structures. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 118–126, 2015.

5 Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

6 Austin R. Benson, David F. Gleich, and Lek-Heng Lim. The spacey random walk: A stochastic process for higher-order data. *SIAM Review*, 59(2):321–345, 2017.

7 Austin R Benson, Ravi Kumar, and Andrew Tomkins. Sequences of sets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1148–1157, 2018.

8 Matthias Bolten, Stephanie Friedhoff, Andreas Frommer, Matthias Heming, and Karsten Kahl. Algebraic multigrid methods for laplacians of graphs. *Linear Algebra and its Applications*, 434(11):2225 – 2243, 2011. Special Issue: Devoted to the 2nd NASC 08 Conference in Nanjing (NSC).

9 Daniel Edler, Ludvig Bohlin, et al. Mapping higher-order network flows in memory and multilayer networks with infomap. *Algorithms*, 10(4):112, 2017.

10 Ernesto Estrada and Grant J. Ross. Centralities in simplicial complexes. applications to protein interaction networks. *Journal of Theoretical Biology*, 438:46 – 60, 2018.

11 Gourab Ghoshal, Vinko Zlatić, Guido Caldarelli, and Mark EJ Newman. Random hypergraphs and their applications. *Physical Review E*, 79(6):066118, 2009.

12 Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature communications*, 10(1):1–9, 2019.

13 Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014.

14 Renaud Lambiotte, Martin Rosvall, Michael Schaub, Ingo Scholtes, and Jian Xu. Beyond graph mining: Higher-order data analytics for temporal network data. In *Hands-On Tutorial at the 24th ACM SIGKDD International Conference on Knowledge Discovery*, volume 38, 2018.

15 Renaud Lambiotte, Martin Rosvall, and Ingo Scholtes. From networks to optimal higher-order models of complex systems. In *Nature physics*, Vol. 15, p. 313-320, March 25 2019

16 Timothy LaRock, Vahan Nanumyan, Ingo Scholtes, Giona Casiraghi, Tina Eliassi-Rad, and Frank Schweitzer. HYPA: Efficient detection of path anomalies in time series data on networks. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 460–468. SIAM, 2020.

17 Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex Networks: Principles, Methods and Applications*. Cambridge University Press, 2017.

18 Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics Reports*, 716-717:1 – 58, 2017. Random walks and diffusion on networks.

19 Huda Nassar, Caitlin Kennedy, Shweta Jain, Austin R. Benson, and David F. Gleich. Using cliques with higher-order spectral embeddings improves graph visualizations. In *Proceedings of the 2020 Web Conference (WWW)*, pages 2927–2933, 2020.

20 Tiago P Peixoto and Martin Rosvall. Modelling sequences and temporal networks with dynamic community structures. *Nature communications*, 8(1):1–12, 2017.

21 Vincenzo Perri and Ingo Scholtes. Hotvis: Higher-order time-aware visualisation of dynamic graphs. In *Proceedings of the 28th International Symposium on Graph Drawing and Network Visualization (Graph Drawing 2020), Vancouver, BC, Canada (to appear)*, 2020.

22 Luka V. Petrovic and Ingo Scholtes. PaCo: Fast Counting of Causal Paths in Temporal Network Data. In *Proceedings of the 11th Temporal Web Analytics Workshop (TempWeb 2021)* held in conjunction with The Web Conference 2021, Ljubljana, Slovenia, April 2021

23 Luka V. Petrovic and Ingo Scholtes. Learning the markov order of paths in a network, 2020. arXiv 2007.02861.

24 Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications*, 5:4630, 2014.

25 Mandana Saebi, Giovanni Luca Ciampaglia, Lance M Kaplan, and Nitesh V Chawla. Honem: Network embedding using higher-order patterns in sequential data. *arXiv preprint arXiv:1908.05387*, 2019.

26 Mandana Saebi, Jian Xu, Lance M. Kaplan, Bruno Ribeiro, and Nitesh V. Chawla. Efficient modeling of higher-order dependencies in networks: from algorithm to application for anomaly detection. *EPJ Data Sci.*, 9(1):15, 2020.

27 Ingo Scholtes. When is a network a network? multi-order graphical model selection in pathways and temporal networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 1037–1046, New York, NY, USA, 2017. Association for Computing Machinery.

28 Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature communications*, 5:5024, 2014.

29 Yingxia Shao, Shiyue Huang, Xupeng Miao, Bin Cui, and Lei Chen. Memory-aware framework for efficient second-order random walk on large graphs. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, page 1797–1812, 2020.

30 Leo Torres, Ann S Blevins, Danielle S Bassett, and Tina Eliassi-Rad. The why, how, and when of representations for complex systems. *arXiv preprint arXiv:2006.02870*, 2020.

31 Francesco Tudisco, Austin R. Benson, and Konstantin Prokopchik. Nonlinear higher-order label spreading. *CoRR*, abs/2006.04762, 2020.

32 Tao Wu, Austin R. Benson, and David F. Gleich. General tensor spectral co-clustering for higher-order data. In *Advances in Neural Information Processing Systems 29*, pages 2559–2567, 2016.

33 Jian Xu, Thanuka L Wickramarathne, and Nitesh V Chawla. Representing higher-order dependencies in networks. *Science advances*, 2(5):e1600028, 2016.

34 Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 555–564, 2017.

35 Yan Zhang, Antonios Garas, and Ingo Scholtes. Higher-order models capture changes in controllability of temporal networks. In Journal of Physics: Complexity, Vol. 2, No. 1, January 29, 2021

2   Table of Contents

3 Overview of Talks

3.1 Inference of Time-ordered Multi-Body Interactions

Unai Alvarez-Rodriguez (Universität Zürich, CH)

Higher-order models are typically specialised in a single class of interaction. Multi-time, multi-system and multi-type modelling approaches have not yet been combined, and therefore there is no framework capable of describing processes that simultaneously manifest different classes of interactions. I argue a unification of higher-order models is necessary to bypass this limitation and to improve our understanding of complex systems. Along these lines, I present preliminary results for extracting time-ordered multi-body interactions from time series of systems composed by multiple interacting elements.

3.2 Cascade Processes in Machine Learning

Rebekka Burkholz (Harvard School of Public Health – Boston, US)

I have proposed to develop a unifying framework to represent higher order network information by parametrising a process that evolves on a network as graph neural network. This could be combined with the design of suitable covariate information that represents the higher order model information and would enable the inference of networks and processes based on data. Yet, in many situations, this unification approach is expected to suffer from overparametrisation leaving the question whether there are better and more parameter efficient representations of higher order structure for a given task. As motivating problem, I have presented recent work about learning the order in which mutations accumulate during cancer progression.

3.3 The Why, How, and When of Representations for Complex Systems

Tina Eliassi-Rad (Northeastern University – Boston, US)

Complex systems, which at the most fundamental level consist of entities and their interactions, describe phenomena in a wide variety of fields, from neuroscience to computer science to economics. The wide variety of applications has led to two key challenges: (1) the

development of many domain-specific strategies for analysing complex systems, and (2) the compartmentalization of representation and analysis within a domain due to inconsistencies in the language for complex systems. In our work, we propose a domain-agnostic language to develop a more coherent vocabulary. We use this language to evaluate each step of the analysis of complex systems. We start with the system under study and its observations in terms of the collected data, and then go through different mathematical frameworks for encoding the observed data (i.e., graphs, simplicial complexes, and hypergraphs) and relevant computational methods for each framework. At each step, we consider different types of dependencies. These are properties of the system that describe how the existence of an interaction between a group of entities in a system can affect the possibility of the existence of another relationship. We discuss how dependencies can arise and how they can change the interpretation of results or the entire analysis pipeline. We conclude with two real-world examples.

3.4   Spreading and Centrality on Hypergraphs

Desmond J. Higham (University of Edinburgh, GB)

We typically interact in groups, not just in pairs. The use of hyperedges naturally allows us to model with a nonlinear rate of transmission, in terms of both the group size and the number of infected group members, as is the case, for example, when social distancing is encouraged. I am therefore interested in individual-level, stochastic disease models on a hypergraph [1, 2]. I am also interested in centrality measures that take account of group interactions, which leads to nonlinear eigenvalue problems, and nonlinear extensions of Perron-Frobenius theory and the power method [3].

References
1   Desmond J. Higham and Henry-Louis de Kergorlay, Epidemics on hypergraphs: Spectral thresholds for extinction, Proceedings of the Royal Society, Series A, 2021
2   Desmond J. Higham and Henry-Louis de Kergorlay, Mean field analysis of hypergraph contagion models, arXiv: 2108.05451, 2021
3   F. Tudisco and Desmond J. Higham, Node and edge eigenvector centrality for hypergraphs, Communications Physics, 2021.

3.5 Interacting Discovery Processes on Complex Networks

Gabriele DiBona (Queen Mary University of London, GB)

In my talk, I focused on the influence of social interactions on collective processes, such as the exploration and the discovery of new content in different contexts. The challenge is now to include group interactions using higher-order methods, with a data-driven approach. This can have implications in phenomena as diverse as user interaction in online social networks, collective decisions in teams, team success and optimal structures, nonlinear random walks, brain analysis in social activities, brain creativity, and diffusion of innovation. A first step in this direction has been done in our recent paper [1].

References
1 Iacopo Iacopini, Gabriele Di Bona, Enrico Ubaldi, Vittorio Loreto, and Vito Latora. *Interacting Discovery Processes on Complex Networks.* Phys. Rev. Lett. 125, 248301, 10 December 2020.

3.6 Dynamical Processes on Higher-Order Networks: Beyond Dyadic Projections

Luca Gallo (University of Catania, IT)

Starting from the current literature about dynamical processes on higher-order networks, I formulate two theoretical questions. (i) Dynamical processes on hypergraphs and on simplicial complexes are usually studied in parallel [3, 2]. Can we produce a general theory of dynamical systems on higher-order networks? In particular, is it possible to point out if and how the absence or the presence of the inclusion requirement impacts the dynamics? (ii) To make the problem analytically feasible, previous efforts in the study of dynamical processes on higher-order structures have relied on the definition of suitable dyadic projections [1, 2, 3], i.e. equivalent weighted networks. However, this method can lose information about the higher-order structure, possibly preventing a complete study of the dynamics [4]. Can we produce an analytical framework that goes beyond projected networks?

References
1 Timoteo Carletti, Duccio Fanelli, and Sara Nicoletti. *Dynamical systems on hypergraphs.* Journal of Physics: Complexity 1(3), 035006 (2020)
2 Guilherme Ferraz de Arruda, Michele Tizzani, and Yamir Moreno. *Phase transitions and stability of dynamical processes on hypergraphs.* Communications Physics 4(1), 1-9, (2021)
3 Lucia Valentina Gambuzza, Francesca Di Patti, Luca Gallo, Stefano Lepri, Miguel Romance, Regino Criado, Mattia Frasca, Vito Latora and Stefano Boccaletti. *Stability of synchronization in simplicial complexes.* Nature communications, 12(1), 1-13, (2021).

4 Anastasiya Salova and Raissa M. D'Souza, *Analyzing states beyond full synchronization on hypergraphs requires methods beyond projected networks.* arXiv preprint arXiv:2107.13712. (2021)

3.7 New Data for Higher-Order Network Research

David F. Gleich (Purdue University – West Lafayette, US)

We discussed some challenges we had in visualising a new higher-order dataset derived as a hypergraph representation of Anthony Fauci's emails regarding the COVID-19 pandemic [2]. Analytic studies of these data show how higher-order features were more stable than their graph counterparts [1]; but in the abstract presentation we highlighted how the lack of hypergraph visualisation tools limited our investigation of the data. This dataset is a more modern counterpart to the famous Karate Club dataset as well as the Enron email dataset. In our working paper, we provide a fully parsed version suitable to derive a number of graph, hypergraph, and other higher-order datasets.

References

1 Austin R. Benson, Nate Veldt, and David F. Gleich. 2021. fauci-email: a json digest of Anthony Fauci's released emails. *arXiv* cs.SI 2108.01239. Published online at `http://arxiv.org/abs/2108.01239` and data available from `https://github.com/nveldt/fauci-email`.
2 Natalie Bettendorf and Jason Leopold. Anthony Fauci's Emails Reveal The Pressure That Fell On One Man. *BuzzFeed News.* June 2021. Published online at `https://www.buzzfeednews.com/article/nataliebettendorf/fauci-emails-covid-response`.

3.8 Combining Higher-Order Graph Models with Expert Knowledge

Christoph Gote (ETH Zürich, CH)

Higher-order correlations facilitate unprecedented insights into system processes. However, to interpret and validate the results, we need both a thorough theoretical understanding of the underlying methods and expert subject matter knowledge. We conjecture that the overlap between groups with expertise regarding both aspects is low. Consequently, we ask how we can increase the visibility and applicability of higher-order methods in other scientific fields.

3.9 Benchmarking and Robustness of Higher-Order Graph Models

Stephan Günnemann (Technical University of Munich, DE)

Machine learning methods taking higher-order network structure into account have the potential to obtain richer and potentially more accurate results by modeling the underlying complex graph data better. To evaluate the real success of such higher-order graph-based ML models, however, fair evaluation and benchmarking principles are required – a non-trivial task even for standard graphs and graph learning models [1]. Indeed, beyond providing suitable benchmark datasets of higher-order graph models, such evaluation practices have to identify common tasks and appropriate baselines specifically tacking the higher-order nature into account and comparing them to standard graph approaches. Moreover, evaluations should not be limited to metrics such as accuracy but specifically the robustness of the models need be considered. While, e.g., standard graph neural networks have been shown to be non-robust [2], it is an open challenge whether higher-order graph structures can make the methods and analysis more reliable and, thus, leading to more robust ML models.

References
1 Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, Stephan Günnemann. Pitfalls of Graph Neural Network Evaluation. *Relational Representation Learning Workshop, Neural Information Processing Systems*, 2018.
2 Stephan Günnemann. Graph Neural Networks: Adversarial Robustness. *Graph Neural Networks: Foundations, Frontiers, and Applications*, Springer, chapter 8, 2021.

3.10 (Knowledge | Hyper) Graphs in Social Media and Text

Andreas Hotho (Julius-Maximilians-Universität Würzburg, DE)

In my talk, I focused on three topics related to the seminar. First, hypergraphs are the underlying structure of FolkSonomies, which are behind tagging systems emerged in the Web 2.0 wave. At that time, we started our system BibSonomy which is still online[1, 7]. All the user generated data are freely available for research. Further, I pointed to a couple of results on hypergraphs, e.g. our analysis of the graph structure [2], the behaviour analysis together with Markus Strohmaier [5] and the emergent semantics in the systems [6]. We also developed new ranking and recommendation algorithm. The second topic is on the edge of graphs and natural languages processing (NLP). I show two showcases, the analysis of plots on German novels and dime novels and the emergent languages and communication patterns in the chat messages of the twitch.tv platform [3]. Third, knowledge graphs well known in the semantic web community and widely adopted in many other areas are another graph structure of interest. By integrating KGs with languages models like BERT or GPT, the graph structure is becoming even more interesting for the higher order graph community [4].

References

1 Dominik Benz, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz, and Gerd Stumme. The social bookmark and publication management system bibsonomy. *The VLDB Journal*, 19(6):849–875, December 2010.

2 Ciro Cattuto, Christoph Schmitz, Andrea Baldassarri, Vito D. P. Servedio, Vittorio Loreto, Andreas Hotho, Miranda Grahl, and Gerd Stumme. Network properties of folksonomies. *AI Communications*, 20(4):245 – 262, 2007.

3 Konstantin Kobs, Albin Zehe, Armin Bernstetter, Julian Chibane, Jan Pfister, Julian Tritscher, and Andreas Hotho. Emote-controlled: Obtaining implicit viewer feedback through emote based sentiment analysis on comments of popular twitch.tv channels. *ACM Transactions on Social Computing*, 3(2):1–34, May 2020.

4 Janna Omeliyanenko, Albin Zehe, Lena Hettinger, and Andreas Hotho. Lm4kg: Improving common sense knowledge graphs with language models. In Jeff Z. Pan, Valentina Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal, editors, *The Semantic Web – ISWC 2020*, pages 456–473, Cham, 2020. Springer International Publishing.

5 P. Singer, D. Helic, A. Hotho, and M. Strohmaier. Hyptrails: A bayesian approach for comparing hypotheses about human trails. In *24th International World Wide Web Conference (WWW2015)*, Firenze, Italy, May 18 – May 22 2015. ACM, ACM.

6 Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic grounding of tag relatedness in social bookmarking systems. In *The Semantic Web – ISWC 2008*, volume 5318 of *Lecture Notes in Computer Science*, pages 615–631. Springer Berlin / Heidelberg, 2008.

7 Daniel Zoller, Stephan Doerfel, Robert Jäschke, Gerd Stumme, and Andreas Hotho. Posted, visited, exported: Altmetrics in the social tagging system bibsonomy. *Journal of Informetrics*, 10(3):732 – 749, 2016.

3.11 Inequalities and Higher-Order Interactions

Fariba Karimi (Complexity Science Hub – Wien, AT)

In my talk I propose to consider the types of inequalities that are hidden in higher order interactions that we would miss if we don't consider them. For example, the presence of multiple groups of various size and mixing patterns between groups may cause certain types of hypergraphs representations and result in specific group dynamics. I am interested in developing network models that would consider higher order interactions and use that to understand the emergence of inequalities in society and algorithms.

3.12 Higher-Order Processes in Complex Systems

Vito Latora (Queen Mary University of London, GB)

Dynamical processes on systems with higher-order interactions and/or systems with higher-order temporal dependencies can help to understand the neural and social components of creativity. In this talk I will show some examples of models of collective exploration [1], of

social interactions [2] and social contagion [3] that can be generalised to take into account higher-order interactions and higher-order temporal dependencies [4]. I will also point to some examples of possible experiments to test the effects of higher-order interactions on the dynamics of social systems.

References

1   Iacopini, Di Bona, Ubaldi, Loreto and Latora, 2020, Interacting Discovery Processes on Complex Networks. Phys. Rev. Lett. 125, 248301.
2   Evolutionary dynamics of higher-order interactions in social networks (2021) Alvarez-Rodriguez, Battiston, Ferraz de Arruda, Moreno, Perc, Latora, Nat. Hum. Behav. 5, 586.
3   Iacopini, Petri, Barrat and Latora, 2019. Simplicial models of social contagion. Nat. Comm. 10(1), pp.1-9.
4   Mazzarisi, Lillo, Williams and Latora, Non-Markovian temporal networks with auto- and cross-correlated link dynamics, arXiv preprint arXiv:1909.08134.

3.13   Simplicial Network Analysis Based on Electrical Networks

Kang-Ju Lee (Seoul National University, KR)

I introduce network invariants based on simplicial electrical networks. Effective resistance also known as resistance distance measures how well currents generated by an edge between two vertices as a battery are resisted. Under d-dimensional Kirchhoff's current and voltage laws, we introduce simplicial effective resistance among d+1 vertices [1]. We make use of our measure to propose a simplicial analogue of current-flow closeness centrality or information centrality. We define the simplicial Kirchhoff index as a robustness measure for simplicial networks [2]. We also propose a high-dimensional generalisation of the concept of the number of connected components.

One of the advantages of using simplicial complexes is that we can utilise tools from algebraic topology. Generalising studies in network theory for 1-cycles or flows to simplicial networks will take advantage of it. Finding data set concerning high-dimensional cycles or flows will support these studies.

References

1   Woong Kook and Kang-Ju Lee. Simplicial networks and effective resistance, *Advances in Applied Mathematics* 100 (2018) 71-86.
2   Woong Kook and Kang-Ju Lee. Kirchhoff index of simplicial networks, *Linear Algebra and its Applications* 626 (2021) 1-19.

3.14 Machine Learning for Networks

Lisi Qarkaxhija (Koper, SI)

I am a recent Master's graduate in the field of Data Science were my main priority was machine learning on networks. Before that, I completed a bachelor's degree in Mathematics. In my lightning talk, I introduced myself as a soon-to-be doctorate researcher in the field of Machine Learning for Complex Networks. As such, I established my interest in research concerning higher-order graphs and took the opportunity to familiarise myself with the topic and to form new connections.

3.15 Dynamical Processes on Higher-Order Models: Future Research

Leonie Neuhäuser (RWTH Aachen, DE)

In my talk, I outline two ways of extending higher-order model research, motivated by my previous work on the interplay of dynamics and multi-body topology [1, 2]. Firstly, we have to consider the practicability of higher-order models. The overall system is often determined by an interplay of many model aspects (topology, temporal ordering, type of dynamics) and we need to detect which of these interactions aspects are qualitatively impacting the specific research question of interest. For this, it is important to consider both domain expert knowledge and model expert knowledge. Another interesting question is the interplay of different higher-order dimensions. Current methods are mainly focusing on one specific higher-order aspect, but different aspects may interact. We have investigated the interplay of temporal and multi-way interactions in [3] and found effects, that differ from their projections. This call for more research on the combination of different higher-order model facets.

References

1 Leonie Neuhäuser, Andrew Mellor, and Renaud Lambiotte. Multibody interactions and nonlinear consensus dynamics on networked systems. Physical Review E 101, 2020.
2 Leonie Neuhäuser, Renaud Lambiotte and Michael T. Schaub. Consensus Dynamics and Opinion Formation on Hypergraphs. arXiv:2105.01369, 2021.
3 Leonie Neuhäuser, Michael T. Schaub and Renaud Lambiotte. Consensus Dynamics on Temporal Hypergraphs. arXiv:2109.04985, 2021.

3.16 How, when, and which Higher-Order Models can we use?

Vincenzo Perri (Universität Zürich, CH)

The use of machine learning tools provides a fruitful way for the analysis of network systems. Unlike standard network models, the application of these tools to higher-order models is neither unified nor straightforward. This difficulty comes from the existence of multiple ways to extend these tasks on higher-order networks and a lack of understanding of the commonalities between the different types of higher-order models. In light of this, I am interested in examining the commonalities between higher-order methods and their possibilities for applications.

3.17 The Role of Higher-Order Interactions in Complex Systems

Giovanni Petri (ISI Foundation – Torino, IT), Federico Battiston (Central European University – Vienna, AT), Ginestra Bianconi (Queen Mary University of London, GB), Vito Latora (Queen Mary University of London, GB), and Yamir Moreno (University of Zaragoza, ES)

Complex networks have become the main paradigm for modelling the dynamics of interacting systems. However, networks are intrinsically limited to describing pairwise interactions, whereas real-world systems are often characterised by higher-order interactions involving groups of three or more units. Higher-order (polyadic) structures are therefore a better tool to map the real organisation of many social, biological and man-made systems. Here I outline key challenges for the physics of higher-order systems.

See [1, 2, 3].

References
1 Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G. and Petri, G., 2020. Networks beyond pairwise interactions: structure and dynamics. Physics Reports, 874, pp.1-92.
2 Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J. and Vaccarino, F., 2014. Homological scaffolds of brain functional networks. Journal of The Royal Society Interface, 11(101), p.20140873.
3 Iacopini, I., Petri, G., Barrat, A. and Latora, V., 2019. Simplicial models of social contagion. Nature communications, 10(1), pp.1-9.

3.18 Data-efficient Model Selection of Higher-order Networks

Luka Petrovic (Universität Zürich, CH)

In my previous work I have focused on statistical inference of higher-order network models for paths. They generally have large parameter spaces, and therefore require large amounts of data for training. We leveraged the fact that many networked systems have topological constraints and devised a Bayesian method to improve data-efficiency of model selection for higher-order network models for paths [1]. We believe that this methodology can improve statistical inference for a broader class of higher-order network models.

References

1 Luka Petrovic, Ingo Scholtes. Learning the Markov order of paths in a network, arXiv:2007.02861, https://arxiv.org/abs/2007.02861

3.19 Efficient Variable-Order Markov Models of Network Flows

Martin Rosvall (University of Umeå, SE), Daniel Edler (University of Umeå, SE), Anton Eriksson (University of Umeå, SE), and Jelena Smiljanic (University of Umeå, SE)

Researchers develop maps that reveal essential patterns in network flows to better understand the flows of ideas or information through social and biological systems. In practice, network flow models have implied memoryless first-order Markov chains. Recently, researchers have introduced higher-order Markov chain models with memory to capture patterns in multi-step pathways, including revealing actual, overlapping community structures. However, higher-order Markov chain models suffer from the curse of dimensionality: their vast parameter spaces require exponentially increasing data to avoid overfitting and therefore make mapping inefficient already for moderate-sized systems. Model selection based on Markov chain state lumping into variable-order Markov chains and cross-validation alleviates this problem but wastes plentiful data. We need more efficient methods for reliably describing higher-order network flows. Two central questions arise: Which algorithm best explores the space of variable-order Markov chain models? How do we incorporate Bayesian methods to select the model that best describes the higher-order network flows?

3.20   What are Higher-Order Models?

Michael Schaub (RWTH Aachen, DE)

In this talk I outlined several different ways in which we may consider higher-order models emerging from considerations of modelling low dimensional geometric structure (modelling nonlinear spaces); higher-order models for modelling non-dyadic relational data (interactions between groups vs interactions between pairs of nodes); and higher-order models for complex data supported on (fixed) domains such as hypergraphs, complexes etc.

References
1   Schaub, M. T., Zhu, Y., Seby, J. B., Roddenberry, T. M., and Segarra, S. *Signal processing on higher-order networks: Livin' on the edge. . . and beyond.* Signal Processing, 187, 108149, 2021
2   Bick, C., Gross, E., Harrington, H. A., and Schaub, M. T. *What are higher-order networks?.* arXiv preprint arXiv:2104.11329, 2021

3.21   Higher-Order Models and Cultural Data Analytics

Maximilian Schich (Tallinn University, EE)

In my talk I first gave a brief intro to the research mission of the CUDAN ERA Chair for Cultural Data Analytics at Tallinn University in Estonia (cf. `https://cudan.tlu.ee`). Second, I have provided some insight into the common roots and shared potential of networks with multiple node and link types, of higher-order topology, and a systematic science of art and culture.

3.22   Opening Talk: The Three Ages of Network Science – A Historical Perspective on Higher-Order Graph Models

Ingo Scholtes (Julius-Maximilians-Universität Würzburg, DE & Universität Zürich, CH)

Starting from a historical perspective on what I propose to consider "three ages" of network science, in the opening talk I gave an overview of different modelling frameworks that address different types of higher-order information and dependencies in complex networks. Addressing the challenge of dyadic interactions with multiple types, a first category of higher-order models includes signed graphs, multiplex networks and multi-layer networks. The second
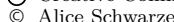
category of models includes simplicial complexes, hypergraphs, and motif-based network models, which can be used to address the challenge of modelling data with polyadic, i.e. non-dyadic, relationships. A third category of models uses higher-order Markov chains, memory networks, or high-dimensional De Bruijn graphs to model higher-order dependencies in time-ordered and sequential data. Following this categorisation, I presented three cross-cutting challenges that require a collaboration between researchers who address these different modelling frameworks. The first challenge addresses the practicality of higher-order models for data science practitioners, e.g., considering computational complexity, data efficiency, model dimensionality, and the need for intuitive and efficient visualisations of high-dimensional models. A second challenge is due to the curse of dimensionality that is common in higher-order models, which introduces the challenge of generalisability and model selection. A third challenge is the development of a unified perspective that combines different higher-order modelling frameworks to address complex data sets like, e.g. time-ordered or multi-type polyadic relationships.

References

1 Jürgen Hackl, Ingo Scholtes, Luka Petrović, Vincenzo Perri, Luca Verginer, Christoph Gote. *Analysis and visualisation of time series data on networks with pathpy*. In Proceedings of the 11th Temporal Web Analytics Workshop (TempWeb 2021) in conjunction with The Web Conference 2021, Ljubljana, Slovenia, April 2021

2 Vincenzo Perri and Ingo Scholtes. *HOTVis: Higher-Order Time-Aware Visualisation of Dynamic Graphs*. In Proceedings of the 28th International Symposium on Graph Drawing and Network Visualization (GD 2020), Vancouver, BC, Canada, September 15-18, 2020

3 Renaud Lambiotte, Martin Rosvall, Ingo Scholtes. *From Networks to Optimal Higher-Order Models of Complex Systems*. In Nature Physics, Vol. 15, p. 313-320, March 25 2019

4 Ingo Scholtes. *When is a Network a Network? Multi-Order Graphical Model Selection in Pathways and Temporal Networks*. In KDD'17 – Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Nova Scotia, Canada, August 13-17, 2017

5 Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio Tessone and Frank Schweitzer. *Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks*, In Nature Communications, Vol. 5, Article 5024, September 24, 2014

3.23 Motifs for Processes on Networks

Alice Schwarze (University of Washington – Seattle, US)

The study of motifs in networks can help researchers uncover links between the structure and function of networks in biology, sociology, economics, and many other areas. Empirical studies of networks have identified feedback loops, feed-forward loops, and several other small structures as "motifs" that occur frequently in real-world networks and may contribute by various mechanisms to important functions in these systems. However, these mechanisms are unknown for many of these motifs. We propose to distinguish between "structure motifs" (i.e., graphlets) in networks and "process motifs" (which we define as structured sets of walks) on

networks and consider process motifs as building blocks of processes on networks. Using the steady-state covariances and steady-state correlations in a multivariate Ornstein–Uhlenbeck process on a network as examples, we demonstrate that the distinction between structure motifs and process motifs makes it possible to gain quantitative insights into mechanisms that contribute to important functions of dynamical systems on networks.

3.24 Higher-Order Models of Group Formation

Frank Schweitzer (ETH Zürich, CH)

Joint work of Frank Schweitzer, Georges Andres
Main reference Frank Schweitzer, Georges Andres: "Social nucleation: Group formation as a phase transition", CoRR, Vol. abs/2107.06696, 2021.
URL https://arxiv.org/abs/2107.06696

I talk about the dynamics of group formation, for a good reason: Group structures can be represented as polyadic interactions and are thus accessible by higher-order network models. But group formation is inherently driven by social mechanisms: homophily, cost/benefit evaluation, restricted access to resources, competition, to name a few. The question is: how are these social mechanisms preserved in a higher-order representation? In my short presentation, I provide a model that works for first-order networks, combining agent-based modelling with rules for network formation. Would similar results be achievable with higher-order models? How should these models look like? Would we gain anything beyond what the first-order network model already provides?

3.25 Higher-Order Models and Responsible Machine Learning

Markus Strohmaier (RWTH Aachen, DE)

In my talk I am exploring issues and challenges related to Responsible Machine Learning on Social Networks. I will argue that traditional methods for evaluating machine learning models need to be expanded to include and consider social challenges such as polarisation, inequality, exclusion or discrimination that are potentially arising from the deployment of machine learning techniques in social settings. I conclude with an outlook of potential avenues for further research.

3.26 Network Evolution and Spacetime Networks as Higher-Order Graphs

Chester Tan (National University of Singapore, SG)

In this lightning talk, I propose the following two questions: (1) Can higher order random network evolution models (e.g. higher order preferential attachment) be represented and analysed as higher order path sequence networks and vice versa. (2) How can higher order networks be represented meaningfully and analyzed usefully in spacetime?

3.27 Generative Models for Higher-Order Interactions

Anatol Wegner (University College London, GB)

The talk briefly introduced generative models for higher order interactions that include interactions that can take the form of any simply connected motifs. These models include a wide variety of higher order structures that go beyond cliques while remaining analytically tractable. I discussed the use of these models in inference based methods that can be used to obtain higher order representations of networks and raised potential applications of such generative models in graph based machine learning.

4 Working groups

4.1 Unification of Higher-Order Models

Unai Alvarez-Rodriguez (Universität Zürich, CH), Ginestra Bianconi (Queen Mary University of London, GB), Natasa Przulj (Barcelona Supercomputing Center, ES), Maximilian Schich (Tallinn University, EE), Alice Schwarze (University of Washington – Seattle, US), Leo Torres (Northeastern University – Boston, US), and Anatol Wegner (University College London, GB)

Despite *unification* being a newcomer to the higher-order jargon, its popularity skyrocketed in Dagstuhl, to the point of being shortlisted as a key discussion topic for the higher-order interactions field. Indeed, it was the preferred option for eight researchers in the working group allocation round, making it the most voted with twice the support of its most successful competitor. These values indicate a substantial acceptance of the unification discourse introduced during the first part of the seminar, where unification was presented as a quest towards a Utopian model merging multi-type, multi-time and multi-body interactions in a single formalism. The minimal desirable purpose of unification would be to develop a shared perspective that clarifies the relation, mutual difference and gap of alternative paradigms, including hypergraphs, simplicial complexes, multi-layer networks and temporal higher-order networks.

The working group on unification was gathered with the goal of highlighting what we stand to gain from the unification of different higher-order models. Our initial exchange was a virtual round table where we took turns to share our views on unification. In this process we identified two opposite aspects to refine in further explorations. The first one is the purely theoretical challenge of merging different types of higher-order approaches for the sake of addressing rich classes of dynamics. The second one is finding practical applications where a unified formalism is preferred over already existing ones. In summary, unification should aim not just for mathematical elegance and instead prove to be useful also for practitioners.

For the remaining days the team organised an asynchronous brain storming (deviating from more-typical synchronous Dagstuhl-style discussion, due to the hybrid nature of the event) to incorporate on-line as well as on-site participants with different time zones. Every member of the group then started a search for potential benefits of unification. This search lead to the following findings:

One of the profitable byproducts of unification is model compatibility. A unified model containing current frameworks as particular cases would provide a common language for higher-order phenomena which may enable combining results obtained independently in different domains of research on higher-order networks.

Another idea that we discussed was the use of unification as a principle for workflow automation. For researchers interested in applying network science to case studies, one of the first important steps is to decide which is the type of higher-order model that best describes a data set. Choosing a wrong model may lead to misleading conclusions about the behaviour of systems. The standard procedure to tackle this problem is employing model selection techniques. A challenge in doing so is that many currently available models are incompatible with each other, because they do not allow mixtures between different features (multi-type,multi-time and multi-body). A unified model would overcome this rigid structure as it would include degrees of freedoms of different features. Furthermore, such model flexibility would also remove the otherwise necessary step of network-type selection and therefore simplify the work of applied researchers.

Knowledge graphs were another topic of the debate. Within the joint discussion, group members Ginestra Bianconi and Maximilian Schich have pointed out the relevance of knowledge graphs, from the perspective of applied mathematics/physics and socio-cultural domain expertise respectively. Within the unification working group, knowledge graphs, and by extension less generalised database models, such as relational databases, have been considered regarding their relevance towards unification of higher-order network research. From a mathematical perspective, knowledge graphs are relevant as their configuration and growth is likely out of sync with existing maximum entropy models of network growth (e.g. adding bespoke link motifs instead of n-simplices). From the perspective of domain experts, ranging from biology to socio-cultural disciplines, developing a deeper understanding of higher-order structure and dynamics of knowledge graphs is a desiderate that seems more or less obvious since about two decades. From this a joint challenge emerges that can now be tackled based on the recent advances of higher-order network science. Consequently a second notion of "unification" emerged in the discussion, where different approaches of higher-order network science can be tested against each other, while looking at knowledge graphs that permeate a broad spectrum of disciplines.

All in all, we were able to ground the original proposal by showcasing specific methodologies that would be improved by a unified model. Our working group concluded that there is a robust motivation for a unification of higher-order models, and that we can anticipate an increase in the research community's efforts towards unification in years to come.

4.2 Social Impact of Higher-Order Models

Leonie Neuhäuser (RWTH Aachen, DE), Fariba Karimi (Complexity Science Hub – Wien, AT), and Markus Strohmaier (RWTH Aachen, DE)

In the working group "Social impact of higher-order models", we discussed which unique challenges and opportunities arise when deploying higher-order approaches to modelling social systems.

First, we identify the potential of higher-order models to better capture the rich subtleties and nuances present in social systems that might be neglected or ignored when modelling social systems with lower order approaches. This might help address issues that can arise from lower order modelling approaches such as conveying or exaggerating biases existing in the data due to oversimplification. We give examples of different scenarios for each of the three main higher-order model streams: multi-way, multi-layer and temporal interactions.

Second, we also identify a potential of higher-order models to introduce new problems themselves that might have negative consequences on social systems, such as disadvantaging certain parts of a social system (groups, communities) or warping and changing the representation of social systems in undesirable ways. We identify two main challenges: data availability and model interpretation. With regard to the first point, additional degrees of freedom for a model creates additional possibility for bias and misrepresentation e.g. due to data availability.

Data resolution affects how well we can infer certain dimensions of a higher-order model in practice and how much the models generalises for certain groups. Additionally, there might be some dimensions that we particularly do not want to include in a model because they introduce bias. Secondly, the interpretation of the results of a higher-order model can be complicated by the higher-order model aspects, which have to be well motivated and backed up with theory. When constructing a model, we want to capture all aspects of a system which are relevant for a specific research question of interest. Additional model dimensions can possibly lead to more insights, but also to more misinterpretation of their meaning is not clear in a specific context.

In summary, higher-order models of social systems have the potential to help overcome limitations of existing lower order approaches, but also introduce new challenges which need to be addressed to avoid introducing unintentional harms that result from information captured by higher order approaches.

4.3 Applications of Higher-Order Models

Vincenzo Perri (Universität Zürich, CH), Gabriele DiBona (Queen Mary University of London, GB), Luca Gallo (University of Catania, IT), Christoph Gote (ETH Zürich, CH), Jürgen Hackl (University of Liverpool, GB), Desmond J. Higham (University of Edinburgh, GB), and Frank Schweitzer (ETH Zürich, CH)

This report summarises the discussion that has taken place during the breakout sessions of the working group focused on the topic of "applications of higher-order networks". The discussions covered a broad range of topics, which we report in what follows.

A practitioner's guide

In our examination we abstract from the details of an application to specific systems. We decided to draft a practitioner's guide such that practitioners, based on their domain knowledge, are equipped to determine if and when higher-order networks should be used and which question they can answer. We identify five questions to address before applying higher-order methods and outline possible answers.

Do I need higher-order models?

To use higher-order networks, we first need to understand if the question has higher-order characteristics. Higher-order patterns do not pertain to the network structure (i.e., which pairwise interactions occur) but emerge from more complicated relationships. We identify three dimensions that result in higher-order patterns:
1. sequential (causal) dependencies
2. group interactions
3. multiple node or edge types

The use of higher-order networks is beneficial only if the problem displays at least one of these characteristics.

Do my data allow the use of higher-order models?

Even if the problem has higher-order characteristics, the data at our disposal might not allow for the use of higher-order methods. The complexity of higher-order interactions leads to models with a higher number of parameters compared to standard methods. Such complexity raises constraints relative to data quantity, as more parameters have to be estimated, and quality, as inconsistencies in the data might lead to cascading effects. Model selection needs to be used to select the optimal model given the available amount of data. Possible cascading effects are still an open issue. We will re-encounter them in the last question.

How do I get the model?

Answering positively to the previous two questions establishes the conditions for a fruitful application of higher-order networks.

Now, the practitioner has to choose the more suitable higher-order formalism from the available ones. The choice might not be straightforward, and it will depend on both the question and the available data. Helping a practitioner answer this question requires the community to provide tools (tutorials, software, etc.) that allow practitioners to understand the use-cases of different higher-order formalisms and eventually compare or mix them.

How to analyse higher-order models

Analysing higher-order networks might not be as straightforward as for standard networks. The first key step is the choice of the data structure to use, which may affect both the flexibility and efficiency of a computational algorithm. Then, performing the analysis requires understanding the meaning of the interactions between the elements of the higher-order network. Additionally, we need to decide whether to use the higher-order information to predict higher-order or standard structures. Finally, one should also consider the problem of data quality underlined above (question 2). While steps forward have been made in considering the impact of incomplete or noisy data on standard networks, this topic has received little attention for higher-order networks.

How to interpret the results

The interpretation of higher-order methods' results is often not as straightforward as that of standard network methods. One challenge is to express patterns identified in the higher-order representation in terms of standard nodes. Depending on how we project from higher to lower order, we will retain more or different types of information. Additionally, the question of interpretation can not be separated from the other topics discussed above. Issues like data quality and quantity have to be considered when interpreting the results in order to be able to separate the model's sensitivity to changes in the data from the system's stability.

Conclusion

In our discussions, we identified the most challenging problems to be the ones regarding the interpretation of the results of a higher-order model in terms of a real-world problem. We suggest thinking of the minimal model that can explain the phenomenon, also considering simple networks. Even if this process has been undertaken when choosing the model, we should continue to question the choice of the model when interpreting the results.

4.4 Learning and Model Selection in Higher-Order Networks

Martin Rosvall (University of Umeå, SE), Rebekka Burkholz (Harvard School of Public Health – Boston, US), Timothy LaRock (Northeastern University – Boston, US), Vito Latora (Queen Mary University of London, GB), Kang-Ju Lee (Seoul National University, KR), Giovanni Petri (ISI Foundation – Torino, IT), Luka Petrovic (Universität Zürich, CH), Michael Schaub (RWTH Aachen, DE), Alice Schwarze (University of Washington – Seattle, US), and Michele Starnini (ISI Foundation, IT)

In the working group *Learning and model selection in higher-order networks*, we discussed the different roles models can have and how the specific task must decide the model choice. Our discussions focused on two model perspectives.

The first perspective concerned *the scope of what we aim to learn.* Are we interested in relational data – the structure of a network? Or are we interested in covariate data constrained by relational data – signals, metadata, and dynamical data? Or both?

Relational data describe interactions between at least two entities, the topology of a possibly higher-order network system. In a standard, dyadic, static setup, edges describe the system's topology, which we aim to model. For example, we observe a network sample and infer a probability distribution over graphs using a stochastic block model as a statistical model of the network.

In higher-order models, the types of relational data we recognise are significantly larger. Instead of considering only a single type of dyadic static relation between two nodes, we consider typed interactions (signed, multiplex, multi-layer), temporal interactions with a path-dependency, polyadic interactions, or any combination of them.

In covariate data, each data point is associated with a node, an edge, or a higher-order simplex. One can obtain covariate data from measurements at a given time or a sequence of time points. One can also obtain covariate data as a function of time from theoretical dynamical systems or computational simulations. Such state variables can capture a dynamical process that takes place on the interacting system. Examples include time series of electric signals at the different cortical areas of the brain measured through EEG brain imaging, infected individuals in the spreading of a disease across a population, or traffic flows in street networks.

The other perspective concerned the *objective of learning.* Are we interested in prediction or classification from incomplete data, or do we seek to discern important mechanisms of the system under study, or both? These objectives come with different trade-offs.

For predicting future interactions or classifying nodes based on incomplete data, we must balance model and data complexity to find a model that accurately describes the available data. When detecting the optimal order of a multi-order network model for pathway data, for example, we need to balance the increase in the likelihood of a more complex model with the increase in the complexity of the model. The richer the data we can collect, the more flexible models we can try to fit. With access to polyadic data, we may successfully fit a statistical hypergraph model. In contrast, we may need more data to infer the polyadic relations from dyadic relations. Similarly, rich covariate data may allow for a more detailed model.

To identify a system's important mechanisms, we should decide what assumptions to use for modelling relational and covariate data. We discussed a scenario in which we study the spread of an epidemic by observing time-stamped interactions between people, potentially augmented with information about the state of the nodes. We could consider these data in at least two ways. First, we may think about them as coming from an epidemic process that spreads on a temporal topology: the process runs continuously on top of each node, but the relations between people are only active for some time. Second, we may interpret our data as events generated by a point process on a latent but largely static interaction topology. In this example, our mechanism of interest and the respective model's ability to describe the data, should guide us in choosing a model. We may prefer a simple model with system-specific assumptions over a flexible model with a potentially better fit to the data than the simple model because a high model complexity obscures the mechanisms that we seek to identify.

We concluded that generalising statistical principles developed for networks to higher-order network models seems promising for trade-offs of model flexibility. By contrast, trade-offs of higher-order models that we develop to gain mechanistic insights are under-explored and require new computational and mathematical methods.

4.5 Benchmark Data and Evaluation Practices

*Ingo Scholtes (Julius-Maximilians-Universität Würzburg, DE & Universität Zürich, CH),
Stephan Günnemann (Technical University of Munich, DE), Andreas Hotho (Julius-Maximilians-
Universität Würzburg, DE), and Jelena Smiljanic (University of Umeå, SE)*

An important open issue that has been raised by several participants during their introductory statements is a lack of commonly used benchmark data and generally accepted practices to evaluate higher-order graph models. Mirroring the diversity of the higher-order graph community, this is a multi-faceted problem. The working group has identified three related challenges which are presented below. We used them to derive four opportunities for the higher-order graph community to improve their evaluation practices and – in turn – increase their impact on applications of network and data science.

4.5.1 Challenges for the Evaluation of Higher-Order Graph Models

Comparing Higher- vs. First-Order Models

An important first challenge is due to the common need to show what we gain by using higher-order graph models, as opposed to methods that are based on first-order graphs. To this end, researchers typically evaluate their models based on a variety of data mining, prediction, and modelling tasks. The choice of those tasks, as well as the choice of data set in which those tasks are addressed, is often informed by specific assumptions of the higher-order models that are being evaluated. This introduces potential issues for the *external validity*, i.e. it is not clear to what extent the obtained results generalise to other settings or data with higher-order characteristics that do *not* match the assumptions of a given modelling framework.

Comparison of Different Higher-Order Models

While the challenge above applies to each "paradigm" of higher-order graph models individually, a second challenge arises due to the growing number of different modelling paradigms that address the same higher-order characteristic of data or systems, e.g., the use of hypergraphs vs. simplicial complexes to model systems with polyadic interactions. To facilitate a fair comparison between such different modelling approaches, the community should establish standard benchmark data sets that exhibit higher-order characteristics, along with a set of clearly defined tasks and evaluation metrics that do not favour one or the other modelling paradigm. This would not only help practitioners to decide which modelling paradigm to choose for a specific system. It is also likely to improve our understanding of the advantages and disadvantages of different paradigms and the sometimes implicit assumptions they are based on.

Comparison of Models for Different Higher-Order Characteristics

As highlighted in the panel discussion, there is no single, *correct* type of higher-order graph model that could be used to model all networked systems. Instead, we are commonly confronted with systems that exhibit multiple higher-order characteristics at once such as, e.g., networks with temporally ordered, multi-typed, and polyadic interactions. A third important challenge for the community is thus to understand what we lose or gain by

using models that capture only one of those characteristics. Given a modelling task in a system with temporally ordered polyadic interactions, is it preferable to use a hypergraph model that ignores the temporal ordering of interactions, or is it preferable to use a model that captures causal path while ignoring the fact that interactions are non-dyadic? To answer such questions, the community needs benchmark data and problems that support a comparison of higher-order graph models that address *different* higher-order characteristics in complex systems. We further need model-independent prediction or modelling tasks like, e.g., the prediction of interactions, node- or graph-level classification tasks, or forecasting the evolution of dynamical processes, that could be used to compare the performance of different higher-order graph models.

4.5.2   Opportunities to Improve Evaluation Practices

Based on the challenges outlined above, the working group has identified three opportunities for the higher-order graph modelling community, which we outline below.

Opportunity 1: Higher-Order Graph Benchmarks

A first opportunity is to establish benchmarks that can be addressed by different types of higher-order graph models, and which should be based on the following ingredients:

- data sets on networked systems with a given higher-order pattern (polyadic, multi-typed, temporal interactions, etc.)
- measure for model performance based on a given prediction or modelling task
- a baseline against which we compare model performance. Depending on the problem, this baseline can either be state-of-the-art techniques or, if we want to reason about the benefit of higher-order models, lower- or first-order versions of a given model.

We note that the following online repositories for network data contain data sets that may have the necessary characteristics for such benchmark data:

- **SNAP** `https://snap.stanford.edu/data/` (temporal, multi-layer, polyadic)
- **netzschleuder** `https://networks.skewed.de/` (temporal, multi-layer, polyadic)
- **Konect** `http://konect.cc` (temporal, multi-layer)
- **Sociopatterns** `http://www.sociopatterns.org` (polyadic, temporal)

Referring to the first challenge, a common goal in the study of higher-order graph models is to assess the advantage over techniques based on first-order graphs. The question which first-order graph model should be chosen to facilitate a fair comparison is non-trivial and has – in some cases – been addressed in an unsatisfactory fashion. As an example, consider a comparison of a model with weighted higher-order interactions with an unweighted first-order graph model. The results of such an experiment do not tell us a lot about the impact of higher-order interactions, since it mixes the effect of a projection to first-order interactions with the effect of reducing a weighted to an unweighted graph model.

How can we define baselines that enable a fair comparison? One possible approach is to apply higher-order graph models to a version of a data set, where the higher-order dependencies have been selectively removed. E.g. for memory networks or De Bruijn graph models of paths in temporally ordered interactions, we can use data where time stamps have been randomly reshuffled, which removes any temporal correlations in the ordering while preserving information on the temporal distribution and the frequency of interactions. Similar randomisation approaches that maintain first-order characteristics but destroy higher-order patterns may be possible for data on polyadic interactions or multi-typed relations.

Opportunity 2: Using Higher-order Models to improve on Standard Graph Mining Problems

We can evaluate higher-order graph models in standard graph mining problems. This allows us to compare higher-order models against state-of-the-art algorithms as well as to different higher-order graph modelling frameworks with each other. Examples that can be potentially addressed based on different types of modelling paradigms include:

- Node ranking, where the ranking is based on different higher-order generalisations of centrality measures
- Node classification, where classes are assigned to nodes in a way that incorporates higher-order patterns
- Link prediction, where dyadic interactions are predicted based on models incorporating higher-order characteristics of the data (e.g. time or multiple types)
- Graph clustering, where clusterings in higher-order graph spaces are projected to clusters in a first-order graph
- Vector-space embedding, where vector representations of nodes or links are derived from a higher-order generalisation of similarity/dissimilarity rankings.

Recently, an extensive evaluation platform for graph mining problems has been proposed in [1]. It would be a worthwhile effort to consider whether a similar set of problems and evaluation practices, as well as convenient solutions for standardised data splitting, sampling and shuffling, could be combined with some of the data sets above to establish a higher-order graph benchmark that is accepted by the community.

Opportunity 3: Defining Novel Benchmark Problems involving Higher-Order Patterns

Apart from evaluating higher-order graph models in terms of standard graph mining and learning tasks, an interesting prospect for the definition of novel evaluation practices is that some of those problems can be naturally and meaningfully translated to the higher-order primitives used by different modelling frameworks. Examples include:

- multi-layer link prediction, where we predict links given a layer, a layer given a link, or both the layer and the link
- hyperedge or k-simplex prediction, which can be easily defined for co-occurrence or co-authorship data
- hyperedge clustering, which can be used to identify, e.g. groups of similar collaboration patterns
- path ranking, where rather than identifying nodes we identify node sequences or sets that are most important, e.g, for spreading patterns or information propagation
- path clustering, where we identify sets of paths observed in a time series data set that are more similar to each other than to other paths
- path prediction or classification, which can be useful for applications in click stream data, information propagation, as well as end-to-end vs. next-element prediction in sequential data

Opportunity 4: Model Dimensionality and Data Sparsity

One of the key challenges that we face in the study of higher-order models is that we often increase the dimensionality of the model, i.e. we add degrees of freedom that – on the one hand – enable us to more accurately model systems but – on the other hand – potentially

(a) $w_2 = 0.5$ solution   (b) $w_2 = 1.5$ solution   (c) $w_2 = 2.5$ solution
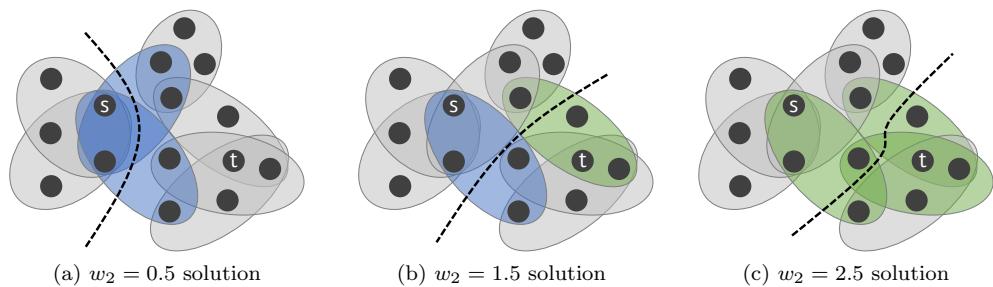
Figure 2 A figure from Veldt et al. [31] that illustrates how a hypergraph can be split into two pieces as a characteristic parameter changes in a simple scenario. Crucial to this drawing is showing which hyperedges are separated in the partition, which is simple in the convex-set drawing of a hypergraph used in the figure. It would be challenging to illustrate this figure with a node-and-edge drawing – even of the bipartite network representation of a hypergraph. This motivates our questions of how to visualise these hypergraphs.

increase computational complexity and pose challenges for the generalisability, robustness, and data efficiency of our models. However, this challenge also introduces opportunities for a definition of evaluation practices that go beyond mere model accuracy. Exemplary aspects that should be incorporated in the evaluation of higher-order graph models include:

- **model robustness**: How robust is a higher-order model against the introduction of noise and how does the inclusion of higher-order primitives specifically change the robustness compared to first-order graphs?
- **model size**: How much memory does a model consume, how many degrees of freedom does it have and how does the model size depend on key system parameters like the number of nodes or the density of (higher-order) interactions?
- **data efficiency**: How much data do we need to reliably model higher-order patterns in a given data set?
- **scalability**: How much time do we need to learn a model or to make predictions and how does the computational complexity depend on the size of the data (in terms of number of observations) or the size of the system?

References

1   Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, Jure Leskovec: *Open Graph Benchmark: Datasets for Machine Learning on Graphs*, arXiv 2005.00687, https://arxiv.org/abs/2005.00687

4.6   Visualisation and Interpretability of Higher-Order Networks

Chester Tan (National University of Singapore, SG), Daniel Edler (University of Umeå, SE), Anton Eriksson (University of Umeå, SE), David F. Gleich (Purdue University – West Lafayette, US), and Lisi Qarkaxhija (Koper, SI)

Visualisation is essential to understanding and *interpreting* data – it facilitates recognising norms and outliers in non-human readable data by representing data in more accessible forms such as graphs. Recognising its importance and infancy in its use with higher-order networks,

we decided to form a working group to discuss exactly the Visualisation and Interpretability of Higher-Order Networks. In this working group we discussed the current state of the art and its limitations, and deduced some notable key opportunities for development.

4.6.1 Current literature

We first sought a common understanding of the state of the art by discussing known data visualisation tools and identifying if they support higher-order visualisations, and the type(s) of higher-order visualisations they facilitate.

Relevant Tools from Workshop Participants

Acknowledging while looking to leverage our biases, we first highlighted the following most relevant tools developed by participants in this Dagstuhl: **pathpy** [26] is a Python package which provides an automated framework to deduce the most likely Markov order for sequential data, and visualise such data as its most probable higher-order De Bruijn Graph. **Infomap Network Navigator** [12] is an interactive web application that generates a zoomable map for networks clustered with InfoMap. While it supports higher-order networks, it doesn't draw the raw network but instead the hierarchical modular network structure using existing tools for force directed layouts, augmented with new constraints, to support the higher order visualisation of state networks. **LocalGraphClustering** [15] is a Python (and Julia) package designed to identify local structures in networks and visualise how their local groups compress into low-rank representations, primarily to highlight differences in the way various algorithms *see or experience* network structures.

Desktop GUI Applications

We then discussed well-known Desktop GUI applications for visualising network data such as **Gephi** [2] and **Cytoscape** [28]. Neither requires any programming, and both support importing a network and associated metadata from various file types. Gephi is an open source cross-platform application that is able to visualise and analyse large networks, while Cytoscape is an open source software platform for visualising complex networks and integrating these with any type of attribute data, focusing on bioinformatics data. Notably, neither have dedicated support for higher-order network representations.

General Graph Toolboxes

While these relatively well-known GUI applications did not support visualising higher-order data, we noted that there are many software libraries designed to draw networks, or support programs that work with networks and produce visualisations, though they all require some programming familiarity: **MuxViz** [10] is an R package for the analysis and visualisation of interconnected multi-layer networks. **NetworkX** [17] is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. It has many functions to help draw networks and a number of functions to compute force directed-like layouts. **D3.js** [7] is a JavaScript library that transforms data to interactive visualisations in the browser. It includes force-directed graph layout algorithms. **iGraph** [9] is a network analysis and visualisation software written in C++ with bindings to R and Python. This includes tools to compute network layouts (coordinates) for each node from a variety of methods. These scale to large graphs with millions of nodes in reasonable time-frames (hours). **GraphViz** [13] is a free and open source graph visualisation program in DOT

language scripts. Like **iGraph**, it supports a variety of layout algorithms. **Graph-tool** [22] is a Python package which is used to produce useful visualisations, statistical analysis and manipulation of networks. It known for its performance since its main algorithms and data structures are written in C++.

4.6.2   The Many Forms of Higher-Order Networks and their Visualizations

From these known tools and frameworks, we identified some common models of higher-order network models and visualisations which elucidated even more higher-order visualisation methods.

Hypergraphs

One existing means of visualising higher-order data is through sets of nodes. Berge used this technique in his book on hypergraphs [5]. A downside to this visualisation approach is that an inaccurate drawing may result in nodes appearing to belong to hyperedges that they do not contain.

Bipartite Networks

Another means of visualizing higher-order data is through a bipartite network. This visualisation corresponds to using the incidence matrix of a graph as the adjacency matrix of a bipartite graph, and is further related to what is called a *star* expansion of a hypergraph structure.

Space Embedded Networks

Some network embeddings [23, 16] produce a set of coordinates for a network by minimising an energy function over small sets sampled from the network. These have been extended to higher-order data as well [11, 29], where the output is typically set of coordinate in a high-dimensional space, with 1 coordinate per node. These can be subsequently processed with tSNE [30] or UMAP [20] or alternative dimension reduction techniques, though these dimension reduction techniques do have their biases and compromises [32]. Some methods to embed networks in spacetime [8] have also been explored.

4.6.3   Key Opportunities and Takeaways

From these discussions, we deduced that network visualisation serves two primary roles: **(1) to elucidate results in studied network data**, as a static figure or a short movie – a form of network data visualisation that has appeared on the cover image of many highly regarded interdisciplinary journals – and **(2) to facilitate the discovery of features in data**, where tools often have interactive graphical components that make it easy to manipulate diverse data.

In an ideal scenario, higher-order graph models and data can simultaneously be visualised and interpreted by existing tools and also pose new challenges and opportunities beyond them. This apparent contradiction follows because, while there are many ways to translate higher-order data into network-like representations, each interpretation has its biases which obscures some properties of the higher order data over others. To illustrate this point, see Figure 2. The contents of the figure require its higher-order data to be expressed in a new or different way. Similar figures, with new visualisation strategies, arise in many papers introducing higher-order topics [21, 25, 3]. Many of these figures use non-standard visual representations of higher-order data that are difficult or impossible to replicate with standard tools.

Dimensions of Higher-Order Network Visualisations

These two primary roles make just one of the many notable *dimensions* we found network visualisations to have, including: **Interactive vs Static** For data exploration and online showcases, interactive visualisations are preferred over static visualisations, while paper figures are typically static, and have stricter requirements for legibility. **Multi-layer vs Single-layer** With usually an ordinary 2D layout of the network in each layer, a multi-layer network visualisation can either be in two or three visual dimensions with each layer drawn as cards side by side. **10 vs 1000 vs millions of nodes** It is easy to visualise entire small networks, but significantly more difficult to interpret visualisations of large networks. **Annotated vs Non-annotated** While annotations in our visualisations aid comprehension of the data and provide additional information about topics that aren't evident to the human eye, having a lot of annotations make visualisations undesirably noisy and cluttered. **Raw Network vs Clusters** In most interactive maps, the level of detail shown depends on the zoom level. A hierarchical clustering algorithm can similarly help us navigate a very large network, overcoming potential graphical or computational limitations. **Node-edge Plots vs Feature Embeddings** A standard way of drawing a network is to plot both nodes and edges, and for that to look nice for a broad range of networks, it typically requires a force-directed layout algorithm, or similar, that minimises edge crossings. On the other hand, if we want to highlight some features of our data, we may, for example, embed nodes in a high-dimensional space of node features, and employ dimensionality reduction techniques to layout nodes in two or three visual dimensions. **Instantaneous vs Evolving** visualisations. **GUI vs Programming Interfaces** Many GUI apps contain various tools that are often more user friendly and less time consuming than their programming counterparts, which, instead, often offer greater customisability and reproducibility.

We note that the seemingly opposing poles in each of these dimensions listed above are not necessarily mutually exclusive – e.g. a GUI app could have a programming interface, and that the current state of the art supports only a small subspace of these dimensions. The following example may further elucidate how the current lack of tools and techniques hampers research: in Gleich's work this past summer studying a set of emails surrounding the US government's response to the COVID pandemic, we sought to use time-varying hypergraphs to represent the email information [4, 6]. Hypergraphs were key to the representation as often a single email will bridge a number of different organisational entities in the strongly hierarchical government agencies. Yet, without current tools supporting them adequately, the team had to implement rudimentary ideas as surrogates for investigations they wished to conduct. This make it significantly more difficult to interpret the data using the growing set of higher-order data tools produced by the community.

A List of Some of the Many Forms of Higher-Order Network Visualisations

As another key takeaway we briefly list some of the common forms of higher-order network visualisations we discussed: **hypergraphs**, **simplicial complexes**, **bipartite networks**, **multi-layer networks**, **multiplex networks**, **higher-order space embedded networks**. This list highlights, among other things, the variety in visualisation methods and language of, which made discussions especially challenging and interesting, and an apparent recurring and arguably most interesting theme throughout this Dagstuhl.

4.6.4 Key Opportunities

Finally, we concluded that the following are two key opportunities in the field for further study and development: **(1) Interactive tools for higher-order data** There is ongoing work on tools to work with higher-order representations of processes on network data by the workshop participants in the Infomap Network Navigator (see the paragraph below). Additional tools have identified similar weaknesses and aspects. See, for instance, open issues on the `Gephi` and `graphviz` software to support hypergraph drawings. **(2) Revisiting fundamental ideas** Many existing network visualisations involve a variety of studies closely related to many applied algorithms. For instance, spectral network drawing was originally proposed as an energy minimisation technique [18] that predated Fiedler's work on Laplacian eigenvectors [14]. There are now higher-order generalisations of many similar ideas [19] (and references therein). Many successful node placement techniques for graph visualisation are based on force simulations (e.g. Force-Atlas, etc.) Higher-order data present novel opportunities to evolve this research. For instance, recent work on force directed placement [1]. Related work includes efficient molecular dynamics simulations [24], which suggest novel types of possible forces for higher-order data.

References

1 Naheed Anjum Arafat and Stéphane Bressan. Hypergraph drawing by force-directed placement. In Djamal Benslimane, Ernesto Damiani, William I. Grosky, Abdelkader Hameurlain, Amit Sheth, and Roland R. Wagner, editors, *Database and Expert Systems Applications*, pages 387–394, Cham, 2017. Springer International Publishing.

2 Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*, 2009.

3 Austin Benson, David F. Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

4 Austin Benson, Nate Veldt, and David F. Gleich. fauci-email: a json digest of anthony fauci's released emails. *arXiv*, cs.SI:2108.01239, 2021. Code and data available from `https://github.com/nveldt/fauci-email`.

5 Claude Berge. *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier, 1984.

6 Natalie Bettendorf and Jason Leopold. Anthony fauci's emails reveal the pressure that fell on one man. BuzzFeed News, `https://www.buzzfeednews.com/article/nataliebettendorf/fauci-emails-covid-response`, June 2021.

7 Mike Bostock. D3.js – data-driven documents, 2012.

8 James R. Clough and Tim S. Evans. Embedding graphs in lorentzian spacetime. *PLOS ONE*, 12(11):e0187301, Nov 2017.

9 Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

10 Manlio De Domenico, Mason A Porter, and Alex Arenas. Muxviz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, 3(2):159–176, 2015.

11 Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. Metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 135–144, New York, NY, USA, 2017. Association for Computing Machinery.

12 D. Edler, A. Eriksson, and M. Rosvall. *The Infomap Software Package*, 2021.

13 John Ellson, Emden Gansner, Lefteris Koutsofios, Stephen North, Gordon Woodhull, Short Description, and Lucent Technologies. Graphviz — open source graph drawing tools. In *Lecture Notes in Computer Science*, pages 483–484. Springer-Verlag, 2001.

14 Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(98):298–305, 1973.

15 Kimon Fountoulakis, David F. Gleich, and Michael W. Mahoney. A short introduction to local graph clustering methods and software. In *Book of Abstracts for 7th International Conference on Complex Networks and Their Applications*, pages 56–59, 2018.

16 Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

17 Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

18 Kenneth M. Hall. An r-dimensional quadratic placement algorithm. *Management Science*, 17(3):219–229, 1970.

19 Pan Li and Olgica Milenkovic. Submodular hypergraphs: p-laplacians, Cheeger inequalities and spectral clustering. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3014–3023, Stockholm Sweden, 10–15 Jul 2018. PMLR.

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

21 Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

22 Tiago P. Peixoto. The graph-tool python library. *figshare*, 2014.

23 Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 701–710, New York, NY, USA, 2014. ACM.

24 Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

25 Martin Rosvall, Alcides V. Esquivel, Andrea Lancichinetti, Jevin D. West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5(4630), 2014.

26 Jürgen Hackl, Ingo Scholtes, Luka V Petrović, Vincenzo Perri, Luca Verginer, Christoph Gote. Analysis and visualisation of time series data on networks with pathpy In *Proceedings of the 11th Temporal Web Analytics Workshop (TempWeb 2021)* held in conjunction with The Web Conference 2021, Ljubljana, Slovenia, April 2021

27 Ingo Scholtes. When is a network a network? multi-order graphical model selection in pathways and temporal networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 1037–1046, New York, NY, USA, 2017. Association for Computing Machinery.

28 Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

29 Justin Sybrandt, Ruslan Shaydulin, and Ilya Safro. Hypergraph partitioning with embeddings. *IEEE Transactions on Knowledge and Data Engineering*, page 1–1, 2020.

30 Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

31 Nate Veldt, Austin R. Benson, and Jon Kleinberg. Hypergraph cuts with general splitting functions, 2020.

32 Martin Wattenberg, Fernanda Viégas, and Ian Johnson. How to use t-sne effectively. *Distill*, 1(10):e2, 2016.

5   Panel discussions

5.1   What are Higher-Order Graph Models?

*Ingo Scholtes (Julius-Maximilians-Universität Würzburg, DE & Universität Zürich, CH)*

Following the sessions with brief introductory statements and a first meeting of the working groups, participants spontaneously engaged in an open-end evening discussion on what they identified as an important open issue in the community: we lack a commonly agreed-upon definition of higher-order graphs and networks. In particular, different researchers use the term "higher-order" to refer to different characteristics of either networked systems, network models, or data.

The discussion revealed that the seminar participants agree that, as a community, we must more clearly distinguish between (i) complex networked systems that consist of many interacting elements, (ii) high-dimensional data that capture those interactions between system elements, and (iii) graph or network *models* of those systems. Commonly used graph models with a single type of dyadic, static links as the simplest possible – but neither the only nor necessarily optimal – graphical representation of data on element-element interactions that can be used to generate insights into complex systems. The analysis of such *first-order* graph model can nevertheless be reasonable if (a) we know that the system exclusively features a single type of interaction between pairs of elements, (b) we only have access to relational data capturing pair-wise interactions even though we know interactions in the system are more complex, or (c) we seek to understand which of the system's characteristics can already be explained by first-order interaction.

A clear distinction whether the term *network* refers to the system to be modelled, the structure of the available data, or the mathematical model used to analyse the data is often missing. This complicates the rigorous definition of higher-order graphs and networks and has – at times – fostered misunderstandings between different communities regarding whether a given type of model should be considered *higher-order* or not. Summarising the results of the panel discussion, in the following we take two perspectives that focus on the characteristics of the *model* and the *data* on the system to be modelled.

Model perspective

A first approach to define *higher-order graph models* considers the mathematical representation used to study the topology, i.e. who can influence whom and how, of a complex system. First-order graph models assume that the topology (and the resulting behaviour) of a complex system can be reduced to a set of dyadic edges, which can be mathematically represented in terms of adjacency, transition, or Laplacian matrices with $\mathcal{O}(n^2)$ entries, where n is the number of elements or vertices in the system. Despite major differences in terms of modelling assumptions, a common feature of all higher-order graph models – be it hypergraphs, simplicial complexes, memory networks, or high-dimensional De Bruijn graphs – is that they require mathematical notations with higher dimensionality than common matrix representations. This characteristic of different higher-order models translates to similar ideas, e.g. the use of tensors and flattened representations of high-dimensional linear operators, as well as common challenges, e.g. computational challenges and dimensionality issues in higher-order graph learning methods.
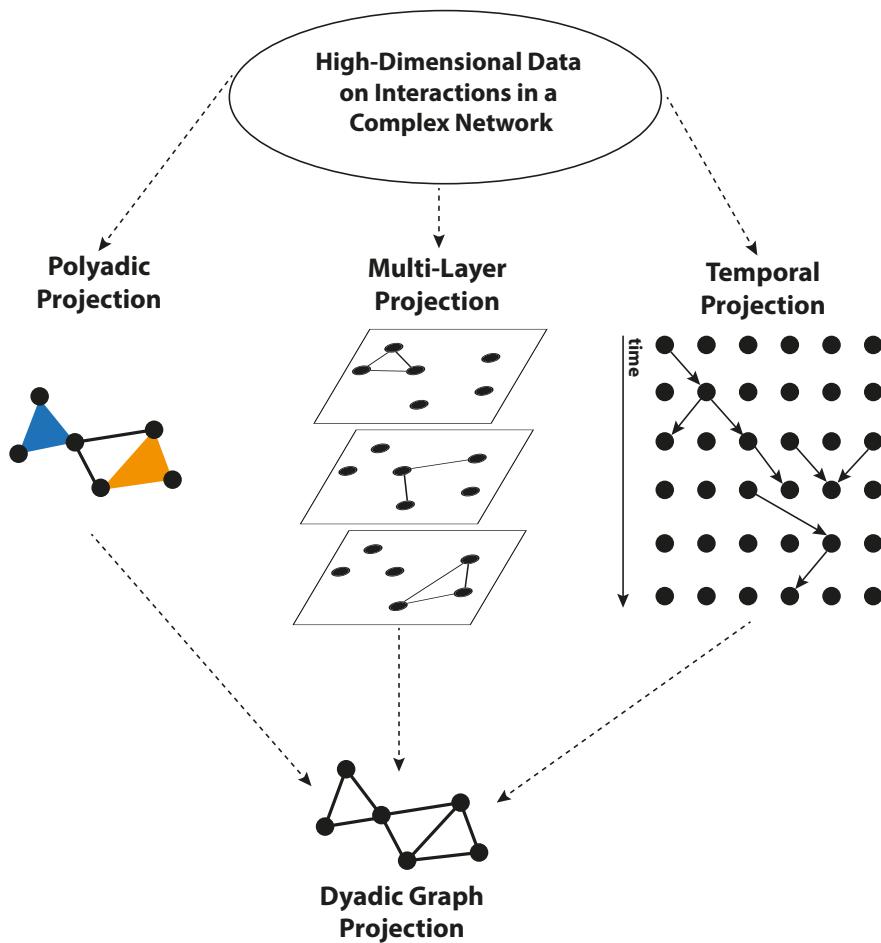
Figure 3 Different higher-order graph models can be viewed as different projections of high-dimensional data on interactions in complex systems along different dimensions, where single-typed dyadic graphs can be viewed as a maximally simple projection of those different higher-order models.

While it may seem intuitive that the use of higher-order graph models either requires networked systems with non-dyadic interactions or data with higher-order characteristics, this is not necessarily the case. Several graph learning techniques make use of higher-order primitives, which – however – are not used to model higher-order structure in the underlying systems or data. A prominent example is node2vec[16], which can be viewed as a random walk in a second-order graph model, but which is usually applied to data on graphs with simple dyadic interactions. Here, the higher-order model is rather used to encode non-local features of the graph topology into a model for a dynamical process on the graph. The question whether we consider such a model as *higher-order* graph model or not highlights that we may need to look beyond the characteristics of the data.

Data perspective

An interesting point raised during the discussion was that it may actually be easier to reach consensus on a definition of higher-order characteristics in data, rather than higher-order characteristics in *graph models*. From the perspective of "first-order" graph theory or network

science, higher-order characteristics in data can be defined as any information that goes beyond the specification of dyadic edges, i.e. any data that gives rise to more than a subset of the Cartesian product of vertices. Examples for such data with higher-order characteristics include but are not limited to:

- multiple sets of edges capturing interactions with different properties (such as multi-typed or signed interactions that invalidate a simple transitive treatment of edges)
- data capturing polyadic interactions, e.g. tuples or sets with a cardinality higher than two
- ordered or time-stamped sequences of dyadic or polyadic interactions

In network science, such higher-order characteristics in data is often reduced to dyads because we want to apply standard graph algorithms or network analysis techniques. In contrast, as higher-order graph models we can define any model that seeks to more faithfully represent (one or more) of the higher-order characteristics present in data on complex systems that influence how nodes can directly or indirectly influence each other. Notably, different types of higher-order graph models can destroy different higher-order characteristics in the data: A hypergraph model of time-stamped polyadic interactions destroys higher-order patterns that are due to the timing and ordering of interactions, while higher-order De Bruijn graph models for temporally-ordered dyadic links destroy patterns that are due to the polyadic nature of the interactions. The combination or unification of different higher-order modelling frameworks to capture multiple higher-order characteristics of data is an important open challenge that must be addressed by the community.

We finally noted that the large popularity of graph models with dyadic links or edges often leads to the unfortunate development that the data collection and engineering process is informed by the features of simple graph models rather than the modelling process being informed by the higher-order characteristics of the system to be modelled. As an example, data on co-authorship networks are often provided in the form of dyadic relationships between authors even though the underlying interactions are fundamentally non-dyadic. Similarly, data is often tailored to the application of time-slice snapshot network models, discarding information that would be important to infer higher-order patterns in the temporal ordering of interactions. This leads to what could be called a "data bottleneck" that hinders the application of higher-order graph structures to model the higher-order characteristics present in many real complex systems.

Participants

- Unai Alvarez-Rodriguez
Universität Zürich, CH
- Luca Gallo
University of Catania, IT
- David F. Gleich
Purdue University –
West Lafayette, US
- Christoph Gote
ETH Zürich, CH
- Jürgen Hackl
University of Liverpool, GB
- Andreas Hotho
Universität Würzburg, DE

- Kang-Ju Lee
Seoul National University, KR
- Leonie Neuhäuser
RWTH Aachen, DE
- Vincenzo Perri
Universität Zürich, CH
- Giovanni Petri
ISI Foundation – Torino, IT
- Luka Petrovic
Universität Zürich, CH
- Martin Rosvall
University of Umeå, SE

- Michael Schaub
RWTH Aachen, DE
- Maximilian Schich
Tallinn University, EE
- Ingo Scholtes
Universität Würzburg, DE &
Universität Zürich, CH
- Frank Schweitzer
ETH Zürich, CH
- Markus Strohmaier
RWTH Aachen, DE

Remote Participants

- Federico Battiston
Central European University –
Vienna, AT
- Ginestra Bianconi
Queen Mary University of
London, GB
- Rebekka Burkholz
Harvard School of Public Health –
Boston, US
- Giulia Cencetti
Bruno Kessler Foundation –
Trento, IT
- Gabriele DiBona
Queen Mary University of
London, GB

- Daniel Edler
University of Umeå, SE
- Tina Eliassi-Rad
Northeastern University –
Boston, US
- Anton Eriksson
University of Umeå, SE
- Stephan Günnemann
TU München, DE
- Heather Harrington
University of Oxford, GB
- Desmond J. Higham
University of Edinburgh, GB

- Fariba Karimi
Complexity Science Hub –
Wien, AT
- Danai Koutra
University of Michigan –
Ann Arbor, US
- Renaud Lambiotte
University of Oxford, GB
- Timothy LaRock
Northeastern University –
Boston, US
- Vito Latora
Queen Mary University of
London, GB

Yamir Moreno
University of Zaragoza, ES

Natasa Przulj
Barcelona Supercomputing
Center, ES

Lisi Qarkaxhija
Koper, SI

Alice Schwarze
University of Washington –
Seattle, US

Jelena Smiljanic
University of Umeå, SE

Michele Starnini
ISI Foundation, IT

Chester Tan
National University of
Singapore, SG

Leo Torres
Northeastern University –
Boston, US

Anatol Wegner
University College London, GB