DSSV 2023
Data Science, Statistics & Visualisation
EC DA 2023 European Conference on Data Analysis

A geometric approach to convergence rates for graphical models based on a single observation of discrete and dependent network and attribute data

A. Michael Schweinberger[1] and B. Jonathan R. Stewart[2]

[2]Department of Statistics, Pennsylvania State University, 326 Thomas Building, University Park, PA 16802, USA

[2]Department of Statistics, Florida State University, 117 N Woodward Ave, Tallahassee, FL 32306-4330, USA

Understanding and predicting how the interconnected and interdependent world of the twenty-first century operates and affects the welfare of billions of people around the world requires statistical procedures for learning from network and attribute data. More often than not, network and attribute data are discrete and dependent, and independent replications of networks are unavailable. In such scenarios, it is natural to base statistical learning on models with conditional independence properties, which facilitate theoretical guarantees for statistical procedures in single-observation scenarios. In other words, it is natural to base statistical learning on graphical models which possess conditional independence properties by construction, and admit exponential-family representations of joint distributions and GLM-representations of conditional distributions. Such models can be viewed as generalizations of GLMs for dependent network and attribute data and are widely used in practice, implemented in more than 20 `R` packages and downloaded more than 2.5 million times from the `RStudio` CRAN server alone. Having said that, probabilists and statisticians have expressed concern about the probabilistic behavior of such models and whether statistical learning is possible based on a single observation of discrete and dependent network and attribute data [e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Building on [12, 13], we introduce a novel geometric approach to obtaining convergence rates for scalable M-estimators of graphical models based on a single observation of discrete and dependent random variables, with applications to discrete and dependent network and attribute data.

[1] M. S. Handcock, "Statistical models for social networks: Inference and degeneracy," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (R. Breiger, K. Carley, and P. Pattison, eds.), pp. 1–12, Washington, D.C.: National Academies Press, 2003.

[2] A. Rinaldo, S. E. Fienberg, and Y. Zhou, "On the geometry of discrete exponential families with application to exponential random graph models," *Electronic Journal of Statistics*, vol. 3, pp. 446–484, 2009.

[3] S. Bhamidi, G. Bresler, and A. Sly, "Mixing time of exponential random graphs," *The Annals of Applied Probability*, vol. 21, pp. 2146–2170, 2011.

[4] M. Schweinberger, "Instability, sensitivity, and degeneracy of discrete exponential families," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1361–1370, 2011.

[5] S. E. Fienberg, "A brief history of statistical models for network analysis and open challenges," *Journal of Computational and Graphical Statistics*, vol. 21, pp. 825–839, 2012.

[6] S. Chatterjee and P. Diaconis, "Estimating and understanding exponential random graph models," *The Annals of Statistics*, vol. 41, pp. 2428–2461, 2013.

[7] M. Schweinberger and M. S. Handcock, "Local dependence in random graph models: characterization, properties and statistical inference," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 77, pp. 647–676, 2015.

[8] S. Lauritzen, A. Rinaldo, and K. Sadeghi, "Random networks, graphical models and exchangeability," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 80, pp. 481–508, 2018.

[9] S. Mukherjee, "Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics," *Bernoulli*, vol. 26, pp. 1016–1043, 2020.

[10] M. Schweinberger, "Consistent structure estimation of exponential-family random graph models with block structure," *Bernoulli*, vol. 26, pp. 1205–1233, 2020.

[11] M. Schweinberger, P. N. Krivitsky, C. T. Butts, and J. R. Stewart, "Exponential-family models of random graphs: Inference in finite, super, and infinite population scenarios," *Statistical Science*, vol. 35, pp. 627–662, 2020.

[12] M. Schweinberger and J. R. Stewart, "Concentration and consistency results for canonical and curved exponential-family models of random graphs," *The Annals of Statistics*, vol. 48, pp. 374–396, 2020.

[13] J. R. Stewart and M. Schweinberger, "Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension," tech. rep., Department of Statistics, Florida State University, 2022. https://arxiv.org/abs/2012.07167.