Biometrics 75, 172–182
March 2019

DOI: 10.1111/biom.12958

Bayesian Inference of Hub Nodes Across Multiple Networks

Junghi Kim,[*] Kim-Anh Do,[**] Min Jin Ha,[***] and Christine B. Peterson[****]

Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.
[*]*email:* jkim30@mdanderson.org
[**]*email:* kimdo@mdanderson.org
[***]*email:* mjha@mdanderson.org
[****]*email:* cbpeterson@mdanderson.org

Summary. Hub nodes within biological networks play a pivotal role in determining phenotypes and disease outcomes. In the multiple network setting, we are interested in understanding network similarities and differences across different experimental conditions or subtypes of disease. The majority of proposed approaches for joint modeling of multiple networks focus on the sharing of edges across graphs. Rather than assuming the network similarities are driven by individual edges, we instead focus on the presence of common hub nodes, which are more likely to be preserved across settings. Specifically, we formulate a Bayesian approach to the problem of multiple network inference which allows direct inference on shared and differential hub nodes. The proposed method not only allows a more intuitive interpretation of the resulting networks and clearer guidance on potential targets for treatment, but also improves power for identifying the edges of highly connected nodes. Through simulations, we demonstrate the utility of our method and compare its performance to current popular methods that do not borrow information regarding hub nodes across networks. We illustrate the applicability of our method to inference of co-expression networks from The Cancer Genome Atlas ovarian carcinoma dataset.

Key words: Bayesian modeling; Gaussian graphical model; Hub node; Multiple networks

1. Introduction

Rather than being determined by a single causal mutation or event, most common diseases arise from dysregulation of the complex networks of interaction that govern cellular behavior. These networks, including gene regulatory, protein interaction, and metabolic networks, have common topological features, including the presence of a few highly connected nodes (Barabasi and Oltvai, 2004; Barabasi et al., 2011). This small set of nodes with a large number of connections, often called hubs, mediate interactions among the large number of weakly connected nodes within the network, and ultimately determine network function (Jeong et al., 2001). In the context of cancer, studies have shown that the proteins of disease-associated genes tend to have higher degree than other nodes in the network (Wachi et al., 2005; Jonsson and Bates, 2006). The identification of hub nodes is of particular interest as they represent attractive targets for treatment due to their critical position within the network structure (Tan et al., 2014).

High-throughput technologies have enabled the acquisition of data sets characterizing the abundances of gene transcripts, proteins, and metabolites, often across a variety of conditions. Given such data, we are interested in inferring both a network for each setting, and in understanding the similarities and differences of the networks across conditions. In a statistical framework, this corresponds to the problem of inferring multiple graphical models, where the edges in the graph correspond to dependence relations among the observed variables. A number of statistical methods have been developed in recent years to address this challenge. Danaher et al. (2014) proposed joint estimation of multiple graphical models using convex penalties. Specifically, they employed two penalty functions: the fused graphical lasso, which encourages both shared structure and shared edge values, and the group graphical lasso, which focuses only on shared structure. In the Bayesian framework, Peterson et al. (2015) proposed a method to estimate multiple Gaussian graphical models via a Markov random field (MRF) prior which flexibly encourages common structure when supported by the data. Some approaches focus on learning differential networks, mainly in the context of two sample groups. Ha et al. (2015) constructed the differential network by subtracting the edge weights of two networks estimated from covariance regression. In the Bayesian framework, Mitra et al. (2016) treated one group's network as the baseline, and defined latent indicators of difference for each edge between the two graphs to infer a differential network. The methods mentioned above assume that similarities and differences between networks are driven by edges. In contrast, Mohan et al. (2014) took a node-based approach, assuming that similarities among multiple graphs are due to the presence of either common hub nodes that are shared across all groups (co-hub node joint graphical lasso), or that differences are attributable to individual nodes that are perturbed across conditions (perturbed-node joint graphical lasso). The methods proposed in Mohan et al. (2014) have several limitations, however, including the challenge of selecting

appropriate tuning parameters, and the fact that co-hub node joint graphical lasso assumes all hubs are shared across all groups.

In this article, we formulate bHUB, a Bayesian hub-based approach to the problem of multiple network inference. We introduce a latent indicator for each node which reflects its status as a hub node, and formulate a prior which links these indicators across multiple graphs. A particularly novel aspect of our model is that prior edge inclusion probabilities depend on whether the edge is connected to a hub node; this contrasts with existing Bayesian approaches, which typically assume an independent Bernoulli prior on edge inclusion. Moreover, in the context of multiple sample groups, we link the selection of hub nodes, rather than individual edges, as in the usual paradigm for this setting. We are able to provide simple and interpretable posterior summaries on the status of individual nodes as hubs, and, in the multiple network setting, an intuitive interpretation in terms of shared or perturbed hub nodes. To our knowledge, this is the first Bayesian modeling approach to incorporate inference on highly connected nodes in graph learning for multiple groups. This article is organized as follows. Section 2 provides background on graphical models and on existing methods for graphical model estimation. Section 3 presents the modeling approach for our proposed method, including the likelihood and the construction of the priors. Section 4 includes posterior inference using Markov chain Monte Carlo (MCMC). Section 5 provides an illustrative example, while Section 6 includes a more realistic simulation study. Section 7 demonstrates the application of our method to the inference of co-expression networks across four subtypes of ovarian cancer. Section 8 concludes the article.

2. Background

2.1. *Graphical Models*

An undirected graph $G = (V, E)$ is composed of a set of p nodes $V = \{1, 2, \ldots, p\}$ and a set of edges $E \subset V \times V$. Graphical models in statistics use an undirected graph to represent conditional dependence relations among a set of variables. Each node in the graph corresponds to a random variable, and two nodes are connected by an edge if and only if the corresponding variables are dependent conditional on all of the remaining variables (Lauritzen, 1996). The lack of an edge between two nodes, therefore, represents a conditional independence relationship. In the context of multivariate normal data, Gaussian graphical models (Dempster, 1972) provide a convenient framework for graphical model estimation. Specifically, consider the multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean, and $\boldsymbol{\Omega}_{p\times p}$ is the precision matrix, the inverse of the covariance matrix. It is a special property of the multivariate normal distribution that $\omega_{ij} = 0$ if and only if variables i and j are conditionally independent, so the pattern of zero and nonzero elements in $\boldsymbol{\Omega}$ corresponds directly to an edge set E. The Gaussian graphical modeling framework has been applied to model various types of biological networks, including gene regulatory and protein interaction networks, where the underlying expression or abundance data have been transformed to achieve normality.

2.2. *Graphical Model Estimation*

The graphical lasso, which relies on an ℓ_1 penalty to achieve sparsity in estimation of the precision matrix, is one of the most common approaches for graphical model estimation (Friedman et al., 2007). Another popular approach, constrained ℓ_1 minimization for inverse matrix estimation (CLIME), has desirable theoretical properties and can be solved using linear programming (Cai et al., 2011). In the Bayesian framework, graphical model estimation can be achieved through the choice of an appropriate prior on $\boldsymbol{\Omega}$. The Bayesian graphical lasso, proposed as the Bayesian analogue to the graphical lasso, incorporates double exponential priors on the off-diagonal entries of the precision matrix (Wang, 2012). While the computational tractability of this approach is attractive, it does not result in sparse estimates of the precision matrix. Alternative approaches in the Bayesian framework utilize a hierarchical formulation, with priors on the precision matrix conditional on a graph structure $p(\boldsymbol{\Omega}|G)$ and on the graph structure $p(G)$, where the graph structure is modeled as a set of binary latent variables. The G-Wishart distribution, the conjugate prior on the precision matrix conditional on a graph G, can be used within this framework (Roverato, 2002; Atay-Kayis and Massam, 2005). Sampling from the G-Wishart distribution is computationally challenging, however, due in part to its intractable normalizing constant. Recently, Wang (2015) proposed the stochastic search structure learning (SSSL) approach, which utilizes a continuous mixture prior on the off-diagonal entries of the precision matrix. This approach, inspired by mixture priors in the context of Bayesian variable selection (George and McCulloch, 1993), is attractive as it allows both direct modeling of the graph structure and computationally efficient sampling.

3. Methods

We develop a Bayesian hierarchical model (bHUB) to address the challenge of inferring multiple graphs. In contrast to most existing joint network estimation methods, which assume edge-wise differences across graphs, we focus on the sharing of hub nodes across networks. Our approach not only allows accurate estimation of the graph structure for each sample group, but also enables inference on the status of biologically important hub nodes, which may be common across all groups, or perturbed and only active in certain groups. In this section, we describe the formulation of the full hierarchical model; we would like to emphasize that our primary innovation lies in the prior formulations described in Sections 3.3 and 3.4, where we propose a novel approach to characterize prior edge probabilities as depending on the status of nodes as hubs, and link the identification of these hubs across multiple groups.

3.1. *Hierarchical Model*

Assume we have K groups, which may correspond to different experimental conditions or subtypes of disease. We assume that the same variables are measured for each group, but the sample sizes may differ, so that for the kth group, n_k samples are observed for p random variables. The likelihood is a group-specific multivariate normal

$$x_{k,i} \sim N(\mathbf{0}, \boldsymbol{\Omega}_k^{-1})$$

for $i = 1, \ldots, n_k$, where $\boldsymbol{\Omega}_k$ is a group-specific precision matrix, and we assume that the data are centered, so $\boldsymbol{\mu}_k = \mathbf{0}$.

Here, we provide an overview of the proposed Bayesian hierarchical model, with more detailed discussion to follow in Sections 3.2–3.4. For group $k = 1, 2, \ldots, K$, we specify the following set of hierarchical priors

$$p(\boldsymbol{\Omega}_k|\mathbf{G}_k),$$
$$p(\mathbf{G}_k|\boldsymbol{h}_k, \beta),$$
$$p(\boldsymbol{h}_k|\tau, \eta),$$

where β, τ, η are fixed hyperparameters. In the first level of the hierarchy, the prior on the precision matrix $\boldsymbol{\Omega}_k$ is conditional on a group-specific graph structure $\mathbf{G}_k$. We express the graph structure $\mathbf{G}_k$ as an adjacency matrix, where $g_{k,ij} = 1$ if nodes i and j are connected in the graph for group k, and 0 otherwise. We introduce a vector of latent indicators $\boldsymbol{h}_k = (h_{k,1}, h_{k,2}, \ldots, h_{k,p})'$ to represent the status of each node as a hub node, where $h_{k,i} = 1$ if node i is a hub in graph k, and 0 otherwise. In the second level of the hierarchy, the prior on the graph $\mathbf{G}_k$ is conditional on the node-specific parameter vector $\boldsymbol{h}_k$ and the hyperparameter β, which affects the prior probability of edge inclusion. In the final level of the hierarchy, the hyperparameters τ and η determine $p(h_{k,i} = 1)$, the prior probability of a node being a hub. Figure 1 shows a graphical representation the bHUB model, where circles indicate parameters, and squares represent observed random variables.

3.2. *Priors on Precision Matrices and Graphs*

For the prior on the precision matrices, we adapt the mixture prior proposed as part of the stochastic search structure learning (SSSL) method in Wang (2015) since it allows us to perform inference on the graph structure while maintaining computational scalability. Specifically, we assume that the off-diagonal elements in the precision matrix follow a normal mixture distribution, where entries in the precision matrix corresponding to true zeros (missing edges) belong to a normal distribution centered at 0 with small variance, and entries corresponding to truly nonzero values (edges) belong to a normal distribution with larger variance. Since they are constrained to be positive, the p diagonal terms in each precision matrix have prior exponential densities. Unlike the stochastic search structure learning approach of Wang (2015), which assumes a Bernoulli prior with the same probability for all edges, we introduce an edge-specific parameter $\pi_{k,ij}$ that depends on whether nodes i and j are considered to be hub nodes. Formally, we express the priors on the precision matrix and graph for each group $k = 1, \ldots, K$ as

$$p(\boldsymbol{\Omega}_k|\mathbf{G}_k, v_0, v_1, \lambda) \propto \prod_{i<j} N(\omega_{k,ij}|0, v_{k,ij}^2) \prod_i \exp\left(\omega_{k,ii}|\frac{\lambda}{2}\right), \quad (1)$$

$$p(\mathbf{G}_k|\{\pi_{k,ij}\}) \propto \prod_{i<j} \{\pi_{k,ij}^{g_{k,ij}}(1-\pi_{k,ij})^{1-g_{k,ij}}\}, \quad (2)$$

where $\boldsymbol{\Omega}_k$ is constrained to belong to the space of symmetric positive definite matrices. In equation (1), $N(\omega|0, v^2)$ is the density function of a normal random variable with mean 0 and variance v^2 evaluated at ω, and $\exp(\omega|\lambda)$ represents the exponential density function. The variance $v_{k,ij}$ is determined by the presence or absence of the corresponding edge:

$$v_{k,ij} = \begin{cases} v_0 & \text{if } g_{k,ij} = 0, \\ v_1 & \text{if } g_{k,ij} = 1, \end{cases} \quad (3)$$

where v_0 is small, and v_1 is relatively large. Wang (2015) finds good performance for graph structure learning given parameters in the range $v_0 \geq 0.01$–$h \leq 1000$, where $v_1 = h \cdot v_0$.

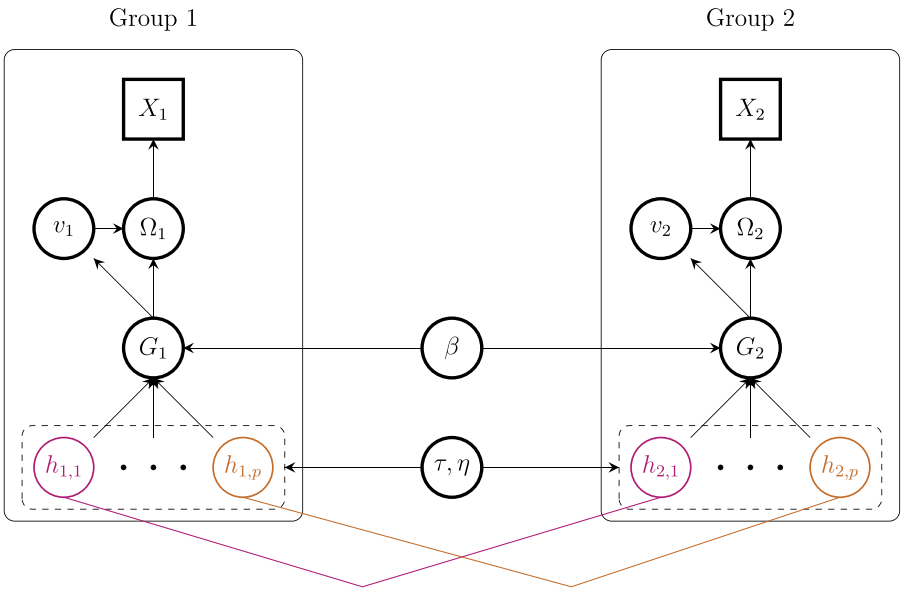


Figure 1. Graphical representation of the bHUB model for $K = 2$ groups. Circles indicate parameters, and squares represent observed random variables. Node-wise information is shared across groups. See Section 3 for a detailed explanation of the model.

3.3. *Prior Probability of Edge Inclusion*

We would like the prior probability of edge inclusion $\pi_{k,ij}$ to depend on whether nodes i and j are hub nodes. In particular, $\pi_{k,ij}$ should be small when neither i nor j is a hub, to encourage overall sparsity of the graph structure, and relatively large when either i or j is a hub, since hub nodes tend to have a large number of connections in the graph. To achieve this goal, we use the following prior formulation:

$$\pi_{k,ij} = \beta_1(1-h_{k,i})(1-h_{k,j}) + \beta_2\{1-(1-h_{k,i})(1-h_{k,j})\}. \quad (4)$$

When neither node is a hub (i.e., $h_{k,i} = h_{k,j} = 0$), the prior edge inclusion probability $\pi_{k,ij}$ is β_1, while when either $h_{k,i}$ or $h_{k,j}$ is a hub (i.e., $h_{k,i} = 1$ or $h_{k,j} = 1$), the prior edge inclusion probability will be β_2. The parameters $\{\beta_1, \beta_2\}$ should be small and large, respectively. To encourage sparsity and avoid the false positive selection of edges, we set $\beta_1 = 0.001$, reflecting the fact that a connection between two non-hub nodes is unlikely. Since edges from hub nodes are considered likely, we would like β_2 to be large. To allow flexibility in its value, we place a beta hyperprior on β_2

$$p(\beta_2) = \frac{1}{B(a,b)} \cdot \beta_2^{a-1}(1-\beta_2)^{b-1}, \quad (5)$$

where a and b are fixed hyperparameters. The parameters a and b should be chosen so that the prior mean of β_2 is larger than β_1, that is, there is separation of the edge inclusion probabilities for hub nodes versus non-hub nodes. Smaller values of β_2 encourage hub nodes with fewer connections and overall sparsity of the graphs $\mathbf{G}_1, \ldots, \mathbf{G}_K$. Larger values of β_2 encourage more highly connected hub nodes, and therefore graphs which are denser.

3.4. *Node-Specific Prior*

In our model, the Bernoulli random variable $h_{k,i}$ indicates whether node i in graph k is a hub or not. We share information on the hub status of each node across graphs in determining $h_{k,i}$. Specifically, we assume that node i in graph k has a larger prior probability of being a hub if node i is a hub in the graphs for other groups (i.e., $h_{m,i} = 1$ for $m \neq k$) since hub nodes are likely to be preserved across settings. The prior probability of $h_{k,i}$ is given by

$$Pr(h_{k,i}|\tau, \eta, h_{-(k,i)}) \propto \exp(h_{k,i}\left(\eta + 2\tau \sum_{m\neq k} h_{m,i}\right).$$

This is consistent with the joint probability across the k groups

$$P(\boldsymbol{h}_i|\tau,\eta) = C(\eta,\tau)^{-1}\exp(\eta \cdot \mathbf{1}_k\boldsymbol{h}_i^T + \tau \cdot \boldsymbol{h}_i^T(1-I)\boldsymbol{h}_i), \quad (6)$$

where $\boldsymbol{h}_i = (h_{1,i}, ..., h_{K,i})^T$ and $\mathbf{1}_k$ is a K-vector of 1s. Here, 1 is a $K \times K$ matrix with all elements one and I is an identity matrix. The normalizing constant in equation (6) is

$$C(\eta,\tau) = \sum_{\boldsymbol{h}_i \in \{0,1\}^K} \exp(\eta \cdot 1_k^T\boldsymbol{h}_i + \tau \cdot \boldsymbol{h}_i^T(1-I)\boldsymbol{h}_i),$$

which depends only on τ and η. The joint prior on the graphs $(\mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_K)$ is the product of the densities for each node:

$$p(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p|\tau,\eta) = \prod_{i=1,\ldots,p} P(\boldsymbol{h}_i|\tau,\eta).$$

Under this prior, the conditional probability for node i to be a hub in graph $\mathbf{G}_k$, given the status of the nodes in the remaining graphs, is

$$p(h_{k,i}|\{h_{m,i}\}_{m\neq k},\tau,\eta) = \frac{\exp\left(h_{k,i}\left(\eta + 2\tau\sum_{m\neq k} h_{m,i}\right)\right)}{1+\exp\left(\eta + 2\tau\sum_{m\neq k} h_{m,k}\right)}$$

where τ is a scalar hyperparameter. Note that when $h_{m,i} = 0$ for all $m \neq k$ (i.e., node i is not selected as a hub in any other graph), we have $\Pr(h_{k,i} = 1) = \exp(\eta)/(1+\exp(\eta))$, which is 0.5 when $\eta = 0$. The inverse logit transformation of η can be interpreted as the base probability for each node to be a hub. To encourage sparsity, we set $\eta < 0$. We further specify a $Ga(a_\tau, b_\tau)$ hyperprior on τ, ensuring the positivity of this parameter.

3.5. *Recommended Parameter Settings*

The hyperparameters v_0, v_1, and λ, which play a role in the prior on the precision matrices given in equation (1), have a similar influence to that described in the sensitivity analysis provided in Wang (2015). In our simulations, we set $v_0^2 = 0.03$, $v_1^2 = 100 \times v_0^2$, and $\lambda = 1$, consistent with the guidelines provided by Wang (2015).

The choice of a and b in the prior on β_2 given in equation (5) affect the prior probability of edge inclusion for edges with a hub node as a terminal node. A larger prior mean for β_2 results in the selection of more edges from hub nodes, and increased posterior probabilities for nodes being hubs. The prior edge inclusion probability for non-hub nodes is set at $\beta_1 = 0.001$. Given this fixed value of β_1, a larger β_2 prior mean results in more separation between edge inclusion probabilities for hub nodes and non-hub nodes.

The hyperparameter η in equation (3.4) should be set so that the baseline probability of a node being a hub is small, since typically only a few nodes within a network function as hubs. For example, if we assume one hub node exists out of 100 nodes, we can define $\text{logit}^{-1}(\eta) = 0.01$ so that the prior probability of each node being a hub is 0.01.

Additional details on sensitivity to these parameters and to the parameters v_0, h and τ are given in Supplementary Section S2.

4. Posterior Inference

Let $\boldsymbol{\Psi}$ denote the set of all parameters $\{\boldsymbol{\Omega}_k, \mathbf{G}_k, h_{k,i}, \beta, \tau, \eta,$ for $k = 1, \ldots, K, i = 1, \ldots, p\}$ and $\mathbf{X}$ denote the observed data for all sample groups. Given the likelihood and priors defined in Section 3, the joint posterior density can be written

$$p(\boldsymbol{\Psi}|\mathbf{X}) \propto \prod_{k=1}^{K} \left\{p(X_k|0,\boldsymbol{\Omega}_k^{-1}) \cdot p(\boldsymbol{\Omega}_k|\mathbf{G}_k) \cdot p(\mathbf{G}_k|h_k,\beta)\right\}$$
$$p(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p|\tau,\eta) \cdot p(\beta) \cdot p(\tau,\eta).$$

Since this density is not analytically tractable, we rely on Markov chain Monte Carlo (MCMC) sampling to generate a posterior sample of the parameters as the basis for inference.

4.1. *MCMC Sampling*

At the top level, our MCMC sampler follows a Gibbs scheme, in which parameters are sampled from their posterior full conditional distributions. To sample the graph and precision matrix for each group, we adapt the method of Wang (2015), which does not require computation of the prior normalizing constant and allows block updates of precision matrices and graphs. The latent indicators $h_{k,i}$ and β_2 are sampled directly from their posterior full conditional distributions, while we use a Metropolis-within-Gibbs step to sample τ. We provide an overview of the sampling scheme here; additional detail is provided in the Supplementary Materials. For each group k, we initialize the graph $\mathbf{G}_k^{(0)} = I$ and the node indicators $h_{k,i}^{(0)} = 0$ for $i = 1, \ldots, p$. We initialize the precision matrix to the identity matrix: $\mathbf{\Omega}_k^{(0)} = I_p$. At each iteration t, we apply following updates:

Step 1. Update $\mathbf{\Omega}_k^{(t)}$ and $\mathbf{G}_k^{(t)}$ using the column-wise block Gibbs sampler of Wang (2015).
Step 2. Update the node-specific parameters ${h_{k,i}}^{(t)}$ for $i = 1, \ldots, p$, $k = 1, \ldots, K$.
Step 3. Update the parameter $\beta_2^{(t)}$ affecting the probabilities of edge inclusion.
Step 4. Update the parameter $\tau^{(t)}$ affecting the probabilities of nodes being hubs.

Following Wang (2015) and others, for the purpose of graph selection, we take the posterior median graph, that is, the graph consisting of all edges with a marginal posterior probability of inclusion (PPI) of at least 0.5. To obtain ROC curves for structure learning across a range of model sizes, we vary the selection threshold on the PPIs in the range $[0, 1]$.

5. An Illustrative Example

To allow an easily interpretable comparison of the performance of bHUB to existing network inference methods, we set up an illustrative example. We construct three graphs on $p = 60$ nodes, shown in Figure 2, and their corresponding precision matrices $\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$, and $\mathbf{\Omega}_3$. The graph for group 1 includes one hub node (node 1) while the graphs for groups 2 and 3 each have two hub nodes (node 1 and node 12). The hubs have differing degree: in particular, node 12 in graph 3 has fewer edges than the other hub nodes. To construct the precision matrices $\mathbf{\Omega}_k$, $k = 1, \ldots, 3$, we sample values from the uniform distribution on $\{[-0.6, -0.4] \cup [0.4, 0.6]\}$ for the nonzero elements. To ensure the positive definiteness, we use an approach similar to that of Danaher et al. (2014), in which we divide each off-diagonal element by the sum of the off-diagonal elements in its row, and then average the matrix with its transpose. To generate the simulated data sets, for each $\mathbf{\Omega}_k$, we sample $n = 100$ subjects from the multivariate normal distribution $N(\mathbf{0}, \mathbf{\Omega}_k^{-1})$ for $k = 1, \ldots, 3$.

We compare the performance of bHUB against existing methods in terms of graph learning and hub identification. We apply methods designed for the inference of a single graph, including the graphical lasso (GLASSO) (Friedman et al., 2007), CLIME (Cai et al., 2011), and SSSL, that is, separate Bayesian inference using Bernoulli edge priors (Wang, 2015), separately for each graph. Since it is the most common method for joint estimation, we also compare to the joint graphical lasso (JGL) (Danaher et al., 2014) to perform joint estimation across multiple graphs. Although the approach proposed in Mohan et al. (2014) is relevant to our work, we did not include Mohan et al. (2014) as a comparison because the software implementation provided is only for the case of two groups, and the selection of tuning parameters is acknowledged to be an open problem, making straightforward application of the method infeasible. The tuning parameters for GLASSO and CLIME are selected using five-fold cross-validation. To select JGL tuning parameters λ_1 and λ_2, as recommended in Danaher et al. (2014), we perform a grid search to find the parameter set that minimizes the Akaike information criterion (AIC). For SSSL, we use the hyperparameter setting $v_0^2 = 0.03$, $v_1^2 = 3$, $\pi = 10/(p-1)$, and $\lambda = 1$. Finally, we apply our proposed method (bHUB) with the parameter setting $v_0^2 = 0.03$, $v_1^2 = 3$, and $\lambda = 1$, so the shared parameters with SSSL have identical settings. We choose a and b in equation (5) so that the prior on β_2 has a mean of 0.025, implying that the edge inclusion probability for hub nodes is 25 times larger than that of non-hub nodes, given that $\beta_1 = 0.001$. For the prior on τ,

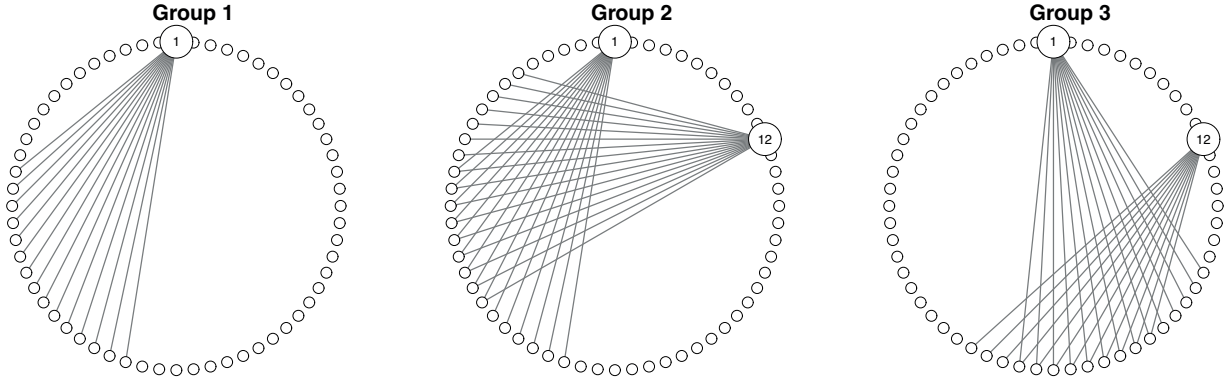
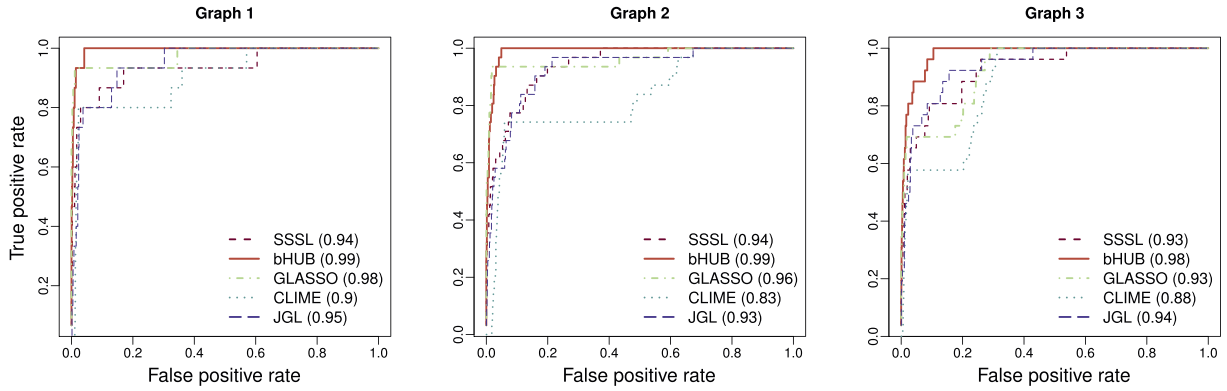


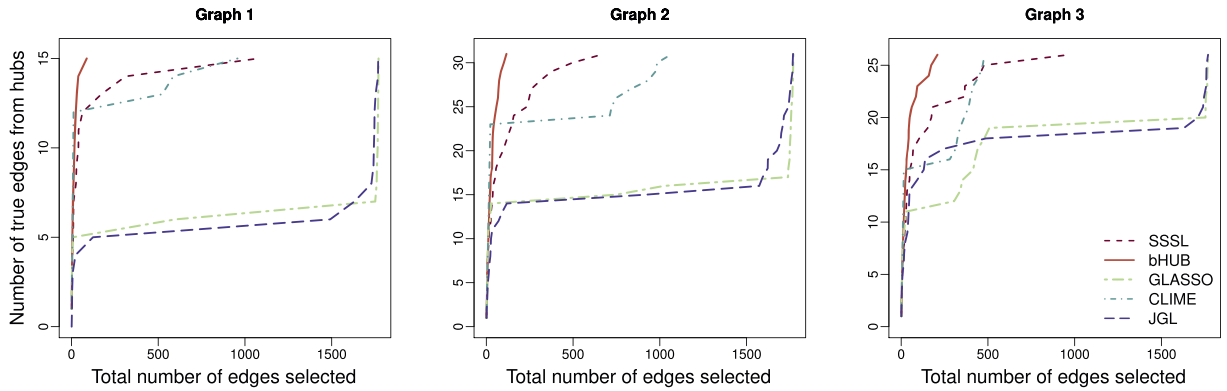
Figure 2. True graph structure for the illustrative example provided in Section 5. Node 1 is a true common hub node in all graphs. Node 12 is a true common hub node for Graph 2 and Graph 3.

which affects the prior probability of a node being a hub, we use a $Ga(a_\tau, b_\tau)$ density with $a_\tau = 1.5$ and $b_\tau = 0.2$, and set $\text{logit}^{-1}(\eta) = 0.01$ so that the default probability of being a hub is 0.01. To obtain a sample from the posterior distribution, we run the MCMC sampler described in Section 4 with 2000 iterations as burn-in and 10,000 iterations as the basis of inference.

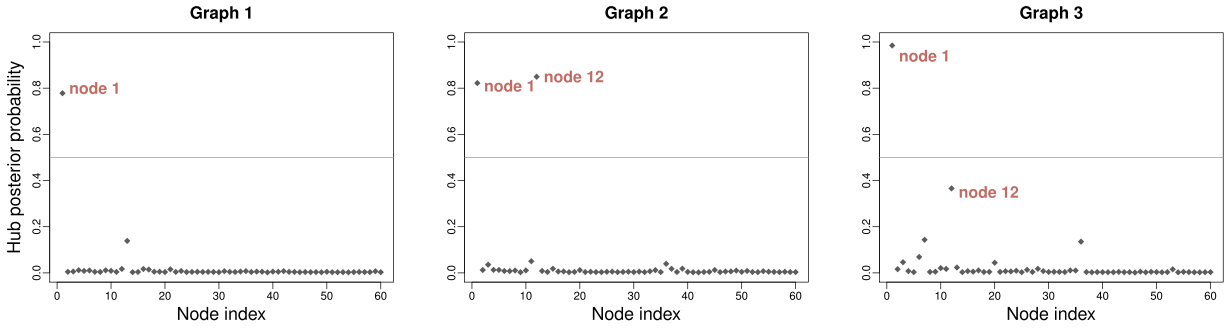
Figure 3 illustrates the performance of each method. The top panel (a) shows ROC curves for edge learning in each graph. The middle panel (b) shows performance for identifying true edges from hub nodes. The bottom panel (c) presents the posterior probability of being a hub for each node. bHUB shows good performance in graph structure learning, with a higher AUC than other methods. In particular, as shown in



(a) ROC curves for graph learning



(b) Performance for hub identification



(c) Hub posterior probability

Figure 3. Performance comparison for the illustrative example provided in Section 5. The top panel (a) shows ROC curves for identifying each graph structure. The middle panel (b) shows performance for identifying true edges from hub nodes, where Graph 1 has one hub with 15 edges, Graph 2 has two hub nodes with 31 true positive edges, and Graph 3 has 26 true positive edges from two hubs. The bottom panel (c) shows the posterior probability of each node being a hub in which node 1 is a true co-hub node in all graphs, and node 12 is a true co-hub for Graph 2 and Graph 3. The horizontal line represents the posterior probability 0.5.

panel (b), it outperforms other methods in identifying true edges from hub nodes. As expected, the posterior probability for node 12 in group 3 is slightly smaller than that of other hubs, but in general, the posterior probability for nodes being hub nodes indicates that the proposed method successfully separates the hub nodes and non-hub nodes. Section S3 in the Supplementary Materials addresses convergence of the sampler. Figure S2 is a trace plot of the parameter τ, which shows good mixing and no strong trends, while Figure S3 indicates posterior convergence as demonstrated by agreement between independent chains. Figure S4 in the Supplementary Materials shows the degree distributions for the networks estimated by each method.

6. Simulation Study

In this section, we present simulation results demonstrating the performance of our proposed method for graph learning and hub identification for a variety of different graph structures more complex than those considered in Section 5.

6.1. *Simulation Set-Up*

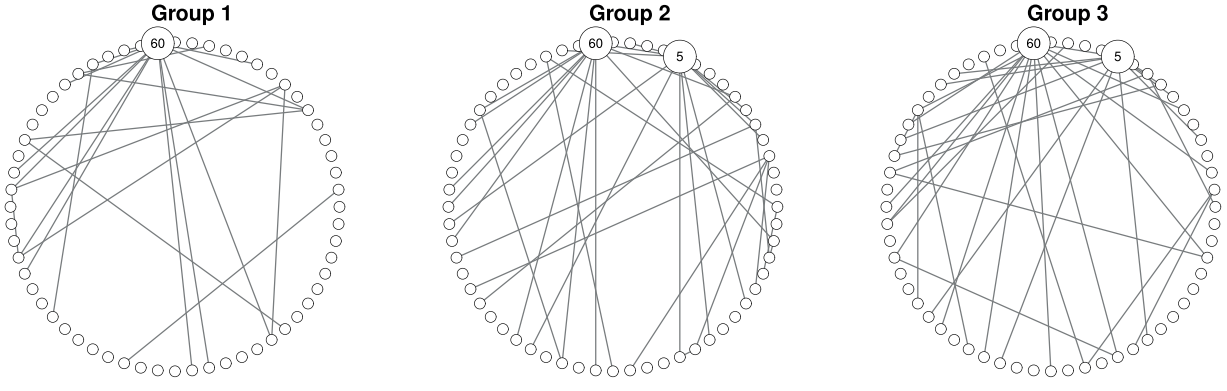
We generate a variety of graph structures: scale-free networks with shared hubs (*Setting 1*), networks with shared edges (*Setting 2*), and community networks (*Setting 3*). Figure 4 (*Setting 1*) and Figure S6 (*Settings 2* and *3*, Supplementary Materials) illustrate the true graph structures for each simulated setting. In all settings, node 60 is a true common hub node across all graphs, and node 5 is a true co-hub for Graph 2 and Graph 3. We describe *Setting 1* in more detail below. The details for *Settings 2* and *3* are given in the Supplementary Materials.

In *Setting 1*, we consider scale-free random graphs, where the node degrees follow a power-law distribution. A characteristic property of scale-free networks is that there tend to be a large number of nodes with low degree, and a small number of highly connected hub nodes. Many biological networks (including protein interaction and metabolic networks) resemble-scale free networks (Albert, 2005). We generate random graphs separately for each group using the R package `igraph` (Csárdi and Nepusz, 2006). Since we would like to measure the performance of learning hub nodes, and there is no clear distinction between hub and non-hub nodes in random networks, we select graphs that include one or two nodes that are connected to more than 10 other nodes out of $p = 60$, and define these nodes as hubs. To create common hub nodes, we permute the node index so that highly connected nodes have a common index across graphs. The nonzero elements of the precision matrices $\mathbf{\Omega}_k$ for each graph $\mathbf{G}_k$, $k = 1, \ldots, 3$ are sampled from the uniform distribution on $\{[-0.6, -0.4] \cup [0.4, 0.6]\}$. In order to make the matrices positive definite, following Mohan et al. (2014), we average the matrix with its transpose to obtain $\bar{\mathbf{E}}_k$, and then update the diagonal elements by setting $\mathbf{\Omega}_k$ to $\bar{\mathbf{E}}_k + (0.1 - \Lambda_{\min}(\bar{\mathbf{E}}_k))\mathbf{I}$, where $\Lambda_{\min}(\bar{\mathbf{E}}_k)$ is the smallest eigenvalue of $\bar{\mathbf{E}}_k$ for $k = 1, 2, 3$. Then we scale each $\mathbf{\Omega}_k$ to have diagonal elements 1. We generate random normal data using $\mathbf{\Omega}_1, \mathbf{\Omega}_2$, and $\mathbf{\Omega}_3$ as the true precision matrices for sample size $n = 100$.

Our simulation results are averaged over 25 simulated data sets. We apply each method with the same parameter settings and procedures as in the illustrative example given in Section 5. For both Bayesian methods, we use 2000 iterations of burn-in followed by 10,000 iterations as the basis for posterior inference.

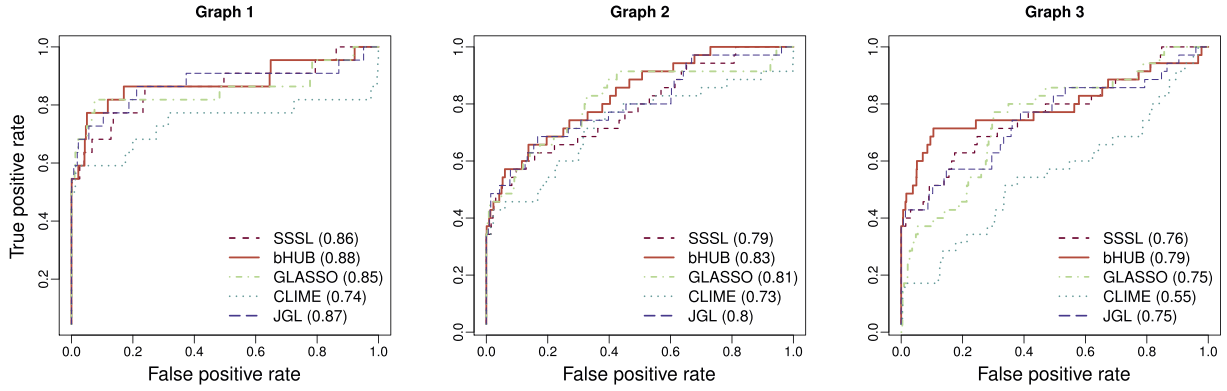
6.2. *Results*

We average the estimates from the 25 simulated datasets for each simulation setting. The results for *Setting 1* (scale-free networks) are given in Figure 5. Those for *Setting 2* (networks with shared edges) and *Setting 3* (community networks) are presented in Figures S7 and S8 (Supplementary Materials). In each figure, the top panel (a) presents ROC curves with AUCs for graph structure learning, the middle panel (b) compares the performance for identifying the edges of hub nodes, and the bottom panel (c) presents whisker plots to summarize posterior inference on the hub status of nodes using bHUB. Although many of the methods performed well in terms of graph structure learning, bHUB achieve the highest average AUC across the settings considered. Moreover, bHUB show improved power in learning the

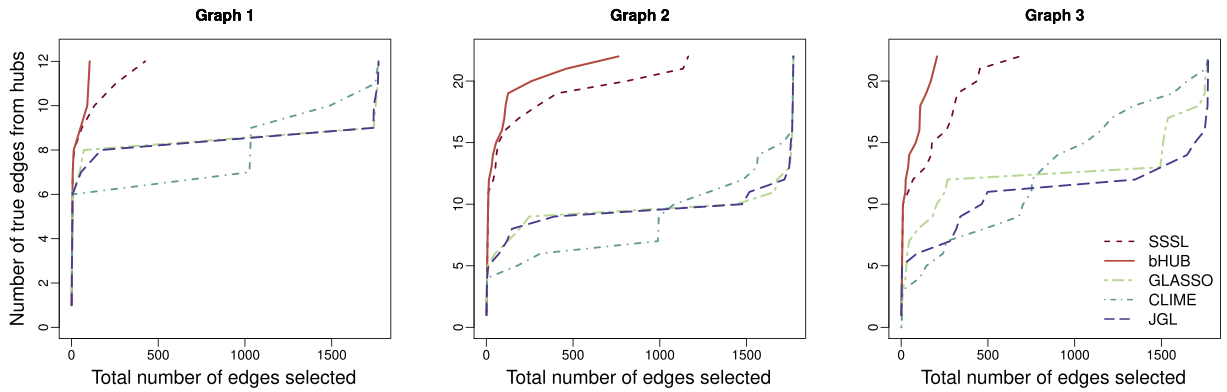


(a) *Setting 1.* Networks for scale free graphs

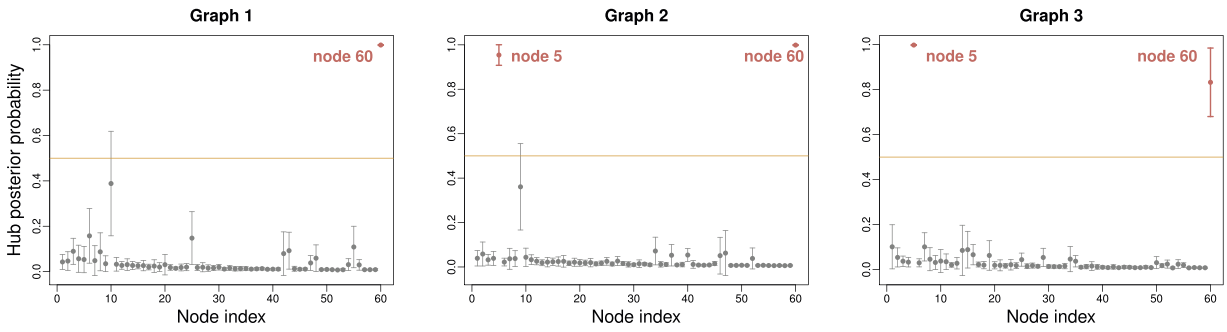
Figure 4. True graph structure for each simulation setting described in Section 6 for *Setting 1* scale free graphs. Node 60 is a true common hub node in all graphs, and node 5 is a true common hub node for Graph 2 and Graph 3.



(a) ROC curves for graph learning



(b) Performance for hub identification



(c) Posterior probability of being a hub

Figure 5. *Setting 1.* Results averaged over 25 simulated datasets. The top panel (a) shows ROC curves for identifying each graph structure, with the area under curve (AUC) presented in parentheses. The middle panel (b) shows performance for identifying true edges from hub nodes: Graph 1 has one hub with 12 edges, while Graph 2 and 3 have two hub nodes with 23 true positive edges respectively. The bottom panel (c) shows whisker plots of the posterior probability of each node being a hub: node 60 is a true co-hub node in all graphs, and node 5 is a true co-hub for Graph 2 and Graph 3. The horizontal line represents the posterior probability 0.5.

edges of highly connected nodes, which are a common feature of biological networks. Finally, bHUB achieve a good separation of the posterior probabilities for hub versus non-hub nodes, demonstrating its utility in providing inference for the identification of hub nodes. As shown in panel (c) of Figure 5, bHUB correctly identifies the true hub nodes in all graphs.

7. Application to Learning Subtype-Specific Networks for Ovarian Cancer

In this section, we illustrate the application of our method to inference of co-expression networks across related sample groups. We consider gene expression data for ovarian carcinoma from The Cancer Genome Atlas (TCGA). It has been demonstrated that high-grade serous ovarian cancer

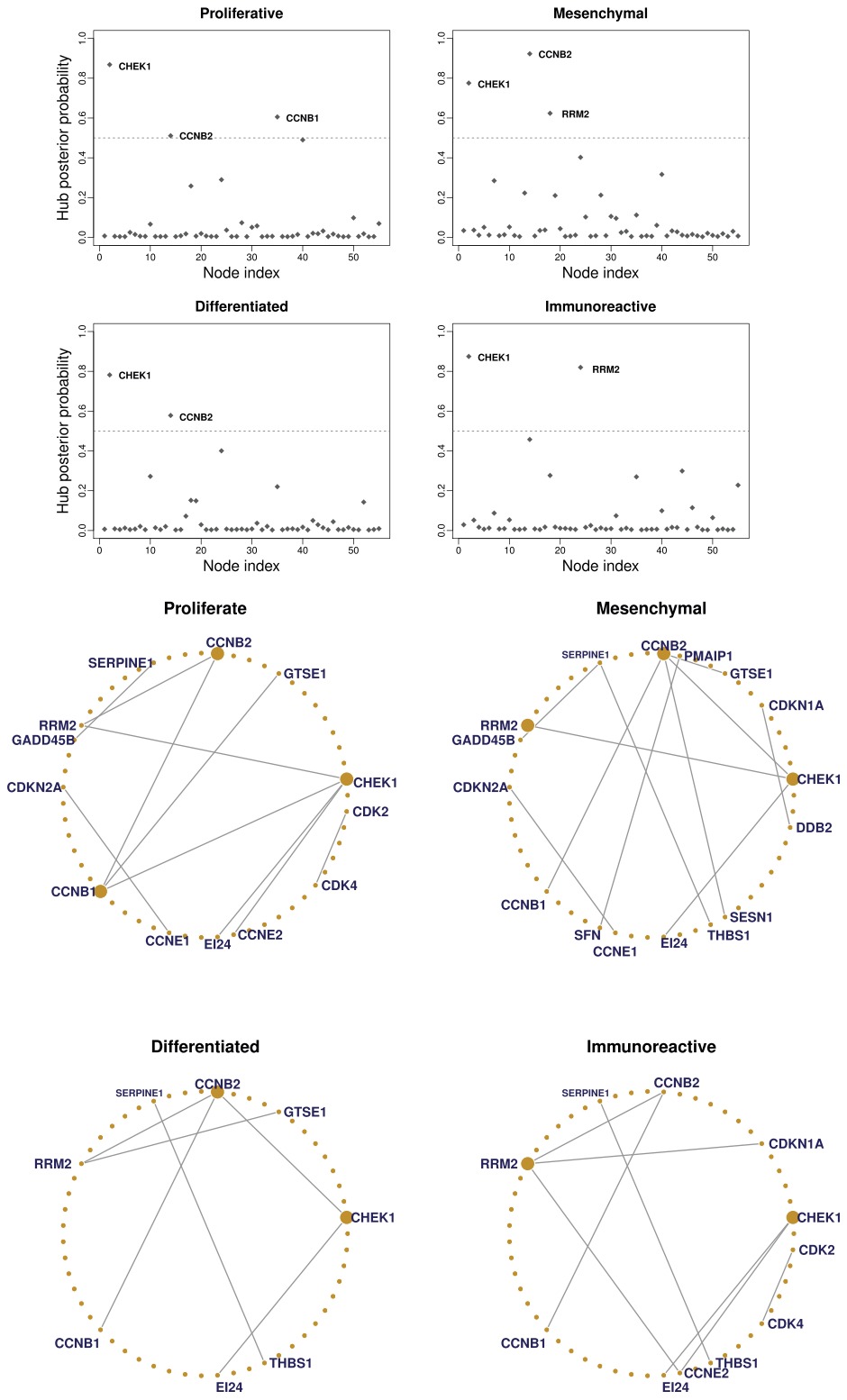


Figure 6. P53 signaling pathway: Inferred posterior probability of being a hub for each node and inferred co-expression networks for the ovarian cancer subtypes. CHEK1 (node 2) is a co-hub for all subtypes. CCNB2 (node 14) is a co-hub gene for proliferative, mesenchymal, and differentiated subtypes, while the mesenchymal and immunoreactive subtypes share a co-hub RRM2 (node 19).

(HGS-OvCa) patients can be classified into four subtypes: proliferative, mesenchymal, differentiated, and immunoreactive (The Cancer Genome Atlas Research Network, 2011). In the TCGA data set, these subtypes have sample sizes of 138, 109, 135, and 107 per group, respectively. Of the 11,864 genes included in the data set, we considered three pathways for network inference, selected for their relevance to ovarian cancer: the P53 signaling pathway, the HR pathway, and the TGF-beta pathway. Based on pathway information from the KEGG database (Kanehisa et al., 2017), these pathways include 55, 26, and 74 genes, respectively. Alterations to the P53 signaling pathway have been linked to cancer pathogenesis and resistance to therapy, and are common in ovarian cancer (Stegh, 2012). The functional status of the homologous recombination (HR) pathway, which is critical in DNA damage repair, has been shown to predict drug response in ovarian cancer (De Picciotto et al., 2016), and recent studies have identified potential therapeutic targets in ovarian cancers that harbor defects in the HR pathway (Ledermann et al., 2016; Konstantinopoulos et al., 2015). Finally, the TGF-beta pathway is also known to modulate the severity of ovarian cancer (Yamada et al., 1999; Baldwin et al., 2003; Yeung et al., 2013).

We applied bHUB using the same parameter settings as in the simulation studies, other than the prior on β_2, which was adjusted based on the number of input variables (see Sections 3.5 and S2 in the Supplementary Materials for guidance on the choice of parameter settings and sensitivity.) For each pathway, we present the resulting posterior probabilities of being a hub for each node in each subgroup, and the inferred gene co-expression network based on the posterior median graph. Figure 6 illustrates the results for the P53 signaling pathway. bHUB identifies the gene CHEK1 (node 2) as a co-hub gene for all subtypes. Bryant et al. (2014), Kobayashi et al. (2015), and Kim et al. (2016) highlight CHEK1 as a therapeutic target for ovarian cancer. In addition, we identify CCNB2 (node 14) as a co-hub shared by the proliferative, mesenchymal, and differentiated subtypes, and RRM2 (node 19) as a co-hub gene for the mesenchymal and immunoreactive subtypes. Fu and Wang (2013) also found that CCNE1 and CCNB2 are significant genes in the P53 pathway. Zhou et al. (2013) showed that inhibition of RRM2 reduce the proliferation of ovarian cancer. We provide a full description of the results for the HR and TGF-beta pathways, including figures and interpretation of the resulting networks, in Section S7 (Supplementary Materials, p. 14–18), where we also provide a comparison with results obtained using separate Bayesian graphical model inference (SSSL). In summary, these results demonstrate the utility of bHUB in learning networks with biologically interpretable features. The identified hub nodes have critical importance to ovarian cancer phenotypes such as proliferation and survival. In general, our results suggest the potential utility of bHUB in identifying treatment targets within biological networks.

8. Discussion

In this work, we have developed a novel modeling approach to hub-based inference of multiple graphs. The proposed model uses a latent indicator for each node to represent its status as a hub. By linking these indicators through a Bayesian hierarchical prior, we are able to encourage co-hubs between groups, and also provide inference on hub nodes, which may be preserved or perturbed across settings, and are of particular interest in biological networks. The hub status of each node reflects the uncertainty over the inclusion of specific edges; this is in contrast to assessing the node degree in the posterior selected graph, which is conditional on the model selection.

Through extensive simulation studies on various types of networks, we have shown that our method offers good performance for structure learning and improved power for discovering edges from highly connected hub nodes. Moreover, we provide accurate inference for the discovery of hub nodes. The practical utility of the method was illustrated through an application to co-expression networks across various subtypes of ovarian cancer, where hub nodes are of particular interest as treatment targets, and understanding whether a gene is functioning as a hub for a particular subtype of disease could potentially provide guidance on selection of the best therapy.

9. Supplementary Materials

Supplementary materials, including details of the MCMC sampling, sensitivity analysis, convergence diagnostics, and the additional simulation and case study results referenced in Sections 3–7, as well as Matlab code implementing the proposed procedure, are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

The authors acknowledge the use of data generated by the TCGA Research Network: `http://cancergenome.nih.gov/`. CBP and KAD are partially supported by NIH/NCI CCSG grant P30CA016672 and by the MD Anderson CLL Moon Shot 710499-80-111996-21. KAD is partially supported by EDRN grant U24CA086368-16, SPORE grant P50CA140388-06A1, and CCTS grant 5UL1TR000371-07. JK is partially supported by start-up and incentive funds to KAD and CBP. MJH is partially supported by NIH grant 1R21CA220299-01A1.

References

Albert, R. (2005). Scale-free networks in cell biology. *Journal of Cell Science* **118**, 4947–4957.

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.

Baldwin, R., Tran, H., and Karlan, B. (2003). Loss of c-myc repression coincides with ovarian cancer resistance to transforming growth factor β growth arrest independent of transforming growth factor β/smad signaling. *Cancer Research* **63**, 1413–1419.

Barabasi, A. and Oltvai, Z. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101–113.

Barabasi, A. L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics* **12**, 56–68.

Bryant, C., Rawlinson, R., and Massey, A. (2014). Chk1 inhibition as a novel therapeutic strategy for treating triple-negative breast and ovarian cancers. *BMC Cancer* **14**, 570.

Cai, T., Kang, J., Kemmer, P., and Guo, Y. (2011). A constrained $\ell 1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106**, 594–607.

Csárdi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1–9.

Danaher, P., Wang, W., and Witten, D. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal Royal Statistics Society, Series B* **76**, 373–397.

De Picciotto, N., Cacheux, W., Roth, A., Chappuis, P., and Labidi-Galy, S. (2016). Ovarian cancer: Status of homologous recombination pathway as a predictor of drug response. *Critical Reviews in Oncology Hematology* **101**, 50–59.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–175.

Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441.

Fu, L. and Wang, B. (2013). Investigation of the hub genes and related mechanism in ovarian cancer via bioinformatics analysis. *Journal of Ovarian Research* **6**, 92.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

Ha, M., Baladandayuthapani, V., and Do, K. (2015). Dingo: Differential network analysis in genomics. *Bioinformatics* **31**, 3413–3420.

Jeong, H., Mason, S. P., Barabási, A.-L., and Oltvai Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41–42.

Jonsson, P. and Bates, P. (2006). Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22**, 2291–2297.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases, and drugs. *Nucleic Acids Research* **45**, D353–D361.

Kim, H., George, E., Ragland, R., Rafail, S., Zhang, R., Krepler, C., et al. (2016). Targeting the ATR/CHK1 axis with PARP inhibition results in tumor regression in BRCA mutant ovarian cancer models. *Clinical Cancer Research* **23**, 3097–3108.

Kobayashi, H., Shigetomi, H., and Yoshiomoto, C. (2015). Checkpoint kinase 1 inhibitors as targeted molecular agents for clear cell carcinoma of the ovary (review). *Oncology Letters* **10**, 571–576.

Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I., and D'Andrea, A. D. (2015). Homologous recombination deficiency: Exploiting the fundamental vulnerability of ovarian cancer. *European Journal Cancer* **5**, 1137–1154.

Lauritzen, S. (1996). *Graphical Models*. Oxford: Clarendon Press.

Ledermann, J., Drew, Y., and Kristeleit, R. (2016). Homologous recombination deficiency and ovarian cancer. *European Journal Cancer* **60**, 49–58.

Mitra, R., Müller, P., and Ji, Y. (2016). Bayesian graphical models for differential pathways. *Bayesian Analysis* **11**, 99–124.

Mohan, K., London, P., Fazel, M., Witten, D., and Lee, S. (2014). Node-based learning of Gaussian graphical models. *Journal of Machine Learning Research* **15**, 445–488.

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

Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics* **29**, 391–411.

Stegh, A. (2012). Targeting the p53 signaling pathway in cancer therapy - the promises, challenges, and perils. *Expert Opinion on Therapeutic* **16**, 67–83.

Tan, K., London, P., Mohan, K., Lee, S., Fazel, M., and Witten, D. (2014). Learning graphical models with hubs. *Journal of Machine Learning Research* **15**, 3297–3331.

The Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615.

Wachi, S., Yoneda, K., and Wu, R. (2005). Interactome- transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21**, 4205–4208.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis* **7**, 771–790.

Wang, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Analysis* **10**, 351–377.

Yamada, S., Baldwin, R., and Karlan, B. (1999). Ovarian carcinoma cell cultures are resistant to TGF-β1-mediated growth inhibition despite expression of functional receptors. *Gynecologic Oncology* **75**, 72–77.

Yeung, T., Leung, C., Wong, K., Samimi, G., Thompson, M., Liu, J., et al. (2013). TGF-β modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Research* **73**, 5016–5028.

Zhou, B., Su, L., Hu, S., Hu, W., Yip, M., Wu, J., et al. (2013). A small-molecule blocking ribonucleotide reductase holoenzyme formation inhibits cancer cell growth and overcomes drug resistance. *Cancer Research* **17**, 6484–6493.

Received July 2017. *Revised March* 2018. *Accepted July* 2018.