Statistical Methods & Applications
https://doi.org/10.1007/s10260-021-00600-7

COMMENT

Discussion to: *Bayesian graphical models for modern biological applications* by Y. Ni, V. Baladandayuthapani, M. Vannucci and F.C. Stingo

Michael Schweinberger[1]

Accepted: 5 September 2021

Abstract
It is a pleasure to congratulate Ni et al. (Stat Methods Appl 490:1–32, 2021) on the recent advances in Bayesian graphical models reviewed in Ni et al. (Stat Methods Appl 490:1–32, 2021). The authors have given considerable thought to the construction and estimation of Bayesian graphical models that capture salient features of biological networks. My discussion focuses on computational challenges and opportunities along with priors, pointing out limitations of the Markov random field priors reviewed in Ni et al. (Stat Methods Appl 490:1–32, 2021) and exploring possible generalizations that capture additional features of conditional independence graphs, such as hub structure and clustering. I conclude with a short discussion of the intersection of graphical models and random graph models.

Keywords Graphical models · Random graphs · Random graph priors · Nodewise regression

1 Priors: generalizations of Markov random graph priors

I agree with Ni et al. (2021) that two important advantages of a Bayesian approach are the ability to incorporate prior knowledge and provide a natural measure of uncertainty about unknown quantities of interest, including the conditional independence graphs of graphical models (Lauritzen 1996; Maathuis et al. 2019). Having said that, the Markov random field priors reviewed in Ni et al. (2021) have limitations. I single out two of them and explore possible generalizations that capture additional features of conditional independence graphs, such as hub structure and clustering.

✉ Michael Schweinberger
michael.schweinberger@missouri.edu

[1] Department of Statistics, University of Missouri, Columbia, MO, USA

Published online: 03 November 2021

1.1 Markov random graph priors

The Markov random field priors reviewed in Ni et al. (2021), proposed by Peterson et al. (2015) and others, assume that the prior probability of the conditional independence graphs $\mathcal{G}_1, \ldots, \mathcal{G}_K$ of groups of subjects $1, \ldots, K$ is proportional to

$$\mathbb{P}(\mathcal{G}_1, \ldots, \mathcal{G}_K \mid \boldsymbol{v}, \boldsymbol{\Theta}) \propto \prod_{i<j}^{p} \exp\left(v_{i,j} \sum_{k=1}^{K} g_{k,ij} + \sum_{k=1}^{K} \sum_{l=1}^{K} \theta_{kl}\, g_{k,ij}\, g_{l,ij} \right). \quad (1)$$

While Ni et al. (2021)—in line with Peterson et al. (2015)—call such priors Markov random field priors, it would be more appropriate to call them Markov random graph priors, because such priors are distributions on sets of graphs with Markov properties (Frank and Strauss 1986; Lauritzen et al. 2018). To be sure, Markov random fields and Markov random graphs have common ancestors in the form of Markov random fields in spatial statistics (Bartlett 1955; Whittle 1963; Besag 1974), but linking such priors to Markov random graphs opens the door to constructing more sophisticated priors that capture additional features of conditional independence graphs, such as hub structure and clustering. To explore Markov random graph priors and possible generalizations in more depth, let me first point out two implicit assumptions made by the Markov random graph priors reviewed in Ni et al. (2021).

1.2 Assumption 1: between-group dependence and within-group independence

While accomplishing the stated goal of encouraging a connection between two variables i and j in the conditional independence graph of group k if the same connection is present in the conditional independence graphs of related groups $l \neq k$, the Markov random graph priors reviewed in Ni et al. (2021) assume that the conditional probability of the event that there exists an edge between two variables i and j in group k, given everything else, does not depend on the absence or presence of edges between other pairs of variables:

$$\mathbb{P}(g_{k,ij} = 1 \mid \text{ everything else}) \propto \exp\left(v_{i,j} + \sum_{l \neq k}^{K} \theta_{kl}\, g_{l,ij} \right). \quad (2)$$

The assumption of within-group independence is restrictive, because data on networks have revealed that connections tend to depend on other connections. As a case in point, it is not credible that connections between regions of the human brain are independent, because complex tasks such as walking, writing, and performing mathematical operations may require the coordination of multiple brain regions.

1.3 Assumption 2: short-tailed degree distributions

The assumption of within-group independence and the absence of model components that encourage other relevant features of conditional independence

graphs suggests that the model does not capture many relevant features of conditional independence graphs, such as hub structure. If conditional independence graphs contain hubs, their degree distributions are long-tailed rather than short-tailed. Therefore, priors should accommodate both short- and long-tailed degree distributions.

1.4 Generalizations of Markov random graph priors I: hub structure and clustering

Markov random graph priors are special cases of exponential-family random graph priors, which provide a large and flexible class of priors on sets of graphs with convenient properties (Schweinberger et al. 2020). The class of exponential-family random graph priors constitutes a natural class of priors, because graphical models and exponential-family random graph priors have common ancestors in the form of Markov random fields in spatial statistics (Besag 1974) and share the same mathematical platform: the statistical exponential-family platform (Sundberg 2019).

One of the greatest advantages of exponential-family random graph priors is the fact that within-group dependence can be captured along with addititional features of conditional independence graphs, such as hub structure and clustering. For example, using an exponential-family model with the degrees of nodes as sufficient statistics along with Dirichlet process priors, Schweinberger et al. (2021) constructed a flexible semiparametric Bayesian model that allows some of the nodes to have many connections while others have few connections. Exponential-family random graph priors capturing hub structure along with between- and within-group dependence can be constructed along similar lines, and provide a Bayesian alternative to the non-Bayesian approach of Mohan et al. (2014). An additional feature of many networks is clustering, which can likewise be captured by exponential-family random graph priors. The literature on exponential-family random graphs provides valuable lessons on how to capture such features of conditional independence graphs through exponential-family random graph priors (Hunter et al. 2012; Lusher et al. 2013; Schweinberger et al. 2020).

1.5 Generalizations of Markov random graph priors II: groups of different sizes

Exponential-family random priors open the door to weakening other assumptions made in Ni et al. (2021), e.g., the assumption that the number of variables p is constant across all groups of subjects. If the groups of subjects correspond to patients in various stages of a disease and different stages of the disease are driven by different genes, the assumption that the number of variables is constant across all stages of the disease may be restrictive. Constructing models that allow the number of variables to vary according to the stage of the disease would be interesting, but gives rise to additional challenges: e.g., if edges are assumed to be independent Bernoulli(μ) random variables and $\mu \in (0, 1)$ does not depend on the number of nodes, the expected degrees of nodes grow with the number of nodes. In the applications considered in Ni et al. (2021), the expected degrees of nodes in the

conditional independence graph of group k would depend on the number of nodes in the conditional independence graph of group k, which may be an untenable assumption. It would be more reasonable to assume that the expected degrees of nodes are invariant to the number of nodes in the conditional independence graphs of the groups. Exponential-family random graph priors with such invariance properties have been proposed in the literature on exponential-family random graphs (Krivitsky et al. 2011; Krivitsky and Kolaczyk 2015; Butts and Almquist 2015). Exponential-family random graph priors with other invariance properties can likewise be constructed.

2 Computational challenges and opportunities

While some scalable approaches to Bayesian inference for graphical models have been explored (e.g., Tan et al. 2017; Li et al. 2020), scalable Bayesian inference remains an open problem. I highlight two interesting directions for future research.

The first one is a Bayesian approach adapting the main ideas of the ℓ_1-penalized nodewise regression approach of Meinshausen and Bühlmann (2006), without losing the Bayesian advantage of capturing uncertainty. The ℓ_1-penalized nodewise regression approach of Meinshausen and Bühlmann (2006) reduces the estimation of Gaussian graphical models to ℓ_1-penalized regressions of each variable on all other variables. The ℓ_1-penalized regressions help estimate the neighborhood of each node in the conditional independence graph, and combining the estimated neighborhoods gives rise to an estimate of the entire conditional independence graph. The ℓ_1-penalized regression approach has at least three advantages:

- Computational advantages: ℓ_1-penalized nodewise regression reduces the estimation of Gaussian graphical models to ℓ_1-penalized linear regressions, which can be performed on multi-core computers and computing clusters in parallel, facilitating large-scale computing.
- Theoretical advantages: Since ℓ_1-penalized nodewise regression reduces the estimation of Gaussian graphical models to ℓ_1-penalized regressions and the properties of ℓ_1-penalized regressions have been studied, one can conclude that the neighborhoods of all nodes in the conditional independence graph can be recovered with high probability and, by a union bound, the entire conditional independence graph can be recovered with high probability (provided strong assumptions are satisfied).
- The ℓ_1-penalized nodewise regression approach is simple and can be adapted to non-Gaussian graphical models (e.g., Ravikumar et al. 2010).

It would be worthwhile to explore whether a form of Bayesian nodewise regression helps scale up Bayesian inference for graphical models, without losing the Bayesian advantage of capturing uncertainty.

The second one is a Bayesian approach based on pseudo-likelihoods (Besag 1974) or composite-likelihoods (Lindsay 1988; Lindsay et al. 2011). Despite scepticism in the statistical literature concerned with spatial data (e.g., Besag 2001) and network data (e.g., Corander et al. 2002; van Duijn et al. 2009), the

pseudolikelihood-based approach of Besag (1974) seems to undergo a renaissance, both in the Bayesian literature (e.g., Ghosh et al. 2021) and the non-Bayesian literature (e.g., Amini et al. 2013; Ghosal and Mukherjee 2020; Stewart and Schweinberger 2021). Indeed, while pseudo- and composite-likelihoods may not perform well when the dependence induced by the model is too strong (as Besag pointed out in 2001), such approaches are scalable and can perform well when the dependence is not too strong. It would be interesting to explore Bayesian pseudo- or composite-likelihood approaches to graphical models with a view to scaling up Bayesian inference for graphical models.

3 Graphical models and random graph models

In the discussion of Ni et al. (2021), the authors bring up an interesting connection with a related branch of literature: models of network data (Kolaczyk 2009).

3.1 Graphs representing data structure, model structure, and mathematical operations

To discuss the connection between models of network data and graphical models in more depth, it is useful to distinguish three streams of data science involving graphs:

- Graphs that represent data structure: random graph models (e.g., Holland and Leinhardt 1981; Frank and Strauss 1986; Lauritzen et al. 2018; Schweinberger et al. 2020; Hoff 2021).
- Graphs that represent model structure: graphical models (e.g., Pearl 1988; Lauritzen 1996; Maathuis et al. 2019).
- Graphs that represent mathematical operations: e.g., neural networks (e.g., Goodfellow et al. 2016; Schmidt-Hieber 2020; Fan et al. 2021).

These three streams intersect: e.g., graphical models and random graph models intersect (e.g., Frank and Strauss 1986; Lauritzen et al. 2018) and have common ancestors in the form of Markov random fields in spatial statistics (Bartlett 1955; Whittle 1963; Besag 1974); and graphical models and neural networks intersect, because some generative neural network models have representations as graphical models (e.g., restricted Boltzmann machines, MacKay 2003).

3.2 Learning dependence structure

Ni et al. (2021) suggest that random graph models and graphical models are distinguished by whether networks are observed or unobserved, in which case the networks are learned from data.

The relationship between these two model classes is more complicated. First of all, it is helpful to keep in mind that every distribution $\mathbb{P}$—including every random graph distribution—factorizes with respect to some graph (e.g., the complete graph with all possible edges), although the graph may not be sparse and may not be able to represent all conditional independencies of $\mathbb{P}$. Indeed, many random graph

models possess conditional independencies and, in most applications, the conditional independence structure of random graph models is unknown and specifying the conditional independence structure is challenging. Worse, some of the earliest attempts to specify the conditional independence structure of random graph models turned out to be problematic (e.g., Handcock 2003; Schweinberger 2011; Chatterjee and Diaconis 2013). While models with more reasonable assumptions and properties have been developed (Schweinberger et al. 2020) and some of them have been equipped with provable statistical guarantees (Schweinberger and Stewart 2020; Mukherjee 2020), it would be preferable to learn the conditional independence structure of random graph models from data rather than specifying the conditional independence structure. In the literature on graphical models, considerable progress has been made on learning the conditional independence structure of models from data (e.g., Meinshausen and Bühlmann 2006; Maathuis et al. 2019), but most of those advances require independent replications from the same source (in a well-defined sense). In studies involving network data, independent replications from the same source are hard to obtain, and even when independent replications are available (e.g., in multilevel network data sets, Lazega and Snijders 2016), the replications may not have the same size: e.g., if networks of friendships within school classes can be assumed to be independent across school classes, then independent replications of within-school-class networks are available, but the school classes may not have the same size, because researchers do not have control over the sizes of school classes.

In general, given a single observation of dependent random variables (e.g., a single observation of a population graph with dependent edges), it is challenging to learn the conditional independence structure of models from data. In special cases, one can learn the conditional independence structure provided that the conditional independence structure satisfies strong restrictions, as demonstrated by the work on Csiszár and Talata (2006) concerned with Markov random fields on lattices and the work of Schweinberger (2020) concerned with random graphs satisfying local dependence assumptions. That said, estimating the conditional independence graph based on a single observation of a population graph with dependent edges is non-trivial in many other cases.

Acknowledgements I acknowledge support from the US National Science Foundation in the form of NSF awards DMS-1513644 and DMS-1812119.

References

Amini AA, Chen A, Bickel PJ, Levina E (2013) Pseudo-likelihood methods for community detection in large sparse networks. Ann Stat 41:2097–2122

Bartlett M (1955) An introduction to stochastic processes. Cambridge University Press, Cambridge, UK

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J Roy Stat Soc B 36:192–225

Besag J (2001) Comment: conditionally specified distributions: an introduction. Stat Sci 16:265–267

Butts CT, Almquist ZW (2015) A flexible parameterization for baseline mean degree in multiple-network ERGMs. J Math Sociol 39:163–167

Chatterjee S, Diaconis P (2013) Estimating and understanding exponential random graph models. Ann Stat 41:2428–2461

Corander J, Dahmstrom K, Dahmstrom P (2002) Maximum likelihood estimation for exponential random graph models. In: Hagberg J (ed) Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics. A Festschrift in honour of Ove Frank, University of Stockholm, Department of Statistics, pp 1–17

Csiszár I, Talata Z (2006) Consistent estimation of the basic neighborhood of Markov random fields. Ann Stat 34:123–145

Fan J, Ma C, Zhong Y (2021) A selective overview of deep learning. Stat Sci 36:264–290

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81:832–842

Ghosal P, Mukherjee S (2020) Joint estimation of parameters in Ising model. Ann Stat 48:785–810

Ghosh S, Khare K, Michailidis G (2021) Strong selection consistency of Bayesian vector autoregressive models based on a pseudo-likelihood approach. Ann Stat 49(3):1267–1299

Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press. http://www.deeplearningbook.org

Handcock MS (2003) Statistical models for social networks: inference and degeneracy. In: Breiger R, Carley K, Pattison P (eds) Dynamic social network modeling and analysis: workshop summary and papers. National Academies Press, Washington, pp 1–12

Hoff PD (2021) Additive and multiplicative effects network models. Stat Sci 36:34–50

Holland PW, Leinhardt S (1981) An exponential family of probability distributions for directed graphs. J Am Stat Assoc 76:33–65

Hunter DR, Krivitsky PN, Schweinberger M (2012) Computational statistical methods for social network models. J Comput Graph Stat 21:856–882

Kolaczyk ED (2009) Statistical analysis of network data: methods and models. Springer, New York

Krivitsky PN, Kolaczyk ED (2015) On the question of effective sample size in network modeling: an asymptotic inquiry. Stat Sci 30:184–198

Krivitsky PN, Handcock MS, Morris M (2011) Adjusting for network size and composition effects in exponential-family random graph models. Stat Methodol 8:319–339

Lauritzen S (1996) Graphical models. Oxford University Press, Oxford, UK

Lauritzen S, Rinaldo A, Sadeghi K (2018) Random networks, graphical models and exchangeability. J R Stat Soc Ser B (Stat Methodol) 80:481–508

Lazega E, Snijders TAB (eds) (2016) Multilevel network analysis for the social sciences. Springer, Switzerland

Li ZR, McComick TH, Clark SJ (2020) Using Bayesian latent Gaussian graphical models to infer symptom associations in verbal autopsies. Bayesian Anal 15:781–807

Lindsay BG (1988) Composite likelihood methods. Contemp Math 80:221–239

Lindsay BG, Yi GY, Sun J (2011) Issues and strategies in the selection of composite likelihoods. Stat Sin 21:71–105

Lusher D, Koskinen J, Robins G (2013) Exponential random graph models for social networks. Cambridge University Press, Cambridge, UK

Maathuis M, Drton M, Lauritzen S, Wainwright M (2019) Handbook of graphical models. CRC Press, Boca Raton

MacKay D (2003) Information theory, inference and learning algorithms. Cambridge University Press, Cambridge

Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the LASSO. Ann Stat 34:1436–1462

Mohan K, London P, Fazel M, Witten D, Lee SI (2014) Node-based learning of multiple Gaussian graphical models. J Mach Learn Res 15:445–488

Mukherjee S (2020) Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics. Bernoulli 26:1016–1043

Ni Y, Baladandayuthapani V, Vannucci M, Stingo FC (2021) Bayesian graphical models for modern biological applications. Stat Methods Appl 490:1–32

Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Francisco

Peterson C, Stingo FC, Vannucci M (2015) Bayesian inference of multiple Gaussian graphical models. J Am Stat Assoc 110:159–174

Ravikumar P, Wainwright MJ, Lafferty J (2010) High-dimensional Ising model selection using ℓ_1-regularized logistic regression. Ann Stat 38:1287–1319

Schmidt-Hieber J (2020) Nonparametric regression using deep neural networks with ReLU activation function. Ann Stat 48:1875–1897

Schweinberger M (2011) Instability, sensitivity, and degeneracy of discrete exponential families. J Am Stat Assoc 106(496):1361–1370

Schweinberger M (2020) Consistent structure estimation of exponential-family random graph models with block structure. Bernoulli 26:1205–1233

Schweinberger M, Stewart JR (2020) Concentration and consistency results for canonical and curved exponential-family models of random graphs. Ann Stat 48:374–396

Schweinberger M, Krivitsky PN, Butts CT, Stewart JR (2020) Exponential-family models of random graphs: inference in finite, super, and infinite population scenarios. Stat Sci 35:627–662

Schweinberger M, Bomiriya RP, Babkin S (2021) A semiparametric Bayesian approach to epidemics, with application to the spread of the coronavirus MERS in South Korea in 2015. J Nonparam Stat 1–35, to appear

Stewart JR, Schweinberger M (2021) Pseudo-likelihood-based *M*-estimators for random graphs with dependent edges and parameter vectors of increasing dimension. Technical report, Department of Statistics, Florida State University. arXiv:2012.07167

Sundberg R (2019) Statistical modelling by exponential families. Cambridge University Press, Cambridge

Tan LSL, Jasra A, Iorio MD, Ebbels TMD (2017) Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks. Ann Appl Stat 11:2222–2251

van Duijn MAJ, Gile K, Handcock MS (2009) A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Soc Netw 31:52–62

Whittle P (1963) Stochastic processes in several dimensions. Bull Int Stat Inst 40:974–994