


Article

Testing Differential Gene Networks under Nonparanormal Graphical Models with False Discovery Rate Control

Qingyang Zhang 

Department of Mathematical Sciences, University of Arkansas, Arkansas, AR 72701, USA; qz008@uark.edu

Received: 20 January 2020; Accepted: 30 January 2020; Published: 5 February 2020



Abstract: The nonparanormal graphical model has emerged as an important tool for modeling dependency structure between variables because it is flexible to non-Gaussian data while maintaining the good interpretability and computational convenience of Gaussian graphical models. In this paper, we consider the problem of detecting differential substructure between two nonparanormal graphical models with false discovery rate control. We construct a new statistic based on a truncated estimator of the unknown transformation functions, together with a bias-corrected sample covariance. Furthermore, we show that the new test statistic converges to the same distribution as its oracle counterpart does. Both synthetic data and real cancer genomic data are used to illustrate the promise of the new method. Our proposed testing framework is simple and scalable, facilitating its applications to large-scale data. The computational pipeline has been implemented in the R package *DNetFinder*, which is freely available through the Comprehensive R Archive Network.

Keywords: gene regulatory network; nonparanormal graphical model; network substructure; false discovery rate control

1. Background

Inferring the structural change of a network under different conditions is essential in many problems arising in biology, medicine, and other scientific fields. For instance, in genomics, it is often of importance to study the structural change of a genetic pathway between diseased and normal groups. In the field of brain mapping, it is critical to identify the difference in brain connectivity between groups (for example, the brain connectivity network of normal subjects and patients often possess different structures). Most of these applications have relied on the prevailing Gaussian graphical models (GGMs) because of its good interpretability and computational convenience, and there is a rich and growing literature on learning differential networks under GGMs. To name a few, Guo et al. (2015) [1] introduced a joint estimation for multiple GGMs by a group lasso approach, under the assumption that the GGMs being studied are sparse and only differ in a small portion of edges. Danaher et al. (2014) [2] proposed a fused graphical lasso method which is free from the sparsity assumption on condition-specific networks and only requires the sparsity of the differential network. Zhao et al. (2014) [3] constructed a new estimator which directly estimates the differential network defined as $\Delta = \Sigma_X^{-1} - \Sigma_Y^{-1}$, where Σ_X^{-1} and Σ_Y^{-1} represent the two condition-specific precision matrices and Δ , Σ_X^{-1} , Σ_Y^{-1} have the same dimension. Liu (2017) [4] presented a new test to simultaneously study structural similarities and differences between multiple high-dimensional GGMs, which adopts the partial correlation coefficients to characterize the potential changes of dependency strength between two variables.

Most of the aforementioned algorithms were based upon penalized likelihood maximization. Although some algorithms were consistent under certain regularity conditions, they failed to control

the false discovery rate (FDR) of the substructure detection as it is difficult to choose a tuning parameter to control the FDR at the desired level [1–3]. One exception is Liu (2017), who introduced a hierarchical testing framework to adjust for the multiplicity. Liu’s test was constructed to asymptotically control the FDR while keeping satisfactory statistical power. Simulation studies in [4] have shown that this new test exhibits substantial power gains over existing methods such as graphical lasso. One major drawback that limits the application of Liu’s test is the Gaussian assumption, which is often violated in practice especially in genomics. For instance, some digital measurements of gene expression level such as RNA-Seq data often greatly deviate from normality even after log-transformation or other variance-stabilizing transformations. In this paper, we aim to extend Liu’s work to a more flexible semiparametric framework, namely the nonparanormal graphical models (NPNGMs), where the random variables are assumed to follow a multivariate normal distribution after a set of monotonically increasing transformations. We use a novel rank-based multiple testing method to detect the structural difference between multiple networks from non-Gaussian data. The method is computationally efficient and asymptotically controls the FDR at a desired level. To begin with, we give the formal definition of nonparanormal distribution:

Definition 1. A random vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$ follows a nonparanormal distribution if there exists a set of univariate and monotonically increasing transformations, $\mathbf{f} = (f_1, \dots, f_p)$, such that:

$$(X_1, \dots, X_p) \equiv (f_1(Y_1), \dots, f_p(Y_p)) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance matrix in the multivariate normal distribution, respectively. The distribution of \mathbf{Y} depends on three parameters and it can be generally written as $\mathbf{Y} \sim \text{NPN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{f})$.

By Definition 1 and Sklar’s theorem, it is easy to verify that when the transformation functions f_j ’s are all differentiable, the nonparanormal distribution $\text{NPN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{f})$ is equivalent to a Gaussian copula [5]. As graphical models, the NPNGMs are much more flexible than GGMs in modeling non-Gaussian data while retaining the interpretability of the latter. Some recent studies have established the estimation and properties of high dimensional nonparanormal graphical models. For example, Liu et al. (2009) [5], who first studied high-dimensional NPNGMs, bridged the estimations of GGMs and NPNGMs by a nonparametric and truncated (Winsorized) estimator of the unknown transformation functions. Xue and Zou (2012) [6] proposed to use an adjusted Spearman’s correlation to estimate the structure of high-dimensional NPNGMs, and they showed that the rank-based estimator achieves the same rate of convergence as its oracle counterpart (i.e., assuming known transformation functions). Despite the advances in single NPNGM estimation, to the best of our knowledge, the inference of differential substructure between multiple NPNGMs has not been studied. In this paper, we tackled this problem by embedding the Winsorized estimator into the testing framework of Liu (2017). Under some regularity conditions, we showed that the new test statistic converges to the same distribution as its oracle counterpart does [4].

We begin with the notations and problem formulation. For a vector $\mathbf{a} = (a_1, \dots, a_p)$, we define its ℓ_0 norm as $\|\mathbf{a}\|_{\ell_0} = \sum_{i=1}^p I\{a_i \neq 0\}$, its ℓ_1 norm as $\|\mathbf{a}\|_{\ell_1} = \sum_{i=1}^p |a_i|$, its ℓ_2 norm as $\|\mathbf{a}\|_{\ell_2} = \sqrt{\sum_{i=1}^p a_i^2}$, and its ℓ_∞ norm as $\|\mathbf{a}\|_{\ell_\infty} = \max_i |a_i|$. For a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times q}$, we define its ℓ_0 norm as $\|\mathbf{A}\|_0 = \sum_{i,j} I\{a_{ij} \neq 0\}$, its ℓ_1 norm as $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{ij}|$, its Frobenius norm as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ and its ℓ_∞ norm as $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{ij}|$. Let $\mathbf{A}_{i,-j}$ denote the i th row of \mathbf{A} with its j th entry being removed and $\mathbf{A}_{-i,j}$ denote the j th column with its i th entry being removed. We use $\mathbf{A}_{-i,-j}$ to denote a $(p-1) \times (q-1)$ matrix by removing the i th row and the j th column. For square matrix \mathbf{B} , we let $\lambda_{\max}(\mathbf{B})$ and $\lambda_{\min}(\mathbf{B})$ denote the largest and smallest eigenvalues of \mathbf{B} respectively. In addition, for a given sequence of random variable $\{X_n, n = 1, 2, \dots\}$ and a constant sequence $\{a_n, n = 1, 2, \dots\}$, $X_n = o_p(a_n)$ denotes that X_n/a_n converges to zero in probability as n approaches to infinity and $X_n = O_p(a_n)$ denotes that X_n/a_n

is stochastically bounded. If there are positive constants c and C such that $c \leq X_n/a_n \leq C$ for all $n \geq 1$, we write $X_n \sim a_n$.

To formulate the problem, we let $k \in \{1, 2, \dots, K\}$ be the index of class, p be the dimension, and $(Y_1^{(k)}, \dots, Y_{n_k}^{(k)})$ be a sample of size n_k for class k where $Y_m^{(k)} = (Y_{m1}^{(k)}, \dots, Y_{mp}^{(k)})^T \in \mathbb{R}^p$, $m \in \{1, \dots, n_k\}$. Under $Y_m^{(k)} \sim NPN(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}, \mathbf{f}^{(k)})$, we test the following hypothesis:

$$\begin{aligned} H_{0ij} &: \rho_{ij\cdot}^{(1)} = \rho_{ij\cdot}^{(2)} = \dots = \rho_{ij\cdot}^{(K)}, \\ H_{aij} &: \rho_{ij\cdot}^{(k)} \neq \rho_{ij\cdot}^{(k')}, \text{ for some } k, k' \in \{1, \dots, K\}, \end{aligned}$$

where $1 \leq i, j \leq p$, $\{\boldsymbol{\Sigma}^{(k)}\}^{-1} = \boldsymbol{\Omega}^{(k)} = (\omega_{ij}^{(k)})$, and $\rho_{ij\cdot}^{(k)}$ represents the partial correlation coefficient between $X_i^{(k)}$ and $X_j^{(k)}$ given $\mathbf{X}^{(k)} \setminus (X_i^{(k)}, X_j^{(k)})$, $(X_{m1}^{(k)}, \dots, X_{mp}^{(k)}) = (f_1^{(k)}(Y_{m1}^{(k)}), \dots, f_p^{(k)}(Y_{mp}^{(k)}))$. The edge (i, j) is a differential edge if $\rho_{ij\cdot}^{(k)} \neq \rho_{ij\cdot}^{(k')}$ for some $k, k' \in \{1, \dots, K\}$, and the differential network is defined as the set of all differential edges. As a well-known result in statistics, $\rho_{ij\cdot}^{(k)} = -\omega_{ij}^{(k)} / \sqrt{\omega_{ii}^{(k)} \omega_{jj}^{(k)}}$. Here, we consider an equivalent alternative of the hypothesis testing above. Similar as in [4], let

$$S_{ij}(\boldsymbol{\Omega}) = \sqrt{\sum_{1 \leq k < k' \leq K} (\rho_{ij\cdot}^{(k)} - \rho_{ij\cdot}^{(k')})^2}, \tag{1}$$

then the hypothesis testing can be simplified as

$$\begin{aligned} H_{0ij} &: S_{ij}(\boldsymbol{\Omega}) = 0, \\ H_{aij} &: S_{ij}(\boldsymbol{\Omega}) > 0. \end{aligned}$$

As $S_{ij}(\boldsymbol{\Omega}) = S_{ji}(\boldsymbol{\Omega})$, we define $\mathcal{H}_0 = \{H_{0ij}, 1 \leq i < j \leq p\}$ and $\mathcal{H}_a = \{H_{aij}, 1 \leq i < j \leq p\}$, and the total numbers of tests are $p(p-1)/2$, i.e., $\text{card}(\mathcal{H}_0) = \text{card}(\mathcal{H}_a) = p(p-1)/2$.

The rest of this paper is structured as follows: In Section 2, we introduce the new test statistic and multiple testing procedure. In Section 3 we perform a simulation study to evaluate the finite sample performance of the proposed test in terms of FDR control and statistical power. We then apply the new method to a rich genomic data to study the genetic difference between four breast cancer subtypes. We discuss the strength and shortcomings of the test in Section 5. Technical proof of the asymptotic results is provided in Appendix A.

2. Statistical Methods

2.1. Winsorized Estimator of the Latent Gaussian Variables

In practice, the transformation functions $\mathbf{f}^{(k)} = (f_1^{(k)}, \dots, f_p^{(k)})$ in the nonparanormal distribution are unknown. However, one can use a Winsorized estimator to approximate $\mathbf{f}^{(k)}$, i.e., to impute the latent Gaussian variables (oracle data) $(X_{m1}^{(k)}, \dots, X_{mp}^{(k)})_{1 \leq m \leq n_k}$. To illustrate the Winsorized estimator, we define the following quantile function:

$$\hat{h}_j^{(k)}(t) = \Phi^{-1}(\tilde{F}_j^{(k)}(t)), \quad 1 \leq j \leq p,$$

where $\tilde{F}_j^{(k)}$ is some estimator of the cumulative distribution function of $Y_j^{(k)}$, and a natural choice for $\tilde{F}_j^{(k)}$ would be the empirical cumulative distribution function (eCDF)

$$\hat{F}_j^{(k)}(t) = \frac{1}{n_k} \sum_{m=1}^{n_k} I\{Y_{mj}^{(k)} \leq t\}.$$

One major drawback of the eCDF above is that under high dimensionality, the variance of $\hat{F}_j^{(k)}(t)$ could be too large. To overcome the problem, Liu et al. (2009) considered a truncated (Winsorized) estimator as follows:

$$\tilde{F}_j^{(k)} = \begin{cases} \delta_n & \hat{F}_j^{(k)}(t) < \delta_n \\ \hat{F}_j^{(k)}(t) & \delta_n \leq \hat{F}_j^{(k)}(t) \leq 1 - \delta_n, \\ 1 - \delta_n & \hat{F}_j^{(k)}(t) > 1 - \delta_n \end{cases}$$

where δ_n serves as the truncation parameter that should be carefully chosen. Liu et al. (2009) [5] suggested $\delta_n = 1/(4n^{1/4}\sqrt{\pi \log n})$ to balance the bias and variance of eCDF, and so we will use this value in our calculations. To estimate the transformation functions and impute the latent Gaussian variable \mathbf{X} , we define

$$X_{mj}^{(k)*} = \tilde{f}_j^{(k)}(Y_{mj}^{(k)}) = \hat{\mu}_j^{(k)} + \hat{\sigma}_j^{(k)} \tilde{h}_j^{(k)}(Y_{mj}^{(k)}),$$

where $\tilde{h}_j^{(k)}(t)$, $\hat{\mu}_j^{(k)}$ and $\hat{\sigma}_j^{(k)}$ are given below:

$$\begin{aligned} \tilde{h}_j^{(k)}(t) &= \Phi^{-1}(\tilde{F}_j^{(k)}(t)), \\ \hat{\mu}_j^{(k)} &= \frac{1}{n_k} \sum_{m=1}^{n_k} Y_{mj}^{(k)}, \\ \hat{\sigma}_j^{(k)} &= \sqrt{\frac{1}{n_k} \sum_{m=1}^{n_k} (Y_{mj}^{(k)} - \hat{\mu}_j^{(k)})^2}. \end{aligned}$$

The Winsorized estimator $X_{mj}^{(k)*}$ generally works well in approximating the unknown $X_{mj}^{(k)}$, and it could be used to estimate the oracle sample covariance. Let $\hat{\Sigma}^{(k)}$ be the sample covariance matrix by the oracle data, and $\tilde{\Sigma}^{(k)}$ be the sample covariance matrix by $(X_1^{(k)*}, \dots, X_p^{(k)*})$, that is

$$\tilde{\Sigma}^{(k)} = \frac{1}{n_k} \sum_{m=1}^{n_k} (X_m^{(k)*} - \tilde{\boldsymbol{\mu}}^{(k)})(X_m^{(k)*} - \tilde{\boldsymbol{\mu}}^{(k)})^T,$$

where $\tilde{\boldsymbol{\mu}}^{(k)} = (1/n_k) \sum_{m=1}^{n_k} X_m^{(k)*}$. Liu et al. (2009) established the following consistency results under mild regularity conditions:

$$\|\tilde{\Sigma}^{(k)} - \hat{\Sigma}^{(k)}\|_{\infty} = O_p\left(\sqrt{\frac{\log p \log^2 n_k}{n_k^{1/2}}}\right).$$

When estimating the precision matrix $\boldsymbol{\Omega}^{(k)}$, one can consider a modified graphical lasso based on imputed data, i.e.,

$$\tilde{\boldsymbol{\Omega}}_{glasso}^{(k)} = \arg \min_{\boldsymbol{\Omega}} \left\{ \text{tr}(\boldsymbol{\Omega} \tilde{\Sigma}^{(k)}) - \log |\boldsymbol{\Omega}| + \lambda \|\boldsymbol{\Omega}\|_1 \right\}. \tag{2}$$

Liu et al. (2009) showed the following convergence, which elucidated the asymptotic equivalence between the oracle data and imputed data in the structural estimation of NPNGM

$$\|\tilde{\boldsymbol{\Omega}}_{glasso}^{(k)} - \boldsymbol{\Omega}^{(k)}\|_F = O_p\left(\sqrt{\frac{(\|\boldsymbol{\Omega}^{(k)}\|_0 + p) \log p \log^2 n_k}{n_k^{1/2}}}\right).$$

2.2. Asymptotic Results for a Single Class

To extend Liu’s test to a nonparanormal case, we first consider the problem of single GGM estimation based on oracle data, i.e., $(X_{m1}^{(k)}, \dots, X_{mp}^{(k)})_{1 \leq m \leq n_k} \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, in the following regression framework

$$X_{mj}^{(k)} = \alpha_j^{(k)} + \mathbf{X}_{m,-j}^{(k)'} \boldsymbol{\beta}_j^{(k)} + \epsilon_{mj}^{(k)}. \tag{3}$$

It is not hard to show that the regression coefficients $\boldsymbol{\beta}_j^{(k)} = (\beta_{j,1}^{(k)}, \dots, \beta_{j,j-1}^{(k)}, \beta_{j,j+1}^{(k)}, \beta_{j,p}^{(k)})$ and the error term $\epsilon_{mj}^{(k)}$ satisfy

$$\boldsymbol{\beta}_j^{(k)} = -(\omega_{jj}^{(k)})^{-1} \boldsymbol{\Omega}_{-j,j}^{(k)}, \text{ cov}(\epsilon_{mi}^{(k)}, \epsilon_{mj}^{(k)}) = \frac{\omega_{ij}^{(k)}}{\omega_{ii}^{(k)} \omega_{jj}^{(k)}}.$$

As the oracle data $(X_{m1}^{(k)}, \dots, X_{mp}^{(k)})_{1 \leq m \leq n_k}$ in Equation (3) are generally unknown, we consider a new regression model based on Winsorized imputations:

$$X_{mj}^{(k)*} = \hat{\alpha}_j^{(k)} + \mathbf{X}_{m,-j}^{(k)*'} \hat{\boldsymbol{\beta}}_j^{(k)} + \epsilon_{mj}^{(k)*}. \tag{4}$$

In solving the problem of single GGM estimation, Liu (2017) proposed an elegant test based on a bias-corrected sample covariance. This has motivated us to construct the following new statistic

$$S_{ij}^{(k)*} = \sqrt{\frac{1}{n_k r_{ii}^{(k)*} r_{jj}^{(k)*}} \left(\sum_{m=1}^{n_k} \epsilon_{mi}^{(k)*} \epsilon_{mj}^{(k)*} + \sum_{m=1}^{n_k} \{\epsilon_{mi}^{(k)*}\}^2 \hat{\beta}_{i,j}^{(k)} + \sum_{m=1}^{n_k} \{\epsilon_{mj}^{(k)*}\}^2 \hat{\beta}_{j,i}^{(k)} \right)}, \tag{5}$$

where $r_{ij}^{(k)*} = (1/n_k) \sum_{m=1}^{n_k} \epsilon_{mi}^{(k)*} \epsilon_{mj}^{(k)*}$. By letting $\bar{\boldsymbol{\epsilon}}^{(k)} = (1/n_k) \sum_{m=1}^{n_k} \boldsymbol{\epsilon}_m^{(k)}$, $(\hat{\sigma}_{ij,\epsilon}^{(k)})_{1 \leq i,j \leq p} = (1/n_k) \sum_{m=1}^{n_k} (\boldsymbol{\epsilon}_m^{(k)} - \bar{\boldsymbol{\epsilon}}^{(k)})(\boldsymbol{\epsilon}_m^{(k)} - \bar{\boldsymbol{\epsilon}}^{(k)})'$, $b_{ij}^{(k)} = \omega_{ii}^{(k)} \hat{\sigma}_{ii,\epsilon}^{(k)} + \omega_{jj}^{(k)} \hat{\sigma}_{jj,\epsilon}^{(k)} - 1$, we will prove that, under mild conditions (see a detailed proof in Appendix A)

$$S_{ij}^{(k)*} + b_{ij}^{(k)} \frac{\omega_{ij}^{(k)}}{\omega_{ii}^{(k)} \omega_{jj}^{(k)}} \xrightarrow{D} N\left(0, 1 + \frac{\{\omega_{ij}^{(k)}\}^2}{\omega_{ii}^{(k)} \omega_{jj}^{(k)}}\right). \tag{6}$$

Similar as in [4], the estimated coefficients $\hat{\boldsymbol{\beta}}_j^{(k)}$ must satisfy the following conditions:

$$\|\hat{\boldsymbol{\beta}}_j^{(k)} - \boldsymbol{\beta}_j^{(k)}\|_{\ell_1} = O_p(a_n^{(k)}),$$

$$\min \left\{ \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}^{(k)}) \|\hat{\boldsymbol{\beta}}_j^{(k)} - \boldsymbol{\beta}_j^{(k)}\|_{\ell_2}, \max_{1 \leq j \leq p} \sqrt{(\hat{\boldsymbol{\beta}}_j^{(k)} - \boldsymbol{\beta}_j^{(k)})^T \hat{\boldsymbol{\Sigma}}_{-j,-j}^{(k)} (\hat{\boldsymbol{\beta}}_j^{(k)} - \boldsymbol{\beta}_j^{(k)})} \right\} = O_p(b_n^{(k)}),$$

where

$$a_n^{(k)} = o(\sqrt{\log p / n_k}), \text{ and } b_n^{(k)} = o(n_k^{-1/4}). \tag{7}$$

Equation (6) is our main result, which is essentially a counterpart of Proposition 3.1 in [4]. The detailed proof is given in Appendix A. The asymptotic result we obtained here suggested that, by an appropriate choice of regression coefficients $\hat{\boldsymbol{\beta}}_j^{(k)}$, Liu’s test can be readily extended to a nonparanormal framework by Winsorized imputation. Under GGMs, the condition (7) can be satisfied by several popular shrinkage estimators including lasso estimator and Dantzig selector. For the choice of $\boldsymbol{\beta}_j^{(k)}$ under NPNGMs, one can use the rank-based method introduced by Xue and Zou (2012) [6]. Xue and Zou (2012) showed that the rank-based estimator (e.g., rank-based lasso and rank-based Dantzig selector) achieved exactly the same convergence rate as its oracle counterpart, therefore, it also satisfies our condition (7).

2.3. Multiple Testing Procedure for FDR Control

Now we introduce the multiple testing procedure for FDR control based on the single-class result from Equation (6). As suggested in [4], the partial correlation coefficient can be well estimated by a thresholding estimator

$$\hat{\rho}_{ij}^{(k)} = S_{ij}^{(k)} I\{|S_{ij}^{(k)}| \geq 2\sqrt{\frac{\log p}{n_k}}\},$$

and we define the following two-sample test statistics

$$S_{ij}^{(k,k')} = \frac{S_{ij}^{(k)} - S_{ij}^{(k')}}{\sqrt{\frac{1}{n_k}(1 - \{\hat{\rho}_{ij}^{(k)}\}^2)^2 + \frac{1}{n_{k'}}(1 - \{\hat{\rho}_{ij}^{(k')}\}^2)^2}}.$$

In the multi-sample case $\mathbf{S}_{ij} = (S_{ij}^{(k,k')})_{1 \leq k < k' \leq K}$, we consider a sum squared test statistics

$$\mathbb{S}_{ij} = \sqrt{\sum_{k < k'} \{S_{ij}^{(k,k')}\}^2}.$$

Motivated by [4] (Equations 2.6 and 2.7) and [7], we define the following statistic

$$T_{ij} = \Phi^{-1}P\left(\sqrt{\sum_{i=1}^M \lambda_i Z_i^2} \leq \mathbb{S}_{ij}\right),$$

and constant $A = (P_0 - \hat{P}_0)/Q_0$, where $Z_i, i = 1, \dots, M$ represent a sequence of M i.i.d. standard normal random variables, $P_0 = 2\Phi(1) - 1$, $\hat{P}_0 = 2 \sum_{1 \leq i < j \leq p} I\{|T_{ij}| \leq 1\}/(p^2 - p)$, $Q_0 = \sqrt{2}\phi(1)$ and $A(t) = (1 + |A| \frac{|t|\phi(t)}{\sqrt{2}(1-\Phi(t))})^{-1}$. For a given $0 < \alpha_0 < 1$, let

$$t(\alpha_0) = \inf \left\{ t \in \mathbb{R}, 1 - \phi(t) \leq \frac{\alpha_0 A(t) \max\{1, \sum_{1 \leq i < j \leq p} I\{T_{ij} \geq t\}\}}{(p^2 - p)/2} \right\}.$$

the FDR can be controlled at level α , if we reject $H_{0ij} : S_{ij}(\mathbf{\Omega}) = 0$ when $T_{ij} \geq t(\alpha_0)$. One may refer to [7] for the detailed proof about this testing procedure.

Our proposed computational pipeline consisted of three steps: (1) Winsorized imputation for the latent Gaussian variables; (2) rank-based estimation of regression coefficients, and (3) multiple testing with FDR control. On the whole, we put forward a simple procedure to estimate the structural difference between multiple nonparanormal graphical models. The computational pipeline for a two-sample comparison has been implemented in the R package *DNetFinder*, which can be downloaded from the Comprehensive R Archive Network (CRAN).

3. Numerical Study

We performed a simulation study to evaluate the finite sample performance of the proposed procedure. In particular, we evaluated the empirical false discovery rate (eFDR) as well as the statistical power under two classes, i.e., $K = 2$. The dimension and sample size were set to be $p = 200$ and $n_1 = n_2 = 100$. We consider two commonly used graph-generating models including the band graph and Erdős–Rényi (ER) graph, and two estimators for regression coefficients including lasso estimator and Dantzig selector. Detailed set-up for precision matrices $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are given below:

- **Band graph:** $\Omega_1 = (\omega_{ij})_{1 \leq i, j \leq p}$ was obtained by the following assignments

$$\omega_{ij} = \begin{cases} 1 & |i - j| = 0 \\ 0.6 & |i - j| = 1 \\ 0 & |i - j| \geq 2 \end{cases}$$

We then randomly picked 50 edges in Ω_1 as the differential edges and changed their signs in Ω_2 . To ensure positive definiteness, we added $\max(|\lambda_{\min}(\Omega_1)|, |\lambda_{\min}(\Omega_2)|) + 0.05$, to the diagonal of Ω_1 and Ω_2 .

- **Erdős–Rényi (ER) graph:** Each node pair (i, j) were randomly connected with probability 5%. A correlation coefficient is generated for each edge in the network from a two-part uniform distribution $[-1/2, -1/4] \cup [1/4, 1/2]$. To ensure positive-definiteness, we shrunk the correlations by a factor of 5 and the diagonals were set to be one for Ω_1 . We then randomly selected 5% of the edges as the differential edges, and changed their signs in Ω_2 .

For each graph, we generated the latent Gaussian data (oracle data) from $N(\mathbf{0}, \Omega^{-1})$, $\Omega \in \{\Omega_1, \Omega_2\}$, and a Winsorized estimator with truncation parameter $\delta_n = 1/(4n^{1/4} \sqrt{\pi \log n})$ was used to implement our test. The performance of the proposed method was then evaluated in two aspects: false discovery rate control and statistical power. In particular, we compared the results based on oracle data and imputed data by the Winsorized estimator. Two estimators including the lasso estimator and Dantzig selector were used to estimate coefficients $\hat{\beta}$. For oracle data, we applied the R package *flare* to calculate the solution path over a sequence of 20 candidate λ 's and tune by Akaike information criterion (AIC). For imputed data, we adopted the rank-based methods introduced by [6], i.e., the rank-based lasso and rank-based Dantzig selector. The simulation was repeated for 100 times for each FDR level ($\alpha \in \{0.05, 0.10, 0.15, \dots, 0.50\}$) and the average empirical FDR and statistical power were summarized.

Figures 1 and 2 compared the empirical false discovery rate (eFDR) with the desired level α under the band graph and ER graph. It can be seen that the empirical FDR based on imputed data is close to the one by oracle data, both close to the desired level of α , suggesting that the FDRs were controlled quite well for both cases. The lasso estimator works almost equally well as Dantzig selector in both settings.

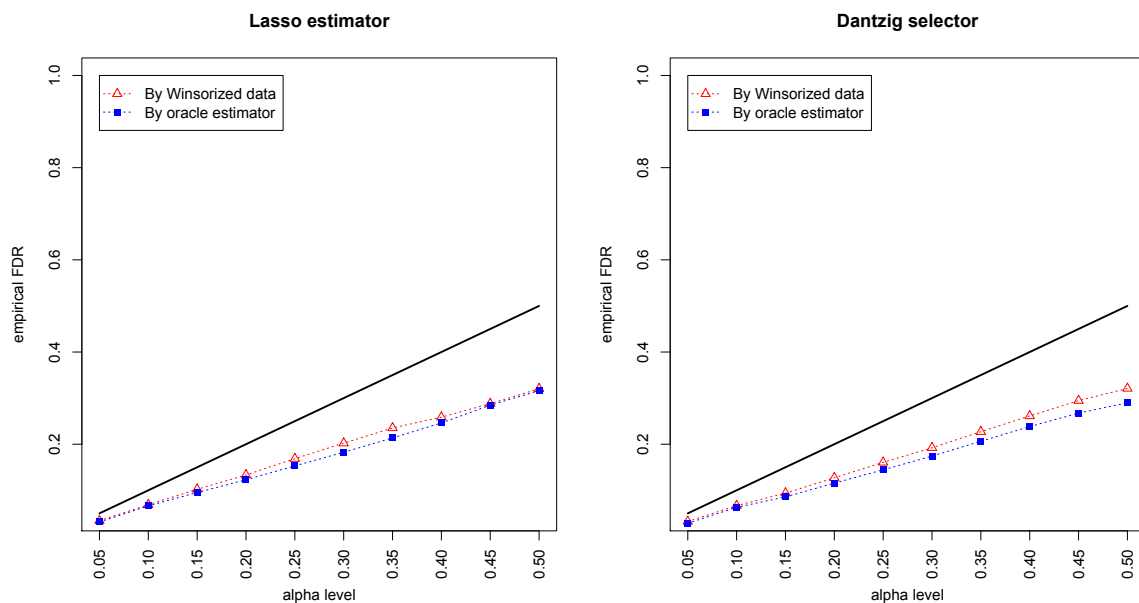


Figure 1. Empirical false discovery rates (eFDRs) by oracle data and Winsorized imputations under the band graph setting. The x-axis represents the desired FDR levels from 0.05 to 0.5, and the solid line is $y = x$.

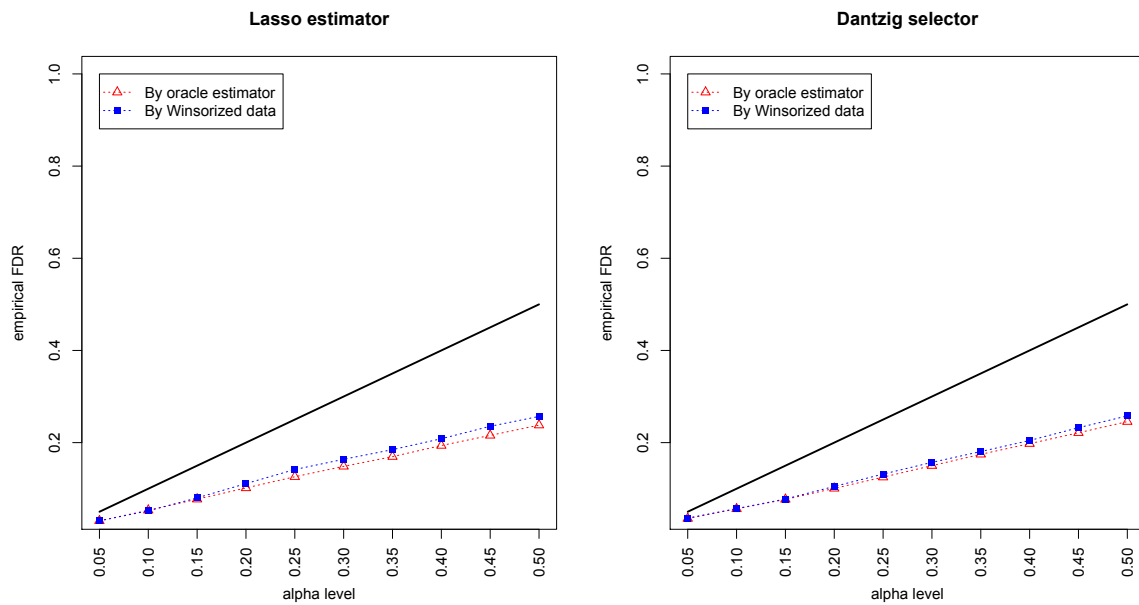


Figure 2. Empirical FDRs (eFDRs) by oracle data and Winsorized imputations under the Erdős–Rényi (ER) graph setting. The x -axis represents the desired FDR levels from 0.05 to 0.5, and the solid line is $y = x$.

Figures 3 and 4 summarized the statistical power of the test for the band graph and ER graph. As can be seen, the power for ER graph is substantially lower than the band graph, indicating that the complexity and denseness of the underlying differential network may significantly decrease the power of our test. The test based on oracle data performs slightly better than the imputed data, which is due to the loss of information during Winsorized imputation. Similar as we observed from Figures 1 and 2, the lasso estimator works almost equally well as Dantzig selector.

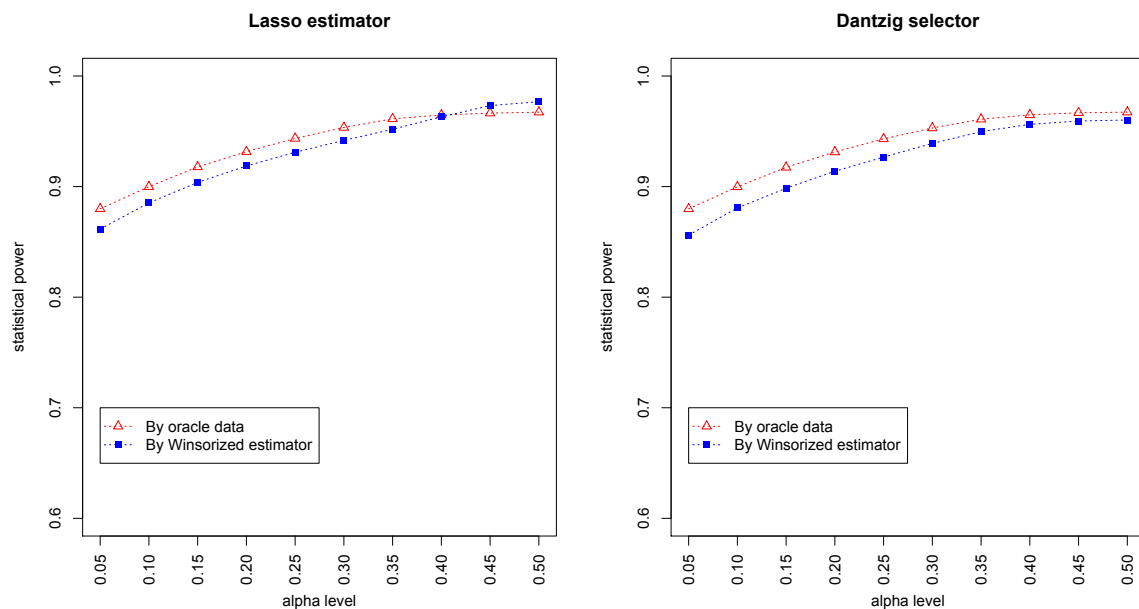


Figure 3. Statistical powers by oracle data and Winsorized imputations under the band graph setting. The x -axis represents the desired FDR levels from 0.05 to 0.5.

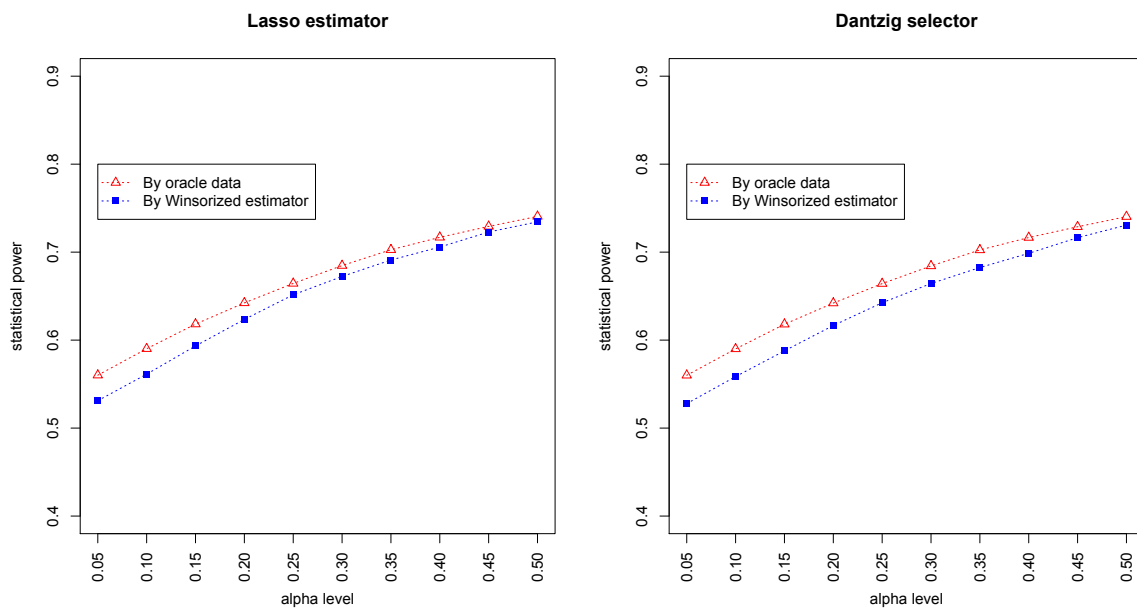


Figure 4. Statistical powers by oracle data and Winsorized estimator under the ER graph setting. The *x*-axis represents the desired FDR levels from 0.05 to 0.5.

In addition, we compared the proposed test with a direct estimator, recently developed by Zhang (2019) [8]. The direct estimator is a rank-based estimator and can be solved by a parametric simplex algorithm. We simulated the data from the Erdős–Rényi (ER) graph with different sample sizes ($n = 25, 50, 100, 150$) and numbers of dimensions ($p = 40, 60, 90, 120$). As the direct estimator does not control the false discovery rate, we set the FDR level at 0.05 for our proposed test. Figure 5 summarized the empirical FDR and statistical power under different sample sizes (with dimension fixed at 100) and different dimensions (with sample size fixed at 100). It can be seen that the two methods have comparable performance and our proposed test achieves lower FDR but slightly lower statistical power. However, it is noteworthy that the direct estimator is computationally expensive and becomes impractical when the dimensions exceed 150. Table 1 summarized the running time of the two methods, where it can be seen that our test is much faster than the direct estimator, especially for relatively high dimensions. For instance, when $p = 120$, the direct estimator takes hours while our test takes less than 10 seconds. As the core part of the proposed algorithm is the estimation of regression coefficients, the time complexity is the same as the linear regression. For instance, with LASSO and $p > n$, the time complexity is $O(np^2)$, while the direct estimator by Zhang (2019) has a time complexity $O(np^4)$.

Table 1. Running time of the proposed test and direct estimator (in seconds).

<i>n/p</i>	40	60	90	120
25	0.88 (7.0)	1.59 (110)	3.79 (1936)	6.49 (23,066)
50	1.19 (7.7)	1.93 (127)	4.15 (1973)	6.83 (23,119)
100	1.87 (9.1)	2.61 (146)	5.00 (2016)	7.80 (23,153)
200	2.11 (11)	3.04 (165)	6.22 (2055)	9.61 (23,201)

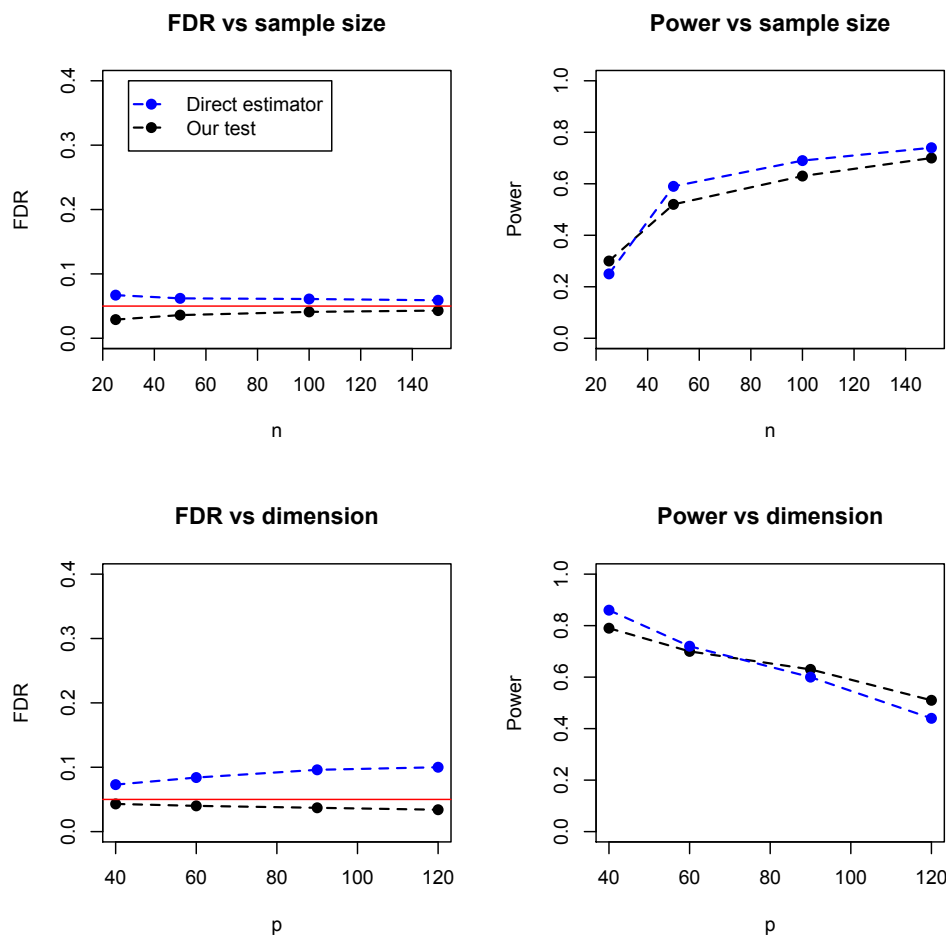


Figure 5. Comparison of the proposed test and direct estimator by Zhang (2019), in terms of empirical FDR and statistical power under different sample sizes and dimensions.

4. A Genomic Application

In this part, we applied the proposed test to the Cancer Genome Atlas data (TCGA, [9]) to study the different roles of the cell cycle pathway in the two subtypes of breast cancer including luminal A subtype and basal-like subtype. The cell cycle pathway is known to play a critical role in the initiation and progression of many human cancers including breast cancer and ovarian cancer [10,11]. For instance, the cell cycle pathway provided by KEGG (Kyoto Encyclopedia of Genes and Genomes, [12]) contains 128 important genes that co-regulate cell proliferation, including *ATM*, *RB1*, *CCNE1*, and *MYC*. Abnormal regulation among these genes may cause the over-proliferation of cells and an accumulation of tumor cell numbers [11].

The transcriptome profiling data for breast cancer were downloaded through the Genomic Data Commons portal [13] in January 2017. The expression level of each gene was quantified by the count of reads mapped to the gene. The quantifications were done by software *HTSeq* of version 0.9.1 [14]. In our analysis, we excluded 43 subjects including 12 male subjects and 31 subjects with >1% missing values. In addition, we removed the effects due to different age groups and batches using a median-matching and variance-matching strategy [10,15,16]. For example, the batch effect can be removed in the following way:

$$g_{ijk}^* = M_i + (g_{ijk} - M_{ij}) \frac{\hat{\sigma}_{g_i}}{\hat{\sigma}_{g_{ij}}},$$

where g_{ijk} refers to the expression value for gene i from sample k in batch j ($j = 1, 2, \dots, J; k = 1, 2, \dots, n_j$), M_{ij} represents the median of $g_{ij} = (g_{ij1}, \dots, g_{ijn_j})$, M_i refers to the median of $g_i = (g_{i1}, \dots, g_{in_i})$, $\hat{\sigma}_{g_i}$ and $\hat{\sigma}_{g_{ij}}$ stand for the standard deviations of g_i and g_{ij} , respectively.

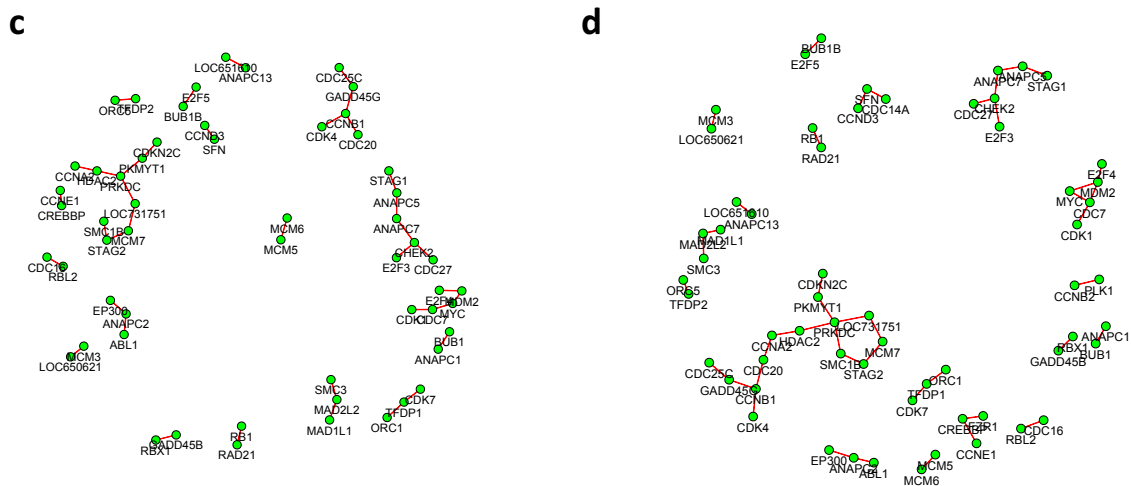


Figure 6. The inferred differential networks between the LumA and Basal-like subtypes under different desired FDR levels: (a) 0.05; (b) 0.10; (c) 0.15; (d) 0.20, with all isolated genes being removed. Each connection in the network represents an identified differential edge.

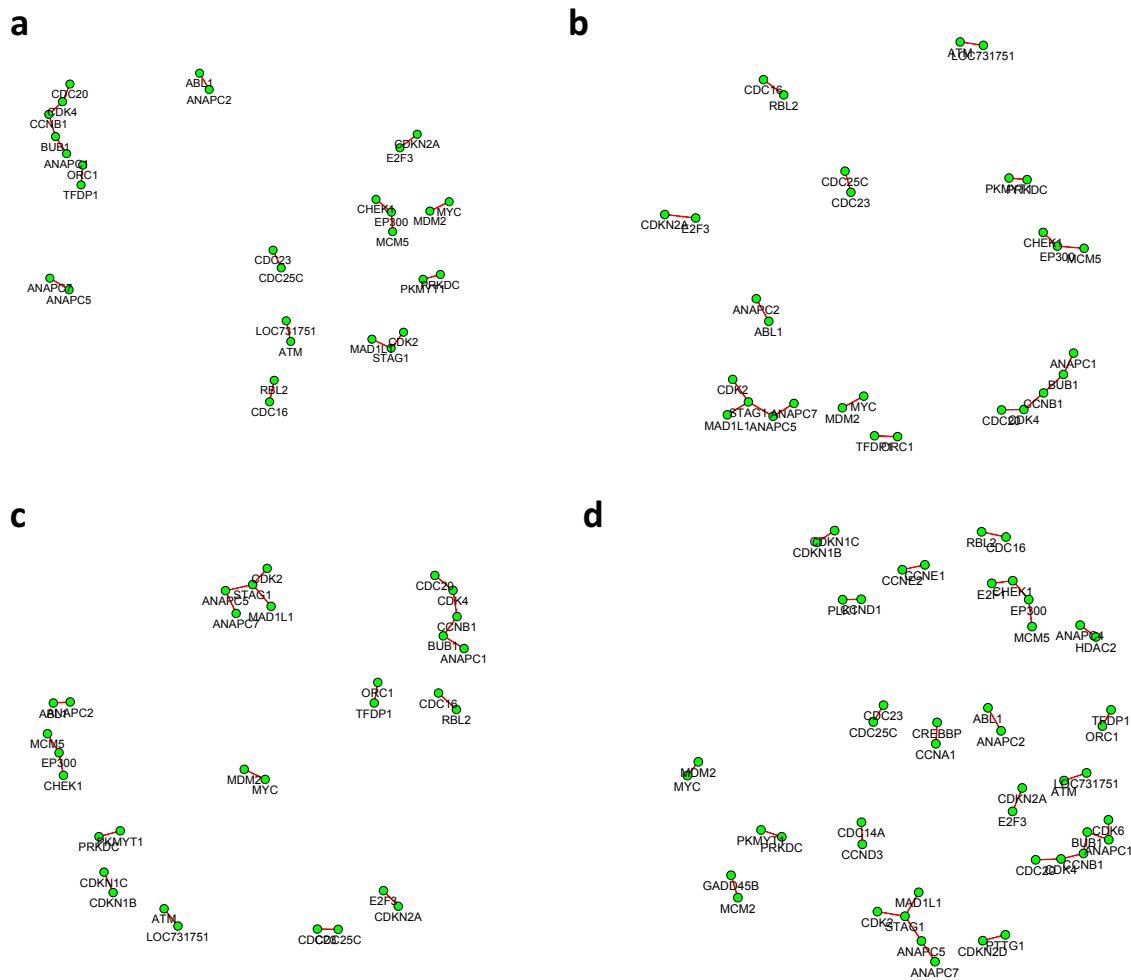


Figure 7. The inferred differential networks between the LumB and Basal-like subtypes under different desired FDR levels: (a) 0.05; (b) 0.10; (c) 0.15; (d) 0.20, with all isolated genes being removed. Each connection in the network represents an identified differential edge.

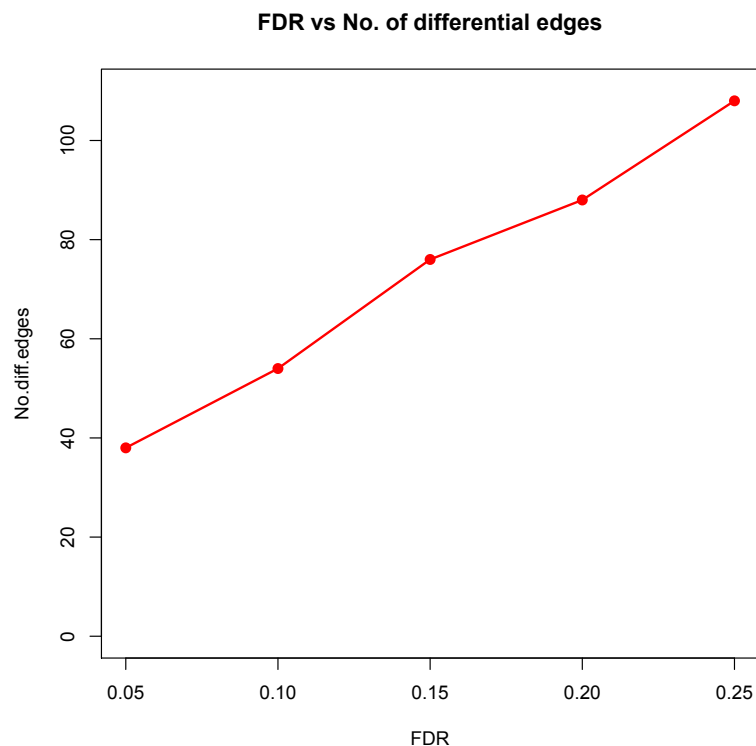


Figure 8. Desired FDR level against the number of differential edges.

5. Discussion

Detecting the differential substructure on multiple graphical models is a fundamental and challenging problem in statistics. Liu (2017) studied the problem under the Gaussian framework and introduced an elegant hierarchical test based on the estimation of single GGM. Unlike most existing methods, Liu’s approach asymptotically controlled the false discovery rate at a nominal level, which guarantees the quality of the estimated differential network. In this work, we further extended Liu’s test to a more flexible semiparametric framework, namely the nonparanormal graphical models. Our test is built upon a Winsorized estimator of the unknown transformation functions and it enjoys similar asymptotic properties as its oracle counterpart does.

Although the new test holds great promise in many applications such as genetic network modeling, it has some practical limitations. First, as we see from the theoretical derivation, the good performance of the test relied on the sparsity assumption on the differential network. Although the sparsity assumption is reasonable in many cases, it still could be violated in some applications. For instance, some genetic pathways may exhibit a global change of gene–gene regulations between different phenotypes. When the differential network is dense or locally dense, the method may fail to control the FDR. To solve the problem, a new test needs to be defined to evaluate the level of the sparseness of the change between two conditions. However, there is still a gap on the literature of this topic.

Second, one key assumption in NPNGMs is that the transformed variables follow a joint Gaussian distribution. This assumption also needs to be checked in real-world applications. Under low dimensions, one can employ some popular normality tests, including the Anderson–Darling test and Shapiro–Wilk test, on the imputed data or other normal scores. However, most of these tests fail to detect non-normality for high-dimension data. The normality test under high dimension is still an open and challenging problem and we left it for future research.

It is also noteworthy to mention that the new test relied on an accurate estimator for the coefficients β . Motivated by [6], we chose two popular estimators including lasso estimator and Dantzig selector based on the adjusted Spearman’s rank, which satisfies Condition (7). In fact, some other estimators

also satisfy the conditions, for instance, the rank-based adaptive lasso [6,25] and square-root lasso estimator [6,26]. These estimators can also be incorporated into our testing framework.

6. Conclusions

We have introduced a novel statistical test to detect the structural difference between the two nonparanormal graphical models. The proposed test dropped the Gaussian assumption and can be potentially applied to many non-Gaussian data for differential network analysis. For instance, some digital gene expression data (e.g., RNA-seq data) do not follow Gaussian distribution even after log transformation or other variance-stabilizing transformations. In such cases, one can model the data with a nonparanormal graphical model and apply our test to find differential edges between two or multiple phenotypic conditions. The proposed test may also be used to detect the difference between normal and disease populations in the brain connectivity network.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FDR	False discovery rate
NPNGM	Nonparanormal graphical model
GGM	Gaussian graphical model
TCGA	The Cancer Genome Atlas

Appendix A. Proof of Equation (6)

Define the estimated residuals based on Winsorized estimator as:

$$\epsilon_{mj}^* = X_{mj}^* - \bar{X}_j^* - (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*)\hat{\boldsymbol{\beta}}_j,$$

where $\bar{X}_j^* = 1/n \sum_{m=1}^n X_{mj}^*$ and $\bar{\mathbf{X}}_{-j}^* = 1/n \sum_{m=1}^n \mathbf{X}_{m,-j}^*$. The choice of $\hat{\boldsymbol{\beta}}_j$ must satisfy the following two conditions:

$$\begin{aligned} & \|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_{\ell_1} = O_p(a_n), \\ & \min \left\{ \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}) \|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_{\ell_2}, \max_{1 \leq j \leq p} \sqrt{(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)^T \hat{\boldsymbol{\Sigma}}_{-j,-j} (\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)} \right\} = O_p(b_n), \end{aligned}$$

$$\text{where } a_n^{(k)} = o(\sqrt{\log p/n_k}), \text{ and } b_n^{(k)} = o(n_k^{-1/4}).$$

It is noteworthy to mention that the conditions above are slightly different from the conditions in [4] due to the different convergence rates by oracle data and imputed data. The conditions above can be satisfied by the rank-based estimators introduced in [6], e.g., rank-based lasso estimator or rank-based Dantzig selector. By letting $\tilde{\epsilon}_{mj} = \epsilon_{mj} - \bar{\epsilon}_i$, we have:

$$\epsilon_{mi}^* \epsilon_{mj}^* = \tilde{\epsilon}_{mi} \tilde{\epsilon}_{mj} - \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \hat{\boldsymbol{\beta}}_j - (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) \boldsymbol{\beta}_j \} \quad (\text{A1})$$

$$- \tilde{\epsilon}_{mj} \{ (\mathbf{X}_{m,-i}^* - \bar{\mathbf{X}}_{-i}^*) \hat{\boldsymbol{\beta}}_i - (\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i}) \boldsymbol{\beta}_i \} \quad (\text{A2})$$

$$+ \{ \hat{\boldsymbol{\beta}}_i^T (\mathbf{X}_{m,-i}^* - \bar{\mathbf{X}}_{-i}^*)^T (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_i^T (\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i})^T (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) \boldsymbol{\beta}_j \}. \quad (\text{A3})$$

First, for term (A3), we have:

$$\begin{aligned} & \left| \frac{1}{n} \sum_{m=1}^n \{ \hat{\beta}_i^T (\mathbf{X}_{m,-i}^* - \bar{\mathbf{X}}_{-i}^*)^T (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \hat{\beta}_j - \beta_i^T (\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i})^T (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) \beta_j \} \right| \\ &= | \hat{\beta}_i^T (\tilde{\Sigma}_{-i,-j} - \hat{\Sigma}_{-i,-j}) \hat{\beta}_j + (\hat{\beta}_i - \beta_i)^T (\tilde{\Sigma}_{-i,-j} - \Sigma_{-i,-j}) (\hat{\beta}_j - \beta_j) + (\hat{\beta}_i - \beta_i)^T \Sigma_{-i,-j} (\hat{\beta}_j - \beta_j) | \\ &\leq \max_{i,j} | \hat{\beta}_i^T (\tilde{\Sigma}_{-i,-j} - \hat{\Sigma}_{-i,-j}) \hat{\beta}_j | + \max_{i,j} | (\hat{\beta}_i - \beta_i)^T (\tilde{\Sigma}_{-i,-j} - \Sigma_{-i,-j}) (\hat{\beta}_j - \beta_j) | + \max_{i,j} | (\hat{\beta}_i - \beta_i)^T \Sigma_{-i,-j} (\hat{\beta}_j - \beta_j) |, \end{aligned}$$

where the last term can be bounded as follows:

$$\max_{i,j} | (\hat{\beta}_i - \beta_i)^T \Sigma_{-i,-j} (\hat{\beta}_j - \beta_j) | = O_p(\lambda_{\max}(\Sigma) \max_{1 \leq i \leq p} \|\hat{\beta}_i - \beta_i\|_{\ell_2}^2) = O_p(b_n^2).$$

It is not hard to show that:

$$\|\tilde{\Sigma} - \Sigma\|_{\infty} = O_p\left(\sqrt{\frac{\log p}{n}}\right),$$

therefore, the second term can also be bounded

$$\max_{i,j} | (\hat{\beta}_i - \beta_i)^T (\tilde{\Sigma}_{-i,-j} - \Sigma_{-i,-j}) (\hat{\beta}_j - \beta_j) | = O_p\left(a_n^2 \sqrt{\frac{\log p}{n}}\right).$$

Under some mild regularity conditions (stated in [6]), we have

$$\|\tilde{\Sigma} - \hat{\Sigma}\|_{\infty} = O_p\left(\sqrt{\frac{\log p \log^2 n}{n^{1/2}}}\right),$$

thus under the condition that $\max_{i,j} |\beta_{i,j}| \leq C_1$ and $\lambda_{\min}(\Sigma) = o((\log p/n)^{\frac{3}{4}})$, the first term can be bounded as follows:

$$\begin{aligned} & \max_{i,j} | \hat{\beta}_i^T (\tilde{\Sigma}_{-i,-j} - \hat{\Sigma}_{-i,-j}) \hat{\beta}_j | \\ &\leq \max_{i,j} | \beta_i^T (\tilde{\Sigma}_{-i,-j} - \hat{\Sigma}_{-i,-j}) \beta_j | + \max_{i,j} | (\hat{\beta}_i - \beta_i)^T (\tilde{\Sigma}_{-i,-j} - \hat{\Sigma}_{-i,-j}) (\hat{\beta}_j - \beta_j) | \\ &= O_p\left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}}\right). \end{aligned}$$

Combining the three terms above, we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{m=1}^n \{ \hat{\beta}_i^T (\mathbf{X}_{m,-i}^* - \bar{\mathbf{X}}_{-i}^*)^T (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \hat{\beta}_j - \beta_i^T (\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i})^T (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) \beta_j \} \right| \\ &= O_p\left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}} + b_n^2\right). \end{aligned}$$

Next, we bound term (A1), which can be rewritten as:

$$\tilde{\epsilon}_{mi} (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) (\hat{\beta}_j - \beta_j) + \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) - (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \} \hat{\beta}_j,$$

where the first term can be further decomposed into two parts,

$$\tilde{\epsilon}_{mi} (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) (\hat{\beta}_j - \beta_j) = \tilde{\epsilon}_{mi} (X_{mi} - \bar{X}_i) (\hat{\beta}_{i,j} - \beta_{i,j}) I\{i \neq j\} + \sum_{l \neq i,j} \tilde{\epsilon}_{mi} (X_{ml} - \bar{X}_l) (\hat{\beta}_{l,j} - \beta_{l,j}).$$

To bound $\sum_{l \neq i, j} \tilde{\epsilon}_{mi}(X_{ml} - \bar{X}_l)(\hat{\beta}_{l,j} - \beta_{l,j})$, we use the independence between ϵ_{mi} and $\mathbf{X}_{m,-i}$. It is easy to show that

$$\max_{l \neq i} \left| \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(X_{ml} - \bar{X}_l) \right| = O_p \left(\sqrt{\frac{\log p}{n}} \right),$$

which indicates that

$$\max_{i,j} \left| \frac{1}{n} \sum_{m=1}^n \left(\sum_{l \neq i,j} \tilde{\epsilon}_{mi}(X_{ml} - \bar{X}_l)(\hat{\beta}_{l,j} - \beta_{l,j}) \right) \right| = O_p \left(a_n \sqrt{\frac{\log p}{n}} \right).$$

By the independence between ϵ_{mi} and $\mathbf{X}_m^* - \mathbf{X}_m$, it is not hard to show

$$\begin{aligned} & \left| \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) - (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \} \hat{\beta}_j \right| \\ & \leq \left| \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) - (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \} \beta_j \right| + \left| \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) - (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \} (\beta_j - \hat{\beta}_j) \right| \\ & = O_p \left(\frac{\log p}{n} + a_n \sqrt{\frac{\log p}{n}} \right). \end{aligned}$$

Combing term (A1) and term (A2), we have

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \{ (\mathbf{X}_{m,-j}^* - \bar{\mathbf{X}}_{-j}^*) \hat{\beta}_j - (\mathbf{X}_{m,-j} - \bar{\mathbf{X}}_{-j}) \} \beta_j &= \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(X_{mi} - \bar{X}_i)(\hat{\beta}_{i,j} - \beta_{i,j}) I\{i \neq j\} \\ &+ O_p \left(\frac{\log p}{n} + a_n \sqrt{\frac{\log p}{n}} \right). \end{aligned}$$

Summarizing all the results above, by Equations (22) and (23) of Liu (2013), we have

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \epsilon_{mi}^* \epsilon_{mj}^* &= \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \tilde{\epsilon}_{mj} - \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(X_{mi} - \bar{X}_i)(\hat{\beta}_{i,j} - \beta_{i,j}) I\{i \neq j\} \\ &\quad - \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mj}(X_{mj} - \bar{X}_j)(\hat{\beta}_{j,i} - \beta_{j,i}) I\{i \neq j\} \\ &\quad + O_p \left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}} + a_n \sqrt{\frac{\log p}{n}} + b_n^2 \right). \end{aligned}$$

As $\frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(X_{mi} - \bar{X}_i) = \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}^2 + \frac{1}{n} \tilde{\epsilon}_{mi}(\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i})\beta_i$, and $\text{Var}(\mathbf{X}_{m,-i}\beta_i) = (\sigma_{ii}\omega_{ii} - 1)/\omega_{ii} \leq C$, we have

$$\max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(\mathbf{X}_{m,-i} - \bar{\mathbf{X}}_{-i})\beta_i \right| = O_p \left(\frac{\log p}{n} \right),$$

therefore

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}(X_{mi} - \bar{X}_i) &= \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}^2 + O_p(\log p/n) \\ &= \frac{1}{n} \sum_{m=1}^n \epsilon_{mi}^{*2} + O_p \left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}} + a_n \sqrt{\frac{\log p}{n}} + b_n^2 \right), \end{aligned}$$

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \epsilon_{mi}^* \epsilon_{mj}^* &= \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi} \tilde{\epsilon}_{mj} - \frac{1}{n} \sum_{m=1}^n \epsilon_{mi}^{*2} (\hat{\beta}_{i,j} - \beta_{i,j}) I\{i \neq j\} - \frac{1}{n} \sum_{m=1}^n \epsilon_{mj}^{*2} (\hat{\beta}_{j,i} - \beta_{j,i}) I\{i \neq j\} \\ &+ O_p \left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}} + a_n \sqrt{\frac{\log p}{n}} + b_n^2 \right). \end{aligned}$$

In addition

$$\frac{1}{n} \sum_{m=1}^n \epsilon_{mi}^{*2} = \frac{1}{n} \sum_{m=1}^n \tilde{\epsilon}_{mi}^2 + O_p \left(\sqrt{\frac{\log^2 p \log^2 n}{n^{3/2}}} + a_n^2 \sqrt{\frac{\log p \log^2 n}{n^{1/2}}} + a_n \sqrt{\frac{\log p}{n}} + b_n^2 \right).$$

Equation (6) follows immediately by central limit theorem.

References

- Guo, J.; Levina, E.; Michailidis, G.; Cai, T.T. Joint estimation of multiple graphical models. *Biometrika* **2015**, *98*, 1–15. [CrossRef]
- Danaher, P.; Wang, P.; Witten, D.M. The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B* **2014**, *76*, 373–397. [CrossRef]
- Zhao, S.; Cai, T.T.; Li, H. Direct estimation of differential networks. *Biometrika* **2014**, *101*, 253–268. [CrossRef]
- Liu, W. Structural similarity and difference testing on multiple sparse Gaussian graphical models. *Ann. Stat.* **2017**, *45*, 2680–2707. [CrossRef]
- Liu, H.; Lafferty, J.; Wasserman, L. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **2009**, *10*, 2295–2328.
- Xue, L.; Zou, H. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Ann. Stat.* **2012**, *40*, 2541–2571. [CrossRef]
- Efron, B. Correlation and large-scale simultaneous significance testing. *J. Am. Stat. Assoc.* **2007**, *102*, 93–103. [CrossRef]
- Zhang, Q. Direct estimation of differential network under high-dimensional nonparanormal graphical models. *Can. J. Stat.* **2019**, *48*, 1–17. [CrossRef]
- Available online: <https://cancergenome.nih.gov> (accessed on 9 December 2019).
- Zhang, Q.; Burdette, J.; Wang, J.-P. Integrative network analysis of tcga data for ovarian cancer. *BMC Syst. Biol.* **2014**, *8*, 1–18. [CrossRef]
- The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **2012**, *490*, 61–70. [CrossRef]
- Available online: <http://www.genome.jp/kegg/pathway> (accessed on 9 December 2019).
- Available online: <https://gdc.cancer.gov> (accessed on 9 December 2019).
- Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [CrossRef] [PubMed]
- Hsu, F.; Serpedin, E.; Hsiao, T.; Bishop, A.; Dougherty, E.; Chen, Y. Reducing confounding and suppression effects in tcga data: An integrated analysis of chemotherapy response in ovarian cancer. *BMC Genom.* **2012**, *13*, S13. [CrossRef] [PubMed]
- Zhang, Q. A powerful nonparametric method for detecting differentially co-expressed genes: Distance correlation screening and edge-count test. *BMC Syst. Biol.* **2018**, *12*, 58. [CrossRef] [PubMed]
- Liu, M.C.; Pitcher, B.N.; Mardis, E.R.; Davies, S.R.; Friedman, P.N.; Snider, J.E.; Vickery, T.L.; Reed, J.P.; DeSchryver, K.; Singh, B.; et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: Correlative analysis of C9741. *Breast Cancer* **2016**, *2*, 15023. [CrossRef]
- Haibe-Kains, B.; Desmedt, C.; Loi, S.; Culhane, A.C.; Bontempi, G.; Quackenbush, J.; Sotiriou, C. A three-gene model to robustly identify breast cancer molecular subtypes. *J. Natl. Cancer Inst.* **2012**, *104*, 311–325. [CrossRef]
- Gendoo, D.M.; Ratanasirigulchai, N.; Schroder, M.S.; Pare, L.; Parker, J.S.; Prat, A.; Haibe-Kains, B. Genefu: An R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **2016**, *32*, 1097–1099. [CrossRef]

20. Ding, K.; Li, W.; Zou, Z.; Zou, X.; Wang, C. CCNB1 is a prognostic biomarker for ER+ breast cancer. *Med. Hypothesis* **2014**, *83*, 359–364. [[CrossRef](#)]
21. Sun, G.; Yang, L.; Dong, C.; Ma, B.; Shan, M.; Ma, B. PRKDC regulates chemosensitivity and is a potential prognostic and predictive marker of response to adjuvant chemotherapy in breast cancer patients. *Oncol. Rep.* **2017**, *37*, 3536–3542. [[CrossRef](#)]
22. Huggett, M.; Tudzarova, S.; Proctor, I.; Loddo, M.; Keane, M.G.; Stoeber, K.; Williams, G.H.; Pereira, S.P. Cdc7 is a potent anti-cancer target in pancreatic cancer due to abrogation of the DNA origin activation checkpoint. *Oncotargets* **2016**, *7*, 18495–18407. [[CrossRef](#)]
23. Desrichard, A.; Bidet, Y.; Uhrhammer, N.; Bignon, Y. CHEK2 contribution to hereditary breast cancer in non-BRCAfamilies. *Breast Cancer Res.* **2011**, *13*, R119. [[CrossRef](#)]
24. Liu, W. Gaussian graphical model estimation with false discovery rate control. *Ann. Stat.* **2013**, *41*, 2948–2978. [[CrossRef](#)]
25. Zou, H. The Adaptive Lasso and Its Oracle Properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
26. Belloni, A.; Chernozhukov, V.; Wang, L. Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **2011**, *98*, 791–806. [[CrossRef](#)]



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).