The Graphical R2D2 Estimator for the Precision Matrices

Dailin Gan[1], Guosheng Yin[2], and Yan Dora Zhang[2,3,*]

[1] *Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA*
[2] *Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong SAR, China*
[3] *Centre for PanorOmic Sciences, The University of Hong Kong, Hong Kong SAR, China*
[*] *To whom correspondence should be addressed. Tel: (852)3917-2473; Email: doraz@hku.hk*

Abstract

Biological networks are important for the analysis of human diseases, which summarize the regulatory interactions and other relationships between different molecules. Understanding and constructing networks for molecules, such as DNA, RNA and proteins, can help elucidate the mechanisms of complex biological systems. The Gaussian Graphical Models (GGMs) are popular tools for the estimation of gene regulatory networks because of their biological interpretability. Nonetheless, reconstructing GGMs from high-dimensional datasets is still challenging and current methods cannot handle the sparsity and high-dimensionality issues arising from datasets very well. Here we developed a new GGM, called the graphical R2D2 (R^2-induced Dirichlet Decomposition), based on the R2D2 priors for linear models. The proposed graphical R2D2 estimator has attractive properties in estimating the precision matrices, such as greater concentration near the origin and heavier tails than current shrinkage priors. When the true precision matrix is sparse and of high dimension, the graphical R2D2 provides the estimates with smallest information divergence from the sampling model. Besides, we also provide a full Gibbs sampler for implementing the graphical R2D2 estimator. R code is available for download from https://github.com/RavenGan/GR2D2. The graphical R2D2 estimator shows superior performance to existing techniques in simulations. We also provide breast cancer gene network analysis using the graphical R2D2 estimator and the important genes recognized from the inferred gene regulatory networks are consistent with biological ground truth.

1 Introduction

Biological networks, which capture the regulatory interactions or other relationships between molecules, such as DNA, RNA and protiens, are important for us to understand the mechanisms of complicated biological systems [1, 11]. The genetic regulatory networks, for

instance, have played important roles of analysing gene expression data, for they can capture the complex gene-gene interactions and make the inferences of the dependency structures of genes possible [26, 27]. The Gaussian Graphical Models (GGMs) have been a popular tool in genomics to describe the functional dependencies between genes of interests in graphs, since these models can circumvent indirect associations between genes by evaluating conditional dependencies in multivariate Gaussian distributions [14, 15, 23, 26]. GGMs have also been successfully applied to different scenarios to infer the genetic regulatory networks [4, 16, 20, 22, 24, 28]. In GGMs, we assume that the multivariate vector of genes follows a multivariate normal distribution with a particular structure of the inverse of the covariance matrix, called the precision matrix. Under this assumption, the conditional independence of two genes is equivalent to a zero value of the corresponding element in the precision matrix. The graph generated by GGMs can provide lucid interpretations of the relationship between genes in which nodes represent genes and edges between them represent interactions. The set of edges, which decides the network topology, can also be used to generate hypotheses about genetic interactions [15].

However, reconstructing GGMs from high-dimensional datasets remain a difficult task. The standard estimation of the partial correlation of genes either estimates the precision matrix or the parameters in p least squares regression problems, where p is the number of genes. If the number n of samples is much smaller than the number p of genes, both methods are inappropriate [13]. Possible alternatives are either using penalization or the Bayesian framework to estimate the precision matrix. Consider n samples from a p-dimensional multivariate normal distribution with zero mean and a $p \times p$ covariance matrix Ω^{-1}, which is

$$\boldsymbol{y}_k \sim \text{Normal}(\mathbf{0}, \Omega^{-1}),$$

where $k = 1, 2, \ldots, n$. Ω is the precision matrix, where the i, j-th off-diagonal element of Ω is the negative of the partial covariance between features i and j, and the i-th feature is regressed on all the other features [21]. Besides, under the multivariate normal assumption, zero off-diagonal elements in Ω are conditionally independent given the remaining features. One major challenge in estimating precision matrices is that the number of parameters increases quadratically with the number of features, leading to a sparse matrix. One approach to deal with sparsity is to penalized the likelihood functions. In graphical lasso [6], they maximize the penalized likelihood

$$\log(|\Omega|) - \text{tr}(S\Omega/n) - \sum_{i<j} \phi_\lambda(|\omega_{ij}|), \tag{1}$$

where $S = \sum_{i=1}^n \boldsymbol{y}_i^T \boldsymbol{y}_i$, $\Omega = (\omega_{ij})$, $\phi_\lambda(|\omega_{ij}|)$ is the l_1 penalty, and λ is a tuning parameter. A Bayesian version of the graphical lasso [25] is proposed afterwards, and the goal is to maximize a posterior estimate of Ω under the prior

$$p(\Omega|\lambda) \propto \prod_{i<j} (\text{DE}(\omega_{ij}|\lambda)) \prod_{i=1}^p (\exp(\omega_{ii}|\lambda/2)) \mathbf{1}_{\mathcal{S}_p},$$

where $\text{DE}(x|\lambda)$ denotes the double exponential distribution with rate λ, $\exp(x|\lambda)$ denotes the exponential distribution with rate λ, and $\mathcal{S}_p$ is the space of $p \times p$ positive definite matrices.

The smoothly clipped absolute deviation (SCAD) penalty [5] was also introduced to estimate the precision matrix, in which they maximized the penalized likelihood in equation 1 where the penalty term has the first-order derivative

$$\phi_{\lambda}^{'}(|x|) = \lambda(\mathbf{1}_{|x|\leq\lambda} + \frac{(a\lambda - |x|)}{(a-1)\lambda}\mathbf{1}_{|x|>\lambda}),$$

with $a > 2$ and $\lambda > 0$. However, the performance of this method relies on the choice of tuning parameters, which cannot be easily implemented in reality.
The graphical horseshoe estimator employs a Bayesian framework and puts horseshoe priors on the off-diagonal elements of precision matrices, and puts uninformative prior on the diagonal elements [17]. The model is summarized in the following

$$\begin{aligned} \omega_{ii} &\propto 1, \\ \omega_{ij;i<j} &\sim \text{Normal}(0, \lambda_{ij}^2\tau^2), \\ \lambda_{ij;i<j} &\sim C^+(0,1), \\ \tau &\sim C^+(0,1) \end{aligned}$$

where $C^+(0,1)$ denotes a half-cauchy random variable with density $p(x) \propto (1+x^2)^{-1}; x > 0$. The distinctive scale parameter λ_{ij} is treated as the local shrinkage parameter, and the scale parameter τ is shared by all dimensions, treated as the global shrinkage parameter.

The graphical lasso, Bayesian graphical lasso, and the graphical SCAD face the problem of accumulated estimation errors due to increasing number of parameters [17], while the graphical horseshoe estimator shows weak performance in high-dimensional settings when the number p of features is much larger than the sample size n. Here, we propose an alternative Bayesian estimator called the graphical R2D2 (R^2-induced Dirichlet Decomposition) estimator that can handle sparse and high-dimensional issues faced by other methods. The R2D2 shrinkage priors have shown promising performance in high-dimensional linear regression models, having greater concentration near the origin and heavier tails than current shrinkage priors [29]. We extended the R2D2 priors from linear models to graphical models and our simulation results show that the graphical R2D2 estimators achieved the state-of-art performance in sparse and high-dimensional settings, compared with the graphical lasso and the graphical horseshoe estimator.

The remaining sections are organized in the following. We outlines the graphical R2D2 estimator with a full Gibbs sampler for efficient sampling. Next we present our simulation studies to show the efficiency of the graphical R2D2 followed by real data analysis.

2 Methods

For Gaussian graphical models, it is usually assumed that the patterns of variation in expression of a gene can be predicted by those of a small subset of other genes, which leads to sparsity in the precision matrix [27]. For such precision matrix, we need a shrinkage method that is able to give zero or very small estimate for the zero elements. Besides, this method should also be able to sustain the nonzero values and shrink them as little as possible. We propose the graphical R2D2 estimator to achieve this goal.

2.1 The Graphical R2D2 Hierarchical Model

The graphical R2D2 model puts R2D2 priors on the off-diagonal elements of the precision matrix. The element-wise priors are specified for $i, j = 1, \ldots, p$ as follows

$$\begin{gathered}\omega_{ii} \propto 1, \quad \omega_{ij} \sim \mathrm{N}(0, \phi_{ij}\psi_{ij}\omega/2), \\ \boldsymbol{\phi} \sim \mathrm{Dir}(\alpha_\pi, \alpha_\pi, \ldots, \alpha_\pi), \quad \psi_{ij} \sim \exp(1/2), \\ \omega|\xi \sim \mathrm{Gamma}(a, \xi), \quad \xi \sim \mathrm{Gamma}(b, 1),\end{gathered}$$

where $\mathrm{N}(0,1)$ denotes a normal random variable with mean 0 and variance 1; $\mathrm{Dir}(\alpha_\pi, \alpha_\pi, \ldots, \alpha_\pi)$ denotes a vector of Dirichlet distributed random variables with a parameterized vector $\boldsymbol{\alpha} = [\alpha_\pi, \alpha_\pi, \ldots, \alpha_\pi]^T$; $\exp(1/2)$ denotes an exponential random variable with rate 1/2; $\mathrm{Gamma}(1,1)$ denotes a Gamma random variable with shape 1 and rate 1. The distinctive scale parameters ϕ_{ij} and ψ_{ij} on each dimension are referred to as the local shrinkage parameters, and the scale parameter ω shared by all dimensions is referred to as the global shrinkage parameter. ξ is used as an auxiliary parameter to help sample ω. Since conditional independence of each entry is guaranteed in precision matrices given the remaining entries under multivariate normal assumption, the prior on Ω under graphical R2D2 model can be written as the product of off-diagonal entries

$$p(\Omega|\boldsymbol{\phi}, \psi_{ij}, \omega) \propto \prod_{i<j} \mathrm{N}(\omega_{ij}|\phi_{ij}, \psi_{ij}, \omega)\mathbf{1}_{\Omega \in \mathcal{S}_p},$$

where $\mathcal{S}_p$ is the space of $p \times p$ positive definite matrices.

In high-dimensional precision matrix estimation using graphical R2D2, the global shrinkage parameter ω adapts to the sparsity of the entire matrix Ω and shrinks the estimates of the off-diagonal elements toward zero. On the flip side, the local shrinkage parameters, ϕ_{ij} and ψ_{ij} for $i < j$, preserve the magnitude of nonzero off-diagonal elements, and ensure that the element-wise biases are not very large.

2.2 A Data-Augmented Block Gibbs Sampler

Posterior samples under the graphical R2D2 hierarchical model are drawn by an augmented block Gibbs sampler, utilizing the scheme proposed by [18] for linear regression. In each iteration, each column and row of Ω and $\Lambda = (\phi_{ij}\psi_{ij})$ are partitioned from a $p \times p$ matrix of parameters and updated in a block. Then the global shrinkage parameter ω and its auxiliary variable ξ are updated.

The following part derives the posterior distribution of the precision matrix. Given data $Y_{n\times p}$ and the shrinkage parameters, we provide necessary derivation in the following. Define the following matrices

$$\Omega = \begin{bmatrix} \Omega_{(-p)(-p)} & \boldsymbol{\omega}_{(-p)p} \\ \boldsymbol{\omega}^T_{(-p)p} & \omega_{pp} \end{bmatrix}, \quad S = \begin{bmatrix} S_{(-p)(-p)} & \boldsymbol{s}_{(-p)p} \\ \boldsymbol{s}^T_{(-p)p} & s_{pp} \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \Lambda_{(-p)(-p)} & \boldsymbol{\lambda}_{(-p)p} \\ \boldsymbol{\lambda}^T_{(-p)p} & \lambda_{pp} \end{bmatrix},$$

where $(-p)$ denotes the set of all indices except for p, and $\Lambda_{(-p)(-p)}$ and $\boldsymbol{\lambda}_{(-p)p}$ have entries $\phi_{ij}\psi_{ij}$. Diagonal elements of $\Lambda_{(-p)(-p)}$ are set to be 1. The posterior distribution of Ω is

$$\begin{aligned} p(\Omega|\Lambda, Y, \omega) &\propto p(S|\Omega, \Lambda, \omega)p(\Omega|\Lambda, \omega) \\ &\propto |\Omega|^{n/2} \exp(-\frac{1}{2}tr(\Omega S)) \prod_{i<j} \exp(-\frac{\omega_{ij}^2}{\psi_{ij}\phi_{ij}\omega}), \end{aligned}$$

where $S = Y^T Y \sim W_p(\Omega^{-1}, n)$. After some calculation, the full conditional of the last column of Ω is

$$\begin{aligned} &p(\boldsymbol{\omega}_{(-p)p}, \omega_{pp}|Y, \Lambda, \omega, \Omega_{(-p)(-p)}) \\ &\qquad \propto (\omega_{pp} - \boldsymbol{\omega}_{(-p)p}^T \Omega_{(-p)(-p)}^{-1} \boldsymbol{\omega}_{(-p)p})^{n/2} \\ &\qquad \times \exp(-\frac{1}{2}(2\boldsymbol{s}_{\boldsymbol{(-p)p}}^T \boldsymbol{\omega}_{\boldsymbol{(-p)p}} + s_{pp}\omega_{pp})) \\ &\qquad \times \exp(-\boldsymbol{\omega}_{(-p)p}^T (\Lambda^* \omega)^{-1} \boldsymbol{\omega}_{(-p)p}), \end{aligned}$$

where $\Lambda^* = \text{diag}(\boldsymbol{\lambda}_{\boldsymbol{(-p)p}})$. Let $\boldsymbol{\beta} = \boldsymbol{\omega}_{(-p)p}$, $\gamma = \omega_{pp} - \boldsymbol{\beta}^T \Omega_{(-p)(-p)}^{-1} \boldsymbol{\beta}$, and $\omega_{pp} = \gamma + \boldsymbol{\beta}^T \Omega_{(-p)(-p)}^{-1} \boldsymbol{\beta}$. The Jacobian of this transformation is a constant, and the full conditional of β and γ is

$$\begin{aligned} &p(\boldsymbol{\beta}, \gamma|Y, \Lambda, \omega, \Omega_{(-p)(-p)}) \\ &\qquad \propto \gamma^{n/2} \exp(-\boldsymbol{\beta}^T (\Lambda^* \omega)^{-1} \boldsymbol{\beta}) \\ &\qquad \times \exp(-\frac{1}{2}(s_{pp}\gamma + s_{pp}\boldsymbol{\beta}^T \Omega_{(-p)(-p)}^{-1} \boldsymbol{\beta} + 2\boldsymbol{S}_{(-p)p}^T \boldsymbol{\beta})) \\ &\qquad \propto \gamma^{n/2} \exp(-\frac{1}{2}s_{pp}\gamma) \\ &\qquad \times \exp(-\frac{1}{2}(\boldsymbol{\beta} + D\boldsymbol{s}_{(-p)p})^T D^{-1} (\boldsymbol{\beta} + D\boldsymbol{s}_{(-p)p})), \end{aligned}$$

where $D = (2(\Lambda^* \omega)^{-1} + s_{pp}\Omega_{(-p)(-p)})^{-1}$. The above formula can be summarized as

$$\begin{aligned} p(\boldsymbol{\beta}, \gamma|Y, \Lambda, \omega, \Omega_{(-p)(-p)}) \propto &\text{Gamma}(\frac{n}{2} + 1, \frac{s_{pp}}{2}) \\ &\times \text{Normal}(-D\boldsymbol{s}_{(-p)p}, D). \end{aligned}$$

Hence, the posterior distribution of the last row and column of Ω is obtained. Since the derivation of the posterior distributions of ψ_{ij}, ω, ξ and $\boldsymbol{\phi}$ is standard and hence we summarize them in the following sampling procedures with proposed algorithm structure.

2.3 Gibbs sampling procedures with initial settings

Based on the results derived in the previous section, we provide the initial settings used in our sampling procedures. We use the same initial settings as [17] and [29], and set p to be the number of rows (or columns) in $S = Y^T Y$, $\Omega = I_{p\times p}$, $\Sigma = I_{p\times p}$, $\Phi = \mathbf{1}_{p\times p}$, $\Psi = \mathbf{1}_{p\times p}$, $\omega = 1$, $\xi = 1$, $b = 0.5$, $r = 1$, $C = 1$, $p_n = \frac{p(p-1)}{2}$, $\alpha_\pi = \frac{C}{p_n^{b/2} n^{rb/2} \log(n)}$, $a = \frac{p(p-1)}{2}\alpha_{pi}$, $\Lambda = \Phi \odot \Psi$, where Φ has entries of ϕ_{ij}, Ψ has entries of ψ_{ij}. The Gibbs sampling procedures are as follows:

(a) Set initial values as above.

(b) Choose the p-th column and do the following until the matrix Ω is updated.

(b.1) Sample $\gamma|Y, \Lambda, \omega, \Omega_{(-p)(-p)} \sim \text{Gamma}(\text{shape} = \frac{n}{2} + 1, \text{rate} = \frac{s_{pp}}{2})$.

(b.2) Sample $\boldsymbol{\beta}|Y, \Lambda, \omega, \Omega_{(-p)(-p)} \sim \text{Normal}(-D\boldsymbol{s}_{(-i)i}, D)$, where $D = (2(\Lambda^*\omega)^{-1} + s_{pp}\Omega_{(-p)(-p)})^{-1}$, $\Lambda^* = \text{diag}(\boldsymbol{\lambda}_{(-p)p})$, and $\lambda_{ij} = \phi_{ij}\psi_{ij}$.

(b.2) Sample $\psi_{ij}^{-1}|\omega_{ij}, \phi_{ij}, \omega \sim \text{IG}(\mu = \sqrt{\phi_{ij}\omega/2}/|\omega_{ij}|, \lambda = 1)$, and then take the reciprocal to get ψ_{ij}.

(c) Sample $\omega|\Omega, \Lambda, \xi \sim \text{giG}(\chi = 2\sum_{i<j} \frac{\omega_{ij}^2}{\phi_{ij}\psi_{ij}}, \rho = 2\xi, \lambda_0 = a - \frac{p(p-1)}{4})$.

(d) Sample $\xi|\omega \sim \text{Gamma}(\text{shape} = a + b, \text{rate} = \omega + 1)$

(e) Sample $\boldsymbol{\phi}|\Omega, \psi, \omega, \xi$. Based on above derivation, if $a = \frac{p(p-1)}{2}\alpha_\pi$, one can draw $T_1, \ldots, T_{\frac{p(p-1)}{2}}$ independently with $T_j \sim \text{giG}(\chi = 2\omega_j^2/\psi_j, \rho = 2\xi, \lambda_0 = \alpha_\pi - \frac{1}{2})$. Then set $\phi_j = T_j/T$ with $T = \sum_{j=1}^{\frac{p(p-1)}{2}} T_j$.

(f) Repeat steps (b)-(g) until convergence.

The algorithm 1 of the graphical R2D2, along with simulation examples, is available at https://github.com/RavenGan/GR2D2.

3 Simulation Studies

In the work of the graphical horseshoe [17], they have compared the performance of the graphical horseshoe with the graphical SCAD[5], frequentist graphical lasso [6] and the Bayesian graphical lasso [25] using diverse assessment criteria and different structures of the precision matrices. Their results showed that the graphical horseshoe estimator achieved the best performance. Therefore, we only compare the graphical R2D2 estimator with the graphical horseshoe estimator and the graphical lasso, where the graphical lasso works as the baseline.

3.1 Simulation Setting and Assessment Criterion

We consider four examples in our simulation studies with different pairs of features and observations. These pairs are (p, n) = (100, 50), (150, 50), (200, 50), and (250, 50). The reason of this setting is that we intend to compare the graphical R2D2 and other estimators when the ratio p/n increases, which corresponds to high-dimensional settings.

The diagonal entries of the precision matrix ω_0 is 1, while the off-diagonal entries are designed to be one of the following two patterns [7]. Actually there are three patterns mentioned in the work of [7], which are *Random* pattern, *Hubs* pattern and *Cliques* pattern. However, we only consider *Hubs* pattern and *Cliques* pattern, since when generating samples from *Random* pattern, the precision matrix may not always be positive definite and hence not invertible. Therefore, we may not get the sample covariance matrices from *Random* pattern,

Algorithm 1 The Graphical R2D2 Sampler

function GR2D2($S, n, burnin, nmc$) ▷ where $S = Y^T Y, n =$ sample size
  Set p to be the number of rows (or columns) in S
  Set initial values $\Omega = I_{p\times p}$, $\Sigma = I_{p\times p}$, $\Phi = \mathbf{1}_{p\times p}$, $\Psi = \mathbf{1}_{p\times p}$
  Set $\omega = 1$, $\xi = 1$, $b = 0.5$, $r = 1$, $C = 1$
  Set $p_n = \frac{p(p-1)}{2}$, $\alpha_\pi = \frac{C}{p_n^{b/2} n^{rb/2}\log(n)}$, $a = \frac{p(p-1)}{2}\alpha_{pi}$, $\Lambda = \Phi \odot \Psi$,
  where Φ has entris of ϕ_{ij}, Ψ has entris of ψ_{ij}
  for $iter = 1$ to (burnin + nmc) **do**
    for $i = 1$ to p **do**
      $\gamma \sim \text{Gamma}(\text{shape} = \frac{n}{2} + 1, \text{rate} = \frac{s_{ii}}{2})$
      $\Omega^{-1}_{(-i)(-i)} = \Sigma_{(-i)(-i)} - \boldsymbol{\sigma}_{(-i)i}\boldsymbol{\sigma}^T_{(-i)i}/\sigma_{ii}$
      $D = (s_{ii}\Omega^{-1}_{(-i)(-i)} + 2\text{diag}(\boldsymbol{\lambda}_{(-i)i}\omega)^{-1})^{-1}$
      $\boldsymbol{\beta} \sim \text{Normal}(-D\boldsymbol{s}_{(-i)i}, D)$ ▷ sample $\boldsymbol{\beta}$
      $\boldsymbol{\omega}_{(-i)i} = \boldsymbol{\beta}$, $\omega_{ii} = \gamma + \boldsymbol{\beta}^T\Omega^{-1}_{(-i)(-i)}\boldsymbol{\beta}$ ▷ variable transformation
      $\psi^{-1}_{ij} \sim \text{InvGaussian}(\mu_{ij} = \frac{\sqrt{\phi_{ij}\omega/2}}{|\omega_{ij}|}, \lambda = 1)$
      Take the reciprocal of ψ^{-1}_{ij} to get ψ_{ij} ▷ sample Ψ
      Save updated Ω
      $\Sigma_{(-i)(-i)} = \Omega^{-1}_{(-i)(-i)} + (\Omega^{-1}_{(-i)(-i)}\boldsymbol{\beta})(\Omega^{-1}_{(-i)(-i)}\boldsymbol{\beta})^T/\gamma$
      $\boldsymbol{\sigma}_{(-i)i} = -(\Omega^{-1}_{(-i)(-i)}\boldsymbol{\beta})/\gamma$, $\sigma_{ii} = 1/\gamma$
      Save updated Σ
    end for
    Motivated by [2], since $a = \frac{p(p-1)}{2}\alpha_\pi$
    $T_{ij} \sim \text{giG}(\chi = 2\omega^2_{ij}/\phi_{ij}, \rho = 2\xi, \lambda = \alpha_\pi - 1/2)$
    $\phi_{ij} = T_{ij}/T$, where $T = \sum_{i<j} T_{ij}$
    $\lambda_{ij} = \phi_{ij}\psi_{ij}$
    Save updated Λ
    $\omega \sim \text{giG}(\chi = \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\frac{2\omega^2_{ij}}{\phi_{ij}\psi_{ij}}, \rho = 2\xi, \lambda = a - \frac{p(p-1)}{4})$ ▷ sample ω
    $\xi \sim \text{Gamma}(\text{shape} = a + b, \text{rate} = \omega + 1)$ ▷ sample ξ
  end for
  Return MC samples Ω
end function

and hence we drop this pattern. Detailed explanations of *Hubs* pattern and *Cliques* pattern used in our simulation studies are provided in the following paragraphs. We take the pair $(p, n) = (100, 50)$ as an example.

Hubs. The rows or columns are partitioned into disjoint groups $\{G_k\}_1^K$. Each group has a row k where off-diagonal elements are taken to be $\omega_{ik} = 0.25$ (corresponding to negative partial correlations) for $i \in G_k$ and $\omega_{ij} =$ otherwise, and we allow four members within each group. We consider 10 groups and 10 members within each group, giving 40 nonzero off-diagonal elements in Ω_0. The topological structure of this pattern is like a star with four members connected with one member in the center.

Cliques. The rows or columns are partitioned into disjoint groups and $\{\omega_{ij} : i, j \in G_k, i \neq j\}$ are set to -0.45 corresponding to positive partial correlation case and set to 0.75 corresponding to negative partial correlation case. We also use 10 groups in this setting but only also three members within each group resulting in 30 nonzero off-diagonal elements in Ω_0. The topological structure of this pattern is like a small circle with only three points connected with each other.

The results of four pairs of observations and features and three patterns, *Hubs*, *Cliques positive* and *Cliques negative* are summarized in Tables 1 and 2. In these tables, Mean (SD), Stein's loss, Frobenius norm, precision, recall, false positive rate, accuracy, and F score of precision matrix estimation are generated accordingly.

For each choice of Ω_0, 30 datasets are generated and Ω is estimated using the graphical lasso, the graphical horseshoe and graphical R2D2 estimators. The graphical horseshoe estimator is implemented in MATLAB (2019a), while the graphical lasso is implemented using the package "glasso" in Rstudio (version 3.6.2). For the graphical horseshoe and the graphical R2D2 estimators, we drop the initial 500 burnin samples and keep the following 500 MCMC samples with convergence. We also use the median of the remaining 500 samples as our estimate for the graphical horseshoe and graphical R2D2.

Stein's loss of the estimated precision matrix $\hat{\Omega}$ is equal to 2 times the Kullback-Leibler divergence of $\hat{\Omega}$ from Ω_0, with the formula

$$\mathcal{L}_S(\hat{\Omega}, \Omega_0) = \mathrm{tr}(\hat{\Omega}\Omega_0^{-1}) - \log|\hat{\Omega}\Omega_0^{-1}| - p.$$

Frobenius norm between $\hat{\Omega}$ and Ω_0 is

$$\mathcal{L}_F(\hat{\Omega}, \Omega_0) = ||\hat{\Omega} - \Omega_0||_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}|\hat{\omega}_{ij} - \omega_{ij}|^2}.$$

Since the graphical horseshoe and the graphical R2D2 are shrinkage methods that do not estimate elements to be exactly equal to zero, we use the 95% quantile intervals in the MCMC samples for variable selection. This is to say, if the 95% quantile intervals of an off-diagonal entry of Ω does not contain zero, this entry is considered as a discovery and the median values across the MCMC samples are used as the estimate, and vice verse. For each measurement, we report the mean and standard deviation (s.d.) computed over 30 datasets.

3.2 Estimation

Based on Tables 1 and 2, the graphical R2D2 estimate always has the smallest Steins' loss and most of the time has the smallest Frobenius norm (F norm) with the smallest standard deviation among all the methods considered. When the ratio p/n is 2, the graphical horseshoe estimators show better performance than graphical R2D2 in F norm in the patterns cliques positive and cliques negative; and when the ratio between p and n is 3, the graphical horseshoe estimators show better performance than graphical R2D2 in F norm in the pattern cliques negative only. When p/n is 4 and 5, the graphical R2D2 shows consistent superior performance to the graphical horseshoe in all three patterns. The graphical R2D2 is expected to perform well when the precision matrix is sparse and the absolute values of scaled nonzero elements are large as mentioned in the work of [30]. Besides, the graphical R2D2 is able to deal with high dimensional datasets and always provides the best results as shown in $p/n = 4$ and $p/n = 5$.

For the simulations of the graphical Lasso, it is expected to perform the worst. Besides, it could be noted that as the ratio between p and n increases, the Stein's loss and F norm increases accordingly, which are also expected since the higher dimension and the sparser the datasets are, the harder it could be to estimate the precision matrices. Considering the standard deviation across 30 datasets, the graphical R2D2 always gives 0 values, indicating that the estimation of graphical R2D2 could provide the most concentrated estimation results with least variation. For the graphical horseshoe estimator, it shows relatively small variation while the graphical lasso provides the biggest variations.

3.3 Variable Selection

We first provide the formula of precision, recall, FPR, accuracy and F score and then discuss the results from Tables 1 to 2. Denote True Positive, False Positive, False Negative and True Negative as TP, FP, FN, and TN respectively. The formulae are

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}}$$
$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}$$
$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

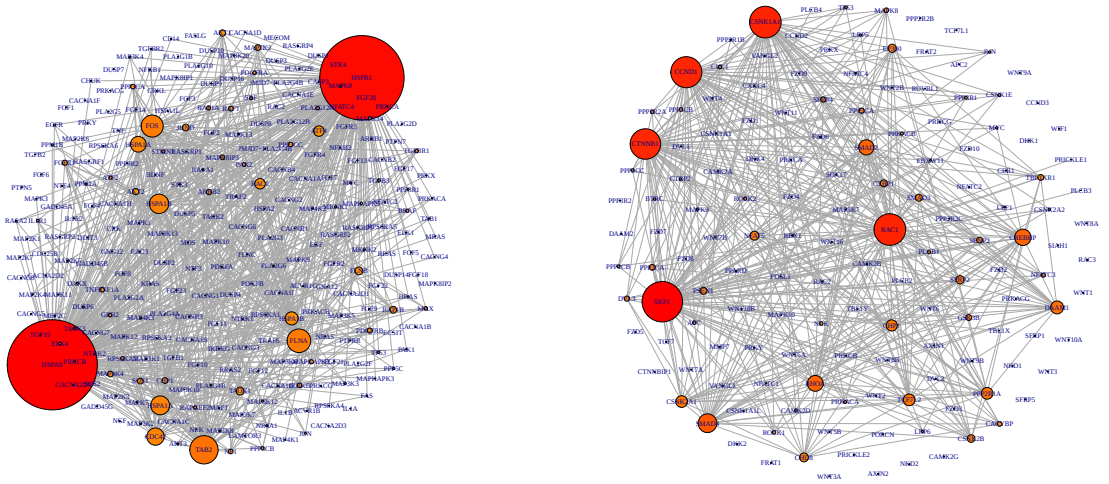
Similar to the results from Estimation section, the graphical R2D2 always provides 0 standard deviation across 30 datasets, indicating concentrated results with least variation. The graphical horseshoe estimators provide comparably small deviations while the graphical lasso provides larger deviations most of the time. Considering the precision score, the graphical R2D2 always gives 1, indicating the all the positive observations are true observations. The graphical horseshoe provides results close to 1, showing comparably good results, while the graphical lasso provides quite small precision scores. For the scores of recall, the graphical lasso always gives the best recall scores while the results of graphical horseshoe and the graphical R2D2 improve as the ration between p and n increases. Besides, the graphical

horseshoe estimators most of the time give better scores than the graphical R2D2. For FPR, the graphical R2D2 always gives 0 values, indicating that there are 0 false positive result across all 30 datasets. The graphical horseshoe estimators provide results quite close to zero across 30 datasets, showing comparably good results to the graphical R2D2. However, the graphical lasso always provides the worst FPR scores with values over 0.7, indicating that this method always provides false positive estimations. For the accuracy scores, both the graphical horseshoe and the graphical R2D2 show comparably good results with values more than 0.99, while the graphical lasso provides poorer accuracy results. One point that is worth mentioning is that since the precision matrices are quite sparse, there are always much less observations than zero values, constructing unbalanced datasets. Therefore, the accuracy scores may not be a good measurement to assess the performance of three estimation methods. Next we turn to F scores, which can cooperate the results of precision and recall, giving measurement free from the effects of unbalanced datasets. According to results of F scores, the graphical horseshoe method always provides the best F score, then the graphical R2D2 and the graphical lasso at last. As the ratio between p and n increases, the F scores of the graphical horseshoe and the graphical R2D2 increase accordingly, while the graphical lasso shows limited increment. It is noteworthy that as the ratio between p and n increases, the difference between the F scores from the graphical horseshoe and graphical R2D2 decreases.

4 Real data analysis

We analyze the breast cancer dataset from The Metastatic Breast Cancer Project downloaded from the cBioPortal [3, 8]. Our analysis includes four signaling pathways, which are MAPK signaling pathway (hsa04010), Wnt signaling pathway (hsa04310), p53 signaling pathway (hsa04115), and cell cycle (hsa04110). All the pathways were downloaded from the Kyoto Encyclopedia of Genes and Genomes database[12]. We use RNA-sequencing data as the raw data which contains 26807 genes and 146 patients. We only use those genes which appear both in the raw data and signaling pathways, resulting in 265 genes for MAPK, 149 genes for Wnt, 68 genes for p53, and 112 genes for cell cycle. We assume the gene expressions of individuals are identically distributed with a multivariate normal distribution. We used 25% quantile intervals in the MCMC samples for variable selection. If 0 is not included in the quantile interval, we treated it as a discovery. For real data analysis, we used burn-in and the number of MCs both equal to 500.

The inferred graphs by graphical R2D2 of four pathways are shown in Figures 1 (a), 1 (b), 2 (a), and 2 (b). There are 672 edges in MAPK, 421 edges in Wnt, 248 edges in p53 and 521 edges in cell cycle. We also reported the top 10 genes which are densely connected with other genes in tables 3 and 4. As shown in Wnt, p53 and cell cycle pathways, the oncogene CCND1 is always among the top ten densely connected genes, indicating patients may have CCND-1 driven breast cancer. The tumor suppressor gene PTEN is also identified in p53 pathway. HSPA8, as identified as the most densely connected gene in MAPK pathway, was found significantly changed in breast cancer [9]. The gene SKP1, which is an essential part of SCF (SKP1-CUL1-F-box protein) E3 ubiquitin ligase complex associated with breast cancer patients [10], was also identified as the most densely connected gene in Wnt pathway.



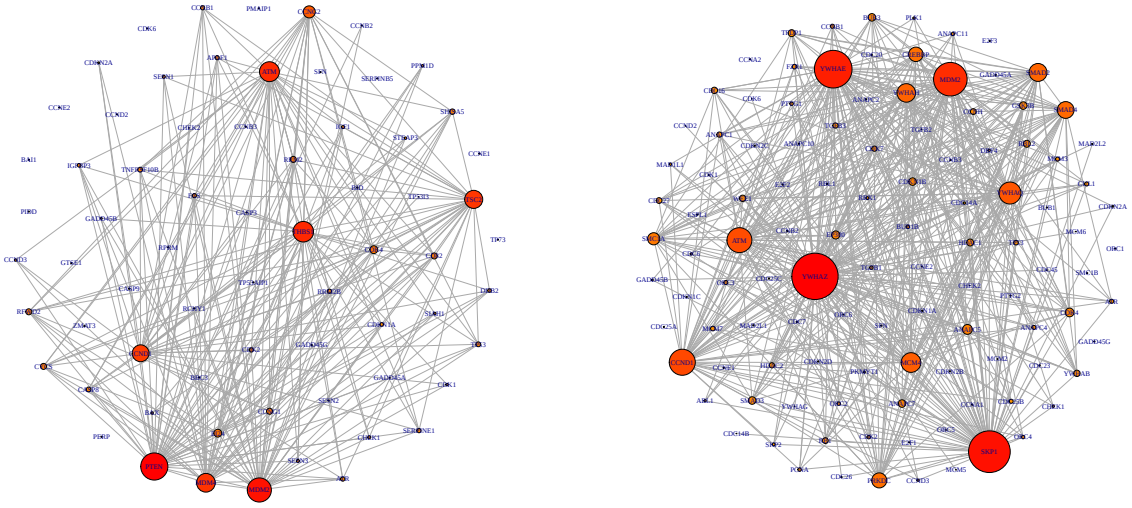
(a) MAPK (b) Wnt

Figure 1: The inferred graph for the breast cancer dataset by the graphical R2D2 (GR2D2) estimates, considering the MAPK and Wnt signaling pathways. The size of node is proportional to degrees within each group.

YWHAZ, identified in cell cycle pathway was found to promote cell migration in breast cancer and has interactions with gene DAAM1, which is recognized in Wnt pathway[19]. Based on our real data analysis and breast cancer gene literature, we can notice that the graphical R2D2 can capture the genetic relationship which has been validated by biological experiments. Moreover, this method can provide insights to those genes which have the potential to affect breast cancer but have not been verified experimentally.

5 Discussion

In high-dimensional data analysis, how to estimate the precision matrix is always an attractive challenge under multivariate Gaussian assumption. In this work, we proposed the graphical R2D2 estimator with easy implementation by a full Gibbs sampler. By using a prior with high density near the origin and a combination of an exponential distributed parameter and a Dirichlet distributed parameter on each dimension, the graphical R2D2 hierarchical model provides estimates close to the true distribution in Kullback-Leibler divergence and with the smallest bias for nonzero elements among current published methods. Simulation studies also confirm that the graphical R2D2 outperforms alternative methods like the graphical horseshoe and the graphical lasso in different settings. Furthermore, real data analysis shows that our graphical R2D2 can capture important genetic relationship and have the ability to provide insights for future experiment design. Therefore, we may conclude that the graphical



(a) p53 (b) Cell cycle

Figure 2: The inferred graph for the breast cancer dataset by the graphical R2D2 (GR2D2) estimates, considering the p53 and Cell cycle signaling pathways. The size of node is proportional to degrees within each group.

R2D2 could provide the best precision matrix estimation in sparse and high-dimensional settings, and infer meaningful genetic relationships.

6 Data availability

Data used in this study were downloaded from public databases that have been described in the main text.

7 Supplementary information and code

Supplementary information are available online with detailed derivation of GR2D2. The source code is available at https://github.com/RavenGan/GR2D2.

$p = 100$ $n = 50$	Hubs GL	Hubs GHS	Hubs GR2D2	Cliques positive GL	Cliques positive GHS	Cliques positive GR2D2	Cliques negative GL	Cliques negative GHS	Cliques negative GR2D2
F norm	82.44 (0.51)	3.44 (0.14)	**3.22** (**0.11**)	86.33 (0.67)	**4.87** (0.22)	5.01 (**0.12**)	86.72 (0.53)	**6.77** (0.21)	6.94 (**0.06**)
Stein's loss	484.85 (4.40)	6.71 (0.43)	**5.86** (**0.34**)	793.87 (16.95)	58.66 (4.71)	**53.12** (**2.65**)	679.98 (10.21)	31.62 (2.00)	**23.95** (**1.18**)
Precision	0.0224 (0.0006)	0.9981 (0.0059)	**1** (**0**)	0.0202 (0.0003)	0.9969 (0.0074)	**1** (**0**)	0.0207 (0.0002)	0.9989 (0.0044)	**1** (**0**)
Recall	**0.9574** (0.0261)	0.5600 (0.0075)	0.5556 (**0**)	**0.9708** (0.0155)	0.7967 (0.0222)	0.6250 (**0**)	**0.9963** (0.0067)	0.7142 (0.0191)	0.6250 (**0**)
FPR	0.7654 (0.0056)	2.04e-05 (6.21e-5)	**0** (**0**)	0.7649 (0.2468)	4.07e-5 (9.84e-5)	**0** (**0**)	0.7667 (0.0056)	1.36e-5 (5.16e-5)	**0** (**0**)
Accuracy	0.2476 0.0055	**0.9921** 1.30e-4	0.9920 **0**	0.2468 (0.0051)	**0.9967** (3.59e-4)	0.9940 (**0**)	0.2456 (0.0055)	**0.9954** (3.10e-4)	0.9940 (**0**)
F score	0.4380 (0.0012)	**0.7170** (0.0058)	0.7143 (**0**)	0.3961 (0.0396)	**0.8854** (0.0137)	0.7692 (**0**)	0.0405 (0.0003)	**0.8327** (0.0132)	0.7692 (**0**)

$p = 150$ $n = 50$	Hubs GL	Hubs GHS	Hubs GR2D2	Cliques positive GL	Cliques positive GHS	Cliques positive GR2D2	Cliques negative GL	Cliques negative GHS	Cliques negative GR2D2
F norm	140.70 (0.39)	4.18 (0.25)	**3.65** (**0.17**)	146.46 (0.62)	5.40 (0.19)	**5.27** (**0.11**)	146.15 (0.63)	**7.08** (0.18)	7.13 (**0.08**)
Stein's loss	1154.85 (4.72)	8.71 (0.71)	**6.92** (**0.49**)	1631.26 (19.65)	61.82 (3.43)	**54.25** (**1.86**)	1468.33 (11.24)	33.34 (1.96)	**24.44** (**1.25**)
Precision	0.0152 (0.0003)	0.9965 (0.0068)	**1** (**0**)	0.0142 (0.0002)	0.9970 (0.0058)	**1** (**0**)	0.0146 (0.0002)	0.9976 (0.0048)	**1** (**0**)
Recall	**0.9497** (0.0197)	0.6545 (0.0051)	0.6522 (**0**)	**0.9643** (0.0118)	0.8435 (0.0192)	0.7143 (**0**)	**0.9912** (0.0090)	0.7911 (0.0214)	0.7143 (**0**)
FPR	0.6355 (0.0022)	2.39e-5 (4.68e-5)	**0** (**0**)	0.6315 (0.0029)	2.39e-5 (4.67e-5)	**0** (**0**)	0.6314 (0.0041)	1.79e-5 (3.65e-5)	**0** (**0**)
Accuracy	0.3704 (0.0022)	**0.9964** (7.38e-5)	**0.9964** (**0**)	0.3741 (0.0029)	**0.9985** (1.94e-4)	0.9973 (**0**)	0.3744 (0.0041)	**0.9980** (1.96e-4)	0.9973 (**0**)
F score	0.0299 (0.0006)	**0.7901** (0.0045)	0.7895 (**0**)	0.0280 (0.0004)	**0.9137** (0.0120)	0.8333 (**0**)	0.0287 (0.0003)	**0.8823** (0.0131)	0.8333 (**0**)

Table 1: Frobenius norm, Mean Stein's loss, precision, recall, false positive rates, accuracy and F scores of precision matrix estimates over 30 datasets generated by multivariate normal distributions with precision matrix Ω, where $p = 100$ and 150, $n = 50$. The precision matrix is estimated by frequentist graphical lasso (GL), the graphical horseshoe (GHS), and the graphical R2D2 (GR2D2). The best performance in eacho row is shown in bold.

$p = 200$ $n = 50$	Hubs GL	Hubs GHS	Hubs GR2D2	Cliques positive GL	Cliques positive GHS	Cliques positive GR2D2	Cliques negative GL	Cliques negative GHS	Cliques negative GR2D2
F norm	191.18 (0.65)	4.97 (0.30)	**3.95** (**0.20**)	197.90 (0.73)	5.98 (0.27)	**5.51** (**0.11**)	197.08 (0.70)	7.70 (0.31)	**7.31** (**0.09**)
Stein's loss	1922.52 (7.89)	11.65 (1.06)	**8.08** (**0.64**)	2528.43 (29.99)	64.99 (4.50)	**55.13** (**3.49**)	2327.16 (13.09)	38.18 (2.79)	**26.32** (**1.21**)
Precision	0.0124 (0.0002)	0.9948 (0.0084)	**1** (**0**)	0.0117 (0.0001)	0.9968 (0.0043)	**1** (**0**)	0.0121 (**0**)	0.9945 (0.0062)	**1** (**0**)
Recall	**0.9524** (0.0155)	0.7169 (0.0040)	0.7143 (**0**)	**0.9650** (0.0105)	0.8751 (0.0124)	0.7692 (**0**)	**0.9920** (0.0066)	0.8277 (0.0141)	0.7692 (**0**)
FPR	0.5357 (0.0024)	2.69e-05 (4.33e-5)	**0** (**0**)	0.5310 (0.0026)	1.85e-5 (2.47e-5)	**0** (**0**)	0.5315 (0.0021)	3.40e-5 (3.40e-5)	**0** (**0**)
Accuracy	0.4678 (0.0024)	**0.9980** (5.10e-5)	**0.9980** (**0**)	0.4722 (0.0026)	**0.9992** (8.67e-5)	0.9985 (**0**)	0.4719 (0.0021)	**0.9989** (9.19e-5)	0.9985 (**0**)
F score	0.0287 (0.0003)	**0.8333** (0.0040)	**0.8333** (**0**)	0.0232 (0.0003)	**0.9320** (0.0075)	0.8696 (**0**)	0.0238 (0.0002)	**0.9034** (0.0084)	0.8696 (**0**)

$p = 250$ $n = 50$	Hubs GL	Hubs GHS	Hubs GR2D2	Cliques positive GL	Cliques positive GHS	Cliques positive GR2D2	Cliques negative GL	Cliques negative GHS	Cliques negative GR2D2
F norm	236.34 (0.66)	5.97 (0.25)	**4.29** (**0.18**)	243.61 (0.69)	6.82 (0.22)	**5.75** (**0.11**)	242.14 (0.97)	8.27 (0.37)	**7.51** (**0.13**)
Stein's loss	2745.06 (8.21)	15.61 (1.04)	**9.12** (**0.57**)	3444.20 (14.67)	69.48 (3.77)	**54.75** (**1.98**)	3220.52 (15.80)	41.84 (3.68)	**27.78** (**1.94**)
Precision	0.0108 (0.0002)	0.9921 (0.0082)	**1** (**0**)	0.0105 (0.0001)	0.9948 (0.0049)	**1** (**0**)	0.0106 (0.0001)	0.9938 (0.0052)	**1** (**0**)
Recall	**0.9520** (0.0169)	0.7590 (0.0031)	0.7576 (**0**)	**0.9701** (0.0097)	0.8946 (0.0137)	0.8065 (**0**)	**0.9889** (0.0084)	0.8576 (0.0106)	0.8065 (**0**)
FPR	0.4617 (0.0018)	3.22e-5 (3.38e-5)	**0** (**0**)	0.4574 (0.0013)	2.36e-5 (2.22e-5)	**0** (**0**)	0.4583 (0.0014)	2.68e-5 (2.25e-5)	**0** (**0**)
Accuracy	0.5405 (0.0018)	**0.9987** (3.43e-5)	**0.9987** (**0**)	0.5446 (0.0013)	**0.9995** (7.27e-5)	0.9990 (**0**)	0.5439 (0.0014)	**0.9993** (6.02e-5)	0.9990 (**0**)
F score	0.0214 (0.0004)	0.8600 (0.0033)	**0.8621** (**0**)	0.0207 (0.0002)	**0.9420** (0.0081)	0.8929 (**0**)	0.0211 (0.0002)	**0.9207** (0.0069)	0.8929 (**0**)

Table 2: Frobenius norm, Mean Stein's loss, precision, recall, false positive rates, accuracy and F scores of precision matrix estimates over 30 datasets generated by multivariate normal distributions with precision matrix Ω, where $p = 200$ and 250, $n = 50$. The precision matrix is estimated by frequentist graphical lasso (GL), the graphical horseshoe (GHS), and the graphical R2D2 (GR2D2). The best performance in eacho row is shown in bold.

MAPK	Degree	Wnt	Degree
HSPA8	172	SKP1	69
HSPB1	161	CSNK1A1	54
TAB2	54	RAC1	54
FLNA	45	CCND1	53
FOS	42	CTNNB1	52
HSPA1B	37	SMAD4	32
HSPA1A	36	CREBBP	28
CDC42	33	RHOA	27
ATF4	20	SMAD2	26
RAC1	20	DAAM1	22

Table 3: Top 10 densely connected genes in MAPK and Wnt

p53	Degree	cell cycle	Degree
PTEN	45	YWHAZ	80
MDM2	40	SKP1	72
THBS1	34	YWHAE	65
ATM	33	MDM2	58
MDM4	31	CCND1	45
TSC2	30	ATM	43
CCND1	28	YWHAQ	38
CCNG2	20	MCM4	34
CDK4	13	YWHAH	32
EI24	13	SMAD2	31

Table 4: Top 10 densely connected genes in p53 and cell cycle

References

[1] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.

[2] Anirban Bhattacharya, Debdeep Pati, Natesh S Pillai, and David B Dunson. Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490, 2015.

[3] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J Byrne, Michael L Heuer, Erik Larsson, et al. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, 2012.

[4] Adrian Dobra, Chris Hans, Beatrix Jones, Joseph R Nevins, Guang Yao, and Mike West. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90(1):196–212, 2004.

[5] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[7] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Applications of the lasso and grouped lasso to the estimation of sparse graphical models. Technical report, Technical report, Stanford University, 2010.

[8] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, et al. Inte-

grative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Science signaling*, 6(269):pl1–pl1, 2013.

[9] Lingmi Hou, Maoshan Chen, Minghao Wang, Xiang Cui, Yanchun Gao, Tianyong Xing, Jingdong Li, Shishan Deng, Jiani Hu, Hongwei Yang, et al. Systematic analyses of key genes and pathways in the development of invasive breast cancer. *Gene*, 593(1):1–12, 2016.

[10] Ye-Fei Huang, Zhe Zhang, Meng Zhang, Yan-Su Chen, Jun Song, Ping-Fu Hou, Hong-Mei Yong, Jun-Nian Zheng, and Jin Bai. Cul1 promotes breast cancer metastasis through regulating ezh2-induced the autocrine expression of the cytokines cxcl8 and il11. *Cell death & disease*, 10(1):1–14, 2018.

[11] Trey Ideker and Nevan J Krogan. Differential network biology. *Molecular systems biology*, 8(1):565, 2012.

[12] Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[13] Nicole Krämer, Juliane Schäfer, and Anne-Laure Boulesteix. Regularized estimation of large-scale gene association networks using graphical gaussian models. *BMC bioinformatics*, 10(1):1–24, 2009.

[14] Jan Krumsiek, Karsten Suhre, Thomas Illig, Jerzy Adamski, and Fabian J Theis. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC systems biology*, 5(1):1–16, 2011.

[15] Gwenaël GR Leday, Mathisca CM De Gunst, Gino B Kpogbezan, Aad W Van der Vaart, Wessel N Van Wieringen, and Mark A Van De Wiel. Gene network reconstruction using global-local shrinkage priors. *The annals of applied statistics*, 11(1):41, 2017.

[16] Hongzhe Li and Jiang Gui. Gradient directed regularization for sparse gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics*, 7(2):302–317, 2006.

[17] Yunfan Li, Bruce A Craig, and Anindya Bhadra. The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, 28(3):747–757, 2019.

[18] Enes Makalic and Daniel F Schmidt. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182, 2015.

[19] Jie Mei, Yan Liu, Xinqian Yu, Leiyu Hao, Tao Ma, Qiang Zhan, Yan Zhang, and Yichao Zhu. Ywhaz interacts with daam1 to promote cell migration in breast cancer. *Cell death discovery*, 7(1):1–12, 2021.

[20] Vasyl Pihur, Somnath Datta, and Susmita Datta. Reconstruction of genetic association networks from microarray data: a partial least squares approach. *Bioinformatics*, 24(4):561–568, 2008.

[21] Mohsen Pourahmadi. Covariance estimation: The glm and regularization perspectives. *Statistical Science*, pages 369–387, 2011.

[22] Juliane Schäfer and Korbinian Strimmer. An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764, 2005.

[23] Juliane Schäfer and Korbinian Strimmer. Learning large-scale graphical gaussian models from genomic data. In *AIP Conference Proceedings*, volume 776, pages 263–276. American Institute of Physics, 2005.

[24] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

[25] Hao Wang et al. Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, 7(4):867–886, 2012.

[26] Huangdi Yi, Qingzhao Zhang, Cunjie Lin, and Shuangge Ma. Information-incorporated gaussian graphical model for gene expression data. *Biometrics*, 2021.

[27] Jianxin Yin and Hongzhe Li. A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The annals of applied statistics*, 5(4):2630, 2011.

[28] Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

[29] Yan Dora Zhang, Brian P Naughton, Howard D Bondell, and Brian J Reich. Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior. *Journal of the American Statistical Association*, pages 1–13, 2020.

[30] Zhen Zhang, Jiajun Bu, Martin Ester, Jianfeng Zhang, Chengwei Yao, Zhi Yu, and Can Wang. Hierarchical graph pooling with structure learning. *arXiv preprint arXiv:1911.05954*, 2019.