Beyond Graph Mining: Higher-Order Data Analytics for Temporal Network Data

Renaud Lambiotte
University of Oxford
United Kingdom
renaud.lambriotte@maths.ox.ac.uk

Martin Rosvall
Umeå University
Sweden
martin.rosvall@umu.se

Michael Schaub
Massachusetts Institute of Technology
Cambridge, MA, USA
mschaub@mit.edu

Ingo Scholtes*
ETH Zürich
Switzerland
ischoltes@ethz.ch

Jian Xu
Citadel LLC
Chicago, USA
Jian.Xu2@citadel.com

ABSTRACT

Network-based data mining techniques such as graph mining, (social) network analysis, link prediction and graph clustering form an important foundation for data science applications in computer science, computational social science, and the life sciences. They help to detect patterns in large data sets that capture *dyadic relations* between pairs of genes, species, humans, or documents and they have improved our understanding of complex networks.

While the potential of analysing *graph* or *network representations* of relational data is undisputed, we increasingly have access to data on networks that contain more than just dyadic relations. Consider, e.g., data on user click streams in the Web, time-stamped social networks, gene regulatory pathways, or time-stamped financial transactions. These are examples for *time-resolved* or *sequential* data that not only tell us *who* is related to *whom* but also *when* and *in which order* relations occur. Recent works have exposed that the timing and ordering of relations in such data can introduce higher-order, non-dyadic dependencies that are not captured by state-of-the-art graph representations [12, 16–18, 20, 26]. This oversimplification questions the validity of graph mining techniques in time series data and poses a threat for interdisciplinary applications of network analytics.

To address this challenge, researchers have developed advanced graph modelling and representation techniques based on higher- and variable-order Markov models, which enable us to model *non-Markovian* characteristics in time series data on networks. Introducing this exciting research field, the goal of this tutorial is to give an overview of cutting-edge higher-order data analytics techniques. Key takeaways for attendees will be (i) a solid understanding of higher-order network modelling and representation learning techniques, (ii) hands-on experience with state-of-the-art higher-order network analytics and visualisation packages, and (iii) a clear demonstration of the benefits of higher-order data analytics in real-world time series data on technical, social, and ecological systems.

*Corresponding author

1 TARGET AUDIENCE

The target audience of this tutorial are data scientists from academia, industry and government, who face the challenge of analysing time series data on complex networks. A number of works have generalized graph mining, network analysis, and graph summarisation techniques to the time dimension [1, 3, 6, 8, 9, 13, 21, 22]. While these works provide scalable methods that can be applied to large data sets, they have in common that they are based on a representation of temporal network data as sequences of *time-aggregated graph snapshots*. This makes them particularly suitable for *low-frequency data*, i.e. data in which links have coarse-grained time stamps that naturally lend themselves to a snapshot representation.

Going beyond the scope of such methods, this tutorial specifically targets data scientists that need to analyze *high-resolution temporal network data*, where links carry fine-grained, possibly unique, time stamps. Such data pose a fundamental challenge for state-of-the-art dynamic graph mining algorithms. Current methods discard information on the microscopic timing and ordering of links, which is, however, the foundation of so-called *time-respecting paths* [6], i.e. it is needed to answer the question *who* can influence *whom*. For a sequence of two links $A \to B$ and $B \to C$, a node A can only influence C via a (transitive) *time-respecting path* via B if $A \to B$ occurs *before* $B \to C$. Such a transitive path from A to C does not exist if the ordering of links is reversed. As shown by a number of recent works, the specific ordering and timing of events in realistic temporal network data can give rise to *non-Markovian* characteristics in the sequence of links [12, 16–18, 20, 26]. This results in complex, non-dyadic dependencies between nodes that are neglected by existing graph mining and network analysis techniques.

Addressing this important issue in the analysis of temporal network data, this tutorial addresses data science researchers and practitioners who (i) wish to adopt a network perspective to reveal higher-order dependencies between the elements of complex systems, (ii) extract higher-order features for machine learning tasks, or (iii) look for methods that go beyond the dyadic perspective taken by existing dynamic graph mining and social network analysis techniques.

Takeaways. Big data practitioners will learn about (i) the emerging challenges of modelling increasingly *complex* data (as opposed to *big* data) as networks, and (ii) state-of-the-art tools for analysing and visualising higher-order dependencies in relational data that are typically studied from a network perspective. Machine learning researchers will see how to construct optimal graphical representations that best preserve information on higher-order dependencies in the raw data. Government urban planners and policymakers will gain hands-on experience with visualisation and interactive exploration tools that enable anomaly detection in time-evolving traffic networks, as well as simulations of species invasion through global shipping networks. Students new to the interdisciplinary field of network science will learn through intuitive examples how cutting-edge network analytics techniques can be used to uncover rich information hidden in time series data on complex systems.

Impact. By introducing higher-order network analytic methods, the tutorial broadens the choice of network-analytic techniques towards methods that account for higher-order dependencies in complex systems. The methods introduced in this tutorial are of high relevance for a wide variety of attendees at SIGKDD. Bridging the *machine learning*, *network science*, *graph theory*, and *interdisciplinary physics* perspectives on pattern recognition in relational data, our tutorial fosters the interdisciplinary exchange between *data mining* and *network science* researchers with different backgrounds.

Prerequisites. Attendees should have a basic understanding of mathematical and statistical concepts equivalent to the high-school level. A basic understanding of graph-theoretic concepts is a plus; however, all theoretical methods and abstractions will be introduced along the way and visually illustrated in examples. For the hands-on `python` exercises, attendees should have basic programming skills (`python` or other scripting languages) at the level of an introductory programming class. A short tutorial on essential concepts needed in the hands-on sessions will be made available to attendees before the tutorial. A pre-configured `jupyter` notebook that simplifies the handling, analysis and visualisation of static and dynamic network data will be made available to attendees before the tutorial.

2 TUTORS' INFORMATION

Dr. Renaud Lambiotte
Affiliation: University of Oxford, United Kingdom
E-mail: renaud.lambriotte@maths.ox.ac.uk
Address: Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom
Phone: +44 1865 280608

Dr. Martin Rosvall
Affiliation: Umeå University, Sweden
E-mail: martin.rosvall@physics.umu.se
Address: Department of Physics, Umeå University, SE-901 87 Umeå, Sweden
Phone: +46 70 239 1973

Dr. Michael Schaub
Affiliation: Massachusetts Institute of Technology, USA
E-mail: mschaub@mit.edu
Address: Institute for Data, Systems, and Society, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Office: E18-425C, Cambridge, MA, 02139
Phone: +1 857 320 5174

Dr. Ingo Scholtes* *corresponding author*
Affiliation: ETH Zürich
E-mail: ischoltes@ethz.ch
Address: ETH Zurich, WEV G 213, Weinbergstrasse 56/58, 8092 Zurich, Switzerland
Phone: +41 44 632 81 98

Dr. Jian Xu
Affiliation: Citadel LLC
E-mail: Jian.Xu2@citadel.com
Address: 28S.319-D, 131 S Dearborn Ave, Chicago, IL, USA
Phone: +1 574 485 7282

3 TUTORS' BIO AND EXPERTISE

Dr. Renaud Lambiotte is Associate Professor at the Mathematical Institute of University of Oxford. He holds a Ph.D. in physics from the Université libre de Bruxelles, and was postdoc at ENS Lyon, Université de Liège, UCLouvain and Imperial College London. He further held a professorship in Mathematics at the University of Namur. His main research interests are the modelling and analysis of processes taking place on large networks, with a particular focus on social and brain networks. He also acts as an academic editor for PLoS One, European Physical Journal B and Advances in Complex Systems and is review editor for Frontiers in Physics. More info at http://xn.unamur.be/

Dr. Martin Rosvall is Associate Professor in Physics at Umeå University, Sweden, and is also the head of the highly interdisciplinary Integrated Science Lab. After a Ph.D. in theoretical physics at the Niels Bohr Institute in Copenhagen (2003–2006), he was Postdoc at University of Washington in Seattle (2006–2008), and Assistant Professor in Physics at Umeå University. He performs high-impact research in network science and complex systems research, and is a leading authority in network inference. He has pioneered mapping flow pathways and develop novel algorithms and tools that take advantage of today's data explosion for revealing important organizational patterns in complex systems. Martin is co-founder of the data analytics start-up Infobaleen. Infobaleen provides custom-fit machine-learning tools that empowers people by making complex data simple to use and understand. More info at http://www.tp.umu.se/~rosvall/

Dr. Michael Schaub is a Marie-Sklodowska-Curie Fellow at the Institute for Data, Systems, and Society at MIT and the Department of Engineering Science, University of Oxford. He obtained his PhD in applied mathematics and an MSc in biomedical engineering from Imperial College London, and his BSc in electrical engineering and information technology from ETH Zurich. Prior to MIT, he was also associated with the Université catholique de Louvain (Belgium), and the University of Namur (Belgium). He is broadly interested in interdisciplinary applications of applied mathematics in engineering, social and biological systems. His research interests include network theory, data science, and dynamical systems. More info at https://michaelschaub.github.io/

Dr. Ingo Scholtes is a Lecturer at ETH Zürich. He has a computer science and mathematics background and his research addresses foundations of algorithmic and statistical data analysis in complex networks. He is an active member of the network science and data mining communities. His works on higher-order network models were published in interdisciplinary outlets like *Physical Review Letters*, *Nature Communications*, and *The European Physical Journal*, as well as at *SIGKDD*. Together with R. Lambiotte and M. Rosvall, he founded a workshop series on higher-order models[1] that is co-located with the flagship network science conference NetSci. He is an editor of *EPJ Data Science*. Through his work at CERN's Large Hadron Collider he has hands-on experience in massive-scale data analysis. His applied data science works were published in areas like software engineering, social science, and scientometrics. At ETH Zürich he teaches a network analytics course that is part of the data science curriculum. He has extensive experience in teaching interdisciplinary audiences of Master and PhD students with backgrounds in computer science, mathematics, physics, life sciences, engineering, and social sciences. His lectures received the highest possible score in official teaching evaluations at ETH Zürich and Karlsruhe Institute of Technology. More info at http://www.ingoscholtes.net

Dr. Jian Xu is a data scientist at Citadel. He has a Ph.D. in computer science from University of Notre Dame, and his research focuses on representing big sequential data as networks, with interdisciplinary applications in social networks, financial markets, and biological networks. His work on higher-order networks was published in *Science Advances*, interdisciplinary applications in *SIGKDD*, and higher-order network visualization software packages in *IEEE PacificVis*. He also collaborated with the U.S. Army Research Lab on anomaly detection using higher-order networks. Through his work at Interdisciplinary Center for Network Science and Applications (iCeNSA) and Citadel's data strategies group, he has hands-on experience designing algorithms and developing software for 100TB-scale data. He has experience in teaching graduate-level machine learning course at University of Notre Dame, and has the Advanced Teaching Scholars certificate. More info at http://www.jianxu.net

[1]http://complexdata.businesscatalyst.com

4 TUTORIAL OUTLINE

We propose a **six-hour hands-on tutorial** that combines an introduction of fundamental higher-order modelling techniques with hands-on exercises in which participants apply these techniques to real-world data sets. In the following, we give a detailed outline of the material covered:

From Sequential Data to Networks (45 min)
In this introductory session, we provide examples of time-stamped and sequential relational data that exhibit non-Markovian characteristics that cannot be modelled from a graph perspective. We motivate the need for higher-order modelling techniques using real-world examples for wrong statements derived from existing network analysis and (temporal) graph mining techniques. We further provide concrete examples for data on technological, social, biological, and economic networks that necessitate higher-order data analytics techniques.

Case studies: (i) travel itineraries of passengers in the London Tube, (ii) dynamic social network of students at a university campus, (iii) trajectories of container ships in a worldwide logistics network, (iv) clickstreams of Web users in the Wikipedia graph, (v) time-stamped citation networks of scholarly articles.

Learning Optimal Higher-Order Representations (90 min)
A fundamental question in the modelling of higher-order dependencies in time series data is what representation is needed to analyse a given data set. In some data sets, a (standard) graph representation may be sufficient while other cases require higher-order representations. In this session, we show how this question can be answered. We introduce representation learning algorithms to (i) decide when standard graph representations of time series data are justified, and (ii) determine the *optimal* order of a higher-order representation for data sets that cannot be modelled as graphs. The first 45 minutes of this session will be used to introduce foundations of higher-order representation learning. We specifically introduce the model selection algorithm introduced in [18][2] and the order detection technique used in [26][3]. The second half of the session will be dedicated to hands-on applications using the implementation of these methods in the software packages `pathpy` and `BuildHON+`. We show how these packages can be used to learn optimal higher-order representations of the time series data sets presented in Session 1.

Case studies: (i) travel itineraries of passengers in the London Tube, (ii) dynamic social network of students at a university campus, (iii) trajectories of container ships in a worldwide logistics network, (iv) clickstreams of Web users in the Wikipedia graph, (v) time-stamped citation networks of scholarly articles.

[2]see high-level explanation at https://www.youtube.com/watch?v=prenHeyUbB4
[3]see examples at http://www.HigherOrderNetwork.com/algorithm

Higher-Order Temporal Graph Clustering (60 min)
In this session we introduce the problem of *community detection* or *graph clustering* in time series data on networks. Existing works in this area have either focused on (i) stable *topological clusters* in sequences of time slice graphs [4, 5, 10, 11, 24], or (ii) *temporal clusters* of node and link activities [23]. Complementing this view, we will show that cluster structures can emerge in time series data on networks where neither the network topology nor temporal activities exhibit clustering. The reason for this phenomenon is a complex interplay between the network topology and temporal correlations that affect path structures in temporal networks. This can lead to *temporal-topological clusters* that strongly influence dynamical processes [12, 16, 20]. Building on the representation learning algorithms presented in Session 2, we show how the inferred higher-order models can be analysed using the cross-platform community detection algorithm `InfoMap`, developed by one of the tutors [14, 15]. For this, we first provide a brief introduction to *flow compression*, the information-theoretic foundation of `InfoMap`. We show how this powerful approach to clustering can be generalised to higher-order models that capture both temporal and topological characteristics of time series data on networks [16]. We demonstrate this method in a large data set on time-stamped citations between scholarly articles. We show that the clusters inferred by this approach better capture ground truth scientific communities than standard clustering algorithms.

Case study: higher-order clustering of academic journals based on time-stamped citation relations between scholarly articles.

Higher-Order Node Ranking (60 min)
In this session, we show how higher-order models can be used to rank nodes based on time series data on networks. We first show that a ranking of nodes based on standard graph centralities such as eigenvector, betweenness, closeness, or PageRank centrality fails to capture the ground truth importance of nodes in time series data. Following the approach developed in [16, 19, 26], we demonstrate how these measures can be generalised to higher-order models that capture non-Markovian characteristics in time series data on networks. We provide a hands-on demonstration of this novel class of centrality measures in a realistic scenario that addresses the ranking of Wikipedia articles based on user click streams. We show that standard ranking algorithms like PageRank discard non-dyadic, higher-order dependencies between articles that are hidden in such sequential data. Using a data set on time-stamped social networks, we further show that a generalisation of path-based centrality measures to higher-order models provides new perspectives for social network analysis and computational social science.

Case studies: (i) ranking Wikipedia articles based on a freely available dataset capturing user click streams, and (ii) identifying central actors in a time-stamped social network.

Higher-Order Spectral Analysis of Time Series Data (60 min)
Extending our coverage of higher-order clustering and ranking algorithms, in this session we introduce advanced spectral techniques to analyse higher-order models of temporal network data. We show how well-known data mining techniques like Laplacian-based spectral partitioning or eigenvector-based centrality measures can be generalised to higher-order graph models. We introduce a statistical framework to model dynamical processes in temporal graphs and show that this framework provides interesting insights into the complex interplay between temporal and topological characteristics in time series data on networks [20]. We show that such a modelling approach helps us to (i) better understand flow processes in transportation networks, and (ii) temporal-topological cluster structures in time-stamped social networks.

Case studies: (i) time series data on passenger flows in a large transportation network, (ii) spectral clustering of actors in a time-stamped social network.

Visualising Higher-Order Models (45 min)
This session will introduce the analysis and visualisation of higher-order dependencies using the platform `HONVis` [25]. While higher-order modelling techniques provide a more accurate description of time series data by creating additional nodes to encode higher-order dependencies, they also add complexity for visualisation, exploration, and interpretation, especially when the higher-order network is time-evolving, exhibits high orders of dependency, or contains rich metadata (e.g. geographical or temporal information). We will demonstrate `HONVis`[4], which features coordinated multiple views that reveal higher-order network structures at three levels of detail (i.e., global, regional, and individual levels), the ability to simulate diffusion on higher-order networks, and to perform anomaly detection in time-evolving higher-order networks.

Case studies: (i) visualising the flow of invasive species through global shipping; (ii) anomaly detection using higher-order models of taxi itineraries in NYC.

5 PREVIOUS EDITIONS

This is the first tutorial of the form outlined above and we have specially tailored it to the interest of SIGKDD attendees. Some of the tutors have given lectures or hands-on tutorials on particular aspects that will also be covered in this tutorial:

- In 2016, Ingo Scholtes gave an invited lecture on the analysis of time series data on networks in the *School* program of NetSci X 2016. This lecture (90 minutes) is also part of his interdisciplinary course on network analytics at ETH Zürich. Different from the material covered in this tutorial, this lecture focuses on theoretical aspects of higher-order models.
- Martin Rosvall, Ingo Scholtes, and Jian Xu have given short (20 min) hands-on live demonstrations of the data

[4]see video demo at http://www.HigherOrderNetwork.com/visualization/

analytics packages `InfoMap`, `pathpy`, and `HONVis` at the 2017 International Conference on Network Science.

Through its unique interdisciplinary perspective, the collaboration of researchers from different institutions and scientific communities that have studied higher-order models from different angles, and the integration of fundamental and applied aspects in the modelling of higher-order dependencies in complex systems, our tutorial provides the – to the best of our knowledge – most comprehensive coverage so far of this important topic. We anticipate that it will raise the interest of a large audience of researchers and practitioners and that it will equip SIGKDD attendees with a new toolkit of cutting-edge data science techniques.

6 COVERED REFERENCES

The tutorial will introduce and practically demonstrate data science and machine learning techniques that have been developed in the following published works:

- D Edler, L Bohlin, M Rosvall: **Mapping Higher-Order Network Flows in Memory and Multilayer Networks with Infomap**. In *Algorithms*, 10(4), September 2017
- I 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 2017
- J Xu, T L Wickramarathne, N V Chawla: **Representing higher-order dependencies in networks**. In *Science Advances*, May 2016
- I Scholtes, N Wider, A Garas: **Higher-Order Aggregate Networks in the Analysis of Temporal Networks: Path structures and centralities**. In *European Physical Journal B*, 89:61, March 2016
- I Scholtes, N Wider, R Pfitzner, A Garas, C Tessone, F 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
- J Tao, J Xu, C Wang, N V Chawla. **HoNVis: Visualizing and exploring higher-order networks**. In *IEEE Pacific Visualization Symposium (PacificVis)*, April 2017.

7 REQUIRED SOFTWARE AND EQUIPMENT

We will bring a laptop and necessary adapters. We need a projector to present slides as well as a power socket. For the hands-on sessions, attendees should bring their laptop with a recent version of Linux, Windows or Mac OS X. Attendees should have access to a power socket. Before the tutorial, the attendees should set up the following software environment:

- a `python` 3.x environment, ideally a recent `Anaconda` distribution
- core `python` data science packages such as `numpy`, `scipy`, `matplotlib`, `numpy`, and `sklearn`
- optional but recommended: `jupyter` notebook
- a `python`-friendly IDE, e.g. `VS Code`, `pyCharm`, or `Rodeo`
- higher-order data analytics packages: `pathpy`, `BuildHON+`, `InfoMap`

The three higher-order data analytics packages listed above have been developed by the tutors and they are available free of charge under an Open Source license. Detailed setup instructions and introductory tutorials will be adapted from existing material[5] and distributed to the attendees prior to the tutorial. We provide a *test python script* that will allow attendees to check whether the environment has been installed correctly. As an additional solution for attendees that do not wish to set up the tools themselves, we will prepare and distribute a docker image that contains all necessary packages and data sets, and which exposes all necessary tools via a `jupyter` notebook server.

REFERENCES

[1] Charu Aggarwal and Karthik Subbian. 2014. Evolutionary Network Analysis: A Survey. *ACM Comput. Surv.* 47, 1, Article 10 (May 2014), 36 pages. https://doi.org/10.1145/2601412

[2] Tanya Y. Berger-Wolf and Jared Saia. 2006. A Framework for Analysis of Dynamic Social Networks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, USA, 523–528. https://doi.org/10.1145/1150402.1150462

[3] Kathleen M. Carley. 2003. Dynamic Network Analysis. In *Summary of the NRC workshop on Social Network Modeling and Analysis*, Ron Breiger Breiger and Kathleen M. Carley (Eds.).

[4] Laetitia Gauvin, André Panisson, and Ciro Cattuto. 2014. Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach. *PLoS ONE* 9, 1 (01 2014), e86028. https://doi.org/10.1371/journal.pone.0086028

[5] Derek Greene, Donal Doyle, and Padraig Cunningham. 2010. Tracking the evolution of communities in dynamic social networks. In *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*. IEEE, 176–183.

[6] Petter Holme. 2015. Modern temporal network theory: a colloquium. *The European Physical Journal B* 88, 9 (2015), 234.

[7] David Kempe, Jon Kleinberg, and Amit Kumar. 2000. Connectivity and inference problems for temporal networks. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. ACM, 504–513.

[8] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2005. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05)*. ACM, New York, NY, USA, 177–187. https://doi.org/10.1145/1081870.1081893

[9] Wei Liu, Andrey Kan, Jeffrey Chan, James Bailey, Christopher Leckie, Jian Pei, and Ramamohanarao Kotagiri. 2012. On Compressing Weighted Time-evolving Graphs. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM '12)*. ACM, New York, NY, USA, 2319–2322. https://doi.org/10.1145/2396761.2398630

[10] Bivas Mitra, Lionel Tabourier, and Camille Roth. 2012. Intrinsically dynamic network communities. *Computer Networks* 56, 3 (Feb. 2012), 1041–1053. https://doi.org/10.1016/j.comnet.2011.10.024

[11] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. 2010. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science* 328, 5980 (2010), 876–878. https://doi.org/10.1126/science.1184819 arXiv:http://www.sciencemag.org/content/328/5980/876.full.pdf

[12] Tiago P Peixoto and Martin Rosvall. 2017. Modelling sequences and temporal networks with dynamic community structures. *Nature Communications* 8, 1 (2017), 582.

[13] Qiang Qu, Siyuan Liu, Christian S. Jensen, Feida Zhu, and Christos Faloutsos. 2014. Interestingness-Driven Diffusion Process Summarization in Dynamic Networks. In *Machine Learning and*

[5] see e.g. https://ingoscholtes.github.io/pathpy/tutorial.html, http://www.mapequation.org/code.html, and https://github.com/xyjprc/hon

Knowledge Discovery in Databases, Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo (Eds.). Lecture Notes in Computer Science, Vol. 8725. Springer Berlin Heidelberg, 597–613. https://doi.org/10.1007/978-3-662-44851-9_38

[14] Martin Rosvall, Daniel Axelsson, and Carl T Bergstrom. 2009. The map equation. *The European Physical Journal-Special Topics* 178, 1 (2009), 13–23.

[15] Martin Rosvall and Carl T Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105, 4 (2008), 1118–1123.

[16] Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. 2014. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications* 5 (2014), ncomms5630.

[17] Vsevolod Salnikov, Michael T Schaub, and Renaud Lambiotte. 2016. Using higher-order Markov models to reveal flow-based communities in networks. *Scientific reports* 6 (2016), 23194.

[18] Ingo Scholtes. 2017. 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)*. ACM, New York, NY, USA, 1037–1046. https://doi.org/10.1145/3097983.3098145

[19] Ingo Scholtes, Nicolas Wider, and Antonios Garas. 2016. Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities. *The European Physical Journal B* 89, 3 (02 Mar 2016), 61. https://doi.org/10.1140/epjb/e2016-60663-0

[20] Ingo Scholtes, Nicolas Wider, Rene Pfitzner, Antonios Garas, Claudio Juan Tessone, and Frank Schweitzer. 2014. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature Communications* 5 (September 2014), 5024. https://doi.org/10.1038/ncomms6024 arXiv:1307.4030

[21] Neil Shah, Danai Koutra, Tianmin Zou, Brian Gallagher, and Christos Faloutsos. 2015. TimeCrunch: Interpretable Dynamic Graph Summarization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. ACM, New York, NY, USA, 1055–1064. https://doi.org/10.1145/2783258.2783321

[22] Jimeng Sun, Dacheng Tao, and Christos Faloutsos. 2006. Beyond streams and graphs: dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 374–383.

[23] John Tang, Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo, and Vito Latora. 2010. Small-world behavior in time-varying graphs. *Physical Review E* 81, 5 (2010), 055101.

[24] Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. 2007. A Framework for Community Identification in Dynamic Social Networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07)*. ACM, New York, NY, USA, 717–726. https://doi.org/10.1145/1281192.1281269

[25] Jun Tao, Jian Xu, Chaoli Wang, and Nitesh V Chawla. 2017. HoNVis: Visualizing and exploring higher-order networks. In *IEEE Pacific Visualization Symposium (PacificVis), 2017*. IEEE, 1–10.

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