The Annals of Applied Statistics
2017, Vol. 11, No. 4, 2222–2251
https://doi.org/10.1214/17-AOAS1076

BAYESIAN INFERENCE FOR MULTIPLE GAUSSIAN GRAPHICAL MODELS WITH APPLICATION TO METABOLIC ASSOCIATION NETWORKS

BY LINDA S. L. TAN[*,1], AJAY JASRA[*,2] MARIA DE IORIO[†] AND TIMOTHY M. D. EBBELS[‡,3]

National University of Singapore, University College London[†] and Imperial College London[‡]*

We investigate the effect of cadmium (a toxic environmental pollutant) on the correlation structure of a number of urinary metabolites using Gaussian graphical models (GGMs). The inferred metabolic associations can provide important information on the physiological state of a metabolic system and insights on complex metabolic relationships. Using the fitted GGMs, we construct differential networks, which highlight significant changes in metabolite interactions under different experimental conditions. The analysis of such metabolic association networks can reveal differences in the underlying biological reactions caused by cadmium exposure. We consider Bayesian inference and propose using the multiplicative (or Chung–Lu random graph) model as a prior on the graphical space. In the multiplicative model, each edge is chosen independently with probability equal to the product of the connectivities of the end nodes. This class of prior is parsimonious yet highly flexible; it can be used to encourage sparsity or graphs with a pre-specified degree distribution when such prior knowledge is available. We extend the multiplicative model to multiple GGMs linking the probability of edge inclusion through logistic regression and demonstrate how this leads to joint inference for multiple GGMs. A sequential Monte Carlo (SMC) algorithm is developed for estimating the posterior distribution of the graphs.

1. Introduction. Technological advances have enabled quantitative measurements and profiling of metabolites (products of metabolic reactions), which is important to the understanding of complex biological systems as well as the diagnosis and monitoring of disease states. A key feature of such data is that a significant number of metabolite levels are often highly interrelated. Analysis of these associations may provide further information about the physiological state of a system and lend insights on complex metabolic relationships [Steuer (2006)]. In this article, we analyze urinary metabolic data acquired using ^{1}H NMR spectroscopy for

Received April 2016; revised June 2017.
[1]Supported by the National University of Singapore Overseas Postdoctoral Fellowship.
[2]Supported by AcRF Tier 1 Grant R-155-000-156-112.
[3]Supported by the EU PhenoMeNal project (Horizon 2020, 654241).
Key words and phrases. Gaussian graphical models, prior specification, multiplicative model, sequential Monte Carlo.

127 individuals. These subjects live close to a lead and zinc smelter at Avonmouth in Bristol, United Kingdom, that produces large quantities of airbourne cadmium [Ellis et al. (2012)]. An extremely toxic metal, cadmium is commonly released through industrial processes and acute exposure poses numerous health risks. Here, we use Gaussian graphical models [GGMs, Dempster (1972)] to investigate the correlation structure of 22 urinary metabolites for each individual in response to cadmium exposure. Differential networks [Valcárcel et al. (2011)], which highlight significant changes in metabolite interactions under different experimental conditions, are inferred jointly with the GGM characterizing different levels of cadmium exposure. This is a strength of our modelling framework as it allows borrowing of strength across different biological conditions. Analysis of such metabolic association networks can point to differences in the underlying biological reactions caused by cadmium exposure.

Gaussian graphical models [GGMs, Dempster (1972)] provide an important tool for studying the dependence structure among a set of random variables. Under the assumption that the variables have a joint Gaussian distribution, a zero in the precision matrix indicates conditional independence between the associated variables. This corresponds to the absence of an edge in the underlying graph, where nodes denote variables and edges represent conditional dependencies [Lauritzen (1996)]. GGMs are widely used, for instance, in biological networks to study the dependence structure among genes from expression data [e.g., Dobra et al. (2004), Chun, Zhang and Zhao (2015)] and financial time series for forecasting and predictive portfolio analysis [e.g., Carvalho and West (2007), Wang, Reeson and Carvalho (2011)]. In applications where the effect of different experimental conditions on the dependence relationships among variables is of interest, multiple GGMs (one for each condition) have to be estimated. Under such circumstances, joint inference can encourage sharing of information across graphs and allow for common structure where appropriate [e.g., Guo et al. (2011), Peterson, Stingo and Vannucci (2015)].

We focus on Bayesian inference for GGMs using the G-Wishart prior [Roverato (2002), Atay-Kayis and Massam (2005)]. The G-Wishart is the family of conjugate distributions for the precision matrix, where entries corresponding to missing edges in the underlying graph are constrained to be zero. The normalizing constant of the G-Wishart can only be computed in closed form for decomposable graphs. In this work, we consider the unrestricted graph space where nondecomposable graphs are allowed. Where necessary, we use the Monte Carlo method of Atay-Kayis and Massam (2005) and the Laplace approximation of Lenkoski and Dobra (2011) to estimate the normalizing constant efficiently.

The main idea of this paper is to propose a prior for p_{ij}, the probability of a link between nodes i and j, that is grounded in the network literature. We start with one of the simplest random graph models; the multiplicative model where

$$p_{ij} = \pi_i \pi_j.$$

This model is additive on a log scale: $\log p_{ij} = \alpha_i + \alpha_j$ where $\alpha_i = \log \pi_i$. Alternatively, and without substantial difference in performance, we could have assumed a logistic/probit model. Incorporating interaction can be achieved by including an extra term. Extension to include more complex structures is in principle straightforward. For example, scale-free models can be achieved by placing a discrete prior on π_i such as the Barabási–Albert model so that the probability that node i has k connections is proportional to $A + k^{\alpha}$. To incorporate a community structure, we could assume

$$\operatorname{logit}(p_{ij}) = \alpha_i + \alpha_j + \theta_{g_i g_j},$$

where g_i denotes the community i belongs to and $\theta_{g_i g_j}$ denotes an offset for node i and j belonging to the same community, with $\theta_{g_i g_j}$ equal to 0 otherwise.

Here, we propose using the multiplicative model of Chung and Lu (2002) as a prior on graphs for estimating GGMs. This prior is further extended to multiple GGMs via logistic regression. To obtain joint posterior inference for multiple GGMs, we develop a novel sequential Monte Carlo (SMC) algorithm [Del Moral, Doucet and Jasra (2006)] which uses tempering techniques. We apply proposed methods to simulated data in addition to the urinary metabolic dataset.

The rest of the paper is organized as follows. Section 2 provides the background and review of existing methods. In Section 3, we introduce the multiplicative model and discuss its degree and clustering properties. Section 4 specifies the model setup for multiple GGMs. Section 5 describes posterior inference and a Laplace approximation for the prior probabilities of graphs. Proposed methods are illustrated using simulations and an application to urinary metabolic data in Section 6. In Section 7, we illustrate the proposed methods using simulations and an application to urinary metabolic data. Conclusions are stated in Section 8.

2. Background. In the absence of any prior belief on the graphical structure, a uniform prior over all graphs is often used in estimating GGMs [e.g., Lenkoski and Dobra (2011), Wang and Li (2012)]. That is, given p nodes, it is assumed that each of the 2^r possible graphs, where $r = p(p-1)/2$, has equal probability of arising. This prior concentrates its mass on graphs with moderately large number of edges and the expected number of edges as well as the mode is $r/2$ (see Figure 1). Thus, this prior may not be appropriate when sparse graphs are desired. Several alternatives have been developed. To encourage sparse graphs, Dobra et al. (2004) and Jones et al. (2005) propose a prior where every edge is included independently with a small probability α so that a graph with x edges has prior probability $\alpha^x(1-\alpha)^{r-x}$. This prior is known as the Erdős–Rényi model in random graph theory and it reduces to the uniform prior when $\alpha = 0.5$. Jones et al. (2005) recommend taking $\alpha = 2/(p-1)$ so that the expected number of edges is p. Carvalho and Scott (2009) treat α as a model parameter rather than a fixed tuning constant. They place a Beta(a, b) prior on α so that a graph with x edges has prior probability $B(a+x, r+b-x)/B(a, b)$, where $B(a, b)$ denotes the Beta

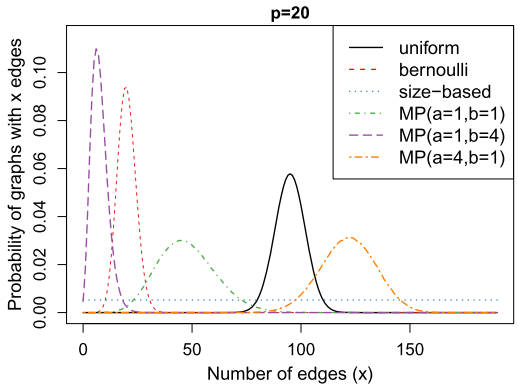

FIG. 1. *Plot shows the probability allocated to graphs with x edges by the uniform prior, the Bernoulli prior* [*Jones et al.* (2005)] *with probability of inclusion of each edge: $\alpha = 2/(p-1)$, the size-based prior* [*Armstrong et al.* (2009)] *or equivalently the Bernoulli prior with $\alpha \sim$ Uniform[0, 1] integrated out* [*Carvalho and Scott* (2009)] *and the multiplicative prior (MP) for different values of a and b.*

function. When $a = b = 1$, this probability simplifies to $\frac{1}{r+1}\binom{r}{x}^{-1}$. This prior is equivalent to the size-based prior [Armstrong et al. (2009)] when the graph space is unrestricted. Under the size-based prior, every size $0, \ldots, r$, has equal probability and every graph of the same size has equal probability. Carvalho and Scott (2009) demonstrate that their proposed prior has strong control over the number of spurious edges and corrects for multiple hypothesis testing automatically, where each null hypothesis corresponds to the exclusion of one edge.

We propose using the multiplicative or Chung–Lu random graph model as a prior on the graphical space of GGMs. Given a desired or expected degree sequence $\{d_1, \ldots, d_p\}$, where d_i denotes the degree (number of neighbours) of node i, the multiplicative model [Chung and Lu (2002)] assumes that the edge between each pair of nodes i and j is formed independently with probability p_{ij} proportional to the product $d_i d_j$. Allowing self-loops and provided $(\max_i d_i)^2 < \sum_i d_i$, the expected degree of node i is exactly d_i. The multiplicative model is able to capture degree distributions which are more diverse (e.g., right-skewed, U-shaped) and closer to that of real-world networks than the Erdős–Rényi model. Notably, the Erdős–Rényi model has a degree distribution that is binomial and can be viewed as a special case of the multiplicative model with a constant expected degree sequence. We consider an alternative parametrization of the multiplicative model introduced by Olhede and Wolfe (2013), which dispenses with self-loops and the normalization constraint by taking $p_{ij} = \pi_i \pi_j$ and $0 < \pi_i < 1$ for each i. They derive degree characteristics and large-sample approximations of this model, which lends insight on the variation attainable in degree structure. Adopting a Bayesian approach, we treat each π_i as a variable with a Beta(a, b) prior. We present degree and clustering properties of the multiplicative model, showing how they depend on

choices of a and b. In the context of GGMs, we show that the multiplicative model provides an avenue to encourage sparsity or graphs that exhibit particular degree patterns based on prior knowledge obtained through expert opinion or past data. We further demonstrate how the multiplicative model can be extended to include covariates and become a prior on joint graphs for multiple GGMs.

Several approaches for joint inference of multiple GGMs have been developed recently. Guo et al. (2011) estimate precision matrices for different groups jointly by parameterizing each entry as a product of two factors: one factor is common across all groups while the other is group specific. A hierarchical l_1 penalty is imposed and optimization is performed using graphical lasso [Friedman, Hastie and Tibshirani (2008)]. Danaher, Wang and Witten (2014) formulate a more general framework called joint graphical lasso and introduce two convex penalty functions; the fused graphical lasso which encourages edge value on top of structural similarity and the group lasso which only encourages a shared sparsity pattern. Chun, Zhang and Zhao (2015) extend the approach of Guo et al. (2011) to a wider class of nonconvex penalty functions. Mohan et al. (2014) consider a node-based approach where multiple GGMs are estimated using a convex regularizer by assuming that the similarities between networks are due to the shared presence of certain highly-connected nodes, which serve as hubs and the differences are due to some nodes whose connectivity changes across conditions. Yajima et al. (2015) compare the Gaussian directed acyclic graphs of two subgroups using Bayesian inference via Gibbs sampling. The strength of association between two variables in the differential group is modeled as the strength in the baseline group plus an edge-specific parameter controlling the difference in association between the two subgroups. They define a prior on the graphical space by centering on a prior graph constructed from a database [Telesca et al. (2012)]. Peterson, Stingo and Vannucci (2015) consider an alternative Bayesian approach, which links graphs from different groups using a Markov random field. The probability of inclusion of each edge in graphs $1, \ldots, K$, is parameterized in terms of a $K \times K$ symmetric matrix which measures the pairwise similarity of groups and is common across all edges, and an edge-specific $K \times 1$ vector, which controls the inclusion probability in each group independently of group relatedness. Priors are further placed on these parameters and a block Gibbs sampler is used for inference.

The approach that we use to link multiple graphs is based on the multiplicative model by expressing the connectivity of each node as a logistic regression function of graph specific covariates. As the multiplicative model decouples the inclusion probability of each edge into the product of the connectivities of the end nodes, the resulting model is parsimonious and scales linearly in the number of variables and graphs. For inference, we develop an SMC sampler for estimating the joint posterior distribution of the graphs. Using tempering techniques [see Del Moral, Doucet and Jasra (2006) and the references therein], we create a sequence of probability distributions from which to sample, moving gradually from a distribution that is easy to sample from, through artificial intermediate distributions towards the posterior distribution of interest.

3. Multiplicative model. Here, we define our notation and present the multiplicative model, followed by a study of its properties. These properties lend insight on the structure and range of networks that can be generated from the multiplicative model. They are also useful in the determination of suitable hyperparameters based on prior understanding of data obtained through expert opinion or a database.

Let $G = (V, E)$ be a simple graph with vertex set $V = \{1, 2, \ldots, p\}$ and edge set $E \subseteq \{(i, j) \in V \times V : i < j\}$. A simple graph is undirected and does not contain self-loops or multiple edges. The adjacency matrix $A = [A_{ij}]$ of G is a $p \times p$ binary matrix where A_{ij} is 1 if an edge is present between nodes i and j, and 0 otherwise for $i, j \in V$. As G is simple, A is symmetric and has zeros on its diagonal.

In the multiplicative model, each edge is modeled independently as

$$\begin{aligned} A_{ij} &\sim \text{Bernoulli}(p_{ij}) \qquad \text{for } 1 \le i < j \le p, \\ p_{ij} &= \pi_i \pi_j \qquad \text{where } 0 \le \pi_i \le 1 \text{ for } i = 1, \ldots, p. \end{aligned} \tag{3.1}$$

Thus, every edge A_{ij} is formed independently with probability p_{ij}, where p_{ij} is a product of the tendencies of nodes i and j to form edges with other nodes. The parameter π_i is characteristic of node i and reflects its activity level. We refer to π_i as the *connectivity* of i. The Erdős–Rényi random graph model arises as a special case when π_i is constant across all i, that is, every link is formed independently with equal probability.

We adopt a Bayesian approach and place an independent Beta prior on each π_i. Let

$$\pi_i \sim \text{Beta}(a, b) \qquad \text{for } i = 1, \ldots, p, \tag{3.2}$$

where $a, b > 0$. We have $p(\pi_i) = \pi_i^{a-1}(1 - \pi_i)^{b-1}/B(a, b)$, where the Beta function $B(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$. Let $\pi = (\pi_1, \ldots, \pi_p)^T$ and $p(\pi) = \prod_{i=1}^{p} p(\pi_i)$. Networks of highly varying densities and structures can be formed by choosing different hyperparameters a and b.

3.1. *Degree and clustering properties.* The degree D_i of a node i is the number of links that involve i or the number of neighbours of i, and is given by $D_i = \sum_{j \ne i} A_{ij}$. The properties below describe the degree distribution and cohesiveness of networks generated from the multiplicative model. Their implications are discussed later in the section. We follow the framework in Rastelli, Friel and Raftery (2015), which is based on probability generating functions [see Newman, Strogatz and Watts (2001)]. Proofs are given in the Supplementary Material [Tan et al. (2017)]. We note that some of these results have been discussed in Olhede and Wolfe (2013) but not with regards to the Beta(a, b) prior. In the following, let $\mu = \frac{a}{a+b}$ and $\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)}$ denote the mean and variance of a Beta(a, b) distribution, respectively.

P1 The probability that a randomly chosen node is a neighbour of a node with connectivity π_i is $\mu\pi_i$.

P2 The degree of a node with connectivity π_i is distributed as Binomial$(p-1, \mu\pi_i)$. Hence its average degree is $(p-1)\mu\pi_i$, which is proportional to π_i.

P3 The probability generating function of the degree of a randomly chosen node is given by

$$G_{D_i}(z) = \sum_{d=1}^{p-1} \mathrm{P}(D_i = d) z^d = \int_0^1 (1 - \mu\pi_i + \mu\pi_i z)^{p-1} p(\pi_i)\, d\pi_i. \tag{3.3}$$

The kth factorial moment of D_i, $\mathrm{E}\{D_i(D_i-1)\cdots(D_i-k+1)\}$, is given by

$$\mathrm{E}\left(\frac{D_i!}{(D_i-k)!}\right) = G_{D_i}^{(k)}(1) = \frac{(p-1)!B(a+k,b)}{(p-1-k)!B(a,b)}\mu^k$$

for any positive integer k.

P4 The average degree of a randomly chosen node is $\mathrm{E}(D_i) = (p-1)\mu^2$ and the variance is $\mathrm{Var}(D_i) = (p-1)\mu^2\{1-\mu^2+(p-2)\sigma^2\}$.

P5 The degree distribution of a randomly chosen node is given by

$$P(D_i = d) = \binom{p-1}{d} \frac{\mu^d}{B(a,b)} \int_0^1 \pi_i^{a+d-1}(1-\mu\pi_i)^{p-1-d}(1-\pi_i)^{b-1} d\pi_i$$

for $d \in \{0, \ldots, p-1\}$. When $b = 1$, $P(D_i = d) = \binom{p-1}{d} aB(\mu, a+d, p-d)/\mu^a$, where $B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}\, dt$ is the incomplete Beta function.

P6 The dispersion index of the degree distribution is given by

$$1 - \frac{a\{a^2 + (b+1)a - (p-2)b\}}{(a+b)^2(a+b+1)}.$$

- When $0 < a < \{\sqrt{b^2 + (4p-6)p + 1} - b - 1\}/2$, the distribution has dispersion index greater than 1 and is over-dispersed.
- When $a = \{\sqrt{b^2 + (4p-6)p + 1} - b - 1\}/2$, the distribution has dispersion index 1 (equal to that of a Poisson distribution).
- When $a > \{\sqrt{b^2 + (4p-6)p + 1} - b - 1\}/2$, the distribution has dispersion index less than 1 (similar to that of a Binomial distribution) and is under-dispersed.

P7 The skewness index or Pearson's moment coefficient of skewness of the degree distribution is $\{E(D_i^3) - 3E(D_i)\,\mathrm{Var}(D_i) - E(D_i)^3\}/\mathrm{Var}(D_i)^{1.5}$, where $\mathrm{E}(D_i^3) = (p-1)\mu^2\{1 + 3(p-2)(\mu^2+\sigma^2) + (p-2)(p-3)\mu E(\pi_i^3)\}$ and $E(\pi_i^3) = \frac{(a+2)(a+1)a}{(a+b+2)(a+b+1)(a+b)}$.

P8 The average degree of the neighbours of a node is independent of the connectivity or degree of that node, and is given by $1 + (p-2)(\mu^2+\sigma^2)$.

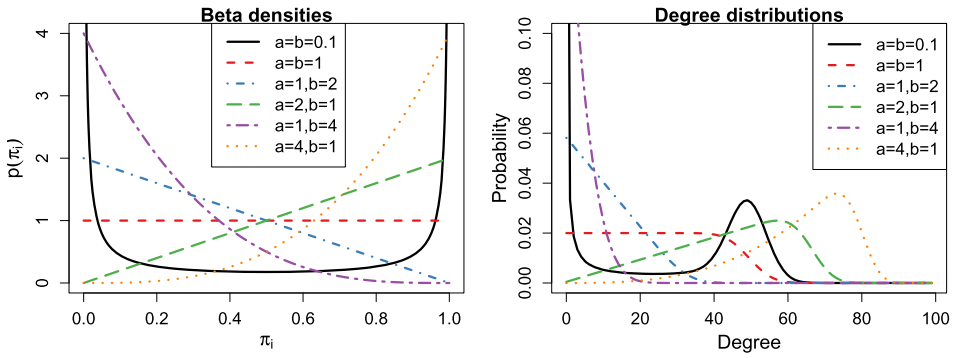

FIG. 2. *Beta densities* (*left*) *and degree distributions of the multiplicative model* (*right*) *corresponding to different hyperparameter settings when* $p = 100$.

P9 The global clustering coefficient, which measures the probability that nodes j and k are linked given that both nodes are linked to i, is given by $\frac{a+1}{a+b+1}$.

In the Erdős–Rényi model, D_i is distributed as Binomial$(p-1, \alpha)$, where α is the probability of inclusion of each edge. When $\alpha = \mu^2$, the mean degree of a randomly chosen node in the multiplicative model is equal to that in the Erdős–Rényi model from P4. However, the variance of the degree distribution in the multiplicative model is greater than that in the Erdős–Rényi model by $(p-1)(p-2)\sigma^2$. Thus, as the number of nodes increases, the multiplicative model can accommodate greater variation in the degree distribution than in the Erdős–Rényi model by $\mathcal{O}(p^2)$ given the same mean degree.

Figure 2 shows the degree distributions of the multiplicative model for graphs with $p = 100$ nodes under different hyperparameter settings. When the degree distributions cannot be computed directly using P5, they are estimated via simulation using 10^5 graphs. Degree distributions of a variety of shapes (e.g., right-skewed, U-shaped) can be obtained from the multiplicative model by varying a and b.

The dispersion index measures how clustered a distribution is compared to standard statistical models. From P6, the degree distribution is over-dispersed when a is small and under-dispersed when a is large. In fact, as $a \to \infty$ (and/or $b \to \infty$), $\sigma^2 \to 0$ and each π_i reduces to a point mass. The multiplicative model thus degenerates and reduces to the Erdős–Rényi model with constant probability of inclusion for every edge. As the degree distribution is over-dispersed for a wide range of hyperparameter values, the multiplicative model is able to represent well heterogeneity in degree sequences.

The skewness index in P7 is useful for identifying asymmetries in degree distributions. Generally, the degree distribution is positively skewed when a is small and b is large and negatively skewed vice versa (plots of the dispersion index and skewness as a function of a and b can be found in the Supplementary Material Figures S1 and S2).

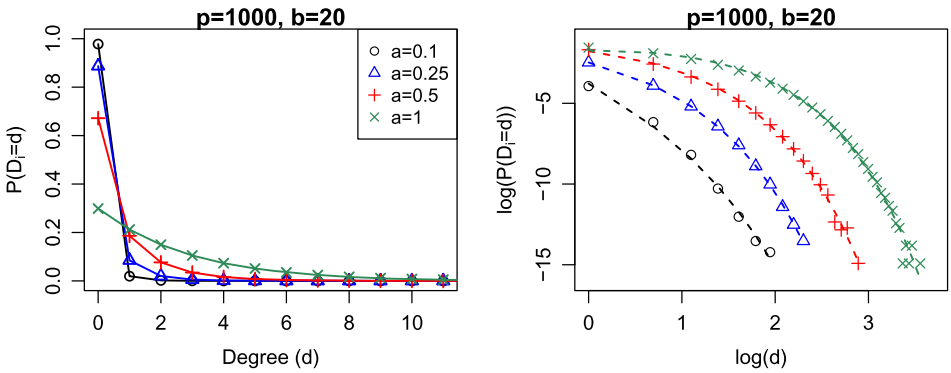

FIG. 3. *Degree distributions for investigating power-law. Right plot is in log-scale.*

Of particular interest are scale-free networks whose degree distributions follow a power law [$P(D_i = d) \propto d^{-\gamma}$ where γ is a positive constant]. Olhede and Wolfe (2013) show that the multiplicative model can lead to networks with power law degree distributions when p is large, the elements of π are ordered such that $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_p$ and a polynomial decay of π_i with i is assumed; $\pi_i \propto i^{-\gamma}$ for $0 < \gamma < 1$. We investigate via simulations the behavior of the degree distributions when $\{\pi_i\}$ is modeled instead as a random sample from a Beta(a, b) prior. As scale-free networks tend to have large positive values for the skewness index, we consider a large $b = 20$ and some small values of $a \in \{0.1, 0.25, 0.5, 1\}$. The left plot in Figure 3 shows the degree distributions obtained via simulation and the right plot shows the relationship between $\log P(D_i = d)$ and $\log d$, which should be a straight line if the power-law is satisfied. We observe that the multiplicative model (with a Beta prior) comes close to but does not quite induce power law networks as the right tail is not sufficiently heavy. However, we find that these points are well-fitted by a power law with an exponential cutoff [$P(D_i = d) \propto d^{-\gamma} \exp(-\tau d)$, Newman (2001)]. Fits of these form are shown as dotted lines in the right plot of Figure 3. In such networks, the power law dominates for small d but turns into an exponential decay for large d. A broad range of empirical data such as protein interaction networks [Jeong et al. (2001), Giot et al. (2003)] and scientific collaboration networks [Fenner, Levene and Loizou (2007)] have been found to exhibit power-laws with exponential cutoffs instead of pure power laws due to finite-size effects such as the physical limitation of the binding sites in the protein structure and the finite working lifetime of a scientist. D'Souza et al. (2007) provides further examples.

4. Gaussian graphical models. Suppose we have a dataset with p variables and K groups or classes. Let $y_h = (y_{h1}, \ldots, y_{hp})$ denote the hth observation of the p variables for $h = 1, \ldots, H$, and S_k be an index set containing the indices of

observations, which belong to group k for $k = 1, \ldots, K$. The number of observations in S_k is denoted by $|S_k|$ and $H = \sum_{k=1}^{K} |S_k|$. Without loss of generality, we assume that the observations in each group are centered at 0 along each variable. We consider

$$y_h|\Omega_k \sim N\big(0, \Omega_k^{-1}\big) \qquad \text{for } h \in S_k, \tag{4.1}$$

where Ω_k is a $p \times p$ precision matrix and $k \in \{1, \ldots, K\}$.

Let $G_k = (V, E_k)$ be a simple graph with vertex set $V = \{1, 2, \ldots, p\}$ and edge set $E_k \subseteq \{(i, j) \in V \times V : i < j\}$ for $k = 1, \ldots, K$. Node $i \in V$ represents the ith variable and each edge $(i, j) \in E_k$ corresponds to $\Omega_{k,ij} \neq 0$. That is, y_{hi} and y_{hj} are conditionally independent (in G_k) given the rest of the elements in y_h if and only if $\Omega_{k,ij} = 0$, or equivalently $(i, j) \notin E_k$. The conjugate prior for Ω_k is the G-Wishart distribution [Atay-Kayis and Massam (2005)], $W_{G_k}(\delta_k, D_k)$, which has density

$$p(\Omega_k|G_k) = \frac{1}{I_{G_k}(\delta_k, D_k)} |\Omega_k|^{(\delta_k - 2)/2} \exp\left\{-\frac{1}{2} \operatorname{tr}(\Omega_k D_k)\right\}.$$

Here, Ω_k is constrained to the cone P_{G_k} of positive definite matrices with entries equal to zero for all $(i, j) \notin E_k$ and $I_{G_k}(\delta_k, D_k)$ is a normalizing constant such that

$$I_{G_k}(\delta_k, D_k) = \int_{\Omega_k \in P_{G_k}} |\Omega_k|^{(\delta_k - 2)/2} \exp\left\{-\frac{1}{2} \operatorname{tr}(\Omega_k D_k)\right\} dK.$$

This normalizing constant is guaranteed to be finite if $\delta_k > 2$ and $D_k^{-1} \in P_{G_k}$ [Diaconis and Ylvisaker (1979)]. The G-Wishart distribution reduces to the Wishart distribution when G_k is complete, and the normalizing constant can then be evaluated in closed form as

$$I_{G_k}(\delta_k, D_k) = 2^{(\delta_k + p - 1)p/2} \Gamma_p\big\{(\delta_k + p - 1)/2\big\} |D_k|^{-(\delta_k + p - 1)/2}, \tag{4.2}$$

where $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=0}^{p-1} \Gamma(a - \frac{1}{2})$ for $a > (p - 1)/2$.

4.1. *Priors over graphs.* We use the multiplicative model to assign prior probabilities to graphs. Let $A_k = [A_{k,ij}]$ be the adjacency matrix of G_k for $k = 1, \ldots, K$. Consider

$$A_{k,ij}|\pi_{k,i}\pi_{k,j} \sim \text{Bernoulli}(\pi_{k,i}\pi_{k,j}), \tag{4.3}$$

where $0 \leq \pi_{k,i} \leq 1$ for $i = 1, \ldots, p$ and $k = 1, \ldots, K$. As in Section 3, the probability that an edge (i, j) is present in E_k is $\pi_{k,i}\pi_{k,j}$, the product of the propensities of nodes i and j to form edges with other nodes in G_k. Priors are further placed on each $\pi_{k,i}$. We consider the cases $K = 1$ and $K > 1$ separately.

4.1.1. *When* $K = 1$. When $K = 1$, there is only one group and the subscript k indicating different groups may be dropped so that $G_1 = G$ and $\pi_{1,i} = \pi_i$ for $i = 1, \ldots, p$. We place a Beta(a, b) prior on each π_i as in (3.2). The prior probability of G with adjacency matrix A is given by

$$\begin{aligned} &p(G|a,b) \\ &= \int p(G|\pi)p(\pi|a,b)\,d\pi \\ &= \frac{1}{B(a,b)^p}\int \prod_{i,j:i<j}(1-\pi_i\pi_j)^{(1-A_{ij})}\prod_{i=1}^{p}\pi_i^{(a+d_i-1)}(1-\pi_i)^{(b-1)}\,d\pi, \end{aligned} \tag{4.4}$$

where d_i denotes the degree of node i.

4.1.2. *When* $K > 1$. We propose a joint prior for $G_1, \ldots, G_K$, which is allowed to depend on covariates specific to each graph. First, we express $\pi_{k,i}$ in terms of a logistic regression as

$$\pi_{k,i} = \frac{\exp(\beta_i^T x_k)}{1+\exp(\beta_i^T x_k)}$$

for $i = 1, \ldots, p$, and $k = 1, \ldots, K$, where $x_k = (x_{k1}, \ldots, x_{kQ})^T$ is a vector of covariates for G_k and $\beta_i = (\beta_{i1}, \ldots, \beta_{iQ})$ is a vector of coefficients specific to node i. Let $x = (x_1, \ldots, x_K)$ and $\beta = (\beta_1^T, \ldots, \beta_p^T)^T$. We consider a normal prior for each β_{iq} such that

$$\beta_{iq}|\sigma_q^2 \sim N(0, \sigma_q^2)$$

for $i = 1, \ldots, p$ and $q = 1, \ldots, Q$. Let $\sigma^2 = (\sigma_1^2, \ldots, \sigma_Q^2)$ be a hyperparameter assumed to be known. The joint prior probability of $G_1, \ldots, G_K$, is given by

$$\begin{aligned} &p(G_1, \ldots, G_K|x, \sigma^2) \\ &= \int p(\beta|\sigma^2)\prod_{k=1}^{K} p(G_k|x_k, \beta)\,d\beta \\ &= \int \prod_{i=1}^{p}\prod_{q=1}^{Q}\left\{\frac{\exp(-\frac{\beta_{iq}^2}{2\sigma_q^2})}{\sqrt{(2\pi\sigma_q^2)}}\right\}\prod_{k=1}^{K}\left\{\prod_{i=1}^{p}\pi_{k,i}^{d_{k,i}}\prod_{i<j}(1-\pi_{k,i}\pi_{k,j})^{1-A_{k,ij}}\right\}d\beta, \end{aligned} \tag{4.5}$$

where $d_{k,i}$ denotes the degree of node i in G_k for $k = 1, \ldots, K$.

As an example, in the application on urinary metabolic data, we consider $K = 2$ and the covariates x_k to include an intercept and an indicator for level of exposure to cadmium (1 if above the median and 0 otherwise). We take $x_1 = (1, 0)$ and $x_2 = (1, 1)$ so that G_1 and G_2 represent the correlation structure of the groups

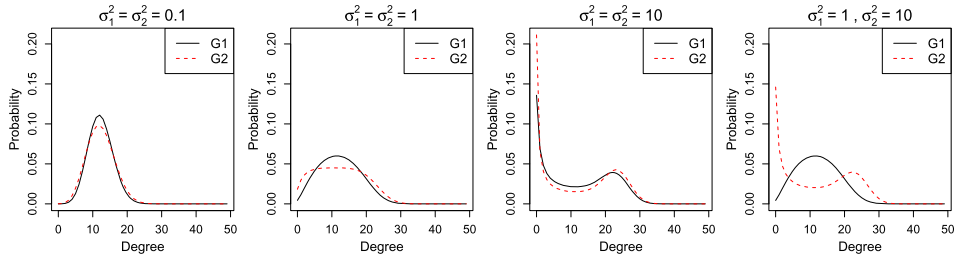

FIG. 4. *Prior degree distributions of G_1 and G_2 for $p = 50$ and different combinations of σ_1^2 and σ_2^2. Covariates for G_1 and G_2 are $(1, 0)$ and $(1, 1)$, respectively.*

with cadmium exposure below or equal to the median, and above the median, respectively. The connectivity of node i is $\pi_{1,i} = \{1 + \exp(-\beta_{i1})\}^{-1}$ in G_1 and $\pi_{2,i} = \{1 + \exp(-\beta_{i1} - \beta_{i2})\}^{-1}$ in G_2. Thus β_{i1} determines the popularity of node i in G_1 while β_{i2} is a differential parameter, which controls the difference in popularity of node i between G_1 and G_2. If β_{i2} is close to zero, the connectivity of node i in G_1 and G_2 is similar. As the magnitude of β_{i2} increases, the difference in connectivity of node i between G_1 and G_2 becomes greater. See illustration in Supplementary Material Figure S3.

Figure 4 shows the prior degree distributions of G_1 and G_2 for $p = 50$ and different values of σ^2. These plots are obtained via simulation of 10^5 joint pairs of graphs in each case. When $\sigma_1^2 = \sigma_2^2 = 0.1$, both β_{i1} and β_{i2} are close to zero, and $\pi_{1,i}$ and $\pi_{2,i}$ are close to 0.5. Thus the degree distribution is shaped like a Binomial curve, resembling the Erdős–Rényi model where each edge is formed independently with constant probability 0.25. As σ_1^2 increases, there is greater variation in the degree sequence of G_1. When σ_1^2 is large, the connectivity of each node tends to the extremes of 0 and 1 (each node has a high probability of being either very connected or isolated). Thus the degree distribution resembles the case where each π_i is allocated a U-shaped Beta$(0.1, 0.1)$ prior as shown in Figure 2. The distinction between the degree distribution of G_1 and G_2 becomes greater as σ_2^2 increases.

We can also add a third covariate say for gender (1 if male and 0 if female) so that $K = 4$ and take $x_1 = (1, 0, 0)$, $x_2 = (1, 0, 1)$, $x_3 = (1, 1, 0)$ and $x_4 = (1, 1, 1)$. Then G_1, for instance, will represent the correlation structure for the group of females with cadmium exposure below or equal to the median level.

5. Posterior distribution. Let $y = (y_1, \ldots, y_H)$. For $K > 1$, the joint distribution of the model is

$$p(y, \Omega_1, \ldots, \Omega_K, G_1, \ldots, G_K, \beta | x, \sigma^2)$$
$$= p(\beta|\sigma^2) \prod_{k=1}^{K} \Big\{ p(G_k|x_k, \beta) p(\Omega_k|G_k) \prod_{h \in S_k} p(y_h|\Omega_k) \Big\}.$$

Integrating out Ω_k, the marginal likelihood $p(\{y_h|h \in S_k\}|G_k)$ can be shown [see, e.g., Atay-Kayis and Massam (2005)] to be given by

$$p(\{y_h|h \in S_k\}|G_k) = (2\pi)^{-p|S_k|/2} I_{G_k}\left(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T\right) \Big/ I_{G_k}(\delta_k, D_k).$$

Integrating out β as well, we have

$$\begin{aligned} &p(G_1, \ldots, G_K|y, x, \sigma^2) \\ &\quad \propto p(G_1, \ldots, G_K|x, \sigma^2) \\ &\qquad \times \prod_{k=1}^{K} I_{G_k}\left(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T\right) \Big/ I_{G_k}(\delta_k, D_k). \end{aligned} \tag{5.1}$$

When $K = 1$, the posterior distribution can be derived similarly. The only difference is that the dependence on x and σ^2 is replaced by the Beta prior hyperparameters a and b. We have

$$p(G|y, a, b) \propto p(G|a, b) I_G\left(\delta + H, D_k + \sum_{h=1}^{H} y_h y_h^T\right) \Big/ I_G(\delta, D). \tag{5.2}$$

For posterior inference, we propose a SMC algorithm to obtain samples from the posterior distribution. To compute the right-hand side of (5.1) and (5.2), note that for any graph G (not necessarily decomposable), normalizing constants of the form $I_G(\delta, D)$ can be evaluated by first factorizing G into its prime components and their separators [see, e.g., Lauritzen (1996)]. Suppose $(\mathcal{P}_1, \ldots, \mathcal{P}_L)$ is a perfect sequence of the prime components of G and $(\mathcal{S}_2, \ldots, \mathcal{S}_L)$ is the corresponding set of separators. Then $I_G(\delta, D) = \prod_{l=1}^{L} I_{G_{\mathcal{P}_l}}(\delta, D) / \prod_{l=2}^{L} I_{G_{\mathcal{S}_l}}(\delta, D)$, where $G_{\mathcal{P}_l}$ and $G_{\mathcal{S}_l}$ denote the subgraphs induced by $\mathcal{P}_l$ and $\mathcal{S}_l$, respectively. As the separators are complete, $I_{G_{\mathcal{S}_l}}(\delta, D)$ can be evaluated as in (4.2). The same applies to $I_{G_{\mathcal{P}_l}}(\delta, D)$ for any complete prime component $\mathcal{P}_l$. Otherwise, we estimate $I_{G_{\mathcal{P}_l}}(\delta, D)$ using the Monte Carlo method of Atay-Kayis and Massam (2005) when δ is small and the Laplace approximation of Lenkoski and Dobra (2011) when δ is large. This combination of using Laplace approximation and Monte Carlo integration to evaluate the normalizing constants is feasible as the size of the graphs considered in this paper is moderately small ($p \leq 22$). When p is large, the size of the Monte Carlo sample has to be increased dramatically to control the variance and Monte Carlo integration becomes a computational bottleneck [see Jones et al. (2005), Wang and Li (2012)]. At this point, techniques that avoid evaluation of prior normalizing constants (and that explore the space of graphs and precision matrices jointly) based on for instance, the exchange algorithm [Murray, Ghahramani and MacKay (2006)] have to be integrated with SMC sampler. The priors on graphs are estimated using Laplace approximation, which is described next.

5.1. *Laplace approximation for prior on graphs.* Evaluating $p(G|a,b)$ or $p(G_1,\dots,G_K|x,\sigma^2)$ via Monte Carlo becomes more computationally intensive as p increases and we estimate these quantities efficiently using Laplace approximation. We consider

$$\int \exp\{f(u)\}\,du \approx (2\pi)^{\frac{n}{2}}\left|-H(u_0)\right|^{-\frac{1}{2}}\exp\{f(u_0)\}, \tag{5.3}$$

where $u=(u_1,\dots,u_d)^T$, u_0 is the mode of $f(u)$ and $H(u_0)$ denotes the Hessian of f evaluated at u_0. The mode u_0 can be found using numerical methods.

For $K=1$, we estimate $p(G|a,b)$ in (4.4) using (5.3) by first making a change of variable and letting $\pi_i=\frac{\exp(u_i)}{1+\exp(u_i)}$. For $K>1$, we estimate $p(G_1,\dots,G_K|x,\sigma^2)$ in (4.5) using (5.3) by taking $u=\beta$. Detailed functional and Hessian expressions are given in the Supplementary Material.

6. Sequential Monte Carlo sampler. We use SMC samplers for posterior inference. Suppose we are interested in sampling from a complex target $\lambda(x)$. The idea is to start with some distribution λ_1 that is easy to sample from and move via a sequence of intermediate distributions, $\lambda_2,\dots,\lambda_{T-1}$, towards the distribution of interest $\lambda_T=\lambda$. At any time t, a large collection of weighted samples $\{W_t^{(n)},X_t^{(n)}|n=1,\dots,N\}$ is maintained, and these particles are used to generate samples from the target distribution at the next time point using sequential importance sampling (SIS) and resampling methods. The motivation is that it would be easier to move the particles from one target to the next if λ_t is close to λ_{t+1}.

6.1. *Review of methodology.* We first review SIS and SMC briefly. Let $\lambda_1,\dots,\lambda_T$, be the target densities, $\gamma_1,\dots,\gamma_T$, be unnormalized densities such that $\lambda_t(x_{1:t})\propto\gamma_t(x_{1:t})$ and η_t be an arbitrary proposal density for $t=1,\dots,T$. In importance sampling, the unnormalized weights are given by

$$w_t(x_{1:t})=\gamma_t(x_{1:t})/\eta_t(x_{1:t}). \tag{6.1}$$

Let $\{X_{1:t}^{(n)}|n=1,\dots,N\}$ be a sample from $\eta_t(x_{1:t})$ and $w_t^{(n)}=w_t(X_{1:t}^{(n)})$. Then

$$W_t^{(n)}=w_t^{(n)}\Big/\sum_{n=1}^{N}w_t^{(n)} \tag{6.2}$$

are the normalized weights. Given $\{W_{1:t}^{(n)},X_{1:t}^{(n)}|n=1,\dots,N\}$ approximating $\lambda_t(x_{1:t})$ at time t, samples $\{X_{1:t+1}^{(n)}\}$ approximating λ_{t+1} at time $t+1$ are obtained in SIS by sampling from $\{X_{1:t}^{(n)}\}$ using a Markov kernel $K_{t+1}(x_t,x_{t+1})$. The proposal density is $\eta_{t+1}(x_{1:t+1})=\eta_t(x_{1:t})K_{t+1}(x_t,x_{t+1})$. From (6.1), the corresponding unnormalized weights can be computed recursively using

$$w_{t+1}(x_{1:t+1})=\frac{\gamma_{t+1}(x_{1:t+1})}{\eta_{t+1}(x_{1:t+1})}=\frac{\gamma_{t+1}(x_{1:t+1})}{\gamma_t(x_t)K_{t+1}(x_t,x_{t+1})}w_t(x_{1:t}).$$

In SMC, artificial joint target distributions $\tilde{\lambda}_t(x_{1:t}) \propto \tilde{\gamma}_t(x_{1:t})$ are introduced, where $\tilde{\gamma}_t(x_{1:t}) = \gamma_t(x_t)\prod_{l=1}^{t-1} L_l(x_{l+1}, x_l)$ and $L_l(x_{l+1}, x_l)$ is an artificial backward in time Markov kernel. Assume $\{W_{1:t}^{(n)}, X_{1:t}^{(n)} | n = 1, \ldots, N\}$ is a weighted sample approximating $\tilde{\lambda}_t(x_{1:t})$ at time t distributed according to $\eta(x_{1:t})$. Moving the samples to $\{X_{1:t+1}^{(n)}\}$ using the Markov kernel K_{t+1}, the unnormalized importance weights can be computed as

$$w_{t+1}(x_{1:t+1}) = \tilde{\gamma}_{t+1}(x_{1:t+1})/\eta_{t+1}(x_{1:t+1}) = w_t(x_{1:t})\tilde{w}_{t+1}(x_t, x_{t+1}), \tag{6.3}$$

where $\tilde{w}_{t+1}(x_t, x_{t+1}) = \gamma_{t+1}(x_{t+1})L_t(x_{t+1}, x_t)/\{\gamma_t(x_t)K_{t+1}(x_t, x_{t+1})\}$ are unnormalized incremental weights. In the proposed algorithm, we take K_{t+1} to be an MCMC kernel of invariant distribution λ_{t+1} and $L_t(x_{t+1}, x_t) = \lambda_{t+1}(x_t) \times K_{t+1}(x_t, x_{t+1})/\lambda_{t+1}(x_{t+1})$. See Del Moral, Doucet and Jasra (2006), Section 3.3.2.3. The unnormalized incremental weights then simplify to

$$\tilde{w}_{t+1}(x_t, x_{t+1}) = \gamma_{t+1}(x_t)/\gamma_t(x_t). \tag{6.4}$$

6.2. *Proposed algorithm.* Our aim is to sample from $p(G_1, \ldots, G_K | y, x, \sigma^2)$ in (5.1) when $K > 1$ and $p(G|y, a, b)$ in (5.2) when $K = 1$. Let $p(G_1, \ldots, G_K|y)$ denote the posterior density generally omitting dependence on covariates and hyperparameters. We have $p(G_1, \ldots, G_K|y) \propto \gamma(G_1, \ldots, G_K|y)$ where

$$\gamma(G_1, \ldots, G_K) = \begin{cases} p(G|a, b) I_G\left(\delta + H, D + \sum_{h=1}^{H} y_h y_h^T\right) \Big/ I_G(\delta, D) \\ \qquad \text{if } K = 1, \\ p(G_1, \ldots, G_K|x, \sigma^2) \prod_{k=1}^{K} \dfrac{I_{G_k}(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T)}{I_{G_k}(\delta_k, D_k)} \\ \qquad \text{if } K > 1. \end{cases}$$

For simplicity, we do not state the dependence of γ on other variables explicitly. To construct the SMC sampler, we devise the following sequence of intermediate target densities:

$$p(G_1, \ldots, G_K|y)^{\phi_1}, p(G_1, \ldots, G_K|y)^{\phi_2}, \ldots, p(G_1, \ldots, G_K|y)^{\phi_T},$$

where $0 < \phi_1 < \phi_2 < \cdots < \phi_T = 1$ is a sequence of user-specified temperatures that can be set adaptively [see Jasra et al. (2011)]. For greater stability, we use tempering to bridge the target densities so that they evolve smoothly. At each time t, we maintain N weighted samples $\{W_t^{(n)}, (G_1, \ldots, G_K)_t^{(n)} | n = 1, \ldots, N\}$ approximating the target $p(G_1, \ldots, G_K|y)^{\phi_t} \propto \gamma(G_1, \ldots, G_K)^{\phi_t}$ and the annealing temperature is raised gradually from 0 to 1.

6.3. *Initialization and computation of weights*. To generate N samples from the initial target $p(G_1, \ldots, G_K|y)^{\phi_1}$ at time $t = 1$, we sample $(G_1, \ldots, G_K)$ uniformly from the joint graphical space. This can be done by sampling each edge in G_k independently with probability 0.5 for $k = 1, \ldots, K$. This process is performed N times independently to obtain $\{(G_1, \ldots, G_K)_1^{(n)} | n = 1, \ldots, N\}$. The weight of each sample can be computed using importance sampling. Let $r = p(p-1)/2$. From (6.1),

$$w_1^{(n)} = \gamma\big((G_1, \ldots, G_K)_1^{(n)}\big)^{\phi_1} 2^{rK}. \tag{6.5}$$

Suppose we increase the temperature from ϕ_{t-1} to ϕ_t at time $t \geq 2$. From (6.3) and (6.4), unnormalized weights for the nth sample can be computed as

$$w_t^{(n)} = w_{(t-1)}^{(n)} \gamma\big((G_1, \ldots, G_K)_{t-1}^{(n)}\big)^{\phi_t - \phi_{t-1}}. \tag{6.6}$$

Normalized weights may be obtained using (6.2).

6.4. *Resampling*. To prevent degeneracy of the particle approximation, we measure the effective sample size, $\mathrm{ESS} = \{\sum_{n=1}^{N} (W_t^{(n)})^2\}^{-1}$, at each time t and resample when the ESS falls below a threshold, say $N_{\text{threshold}} = N/3$. Resampling is performed by drawing N new particles from the multinomial distribution with parameters $(W_t^{(1)}, \ldots, W_t^{(N)})$. In this way, particles with high weights are duplicated multiple times while particles with low weights will be eliminated. Resampled particles are then assigned equal weights.

6.5. *MCMC steps*. Suppose we have weighted samples $\{W_{t-1}^{(n)}, (G_1, \ldots, G_K)_{t-1}^{(n)} | n = 1, \ldots, N\}$. At time t, these samples are moved using an MCMC kernel with invariant distribution $p_t(G_1, \ldots, G_K|y)$ by performing a small number of MCMC steps. This improves mixing and helps to restore the heterogeneity lost during resampling. In this step, candidates for each sample are generated by selecting a small number, say M, of edges uniformly at random from the set of all possible edges and proposing to flip each edge [a 1 (present) to 0 (absent) and vice versa] in turn in each G_k for $k = 1, \ldots, K$. For the selected edge, let $(G_1, \ldots, G_K)_c^{(n)}$ denote the sample obtained after flipping this edge in one of the K graphs in $(G_1, \ldots, G_K)_{t-1}^{(n)}$. As the proposal is symmetric, the candidate is accepted with probability given by

$$\min\big[\big\{\gamma\big((G_1, \ldots, G_K)_c^{(n)}\big)/\gamma\big((G_1, \ldots, G_K)_t^{(n)}\big)\big\}^{\phi_t}, 1\big]. \tag{6.7}$$

If the candidate is accepted, we update the nth sample as $(G_1, \ldots, G_K)_t^{(n)} = (G_1, \ldots, G_K)_c^{(n)}$, otherwise it remains unchanged. The proposed SMC sampler is summarized in Algorithm 1. Note that Algorithm 1 is easily parallelizable as computation of weights as well as the MCMC steps can be performed independently for the N samples.

Algorithm 1 SMC Algorithm for multiple GGMs

At $t = 1$,

- draw $(G_1, \ldots, G_K)_1^{(n)}$ at random uniformly from the joint graphical space for $n = 1, \ldots, N$.
- Compute weights $\{w_1^{(n)}\}$ using (6.5) and obtain normalized weights $\{W_1^{(n)}\}$ using (6.2).

For $t = 2, \ldots, T$,

- Update weights $\{w_t^{(n)}\}$ using (6.6) and obtain normalized weights $\{W_t^{(n)}\}$ using (6.2).
- If ESS $< N_{\text{threshold}}$, resample the particles and set $W_t^{(n)} = 1/N$ for $n = 1, \ldots, N$.
- For $n = 1, \ldots, N$,
  - Randomly select M edges from the set of all possible edges.
  - Set $(G_1, \ldots, G_K)_t^{(n)} = (G_1, \ldots, G_K)_{t-1}^{(n)}$.
  - For $m = 1, \ldots, M$, and $k = 1, \ldots, K$, let $(G_1, \ldots, G_K)_c^{(n)}$ be the sample candidate obtained from $(G_1, \ldots, G_K)_t^{(n)}$ by flipping the mth selected edge in G_k. Accept sample candidate with probability in (6.7). If sample candidate is accepted, set $(G_1, \ldots, G_K)_t^{(n)} = (G_1, \ldots, G_K)_c^{(n)}$.

Algorithm 1 can be fully automated to the extent that one only needs to set the first temperature and MCMC proposal. The rest of the algorithm, such as in [Del Moral, Doucet and Jasra (2012), Jasra et al. (2011), Schäfer and Chopin (2013)] can be made entirely adaptive with stable and mathematically proven convergence [Beskos et al. (2016)].

6.6. *Scalability.* The proposed algorithm scales linearly in K due to the MCMC steps, which have to be performed for each graph. The algorithm does not scale well with respect to the number of nodes p as the computation of the normalizing constants $I_{G_k}(\delta_k, D_k)$ using Monte Carlo integration [Atay-Kayis and Massam (2005)] is computationally expensive when p is large (scales approximately as the cube of p).

7. Results. We discuss the results obtained from simulations and application of the proposed GGM to the urinary metabolic data.

To set the hyperparameters for the multiplicative prior, we suggest using prior data or R packages such as GeneNet [Schaefer, Opgen-Rhein and Strimmer (2015)] or GGMselect [Giraud, Huet and Verzelen (2012)] to obtain a quick preliminary estimate of the degree distribution. The hyperparameters of the multiplicative prior can then be selected so that the shape of the prior degree distribution matches the estimated one. For $K = 1$, one can compute (a, b) using the formulas in P4 so that the mean and variance of the prior degree distribution matches

the estimated one. This procedure is implemented in Section 7.1. Note that the estimated degree distribution may have a variance smaller than that in P4 for any $a > 0$, $b > 0$. In this case, we set b to be large (e.g., 1000) so that the variance is very small and then find a to match the mean degree.

In the following experiments, we take $\delta_k = 3$ and $D_k = I_p$ for $k = 1, \ldots, K$ in the G-Wishart prior. The number of samples used in SMC is $N = 500$. Our code is written in Matlab and is available as part of the Supplementary Material. We run the experiments on HPC (High Performance Computing) where each job is run in parallel across 12 cores.

7.1. *Simulations*. We investigate the performance of the multiplicative model as a prior on graphs for GGMs and compare it with the commonly used uniform prior which assigns equal prior probability to every graph, and the size-based prior [Armstrong et al. (2009)] or equivalently the prior that corrects for multiple hypothesis testing proposed by Carvalho and West (2007) (see Section 2 for details). First, we consider $K = 1$, $p = 20$ nodes and generate data from three different types of networks:

Multiplicative model: Generate $\pi_i \overset{\text{i.i.d.}}{\sim} \text{Beta}(0.1, 0.1)$ and simulate the edges using $A_{ij} \sim \text{Bernoulli}(\pi_i \pi_j)$ for $i < j$.

Scale-free network: A scale-free network with p nodes is generated using the Barabási–Albert model. Starting with a connected network with 2 nodes, new nodes are added one at a time to the network. Each new node is connected to 2 existing nodes with a probability proportional to the degree of existing nodes.

Community network: The p nodes are divided into two communities of equal-size and a network is generated by assuming that the within-community interaction rate is 0.6 and across-community interaction rate to be 0.02.

The generated networks and their degree distributions are shown in Figure 5. For each network we create a $p \times p$ symmetric matrix C where entries corresponding to missing edges are set to zero and nonzero entries are simulated randomly from the uniform distribution on $[-0.6, -0.3] \cup [0.3, 0.6]$. To ensure that the precision matrix Ω is positive, we let c be the smallest eigenvalue of C and set $\Omega = C + (0.1 + |c|)I$, following Mohan et al. (2014). Ten datasets are simulated from the GGM in (4.1) by setting the number of observations $H = 100$ and $K = 1$. The underlying network is regarded as the "true" graph.

Using Algorithm 1, weighted samples from the posterior distribution are obtained for each simulated dataset under the uniform, size-based and multiplicative priors. For the multiplicative model, we consider two settings. For one setting, we set the Beta hyperparameters as $a = b = 1$. For the second setting, we try to find a and b such that the shape of the prior degree distribution resembles that of the true graph. These prior degree distributions are superimposed on the true degree distributions in Figure 5. For the SMC sampler, we set the number of edges

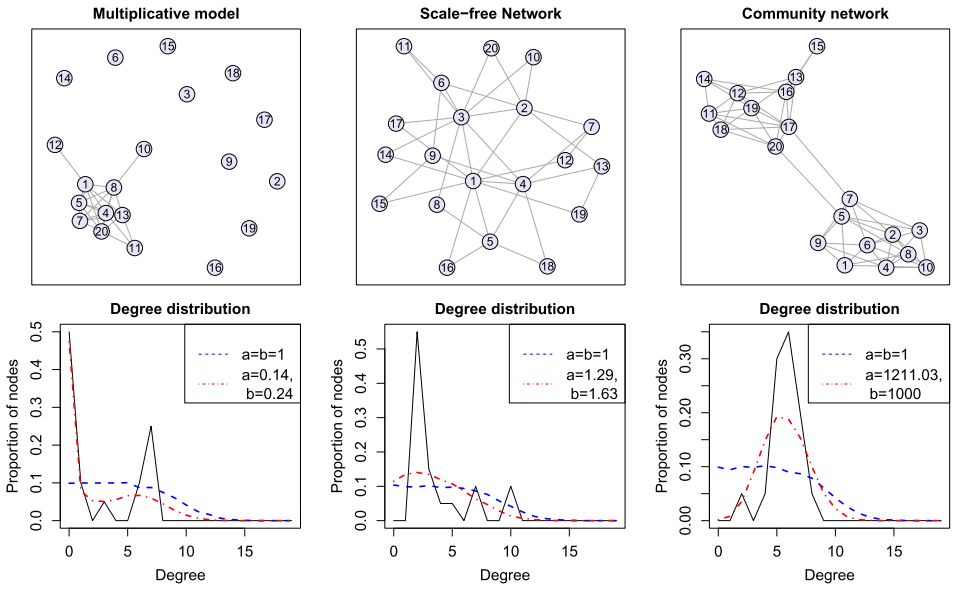

FIG. 5. *Generated networks (top) and corresponding degree distribution (bottom). The degree distributions of the multiplicative model with hyperparameters given by a and b are superimposed.*

flipped at each iteration in the MCMC step, $M = 3$. The sequence $\{\phi_t\}$ is set as $(0.01, 0.02, \ldots, 1)$ with $T = 100$. The CPU time taken on average for each experiment is (6.4 ± 0.5) hours.

Using the weighted samples, we compute the posterior probability of the occurrence of each edge and summarize the results using the area under the ROC curves (AUC). The boxplots of the AUC values are shown in Figure 6. The multiplicative priors performed better than the uniform and especially the size-based prior for data simulated from the multiplicative model. While one may have expected the prior with $a = 0.14$, $b = 0.24$ to perform better than $a = b = 1$, this may not have happened because the proposed method of setting the hyperparameters uses only the mean and variance and may not have captured the true degree distribution well enough. In addition, the relationship between prior probability of a graph and its degree distribution is not straightforward. Note that variability is also involved in the generation of the precision matrix which may represent the underlying true graph better in some cases than others. For data simulated from the scale-free and community networks, the performance of the different priors are quite similar. For these networks, the multiplicative prior performed better if the hyperparameters were tuned to match the degree distribution of the true graph.

Next, we investigate the ability of the multiplicative prior to borrow information across graphs when the nodes have similar connectivity. We simulate 10 datasets each with $H = 100$ observations, $p = 20$ variables and $K = 2$ groups. We assume that there are 50 observations in each group and set the covariates $x_1 = (1, 0)$ and

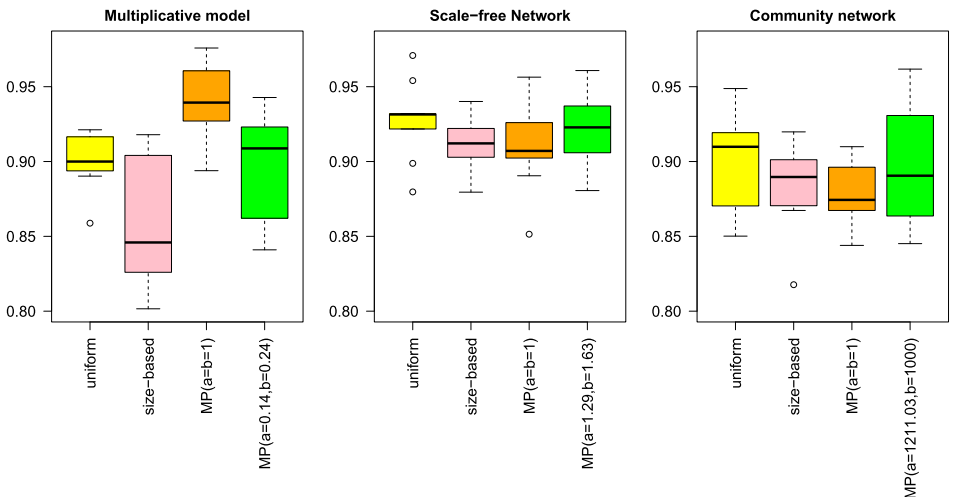

FIG. 6. *Boxplots of AUC values for datasets simulated from different networks obtained using different priors.*

$x_2 = (1, 1)$. The underlying graphs are generated from the multiplicative model where the connectivities are simulated using the model in Section 4.1.2 and precision matrices for each graph are constructed in the same manner as before. Setting $\sigma_1^2 = 10$ and $\sigma_2^2 = 0.01$, the connectivity of the nodes in G_1 may vary over a wide range (since β_{i1} has a large variance) and the connectivity of each node in G_1 and G_2 are similar (since β_{i2} is close to zero). Figure 7 shows the simulated graphs and their degree distributions.

We compare results obtained using (1) the uniform prior which assigns equal prior probability to each pair of graphs, (2) the joint multiplicative prior with $\sigma_1^2 = 10$, $\sigma_2^2 = 0.01$ and (3) independent multiplicative priors for G_1 and G_2 with hyperparameters chosen to match the degree distributions of the true graphs. Using Algorithm 1 with the same setting as before, the average CPU time for the joint prior ($K = 2$) case is (12.4 ± 0.5) hours and for the independent multiplicative priors case ($K = 1$) is (6.5 ± 0.4) hours. The results are summarized using boxplots of the AUC values as shown in Figure 7. The joint multiplicative prior performs better than the uniform prior and the independent multiplicative priors indicating the ability of the multiplicative prior to encourage similarity in connectivity of nodes across graphs.

In the Supplementary Material, we also provide details of a small experiment which shows the significant improvement that SMC provides over standard MCMC. In particular, SMC has a higher acceptance rate and achieves higher average log target density for the same number of MCMC steps.

7.2. *Application to urinary metabolic data.* We analyze urinary metabolic data for $H = 127$ individuals acquired using ^{1}H NMR spectroscopy [see Ellis et al.

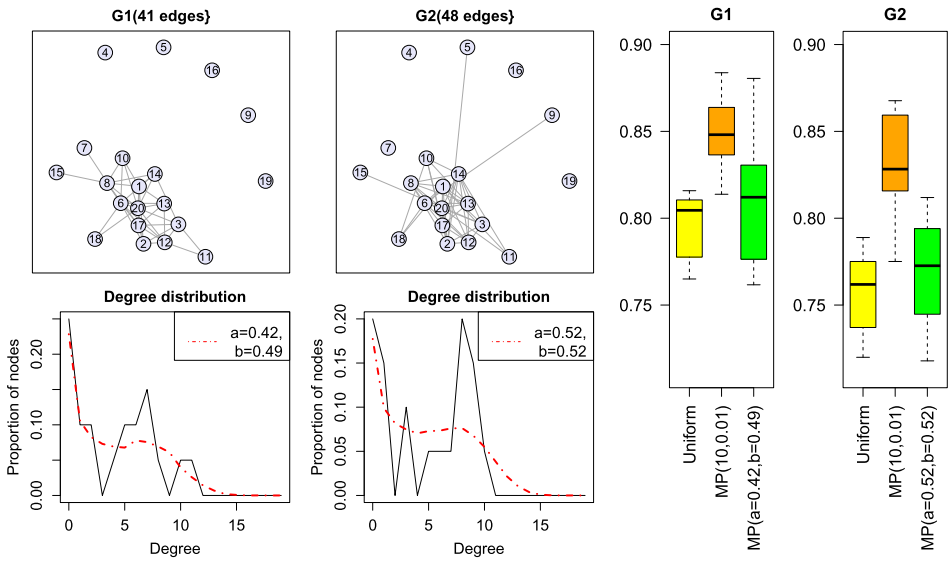

FIG. 7. *Simulated networks* ($K = 2$), *corresponding degree distributions and boxplots of AUC values obtained using different priors.*

(2012) for details]. These individuals live close to a lead and zinc smelter at Avonmouth in Bristol, United Kingdom, which produces large quantities of airborne cadmium (Cd). Here, we investigate the correlation structure of $p = 22$ urinary metabolites listed in Table 1 in response to cadmium exposure through GGMs. This dataset has also been studied by Salamanca, Ebbels and De Iorio (2014) using Bayesian hierarchical models. We perform two analyses. First, we consider the individuals as a homogeneous group. In the second case, we divide the individuals into two groups; S_1 (a control group with cadmium exposure lower than or equal to the median) and S_2 (a diseased group with cadmium exposure higher than the median). In each case, we first use the `R` package `GeneNet` [Schaefer, Opgen-Rhein and Strimmer (2015)] to obtain fast shrinkage estimators of partial correlation in the network. The degree distributions obtained (see Figure 8) can be used as a basis for determining appropriate hyperparameters for the multiplicative model. The observations of each variable are first normalized to have zero mean and standard deviation of one in each group. For the SMC sampler, we set the number of samples $N = 500$, and the number of edges flipped at each iteration in the MCMC step, $M = 5$. The sequence $\{\phi_t\}$ is set as $(0.005, 0.01, \ldots, 1)$ with $T = 200$. The CPU time taken on average for each experiment is (24.7 ± 3.0) hours for $K = 1$ and (48.0 ± 7.5) hours for $K = 2$.

7.3. *Case*: $K = 1$. We study the correlation structure of the metabolites by treating the individuals as one homogeneous group. We compare the performance of four priors on the graphical space. For the multiplicative prior, we obtain

TABLE 1

List of 22 *urinary metabolites and their abbreviations. Columns* 3–6 *and columns* 7–10 *show the weighted mean degree and betweenness, respectively, under the multiplicative model with* $a = 461$, $b = 1000$ [$MP(461, 1000)$] *and* $a = b = 1$ [$MP(1, 1)$], *the size-based prior* (*SB*) *and the uniform prior* (*UF*). *The highest value in each column* (3–10) *is highlighted in bold*

		Degree				Betweenness			
Metabolites	**Abbrev**	***MP* (461, 1000)**	***MP* (1, 1)**	**SB**	**UF**	***MP* (461, 1000)**	***MP* (1, 1)**	**SB**	**UF**
Trimethylamine oxide	TMAO	**4.11**	**4.52**	**4.20**	5.33	**0.21**	**0.15**	**0.20**	0.08
Dimethylamine	DMA	3.47	3.81	3.47	5.68	0.14	0.10	0.12	0.07
P-cresol-sulphate	PCS	3.37	4.33	3.72	**6.39**	0.09	0.10	0.11	**0.09**
Succinate	Suc	3.33	3.95	3.79	5.71	0.16	0.10	0.17	0.07
Creatinine	Creat	2.88	3.54	3.09	4.93	0.08	0.08	0.09	0.06
3-hydroxyisovalerate	3-HV	2.75	2.85	2.74	4.58	0.19	0.10	0.17	0.06
Citrate	Cit	2.60	3.15	2.42	4.84	0.15	0.09	0.12	0.05
Glycine	Gly	2.60	2.70	2.41	4.42	0.13	0.07	0.10	0.05
4-deoxyerythronic acid	4-DEA	2.59	3.52	2.63	5.23	0.12	0.09	0.11	0.08
Pyruvate	Pyr	2.56	3.21	3.01	5.41	0.06	0.05	0.06	0.06
Alanine	Ala	2.29	2.48	2.38	5.45	0.11	0.06	0.10	0.07
Urea	Urea	2.09	2.62	2.09	6.17	0.08	0.07	0.06	0.09
Phenylacetylglutamine	PAG	2.05	2.24	2.16	4.55	0.06	0.02	0.05	0.04
Hippurate	Hip	1.63	1.55	1.76	4.86	0.05	0.02	0.04	0.06
Dimethylgycine	DMG	1.46	1.49	1.44	4.70	0.05	0.02	0.04	0.06
Trimethylamine	TMA	1.25	1.28	1.60	3.28	0.02	0.01	0.04	0.03
Acetate	AcO	1.07	1.93	1.32	5.13	0.02	0.04	0.04	0.05
Lactate	Lac	0.90	1.08	0.96	4.56	0.02	0.02	0.02	0.04
Proline-betaine	PB	0.87	0.64	0.84	2.49	0.01	0.00	0.01	0.01
N-methyl-nicotinic acid	NMNA	0.70	0.56	1.06	3.06	0.01	0.01	0.01	0.02
Formate	For	0.54	0.36	0.43	2.85	0.01	0.00	0.01	0.02
Creatine	Crea	0.25	0.12	0.28	1.94	0.00	0.00	0.00	0.01

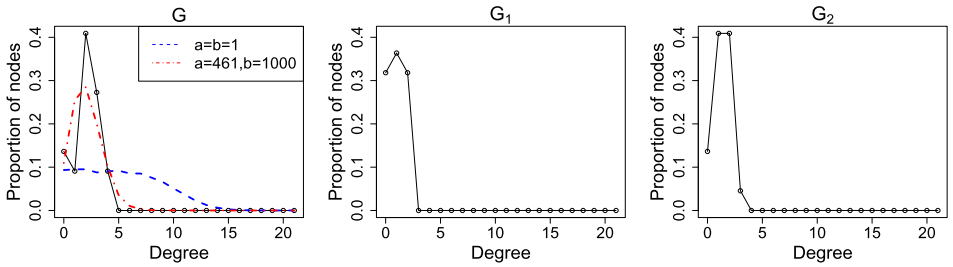

FIG. 8. *Degree distributions estimated using GeneNet for the case where the individuals are treated as one heterogeneous group* (*left*) *and where they are divided into two groups* S_1 (*middle*) *and* S_2 (*right*).

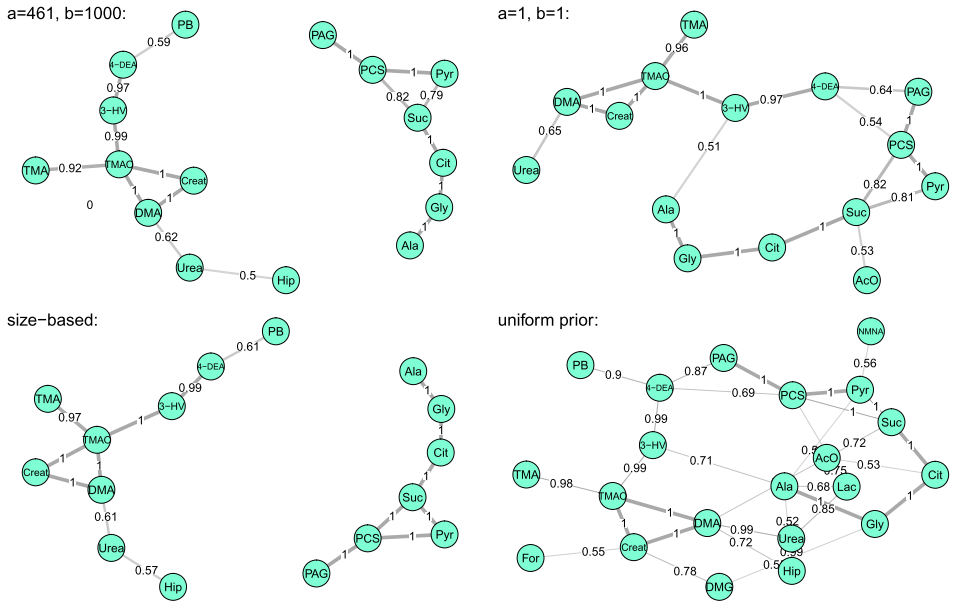

FIG. 9. *Graphs corresponding to different priors. Only edges with posterior weights greater than* 0.5 *are shown.*

$a = 461$, $b = 1000$ by matching the degree distribution from GeneNet. Note that such a prior is highly informative. We also consider for comparison a vague multiplicative prior with $a = b = 1$, the size-based prior and the uniform prior. We fit a GGM to the data using Algorithm 1, obtaining $N = 500$ weighted samples from the posterior distribution in each case. Using these weighted samples, we compute the posterior probability of occurrence of each edge. Figure 9 shows the graphs obtained under each prior. Only edges with posterior probability greater than 0.5 and associated nodes are shown and the width of each edge is proportional to its posterior probability. Graphs showing the full set of nodes and all possible edges are given in the Supplementary Material Figure S5. The graphs obtained under the multiplicative priors and the size-based prior have a high degree of similarity (especially $a = 461$, $b = 1000$) and are much sparser than that of the uniform prior.

Table 1 shows the weighted mean degree and betweenness centrality measures for each metabolite. The metabolites have been sorted in terms of weighted mean degree in decreasing order according to $MP(461, 1000)$. Under the multiplicative and size-based priors, TMAO has the highest degree as well as betweenness. Under the uniform prior, PCS has the highest degree with Urea close behind; these two metabolites also have the highest betweenness.

For the multiplicative model, we can obtain uncertainty measures of the tendency of each node to form connections with other nodes. Figure 10 shows the posterior distributions of the connectivities π_i of each metabolite obtained via simulations. The multiplicative prior with $a = 461$, $b = 1000$ is very informative and

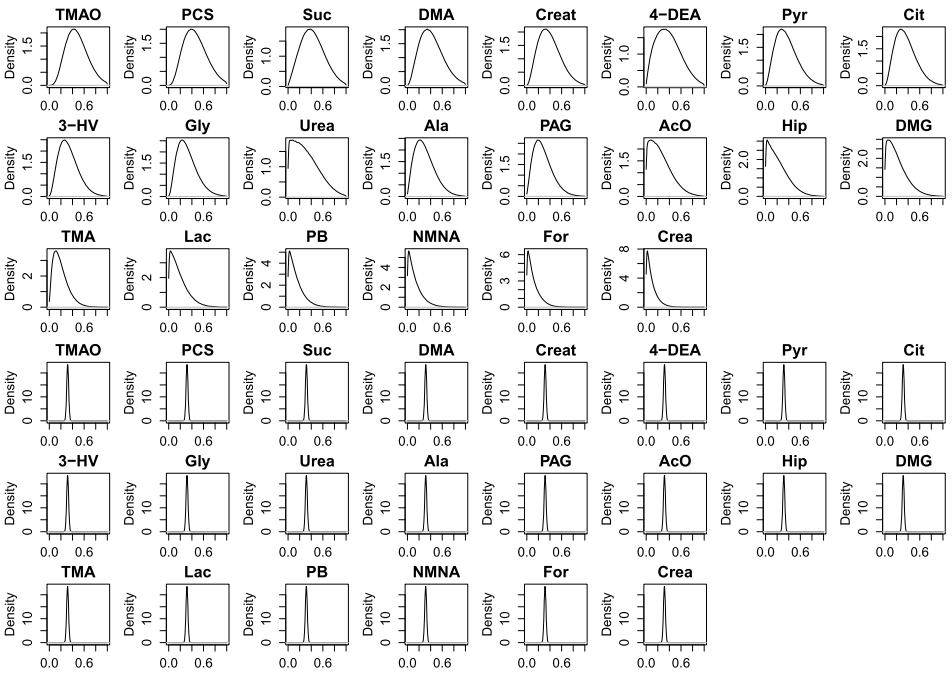

FIG. 10. *Posterior distribution of the connectivity* (π_i) *of different metabolites under the multiplicative model with* $a = b = 1$ (*first* 3 *rows*) *and* $a = 461$, $b = 1000$ (*last* 3 *rows*).

forces the posterior distribution of the connectivity of every node to be almost identical. The multiplicative prior with $a = b = 1$ allows the data to determine the connectivity and there is more distinction among nodes. The mean and 95% credible interval of the connectivity of each metabolite, and the mean covariance matrix corresponding to the multiplicative model with $a = b = 1$ are given in the Supplementary Material Tables S1 and S2.

7.4. *Case*: $K = 2$. Next, we investigate the difference in correlation structure of the urinary metabolites between the two groups of individuals S_1 (with cadmium exposure lower than or equal to the median) and S_2 (cadmium exposure higher than the median). We consider the covariates x_k for the kth graph to include an intercept and an indicator for level of exposure to cadmium (1 if above the median and 0 otherwise) so that $x_1 = (1, 0)$ and $x_2 = (1, 1)$. The difference in graphical structure between G_1 and G_2 due to exposure to urinary cadmium is of interest. We fit a GGM with $K = 2$ to the data using the SMC algorithm under four priors. The first three are the multiplicative model with $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 10$ and $\sigma_1^2 = \sigma_2^2 = 10$, and the last is the uniform prior. From Figure 4 and the preliminary degree distributions obtained using GeneNet (Figure 8), taking $\sigma_1^2 = \sigma_2^2 = 1$ seems appropriate but we wish to investigate if there is any benefit to be

gained by allowing the structure of G_2 to vary more significantly from that of G_1 by taking σ_2^2 to be 10 and whether a prior, which assumes the tendencies to connect are closer to the extremes of 0 and 1 is more appropriate ($\sigma_1^2 = \sigma_2^2 = 10$).

Using Algorithm 1, we obtain weighted samples from the posterior distribution under each of the four priors. The ESS and acceptance rate in the SMC sampler are monitored at each iteration and these plots are given in the Supplementary Material Figure S6 for the multiplicative prior with $\sigma_1^2 = \sigma_2^2 = 1$. Typically, the ESS decreases as the algorithm proceeds until it falls below the threshold, $N_{\text{threshold}} = N/3$, and it bounces back after resampling is performed. Due to bridging of target densities using tempering, the acceptance rate is usually high at the beginning when the temperature ϕ_t is close to zero and proposals have a high probability of being accepted [see (6.7)]. As the temperature increases, the samples becomes more concentrated in the regions of high posterior probability and the acceptance rate falls.

To compare the differences in edges between G_1 and G_2, we construct differential networks which display only edges likely to appear in one graph but not the other. Differential networks serve as powerful tools for exploring the changes in correlation structures across different conditions and have been considered widely in recent research. For instance, Valcárcel et al. (2011) define an edge as differential if the partial correlations estimated via linear shrinkage estimators differ significantly between two graphs while Peterson, Stingo and Vannucci (2015) and Mitra, Müller and Ji (2016) consider the posterior probability of an edge differing across conditions. Here, we adopt another definition, which enables us to differentiate more easily between the edges which are more likely to appear in G_1 than G_2 and vice versa. Let ρ_{ij}^k denote the posterior marginal probability of inclusion of edge (i, j) in G_k for $k = 1, 2$. We estimate ρ_{ij}^k as the proportion of SMC samples for which (i, j) is included in G_k and consider an edge to be differential if $|\rho_{ij}^1 - \rho_{ij}^2| > \kappa$ for some $0 < \kappa < 1$. Figure 11 shows the differential network corresponding to the different priors for $\kappa = 0.5$. The estimates of ρ_{ij}^1 and ρ_{ij}^2 for the edges in the differential networks are given in Table S3 in the Supplementary Material. Due to space limitations, we have also included further detailed results in the Supplementary Material. These include weighted graphs obtained from Algorithm 1 under different priors (Figures S7 and S8), posterior distributions (Figures S9, S10 and S11), betweenness centrality measures, weighted means and 95% credible intervals of the connectivities ($\pi_{i,k}$) and regression coefficients (β_{iq}) of each metabolite (Tables S4, S5 and S6) and the mean covariance matrices corresponding to the multiplicative prior with $\sigma_1^2 = \sigma_2^2 = 1$ (Tables S7 and S8).

The full network in the $K = 1$ case and the differential network in the $K = 2$ case both show similar topological characteristics corresponding to sub-graphs of metabolites. For the case of $K = 2$, the different prior hyperparameters only lead to different levels of shrinkage, but there is a high degree of similarity in terms of biological interpretation. For example, both Figures 9 and 11 show three different sub-graphs linking metabolites with shared metabolic origin. First, a group

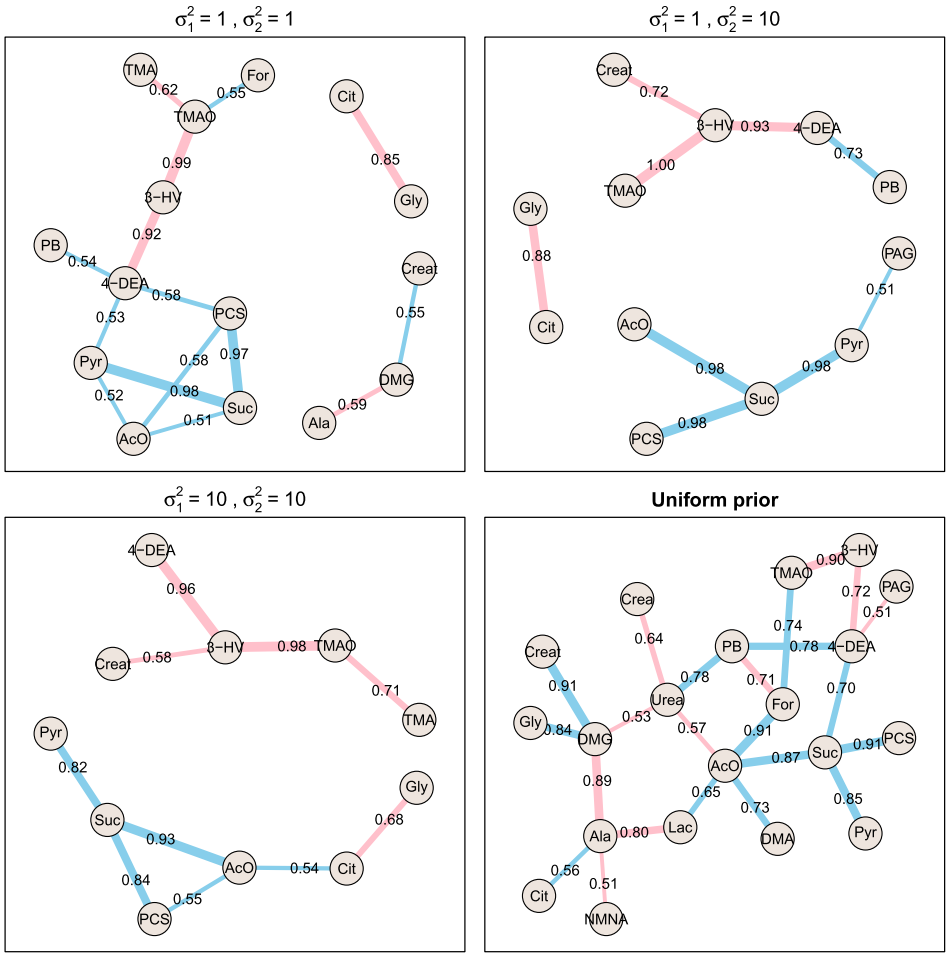

FIG. 11. *Differential network corresponding to the different priors. Edges in blue are likely to appear in G_1 but not in G_2 and pink edges are likely to appear in G_2 but not in G_1. The labels indicate the estimate of $|\rho_{ij}^1 - \rho_{ij}^2|$ for each edge (i, j).*

of organic acids including succinate, pyruvate, acetate and para-cresol sulphate (PCS) are connected, sometimes also with phenylacetylglutamine (PAG). Several of these metabolites (PCS and PAG) are known to be of gut bacterial origin, and Cd stress is known to modulate gut microbiota populations in mice [Liu et al. (2014)]. Increased acetate is a known consequence of renal damage, which could be linked to high Cd levels in this population. The second group contains trimethylamine (TMA) and its oxidation product trimethylamine-N-oxide (TMAO), both part of choline metabolism, plus 3-HV and 4-DEA, which are products of amino acid catabolism. Choline is an essential nutrient and is metabolised primarily in the liver. Due to its long biological half-life, Cd accumulates in human tissues,

especially the liver and kidney, so this observation may point towards a possible mechanistic connection. Moreover, in their original study of this data set, Ellis et al. (2012) reported a positive correlation between urinary Cd and both 4-DEA and 3-HV, though this relationship did not survive correction for age and sex. The third group links citrate and glycine, closely associated via malate and glyoxylate in central carbon metabolism (the network of metabolic reactions essential to life). A strong correlation between Cd and citrate was found by Ellis et al. (2012), while Valcárcel et al. (2011) found a significant deregulation of the dependency network associated with dimethylglycine, a bi-product of the synthesis of glycine from choline. Thus, it is plausible that several of the metabolites found in the networks of Figures 9 and 11 are involved in pathways disregulated due to Cd exposure. However, metabolite associations derive from a variety of factors and many may be indirect, and possibly nonbiochemical in origin, for example, change in expression of membrane transporters. Thus interpretation of dependency networks, such as those generated here, is difficult. Nonetheless, they give us a novel view of the data not exposed in conventional analyses, and may serve to help generate new hypotheses to be investigated by future biochemical experiments.

8. Conclusion. This article proposes using the multiplicative or Chung–Lu random graph model as a prior on the graphical space of GGMs, where the probability of inclusion of each edge is a product of the connectivities of the end nodes. This model can be used to encourage sparsity or particular degree structures, when such prior knowledge is available, say from a database or based on expert opinion. A Bayesian approach is adopted and priors are further placed on the connectivity of the nodes. We study the degree and clustering properties of the multiplicative prior and note that this prior is able to accommodate a wider range of degree structures than the Erdős–Rényi model. For example, we can use it to encourage shrinkage towards the extremes of 0 and 1, or degree distributions that are right-skewed by varying the hyperparameters, We illustrate how this prior can be applied to both single and multiple GGMs and a SMC sampler is developed for posterior inference. We find the performance of this sampler to be stable and consistent in our experiments and it can also be parallelized easily. The multiplicative prior also yields rich posterior inference, enabling a study of the connectivity of each node and how the propensity to connect varies across different experimental conditions in the case of multiple GGMs. This allows deeper exploration into the structure of dependency networks and may aid in the formulation of new scientific hypothesis and in opening further lines of investigations.

Acknowledgments. We are grateful to Drs. Jake Bundy, James Ellis and Hector Keun for provision of the data used in this study. We thank the referees and the Associate Editor for their comments which have helped to greatly improve the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks" (DOI: 10.1214/17-AOAS1076SUPP; .pdf). We provide additional material to support the results in this paper. This include Matlab code, further discussions, detailed derivations and further results on the application to urinary metabolic data.

REFERENCES

ARMSTRONG, H., CARTER, C. K., WONG, K. F. K. and KOHN, R. (2009). Bayesian covariance matrix estimation using a mixture of decomposable graphical models. *Stat. Comput.* **19** 303–316. MR2516221

ATAY-KAYIS, A. and MASSAM, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92** 317–335. MR2201362

BESKOS, A., JASRA, A., KANTAS, N. and THIERY, A. (2016). On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.* **26** 1111–1146. MR3476634

CARVALHO, C. M. and SCOTT, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika* **96** 497–512. MR2538753

CARVALHO, C. M. and WEST, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Anal.* **2** 69–97. MR2289924

CHUN, H., ZHANG, X. and ZHAO, H. (2015). Gene regulation network inference with joint sparse Gaussian graphical models. *J. Comput. Graph. Statist.* **24** 954–974. MR3432924

CHUNG, F. and LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99** 15879–15882. MR1944974

D'SOUZA, R. M., BORGS, C., CHAYES, J. T., BERGER, N. and KLEINBERG, R. D. (2007). Emergence of tempered preferential attachment from optimization. *Proc. Natl. Acad. Sci. USA* **104** 6112–6117.

DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. MR3164871

DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333

DEL MORAL, P., DOUCET, A. and JASRA, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.* **22** 1009–1020. MR2950081

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* 157–175.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

DOBRA, A., HANS, C., JONES, B., NEVINS, J. R., YAO, G. and WEST, M. (2004). Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.* **90** 196–212. MR2064941

ELLIS, J. K., ATHERSUCH, T. J., THOMAS, L. D., TEICHERT, F., PEREZ-TRUJILLO, M., SVENDSEN, C., SPURGEON, D. J., SINGH, R., JARUP, L., BUNDY, J. G. and KEUN, H. C. (2012). Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Medicine* **10** 61.

FENNER, T., LEVENE, M. and LOIZOU, G. (2007). A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff. *Soc. Netw.* **29** 70–80.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

GIOT, L., BADER, J. S., BROUWER, C. et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* **302** 1727–1736.

GIRAUD, C., HUET, S. and VERZELEN, N. (2012). Graph selection with GGMselect. *Stat. Appl. Genet. Mol. Biol.* **11** Art. 3, 52. MR2934907

GUO, J., LEVINA, E., MICHAILIDIS, G. and ZHU, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98** 1–15. MR2804206

JASRA, A., STEPHENS, D. A., DOUCET, A. and TSAGARIS, T. (2011). Inference for Lévy driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22.

JEONG, H., MASON, S. P., BARABASI, A. L. and OLTVAI, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411** 41–42.

JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C. and WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci.* **20** 388–400. MR2210226

LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. MR1419991

LENKOSKI, A. and DOBRA, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *J. Comput. Graph. Statist.* **20** 140–157. MR2816542

LIU, Y., LI, Y., LIU, K. and SHEN, J. (2014). Exposing to cadmium stress cause profound toxic effect on microbiota of the mice intestinal tract. *PLoS ONE* **9** e85323.

MITRA, R., MÜLLER, P. and JI, Y. (2016). Bayesian graphical models for differential pathways. *Bayesian Anal.* **11** 99–124. MR3447093

MOHAN, K., LONDON, P., FAZEL, M., WITTEN, D. and LEE, S.-I. (2014). Node-based learning of multiple Gaussian graphical models. *J. Mach. Learn. Res.* **15** 445–488. MR3190845

MURRAY, I., GHAHRAMANI, Z. and MACKAY, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the* 22*nd Annual Conference on Uncertainty in Artificial Intelligence* (T. Decther and T. Richardson, eds.) 359–366.

NEWMAN, M. E. J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409. MR1812610

NEWMAN, M. E. J., STROGATZ, S. H. and WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* (3) **64**.

OLHEDE, S. C. and WOLFE, P. J. (2013). Degree-based network models. Available at arXiv:1211.6537.

PETERSON, C., STINGO, F. C. and VANNUCCI, M. (2015). Bayesian inference of multiple Gaussian graphical models. *J. Amer. Statist. Assoc.* **110** 159–174. MR3338494

RASTELLI, R., FRIEL, N. and RAFTERY, A. E. (2015). Properties of latent variable network models. Available at arXiv:1506.07806.

ROVERATO, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* **29** 391–411. MR1925566

SALAMANCA, B. V., EBBELS, T. M. D. and DE IORIO, M. (2014). Variance and covariance heterogeneity analysis for detection of metabolites associated with cadmium exposure. *Stat. Appl. Genet. Mol. Biol.* **13** 191–201. MR3181222

SCHAEFER, J., OPGEN-RHEIN, R. and STRIMMER, K. (2015). R package: GeneNet version 1.2.13. Available at https://cran.r-project.org/web/packages/GeneNet/index.html.

SCHÄFER, C. and CHOPIN, N. (2013). Sequential Monte Carlo on large binary sampling spaces. *Stat. Comput.* **23** 163–184. MR3016936

STEUER, R. (2006). Review: On the analysis and interpretation of correlations in metabolomic data. *Brief. Bioinform.* **7** 151–158.

TAN, L. S., JASRA, A., DE IORIO, M. and EBBELS, T. M. (2017). Supplement to "Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks." DOI:10.1214/17-AOAS1076SUPP.

TELESCA, D., MÜLLER, P., PARMIGIANI, G. and FREEDMAN, R. S. (2012). Modeling dependent gene expression. *Ann. Appl. Stat.* **6** 542–560. MR2976482

VALCÁRCEL, B., WÜRTZ, P., SEICH AL BASATENA, N.-K., TUKIAINEN, T., KANGAS, A. J., SOININEN, P., JÄRVELIN, M.-R., ALA-KORPELA, M., EBBELS, T. M. and DE IORIO, M. (2011). A differential network approach to exploring differences between biological states: An application to prediabetes. *PLoS ONE* **6** e24702.

WANG, H. and LI, S. Z. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Stat.* **6** 168–198. MR2879676

WANG, H., REESON, C. and CARVALHO, C. M. (2011). Dynamic financial index models: Modeling conditional dependencies via graphs. *Bayesian Anal.* **6** 639–663. MR2869960

YAJIMA, M., TELESCA, D., JI, Y. and MÜLLER, P. (2015). Detecting differential patterns of interaction in molecular pathways. *Biostatistics* **16** 240–251. MR3365426

L. S. L. TAN
A. JASRA
FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
BLOCK S16, LEVEL 7, 6 SCIENCE DRIVE 2
SINGAPORE 117546
SINGAPORE
E-MAIL: statsll@nus.edu.sg
staja@nus.edu.sg

M. DE IORIO
DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON WC1E 6BT
UNITED KINGDOM
E-MAIL: m.deiorio@ucl.ac.uk

T. M. D. EBBELS
DEPARTMENT OF SURGERY AND CANCER
IMPERIAL COLLEGE LONDON
SOUTH KENSINGTON CAMPUS
LONDON SW7 2AZ
UNITED KINGDOM
E-MAIL: t.ebbels@imperial.ac.uk