1 Introduction

We thank the discussants for their kind and insightful comments and analysis, which have substantially added to the contribution of this issue. Overall, it seems that the discussants agree with the general approach taken to analyze complex biological networks, and have also raised a number of relevant and important issues that we did not emphasize in the paper. Additionally, the discussion also raised many interesting open questions that point to potential research projects. Several common threads emerged from these discussions, which we grouped as follows: graphical model vs random graphs, computation, ordering, heterogeneity, interpretation, models beyond Gaussianity, graph shrinkage priors, and robustness of multiple graphical models. In response, we will provide our comments to the aforementioned general themes; we will point out to the discussants that raised the specific comments.

2 Graphical model vs random graphs

In this section we provide a reply to some of the comments concerning similarities and differences between graphical models and random graphs raised by Michael Schweinberger (MS) and Maria Prosperina Vitale, Giuseppe Giordano, and Giancarlo Ragozini (VGR).

As MS pointed out, graphical models and random graphs, despite their prominent differences in theory, methodology, and computation, can have interesting
intersections. From the modeling point of view, various random graph models can be adapted as prior models for latent graphs in Bayesian graphical models or as structured penalties for the frequentist counterpart. In fact, the uniform graph prior typically used in graphical models puts equal mass on any graph for a fixed number of nodes, which is exactly an Erdős–Rényi random graph with edge connecting probability 0.5. More sophisticated random graph priors may be able to encourage specific graph structures expected from certain applications. For example, in our recent work (Chung et al. 2021), we proposed a treed latent space model as a prior model for estimating microbial association networks, which allowed us to identify plausible microbial communities that were otherwise masked under the Erdős–Rényi random graph prior. In the past, we (Arbel et al. 2017) also briefly tested the generalized gamma process-based random graphs (Caron and Fox 2017) as prior models and they seemed to outperform the Erdős–Rényi random graph prior in the sparse graph setting. Other examples exist in this direction such as those incorporating hub structures (Mohan et al. 2014; Kim et al. 2019).

From the application point of view, VGR conducted in their discussion an interesting bibliometric analysis to link the main themes characterizing graphical models and random graph models. Specifically, the co-word network analysis found that systems biology and gene regulatory networks are overlapping applications of both areas. This is largely consistent with our understanding of the field that graphical models are useful for inferring an unknown gene regulatory network from multi-omic data and then random graph models can be subsequently applied to the inferred gene regulatory network to extract useful information such as node degree distribution and communities. In fact, this two-step approach could be replaced by a one-step approach, i.e., jointly estimating latent graphs and extracting/encouraging certain network structure, via e.g., random graph priors as mentioned above.

3 Computation

In this section we provide a reply to some of the comments concerning computational aspects of the discussed methods raised by Michael Schweinberger (MS), Anindya Bhadra (AB), and Yize Zhao, Zhe Sun and Jian Kang (ZSK).

A noteworthy limitation of the fully Bayesian inference approaches for graphical models that we reviewed is the scalability as pointed out by MS, AB, and ZSK. To scale up these methods, one may replace the proper likelihood function by a pseudo-likelihood function based on nodewise regression, which is a trade off with respect to the propriety of the probability model for scalability (Ha et al. 2020). Alternatively, it is also possible to develop optimization-based computation algorithms such as Variational Bayes (VB) or Expectation–maximization (EM) algorithms, which allow one to obtain quick point estimates of graphs. For example, Li and McCormick (2019) consider a formulation of the continuous spike-and-slab prior of Wang (2015) that includes additional scaling parameters and propose an EM algorithm for efficient estimation. Yang et al. (2021) employ EM in a Bayesian approach to multiple graph estimation. Also, for the case of compositional count data, Osborne et al. (2022) propose a method to simultaneously estimate network
interactions and associations to relevant covariates that employs a hierarchical model with latent layers and spike-and-slab priors for both edge and covariate selection. For posterior inference, the authors develop a novel hybrid scheme that combines VB steps on the covariate selection parameters with EM estimation of the latent variables that define the graph. We note that, generally speaking, optimization-based methods do not quantify the uncertainty associated with the point estimates they produce; in other words, they trade off accuracy of uncertainty quantification for scalability. Fully Bayesian approaches, despite their scalability limitation, are still by far the most popular Bayesian inference approach as a coherent, principled framework.

4 Topological ordering

In this section we provide a reply to some of the comments concerning topological ordering raised by Federico Castelletti, Guido Consonni and Luca La Rocca (CCLR).

For some of the methods described in our review paper, we assume a known ordering of the variables; as correctly pointed out by CCLR this is a crucial assumption that can be made only in the presence of substantive prior knowledge. CCLR wondered whether specifying an ordering becomes more problematic as the number of variables grows; fortunately, at least for the applications we dealt with, the ordering comes as whole piece of information (e.g., from known biological signaling pathways in genomics), that is either present or not. With a convincing simulation study, CCLR demonstrate that leaving the ordering unspecified leads to superior performance in terms of recovering the true underlying DAG relative to fixing an incorrect ordering; in practice, either a given ordering can be assumed with very good confidence or the node ordering should be considered a random variable. Moreover, CCLR discuss how to use DAGs for causal inference. Specifically, CCLR introduce an approach based on the interventional distribution of a DAG that can be used to estimate causal effects and discriminate between DAGs that belong to the same Markov equivalent class; this line of research, that includes important contributions from CCLR (Castelletti and Consonni 2019, 2021), is of great interest for the statistical community but goes beyond the scope of this review paper.

5 Heterogeneity

In this section we provide a reply to some of the comments concerning modeling approaches that can be used to account for heterogeneous observations raised by Federico Castelletti, Guido Consonni and Luca La Rocca (CCLR) and Yize Zhao, Zhe Sun and Jian Kang (ZSK).

CCLR raise several important points regarding covariate-dependent graphical models, wherein the covariates provide additional information regarding the heterogeneity in the samples. As they alluded to (in their Section 2) the covariates can enter the models either through: (i) mean structure; (ii) directly in the graphical structure or (iii) both. Scenario (iii) is an excellent point: whether it would be
possible to have covariates enter the graphical (regression) model through both mean and graph (i.e. precision matrix). It is indeed possible and meaningful in certain contexts. For instance, the graphical regression model reviewed in Section 4.1 can be extended as:

\[ y_{ij} = \sum_{j=1}^{p} g_j(x_t) + \sum_{k \in pa(j)} y_{ik} \theta_{jk}(x_t) + \epsilon_j, \]

where

- The first term: \( g_j(\cdot) \) is some function of the \( p \)-dimensional covariates \( (X) \);
- The second term: \( \theta_{jk}(\cdot) \) models the graphical structure as a function of covariates.

This extended model can be interpreted as a graphical model on both \( Y \)- and \( X \)-variables where \( X \)-variables are (potential) parents of \( Y \)-variables and their edge strengths do not change across observations whereas the edge strengths among \( Y \)-variables could vary with \( X \)-variables. The formulation of priors presented in Section 4.1 could be potentially extended for a coherent Bayesian specification of this model; however it does need some careful considerations regarding mean vs variance estimation trade-offs.

CCLR also raise an interesting future avenue of research regarding mixed/bidirected graphical models. Most methods in graphical regression, to the best of our knowledge, are (i) for continuous data and (ii) assume either an undirected (GGMs) or a directed (DAG) structure. Extension of these methods to discrete data can be done using latent variable models to model various types of discrete data e.g. binary, multi-category and count (or mixtures of these). However, this does raise some interpretational issues, since the graph is on a latent scale as opposed to the observed data scale (Bhadra et al. 2018). The bidirected graphical regression models case is an interesting one—since the conditioning would occur both at the edge and covariate level. The causal interpretations emerging from these graphs require some careful considerations given the interplay between the node level as well as edges level dependencies. Albeit, we do agree these are interesting avenues for future research.

CCLR as well as ZSK raise some interesting points regarding mixtures of graphs, i.e., presence of heterogeneous sub-groups/types in the data. While ZSK focus on the supervised case (where the subgroups are known), CCLR provide a nice summary on unsupervised learning of sub-groups as well as the graphical models. The former is a special case of the graphical regression models as presented in Section 4.1, where the covariate(s) are discrete. This essentially reduces the model to a multiple-graphical model, as nicely summarized by ZSK, albeit with an alternative characterization. The unsupervised case is an interesting one. Mixtures of graphs have been proposed in the literature using infinite mixtures mostly via Bayesian nonparameteric approaches (see CCLR and the references therein) or finite mixtures (e.g. Talluri et al. (2014)). However, we are not aware of any works that coherently allow mixtures to vary across multiple covariates and scales:
discrete, binary or combinations of these. While this would accompany substantial modeling and computational challenges, we envision there could be many potential applications, where multiple covariates (e.g. demographic, clinical, molecular) are assayed on the same individuals (e.g. see case studies in Ha et al. (2018); Bhattacharyya et al. (2020)), which can inform mixtures of graphs, either directed or undirected.

6 Interpretation

In this section we provide a reply to some of the comments concerning interpretation of posterior inference quantities raised by Alberto Roverato (AR) and David Marcano and Adrian Dobra (MD).

AR correctly points out that the interpretation of a graph becomes more problematic with increasing number of nodes; on the other hand, everything is harder in high dimensions and graphical models remain a tool that can help the investigators understand complex interaction patterns, even if proper conditional independence statements are more cumbersome to derive. More importantly, AR comments that inference of large graphs is typically unstable; this is partially true, since it depends on the sample size, strength of the signal, and true underlying graph structure. Moreover, we agree that selection of high-dimensional graphs is challenging and the resulting graph (if selected using marginal probability of edge inclusion, as often done in practice) does not fully retain the theoretical Markov properties; nevertheless, we think that graphical models are still useful exploratory tools that can guide follow-up analyses and experiments. Moreover, the Bayesian framework results in posterior probabilities, of both entire graphs and single edges, that can be interpreted as measure of uncertainty of the selected graphs.

We agree with MD that our review paper does not cover all important aspects related to inference and interpretation of graphical models. Specifically, the analysis and interpretation of all paths of dependencies becomes more and more cumbersome as the number of nodes and edges increases; some recent and relevant works in the context of concentration graphs includes Roverato and Castelo (2020). This type of analysis becomes even more complex for the graphical models reviewed in this paper, and we think that MD are correct in pointing out that this area of research should be of primary importance for both its methodological and practical relevance.

7 Models beyond Gaussianity

In this section we provide a reply to some of the comments concerning methods for non-Gaussian data raised by Luigi Augugliaro, Veronica Vinciotti and Ernst Wit (AVW) and Anindya Bhadra (AB).

We cannot agree more with AVW and AB that going beyond Gaussianity is a critical step towards robustifying graphical models in real-world applications where data are almost never perfectly Gaussian. There are indeed quite a few existing methods in this direction, e.g., via copula (Liu et al. 2009, 2012; Dobra and
Lenkoski 2011; Yoon et al. 2019; Chung et al. 2021) and via latent scaling (Finegold and Drton 2011; Bhadra et al. 2018; Cremaschi et al. 2019). We believe that both copula and latent scaling can be used to robustify multiple graphical models and graphical regression.

For multiple (undirected) graphical models (Peterson et al. 2015), it seems that copula or latent scaling can be directly applied to the sampling model or the likelihood in each group since the links between the groups are enforced only via the prior distribution. For (directed) graphical regression models, it is still possible even though the aforementioned copula and latent scaling methods were all developed for undirected graphs because conditional on covariates, the likelihood is a multivariate Gaussian distribution with a covariance/precision matrix induced from the graphical regression parameters.

We believe that the extension of multiple graphical models and graphical regression beyond Gaussianity could be excellent future exploration! One caveat worth mentioning is interpretability (which was one of the key discussion points raised by AR and MD from different perspectives) when moving from Gaussian to discrete data, for which the conditional independence interpretation may be lost. For example, with the Gaussian copula approach, the conditional independencies do not translate from the latent Gaussian variables to the observed discrete variables due to non-injective transformations.

8 Graph shrinkage priors and posterior concentration property

In this section we provide a reply to some of the comments concerning alternative shrinkage priors raised by Anindya Bhadra (AB).

Again, we thank AB for his stimulating and useful comments. We definitely agree with him that shrinkage priors are viable options that can be used as alternative to point mass mixture depending on the inferential goals, such as the case of estimation of precision matrices. Before going into the specific comments, we want to restate that many of the methods discussed in this review paper do not require to assume that the graph is decomposable, an assumption that indeed can be too restrictive for many applications. As correctly pointed out by AB, among shrinkage priors, the horseshoe prior has many desirable properties, and its successful application to graphical models is not surprising; we refer to the discussion by AB for a comprehensive list of references. Horseshoe priors can be certainly applied to graphical models for complex biological applications, both directed and undirected as well as chain graphs, a line of research that we are currently investigating.

9 Robustness of multiple graphical models

In this final section we provide a reply to some of the comments concerning robustness of the discussed methods raised by Yize Zhao, Zhe Sun and Jian Kang (ZSK).
ZSK investigate the performances of the MRF prior for multiple graph when the joint distribution of the edge selection indicators does not follow an MRF model or the similarity between graphs is not strong. As expected, the performances of the proposed fully Bayesian approach suffers from this type of prior-data conflicts. As ZSK pointed out, a future line of research should consist on the development and implementation of prior distributions that are flexible enough to model complex data generating mechanisms such as the ones investigated by ZSK. On a side note, simulation performances may be affected by many other factors; for example, when the number of edges increases, the strength of the connections becomes weaker (this is a side effect of the mechanism used to ensure that the true precision matrices are positive definite), and graph structure recovery becomes harder, regardless of the prior distributions used for the analysis.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

References

Arbel J, Battiston M, Favaro S, Lijoi A, Prünster I, Mena RH, Ni Y, Müller P (2017) Discussions of the paper sparse graphs using exchangeable random measures by F. Caron and E. Fox. arXiv preprint arXiv:170700842

Bhadra A, Rao A, Baladandayuthapani V (2018) Inferring network structure in non-normal and mixed discrete-continuous genomic data. Biometrics 74(1):185–195

Bhattacharyya R, Ha MJ, Liu Q, Akbani R, Liang H, Baladandayuthapani V (2020) Personalized network modeling of the pan-cancer patient and cell line interactome. JCO Clin Cancer Inf 4:399–411

Caron F, Fox EB (2017) Sparse graphs using exchangeable random measures. J R Stat Soc Ser B Stat Methodol 79(5):1295

Castelletti F, Consonni G (2019) Objective Bayes model selection of Gaussian interventional essential graphs for the identification of signaling pathways. Ann Appl Stat 13(4):2289–2311

Castelletti F, Consonni G (2021) Bayesian inference of causal effects from observational data in Gaussian graphical models. Biometrics 77:136–149

Chung HC, Gaynanova I, Ni Y (2021) Phylogenetically informed Bayesian truncated copula graphical models for microbial association networks. arXiv preprint arXiv:210505082

Cremaschi A, Argiento R, Shoemaker K, Peterson C, Vannucci M (2019) Hierarchical normalized completely random measures for robust graphical modeling. Bayesian Anal 14(4):1271–1301

Dobra A, Lenkoski A (2011) Copula gaussian graphical models and their application to modeling functional disability data. Ann Appl Stat 5(2A):969–993

Finegold M, Drton M (2011) Robust graphical modeling of gene networks using classical and alternative t-distributions. Ann Appl Stat 5:1057–1080
Ha MJ, Banerjee S, Akbani R, Liang H, Mills GB, Do KA, Baladandayuthapani V (2018) Personalized integrated network modeling of the cancer proteome atlas. Sci Rep 8(1):14924
Ha MJ, Stingo FC, Baladandayuthapani V (2020) Bayesian structure learning in multi-layered genomic networks. J Am Stat Assoc (forthcoming)
Kim J, Do KA, Ha MJ, Peterson CB (2019) Bayesian inference of hub nodes across multiple networks. Biometrics 75(1):172–182
Li ZR, McCormick TH (2019) An expectation conditional maximization approach for gaussian graphical models. J Comput Graph Stat 28(4):767–777
Liu H, Lafferty J, Wasserman L (2009) The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J Mach Learn Res 10(10):2295–2328
Liu H, Han F, Yuan M, Lafferty J, Wasserman L (2012) High-dimensional semiparametric Gaussian copula graphical models. Ann Stat 40(4):2293–2326
Mohan K, London P, Fazel M, Witten D, Lee SI (2014) Node-based learning of multiple Gaussian graphical models. J Mach Learn Res 15(1):445–488
Osborne N, Peterson CB, Vannucci M (2022) Latent network estimation and variable selection for compositional data via variational EM. J Comput Graph Stat 31(1):163–175
Peterson CB, Stingo F, Vannucci M (2015) Bayesian inference of multiple Gaussian graphical models. J Am Stat Assoc 110(509):159–174
Roverato A, Castelo R (2020) Path weights in concentration graphs. Biometrika 107(3):705–722
Talluri R, Baladandayuthapani V, Mallick BK (2014) Bayesian sparse graphical models and their mixtures. Stat 3(1):109–125
Wang H (2015) Scaling it up: stochastic search structure learning in graphical models. Bayesian Anal 10(2):351–377
Yang X, Gan L, Narisetty NN, Liang F (2021) Gembag: group estimation of multiple Bayesian graphical models. J Mach Learn Res 22:1–48
Yoon G, Gaynanova I, Müller CL (2019) Microbial networks in spring-semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. Front Genet 10:516

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

Authors and Affiliations

Yang Ni1 • Veerabhadran Baladandayuthapani2 • Marina Vannucci3 • Francesco C. Stingo4

Francesco C. Stingo
francescoclaudio.stingo@unifi.it

Yang Ni
yni@stat.tamu.edu

Veerabhadran Baladandayuthapani
veerab@umich.edu

Marina Vannucci
marina@rice.edu

1 Department of Statistics, Texas A&M University, College Station, USA
2 Department of Biostatistics, University of Michigan, Ann Arbor, USA
3 Department of Statistics, Rice University, Houston, USA
4 Department of Statistics, Computer Science, Applications “G. Parenti”, The University of Florence, Florence, Italy