A Bayesian Approach for Learning Gene Networks Underlying Disease Severity in COPD

Elin Shaddox[1], **Francesco C. Stingo**[2], **Christine B. Peterson**[3], **Sean Jacobson**[4], **Charmion Cruickshank-Quinn**[5], **Katerina Kechris**[6], **Russell Bowler**[4], **Marina Vannucci**[1]

[1]Department of Statistics, Rice University, Houston, USA

[2]Dipartimento di Statistica, Informatica, Applicazioni "G.Parenti", University of Florence, Florence, Italy

[3]Department of Biostatistics, UT MD Anderson Cancer Center, Houston, USA

[4]Department of Medicine, National Jewish Health, Denver, CO, USA

[5]Department of Pharmaceutical Sciences, School of Pharmacy, University of Colorado Denver, Denver, CO, USA

[6]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Denver, CO, USA

Abstract

In this paper, we propose a Bayesian hierarchical approach to infer network structures across multiple sample groups where both shared and differential edges may exist across the groups. In our approach, we link graphs through a Markov random field prior. This prior on network similarity provides a measure of pairwise relatedness that borrows strength only between related groups. We incorporate the computational efficiency of continuous shrinkage priors, improving scalability for network estimation in cases of larger dimensionality. Our model is applied to patient groups with increasing levels of chronic obstructive pulmonary disease severity, with the goal of better understanding the break down of gene pathways as the disease progresses. Our approach is able to identify critical hub genes for four targeted pathways. Furthermore, it identifies gene connections that are disrupted with increased disease severity and that characterize the disease evolution. We also demonstrate the superior performance of our approach with respect to competing methods, using simulated data.

Keywords

Gaussian graphical model; Bayesian inference; Markov random field prior; Spike-and-slab prior; Gene network; Chronic obstructive pulmonary disease (COPD)

Francesco C. Stingo, f.stingo@disia.unifi.it.

1 Introduction

1.1 General Motivation for Network Analysis in Genomics

Bayesian hierarchical models are becoming increasingly popular for inference with genomic data. These methods are powerful tools to understand the structure of complex diseases and to evaluate patterns of variable association, particularly for the analysis of studies with a small sample size. As complex diseases are multi-level illnesses defined by changes at the cellular level [17,32], we can apply network-based inference to genes and their products in order to better understand the underlying biological mechanisms and thereby develop more targeted treatments. In order to accomplish this goal, it is important to develop flexible and computationally efficient models which can adequately analyze the dependence structure of these highly dimensional datasets. A common approach to describe conditional dependence relationships of random variables is graphical models, which have been successfully applied to protein–protein interaction, co-expression, and gene regulatory networks [10,25,41,42].

1.2 Introduction to Statistical Methods for Networks Analysis

Bayesian approaches to network estimation have been found to be successful for both decomposable and unrestricted graphical models. These approaches have the critical advantage of quantifying the uncertainty associated to network estimation. For the decomposable setting, implementation of hyper-inverse Wishart priors enables the development of efficient stochastic search procedures to estimate network structure. [7] used this approach to determine explicitly the marginal likelihoods of the graph. This method was extended to Bayesian variable selection for both high-dimensional decomposable and nondecomposable undirected Gaussian graphical models [19]. [35] described a feature-inclusion stochastic search algorithm, which uses online estimates of edge-inclusion probabilities to guide Bayesian model determination for decomposable Gaussian graphical models. When compared to Markov Chain Monte Carlo, Metropolis-based searches, and lasso methods, their algorithm was found to be superior in both speed and stability.

In the context of biological networks, it is often inappropriate to restrict the model space to only decomposable graphs [26]. Efficient and flexible Bayesian methods for nondecomposable Gaussian graphical models were proposed using the G-Wishart prior by [2,11]. Rather than computing the normalizing constant of marginal likely-hoods analytically, as is the case for decomposable graphs, Markov Chain Monte Carlo methods are used to sample over the joint spaces of precision matrices and graphs in order to avoid posterior normalizing constant computation. Further improvement was proposed by [45] with the implementation of a new exchange algorithm requiring neither proposal tuning nor evaluation of normalizing constants for the G-Wishart distribution. Reduced computational complexity and greater flexibility in prior specification was described by [39] with graph theory results for local updates that facilitate fast exploration of the graph space.

In recent years, such evidence of successful single network structure estimation has led to extensions of the methods to inference for multiple graphical models. Approaches for multiple graphical models are particularly appropriate when the biological network evolves with respect to clinical features, such as disease stage. [16] extended the graphical lasso to

multiple undirected graphs sharing the same variables with similar dependence structures. They propose a method which preserves common structure while allowing for differences through a hierarchical penalty targeting removal of common zeros in the precision matrices. [9] proposed the more general joint graphical lasso approach based on maximizing a penalized log likelihood. Their approach explores the properties of two penalty structures: the fused graphical lasso encouraging shared edge values and shared structure, and the group graphical lasso which supports shared structure but not shared edge values. [46] described a Bayesian approach assessing heterogeneous patterns of association between Gaussian directed graphs for related samples. Another Bayesian approach was proposed by [29] linking graph structure estimation with a Markov Random Field prior favoring an edge if the same edge is included in related sample group graphs. In their method, subgroups are not assumed related and shared structure is learned by defining a spike-and-slab prior on network relatedness parameters.

Computational burden is a major challenge for Bayesian graphical models, motivating the development of methods which are more efficient and have greater scalability. Successful developments in related problems have come from the use of continuous shrinkage priors. [13] used these priors in the form of a two-component normal mixture model in regression analysis. These priors have received even more attention as alternatives for regularizing regression coefficients, see [1,15,27]. When used in estimating covariance matrices through regularizing concentration elements, continuous shrinkage priors have been shown to result in fast and accurate estimation [21,43]. [44] developed a stochastic search structure learning algorithm for undirected graphical models. His method uses continuous shrinkage priors indexed by latent binary indicators, and allows for efficient block updates of the network parameters.

In this paper, we propose a new approach for multiple network analysis which builds on earlier methods [29] by improving scalability with a continuous shrinkage prior in the spirit of [44]. This results in a computationally more efficient approach that can be applied to larger networks. In particular, our work is motivated by the problem of analyzing network evolution of gene networks underlying the complex chronic obstructive pulmonary disease (COPD). Our paper is organized as follows: Section 2 provides a description of our motivating problem and the details of the dataset we apply our method to. Section 3 presents an introduction to Bayesian graphical models and introduces our proposed method, the prior models and the method for posterior inference—in addition to an outline of our Markov chain Monte Carlo method. Section 4 outlines our simulation studies, and section five describes the application of our method to four selected gene pathways involved in COPD. Section 6 concludes the paper.

2 The ECLIPSE COPD Cohort Study

Chronic obstructive pulmonary disease (COPD) is the 3rd leading cause of death in the US [37] and acute exacerbations of COPD (AECOPDs) are the 2nd leading cause of hospital stays [28,37]. Although 90 % of COPD patients are smokers, about 75 % of smokers do not develop COPD. There is a poor understanding of the risk factors that account for disease

susceptibility or resistance to cigarette smoke (CS), as well as of the pathogenic mechanisms underlying the development of emphysema and airway inflammation.

Whole-blood gene expression data from 226 subjects were generated within the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) cohort using the Affymetrix Human Genome U133 Plus 2.0 Array and are available at NCBI GEO GSE22148 [12,36]. Raw data (CEL files) were log-transformed and normalized using the RMA method [18] in the affy R package. Probesets were filtered so that there were present calls in all samples for a final set of 12525 probesets.

Subjects were classified into four groups by severity of radiologic emphysema, a subtype of COPD; [0–5) percent emphysema ($n = 61$), [5–10) percent emphysema ($n = 43$), [10,20) percent emphysema ($n = 46$), and [20+] percent emphysema ($n = 51$). Twenty-five subjects had missing values for percent emphysema and were not used in subsequent analyses.

We examined four candidate pathways that were selected based on the analysis of genomic and metabolomic data from an independent study on the genetic epidemiology of COPD called COPDGene [31]. In the COPDGene cohort, gene expression data from peripheral blood mononuclear cells (PBMCs) were generated on 131 subjects using the same Affymetrix platform as the ECLIPSE data [3]. On those same subjects, plasma metabolite abundance was generated using liquid chromatography/mass spectrometry [4]. Differently expressed genes and differently abundant metabolites were identified for airflow obstruction (FEV1pp forced expiratory volume in 1 second percent predicted) correcting for age, sex, body mass index, and current smoking status. KEGG pathways [20] that showed enrichment of the significant genes and metabolites were prioritized and the top four candidate pathways were used to explore their role in emphysema for this work: glycerophospholipid metabolism (GPL), oxidative phosphorylation (OxPhos), regulation of autophagy (RegAuto), and Fc γ R-mediated phagocytosis (FcyR). For each of the pathways, there were 60 (GPL), 83 (OxPhos), 28 (RegAuto), and 104 (FcyR) probesets that were collapsed to 41 (GPL), 62 (OxPhos), 20 (RegAuto), and 57 (FcyR) unique genes by selecting the probeset with the strongest association with emphysema. These 4 pathways may play a role in the response to cigarette smoke exposure and are interesting candidates for more detailed exploration in emphysema.

3 Proposed Method

The goal of our work is to model and infer the network structure of multiple pathways relevant to COPD. For each pathway, we are interested in understanding how connections between genes break down between the four sample groups defined by the severity of emphysema. We achieve this via a Bayesian hierarchical model, described in 3.1 and 3.2, that allows to jointly estimate a separate network for each group while comparing networks across groups in order to determine pairwise relatedness.

For each given pathway, we observe four $n_k \times p$ data matrices X_k, where $k = 1, \ldots, K = 4$ indexes the group, n_k is the sample size for group k, and p is the total number of genes in the

pathway. Assuming samples are independent and identically distributed within each of the K groups, we can write the likelihood for subject i in group k as

$$X_{k,i} \sim N\left(\mu_k, \Omega_k^{-1}\right), \quad i = 1, ..., n,$$

where $\mu_k \in \mathbb{R}^P$ is the mean vector for group k and $\Omega_k = \Sigma_k^{-1} = \left(\omega_{i,j,k}\right)$ is the precision matrix for group k, a symmetric positive definite matrix constrained to a set of restrictions $\omega_{i,j,k} = 0$, as defined by a graph G_k which is an undirected graphical model representing the conditional dependence relationships existing between the p genes. Each G_k is a mathematical object consisting of two sets, vertices $V = \{1, ..., p\}$ and edges $E \in V \times V$, so $G = (V, E)$. In an undirected graph, an edge exists between vertices i and j if $(i, j) \in E$ and $(j, i) \in E$. In the context of our application, each vertex in G_k corresponds to a gene. An edge is included in the network if the two corresponding genes are conditionally dependent, while the absence of an edge between two vertices means the two corresponding genes are conditionally independent given the remaining genes. For each group k, graph G_k can be thought of as a symmetric binary matrix where each off-diagonal element $g_{k,i,j}$ denotes the inclusion of edge (i, j) in G_k.

3.1 Continuous Shrinkage Prior

In the context of Bayesian analysis of large networks, one of the main challenges is to define a prior distribution on Ω_k. The most common approach is to assign a G-Wishart prior, which is the Wishart distribution restricted to the space of precision matrices where zeros are specified by either a decomposable or nondecomposable graph [33]. While this provides a flexible formulation for modeling, both the prior and posterior normalizing constants are intractable, limiting the method in scalability and computation. We address these difficulties with a recent approach that overcomes these issues, and propose for each network a continuous shrinkage prior as defined by [44]. Let $\Omega k = \left(\omega_{i,j,k}\right)_{p \times p}$ be the p-dimensional concentration matrix for gene interactions for each group. Our prior is a product of $p(p-1)/2$ two-component normal mixture densities, on the off-diagonal elements, and p exponential densities, on the diagonal elements, of the type

$$p(\Omega_k | G) = \{C(\theta)\}^{-1} \prod_{i < j} \left\{(1 - \pi)N\left(\omega_{i,j} | 0, v_0^2\right) + \pi N\left(\omega_{i,j} | 0, v_1^2\right)\right\}$$
$$\times \prod_i \mathrm{Exp}\left(\omega_{i,i} | \frac{\lambda}{2}\right) I_{\Omega_k \in M^+} \propto \prod_{i < j} N\left(\omega_{i,j} | 0, v_{g_{i,j}}^2\right) \prod_i \mathrm{Exp}\left(\omega_{i,i} | \frac{\lambda}{2}\right),$$

Where $v_{g_{i,j}}^2 = v_1^2$ if edge (i, j) is a connection in the network, i.e., $g_{i,j} = 1$, and $v_{g_{i,j}}^2 = v_0^2$ if $g_{i,j} = 0$ and the connection is not in the network. The two-component normal mixture model has been shown to be a successful prior in the context of variable selection, which in our case is equivalent to edge selection, and the choice of hyperparameters v_0^2 and v_1^2 has been closely studied by George and McCulloch (1993, 1997). The hyperparameter spaces for $\theta = \{v_0, v_1, \pi, \lambda\}$ are $v_0 > 0$, $v_1 > 0$, $\lambda > 0$, and $\pi \in (0,1)$ and v_0 and v_1 can be set as either small or large, resulting in a spike-and-slab prior. If for example, v_0 is chosen to be small,

the event $g_{i,j} = 0$ indicates that the edge $\omega_{i,j}$ comes from the $N\left(0, v_0^2\right)$ or diffuse component of the mixture, and consequently $\omega_{i,j}$ is closer to zero and can be estimated as zero. In contrast, if v_1 is chosen to be large, the event $g_{i,j} = 1$ means $\omega_{i,j}$ comes from the other component $N\left(0, v_1^2\right)$ and $\omega_{i,j}$ can then be thought of as substantially different from zero. $C(\theta)$ and the indicator function ensure that the density function integrated over the space M^+ is one. We define this prior by introducing binary latent variables which can be viewed as edge-inclusion indicators $\mathbf{G} \equiv \left(g_{ij}\right)_{i<j} \in \mathscr{G} \equiv \{0,1\}^{p(p-1)/2}$, creating a hierarchical model defined by

$$p(\Omega_k | \mathbf{G}, \theta) = \{C(\mathbf{G}, v_0, v_1, \lambda)\}^{-1} \prod_{\mathbf{i<j}} \mathbf{N}\left(\omega_{\mathbf{i,j}} | 0, v_{\mathbf{g_{i,j}}}^2\right) \prod_{\mathbf{i}} \mathrm{Exp}\left(\omega_{\mathbf{i,i}} \frac{\lambda}{2}\right).$$

and by the prior $p(\mathbf{G}|\theta)$, that is outlined in Sect. 3.2. The constant $C(\mathbf{G}, v_0, \ v_1, \lambda) \in (0, 1)$ is a normalizing constant which ensures proper distributions. Further details on this constant can be found in [44].

3.2   Linking Graphs with a Markov Random Field Prior

To encourage selection of similar edges in related graphs, we define a Markov random field (MRF) prior on the graph structures. In Bayesian variable selection, MRF priors have been used to model dependencies between covariates in regression models [23,30,40]. Our prior follows a similar structure, but it is imposed on the indicators of edge inclusion contrary to indicators of variable inclusion. Each random variable in the set $\mathbf{g_{i,j}} = \{g_{1,i,j}, \ldots, g_{k,i,j}\}$ is then binary and an indicator of edge inclusion within the model. Consequently, each $g_{k,i,j}$ could be modeled by a Bernoulli prior. If all $g_{k,i,j}$ were independent, a product of Bernoulli distributions could be used to model this binary vector. A MRF prior is introduced to capture and model the dependence structure between these binary random variables. A MRF distribution can be seen as a generalization of a set of independent Bernoulli distributions in a multivariate setting. For the binary vector of edge-inclusion indicators $\mathbf{g_{i,j}} = (g_{1,i,j}, \ldots, g_{k,i,j})^T$ where $1 \quad i<j \quad p$, we define a MRF prior distribution as

$$p\left(\mathbf{g_{ij}} | v_{i,j}, \Theta\right) = C\left(v_{i,j}, \Theta\right)^{-1} \exp\left(v_{i,j} \mathbf{1}^T \mathbf{g_{i,j}} + \mathbf{g_{i,j}}^T \Theta \mathbf{g_{i,j}}\right),$$

where $v_{i,j}$ is a specific parameter for each set of edges $\mathbf{g_{i,j}}$, Θ is a $K \times K$ symmetric matrix denoting pairwise relatedness for each sample group's graph, and $\mathbf{1}$ is the unit vector of dimension K. The off-diagonal elements of Θ, θ_{km}, allow us to share information between sample groups k and m, when appropriate, as well as to obtain a measure of relative network similarity across groups. The normalizing constant is defined as

$$C\left(v_{i,j}, \Theta\right) = \sum_{\mathbf{g_{i,j}} \in \{0,1\}^K} \exp\left(v_{i,j} \mathbf{1}^T \mathbf{g_{i,j}} + \mathbf{g_{i,j}}^T \Theta \mathbf{g_{i,j}}\right).$$

As long as the number of sample groups K is reasonably small, the computation of the normalizing constant is straightforward. From the probability of the binary vector of edge-

inclusion indicators, we can see that the prior probability of an edge (i, j) being absent from all K graphs is $p(\mathbf{g_{i,j}} = 0 | v_{i,j}, \Theta) = 1/C(v_{i,j}, \Theta)$.

The joint prior on the graphs $(G_1, \ldots, G_K)$ is the product of the densities for each edge

$$p(G_1, \ldots, G_k | v, \Theta) = \prod_{i < j} p(\mathbf{g_{ij}} | v_{i,j}, \Theta),$$

where $v = \{ v_{i,j} | 1 \quad i < j \quad p \}$. The conditional probability of the inclusion of edge (i, j) in G_k, given the inclusion of that edge in all remaining graphs, is then

$$p(g_{k,i,j} | \{g_{m,i,j}\}_{m \neq k}, v_{i,j}, \Theta) = \frac{\exp(g_{k,i,j}(v_{i,j} + 2\sum_{m \neq k} \theta_{k,m} g_{m,i,j}))}{1 + \exp(v_{i,j} + 2\sum_{m \neq k} \theta_{k,m} g_{m,i,j})}.$$

We also define prior distributions on v and Θ to reduce false selection of edges to account for a lack of correction for multiple testing from a fixed prior probability of inclusion, as noted by [34] in the setting of variable selection. This approach also allows us to obtain posterior estimates of these parameters, reflecting more information learned from the data.

3.3 Prior on Network Similarity

We are interested in a measure of network similarity that characterizes the relatedness of the gene network between the disease subgroups, and allows us to study the disruption and conservation of gene pathways as COPD evolves.

We define Θ as a $K \times K$ symmetric super-graph with nonzero off-diagonal elements $\theta_{k,m}$ capturing similarity between group k and group m. Consequently, the magnitude of $\theta_{k,m}$ indicates pairwise similarity of the two graphs G_k and G_m. Then, we select our prior following [29] as a spike-and-slab prior on the off-diagonal entries $\theta_{k,m}$. Because we want the "slab" portion defined on the positive domain to comply with positive values of $\theta_{k,m}$ for related networks, we desire a density in the positive domain that allows discrimination between zero and nonzero values. Since the Gamma$(x | \alpha, \beta)$ probability density function is equal to 0 at $x = 0$ and is nonzero for $x > 0$ and $\alpha > 1$, it is an appropriate choice for the density of the slab portion of our mixture model prior. Our prior on the network relatedness parameters is then defined as

$$p(\theta_{k,m} | \gamma_{k,m}) = (1 - \gamma_{k,m})\delta_0 + \gamma_{k,m} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_{k,m}^{\alpha - 1} e^{-\beta \theta_{k,m}},$$

with fixed hyperparameters α and β and latent indicator variable $\gamma_{k,m}$ which indicates the event that graph k is related to graph m. We define an independent Bernoulli prior on the $\gamma_{k,m}$s, with hyperparameter $w \in [0,1]$,

$$p(\gamma_{k,m} | w) = w^{\gamma_{k,m}} (1 - w)^{(1 - \gamma_{k,m})}.$$

This prior borrows strength between groups when appropriate without enforcing similarity if groups have different network structures. Our joint priors for the off-diagonal entries of the super-graph and for γ are then

$$p(\Theta \mid \gamma) = \prod_{k < m} p(\theta_{k,m} \mid \gamma_{k,m}),$$

$$p(\gamma) = \prod_{k < m} p(\gamma_{k,m} \mid w).$$

3.4 Edge-Specific Prior

We can specify a prior for the edge-inclusion probability v_{ij} to encourage sparsity of the graphs $G_1, \ldots, G_k$. This same prior can be used in order to incorporate prior knowledge of connections between genes. Negative values of $v_{i,j}$ will reduce the prior probability of inclusion for edge (i, j) in all graphs G_k, and consequently a prior favoring smaller values of v will lead to a preference for model sparsity, which can be attractive in applications where it is beneficial to reduce the number of parameters and make results more interpretable. In contrast, larger values of $v_{i,j}$ make edge (i, j) more likely to be selected according to whether or not it has been selected in other graphs. If we are given a known reference network, say G_0, we can use this network to define a prior which encourages higher selection probabilities for those edges in G_0. If $\theta_{k,m} = 0$ for all $m \neq k$, or if for nonzero $\theta_{k,m}$ no edges $g_{m,i,j}$ are selected, the probability of inclusion of edge (i, j) in G_k can be written as

$$p(g_{k,i,j} \mid v_{i,j}) = \frac{e^{v_{i,j}}}{1 + e^{v_{i,j}}} = q_{i,j}.$$

Then we can impose a prior on $q_{i,j}$ which reflects the belief that graphs having similarities to the reference network $G_0 = (V, E_0)$ are more likely than those with differing edges

$$q_{i,j} = \begin{cases} \text{Beta}(1 + c, 1) & \text{if } (i, j) \in E_0 \\ \text{Beta}(1, 1 + c) & \text{if } (i, j) \notin E_0, \end{cases}$$

where $c > 0$. Then, because $v_{i,j} = \text{logit}(q_{i,j})$, we can apply a univariate transformation to the Beta(a, b) prior on $q_{i,j}$ to write the prior on $v_{i,j}$ as

$$p(v_{i,j}) = \frac{1}{B(a,b)} \frac{e^{a v_{i,j}}}{\left(1 + e^{v_{i,j}}\right)^{a+b}}$$

If dealing with a case where there is no prior knowledge on the graph structure, one can choose a prior favoring lower values to encourage sparsity, such as $q_{i,j} \sim \text{Beta}(1,4)$ for all edges (i, j). In the case where most edges are believed missing for all graphs but those edges present in one graph tend to be included in all other graphs, a prior favoring larger values for $\theta_{k,m}$ can be chosen with a prior favoring smaller values for $v_{i,j}$.

3.5   Posterior Inference

Defining Ψ as the set of all parameters and X as our observed data for all sample groups, our joint posterior is

$$p(\Psi \mid X) \propto \prod_{k=1}^{4} [p(X_k \mid \mu_k, \Omega_k) p(\mu_k \mid \Omega_k) p(\Omega_k \mid G_k)]$$
$$\prod_{i<j} [p(\mathbf{g_{i,j}} \mid v_{i,j}, \Theta) \cdot p(v_{i,j})] p(\Theta \mid \gamma) p(\gamma).$$

This distribution is analytically intractable, so in order to obtain our posterior sample we construct a Markov Chain Monte Carlo (MCMC) sampler.

3.5.1   MCMC Sampling Scheme—Our MCMC scheme begins with a block Gibbs sampler in which we sample network-specific parameters Ω_k and G_k from full conditionals of their posterior distributions. Then, we sample the graph similarity parameters Θ and γ from their conditional posterior distributions using a Metropolis–Hastings method that is equivalent to a reversible jump and incorporates between-model and within-model moves.

The main advantages of our prior on the precision matrices and the latent graphs are that (1) simultaneous block updates of all $p(p-1)/2$ edge-inclusion indicators are enabled and (2) no Markov chain approximation of intractable normalizing constants is required. Our Gibbs sampler can be viewed as a p-coupled stochastic search variable selection algorithm in the spirit of [13]. The generic iteration t of our algorithm can be summarized as follows:

   a.   Update graph $G_k^{(t)}$ and precision matrix $\Omega_k^{(t)}$ for each group $k = 1, \ldots, 4$.

   b.   Update the network relatedness parameters $\theta_{k,m}^{(t)}$ and $\gamma_{k,m}^{(t)}$ for $1 \leq k < m \leq 4$.

Details on *Step a* and *Step b* of our algorithm are provided in the Appendix.

3.5.2   Model Selection—There are two approaches for making inference on the graph structure. The first is to use a maximum a posteriori (MAP) estimate representing the mode of the posterior distribution for each sample group's graph. However, since the space of possible graphs is so large and we may only visit a particular graph a few times during the MCMC, this approach is generally not preferred in the context of large networks. Here, to infer gene connections, we use a more practical approach and estimate the posterior marginal probability (MPP) of edge inclusion for edge $g_{k,i,j}$ as the proportion of MCMC iterations after burn-in where edge (i, j) was included in graph G_k. Following [29], we then select those edges with marginal posterior probability of inclusion (MPP) > 0.5 for each of the four sample groups.

4   Simulation Studies

We use simulated data with related graph structures to assess the performances of the proposed approach. We also compare performances with alternate approaches. We consider two scenarios. The first scenario includes $p = 25$ nodes, the second scenario includes $p = 50$ nodes, to investigate how the method works for a larger scale problem. We begin by

constructing four precision matrices Ω_1, Ω_2, Ω_3, and Ω_4, each corresponding, respectively, to graphs G_1, G_2, G_3, and G_4. For each graph, there are $p \times (p-1)/2$ possible edges to be predicted. Ω_1 is set to the $p \times p$ symmetric matrix with entries $\omega_{i,i} = 1$ for $i = 1, \ldots, p$, entries $\omega_{i,i+1} = \omega_{i+1,i} = 0.5$ for $i = 1, \ldots, p-1$, and $\omega_{i,i+2} = \omega_{i+2,i} = 0.4$ for $i = 1, \ldots, p-2$. For Ω_2, we randomly generated a matrix with 70 % of the edges in Ω_1. We constructed Ω_3 by randomly changing five zero entries of Ω_1 to be nonzero. Lastly, Ω_4 was generated as a symmetric matrix with entries $\omega_{i,i} = 1$ for $i = 1, \ldots p$, $\omega_{i,i+1} = \omega_{i+1,i} = 0.5$ for $i = 1, \ldots, p-1$, and entries $\omega_{1,p} = \omega_{p,1} = 0.4$. Graph structures for the four groups in the 25-node scenario are shown in Fig. 1. To ensure that each generated precision matrix was positive definite, we used a similar approach to that of [9] where each off-diagonal element is divided by the sum of the off-diagonal elements in its row, and then the matrix is averaged with its transpose. Consequently, Ω_2 and Ω_3 are symmetric and positive definite but with off-diagonal elements with values less than half of those for Ω_1 and Ω_4. As a result, the true value of connection strength is weaker which resulted in worse performance of any method for groups two and three. We generated the data matrices X_k of size $n = 100$, for $k = 1, \ldots, 4$, from normal distributions $N(0, \Omega_k^{-1})$, characterized by $\Omega_1, \ldots, \Omega_1$ as the true precision matrices. For the 25-node scenario, edge counts were 47, 43, 52, and 25 for group one to group four, respectively, and pairwise shared edges were

$$\text{Counts of shared edges } = \begin{pmatrix} \cdot & 43 & 47 & 24 \\ & \cdot & 43 & 22 \\ & & \cdot & 24 \\ & & & \cdot \end{pmatrix}$$

For the 50-node scenario, edge counts were 97, 89, 102, and 50 for group one to group four, respectively, and pairwise shared edges were

$$\text{Counts of shared edges } = \begin{pmatrix} \cdot & 89 & 97 & 49 \\ & \cdot & 89 & 47 \\ & & \cdot & 49 \\ & & & \cdot \end{pmatrix}.$$

For prior specification, we used a Gamma(α, β) density with $\alpha = 1$ and $\beta = 9$ for the slab portion of the mixture prior on $\theta_{k,m}$ which results in a prior with mean 0.111. The tail probability is $1 - P(\theta_{k,m} \quad 1) = 0.04$, thereby avoiding assigning weight to larger values of $\theta_{k,m}$ and allowing for better discrimination between zero and nonzero values. To include the prior belief that the networks could be related, we set the hyperparameter $w = 0.5$ in the Bernoulli prior defining the network relatedness latent indicator $\gamma_{k,m}$. Parameters a and b were set to be $a = 1$ and $b = 19$ for all pairs (i, j) in the prior for $v_{i,j}$, resulting in a prior probability of edge inclusion around 5%. Hyperparameters v_0 and v_1 are set to be $v_0 = 0.02$ and $v_1 = 1$ according to published guidelines [44], ensuring the MCMC converges quickly and mixes well. The MCMC was run as described in Sect. 3 with 20,000 iterations of burn-in and 40,000 iterations as a basis for posterior inference. Marginal posterior probability of inclusion (MPP) for each edge $g_{k,i,j}$ is estimated as the percentage of MCMC samples post burn-in which include edge (i, j) in graph k. In order to assess accuracy, we report results for 25 simulated data sets: the 25-node simulated scenario is presented in Table 1 and the 50-

node case in Table 3. We report the true positive rate (TPR), the false positive rate (FPR), and the Matthews correlation coefficient (MCC) using a threshold of 0.5 for edge selection, and the area under the curve (AUC) (Table 2). The MCC is defined as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\,\mathrm{TP+FP}\,)(\,\mathrm{TP+FN}\,)(\,\mathrm{TN+FP}\,)(\,\mathrm{TN+FN}\,)}},$$

where TP, TN, FP, and FN stand for the true positives, true negatives, false positives, and false negatives, respectively. MCC takes values between -1 (total disagreement) and $+1$ (perfect selection) and measures the quality of the edge selection for a given threshold. A value of 0 suggests that the network reconstruction approach is no better than tossing a coin. Results in Tables 1 and 3 suggest that the TPR is higher in groups one and four, accounting for the fact that the magnitudes of the nonzero entries of Ω_1 and Ω_4 are greater than those of Ω_2 and Ω_3. The perfect AUC values of 1.00 for group 1 and group 4 illustrate that the marginal posterior probabilities of edge inclusion successfully provide an accurate means for learning the graph structure. The overall expected false discovery rate, or FDR, for edge selection is 0.12 for both the 25- and 50-node scenarios (Table 4).

The average marginal posterior probability for elements of Θ is estimated as the percentages of MCMC samples with $\theta_{k,m} > 0$ or $\gamma_{km} = 1$ for $1 \leq k < m \leq K$. For the 25-node scenario, averaged MPPs for $\theta_{k,m}$ and their standard errors (SE) were

$$\mathrm{MPP}(\Theta) = \begin{pmatrix} \cdot & 0.9984(.0027) & 0.9917(.00155) & 0.9999(.0001) \\ & \cdot & 0.6496(.0646) & 0.5830(.0468) \\ & & \cdot & 0.5157(.0344) \\ & & & \cdot \end{pmatrix}.$$

For the 50-node scenario, averaged MPPs for $\theta_{k,m}$ and their standard errors (SE) were

$$\mathrm{MPP}(\Theta) = \begin{pmatrix} \cdot & 1.000(.0000) & 0.9983(.0031) & 1.000(.0000) \\ & \cdot & 0.6557(.0665) & 0.4613(.0488) \\ & & \cdot & 0.4618(.0656) \\ & & & \cdot \end{pmatrix}.$$

To emphasize scalability of our method, we expanded our simulation study to a 100 node scenario using the same data-generating mechanisms implemented in the smaller simulations. After running ten replicates of our method, averaged MPPs for $\theta_{k,m}$ and their standard errors (SE) were

$$\mathrm{MPP}(\Theta) = \begin{pmatrix} \cdot & 1.000(.0000) & 0.9999(.0002) & 1.000(.0000) \\ & \cdot & 0.6779(.0478) & 0.4674(.0698) \\ & & \cdot & 0.3785(.0547) \\ & & & \cdot \end{pmatrix}.$$

Increasing the network size had no impact on the overall expected false discovery rate for our method. While true positive rate decreases slightly, other measures of method

performance seem to remain about the same. Further results for the 100-node scenario are shown in Table 5.

We compared the performances of our approach with two alternative multiple network methods. First, using the R package JGL [8], we applied the fused and joint graphical lasso methods of [9]. Accuracy of structure learning is given in Tables 2, 4, and 6 for the 25-, 50-, and 100-node scenarios, in terms of TPR, FPR, MCC, and AUC. For the lasso methods, AUC estimates were obtained by varying the sparsity parameter, while the similarity parameter was fixed. Results reported are the maximum for the sequence of similarity parameter values tested. Results indicate that the fused and group graphical lasso methods are quite good at the identification of true edges and seem to perform better for larger networks, but generally have high false positive rates. Our proposed method on the other hand has much lower sensitivity and achieves the best overall performance as measured by the AUC for the 25- and 50-node settings. For the 100-node setting, using the optimal penalty parameters for each replicate of the lasso methods resulted in an AUC slightly higher than that of our proposed method for fused lasso; however, false positive rates are significantly greater than that for our method.

5 Case Study on Disease Severity in COPD

This section illustrates the application of our method to infer the evolution of gene pathways in COPD subjects as emphysema increases in severity. We applied the proposed joint graphical model estimation method using hyperparameters $\alpha = 4$ and $\beta = 5$ for the slab portion of the mixture prior on $\theta_{k,m}$ which results in a prior with mean 0.4. Because we had no prior reference network or knowledge of graph structure, for our edge-specific prior we chose $q_{i,j} \sim$ Beta(1,9) for all edges (i, j). Other hyperparameters were set the same as in simulations. The MCMC sampler was run for 40,000 burn-in iterations followed by 80,000 iterations used for inference. For posterior inference, we selected those edges with marginal posterior probability of inclusion greater than 0.5. To verify convergence of our chains, we compared correlations of resulting MPP from two chains with different starting points. Pearson correlations were in the range of .9971–.9989, and Spearman correlations were in the range of .9705–.9967. The inferred network structures are shown in Figs. 2, 3, 4 and 5 for each of the four selected pathways, respectively. Network similarity across groups for each of the four pathways was estimated as

$$\text{MPP}(\Theta)_{RegAuto} = \begin{pmatrix} \cdot & .76 & .81 & .83 \\ & \cdot & .73 & .84 \\ & & \cdot & .77 \\ & & & \cdot \end{pmatrix}, \ \text{MPP}(\Theta)_{GPL} = \begin{pmatrix} \cdot & .96 & .94 & .95 \\ & \cdot & .89 & .91 \\ & & \cdot & .96 \\ & & & \cdot \end{pmatrix}$$

$$\text{MPP}(\Theta)_{OxPhos} = \begin{pmatrix} \cdot & 1.0 & 1.0 & 1.0 \\ & \cdot & 1.0 & 1.0 \\ & & \cdot & 1.0 \\ & & & \cdot \end{pmatrix}, \ \text{MPP}(\Theta)_{FcyR} = \begin{pmatrix} \cdot & 1.0 & 1.0 & 1.0 \\ & \cdot & 1.0 & 1.0 \\ & & \cdot & 1.0 \\ & & & \cdot \end{pmatrix}$$

For all final results, hub genes are defined as genes with at least four edges. Hub genes and edges were further examined for protein–protein interactions and disease-related gene annotation. protein–protein interactions were obtained from Biological General Repository for Interaction Datasets (BioGrids) v. 3.4.132 [5]. Disease annotation information was obtained from GeneCards [38] with the search term "lung" or "pulmonary" in the "Publications" search engine for GeneCards.

To further the comparison of our method with other methods, we also applied fused and joint graph lasso methods to the Reg Auto and GPL pathways from the ECLIPSE COPD dataset, using AIC to find optimal parameters. Overall, lasso methods seemed to have similar results to our proposed method with much denser networks due to higher false positive rates. This was expressed in particular by the Reg Auto pathway because every possible unique edge was selected by the lasso method. A detailed description of this comparison can be found in the Appendix.

5.1 Disrupted Interactions Due to Disease Severity

For each of the 4 pathways, we further examined all pairs of inferred gene interactions. Table 7 shows the total number of inferred pair interactions for each one of the 4 pathways, together with the number of those pairs that show evidence of disrupted interactions across disease severity. In the table, for each pathway, the four disease groups, ordered from least to most severe, are coded with 0's and 1's, with 1 indicating high MPP. For example, 1000 indicates that a pair has a MPP 0.50 in the least severe emphysema group (first group is indicated by 1) but not the others (last three groups are indicated by 0), while 0011 indicates that a pair has a MPP 0.50 in the two most severe groups (last two groups are indicated by 1), but not in the less severe disease cases (first two groups are indicated by 0).

For all 4 pathways, we see larger numbers of disrupted pairs in the most extreme case, where the interactions are strongest either for the controls or the most severe emphysema. That is, we find that a pair has a high MPP (0.50) in the no emphysema subjects but low MPP (<0.50) in the mild-to-severe emphysema subjects; or vice versa, a high MPP in the more severe disease subjects but not in the less severe emphysema and control subjects. This observation highlights two interesting sets of interactions for further investigation; interactions that are disrupted even with mild levels of emphysema and interactions that only develop for the most severe emphysema cases.

Interestingly, some of the pairs identified in Table 7 are known to have protein–protein interactions (PPI). In Table 7, we report the numbers of such pairs in parentheses and list the actual pairs in Table 8. One notable edge that changes based on disease severity is the gene pair ATG5-ATG3 in the RegAuto pathway. The MPP of this interaction is higher for the less severe COPD group (0.52–0.59) but then decreases for the most severe COPD group (0.476), indicating a disrupted interaction associated with disease severity. ATG5 (autophagy-related gene 5) is also one of the top hub genes (discussed below) and is associated with the GO category of innate immune response.

5.2 Hub Genes

Hub genes are highly connected genes and, as such, are expected to play an important role in biology. Here, we explored all genes in our inferred networks with at least 2 edges appearing in any of the disease groups [22]. Table 9 indicates the hub genes for each of the 4 pathways, together with the numbers of disrupted pairs and the number of known PPIs involving hub genes. For RegAuto, two of the hub genes have also been discussed above (ATG5-ATG3). For this pathway, increased autophagy has been observed in lung tissue from COPD patients with increased activation of autophagic proteins including protein products of hub genes found in ATG5 and ATG4B [6]. Another gene of interest is PIK3CD in the FcγR pathway. Expression and signaling of this gene is increased in the lungs of patients with COPD and is associated with reduced glucocorticoid responsiveness. Some authors have suggested that selective inhibition of the protein product PI3Kdelta might restore glucocorticoid function in patients with COPD, therefore representing a potential therapeutic target [24].

6 Conclusion

Motivated by the study of four critical pathways in COPD, we have proposed and implemented a novel approach to study how gene networks change with disease progression. We have introduced a novel Bayesian approach for multiple graphical models based on shrinkage and MRF priors. The combination of these two priors has allowed us to develop a computationally efficient algorithm and to perform a fully Bayesian analysis of the four targeted pathways. The proposed modeling approach allows to share information between sample groups, when appropriate, as well as to obtain a measure of relative network similarity across groups.

We have applied our approach to the ECLIPSE COPD dataset. Pathway enrichment of significant genes is often used in genomic research to identify candidate pathways but does not give additional information on how specific interactions within pathways are altered with disease severity. Using our Bayesian hierarchical approach, we were able to infer gene networks within 4 selected pathways. Our method has identified critical hub genes for all the four targeted pathways. Furthermore, several gene connections appeared to be disrupted with increased disease severity and constitute interesting candidates for further investigation, in an effort to characterize the disease evolution. Our analysis has clearly suggested that the autophagy-related gene ATG5 plays a critical role in COPD progression, highlighting critical interactions and highly connected genes that represent interesting targets for therapeutic targets. We also found several genes and gene interactions (ATG5, ATG3, and PIK3CD) that have already been associated with COPD. Further investigation of additional interactions such as UQCRC2-NDUFA1, which shows disruption based on disease severity, is the goal of future work. Using simulation studies, we have demonstrated the superior performance of our approach in comparison with competing methods.

Appendix

Details on our MCMC Algorithm

In this section, we provide a detailed description of *Step a* and *Step b* of our MCMC algorithm.

Step a. By partitioning Ω into $V = \left(v_{i,j}^2\right)$, a $p \times p$ symmetric matrix with zeroed diagonal entries and $\left(v_{i,j}^2\right)_{i\,<\,j}$ in the upper diagonal entries and setting $S = X'X$, we can focus on the last column and row to acquire

$$\Omega = \begin{pmatrix} \Omega_{1,1} & \omega_{1,2} \\ \omega_{1,2}' & \omega_{2,2} \end{pmatrix}, \quad S = \begin{pmatrix} S_{1,1} & s_{1,2} \\ s_{1,2}' & s_{2,2} \end{pmatrix}, \quad V = \begin{pmatrix} V_{1,1} & v_{1,2} \\ v_{1,2}' & 0 \end{pmatrix}.$$

Changing variables from $(\omega_{1,2}, \omega_{2,2})$ to $\left(u = \omega_{1,2}, v = \omega_{2,2} - \omega_{1,2}'\Omega^{-1}\omega_{1,2}\right)$, we have full conditionals

$$u| \cdot \sim N\left(-Cs_{1,2}, C\right) \quad \text{and} \quad v| \cdot \sim \text{Gamma}\left(\frac{n}{2} + 1, \frac{s_{2,2} + \lambda}{2}\right),$$

where $C = \left\{(s_{2,2} + \lambda)\Omega_{1,1}^{-1} + \text{diag}\left(v_{1,2}^{-1}\right)\right\}^{-1}$. Using this method, we can permute any column to attain the full conditional used to generate $\Omega|\mathbf{G}, X$. Our full conditional on $\mathbf{G}$ is then an independent Bernoulli of the form

$$P\left(g_{i,j} = 1|\Omega, X\right) = \frac{N\left(\omega_{i,j}|0, v_1^2\right)\pi}{N\left(\omega_{i,j}|0, v_1^2\right)\pi + N\left(\omega_{i,j}|0, v_0^2\right)(1 - \pi)},$$

where the quantity $\frac{\pi}{1-\pi}$ is determined by the MRF prior on the graph structure such that

$$\frac{\pi}{1-\pi} = \frac{p\left(G_k'|v_{i,j}, \Theta, \{G_m\}_{m \neq k}\right)}{p\left(G_k|v_{i,j}, \Theta, \{G_m\}_{m \neq k}\right)} = \exp\left\{-v_{i,j} + 2\sum_{m \neq k} \theta_{k,m}g_{m,i,j}\right\},$$

for proposed new graph G_k' which differs from the current graph G_k only in that edge (i, j) is excluded from G_k' and included in G_k.

Step b. In order to update $\theta_{k,m}$ and $\gamma_{k,m}$, we must consider the full conditional distribution. Considering only the terms of the joint prior for graphs $G_1, ..., G_k$ which include $\theta_{k,m}$, we can see that

$$p(G_1, ..., G_4|v, \Theta) = \prod_{i\,<\,j} C\left(v_{i,j}, \Theta\right)^{-1}\exp\left(v_{i,j}\mathbf{1}^T\mathbf{g_{i,j}} + \mathbf{g_{i,j}}^T\Theta\mathbf{g_{i,j}}\right)$$

$$\propto \prod_{i\,<\,j} C\left(v_{i,j}, \Theta\right)^{-1}\exp\left(2\theta_{k,m}g_{k,i,j}g_{m,i,j}\right).$$

The full conditional distribution of $\theta_{k,m}$ and $\gamma_{k,m}$ can then be written as

$$p(\theta_{k,m}, \gamma_{k,m} | \cdot) = p(G_1, \ldots, G_k | v, \Theta) p(\theta_{k,m} | \gamma_{k,m}) p(\gamma_{k,m} | w)$$

$$\alpha \left(\prod_{i<j} C(v_{i,j}, \Theta)^{-1} \exp(2\theta_{k,m} g_{k,i,j} g_{m,i,j}) \right)$$

$$\times \left((1 - \gamma_{k,m}) \delta_0 + \gamma_{k,m} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_{k,m}^{\alpha-1} e^{-\beta \theta_{k,m}} \right)$$

$$\times \left(w^{\gamma_{k,m}} (1-w)^{(1-\gamma_{k,m})} \right).$$

Because the normalizing constant from the joint prior on the graphs is analytically intractable, we use Metropolis–Hastings step to sample from $\theta_{k,m}$ and $\gamma_{k,m}$ for each pair of $(k,m), 1 \leq k < m \leq 4$ from the joint full conditional distribution. Each iteration has two steps based on the approach described by [14] to sample from mutually singular distribution mixtures. First, we perform a between-model move. If the current state is $\gamma_{k,m} = 1$, we propose $\gamma_{k,m}^\star = 0$ and $\theta_{k,m}^\star = 0$ resulting in the Metropolis–Hastings ratio

$$r = \frac{p(\theta_{k,m}^\star, \gamma_{k,m}^\star | \cdot) \times q(\theta_{k,m})}{p(\theta_{k,m}, \gamma_{k,m} | \cdot)} = \frac{\Gamma(\alpha)}{\Gamma(\alpha^\star)} \frac{(\beta^\star)^{\alpha^\star}}{\beta^\alpha} (\theta_{k,m})^{\alpha^\star - \alpha} e^{(\beta - \beta^\star)\theta_{k,m}}$$

$$\times \prod_{i<j} \frac{C(v_{i,j}, \Theta) \exp(-2\theta_{k,m} g_{k,i,j} g_{m,i,j})}{C(v_{i,j}, \Theta^\star)} \frac{1-w}{w},$$

Where $\Theta^\star$ represents the network similarity matrix Θ with entry $\theta_{k,m} = \theta_{k,m}^\star$. If moving instead from $\gamma_{k,m} = 0$ to $\gamma_{k,m}^\star = 1$, the ratio is

$$r = \frac{p(\theta_{k,m}^\star, \gamma_{k,m}^\star | \cdot)}{p(\theta_{k,m}, \gamma_{k,m} | \cdot) \times q(\theta_{k,m})} = \frac{\Gamma(\alpha^\star)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta^\star)^{\alpha^\star}} \times (\theta_{k,m})^{\alpha - \alpha^\star} e^{(\beta^\star - \beta)\theta_{k,m}}$$

$$\times \prod_{i<j} \frac{C(v_{i,j}, \Theta) \exp(-2\theta_{k,m}^\star g_{k,i,j} g_{m,i,j})}{C(v_{i,j}, \Theta^\star)} \frac{w}{1-w}.$$

Next, we perform the within-model move if the value of $\gamma_{k,m}$ sampled from the between-model move is 1. Here, we propose a new value using the same proposal density as before, for $\theta_{k,m}$. Our Metropolis–Hastings ratio is

$$r = \frac{p(\theta_{k,m}^\star, \gamma_{k,m}^\star | \cdot) \cdot q(\theta_{k,m})}{p(\theta_{k,m}, \gamma_{k,m} | \cdot) \cdot q(\theta_{k,m}^\star)} = \left(\frac{\theta_{k,m}^\star}{\theta_{k,m}} \right)^{\alpha - \alpha^*} \cdot e^{(\beta^\star - \beta)(\theta_{k,m}^\star - \theta_{k,m})}$$

$$\times \prod_{i<j} \frac{C(v_{i,j}, \Theta) \exp(2(\theta_{k,m}^\star - \theta_{k,m}) g_{k,i,j} g_{m,i,j})}{C(v_{i,j}, \Theta^\star)}.$$

In our last step of the MCMC, we sample from the full conditional distribution of $v_{i,j}$. The terms of the joint prior on the graphs including $v_{i,j}$ are

$$p(G_1, ..., G_k | v, \Theta) = \prod_{i < j} C(v_{i,j}, \Theta)^{-1} \exp\left(v_{i,j} \mathbf{1}^T \mathbf{g_{i,j}} + \mathbf{g_{i,j}}^T \Theta \mathbf{g_{i,j}}\right)$$

$$\propto C(v_{i,j}, \Theta)^{-1} \exp\left(v_{i,j} \mathbf{1}^T \mathbf{g_{i,j}}\right).$$

Given the prior on $v_{i,j}$, we can attain the posterior full conditional given the data and all remaining parameters

$$p(v_{i,j} | \cdot) \propto \frac{\exp(a v_{i,j})}{\left(1 + e^{v_{i,j}}\right)^{a+b}} C(v_{i,j}, \Theta)^{-1} \exp\left(v_{i,j} \mathbf{1}^T \mathbf{g_{i,j}}\right)$$

$$= \frac{\exp\left(v_{i,j}\left(a + \mathbf{1}^T \mathbf{g_{i,j}}\right)\right)}{C(v_{i,j}, \Theta) \cdot \left(1 + e^{v_{i,j}}\right)^{a+b}}.$$

We then propose a value $q^\star$ from the density Beta(2,4) for each pair (i, j) where $1 \quad i < j \quad p$ and set $v^\star = \text{logit}(q^\star)$. We can write our proposal density in terms of $v^\star$ as

$$q(v^\star) = \frac{1}{B(a^\star, b^\star)} \frac{e^{a^\star v^\star}}{\left(1 + e^{v^\star}\right)^{a^\star + b^\star}},$$

with Metropolis–Hastings ratio

$$r = \frac{p(v^\star | \cdot)}{p(v_{i,j} | \cdot)} \frac{q(v_{i,j})}{q(v^\star)}$$

$$= \frac{\exp\left(\left(v^\star - v_{i,j}\right) \cdot \left(a - a^\star + \mathbf{1}^T \mathbf{g_{i,j}}\right)\right) \cdot C(v_{i,j}, \Theta) \cdot \left(1 + e^{v_{i,j}}\right)^{a+b-a^\star-b^\star}}{C(v^\star, \Theta) \times \left(1 + e^{v^\star}\right)^{a+b-a^\star-bI^\star}}.$$

Case Study: Comparison to the Fused and Joint Graphical Lasso

In this section, we compare the proposed Bayesian approach to the fused and joint graphical lasso in terms of the findings obtained from the analysis of the ECLIPSE dataset. Specifically, we focused on the Reg Auto and GPL pathways. For both the fused and joint graphical lasso, we selected the penalty parameters that minimized the AIC, as recommended by [9]. For the Reg Auto pathway, the fused graphical lasso penalty parameters were selected as $\lambda_1 = 0.015$ and $\lambda_2 = 0.0001$, and for the group lasso were selected as $\lambda_1 = 0.015$ and $\lambda_2 = 0$ (this value was selected after an extensive grid search with step size of .0000005). For the GPL pathway, penalty parameters were selected as $\lambda_1 = 0.02$ and $\lambda_2 = 0.0005$ for the fused lasso, and $\lambda_1 = 0.02$ and $\lambda_2 = 0.0$ for the group lasso. Results are summarized in the two tables below.

Reg auto: method edge count comparison

	Proposed method	Group fused lasso	Joint group lasso
Group 1 edge count	98	159	159
Group 2 edge count	95	155	155
Group 3 edge count	89	155	155
Group 4 edge count	98	146	146
Unique edge count	153	190	190

GPL: method edge count comparison

	Proposed method	Group fused lasso	Joint group lasso
Group 1 edge count	312	560	560
Group 2 edge count	255	553	553
Group 3 edge count	288	545	545
Group 4 edge count	314	536	536
Unique edge count	539	802	802

For the Reg Auto pathway, it can be seen that edge counts were equivalent for the fused lasso and the group lasso. Both lasso methods selected all the possible 190 edges; this illustrates the issue corresponding to high false positive rates for lasso methods and consequently hints at more difficult interpretation of results. Percentage overlap of unique edges for Reg Auto was computed as

$$\frac{\text{Unique Edges in Proposed and Lasso Method}}{\text{Unique Lasso Edge Count}},$$

and resulted in an overlap of 80 %. Lasso methods identified the same hub genes as the proposed Bayesian approach, plus ATG10 and ULK3.

Similar conclusions can be derived from the analysis of the GPL pathway. The same edges were selected by both the group and fused lasso for all disease groups; 802 out of 820 possible unique edges were selected. Of the 18 edges remaining which were not selected by the lasso methods, five were selected by our proposed method. This resulted in a percentage eoverlap of unique edges for GPL 67%. The lasso methods identified the same hub genes as our proposed method in addition to DGKE, DGKQ, and MBOAT1. Overall, the lasso methods have similar results to our proposed approach, but result in much more dense networks due to their higher false positive rates. The proposed Bayesian approach provides sparser solutions that can be more easily interpreted.

References

1. Armagan A, Dunson D, Lee J (2013) Generalized double pareto shrinkage. Stat Sin 23(1):119 [PubMed: 24478567]

2. Atay-Kayis A, Massam H (2005) The marginal likelihood for decomposable and non-decomposable graphical gaussian models. Biometrika 92:317–355

3. Bahr T et al. (2013)Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 49(2):316–23 [PubMed: 23590301]

4. Bowler R et al. (2014) Plasma sphingolipids associated with copd phenotypes. Am J Respir Crit Care Med 191(3):275–284

5. Chatr-Aryamontri A, Breitkreutz B, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, Winter A, Sellam A, Chang C, Hirschman J, Theesfeld C, Rust J, Livstone MS, Dolinski K, Tyers M (2015) The biogrid interaction database: 2015 update. Nucleic Acids Res 43(Database issue):470–478 [PubMed: 25452337]

6. Chen Z, Kim H, Sciurba F, Lee S, Feghali-Bostwick C, Stolz D, Dhir R, Landreneau R, Schuchert M, Yousem S, Nakahira K, Pilewski J, Lee J, Zhang Y, Ryter S, Choi A (2008) Egr-1 regulates autophagy in cigarette smoke-induced chronic obstructive pulmonary disease. PLoS ONE 3(10):3316

7. Clyde M, George E (2004) Model uncertainty. Stat Sci 19(1):81–94

8. Danaher P (2012) Jgl: performs the joint graphical lasso for sparse inverse covariance estimation on multiple classes. http://CRAN.R-project.org/package=JGL

9. Danaher P, Wang P, Witten D (2014) The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc B 76(2):373–397

10. Dobra A, Jones B, Hans C, Nevins J, West M (2004) Sparse graphical models for exploring gene expression data. J Multivar Anal 90:196–212

11. Dobra A, Lenkoski A, Rodriguez A (2012) Bayesian inference for general gaussian graphical models with application to multivariate lattice data. J Am Stat Assoc 106:1418–1433

12. GEO (2015) Gene expression omnibus. http://www.ncbi.nlm.nih.gov/geo

13. George E, McCulloch R (1993) Variable selection via Gibbs sampling. J Am Stat Assoc 88:881–889

14. Gottardo R, Raftery A (2008) Markov chain Monte Carlo with mixtures of mutually singular distributions. J Comput Graph Stat 17(4):949–975

15. Griffin J, Brown P (2010) Inference with normal-gamma prior distributions in regression problems. Bayesian Anal 5(1):171–188

16. Guo J, Levina E, Michailidis G, Zhu J (2011) Joint estimation of multiple graphical models. Biometrika 98(1):1–15 [PubMed: 23049124]

17. Hanahan D, Weinberg R (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674 [PubMed: 21376230]

18. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of affymetrix genechip probe level data nucleic acids research. Nucleic Acids Res 31(4):e15 [PubMed: 12582260]

19. Jones B, Carvalho C, Dobra A, Hans C, Carter C, West M (2005) Experiments in stochastic computation for high dimensional graphical models. Stat Sci 20(4):388–400

20. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in kegg. Nucleic Acids Res 42:199–205

21. Khondker Z, Zhu H, Chu H, Lin W, Ibrahim J (2013) The Bayesian Covariance Lasso. Stat Its Interface6(2):243

22. Langfelder P, Mischel SHP (2013) When is hub gene selection better than standard meta-analysis? PLoS ONE 8(4):e61505 [PubMed: 23613865]

23. Li F, Zhang N (2010) Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. J Am Stat Assoc 105(491):1202–1214

24. Marwick J, Caramori G, Casolari P, Mazzoni F, Kirkham P, Adcock I, Chung K, Papi A (2010) A role for phosphoinositol 3-kinase delta in the impairment of glucocorticoid responsiveness in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol 125(5):1146–53 [PubMed: 20381852]

25. Mukherjee S, Speed T (2008) Network inference using informative priors. Proc Natl Acad Sci 105(38):14,313–14,318

26. Ni Y, Marchetti G, Baladandayuthapani V, Stingo F (2015) Bayesian approaches for large biological networks. In: Mitra R, Muller P (eds) Nonparametric Bayesian methods in biostatistics and bioinformatics. Springer, New York

27. Park T, Casella G (2008) The Bayesian lasso. J Am Stat Assoc 20(1):140–157

28. Parshall M (1999) Adult emergency visits for chronic cardiorespiratory disease: does dyspnea matter? Nurs Res 48(2):62–70 [PubMed: 10190832]

29. Peterson C, Stingo F, Vannucci M (2015) Bayesian inference of multiple Gaussian graphical models. J Am Stat Assoc 110(509):159–174 [PubMed: 26078481]

30. Peterson C, Stingo F, Vannucci M (2016) Joint bayesian variable and graph selection for regression models with network-structured predictors. Stat Med 35(7):1017–1031 [PubMed: 26514925]

31. Regan EA et al. (2010) Genetic epidemiology of copd (copdgene) study design. COPD 7(1):32–43 [PubMed: 20214461]

32. Reimand J, Wagih O, Bader G (2013) The mutational landscape of phosphorylation signaling in cancer. Sci Rep. doi:10.1038/srep02651

33. 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

34. Scott J, Berger J (2010) Bayes and empirical Bayes multiplicity adjustment in the variable-selection problem. Ann Stat 38(5):2587–2619

35. Scott J, Carvalho C(2008) Feature-inclusion stochastic search for Gaussian graphical models. J Comput Graphical Stat 17:790–808

36. Singh D et al. (2014) Altered gene expression in blood and sputum in copd frequent exacerbators in the eclipse cohort. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0107381

37. Skrepnek G, Skrepnek S (2004) Epidemiology, clinical and economic burden, and natural history of chronic obstructive pulmonary disease and asthma. AM J Manag Care 10(5):S129–38 [PubMed: 15354678]

38. Stelzer G, Dalah I, Stein T, Satanower Y, Rosen N, Nativ N, Oz-Levi D, Olender T, Belinky F, Bahir I, Krug H, Perco P, Mayer B, Kolker E, Safran M, Lancet D (2011) In-silico human genomics with genecards. Hum Genomics 5(6):709–717 [PubMed: 22155609]

39. Stingo F, Marchetti G (2015) Efficient local updates for undirected graphical models. Stat Comput 25:159–171

40. Stingo F, Vannucci M (2011) Variable selection for discriminant analysis with markov random field priors for the analysis of microarray data. Bioinformatics 27(4):495–501 [PubMed: 21159623]

41. Stingo F, Chen Y, Vannucci M, Barrier M, Mirkes P (2010) A Bayesian graphical modeling approach to microRNA regulatory network inference. Ann Appl Stat 4(4):2024 [PubMed: 23946863]

42. Telesca D, Mueller P, Kornblau S, Suchard M, Ji Y (2012) Modeling protein expression and protein signaling pathways. J Am Stat Assoc 107(500):1372–1384

43. Wang H (2012) The Bayesian graphical lasso and efficient posterior computation. Bayesian Anal 7(2):771–790 [PubMed: 27375829]

44. Wang H (2015) Scaling it up: stochastic search structure learning in graphical models. Bayesian Anal 10(2):351–377

45. Wang H, Li Z (2012) Efficient gaussian graphical model determination under g-wishart prior distributions. Electron J Stat 6:168–198

46. Yajima M, Telesca D, Ji Y, Muller P (2015) Detecting differential patterns of interaction in molecular pathways. Biostatistics 16(2):240–251 [PubMed: 25519431]
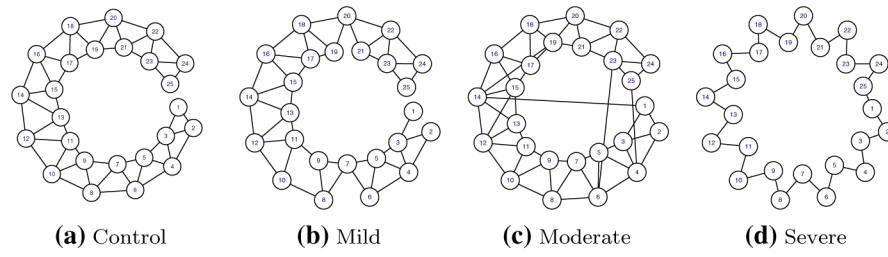
(a) Control   **(b)** Mild   **(c)** Moderate   **(d)** Severe

Fig. 1.
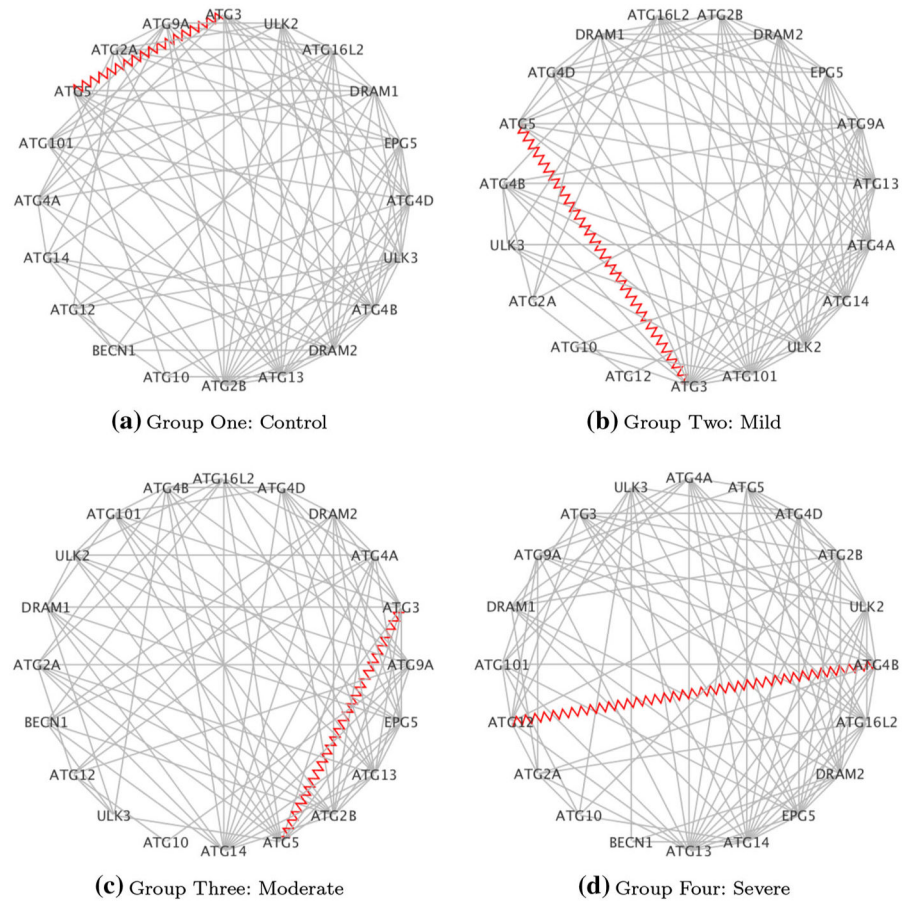Simulation study: true graph networks for the 25-node setting

Fig. 2.

Case study on COPD: estimated networks for the Reg Auto pathway: red zig-zag edges denote known protein–protein interactions (PPI)
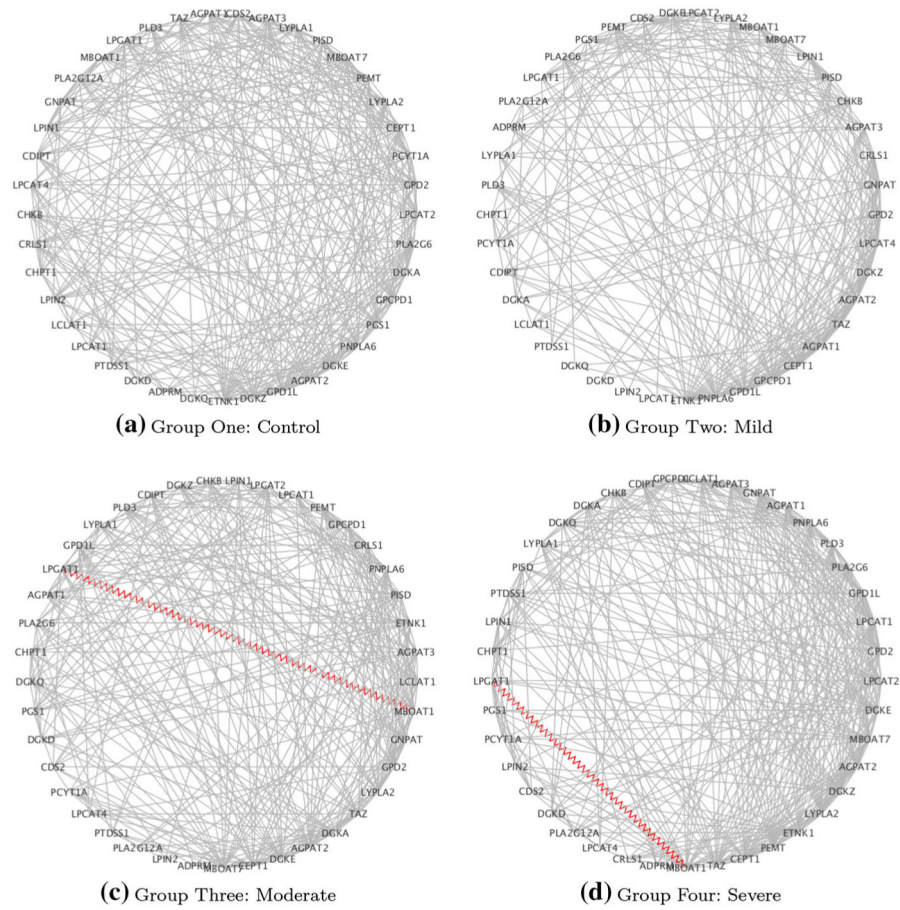
Fig. 3.
Case study on COPD: estimated networks for the GPL pathway: red zig-zag edges denote
known protein–protein interactions (PPI).eps

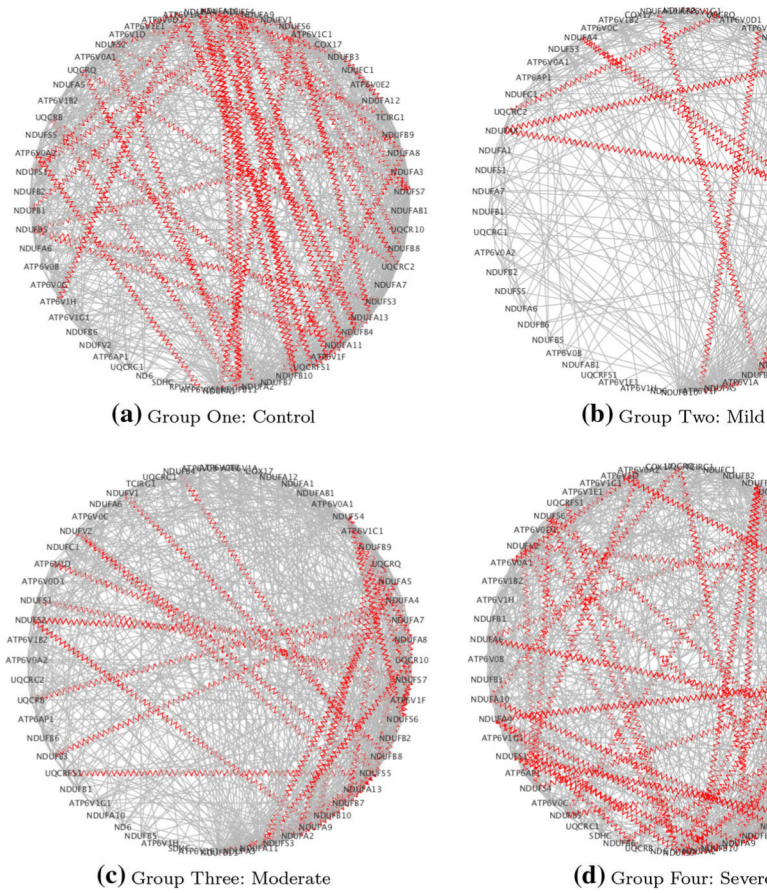**(a)** Group One: Control

(b) Group Two: Mild

(c) Group Three: Moderate

(d) Group Four: Severe

Fig. 4.
Case study on COPD: estimated networks for the OxPhos pathway: red zig-zag edges denote known protein–protein interactions (PPI)

(a) Group One: Control

(b) Group Two: Mild

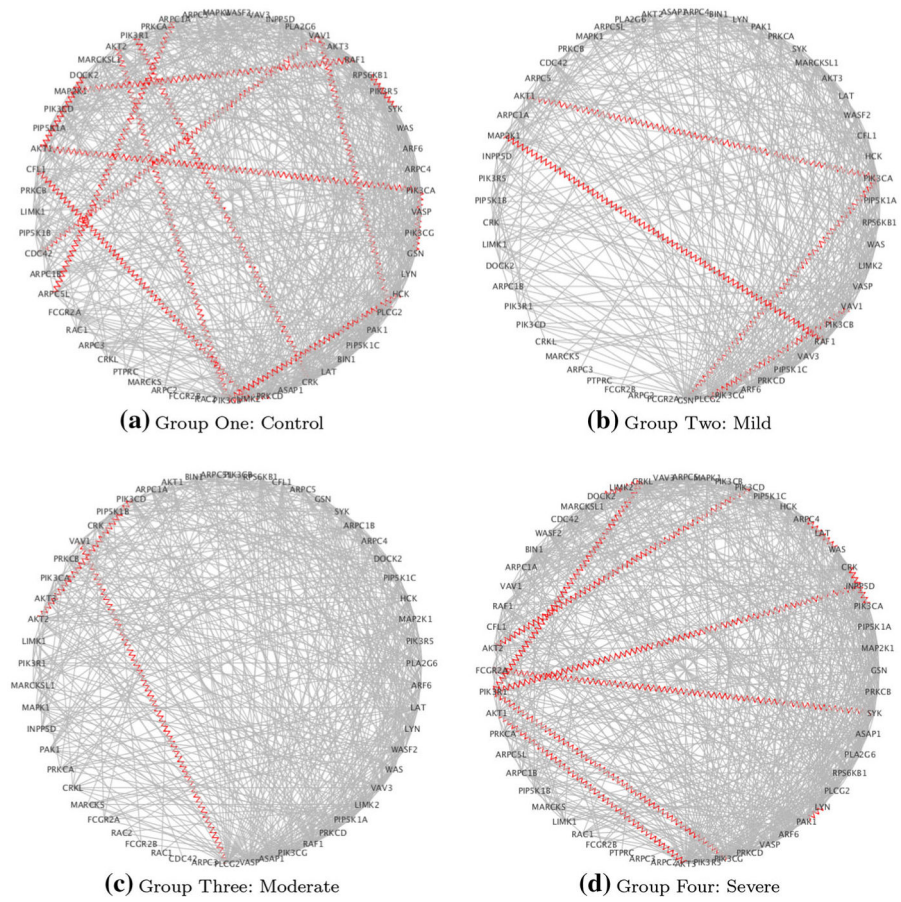**(c)** Group Three: Moderate

(d) Group Four: Severe

Fig. 5.
Case study on COPD: estimated networks for the FcγR pathway: red zig-zag edges denote known protein–protein interactions (PPI)

Table 1

Simulation study: results of our method for the 25-node setting across 25 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Group 1	1.000 (.0000)	0.0038 (.0045)	0.9883 (.0137)	1.0000 (.0000)
Group 2	0.4967 (.0966)	0.0168 (.0064)	0.5990 (.0798)	0.9272 (.0236)
Group 3	0.3838 (.0531)	0.0165 (.0558)	0.5133 (.0558)	0.8841 (.0259)
Group 4	1.000 (.0000)	0.0010 (.0017)	0.9941 (.0097)	1.0000 (.0000)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient and area under curve (AUC), with associated standard error (SE)

Table 2

Simulation study: results from competing methods for the 25-node setting across 25 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Fused graphical lasso	0.96 (.0149)	0.4737 (.0433)	0.3386 (.0256)	0.8993 (.0065)
Group graphical lasso	0.9598 (.0152)	0.4749 (.0498)	0.3378 (.0282)	0.8489 (.0101)
Proposed method	0.5806 (.4264)	0.0006 (.0008)	0.6760 (.3354)	0.9528 (.0527)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient (MCC), and area under curve (AUC) with standard errors (SE)

Table 3

Simulation study: results of our method for the 50-node setting across 25 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Group 1	1.000 (.0000)	0.0029 (.0019)	0.9823 (.0115)	1.000 (.0000)
Group 2	0.5276 (.0416)	0.0091 (.0031)	0.6379 (.0418)	0.9297 (.0145)
Group 3	0.3843 (.0481)	0.0091 (.0036)	0.5268 (.0467)	0.8967 (.0229)
Group 4	1.000 (.0000)	0.0007 (.0007)	0.9915 (.0089)	1.000 (.0000)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient, and area under curve (AUC), with associated standard error (SE)

Table 4

Results from competing methods for the 50-node setting across 25 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Fused graphical lasso	0.9289 (.0154)	0.3301 (.0338)	0.3150 (.0181)	0.9462 (.0033)
Group graphical lasso	0.9285 (.0156)	0.3279 (.0339)	0.3163 (.0179)	0.8834 (.0071)
Proposed method	0.7280 (.2798)	0.0055 (.0045)	0.7846 (.2095)	0.9566 (.0471)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient (MCC), and area under the curve (AUC), with standard errors (SE)

Table 5

Simulation study: results of our method for the 100-node setting across 10 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Group 1	1.000 (.0000)	0.0031 (.0008)	0.9630 (.0091)	1.000 (.0000)
Group 2	0.4553 (.0280)	0.0088 (.0013)	0.5344 (.0179)	0.9182 (.0117)
Group 3	0.4020 (.0357)	0.0082 (.0011)	0.5059 (.0302)	0.9131 (.0107)
Group 4	1.000 (.0000)	0.0007 (.0003)	0.9836 (.0068)	1.000 (.0000)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient, and area under curve (AUC), with associated standard error (SE)

Table 6

Results from competing methods for the 100-node setting across 10 simulated datasets

	TPR (SE)	FPR (SE)	MCC (SE)	AUC (SE)
Fused graphical lasso	0.8658 (.0479)	0.2620 (.2632)	0.2775 (.0940)	0.9688 (.0016)
Group graphical lasso	0.8496 (.0095)	0.1739 (.0028)	0.3089 (.0042)	0.8982 (.0048)
Proposed method	0.7135 (.2929)	0.0053 (.0037)	0.7453 (.2334)	0.9578 (.0434)

We report averaged true positive rate (TPR), false positive rate (FPR), Matthews correlation coefficient (MCC) and area under the curve (AUC), with standard errors (SE)

Table 7

Case study on COPD: numbers of total pairs of unique gene interactions and numbers of disease-disrupted pairs based on disease severity, for each one of the 4 selected pathways

	Total pairs	1000	1100	1110	0111	0011	0001	Total disrupted
GPL	539 (1)	58	26	28	21	40 (1)	59	232 (1)
FcyR	892 (50)	102 (8)	34 (3)	30 (1)	31	40 (1)	125 (9)	362 (22)
OxPhos	1072 (275)	127 (27)	37 (7)	25 (6)	23 (6)	62 (17)	120 (25)	394 (88)
RegAuto	153 (9)	13	2	11 (1)	8	4	11 (1)	49 (2)

There are four emphysema classes ordered by severity, with the first group being the no emphysema group and the last one the most severe emphysema group. For each pathway, the 4 groups are coded with 0's and 1's, with 1 indicating high MPP. For example, 1000 indicates that a pair has a MPP 0.50 in the least severe emphysema group (first group is indicated by 1) but not the others (last three groups are indicated by 0), while 0011 indicates that a pair has a MPP 0.50 in the two most severe groups (last two groups are indicated by 1), but not in the less severe disease cases (first two groups are indicated by 0). The number of pairs with known protein–protein interactions (PPI) is indicated in parentheses and listed further in Table 7

Table 8

Case study on COPD: subset of the disease-disrupted pairs in Table 7 with known protein–protein interactions

	Interaction in control (1000, 1100, 1110)	Interaction in disease (0111, 0011, 0001)
OxPhos	ATP6V0E1-ATP6V0A2,ATP6V1A-ATP6V0D1, ATP6V1A-ATP6V1E1, ATP6V1A-ATP6V1H, ATP6V1F-ATP6V1A, NDUFA1-NDUFS6, NDUFA10-NDUFA11, NDUFA10-NDUFB11, NDUFA4-NDUFB10, NDUFA4-NDUFB4, NDUFB1-NDUFA9, NDUFB3-NDUFA12, NDUFB3-NDUFA4,NDUFB5-NDUFA11, NDUFB5-NDUFB9, NDUFB7-NDUFA8, NDUFB8-NDUFV1, NDUFS1-NDUFA4, NDUFS1-NDUFA9, NDUFS2-NDUFB11, NDUFS2-NDUFS4, NDUFS2-NDUFS7, NDUFS3-NDUFB5, NDUFS4-NDUFA8, NDUFS4-NDUFB10, NDUFS4-NDUFB11, NDUFS4-NDUFB4, NDUFS4-NDUFS7, NDUFS6-NDUFA13, NDUFS6-NDUFA4, NDUFS6-NDUFS7, TCIRG1-ATP6V1E1, UQCRB-NDUFA11, UQCRB-NDUFB11, UQCRC2-NDUFA9, UQCRC2-UQCRQ, UQCRFS1-NDUFA11, UQCRFS1-NDUFA13, UQCRFS1-NDUFA4, UQCRQ-NDUFA2	ATP6AP1-ATP6V0A2, ATP6V0A1-ATP6V0A2, ATP6V1F-ATP6V1D,NDUFA2-NDUFB9, NDUFA3-NDUFA13, NDUFA5-NDUFA11, NDUFA6-NDUFA11,NDUFA6-NDUFA12, NDUFA8-NDUFB4, NDUFA8-NDUFB9, NDUFA9-NDUFA2, NDUFA9-NDUFA4, NDUFA9-NDUFB11, NDUFA9-NDUFB4, NDUFA9-NDUFB9, NDUFB5-NDUFA8, NDUFB6-NDUFA11, NDUFB7-NDUFA9, NDUFB8-NDUFS7, NDUFS1-NDUFB10, NDUFS1-NDUFS7, NDUFS2-NDUFA8, NDUFS2-NDUFA9, NDUFS3-NDUFA10, NDUFS3-NDUFS2, NDUFS3-NDUFV1, NDUFS5-NDUFA11, NDUFS5-NDUFA2, NDUFS5-NDUFA4, NDUFS5-NDUFA7, NDUFS5-NDUFB10, NDUFS5-NDUFB4, NDUFS5-NDUFB7, NDUFS5-NDUFS2, NDUFS5-NDUFS6, NDUFS5-NDUFV2, NDUFS5-UQCRFS1, NDUFS6-NDUFA8, NDUFS6-NDUFS4, NDUFV1-NDUFA4, NDUFV1-NDUFS4, NDUFV2-NDUFA13, NDUFV2-NDUFB4, NDUFV2-NDUFS4, UQCRB-NDUFA4, UQCRC2-NDUFB5, UQCRQ-NDUFA10, UQCRQ-NDUFB7
FcyR	ARPC1A-ARPC5L, CFL1-LIMK2, CRK-PIK3R1, GSN-PIK3CA, HCK-PIK3CB, PIK3CA-AKT1, PIK3CB-AKT2, PLCG2-VAV1, PRKCD-PIK3CB, RAF1-MAP2K1, RPS6KB1-SYK, VAV1-CDC42	AKT1-AKT3, ARPC4-WAS, CRK-PIK3CA, CRKL-DOCK2, CRKL-PIK3R1, FCGR2A-SYK, INPP5D-PIK3R1, LYN-PAK1, PIK3CD-AKT2, PIK3R5-PIK3CG
RegAuto	ATG5-ATG3	ATG4B-ATG12
GPL		LPGAT1-MBOAT1

Table 9

Case study on COPD: summary of top hub genes. Disrupted pairs which are also known PPI are listed in Table 8

Pathway	Hub genes	Disrupted pairs with hub gene	PPI with hub gene	Disrupted pairs and PPI with hub gene
GPL	ADPRM, AGPAT1, AGPAT3, CDIPT, CDS2, CEPT1, CHKB, CHPT1, CRLS1, DGKA, DGKD, DGKZ, ETNK1, GNPAT, GPCPD1, GPD1L, GPD2, LCLAT1, LPCAT1,LPCAT2, LPGAT1, LPIN1, LPIN2, LYPLA1, LYPLA2, MBOAT7, PCYT1A, PEMT, PGS1, PISD, PLA2G6, PLD3, PNPLA6, PTDSS1	226	1	1
FcyR	AKT1, AKT3, ARF6, ARPC1A, ARPC1B, ARPC3, ARPC4, ARPC5, BIN1, CDC42, CFL1, CRK, CRKL, DOCK2, FCGR2A, GSN, HCK, INPP5D, LAT, LIMK1, LIMK2 LYN, MAP2K1, MAPK1, MARCKS, MARCKSL1, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R5, PIP5K1A, PIP5K1B, PIP5K1C, PLA2G6,PLCG2, PRKCA, PRKCB, PRKCD, PTPRC, RAC1, RAF1, FPS6KB1,,SYK, VASP, VAV1, VAV3	354	50	22
OxPhos	ATP6AP1, ATP6V0A1, ATP6V0B, ATP6V0D1, ATP6V0E2, ATP6V1A, ATP6V1B2, ATP6V1C1, ATP6V1D, ATP6V1E1, ATP6V1F, ATP6V1G1, COX17, ND6, NDUFA1, NDUFA10, NDUFA13, NDUFA2, NDUFA3, NDUFA4, NDUFA5, NDUFA6, NDUFA7, NDUFA8, NDUFA9, NDUFAB1, NDUFB1, NDUFB11 NDUFB3, NDUFB4, NDUFB5, NDUFB6, NDUFB7, NDUFB8, NDUFC1, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS5, NDUFS6, NDUFS7, NDUFV1, NDUFV2, SDHC, TCIRG1, UQCR10, UQCRB, UQCRC1, UQCRC2, UQCRFS1, UQCRQ	384	270	88
RegAuto	ATG101, ATG12, ATG13, ATG14, ATG3, ATG4A, ATG4B, ATG5, ATG9A, BECN1, DRAM1, DRAM2, ULK2	43	8	2