Discussion of *"Bayesian Graphical Models for Modern Biological Applications"* by Ni et al.

Anindya Bhadra
Department of Statistics, Purdue University, West Lafayette, IN 47907-2066, USA
Email: bhadra@purdue.edu

The authors are to be congratulated for this timely and important review of graphical models, which remains a fundamental building block for network estimation in diverse applications, including in genomics and neuroimaging, which the present article highlights. Throughout, the authors focus on Bayesian methodology, which makes probabilistic quantification of uncertainty automatic, but also typically results in more challenging computation. Within the Bayesian paradigm, the authors' treatment is comprehensive and current for the most part. In what follows, I comment on a few specific aspects of the article and point out a few more related and recent developments.

1 Structured versus unstructured sparsity in undirected Gaussian graphical models and on the question of posterior concentration

Just as in the case of Bayesian variable selection in linear models using point mass mixture priors with a delta function spike at zero, a distinction can be made among possible approaches for graphical models depending on whether one desires exact zeros in the inferred precision matrix. Exact zeros are pleasing and easier to interpret, but their practical implementation often calls for additional assumptions such as *decomposability*. For undirected graphical models, this assumption is tantamount to: *a chordless cycle of length four or more is not allowed* (Lauritzen, 1996, p. 9). I do not know and cannot think of any biological motivation behind such a restriction. Indeed, as the authors point out, this restriction stems from the fact that calculating the normalizing constant for general graphs is NP-complete whereas for certain decomposable graphs (such as trees) this can be accomplished in polynomial time. If one is willing to dispense with exact zeros, an alternative to modeling without this structural assumption of decomposability is with continuous shrinkage priors, of which the authors refer to the Bayesian graphical lasso (BGL) model and the associated sampling algorithm of Wang (2012), which is a graphical application of the popular Bayesian lasso prior of Park and Casella (2008). While it is true that the posterior mode of the Bayesian graphical lasso corresponds to the frequentist graphical lasso estimate (Friedman et al., 2008), the picture is far less rosier if one takes a more nuanced look at the entire posterior. A useful warning, although in the context of linear regression models rather than in graphical models, is by Castillo et al. (2015), who have this to say: *"the LASSO is essentially non-Bayesian, in the sense that the corresponding full posterior distribution is a useless object."* Thus, it should come as no surprise that these poor posterior

concentration properties might be inherited by BGL, as Banerjee and Ghosal (2015) indeed suggest. Li et al. (2019) recently proposed the graphical horseshoe (GHS) prior, an application of the popular global-local horseshoe prior (Carvalho et al., 2010) to the problem of unstructured precision matrix estimation in Gaussian graphical models (GGMs). These priors have been extensively researched over the last decade; see Bhadra et al. (2019b, 2020) for through reviews. The empirical observation of Li et al. (2019) is that GHS demonstrates considerably superior performance over both state of the art frequentist (e.g., glasso) and Bayesian (e.g., BGL) alternatives. However, they did not study posterior concentration properties of the GHS. Uncertainty quantification and frequentist properties of the horseshoe prior in linear models have been established recently in a series of extensive studies (van der Pas et al., 2017; van der Pas et al., 2016), which makes one wonder whether these properties hold in graphical models as well. This is indeed the case, and Sagar et al. (2021) establish posterior concentration results under both horseshoe and the closely related horseshoe-like (Bhadra et al., 2019a) priors in GGMs. They also provide fast point estimation algorithms under the same prior–penalty dual. Zhang et al. (2021) also consider the GHS model, and establish the consistency of the pseudo posterior, as opposed to the full posterior handled by Sagar et al. (2021). Both Zhang et al. (2021) and Sagar et al. (2021) work under the framework of posterior concentration as framed in Banerjee and Ghosal (2015) for point mass mixture priors, and remain among the first works to address the question of posterior concentration for unstructured sparse undirected GGMs.

Similar questions on posterior concentration could be asked in *directed* graphical models under unstructured sparsity, and I believe this area to be still open to further exploration at present. Banerjee and Ghosal (2015) developed the framework of posterior concentration under which both Zhang et al. (2021) and Sagar et al. (2021) operate. However, as the authors noted, estimation of directed graphical models usually proceeds via node conditional regressions, where the coefficients may or may not be covariate dependent. Thus, I expect there might be some additional hurdles to overcome before an application of the Banerjee and Ghosal (2015) approach is possible for partial regression based estimates of directed GGMs.

2 Bayesian inference in multiple GGMs: scalability in terms of the number of categories

The authors do an excellent job of reviewing the current state of the art in Bayesian estimation of multiple GGMs. The approaches based on Markov random field (MRF) priors to connect the multiple graphs are among the most natural and indeed philosophically pleasing, similar to a spike-and-slab prior with a point mass at zero for sparse linear regression problems. However, for the fully Bayesian approaches such as Peterson et al. (2015), scalability in terms of the number of groups K appears to remain a major hurdle. Most papers the authors cited, and the multiple myeloma application considered in this paper, concern 3 or 4 categories. The authors point out the main reason for this problem: the computation of the normalizing constant is intractable under an MRF prior when K is large, the lack of which unfortunately makes model comparison via Bayes factors very difficult or even impossible. The doubly intractable procedures do offer a possible workaround, but I am not completely convinced regarding their scalability at present.

A recent alternative is by Yang et al. (2021), who seem to have gotten around this problem using the spike-and-slab lasso prior. However, their approach only gives the posterior mode via an expectation–maximization (EM) algorithm and cannot be labeled as fully Bayesian. Thus, it appears there remains a distinct lack of scalable (in terms of K) approaches for multiple

undirected GGMs in the Bayesian literature and some more thoughts or suggestions from the authors regarding how to approach this bottleneck would have been very useful.

3 Moving beyond an assumption of normality

It is easy enough to be seduced by the rich literature on *Gaussian* graphical models to forget that normality is in fact a crucial assumption that is often violated in practice. Indeed, my experience is that simple marginal q–q plots or Kolmorov–Smirnov tests often indicate significant departures from normal marginals for genomic data, which also invalidate an assumption of multivariate normality. The authors refer to some Bayesian literature in undirected models robust to an assumption of normality; prominent among them are Finegold and Drton (2011, 2014) who worked with t-distributed marginals and Bhadra et al. (2018), who introduced a flexible framework of conditional sign independence via random scale transformations of non-normal margins. Conditional sign independence is a weaker form of Markov property compared to conditional independence (the strongest) and conditional uncorrelatedness (intermediate), where one can make statements concerning the signs of the random variables, but not necessarily regarding their magnitudes. However, the framework of Bhadra et al. (2018) is very flexible in two main ways: (a) each marginal is modeled as a normal scale mixture in a data-dependent manner, independently of each other and (b) the inferred sign independence holds among the *observed* variables, as opposed to through some underlying *latent* GGMs (e.g. Pitt et al., 2006). Under certain special cases, such as where the same scaling variable is used across all margins, one recovers a conditional correlation network as in Finegold and Drton (2011). If all scaling variables are identically equal to one, the model reduces to the usual GGM. The fact that the strongest Markov property of conditional independence is only achievable under the strictest assumption of multivariate normality is not surprising, but often overlooked.

However, similar departures from normality can certainly occur in several other problems highlighted by the authors, such as in directed or multiple graphical models. A notable recent work in this direction is by Chakraborty et al. (2021), who introduced the framework of sign independence of Bhadra et al. (2018) into the chain graph models considered by Ha et al. (2021), thereby allowing for non-normal marginals in modeling multiplatform genomic data arising from non small cell lung cancer. This is an important step, since for multi-platform genomic data that respect a chain graph hierarchy, Chakraborty et al. (2021) fit a *joint* model for copy number aberrations, mRNA expressions, protein expressions and drug responses; and non-normality can appear in multiple layers. Barring a few exceptions such as this, it is my opinion that inference robust to an assumption of normality is a highly untapped subfield of graphical models research, where methodological breakthroughs might lead to real biological impact.

4 Concluding remarks

I conclude by reiterating that I thoroughly enjoyed reading this timely and comprehensive review article. As a fellow Bayesian, I commiserate on the feeling that we are always playing "catch up" in terms of computational speed and scalability, never mind that the focus is not on obtaining merely a point estimate. However, I also believe these very same computational and methodological challenges keep us in business! Many of these leads and challenges in the context of graphical models have been expertly outlined by the authors.

Acknowledgements

This research is partially supported by US National Science Foundation grant DMS-2014371.

References

Banerjee, S. and Ghosal, S. (2015). Bayesian structure learning in graphical models. *Journal of Multivariate Analysis* **136,** 147–162.

Bhadra, A., Datta, J., Li, Y., and Polson, N. (2020). Horseshoe regularisation for machine learning in complex and deep models (with discussion). *International Statistical Review* **88,** 302–320.

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. T. (2019a). The horseshoe-like regularization for feature subset selection. *Sankhya B* pages 1–30.

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. T. (2019b). Lasso meets horseshoe: A survey. *Statistical Science* **34,** 405–427.

Bhadra, A., Rao, A., and Baladandayuthapani, V. (2018). Inferring network structure in non-normal and mixed discrete-continuous genomic data. *Biometrics* **74,** 185–195.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97,** 465–480.

Castillo, I., Schmidt-Hieber, J., and van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *The Annals of Statistics* **43,** 1986–2018.

Chakraborty, M., Baladandayuthapani, V., Bhadra, A., and Ha, M. J. (2021). Bayesian robust learning in chain graph models for integrative pharmacogenomics. *Technical Report, The University of Texas MD Anderson Cancer Center* .

Finegold, M. and Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *The Annals of Applied Statistics* **5,** 1057–1080.

Finegold, M. and Drton, M. (2014). Robust Bayesian graphical modeling using dirichlet t-distributions. *Bayesian Analysis* **9,** 521–550.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9,** 432–441.

Ha, M. J., Stingo, F. C., and Baladandayuthapani, V. (2021). Bayesian structure learning in multilayered genomic networks. *Journal of the American Statistical Association* **116,** 605–618.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford, United Kingdom.

Li, Y., Craig, B. A., and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics* **28,** 747–757.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686.

Peterson, C., Stingo, F. C., and Vannucci, M. (2015). Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association* **110,** 159–174.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* **93,** 537–554.

Sagar, K., Banerjee, S., Datta, J., and Bhadra, A. (2021). Precision matrix estimation under the horseshoe-like prior-penalty dual. *arXiv preprint arXiv:2104.10750* .

van der Pas, S., Salomond, J.-B., and Schmidt-Hieber, J. (2016). Conditions for Posterior Contraction in the Sparse Normal Means Problem. *Electronic Journal of Statistics* **10,** 976–1000.

van der Pas, S., Szabó, B., and van der Vaart, A. (2017). Uncertainty quantification for the horseshoe (with discussion). *Bayesian Analysis* **12,** 1221–1274.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis* **7,** 867–886.

Yang, X., Gan, L., Narisetty, N. N., and Liang, F. (2021). GemBag: Group estimation of multiple Bayesian graphical models. *Journal of Machine Learning Research* **22,** 1–48.

Zhang, R., Yao, Y., and Ghosh, M. (2021). Contraction of a quasi-Bayesian model with shrinkage priors in precision matrix estimation. *arXiv preprint arXiv:2104.12060* .