



Special issue on statistical analysis of networks

Preface by the guest editors

Michael Schweinberger¹ · Francesco C. Stingo² · Maria Prosperina Vitale³

Published online: 9 November 2021

© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The special issue on *Statistical Analysis of Networks* aspires to convey the breadth and depth of statistical learning with networks, ranging from networks that are observed to networks that are unobserved and learned from data. It includes ten select papers with methodological and theoretical advances, and demonstrates the usefulness of the network paradigm by applications to current problems.

Keywords Network data · Graphical model · Neural network · Multivariate network · Multiplex network · Multipartite network

Mathematics Subject Classification 05C80 · 62B05 · 62F10 · 91D30

1 Preface

The special issue on *Statistical Analysis of Networks* aspires to convey the breadth and depth of statistical learning with networks. Networks have intrigued mathematicians and scientists since at least 1736, starting with the thought-provoking work of Euler on the *Seven Bridges of Königsberg*. In modern times, at least three streams of research involving statistical learning with networks have emerged. One stream is concerned with models of network data where edges between nodes

✉ Francesco C. Stingo
francescoaudio.stingo@unifi.it

Michael Schweinberger
michael.schweinberger@missouri.edu

Maria Prosperina Vitale
mvitale@unisa.it

¹ University of Missouri, Columbia, USA

² University of Florence, Florence, Italy

³ University of Salerno, Fisciano, Italy

represent observed relationships between actors (e.g., Wasserman and Faust 1994; Kolaczyk 2009), which started with Moreno's work on social networks in the 1930s and 1940s, the work of Gilbert, Erdős and Rényi on random graphs in the 1950s, and the work of Holland, Leinhardt, Fienberg, Wasserman and others on statistical learning from social network data in the 1980s. A second stream is concerned with *graphical models* where edges between nodes represent unobserved conditional dependence relationships between random variables (e.g., Pearl 1988; Lauritzen 1996; Maathuis et al. 2019), with roots in the 1960s and 1970s work of Bartlett and Whittle, Hammersley and Clifford, Besag and others on spatial Markov random fields, and the 1980s work of Lauritzen and Pearl and others on graphical models. A third stream is concerned with *neural networks* (e.g., Goodfellow et al 2016; Fan et al 2021), which represent mathematical operations performed by an artificial neural network processing input and producing output, an idea that can be traced back to the pioneering work of McCulloch and Pitts, Widrow and Hoff, Rosenblatt and others in artificial intelligence in the 1940s and 1950s. All of these streams have experienced ups and downs, but the idea that networks represent a useful set of tools for learning and predicting how the world operates has prevailed, and has found widespread application across a staggering number of domains. The special issue on *Statistical Analysis of Networks* includes ten select papers¹ with methodological and theoretical advances in those areas, with a focus on the first two streams: network data and graphical models.

The first paper on these topics, the paper of Anna Malinovskaya and Philipp Otto on "*Online network monitoring*", develops a statistical network monitoring approach to detecting anomalous behavior, leveraging ideas from the literature on network models and statistical process control. The proposed approach applies multivariate control charts based on exponential smoothing and cumulative sums for monitoring networks over time, generated by temporal exponential random graph models. The performance of the approach is demonstrated by monitoring flight patterns in the United States, detecting substantial changes in air travel before the COVID-19 pandemic and the disruption caused by the pandemic.

The second paper, "*Bayesian dynamic network actor models with application to South Korean COVID-19 patient movement data*" by Antonio M. Arrizza and Alberto Caimo, develops dynamic network actor models, with the goal of studying the movement of population members infected by Sars-CoV-2. The model helps quantify the uncertainty about when and where infected population members move, and provides interesting insights into the spread of Sars-CoV-2 in South Korea in 2020.

The following three papers are devoted to methodological and theoretical developments of graphical models for representing network structures in modern applications.

The theoretical results in Alberto Roverato's paper "*On the interpretation of inflated correlation path weights in concentration graphs*" present one such

¹ The special issue hosts some contributions presented at the ARS conference held in 2019. This biennial scientific meeting, organized in Italy since 2007, is dedicated to the most relevant results and the most recent methodological developments in this research area.

example, focusing on the inflated correlation coefficient and the properties of the corresponding path weights of undirected graphical models. A coefficient's weight can be factorized into terms associated with a node or an edge of the path. The factorization, on the one hand, helps provide insight into the role played by a path in the network and, on the other hand, can be exploited for computing centrality measures.

Federico Castelletti and Alessandro Mascaro propose in their paper "*Structural learning and estimation of joint causal effects among network-dependent variables*" a Bayesian methodology that combines structural learning of Gaussian directed acyclic graphical models and statistical learning of causal effects, arising from simultaneous interventions on a given set of variables in the system. The Bayesian approach captures the uncertainty about the network structure as well as the causal relationships.

The paper of Xiaoyi Yang, Nynke M.D. Niezink and Rebecca Nugent on "*Learning social networks from text data using covariate information*" presents encouraging results on the estimation of social networks from text data using a Local Poisson Graphical Lasso approach with penalties incorporating covariate information. The numerical examples include simulated data with characteristics of historical networks and biographical data of individuals who lived in Britain between 1500 and 1575.

The remaining papers are concerned with the statistical analysis of complex network data structures (e.g., multiplex networks, multimode networks) as well as the identification of groups of nodes using blockmodels and the clustering of ego-centered networks.

The paper of Termeh Shafie and David Schoch, entitled "*Multiplexity analysis of networks using multigraph representations*", proposes a framework for network analysis of aggregated multigraphs, aimed at detecting relevant associations between node and edge attributes. First, the exploratory analysis is performed by comparing frequencies of the different edges within and between aggregated node categories; second, for the confirmatory analysis some statistics are defined under different random multigraph models by considering the distribution of edge multiplicities and providing information on the covariation and dependencies of different edges given node attributes.

The paper on "*A network analysis of student mobility patterns from high school to master's*" by Vincenzo G. Genova, Michele Tumminello, Fabio Aiello and Massimo Attanasio is concerned with multipartite network data arising in the study of student mobility. The authors develop multipartite network methods and leverage them for studying the mobility patterns of students from Sicily, with a view to understanding where Sicilian students move to obtain a Master's degree, depending on where the students come from and where the students obtained a Bachelor's degree.

Tin L.J. Ng and Thomas B. Murphy extend in "*Weighted stochastic block model*" the classic stochastic block model to weighted edges. The parameters of the weighted stochastic block model are estimated by variational methods, and the asymptotic properties of the resulting approximate maximum likelihood estimators

is studied, assuming the number of blocks is fixed. The proposed model is applied to simulated data and data on a network of bike sharing stations in Washington, D.C.

In the paper “*The relative fit measure for evaluating a blockmodel*”, Marjan Cugmas, Ales Ziberna, and Anuska Ferligoj introduce a relative fit measure for selecting an appropriate blockmodel along with the number of blocks. The relative measure’s properties and its usefulness are demonstrated by simulations and two real-data examples.

Last, but not least, the paper on “*A clustering procedure for mixed-type data to explore ego network typologies. An application to elderly people living alone in Italy*” by Elvira Pelle and Roberta Pappadà presents a distance-based clustering procedure for identifying homogeneous groups of ego networks extracted from the Family and Social Subjects Survey conducted by the Italian National Statistical Institute. The authors introduce a dissimilarity measure for mixed-type data and exploit hierarchical clustering for grouping ego networks according to their composition highlighting gender differences.

Acknowledgements We are grateful to the Editors of Statistical Methods & Applications, first and foremost Carla Rampichini, Alessio Farcomeni and Stefano Campostrini, for giving us the opportunity to prepare the special issue. We would like to thank the large number of distinguished scholars who have been generous enough to review and help improve the papers submitted to the Special Issue by providing constructive feedback.

References

- Barnes JA, Harary F (1983) Graph theory in network analysis. *Social Netw* 5(2):235–244
- Fan J, Ma C, Zhong Y (2021) A selective overview of deep learning. *Stat Sci* 36:264–290
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
- Kolaczyk ED (2009) *Statistical analysis of network data: methods and models*. Springer, New York
- Lauritzen SL (1996) *Graphical models*, vol 17. Clarendon Press, Oxford
- Lusher D, Koskinen J, Robins G (eds) (2013) *Exponential random graph models for social networks: theory, methods, and applications*, vol 35. Cambridge University Press, Cambridge
- Maathuis M, Drton M, Lauritzen S, Wainwright M (2019) *Handbook of graphical models*. CRC Press, Boca Raton
- Pearl J (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, San Francisco
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.