Journal of Statistical Planning and Inference 190 (2017) 1–14

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Multilevel Gaussian graphical model for multilevel networks

Lulu Cheng [a], Liang Shan [b], Inyoung Kim [b,*]

[a] Regulatory Statistics Technology Center, Monsanto Company, St. Louis, MO, USA
[b] Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Received 30 March 2016
Received in revised form 28 October 2016
Accepted 6 May 2017
Available online 20 May 2017

Keywords:
Gaussian graphical model
Graphical LASSO
Pathway
Penalized likelihood
Sparse

ABSTRACT

Gaussian graphical models have become a popular tool to represent networks among variables such as genes. They use the conditional correlations from the joint distribution to describe the dependencies between gene pairs, and employ the precision matrix of the genes. Because of the sparse nature of the gene networks and small sample sizes in high dimensional genetic data, regularization approaches attracted much attention in aim at obtaining the shrinkage estimates of the precision matrix. However, existing methods have been focused on the Gaussian graphical model among genes; that is, they are only applicable to a single level Gaussian graphical model. It is known that pathways are not independent of each other because of shared genes and interactions among pathways. Developing multipathway analysis has been a challenging problem because of the complex dependence structure among pathways. By considering the dependency among pathways as well as the genes within each pathway, we propose a multilevel Gaussian graphical model (MGGM) in which one level describes the networks for genes and the other for pathways. We have developed a multilevel L1 penalized likelihood approach to achieve the sparseness on both levels. In addition, we have developed an iterative weighted graphical LASSO algorithm for MGGM. Our simulation results supported the advantages of our approach; our method estimated the network more accurately on the pathway level and sparser on the gene level. We also demonstrated the usefulness of our approach using a canine genes-pathways data set.

1. Introduction

Gaussian graphical models (Dempster, 1972), known as "covariance selection" or "concentration graph" models, have recently become a popular tool to learn gene association networks. It assumes the nodes (i.e., gene expression data observed in our study) are randomly sampled observational or experimental data from a multivariate Gaussian distribution. That is, let $V = \{v_1, \ldots, v_p\}$ be the set of nodes (genes), and $X_1, \ldots, X_p$ denote the expression data for the p genes; we assume that $(X_1, \ldots, X_p) \sim N(0, \Sigma)$ with positive definite variance–covariance matrix $\Sigma = (\sigma_{ij})$ and precision matrix $\Omega = \Sigma^{-1} = (\omega_{ij})$. Then, the Gaussian graphical model uses the precision matrix Ω as the adjacent matrix (i.e. $\omega_{ij} \neq 0$ implies an association between the gene pair and $\omega_{ij} = 0$ implies no association between the gene pair). A related but completely different concept are the so-called gene "relevance networks", which are based on the covariance matrix Σ. The simple reason why Gaussian graphical models should be preferred over relevance networks for the identification of gene networks is that the off-diagonal elements of Ω are proportional to partial correlations, while the off-diagonal elements of Σ are proportional to marginal correlations. In the latter, interactions are defined through standard correlation coefficients so that missing edges denote

* Corresponding author. Fax: +1 540 231 3863.
E-mail address: inyoungk@vt.edu (I. Kim).

http://dx.doi.org/10.1016/j.jspi.2017.05.003
0378-3758/

marginal independence only. The correlation coefficient is a weak criterion for measuring dependence, because marginally, i.e. directly and indirectly, more or less all genes will be correlated. This implies that zero marginal correlation is in fact a strong indicator for independence. On the other hand, partial correlation coefficients do provide a strong measure of dependence and, correspondingly, offer only a weak criterion of independence as most partial correlation coefficients usually vanish. And more often, with high dimension of genetic data, one would prefer concentrating the network size rather than trapping in a large amount of relevances resulting from relevance networks.

A number of studies have worked on estimating Ω. A popular way to estimate the precision matrix for Gaussian graphical models with small sample modeling is to introduce a penalty to the off-diagonal elements in Ω, which is feasible in computing when $n < p$ and which allows us to estimate the off-diagonal element simultaneously. The sparsity of the obtained precision matrix would be able to take the nature of the genetic networks into account. Due to the small sample size in gene expression data, researchers usually take a penalized log-likelihood approach and solve the following objective function,

$$\max_{\Omega}\big[\log\{\det(\Omega)\} - tr(S\Omega) - \lambda P(\Omega)\big],$$

where λ is a non-negative penalty parameter and $P(\cdot)$ is a penalty function on the precision matrix elements. A popular penalty function is to use the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996; Friedman et al., 2008; Yuan and Lin, 2007; Levina et al., 2008), which can be applied to shrink the off-diagonal elements in the precision matrix exactly to zero. Friedman et al. (2008), which is based on a coordinate descent procedure, is fast and can be adopted easily by many extensions of LASSO. For example, to remedy the bias issue in LASSO, Zou (2006) proposed the adaptive LASSO penalty and used the reciprocal of the absolute value of a consistent estimator raised to some power as the weight for each component. The solution can be obtained iteratively using weighted GLASSO. Another example is for the joint estimation of multiple graphical models (Guo et al., 2011). They proposed a factor across data categories for each off-diagonal element to represent the homogeneity network structure and put LASSO penalty on both the elements and factors. Their solution could also be obtained from an iterative weighted GLASSO algorithm.

However, these recent studies only work on association among genes. That is, these methods can describe the association between single genes only. It is known that pathways are sets of genes that serve a particular cellular or physiological function. Hence pathways are not independent of each other because of shared genes and interactions among them. Multi-pathway analysis has been a challenging problem because of the complex dependence structure among pathways. On the other hand, subtle connections between genes in two pathways may indicate strong connection between two pathways but can be ignored by individual gene network analysis. The main goal of our study is to develop a Gaussian graphical model for the gene and pathway network. Thus, by considering the dependency among pathways as well as genes within each pathway, we have proposed a multilevel Gaussian graphical model: one level is for pathway network structure and the second level is for gene network structure. We will propose a hierarchically structured graphical model for this in Section 2.

This paper is organized as follows. In Section 2, we propose a multilevel Gaussian graphical model for the gene and pathway network. Section 3 contains the penalized log-likelihood approach and the development of the algorithm for the solution. In Section 4 we compare our method with GLASSO method for individual gene networks based on several criteria. We introduce a definition of the degree of pathway-level connection. We also give a real data analysis in Section 5. Section 6 contains the conclusion and discussion.

2. Multilevel Gaussian graphical model

In this Section, we describe how to build a multilevel Gaussian graphical model for gene and pathway networks. First, we will provide the precision matrix for the Gaussian graphical model, then we will explain how to extract the pathway network information, and finally we will give a graphical illustration of the multilevel network model. Suppose we have p genes, the expression data for each were denoted by $X_1, \ldots, X_p$; the whole gene network can be represented by the precision matrix Ω,

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1p} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \cdots & \omega_{pp} \end{pmatrix}.$$

In this setting, if the off-diagonal element $\omega_{ij} = 0$, it means the ith and jth genes are conditionally independent. Furthermore, suppose these genes are in k predefined pathways, denoted by $P_1, \ldots, P_k$. Without loss of generality, we can re-denote the genes as: $X_{11}, \ldots, X_{1p_1}, X_{21}, \ldots, X_{2p_2}, \ldots, X_{k1}, \ldots, X_{kp_k}$, where $p_1, p_2, \ldots, p_k$ are the number of genes in each pathway. The conditional correlations among genes in the k and k'th pathways can be rewritten as a p_k-by-$p_{k'}$ sub-block precision matrix $\Omega_{kk'}$,

$$\Omega_{kk'} = \begin{pmatrix} \omega_{11}^{kk'} & \omega_{12}^{kk'} & \cdots & \omega_{1p_{k'}}^{kk'} \\ \omega_{21}^{kk'} & \omega_{22}^{kk'} & \cdots & \omega_{2p_{k'}}^{kk'} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p_k1}^{kk'} & \omega_{p_k2}^{kk'} & \cdots & \omega_{p_kp_{k'}}^{kk'} \end{pmatrix}.$$

where $\omega_{ij}^{kk'}$ is defined as follows. Let us reparameterize the precision matrix by introducing a pathway level factor $\theta_{kk'}$ for kth and k'th pathways. Then the partial correlation between jth gene in the kth pathway and $j'$the gene in the k'th pathway is $\omega_{j_k j'_{k'}}^{kk'} = \theta_{kk'}\gamma_{j_k j'_{k'}}^{kk'}$, where $\theta_{kk'} \geqslant 0$.

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1K} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{K1} & \Omega_{K2} & \cdots & \Omega_{KK} \end{pmatrix} = \begin{pmatrix} \theta_{11}\Gamma_{11} & \theta_{12}\Gamma_{12} & \cdots & \theta_{1K}\Gamma_{1K} \\ \theta_{21}\Gamma_{21} & \theta_{22}\Gamma_{22} & \cdots & \theta_{2K}\Gamma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{K1}\Gamma_{K1} & \theta_{K2}\Gamma_{K2} & \cdots & \theta_{KK}\Gamma_{KK} \end{pmatrix}$$

where

$$\Gamma_{kk'} = \begin{pmatrix} \gamma_{11}^{kk'} & \gamma_{12}^{kk'} & \cdots & \gamma_{1p_{k'}}^{kk'} \\ \gamma_{21}^{kk'} & \gamma_{22}^{kk'} & \cdots & \gamma_{2p_{k'}}^{kk'} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p_k 1}^{kk'} & \gamma_{p_k 2}^{kk'} & \cdots & \gamma_{p_k p_{k'}}^{kk'} \end{pmatrix}.$$

For identifiability, we give the following constraints:

$$\begin{aligned} &\Omega_{kk'} = \theta_{kk'}\Gamma_{kk'}, \quad \theta_{kk'} > 0, \ 1 \leq k, k' \leq K; \\ &\theta_{kk'} = \theta_{k'k}, \qquad \Gamma_{kk'} = \Gamma'_{k'k}, \quad 1 \leq k \neq k' \leq K; \\ &\theta_{kk} = 1, \qquad \Gamma_{kk} = \Omega_{kk}, \quad 1 \leq k \leq K. \end{aligned}$$

In our model, parameter $\theta_{kk'}$ indicates an association between pathways, and thus we call it the "pathway-level connection factor". Another parameter $\gamma_{j_k j'_{k'}}^{kk'}$ represents connections between individual genes, we call it "gene-level connection factor". That is, if the off-diagonal factor $\theta_{kk'} = 0$, it means the k and k'th pathways are conditionally independent. Therefore, none of the genes in the kth pathway have any connection to the genes in the other pathway. If $\theta_{kk'} \neq 0$, there is a connection between these two pathways to some extent. We further provide a definition of connection degree in Section 4 to measure how strong the connection is. This measurement is based on the individual gene-level connection between genes in these two pathways, i.e. $\gamma_{j_k j'_{k'}}^{kk'}$. The diagonal matrix Γ_{kk} represents the gene networks within the kth pathway, one should check the off-diagonal elements within the pathway networks.

Fig. 1 is an illustration of the proposed multilevel Gaussian graphical model. The large circles represent pathways, and each dot in a circle represents a gene. The edges among circles represent dependencies among the pathways, and the lines between dots represent dependencies among genes. With this graph, we could see which pathways are associated with each other, at the same time, we could see the gene networks within each pathway.

3. Estimation of multilevel Gaussian graphical model

In this section, we provide a penalized likelihood approach to estimate the parameters and develop algorithm for the solution. First we define some notation. Let A be a $p \times p$ matrix. We denote $\det(A)$ as the determinant of A, $tr(A)$ as the trace of A, $\phi_{\max}(A)$ and $\phi_{\min}(A)$ as the maximum and minimum eigenvalues of A, respectively, and A^+ as the diagonal matrix with the same diagonal elements of A. We further define as follows: $A^- = A - A^+$, $\|A\|_F^2 = \sum_{ij} a_{ij}^2$, $\|A\|^2 = \phi_{\max}(AA^T)$, and $|A|_1 = \sum_{ij}|a_{ij}|$.

3.1. The penalized approach

To estimate the multilevel Gaussian graphical model, and in order to obtain a sparse network at both gene and pathway levels, we propose the following penalized log-likelihood and denote it as **Q1**:

$$\min_{\{\Theta, \Gamma_{kk'}, \Gamma_{kk}\}} \Big[-\log L(\Omega) + \lambda P(\Omega) \Big] = -\log\{\det(\Omega)\} + tr(S\Omega) + \eta_1|\Theta^-|_1 + \eta_2 \sum_{k \neq k'} |\Gamma_{kk'}|_1 + \eta_3 \sum_k |\Gamma_{kk}^-|_1$$

subject to

$$\begin{aligned} &\Omega_{kk'} = \theta_{kk'}\Gamma_{kk'}, \quad \theta_{kk'} > 0, \ 1 \leq k, \ k' \leq K; \\ &\theta_{kk'} = \theta_{k'k}, \qquad \Gamma_{kk'} = \Gamma_{k'k}^t, \quad 1 \leq k \neq k' \leq K; \\ &\theta_{kk} = 1, \qquad \Gamma_{kk} = \Omega_{kk}, \quad 1 \leq k \leq K. \end{aligned}$$

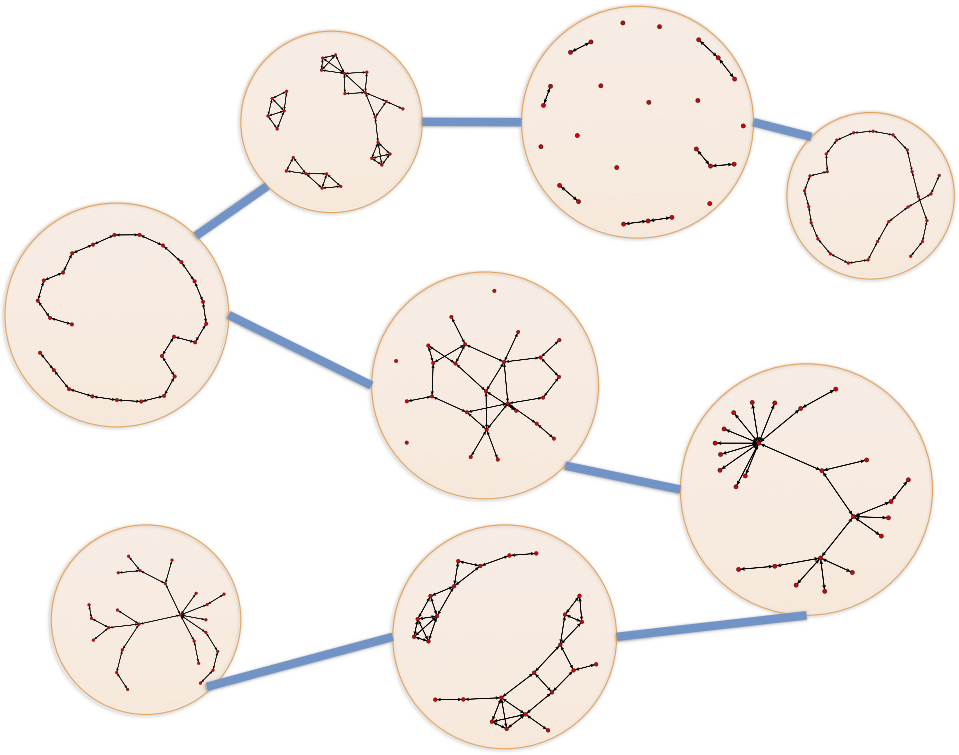


Fig. 1. An illustration of the proposed multi-level Gaussian graphical model. The circles represent pathways and the points in the circles represent genes.

Note that there are three penalty functions applied to different components of Ω. The first one, $\eta_1|\Theta^-|_1$, is used to shrink the off-diagonal pathway-level connection factors. It will effectively remove the really weak edges between pathways. The second penalty, $\eta_2\sum_{k\neq k'}|\Gamma_{kk'}|_1$, controls sparsity of the network between genes in different pathways. That is, if some $\theta_{kk'}$ is not shrunk to zero, there would be connection between these two pathways. With this second penalty function, one can still obtain disconnection between the genes in these two pathways. The third penalty, $\eta_3\sum_k|\Gamma_{kk}^-|_1$, helps obtain a sparse gene network within each pathway. Although three penalty parameters provide more flexibility, tuning these penalty parameters dramatically increases computing time. Thus we show that our problem can be simplified to the following version in Lemma 1.

Lemma 1. ***Q1*** *is equivalent to the following* ***Q2****:*

$$\min_{\{\Theta,\Gamma_{kk'},\Gamma_{kk}\}}\Big[-\log L(\Omega)+\lambda P(\Omega)\Big]=-\log\{\det(\Omega)\}+tr(S\Omega)+|\Theta^-|_1+\eta\sum_{k\neq k'}|\Gamma_{kk'}|_1+\eta_3\sum_k|\Gamma_{kk}^-|_1$$

subject to

$$\begin{aligned}&\Omega_{kk'}=\theta_{kk'}\Gamma_{kk'},\quad \theta_{kk'}>0,\ 1\le k,\ k'\le K;\\&\theta_{kk'}=\theta_{k'k},\qquad \Gamma_{kk'}=\Gamma_{k'k}^t,\quad 1\le k\neq k'\le K;\\&\theta_{kk}=1,\qquad \Gamma_{kk}=\Omega_{kk},\quad 1\le k\le K.\end{aligned}$$

To prove this, we need to show that if $(\hat\Theta^{**},\{\hat\Gamma_{kk'}^{**}\},\{\hat\Gamma_{kk}^{**}\})$ *is a local minimizer of* ***Q1****, then there exists a local minimizer of* ***Q2****, denoted as* $(\hat\Theta^{*},\{\hat\Gamma_{kk'}^{*}\},\{\hat\Gamma_{kk}^{*}\})$*, such that* $\hat\theta_{kk'}^{**}\cdot\hat\Gamma_{kk'}^{**}=\hat\theta_{kk'}^{*}\cdot\hat\Gamma_{kk'}^{*}$ *and vice versa. This proof is shown in Appendix A of the Supplementary Materials.*

With this result, we have only two parameters η and η_3, instead of the three, which significantly reduces computing efforts.

3.2. The algorithm

To make the optimization problem more achievable, **Q2** can be further reformulated.

Lemma 2. *The **Q2** is equivalent to the following **Q3**:*

$$\min_{\{\Theta,\Gamma_{kk'},\Gamma_{kk}\}}\Big[-\log L(\Omega)+\lambda P(\Omega)\Big]=-\log\{\det(\Omega)\}+tr(S\Omega)+\lambda\sum_{k\neq k'}\sqrt{|\Omega_{kk'}|_1}+\eta_3\sum_k|\Omega_{kk}^-|_1$$

where $\lambda=2\sqrt{\eta}$.

To prove this, we need to show that if $\hat{\Omega}_{kk'}$ is a local minimizer of **Q3**, then there exists a local minimizer of **Q2**, denoted as $(\hat{\Theta},\{\hat{\Gamma}_{kk'}\},\{\hat{\Gamma}_{kk}\})$, such that $\hat{\Omega}_{kk'}=\hat{\theta}_{kk'}\hat{\Gamma}_{kk'}$ and vice versa. We summarize the detail of the proof in Appendix B of the Supplementary Materials.

The square root in the term $\lambda\sum_{k\neq k'}\sqrt{|\Omega_{kk'}|_1}$ is like an umbrella penalty function that shades every item under it. That is, it helps shrink all elements in $\Omega_{kk'}$ to zero at the same time. Therefore, it plays a role in the pathway-level connection factor and also provides less computation effort in practice. The solution of **Q3** can be obtained through an interactive approach based on the local linear approximation (Zou and Li, 2008). Letting $\Omega^{(t)}=(\Omega_{kk'}^{(t)})$ be the estimation of the precision matrix at the tth iteration, we can write the approximation as follows:

$$\sqrt{|\Omega_{kk'}|_1}\approx\frac{|\Omega_{kk'}|_1}{\sqrt{|\Omega_{kk'}^{(t)}|_1}}.$$

Then, at the $(t+1)$ iteration, the problem for **Q3** can be written as **Q4**:

$$\min_{\{\Omega_{kk'},\Gamma_{kk}\}}\Big[-\log L(\Omega)+\lambda P(\Omega)\Big]=-\log\{det(\Omega)\}+tr(S\Omega)+\lambda\sum_{k\neq k'}\tau_{kk'}|\Omega_{kk'}|_1+\eta_3\sum_k|\Omega_{kk}^-|_1$$

where $\tau_{kk'}=1/\sqrt{|\Omega_{kk'}^{(t)}|_1}$. The solution can be computed efficiently using the weighted GLASSO algorithm of Friedman et al., (2008). Since $\sqrt{|\Omega_{kk'}^{(t)}|_1}$ could be zero, we make 10^{-10} the threshold for computing stability. The summary of the proposed algorithm is as follows:

1 Initialize the precision matrix as $(\hat{\Sigma}+\nu I)^{-1}$ to guarantee a positive definite start.

2 Update $\Omega^{(t)}$ by **Q4** using weighted GLASSO;

3 Stop until convergence.

In addition, we can provide two asymptotic properties: the first one is consistency and the second one is sparsistency. The asymptotic properties can similarly be shown by following those of Guo et al. (2011). The first one implies that when the sample size n and model size p go to infinity and the tuning parameters go to 0 at a certain rate, the estimated model goes to the true model at certain rate. The second one is more essential since one would prefer to estimate zeros according to the covariance estimation.

3.3. Model selection on the penalty parameter λ

For model selection on the penalty parameters, BIC criteria are often used. As the model is estimated through likelihood-based approach, the BIC criteria are believed to be better than cross-validation. In the Gaussian graphical model, it is defined as

$$BIC(\lambda,\eta_3)=tr(S\hat{\Omega})-\log|\hat{\Omega}|+\frac{\log(n)}{n}|\hat{\Omega}^-|_0,$$

where n is the sample size, S is the sample variance–covariance matrix, $\hat{\Omega}$ is the estimated precision matrix, and $|\hat{\Omega}^-|_0$ represents the number of nonzero elements in the off-diagonal of $\hat{\Omega}$. The simplicity and effectiveness of the BIC have made it very attractive, and it is well known to lead to asymptotically consistent model selection in the setting of a fixed number of variables p and a growing sample size n. However, in a scenario where p grows moderately with n, it is usually observed to be too liberal to select a model with many spurious covariates (Bogdan et al., 2004; Broman and Speed, 2002; Siegmund, 2004). Chen and Chen (2008) have proposed an extended BIC for model selection with large model spaces. Drton and Foygel (2010) have shown that in chain graphical structures, these criteria perform better than either cross-validation or the ordinary BIC criteria in terms of incurring a small loss in the positive selection rate but tightly controlling the false discovery rate. The extended BIC in the graphical model is defined as

$$EBIC(\lambda,\eta_3)=tr(S\hat{\Omega})-\log|\hat{\Omega}|+\frac{\log(n)}{n}|\hat{\Omega}^-|_0+4\xi\frac{\log(p)}{n}|\hat{\Omega}^-|_0,$$

where p is the number of nodes in the graphical model.

There is a trade-off between the positive selection rate and the false discovery rate based on the choice of the positive parameter ξ. From the simulations by Drton and Foygel (2010), $\xi = 0.5$ provides a good balance between the two evaluations. While the Gaussian graphical model is focused on the number of nodes in the graph, our MGGM is more interested in off-diagonal elements in the precision matrix. Hence, we modify the extended BIC. We would like to use $\xi = 0.5$ in the extended BIC, but with a modification of the degree of freedom part related to the number of parameters to be estimated. It is denoted as EBIC_m and is defined as

$$EBIC_m(\lambda, \eta_3) = tr(S\hat{\Omega}) - \log(n)|\hat{\Omega}| + \frac{\log(n)}{n}|\hat{\Omega}^-|_0 + 4\xi\frac{\log\{p(p-1)/2\}}{n}|\hat{\Omega}^-|_0.$$

where $\log\{p(p-1)/2\}$ is the total number of off-diagonal elements in the precision matrix. The extra degree of freedom results in the stronger penalty of large graphs, which is intuitively applicable to our scenario of sparsity at both gene and pathway level. We will demonstrate its improvements for graph estimation at the pathway levels as well in Section 4.

4. Simulation study

In this section, we conduct a simulation to compare the performance of our multilevel Gaussian graphical model with that of the regular graphical LASSO method.

4.1. Simulation settings

The elements of the off-diagonal are set to be $\exp(-\delta_g|i-j|)\exp(-\delta_p|p_i - p_j|)$, where δ_g and δ_p are set to ensure a well-decayed positive definite matrix, i and j are the gene indexes and p_i and p_j are the pathway indexes. We set the gene disconnection rate $P(\gamma = 0) = 0.95$ and vary the pathway disconnection rate between $P(\theta = 0) = 0.9$ and 0.75. A sample size is studied at $n = 50$. The number of pathways are also varying from $K = 10$, 15, to 20, with the number of genes in each pathway varied from $p_k = 5$, 15, to 30. For each combination, we generated 100 data sets. We compare our approach with maximize likelihood method and the regular graphical LASSO using 10 criteria. The first class of comparisons is the loss functions: entropy loss (EL), quadratic loss (QL), and Frobenius loss (FL) function. They are defined as follows:

$$\begin{aligned}
\text{EL} &= tr(\Omega^{-1}\hat{\Omega}) - \log|\Omega^{-1}\hat{\Omega}| - p;\\
\text{QL} &= tr(\Omega^{-1}\hat{\Omega} - I)^2;\\
\text{FL} &= tr\{(\Omega^{-1}\hat{\Omega} - I)^T(\Omega^{-1}\hat{\Omega} - I)\}.
\end{aligned}$$

The second class of comparison is the sparsity which is the number of zeros estimated in the graphical model. The third class of comparison is the false selection rate at gene level: the false positive error (FP) and error rate (FPR), false negative error (FN) and error rate (FNR), which are defined as follows:

$$\begin{aligned}
FP &= \#\text{ non-zeros estimated if the truth is zero};\\
FPR &= \frac{FP}{\#\text{ zeros in the true network}};\\
FN &= \#\text{ zeros estimated when the truth is not zero};\\
FNR &= \frac{FN}{\#\text{ non-zeros in the true network}}.
\end{aligned}$$

Lastly, the fourth class of comparison is the assessment of pathway-level selection accuracy using several measures. The first one we call "pathway connection degree bias (PCDB)". To obtain it, we first define "pathway connection degree (PCD)": for the pathway k and k', there are $p_k \cdot p_{k'}$ possible edges, and the PCD is defined as the proportion of edges among these possible positions. When we consider all pairs of pathways, we can obtain a $K-$by$-K$ PCD matrix, of which the diagonal elements are 1 and the off-diagonal elements are the PCDs for each pathway pair. Next we take the PCD matrix for both the true and estimated gene networks and sum up the squared difference for the elements where the true PCD is positive, i.e. where the true pathway pairs are connected, and one gets defined as PCDB. Similarly, we define the "pathway disconnection degree bias (PDDB)". The difference is that it is based on the "pathway disconnection degree (PDD)", and only sums up the squared PDD over the disconnected pathway pairs in the true precision matrix. Let $PCD_{0,kk'}$ and $PDD_{0,kk'}$ represent true PCD and PDD. We then formulate these measures as follows:

$$\begin{aligned}
PCD_{kk'} &= P(\omega_{jj'}^{kk'} \neq 0) = \frac{\#\omega_{jj'}^{kk'} \neq 0}{p_k \cdot p_{k'}};\\
PCDB &= \sum_{k\neq k'}(PCD_{kk'} - PCD_{0,kk'})^2\mathbf{1}_{\{PCD_{0,kk'}>0\}};\\
PDD_{kk'} &= P(\omega_{jj'}^{kk'} = 0) = \frac{\#\omega_{jj'}^{kk'} = 0}{p_k \cdot p_{k'}};\\
PDDB &= \sum_{k\neq k'}(PDD_{kk'} - PDD_{0,kk'})^2\mathbf{1}_{\{PDD_{0,kk'}=1\}}.
\end{aligned}$$

4.2. Simulation results

All of our simulations provided similar results; thus we summarize six common situations for demonstration:

- Case 1: We generate 10 pathways ($K = 10$). Within each pathway, there are 30 genes ($p_k = 30$). The probability of the disconnection for each pair of genes, $P_{\omega=0} = 0.9$, and the probability of the disconnection for each pair of pathways is $P_{\theta=0} = 0.9$;
- Case 2: The same as Case 1, except the pathway disconnection rate is $P_{\theta=0} = 0.75$;
- Case 3: We generate 15 pathways ($K = 15$). Within each pathway, there are 15 genes ($p_k = 15$). The probability of the disconnection for each pair of genes is $P_{\omega=0} = 0.9$, and the probability of the disconnection for each pair of pathways is $P_{\theta=0} = 0.9$;
- Case 4: The same as Case 3, except the pathway disconnection rate is $P_{\theta=0} = 0.75$;
- Case 5: We generate 20 pathways ($K = 20$). Within each pathway, there are 5 genes ($p_k = 5$). The probability of the disconnection for each pair of genes is $P_{\omega=0} = 0.9$, and the probability of the disconnection for each pair of pathways is $P_{\theta=0} = 0.9$;
- Case 6: The same as Case 5, except the pathway disconnection rate is $P_{\theta=0} = 0.75$.

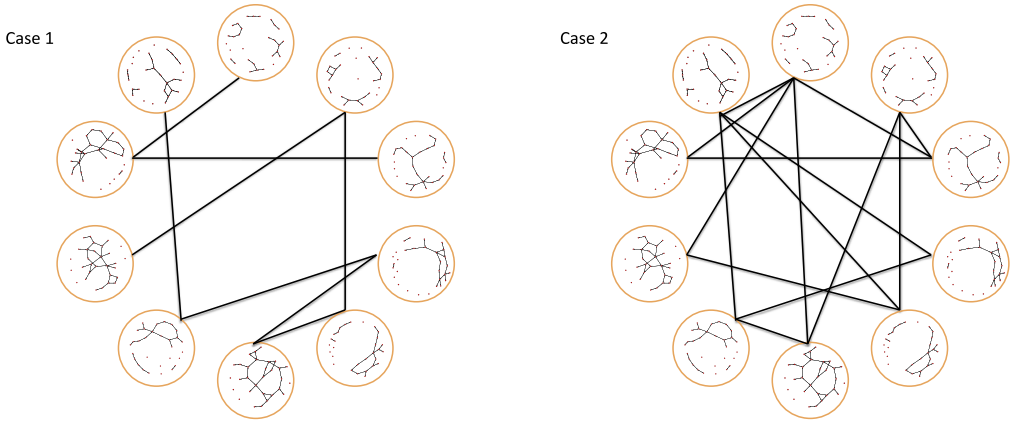
As illustrated by Fig. 2, Cases 1 and 2 represent situations in which we have a small number of pathways but a large number of genes in each pathway. Cases 3 and 4 represent situation in which we have compared a number of genes within each pathway, relative to the number of pathways. Cases 5 and 6 represent we have more pathways but small number of genes in each pathway. In Case 1, 3, and 5, the probability of the disconnection for each pair of pathways is $P_{\theta=0} = 0.9$, which means that we have a sparser pathway-level network than Cases 2, 4, and 6, where the pathway disconnection rate is $P_{\theta=0} = 0.75$.

In Section 3.3, we mentioned the preference for using the extended BIC to select the penalty parameters for the model. The advantage of the extended BIC has been demonstrated by Drton and Foygel (2010) in their simulation of a one-level chain network structure. Compared to the ordinary BIC, it will select a model with a very small loss of positive selection rate but a better control of the false discovery rate. We would demonstrate that in our simulation settings, the extended BIC (with pathway-based modified degree of freedom) also benefits the model selection by reducing the bias of the degree of pathway level connection or disconnection degree. Fig. 3 displays the scatter plot between the pathway connection degree bias (PCDB) obtained from the estimated models selected by EBIC_m, the modified extended BIC, versus those from the estimated models selected by the ordinary BIC. The smaller PCDB is, the better it is. Each point in the figure represents the PCDB obtained from each case in a simulated data set. The diagonal line means that the EBIC_m and BIC will select a model producing the same PCDB; that is, they perform the same when estimating the overall connection strength for pathway pairs. If a point falls under the diagonal line, it means that the EBIC_m-selected model results in a smaller bias when estimating the connection degree at the pathway level and thus better than BIC. Similarly, if a point falls above the diagonal line, it means that EBIC_m selects a model with larger pathway connection degree bias for this data set and thus is worse than BIC.

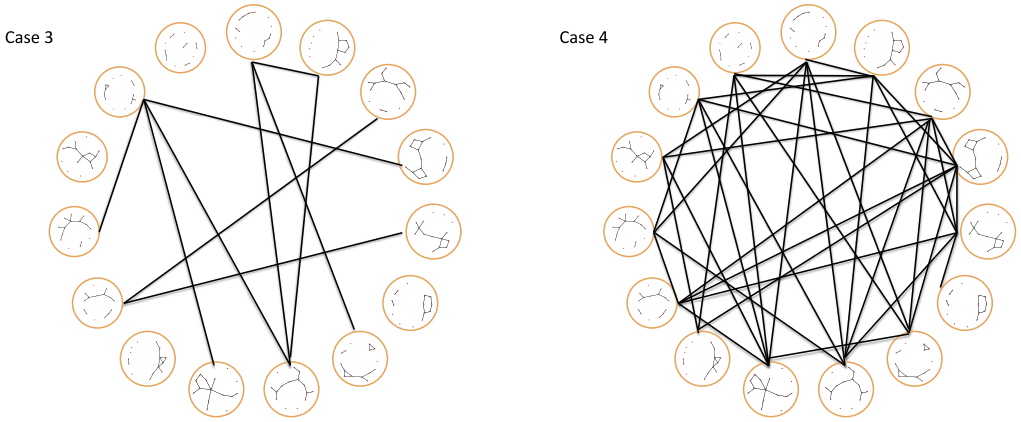
We first compare PCDB among cases that have three different combinations of pathway and gene sizes. We could see that in Cases 1 and 2 (10 pathways with 30 genes in each one), the PCDB from using BIC and EBIC_m is the smallest and almost the same as each other. In Cases 3 and 4 (15 pathways with 15 genes in each one), the PCDB is generally larger than those in Cases 1 and 2, and the comparison trend is parallelly shifted a little bit upward on the diagonal line, which indicates a small but acceptable increase that PCDB incurs when using EBIC_m. In Cases 5 and 6 (20 pathways with 5 genes in each one), we could see that the PCDB is larger than in the other cases, and its variance from data set to data set is also the largest in all cases. There are more points under the diagonal line, indicating that using EBIC_m will be more likely to provide a smaller PCDB than using BIC.

In general, the accuracy of EBIC_m and BIC is similar for the inference of connection at the pathway-level since they provide similar bias of pathway connection degree. Furthermore, we notice that when the pathway connection rate is higher, represented by points with cold tones (black, dark blue, and blue), the PCDB is smaller, which shows that for a sparser pathway network, the bias of the pathway connection degree will be smaller. Similarly, Fig. 4 displays the scatter plot between the pathway disconnection degree bias (PDDB) obtained from the estimated models selected by EBIC_m and those from the estimated models selected by the ordinary BIC. The smaller the PDDB is, the better it is. Each point in the figure represents the PDDB obtained from each case in a simulated data set. The diagonal line means that the EBIC_m and BIC will select a model producing the same PDDB; that is, they perform the same when estimating the overall degree of disconnection for pathway pairs. If a point falls under the diagonal line, it means the EBIC_m-selected model results in a smaller bias when estimating the disconnection degree at the pathway level, and thus better than BIC. Similarly, if a point falls above the diagonal line, it means that the EBIC_m selects a model with a larger pathway connection degree bias for this data set and thus is worse than BIC.

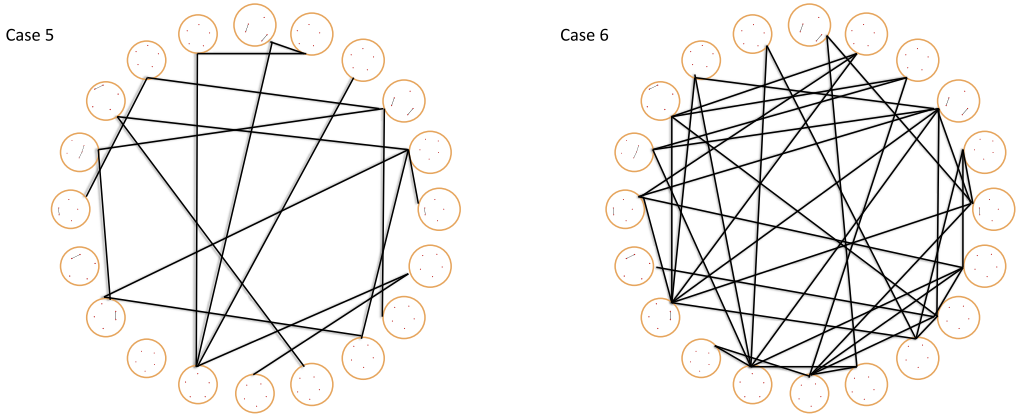
We then compare the PDDB among cases that have three different combinations of pathway and gene sizes. We could see that in Cases 1–4 (10 pathways with 30 genes in each one and 15 pathways with 15 genes in each one) the PDDB is smaller than in Cases 5 and 6, and the comparison trend turns underneath the diagonal line, with a rotation manner, not a parallel shifting, which indicates the significant increase that PDDB incurs when using BIC; in Cases 5 and 6 (20 pathways with 5 genes in each one), the PDDB is the largest and most scattered around among the pathways and gene size settings, with a similar comparison result to the Cases 1–4, indicating that using EBIC_m provides significantly smaller PDDB than using



(a) Case 1 and 2: 10 pathways with 30 genes in each pathway.



(b) Case 3 and 4: 15 pathways with 15 genes in each pathway.



(c) Case 5 and 6: 20 pathways with 5 genes in each pathway.

Fig. 2. Simulated pathway and gene network for the simulation cases. The circles represent pathways and the points in the circles represent genes. Each row shows a different setting of the gene and pathway size. Left column shows sparser pathway networks with disconnection rate $P_{\theta=0} = 0.9$ while right column shows denser ones with $P_{\theta=0} = 0.75$.

BIC. In general, the accuracy of the selected model by EBIC_m is much better than the one selected by BIC for the inference of disconnection at the pathway level since all the points are under the diagonal line.

Furthermore, unlike PCDB, the PDDB is similar in all cases no matter the pathway connection rate is 0.9 or 0.75. These results are shown in Fig. 4. We can observe that the points are overlapped between two cases of different pathway connection

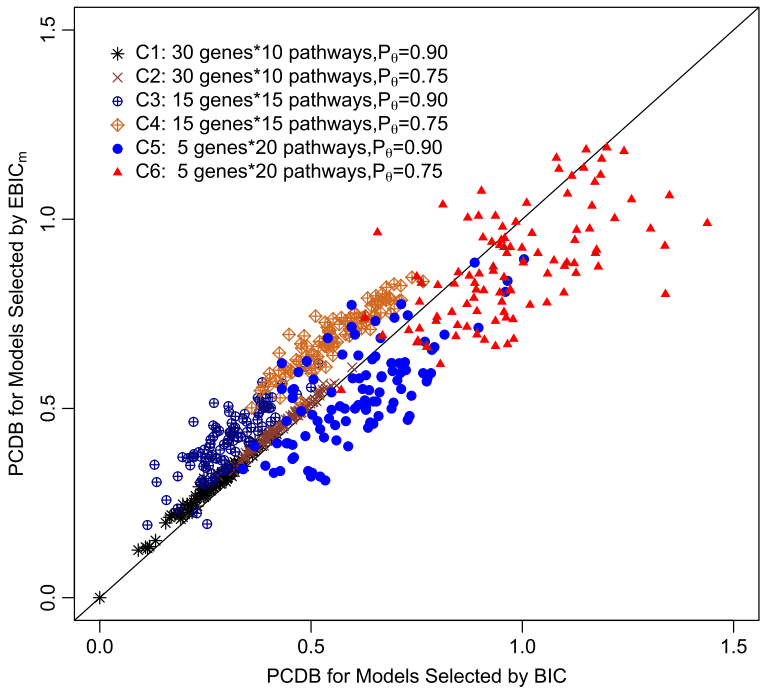


Fig. 3. Pathway Connection Degree Bias (PCDB) comparison for model selection by the modified extended BIC ($EBIC_m$) and BIC criteria; C1–C6 represent Case 1–Case 6 which are considered in simulation study; Points are PCDBs obtained from different cases.

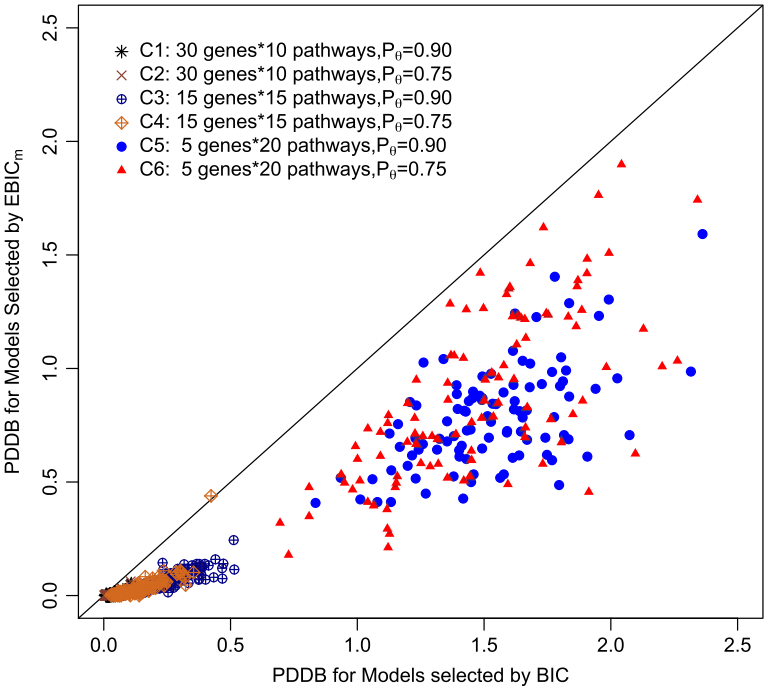


Fig. 4. Pathway Disconnection Degree Bias (PDDB) comparison for model selection by the modified extended BIC ($EBIC_m$) and BIC criteria; C1–C6 represent Case 1–Case 6 which are considered in simulation study; Points are PDDBs obtained from different Cases.

rates under each setting of pathway and gene sizes. The results of comparison of the maximum likelihood estimates (MLE), the graphical LASSO method (Glasso) and the multilevel graphical LASSO (MGlasso) based on the simulations are summarized in Tables 1–6. Tables 1 and 2 show situations in which a small number of pathways and large number of genes within each

Table 1
$K = 10$, $p_k = 30$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.9$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	FP
MLE	−INF	1.83e[33]	1.06e[18]	0	44 787
Glasso	515.79	237.06	29.39	0.9997	9
MGlasso	221.60	117.22	22.16	0.9991	25
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	9.27	1.9767
Glasso	59	0.0002	0.27	0.00	0.0205
MGlasso	50	0.0006	0.23	0.00	0.0005

Table 2
$K = 10$, $p_k = 30$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.75$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	
MLE	−INF	4.35e[33]	2.60e[18]	0	43 425
Glasso	INF	256.50	75.95	1.000	0
MGlasso	365.80	159.75	67.67	0.9979	52
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	8.12	4.4296
Glasso	1425	0.0000	0.90	0.00	0.4707
MGlasso	1386	0.0012	0.88	0.00	0.4526

Table 3
$K = 15$, $p_k = 15$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.9$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	FP
MLE	−INF	3.58e[31]	6.73e[16]	0	24 793
Glasso	124.99	44.14	17.73	0.9939	83
MGlasso	118.79	42.07	17.52	0.9930	19
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	13.71	4.1819
Glasso	338	0.0033	0.69	0.09	0.5200
MGlasso	368	0.0008	0.65	0.00	0.4529

pathway. Tables 3 and 4 show situations of the number of genes within pathway is comparable to the number of pathways. Tables 5 and 6 show situations in which larger number of pathways but fewer genes within each pathway. By checking all the cases, we can see that the maximum likelihood estimator almost always gives the smallest entropy loss because it does not shrink the parameters. Our method, the multilevel graphical LASSO, has the advantage of reducing the quadratic and Frobenius losses. The maximum likelihood estimator would never give a sparse network estimation in terms of having no zero appear on the off-diagonal in the estimated precision matrix, while the graphical LASSO and our method give similar sparsity, with our method being a little bit sparser overall. GLASSO and our method have comparable false positive rates. But our method gives a much better false negative rate.

For the pathway level, our method gives a disconnection degree bias comparable to the graphical LASSO, but has a smaller bias for the connection degree. In general, our method is better for the gene level covariance selection. Additionally, by comparing Tables 1 and 2, which show the same setting of the pathway and gene size but with a different sparsity on the pathway network, we can see that when the pathway-level network is sparser ($P_\theta = 0.9$), the more obvious advantage can be seen for our method in terms of the pathway-level criteria for the PCDB. A similar story can be found by comparing Tables 3 and 4, or Tables 5 and 6. By comparing Tables 2, 4 and 6, which show situations in which there is an increasing number of pathways but a reduced number of genes within each pathway, we found that the pathway disconnection degree bias (PDDB) of our method performed better in the cases where the number of pathways is larger, because PDDBs become smaller than those of the graphical LASSO. A similar story can be found by comparing Tables 1, 3 and 5.

Table 4
$K = 15$, $p_k = 15$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.75$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	FP
MLE	−INF	4.64e[31]	1.16e[17]	0	24 540
Glasso	122.31	54.71	20.38	0.9934	11
MGlasso	98.74	30.89	18.82	0.9941	4
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	12.72	6.2167
Glasso	339	0.0004	0.82	0.009	0.7294
MGlasso	345	0.0002	0.81	0.000	0.7144

Table 5
$K = 20$, $p_k = 5$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.9$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	FP
MLE	−22.98	127.51	50.60	0	4899
Glasso	−51.55	40.86	15.80	0.9977	5
Mglasso	−55.82	38.98	15.33	0.9901	24
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	18.97	3.9642
Glasso	22	0.0010	0.44	0.00	0.6145
Mglasso	13	0.0049	0.25	0.00	0.5713

Table 6
$K = 20$, $p_k = 5$, $P_{\omega=0} = 0.9$, $P_{\theta=0} = 0.75$; MLE = Maximum likelihood estimator; Glasso = Graphical lasso; MGlasso = our multilevel Graphical lasso; EL = entropy loss; QL = quadratic loss; FL = Frobenius loss; $P_{\hat{\omega}=0}$ = relative frequency of estimated zeros; FP = false positive error; FN = false negative error; FPR = false positive error rate; FNR = false negative error rate; PDDB = pathway disconnection degree bias; PCDB = pathway connection degree bias.

Method	EL	QL	FL	$P_{\hat{\omega}=0}$	FP
MLE	−28.47	120.33	47.35	0	4811
Glasso	−52.51	71.79	20.77	0.9985	3
Mglasso	−39.22	34.54	17.14	0.9947	15
Method	FN	FPR	FNR	PDDB	PCDB
MLE	0	1.0000	0.00	16.79	8.9248
Glasso	135	0.0006	0.71	0.05	1.0822
Mglasso	128	0.0031	0.67	0.00	1.0598

5. Real data analysis

In this section we give an example of real data analysis example. The canine data set comes from Pang et al. (2006). The goal of this study is to identify association among top-ranked pathways and to distinguish between does with lesion disease and without lesions. There were 15 dogs with lesions and 14 without, with a microarray of gene expression measurements. There was a total of 6592 genes. The canine data set was generated from investigative toxicology studies designed to identify the molecular pathogenesis of drug-induced vascular injury in coronary arteries of dogs treated with adenosine recept eragonist CI-947. Then, the canine genes are mapped to human orthologs for pathway analysis. The human orthologs for dogs are generated by matching the gene sequences using BLASTx (Enerson et al., 2006). Pang et al. (2006) have provided the rank of pathways based on results from the random forest model (Pang et al., 2006). We are interested in the gene and pathway networks for the top 10 pathways provided by Pang et al. for the two dog categories (healthy dog and unhealthy dog), are displayed in Fig. 5.

We found one pair of pathways that are connected in healthy dogs but the connection breaks in unhealthy dogs. This pair of pathways includes "Androgen and estrogen metabolism" and "Rac1 cell motility signaling pathway (BC)". The pathway "Androgen and estrogen metabolism" describes the inactivation and catabolism of male (androgen) and female (estrogen) hormones. It is related to vascular diseases since the estrogen has been demonstrated in many experiments to have vasodilator effects on endothelial cells, vascular smooth muscle, and extracellular matrix (Mendelsohn, 2002; Smiley and Khalil, 2009). It has been also reported that androgen and estrogen metabolism has effect on the inflammatory process,

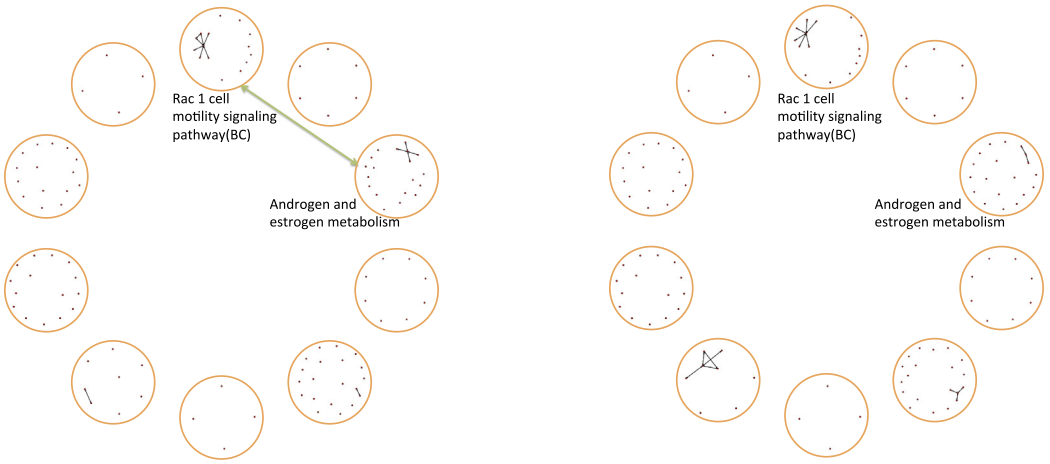


Fig. 5. Network for pathways ranked top 10 using random forest classification (Left:Healthy dog; right: unhealthy dog).

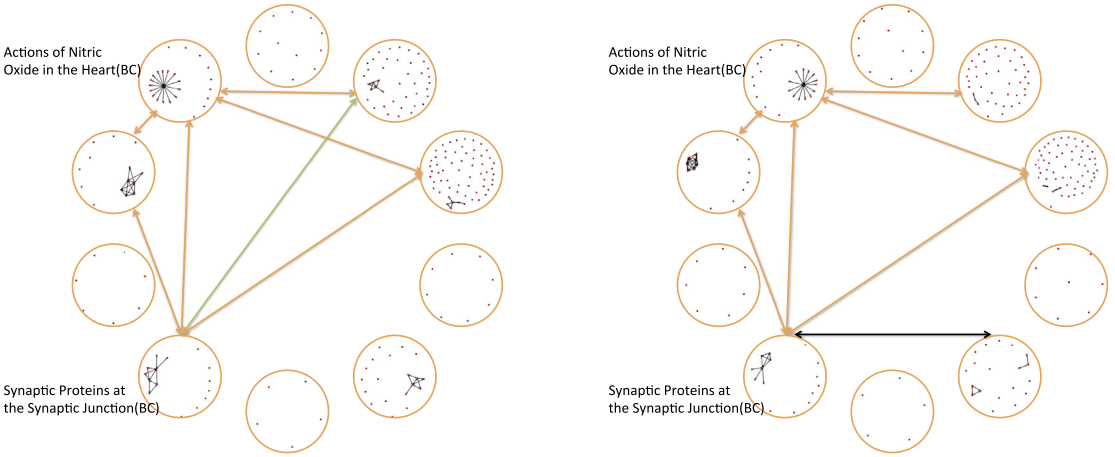


Fig. 6. Network for pathways ranked top 11–20 using random forest classification (Left:Healthy dog; right: unhealthy dog).

oxidative stress, and angiogenesis (Miller and Duckles, 2008). The association of estrogen with white adipose tissue mass with consequent changes in circulating lipid levels and inflammatory cytokines points to another mechanism by which estrogen might benefit vascular health (Simpson et al., 2005). Masood et al. (2010) have also learned the impact of sex hormone metabolism on vascular effects of hormone therapy in cardiovascular disease. Rac1 is a small G-protein in the Rho family that regulates cell motility in response to extracellular signals. Several changes in cytoskeletal structure and other aspects of cell structure are involved in cell motility. It has been indicated that the characteristics of vascular smooth muscle cell phenotypes as they relate to cell migration (Louis and Zahradka, 2010) and that the cell migration process is driven by the small GTPases from the Rho family, primarily by Rac1 and Cdc42 (Coso et al., 1995; Olson et al., 1995; Pankova et al., 2010).

There is some potential support for the interaction between these two pathways. It was examined in vitro, vivo, and observed in human mononuclear cells that the Rac1 GTPase gene-transcription activity is down-regulated by an estrogen, 17 beta-estradiol which may be an important molecular mechanism contributing to the cardiovascular effects of estrogens (Laufs et al., 2003; Li et al., 2011). On the other hand, inhibition of the Rac1 decreases estrogen receptor levels (Rosenblatt et al., 2011). By examining the pathways ranked 11–20 for healthy and unhealthy dogs, as shown in Fig. 6, we found two pathways that are important from another view: "Actions of Nitric Oxide in the Heart" and "Synaptic Proteins at the Synaptic Junctions (BC)". In the network, they are connected to many other top-ranked pathways like a "hub".

That the pathway "Actions of Nitric Oxide in the Heart" plays an important role in vascular diseases is easy to understand since nitric oxide is an important regulator and mediator of numerous processes in the nervous, immune, and cardiovascular systems, including vascular smooth muscle relaxation, resulting in arterial vasodilation and increased blood flow. For the pathway "Synaptic Proteins at the Synaptic Junctions (BC)", research literatures related to vascular activity are

limited. However, some research has found that neurexins and neuroligins, which constitute large and complex families of fundamental players in synaptic activity, are produced and processed by endothelial and vascular smooth muscle cells throughout the vasculature. Moreover, they are dynamically regulated during vessel remodeling and form endogenous complexes in large vessels as well as in the brain (Bottos et al., 2009). Based on the large variety of biofunctions of nitric oxide and the extensive places of synapses in the animal or human body with their central station-styled transmission role, the active connections to other pathways could be expected.

6. Conclusion and discussion

In this paper, we have proposed a multilevel Gaussian graphical model (MGGM) in which one level describes the networks for genes and the other for pathways. A penalized likelihood approach is developed to achieve the sparseness on both levels. We provided an iterative weighted graphical LASSO algorithm for MGGM. In addition to some common criteria people used to evaluate model estimation and selection, we developed two criteria; pathway connection degree bias and pathway disconnection degree bias, to evaluate the performance of a model for the data analysis at the pathway level. Our simulation results supported the advantages of our approach; our method estimated the network more accurately on the pathway-level and more sparsely on the gene level. We also demonstrated the usefulness of our approach using a canine genes-pathways data set.

In our multilevel Gaussian graphical model, gene-level connection will contribute to the pathway-level connection. Notice that in practice, some genes may appear in many pathways, i.e. several pathways may share certain genes. In such situations, we may initially expect the gene connection to be numerically strong with a large correlation coefficient when their expression levels are identical. An initial guess will be that the two pathways may have a strong connection. However, this is not so obvious. Firstly, in a Gaussian graphical model, the gene and pathway dependencies are conditional on the association among other genes and pathways. For example, if pathways 1 and 2 have a shared gene A, the shared gene A will provide a contribution to the pathway effect. But how gene A connects with other genes in pathways 1, 2 and genes in other pathways will also affect its contribution to the connection between pathways 1 and 2. If gene A is weakly associated with all genes in pathways 1 and 2 but strongly associated with many genes in some other pathways, then the conditional dependency between gene A and itself is diminished. Then its contribution to the pathway connections maybe neglected. Secondly, the pathway connection depends on all gene connection between pathway pairs. This can be seen from the weighted graphical LASSO algorithm, where we add up the connection strength of all gene pairs between two pathways and take the square root of it and the reciprocal to get the penalty weight. Even if two pathways share a few genes, if other genes pairs between them are independent of each other, the sum of the connection strength will still be small and the weight will be large, which results in a large penalty on the pathway connection and thus the two pathways can still be weak. We note that our approach is based on a multilevel Gaussian graphical model for a single class. For example, we analyze the healthy and unhealthy dogs separately. Developing our approach with multiple classes will be worthwhile for the future research.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jspi.2017.05.003.

References

Bogdan, M., Doerge, R., Ghosh, J.K., 2004. Modifying the schwarz bayesian information criterion to locate multiple interacting quantitative trait loci. Genetics 167, 989–999.
Bottos, A., Destro, E., Rissone, A., Graziano, S., Cordara, G., Assenzio, B., Cera, M.R., Mascia, L., Bussolino, F., Arese, M., 2009. The synaptic proteins neurexins and neuroligins are widely expressed in the vascular system and contribute to its functions. Proc. Natl. Acad. Sci. 106, 20782–20787.
Broman, K.W., Speed, T.P., 2002. A model selection approach for the identication of quantitative trait loci in experimental crosses. JRSSB 64, 641–656.
Chen, J., Chen, Z., 2008. Extended bayesian information criteria for model selection with large model spaces. Biometrika 95, 759–771.
Coso, O.A., Chiariello, M., Yu, J.C., Teramoto, H., Crespo, P., Xu, N., Miki, T., Gutkind, J.S., 1995. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell 81, 1137–1146.
Dempster, A.P., 1972. Covariance selection. Biometrics 28, 157–175.
Drton, M., Foygel, R., 2010. Extended bayesian information criteria for gaussian graphical models. Adv. Neural Inf. Process. Syst. 23, 2020–2028.
Enerson, B.E., Lin, A., Lu, B., Zhao, H., Lawton, M.P., Floyd, E., 2006. Acute drug-induced vascular injury in beagle dogs: pathology and correlating genomic expression. Toxicol. Pathol. 34, 27–32.
Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.
Guo, J., Levina, E., Michailidis, G., Zhu, J., 2011. Joint estimation of multiple graphical models. Biometrika 98, 1–15.
Laufs, U., Adam, O., Strehlow, K., Wassmann, S., Konkol, C., Laufs, K., Schmidt, W., Bohm, M., Nickenig, G., 2003. Down-regulation of rac-1 GTPase by estrogen. J. Biol. Chem. 278, 5956–5962.
Levina, E., Rothman, A.J., Zhu, J., 2008. Sparse estimation of large covariance matrices via a nested lasso penalty. Ann. Appl. Stat. 2, 245–263.
Li, Y., Wang, J., Santen, R., Kim, T., Park, H., Fan, P., Yue, W., 2011. Estrogen stimulation of cell migration involves multiple signaling pathway interactions. Endocrinology 151, 5146–5156.
Louis, F.S., Zahradka, P., 2010. Vascular smooth muscle cell motility: from migration to invasion. Exp. Exp. Clin. Cardiol. 15, 175–185.
Masood, D.E., Roach, E.C., Beauregard, K.G., Khalil, R.A., 2010. Impact of sex hormone metabolism on the vascular effects of menopausal hormone therapy in cardiovascular disease. Curr. Drug Metab. 11 (8), 693–714.
Mendelsohn, E.M., 2002. Genomic and nongenomic effects of estrogen in the vasculature. Amer. J. Cardiol. 90, F3–F4.
Miller, M.V., Duckles, P.S., 2008. Vascular actions of estrogens: functional implications. Pharmacol. Rev. 60, 210–241.

Olson, M.F., Ashworth, A., Hall, A., 1995. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. Science 269, 1270–1272.
Pang, H., Lin, A., Holford, M., Enerson, B.E., Lu, B., Lawton, M.P., Floyd, E., Zhao, H., 2006. Pathway analysis using random forests classification and regression. Bioinformatics 22, 2028–2036.
Pankova, K., Rosel, D., Novotny, M., Brabek, J., 2010. The molecular mechanisms of transition between mesenchymal and amoeboid invasiveness in tumor cells. Cell. Mol. Life Sci. 67, 63–71.
Rosenblatt, A.E., Garcia, M.I., Lyons, L., Xie, Y., Maiorino, C., Desire, L., Slingerland, J., Burnstein, K.L., 2011. Inhibition of the Rho GTPase, Rac1, decreases estrogen receptor levels and is a novel therapeutic strategy in breast cancer. Endocr.-Relat. Cancer. 18, 207–219.
Siegmund, D., 2004. Model selection in irregular problems: application to mapping quantitative trait loci. Biometrika 91, 785–800.
Simpson, E.R., Misso, M., Hewitt, K.N., Hill, R.A., Boon, W.C., Jones, M.E., Kovacic, A., Zhou, J., Clyne, C.D., 2005. Estrogen–the good, the bad, and the unexpected. Endocr. Rev. 26, 322–330.
Smiley, A.D., Khalil, A.R., 2009. Estrogenic compounds, estrogen receptors and vascular cell signaling in the aging blood vessels. Curr. Med. Chem. 16, 1863–1887.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. 58, 267–288.
Yuan, M., Lin, Y., 2007. Model selection and estimation in the gaussian graphical model. Biometrika 94, 19–35.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429.
Zou, H., Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models. Ann. Statist. 36 (4), 1509–1533.