Statistica Sinica **35** (2025), 1759-1762

Foreword

SPECIAL ISSUE ON NETWORK DATA ANALYSIS

Network data analysis is a powerful tool for understanding complex relationships and interactions across various domains (Goldenberg et al., 2010). For instance, online social networks provide insights into individuals' political orientations (Adamic and Glance, 2005); protein-protein interaction networks aid in predicting effective drug combinations (Cheng, Kovács and Barabási, 2019); co-authorship and citation networks reveal trends in academic research (Ji et al., 2021); international trade networks help analyze global trading relationships and economic movements (Anderson, 2011); and network structures also play a crucial role in estimating causal effects (Aronow and Samii, 2017).

With the growing availability of large-scale networks and the expanding range of network-related applications, statistical methodology and theory for network analysis have seen significant advancements recently. This special issue brings together a collection of papers that explore timely and relevant topics in the field.

Most statistical network models assume that the adjacency matrix is equal to a low-rank "signal" matrix plus a "noise" matrix, where the "signal" matrix carries information of latent communities (Holland, Laskey and Leinhardt, 1983; Hoff, Raftery and Handcock, 2002; Young and Scheinerman, 2007; Jin et al., 2025). How to uncover the latent community structure has been one of the core problems in network analysis. In the special issue, three papers present new community detection techniques.

Cerqueira and Levina propose a pseudo-likelihood approach for community detection in weighted networks, extending methods originally designed for binary networks. Their algorithm, based on the weighted stochastic block model (WSBM), balances efficiency, scalability, and flexibility while incorporating real-valued edge weights. They establish the theoretical consistency of their estimator under a homogeneous WSBM and show that a single iteration achieves weak consistency. Applying their method to fMRI-based brain connectivity networks, they identify meaningful community structures in both healthy and schizophrenic populations.

Yuan, Liu, Li, and Ma address community extraction in network data under stochastic block models, focusing on distinguishing core communities from background nodes. They develop efficient algorithms with polynomial complexity for

large-scale networks and prove their methods' theoretical optimality by showing that the estimators achieve the asymptotic minimax risk. The authors demonstrate the approach's effectiveness through simulations and an application to a political blog network, revealing hidden structures and differentiating between core and non-core members.

Subhadeep and Chen presents a method for community detection using Orthogonal Symmetric Non-negative Matrix Tri-Factorization (OSNTF) applied to the adjacency and normalized Laplacian matrices. They establish a connection between the factorization and a non-negative basis of an invariant subspace, similar to spectral clustering. Since exact OSNTF may not always be computable, they propose an approximate version that solves an optimization problem. The method is shown to be consistent for detecting communities in networks from the stochastic block model and its degree-corrected version.

Many network datasets include additional attributes or features associated with nodes, such as demographic information, behavioral data, or other relevant characteristics. These nodal covariates offer valuable insights into the network's structure and dynamics. Incorporating nodal information enables more advanced models and algorithms and enhances the ability to capture the complexity of real-world networks. In response to this trend, two papers in this special issue address the challenge of integrating nodal covariates with network data.

Jones, Ward, and Jiang proposes a mini-batch strategy for fitting the mixed membership stochastic blockmodel (MMSB) on networks. Their approach leverages nodal information and aggregated relational data and is computationally feasible for large-scale networks. They also present a scalable inference method that incorporates this additional nodal data, allowing for parallel stochastic variational inference. Applied to a citation network with over two million nodes, the method captures explainable structure and demonstrates improved convergence and parameter recovery compared to traditional MMSB approaches.

Hehir, Niu, and Slavkovic extends spectral clustering to networks with discrete node covariates, where the network structure is influenced by both a latent block model and homophily on observed covariates. The method achieves perfect clustering with high probability on large, sparse networks with discrete covariates. The method separates latent network structure from homophily on observed covariates and is applied to a network of online friendships among university students to uncover community structure not explained by covariates.

In dynamic (longitudinal) networks, the relationships between nodes change over time, so are the nodal attributes. Analyzing dynamic networks involves tracking the emergence, disappearance, or transformation of edges and nodes over time, providing insights into the temporal dynamics of the system. This

is a crucial topic for fields like epidemiology, social science, and communication studies, where temporal dependencies play a critical role in shaping network behavior and structure. In this special issue, three papers involve temporal network analysis.

Zhang, Zhang, Xi, and Song address the challenge of monitoring COVID-19 contagion using sewage water systems, given the limitations of at-home testing. The authors propose a supervised learning approach using time-series transitional models and mixed integer programming to identify key sewage manholes for sampling, maximizing cost-effectiveness while predicting infection rates. They introduce a multi-compartment dynamic model to simulate viral loads in wastewater, ensuring the consistency of sampling site selection. The method is demonstrated through a real-world data analysis example.

Grieshop, Feng, Hu, and Schweinberger develops a continuous-time stochastic process to model the interactions among players in team-based sports like basketball and football, where player interactions are indexed by space and time. The model helps understand and predict team performance, emphasizing that a team's strength depends not only on individual players but also on their interactions. They use a Bayesian approach to learn this stochastic process from high-resolution network data and present simulation results, including applications to football clubs like Juventus Turin and Inter Milan in the Italian premier soccer league.

Jiang and Xu introduce a multi-scale testing framework for point processes and longitudinal networks, addressing challenges in two-sample testing and community detection over time. Their method uses a hierarchical discretization scheme for global and local hypothesis testing, identifying differences in intensity functions and key community structures. The authors prove the minimax optimality of their test and validate it through simulations and real-world applications. Their framework, which uses eigenvalue-based and subgraph count-based tests, provides a computationally efficient alternative to kernel and wavelet-based methods.

Network data analysis has emerged as an active and rapidly expanding field, driven by a wealth of exciting developments in modeling, methodology, and practical applications. As the complexity and scale of network data continue to grow, new approaches are being introduced to address various challenges in understanding and interpreting the intricate relationships within these networks. The papers featured in this special issue offer insights into the current trends and innovations in statistical network analysis. We hope that they not only provide valuable perspectives on the state of the field but also spark further exploration and inspire future research.

Jiashun Jin[1]
Zheng Tracy Ke[2]
Emma Jingfei Zhang[3]
Ji Zhu[4]
Guest Editors (in alphabetical order) for this special issue

[1]Carnegie Mellon University
[2]Harvard University
[3]University of Miami
[4]University of Michigan

References

Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, 36–43.

Anderson, J. (2011). The gravity model. *Annual Review of Economics* **3**, 133–160.

Aronow, P. M. and Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics* **11**, 1912–1947.

Cheng, F., Kovács, I. A. and Barabási, A.-L. (2019). Network-based prediction of drug combinations. *Nature Communications* **10**, Article no. 1197.

Goldenberg, A., Zheng, A., Fienberg, S. and Airoldi, E. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning* **2**, 129–233.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098.

Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137.

Ji, P., Jin, J., Ke, Z. T. and Li, W. (2021). Co-citation and co-authorship networks of statisticians. *Journal of Business & Economic Statistics* **40**, 469–485.

Jin, J., Ke, Z. T., Tang, J. and Wang, J. (2025). Network goodness-of-fit for the block-model family. *arXiv:2502.08609*.

Young, S. J. and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-Graph*, 138–149.