Testing Differential Microbial Networks based on the Ising Graphical Models

January 14, 2017

Abstract

Microorganisms such as bacteria do not exist in isolation but form complex ecological interaction networks. Conventional methods such as Gaussian graphical models cannot be used to study the conditional independence among microbes, because the measurements taken are of relative bacterial abundances which are not Gaussian distributed. Alternatively, the Ising model can be used for modeling the conditional independence structure for binary data if one is only interested in the dependency structure among presence/absence of certain bacterial taxa. We propose a testing framework based on the Ising model to detect whether the microbial network structures under different treatment conditions are the same. If they are not the same, our method also conducts multiple testing of network interactions to detect microbe-microbe interactions associated with a binary trait such as presence of a disease.

1 Introduction

Microorganisms such as bacteria do not exist in isolation but form complex ecological interaction networks (Faust et al., 2012). A flexible framework for studying the dependence relationships among microbial taxa as well as detecting differential microbial interactions are important for a better understanding of the human microbiome and its association with complex traits such as diseases. There has been little attention focused on using sequencing data to learn the direct or indirect interactions among microbial taxa and infer differential microbial networks. The challenges lie in the compositional nature of the data collected, that is, the data vector lies in the high-dimensional simplex. Methods based on Gaussian graphical models thus do not apply to such compositional data. Recent work by Cao et al. (2016) proposed a method to estimate the covariance structure of compositional data via composition-adjusted thresholding, yet it does not address the question on conditional independence. Other works on modeling microbial co-occurrence relationships include Faust et al. (2012); Barberán et al. (2012), where pairwise correlations calculated using ensemble methods are used to quantify microbial co-occurrence.

In the present paper, we propose a novel approach for testing differential microbial co-occurrence networks based on the framework of discrete Markov random fields. A Markov random field is a graphical model of a joint probability distribution and offers a flexible framework to characterize the dependence relationships among variables of interest. In particular, discrete Markov random fields are those whose underlying variables take discrete values, and such models have been commonly used to study ... Formally, let $X = (X_1, \ldots, X_p)$ denote a random vector with each variable X_r taking values from the set $\mathcal{X} \in \{0, 1, \ldots, m-1\}$ for $m \geq 2$. The relationships among the variables in X can be captured by an undirected graph $G = (V, E)$ whose vertex set $V = \{1, \ldots, p\}$ and edge set E corresponds to conditional dependence. The pairwise Markov random field associated with graph G is the family of distributions that can be factorized as

$$\mathrm{pr}(X) \propto \exp\left\{ \sum_{(r,t)\in E} \phi_{rt}(X_r, X_t) \right\},$$

where for each edge $(r, t) \in E$, ϕ_{rt} is a mapping from pairs $(X_r, X_t) \in \mathcal{X}_s \times \mathcal{X}_t$ to $\mathbb{R}$. The binary Markov network (a.k.a. Ising model) is a special case with $\mathcal{X} = \{-1, 1\}$ and

$$\mathrm{pr}_\Theta(X) = \frac{1}{Z(\Theta)} \exp\left\{ \sum_{(r,t)\in E} \theta_{rt} X_r X_t \right\} = \exp\left\{ \sum_{(r,t)\in E} \theta_{rt} X_r X_t - \log Z(\Theta) \right\}. \tag{1}$$

Here $Z(\Theta)$ is the normalizing constant to ensure that the probability mass function sums to 1 over all possible X. Given a set of independently and identically distributed observations generated from the distribution pr_Θ, one can thus reconstruct the graph G via estimating the symmetric matrix $\Theta = (\theta_{rt})_{1\leq r,t\leq p}$.

To detect differential microbial co-occurrence networks associated with a binary trait, our proposal first estimates the conditional dependence among microbial taxa using discretized relative abundances. Discretization of the relative abundances allows us to surpass the challenges of having an excess of low abundances, corresponding to rare species, in analyzing compositional data. Let Θ_k be the parameter matrix that represents the microbial co-occurrence network among discretized taxa abundances under condition k for $k = 1, 2$. The difference $\Delta = \Theta_1 - \Theta_2$ is called the differential network. We are interested in global testing of differential microbial networks:

$$H_0 : \Theta_1 = \Theta_2, \quad v.s. \quad H_1 : \Theta_1 \neq \Theta_2, \tag{2}$$

and simultaneous testing of entry-wise changes in microbial interactions:

$$H_{0,rt} : \theta_{rt,1} - \theta_{rt,2} = 0, \quad v.s. \quad H_{1,rt} : \theta_{rt,1} - \theta_{rt,2} \neq 0, \quad 1 \leq r < t \leq p, \tag{3}$$

while controlling the false discovery rate at a pre-specified level.

We provide a rigorous framework for testing differential microbial co-occurrence networks, a problem that has received little attention in the literature. In addition, two-sample testing of discrete Markov networks is of independent interest. Compared to Ren et al. (2016), we are interested in two-sample global and multiple testing of binary Markov networks, whereas the main focus of Ren et al. (2016) is inference for individual entries in one matrix Θ. In addition, we provide a multiple testing procedure for the one-sample case. On the other hand, the main difference between our work and Xia et al. (2015) lies in the use of node-wise logistic regressions as opposed to node-wise linear regressions used in Xia et al. (2015). The use of node-wise logistic regressions requires novel approaches in constructing the standardized test statistics and subsequently characterizing the dependencies among these test statistics.

1.1 Organization of the paper

2 Global Testing of Differential Networks

In this section, we provide a testing procedure for the global hypothesis in (2) and study its theoretical properties. We begin with notations. Let $\{X^{(1)}, \ldots, X^{(n_1)}\}$ be the n_1 observations from the first sample and $\{Y^{(1)}, \ldots, Y^{(n_2)}\}$ the n_2 observations from the second sample, which are often conveniently written in matrix notations as $X \in \mathbb{R}^{n_1 \times p}$ and $Y \in \mathbb{R}^{n_2 \times p}$, respectively. For a matrix $X \in \mathbb{R}^{n_1 \times p}$, X_{-r} corresponds to the $n_1 \times (p-1)$ submatrix with the r-th column removed. For a matrix $\Theta_k = (\theta_{rt,k})_{1 \leq r,t \leq p}$, denote by $\theta_{r,k} = \{\theta_{rt,k}, t \in V \backslash r\}$ the $(p-1)$-dimensional subvector of parameters, where $k = 1, 2$. For a symmetric matrix A, we use $\phi_{\max}(A)$ and $\phi_{\min}(A)$ to denote respectively the largest and smallest eigenvalue of A. For a vector $\boldsymbol{x} \in \mathbb{R}^p$, the usual vector ℓ_1, ℓ_2 and ℓ_∞ norms are denoted respectively by $\|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_\infty$. The hyperbolic functions are defined as $\sinh(x) = (e^x - e^{-x})/2$ and $\cosh(x) = (e^x + e^{-x})/2$. The indicator function is denoted by $I(\cdot)$.

2.1 The testing procedure

In view of the parametrization of the Ising model in (1), the null hypothesis in (2) is equivalent to

$$H_0 : \max_{1\le r<t\le p} |\theta_{rt,1} - \theta_{rt,2}| = 0. \tag{4}$$

A natural solution to constructing the test statistics for the null in (4) is to obtain estimates of $\theta_{rt,k}$ ($k = 1, 2$) and calculate the standardized entry-wise differences. Define the function

$$f(u) = \log(e^u + e^{-u}).$$

The first and second-order derivatives of $f(u)$ are, respectively,

$$\dot f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} = \frac{\sinh(u)}{\cosh(u)}, \quad \ddot f(u) = \frac{4\exp(2u)}{\{\exp(2u)+1\}^2}.$$

The joint distribution of $X_1, \ldots, X_p$ in (1) implies that the node-wise conditional distribution $X_r \mid X_{-r}$ has the following form

$$\mathrm{pr}(X_r \mid X_{-r}) = \frac{\exp(X_r \sum_{j\neq r} X_j\theta_{rj})}{\exp(-X_r \sum_{j\neq r} X_j\theta_{rj}) + \exp(X_r \sum_{j\neq r} X_j\theta_{rj})}. \tag{5}$$

Thus one obtains immediately the conditional expectation of X_r given all remaining variables as $\mathbb{E}[X_r \mid X_{-r}] = \dot f(X_{-r}\theta_{r,1})$. Similarly we have $\mathbb{E}[Y_r \mid Y_{-r}] = \dot f(Y_{-r}\theta_{r,2})$. This motivates us to consider the following data generating models for the node-wise logistic regressions, i.e. for every node $r = 1, \ldots, p$,

$$X_r^{(i)} = \dot f(X_{-r}^{(i)}\theta_{r,1}) + \varepsilon_{r,1}^{(i)}, \quad i = 1, \ldots, n_1, \tag{6}$$

$$Y_r^{(i)} = \dot f(Y_{-r}^{(i)}\theta_{r,2}) + \varepsilon_{r,2}^{(i)}, \quad i = 1, \ldots, n_2, \tag{7}$$

where $\varepsilon_{r,k}^{(i)}$ are random variables satisfying $\mathbb{E}[\varepsilon_{r,1}^{(i)} \mid X_{-r}^{(i)}] = 0, \mathbb{E}[\varepsilon_{r,1}^{(i)} \mid Y_{-r}^{(i)}] = 0, \mathbb{E}[(\varepsilon_{r,1}^{(i)})^2 \mid X_{-r}^{(i)}] = \ddot f(X_{-r}^{(i)}\theta_{r,1})$, and $\mathbb{E}[(\varepsilon_{r,2}^{(i)})^2 \mid Y_{-r}^{(i)}] = \ddot f(Y_{-r}^{(i)}\theta_{r,2})$.

Let $\hat\theta_{r,1}$ and $\hat\theta_{r,2}$ be two initial estimators satisfying

$$\max_{1\le r\le p} \|\hat\theta_{r,k} - \theta_{r,k}\|_1 = o_{\mathrm{p}}\left(\{\log p\}^{-1}\right), \quad k = 1, 2 \tag{8}$$

$$\max_{1\le r\le p} \|\hat\theta_{r,k} - \theta_{r,k}\|_2 = o_{\mathrm{p}}\left(\{n_k \log p\}^{-1/4}\right), \quad k = 1, 2. \tag{9}$$

Note the conditions in (8) and (9) are stronger than those required for entry-wise normality in Ren et al. (2016), and are necessary for establishing the limiting null distribution for the global test in (4). Such estimators can be obtained using the ℓ_1 penalized node-wise logistic regression (Ravikumar et al., 2010) under the sparsity regime $d_{1,k} = o(n_k^{1/2}/(\log p)^{3/2})$. Here $d_{1,k}$ is defined as the *effective sparsity* of Θ_k (Ren et al., 2016):

$$d_{1,k}(\Theta_k) = \max_{1\le t\le p} \sum_{r\neq t} \min\{1, |\theta_{rt,k}|/\lambda\}, \tag{10}$$

where $\lambda > 0$ can be viewed as the signal-to-noise ratio.

The estimators $\hat{\theta}_{r,k}$ are biased. To obtain a nearly unbiased estimator of $\theta_{r,k}$, we adopt the projection-based de-biasing approach with suitably defined score vectors $v_{rt,k}^{(i)}$ for every $t \neq r$ (Ren et al., 2016). Specifically, for any $t \neq r$, the score vectors $v_{rt,k}^{(i)}$ $(k = 1, 2)$ are defined as

$$v_{rt,1}^{(i)} = \delta_{t,1}^{(i)} - g(X_{-\{r,t\}}^{(i)}, \hat{\theta}_{r,1}, \hat{\theta}_{t,1}), \quad (i = 1, \ldots, n_1), \tag{11}$$

$$v_{rt,2}^{(i)} = \delta_{t,2}^{(i)} - g(Y_{-\{r,t\}}^{(i)}, \hat{\theta}_{r,2}, \hat{\theta}_{t,2}), \quad (i = 1, \ldots, n_2). \tag{12}$$

Here $\delta_{t,1} = (X_t + 1)/2$ and the function g defined based on the X sample is

$$g(X_{-\{r,t\}}, \theta_{r,1}, \theta_{t,1}) = \frac{\mathbb{E}_{\theta_{r,1},\theta_{t,1}}\left[\delta_{t,1}\ddot{f}(X_{-r}\theta_{r,1}) | X_{-\{r,t\}})\right]}{\mathbb{E}_{\theta_{r,1},\theta_{t,1}}\left[\ddot{f}(X_{-r}\theta_{r,1}) | X_{-\{r,t\}})\right]}. \tag{13}$$

One can similarly define $\delta_{t,2}$ and $g(Y_{-\{r,t\}}, \theta_{r,2}, \theta_{t,2})$ based on the Y sample. Intuitively, the score vectors $v_{rt,k} \in \mathbb{R}^{n_k}$ resemble the residuals for regressing X_t on $X_{-\{r,t\}}$ (Y_t on $Y_{-\{r,t\}}$). The choice of $v_{rt,k}^{(i)}$ in the form of (11) and (12) ensures that they have the smallest variance and satisfies the following conditions

$$\mathbb{E}[v_{rt,k}^{(i)}\varepsilon_{r,k}^{(i)}] = 0, \quad \langle v_{rt,1}^{(i)}, X_t^{(i)}\rangle = 1, \quad \langle v_{rt,2}^{(i)}, Y_t^{(i)}\rangle = 1.$$

Let $\hat{u}_{r,1}^{(i)} = X_{-r}^{(i)}\hat{\theta}_{r,1}$ $(i = 1, \ldots, n_1)$ and $\hat{u}_{r,2}^{(i)} = Y_{-r}^{(i)}\hat{\theta}_{r,2}$ $(i = 1, \ldots, n_2)$. Then the de-biased estimators are defined as, respectively,

$$\check{\theta}_{rt,1} = \hat{\theta}_{rt,1} + \frac{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\{X_r^{(i)} - \dot{f}(\hat{u}_{r,1}^{(i)})\}}{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}}, \quad \check{\theta}_{rt,2} = \hat{\theta}_{rt,2} + \frac{\sum_{i=1}^{n_2} v_{rt,2}^{(i)}\{Y_r^{(i)} - \dot{f}(\hat{u}_{r,2}^{(i)})\}}{\sum_{i=1}^{n_1} v_{rt,2}^{(i)}\ddot{f}(\hat{u}_{r,2}^{(i)})Y_t^{(i)}}. \tag{14}$$

The basic idea of the above bias correction comes from Zhang and Zhang (2014) and local Taylor expansion, which we briefly explain as follows using the X sample. Applying local Taylor expansion to $\dot{f}(u_{r,1}^{(i)})$ around $\hat{u}_{r,1}^{(i)}$, the data generating model in (6) can be rewritten as

$$X_r^{(i)} - \dot{f}(\hat{u}_{r,1}^{(i)}) + \ddot{f}(\hat{u}_{r,1}^{(i)})X_{-r}^{(i)}\hat{\theta}_{r,1} = \ddot{f}(\hat{u}_{r,1}^{(i)})X_{-r}^{(i)}\theta_{r,1} + Re_i + \varepsilon_{r,1}^{(i)},$$

where Re_i is the residual term from the Taylor expansion and is in the order of $o_{\mathrm{p}}(\{n_1 \log p\}^{-1/2})$ under the conditions in (8) and (9). An equivalent definition of $\check{\theta}_{rt,1}$ to that in (14) is thus

$$\check{\theta}_{rt,1} = \theta_{rt,1} + \frac{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}}{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} + \frac{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_{-\{r,t\}}^{(i)}\Delta_{\{r,-t\},1}}{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} + o_{\mathrm{p}}(\{n_1 \log p\}^{-1/2}), \tag{15}$$

where $\Delta_{r,k} = \hat{\theta}_{r,k} - \theta_{r,k}$. Since $\mathbb{E}[v_{rt,k}^{(i)}\varepsilon_{r,k}^{(i)}] = 0$, the bias of $\check{\theta}_{rt,1}$ is approximately

$$\frac{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_{-\{r,t\}}^{(i)}\Delta_{\{r,-t\},1}}{\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} + o_{\mathrm{p}}(\{n_1 \log p\}^{-1/2}).$$

In the low-dimensional setting where $\mathrm{rank}(X) = p < n_1$, one can choose $v_{rt,1} = X_t^{\perp}$ as the projection of X_t onto the orthogonal complement of the column space of $X_{-\{r,t\}}$. The resulting estimator $\check{\theta}_{rt,1}$ in (14) is nearly unbiased under the conditions (8) and (9) on the initial estimators. However, in the high-dimensional regime where $p \ll n$, $X_t^{\perp}$ is no longer a valid choice of the score vector as $\mathrm{rank}(X) < p$, and the condition $v_{rt,1} \perp X_{-\{r,t\}}$ forces $v_{rt,1} = \mathbf{0}$. The idea in (14) is to project the residual $X_r - \dot{f}(\hat{u}_{r,1})$ to the direction of $v_{rt,1}$ for bias correction, which yields nearly unbiased $\check{\theta}_{rt,1}$ with appropriate choice of $v_{rt,1}$.

The estimators $\check{\theta}_{rt,k}$'s are heteroscedastic, and therefore it is necessary to quantify the variance of $\check{\theta}_{rt,k}$ before constructing the test statistic for the global null hypothesis in (4). Lemma A.1 shows that

$$\begin{aligned}\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\ddot{f}(u_{r,1}^{(i)})X_t^{(i)} - F_{rt,1}/2\right| =& O_{\mathrm{p}}\left(\sqrt{\frac{\log p}{n_1}}\right),\\ \left|\frac{1}{n_2}\sum_{i=1}^{n_2} v_{rt,2}^{(i),o}\ddot{f}(u_{r,2}^{(i)})Y_t^{(i)} - F_{rt,2}/2\right| =& O_{\mathrm{p}}\left(\sqrt{\frac{\log p}{n_2}}\right),\end{aligned}$$

where $v_{rt,k}^{(i),o}$ corresponds to the oracle score vector calculated based on $\theta_{r,k}, \theta_{t,k}$, and

$$F_{rt,k} = 4\mathbb{E}_{\Theta_k}\left[(v_{rt,k}^o)^2\ddot{f}(u_{r,k})\right]. \tag{16}$$

By (15) and Lemma A.2, the following estimators

$$\tilde{\theta}_{rt,1} = \theta_{rt,1} + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{2v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}}{F_{rt,1}}, \quad \tilde{\theta}_{rt,2} = \theta_{rt,2} + \frac{1}{n_2}\sum_{i=1}^{n_2}\frac{2v_{rt,2}^{(i),o}\varepsilon_{r,2}^{(i)}}{F_{rt,2}}, \tag{17}$$

satisfy that

$$|\check{\theta}_{rt,k} - \tilde{\theta}_{rt,k}| = o_{\mathrm{p}}\left(\{n_k \log p\}^{-1/2}\right), \quad k = 1, 2. \tag{18}$$

Consequently, one can approximate $\mathrm{Var}(\check{\theta}_{rt,k})$ using $\mathrm{Var}(\tilde{\theta}_{rt,k})$. Denote by $s_{rt,k} := \mathrm{Var}\left(2v_{rt,k}^o\varepsilon_{r,k}/F_{rt,k}\right) = 1/F_{rt,k}$. Given the two samples X and Y, $s_{rt,k}$ can be, respectively, estimated by

$$\check{s}_{rt,1} = \left(4n_1^{-1}\sum_{i=1}^{n_1}(v_{rt,1}^{(i)})^2\ddot{f}\big(X_{-r}^{(i)}\hat{\theta}_{r,1}\big)\right)^{-1}, \quad \check{s}_{rt,2} = \left(4n_2^{-1}\sum_{i=1}^{n_2}(v_{rt,2}^{(i)})^2\ddot{f}\big(Y_{-r}^{(i)}\hat{\theta}_{r,2}\big)\right)^{-1}. \tag{19}$$

Finally, define the standardized statistic

$$W_{rt} =: \frac{\check{\theta}_{rt,1} - \check{\theta}_{rt,2}}{\sqrt{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2}}, \quad 1 \le r < t \le p. \tag{20}$$

We use the following test statistic for the global test $H_0 : \Theta_1 = \Theta_2$

$$M_{n,p} = \max_{1\le r<t\le p} W_{rt}^2 = \max_{1\le r<t\le p}\frac{(\check{\theta}_{rt,1} - \check{\theta}_{rt,2})^2}{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2}. \tag{21}$$

Details of the testing procedure are provided in Algorithm 1.

Algorithm 1 Global testing

1: Calculate W_{rt} as in (20) for $1 \le r < t \le p$.

2: Form the test statistic for the global test $H_0 : \Theta_1 = \Theta_2$

$$M_{n,p} = \max_{1 \le r < t \le p} W_{rt}^2 = \max_{1 \le r < t \le p} \frac{(\check{\theta}_{rt,1} - \check{\theta}_{rt,2})^2}{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2}. \tag{22}$$

3: Reject H_0 if $\Psi_\alpha = 1$, where

$$\Psi_\alpha = I(M_{n,p} \ge q_\alpha + 4\log p - \log\log p), \tag{23}$$

and q_α is the $1-\alpha$ quantile of the type I extreme value distribution with the cumulative distribution function $\exp\{-(8\pi)^{-1/2} e^{-z/2}\}$ and $I(\cdot)$ denotes the indicator function.

Remark 1. The construction of de-biased estimate $\check{\theta}_{rt,k}$ in (17) is different from the method used in Cai et al. (2013) and Xia et al. (2015) for the Gaussian case. The rejection rule in (23) follows from the limiting null distribution of $M_{n,p}$, which will be shown in Section 2.2.

Remark 2. As mentioned earlier, the initial estimators $\hat{\theta}_{rt,k}$ can be obtained via the ℓ_1 penalized node-wise logistic regression, with tuning parameters selected via the extended Bayesian information criterion (EBIC) (Barber and Drton, 2015). The score vectors can be approximated using their empirical versions. Specifically, by definition, the population score vector $v_{rt,1}$ is equivalent to

$$\delta_{t,1} - \frac{\exp\left(X_{-\{r,t\}}\theta_{\{t,-r\},1}\right)\cosh(-\theta_{rt,1} + X_{-\{r,t\}}\theta_{\{r,-t\},1})}{\sum_{\omega=\pm 1}\exp\left(\omega X_{-\{r,t\}}\theta_{\{t,-r\},1}\right)\cosh(-\omega\theta_{rt,1} + X_{-\{r,t\}}\theta_{\{r,-t\},1})}.$$

Given the initial estimates $(\hat{\theta}_{rt,1})_{1\le r,t\le p}$, one can easily obtain estimates of $v_{rt,1}^{(i)}$ for $i = 1, \ldots, n_1$. Calculations of $v_{rt,2}^{(i)}$ for $i = 1, \ldots, n_2$ follow similarly.

2.2 Limiting null distribution

In this section, we derive the theoretical properties of $M_{n,p}$ under the null hypothesis in (2), specifically, its limiting distribution and rate-optimal power.

Let $\mathcal{A} = \{(r,t) : 1 \le r < t \le p\}$ be the pairwise index set. Throughout the section, we assume the sample sizes n_1 and n_2 are comparable such that their ratio can be treated as a constant. For distinct indices r, t, l, m, let $\mathcal{E}_{r,t,l,m} = \{[p]\backslash\{r\}, [p]\backslash\{r,t\}, [p]\backslash\{r,t,l\}, [p]\backslash\{r,t,l,m\}\}$. For each $k = 1, 2$ and node $r = 1, \ldots, p$, denote by $Q_{r,k}$ the Hessian of the local likelihood function associated with the penalized logistic regression, which can be written analytically as (Ren et al., 2016)

$$Q_{r,1} = \mathbb{E}_{\Theta_1}[\ddot{f}(X_{-r}\theta_{r,1})X_{-r}X_{-r}^T], \quad Q_{r,2} = \mathbb{E}_{\Theta_2}[\ddot{f}(Y_{-r}\theta_{r,2})Y_{-r}Y_{-r}^T].$$

We make the following assumptions.

(A1). Assume that $\log p = o(n_k^{1/5})$ and $n_1 \asymp n_2$. For each $r = 1, \ldots, p$, there exist constants $C_{\min}, C_{\max} >$

0 such that

$$0 < C_{\min} \leq \phi_{\min}(Q_{r,k}) \leq \phi_{\max}(Q_{r,k}) \leq C_{\max} < \infty.$$

Further, there exists $c^* > 0$ such that $\min_{1\leq r<t\leq p} F_{rt,k} > c^* > 0$ for $k = 1, 2$.

(A2). Suppose that for some constant $\zeta > 0$, we have $\max_{1\leq r<t\leq p} |\theta_{rt,k}| \leq \zeta < \infty$ for $k = 1, 2$. For a constant $\xi > 0$, denote by

$$\mathcal{A}_t(\xi) = \{r : |\sinh(2\theta_{rt,1})| \geq (\log p)^{-2-\xi} \text{ or } |\sinh(2\theta_{rt,2})| \geq (\log p)^{-2-\xi}\}.$$

Assume that the size of $\mathcal{A}_t(\xi)$ is small such that $\max_{1\leq t\leq p} |\mathcal{A}_t(\xi)| = o(p^\gamma)$ for $0 < \gamma < \min\{1/3, (1-\zeta)/(1+\zeta)\}$.

(A3). For any $r, t, l, m \in [p]$ and any $\Lambda \subset \mathcal{E}_{r,t,l,m}$, there exists a constant $C_w > 0$ such that

$$\mathbb{E}[\exp(2\sum_{j\in\Lambda} \theta_{rj,1}X_j) + \exp(-2\sum_{j\in\Lambda} \theta_{rj,1}X_j)] < C_w < \infty,$$
$$\mathbb{E}[\exp(2\sum_{j\in\Lambda} \theta_{rj,2}Y_j) + \exp(-2\sum_{j\in\Lambda} \theta_{rj,2}Y_j)] < C_w < \infty.$$

The first part of Assumption (A1) is necessary to ensure a good initial estimator as stated in (8) and (9). Note the matrix $Q_{r,k}$ corresponds to the Fisher information matrix associated with the local conditional probability distribution, and is analogous to the $(p-1)\times(p-1)$ submatrix of the covariance matrix in Gaussian graphical models. The second part of Assumption (A1) guarantees that the variance of $\tilde{\theta}_{rt,k}$ is finite. Indeed, as mentioned in Ren et al. (2016), if $\mathbb{E}[\exp(CX_{-r}\theta_{r,1}) + \exp(-CX_{-r}\theta_{r,1})] < C_w < \infty$ and $\mathbb{E}[\exp(CY_{-r}\theta_{r,2}) + \exp(-CY_{-r}\theta_{r,2})] < C_w < \infty$ for some constant $C > 4$, then $\min_{1\leq r<t\leq p} F_{rt,k}$ is bounded away from zero. Variants of Assumption (A2) are commonly used in previous work (Cai et al., 2013; Cai and Liu, 2016; Xia et al., 2015). Intuitively, this assumption says that the pairwise partial correlations are not too large and ensures that the test statistic W_{rt}'s are weakly dependent on each other. In the case of Gaussian graphical models, Xia et al. (2015) made assumptions directly in terms of entries in the precision matrix. In the current context of Ising graphical models, the dependencies among the test statistic W_{rt}'s can be approximately quantified by $\mathrm{Cov}(v_{rt,k}\varepsilon_k, v_{r't',k}\varepsilon_k)$ for $(r,t) \neq (r',t')$ and $k = 1, 2$. Assumption (A3) implies that for any $\Lambda \subset \mathcal{E}_{r,t,l,m}$, $\sum_{j\in\Lambda} \theta_{rj,1}X_j$ and $\sum_{j\in\Lambda} \theta_{rj,2}Y_j$ are sub-exponential random variables, and together with Assumption (A2) ensure that $\mathrm{Cov}(v_{rt,k}\varepsilon_k, v_{r't',k}\varepsilon_k)$ for $(r,t) \neq (r',t')$ is not too large.

We are ready to state our first main result regarding the asymptotic distribution of the global test in Algorithm 1.

Theorem 2.1. *Suppose Assumptions (A1), (A2), (A3),* (8) *and* (9) *hold. Then under the null hypothesis in* (2)*, for any* $z \in \mathbb{R}$,

$$\mathrm{pr}(M_{n,p} - 4\log p + \log\log p \leq z) \to \exp\{-(8\pi)^{-1/2}e^{-z/2}\}, \quad n_1, n_2, p \to \infty, \tag{24}$$

where $M_{n,p}$ is defined in (21). *Under the null, the convergence in* (24) *is uniform for all X and Y satisfying Assumptions (A1), (A2), (A3),* (8) *and* (9).

Remark 3. The main steps in the proof of Theorem 2.1 include approximating the test statistic $M_{n,p}$ by the following one calculated based on $\tilde{\theta}_{rt,k}$:

$$\tilde{M}_{n,p} = \max_{1 \le r < t \le p} \frac{(\tilde{\theta}_{rt,1} - \tilde{\theta}_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2},$$

since $\tilde{\theta}_{rt,k}$ is close to $\check{\theta}_{rt,k}$ for all $1 \le r < t \le p$ and $k = 1, 2$. By definition of $\tilde{\theta}_{rt,k}$ in (17) and Assumption (A3), the dependence between $\tilde{\theta}_{rt,k}$ and $\tilde{\theta}_{r't',k}$ for $(r', t') \ne (r, t)$ is upper bounded in magnitude by

$$\big\{|\sinh(2\theta_{rr',k})| + |\sinh(2\theta_{rt',k})|\big\}\big\{|\sinh(2\theta_{tt',k})| + |\sinh(2\theta_{tr',k})|\big\},$$

up to some constants that are independent of Θ_k. Assumption (A2) implies that such dependencies are weak, and thus allows us to apply strategies developed in Xia et al. (2015) for the Gaussian case. Details regarding the proof of Theorem 2.1 are available in the Appendix.

2.3 Asymptotic power

Besides the limiting distribution of $M_{n,p}$ under the null hypothesis, our proposed testing procedure in Algorithm 1 also maintains rate-optimal power. For a pre-specified $\alpha \in (0, 1)$, let q_α be the $1-\alpha$ quantile of the type I extreme value distribution with the cumulative distribution function $\exp\{-(8\pi)^{-1/2}e^{-z/2}\}$. Consider the following class of partial correlation matrices

$$\mathcal{U}(c) = \Big\{(\Theta_1, \Theta_2) : \max_{1 \le r < t \le p} \frac{|\theta_{rt,1} - \theta_{rt,2}|}{(s_{rt,1}/n_1 + s_{rt,2}/n_2)^{1/2}} \ge c(\log p)^{1/2}\Big\}. \qquad (25)$$

The next theorem shows that the test Ψ_α defined in (23) rejects H_0 in (2) with high probability when $(\Theta_1, \Theta_2) \in \mathcal{U}(4)$.

Theorem 2.2. *Consider the test Ψ_α in* (23). *Suppose that Assumptions (A1), (A2), (A3),* (8) *and* (9) *hold. Then*

$$\inf_{(\Theta_1, \Theta_2) \in \mathcal{U}(4)} \mathrm{pr}(\Psi_\alpha = 1) \to 1, \quad n_1, n_2, p \to \infty.$$

The following result shows that the lower bound in (25) is rate-optimal. Let $\mathcal{T}_\alpha$ be the set of all α-level tests, that is, $\mathrm{pr}(T_\alpha = 1) \le \alpha$ for any $T_\alpha \in \mathcal{T}_\alpha$.

Theorem 2.3. *Suppose that $\log p = o(n_k)$ for $k = 1, 2$. Let $\alpha, \beta > 0$ and $\alpha + \beta < 1$. Then there exists a constant $c_0 > 0$ such that for all sufficiently large n_1, n_2 and p,*

$$\inf_{(\Theta_1, \Theta_2) \in \mathcal{U}(c_0)} \sup_{T_\alpha \in \mathcal{T}_\alpha} \mathrm{pr}(T_\alpha = 1) \le 1 - \beta.$$

Remark 4. Theorem 2.3 says that the lower bound in the order of $(\log p)^{1/2}$ in (25) cannot be improved, because for any $c_0 > 0$ sufficiently small, there always exists an α-level test that is unable to reject the global null hypothesis in (2) with probability tending to one.

3 Multiple Testing with False Discovery Rate Control

If the global null $\Theta_1 = \Theta_2$ is rejected, simultaneous testing of pairwise partial correlations is often of interest to identify the differential network $\Delta = \Theta_1 - \Theta_2$. To this end, we consider multiple testing of all entries in Δ, i.e. for every $1 \le r < t \le p$,

$$H_{0,rt} : \theta_{rt,1} - \theta_{rt,2} = 0, \quad v.s. \quad H_{1,rt} : \theta_{rt,1} - \theta_{rt,2} \neq 0. \tag{26}$$

The test statistic for (26) is

$$W_{rt} =: \frac{\check{\theta}_{rt,1} - \check{\theta}_{rt,2}}{\sqrt{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2}}, \quad 1 \le r < t \le p.$$

Before introducing the multiple testing procedure, it is helpful to understand the properties of the test statistics W_{rt}. Under the null hypothesis (26) and some regularity conditions, one can show that

$$\sup_{0 \le \tau \le c\sqrt{\log p}} \left| \frac{\mathrm{pr}(|W_{rt}| \ge \tau)}{2 - 2\Phi(\tau)} - 1 \right| \to 0, \quad n_1, n_2 \to \infty, \tag{27}$$

uniformly for all $1 \le r < t \le p, p = n_k^{\gamma}$, for any $c > 0$ and any $\gamma > 0$. Here $\Phi(\tau)$ is the standard normal cumulative distribution function. Denote the set of true nulls by $\mathcal{H}_0 = \{(r,t) : \theta_{rt,1} = \theta_{rt,2}, 1 \le r < t \le p\}$. Since the asymptotic null distribution of every W_{rt} is standard normal, it is easy to see that

$$\mathrm{pr}\left(\max_{(r,t)\in\mathcal{H}_0} |W_{rt}| \ge 2\sqrt{\log p} \right) \to 0, \quad n_1, n_2, p \to \infty. \tag{28}$$

We are now ready to present the multiple testing procedure. For any given threshold level $\tau > 0$, the null hypothesis $H_{0,rt}$ is rejected if $|W_{rt}| \ge \tau$. Let $R_0(\tau) = \sum_{(r,t)\in\mathcal{H}_0} I(|W_{rt}| \ge \tau)$ be the total number of false positives and $R(\tau) = \sum_{1\le r<t\le p} I(|W_{rt}| \ge \tau)$ the total number of rejections. The false discovery proportion and false discovery rate are defined, respectively, as

$$\mathrm{FDP}(\tau) = \frac{R_0(\tau)}{\max\{R(\tau), 1\}}, \quad \mathrm{FDR}(\tau) = \mathbb{E}[\mathrm{FDP}(\tau)]. \tag{29}$$

For a pre-specified level α, an ideal choice of τ that is able to control the false discovery proportion and false discovery rate is

$$\tau_0 = \inf\left\{0 \le \tau \le 2(\log p)^{1/2} : \mathrm{FDP}(\tau) \le \alpha\right\},$$

where the upper bound for τ is due to (28). Since $\mathcal{H}_0$ is unknown, one can estimate $R_0(\tau)$ by $2\{1 - \Phi(\tau)\}|\mathcal{H}_0|$ as in Cai and Liu (2016), where $|\mathcal{H}_0|$ is estimated by $q = (p^2 - p)/2$ due to the sparsity of Δ.

In fact, Cai and Liu (2016) showed that

$$\sup_{0\le\tau\le b_p}\left|\frac{R_0(\tau)}{\{2-2\Phi(\tau)\}|\mathcal{H}_0|}-1\right|\to 0, \tag{30}$$

in probability for $b_p = \sqrt{4\log p - 2\log(\log p)}$ as $n_1, n_2, p \to \infty$. Thus we shall estimate $\mathrm{FDP}(\tau)$ by $\frac{\{2-2\Phi(\tau)\}(p^2-p)/2}{\max\{R(\tau),1\}}$.

The multiple testing procedure is summarized in Algorithm 2.

Algorithm 2 Multiple testing with FDR control

1: Calculate the test statistic W_{rt} for $1 \le r < t \le p$.

2: For given $0 \le \alpha \le 1$, calculate

$$\hat{\tau} = \inf\left\{0 \le \tau \le \sqrt{4\log p - 2\log(\log p)} : \frac{2\{1-\Phi(\tau)\}(p^2-p)/2}{\max\{R(\tau),1\}} \le \alpha\right\}.$$

If $\hat{\tau}$ does not exist, set $\hat{\tau} = \sqrt{4\log p}$.

3: For $1 \le r < t \le p$, reject $H_{0,rt}$ if $|W_{rt}| \ge \hat{\tau}$.

Remark 5. The multiple testing procedure in Algorithm 2 relies on the approximation in (30), which does not hold when $\tau > b_p$. In this case, direct thresholding on W_{rt}'s with $\hat{\tau} = \sqrt{4\log p}$ is used to control the false discovery proportion. Another way of approximating $R_0(\tau)$ is to use bootstrap, as done in Cai and Liu (2016).

The following theorem shows that the above procedure controls the false discovery proportion and false discovery rate at the pre-specified level α asymptotically. Let $q_0 = |\mathcal{H}_0|, q = (p^2-p)/2$, and

$$\mathcal{S}_\rho = \left\{(r,t) : 1 \le r < t \le p, \frac{|\theta_{rt,1}-\theta_{rt,2}|}{(s_{rt,1}/n_1 + s_{rt,2}/n_2)^{1/2}} \ge (\log p)^{1/2+\rho}\right\}.$$

Theorem 3.1. *Suppose $|\mathcal{S}_\rho| \ge \{\frac{1}{(8\pi)^{1/2}\alpha} + \zeta\}(\log\log p)^{1/2}$ for some $\rho > 0$ and $\zeta > 0$. Assume further that $q_0 \ge c_1 p^2$ for some $c_1 > 0$ and $p \le c_2 n_k^\gamma$ for some $c_2 > 0$ and $\gamma > 0$. Then under Assumption (A1), (A2), (A3),* (8) *and* (9)*, we have*

$$\frac{\mathrm{FDR}(\hat{\tau})}{\alpha q_0/q} \to 1, \quad \frac{\mathrm{FDP}(\hat{\tau})}{\alpha q_0/q} \to 1$$

in probability, as $n_1, n_2, p \to \infty$.

Remark 6. Proof of Theorem 3.1 follows the same strategy used in Xia et al. (2015). Again this requires knowledge of the dependence among the W_{rt}'s.

4 Simulations

In this section, we evaluate the performance of the proposed test, including the size and power of the global test, and the false discovery rate controlled multiple testing procedure.

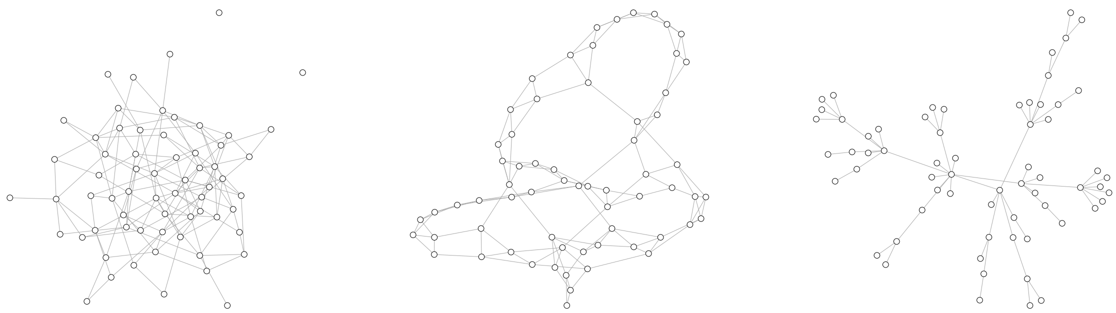

Figure 1: Illustrations of different graph classes used in simulations: Erdős-Rényi random graph (left), Watts-Strogatz small-world network (middle), and Barabasi-Albert scale-free network (right).

4.1 Data and model selection

Consider the following three different classes of graphs: (a) the Erdős-Rényi random graph $G_{\mathrm{ER}}(p, d/p)$ (Erdős and Rényi, 1960) with an average degree $d = 4$, (b) the Watts-Strogatz model G_{WS} (Watts and Strogatz, 1998) forming networks with small-world properties, and (c) the Barabasi-Albert scale-free network model G_{BA} (Barabási and Albert, 1999), as illustrated in Figure 1. In all comparisons, the graph sizes $p \in \{100, 200\}$ and the sample sizes $n_1 = n_2 \in \{100, 200\}$. For a given graph, we examined the performance of models with *mixed couplings* such that the nonzero entries $\theta_{rt,k}$ were drawn uniformly from $[-0.5, -0.1] \cup [0.1, 0.5]$. Given the two matrices Θ_1 and Θ_2, we generated the data $\{X^{(i)}\}_{i=1}^{n_1} \sim \mathrm{P}_{\Theta_1}$ and $\{Y^{(i)}\}_{i=1}^{n_2} \sim \mathrm{P}_{\Theta_2}$ by Gibbs sampling.

To get an initial estimator $\hat{\Theta}_k$ of Θ_k, we first run node-wise ℓ_1 regularized logistic regression with penalty parameter $\lambda_{r,k}$, using the R-package `glmnet` from Friedman et al. (2010). Symmetric partial correlation estimates are obtained by aggregating the node-wise estimates of regression coefficients using the AND rule, i.e. $\hat{\theta}_{rt,k} := (\hat{\theta}_{rt,k} + \hat{\theta}_{tr,k})/2$. The optimal penalty parameter $\lambda_{r,k}$'s were chosen in a way so as to maximize the performance of the global testing and the multiple testing procedures. For the global testing, the tuning parameter $\lambda_{r,k}$ in each local regression was selected based on the extended Bayesian information criterion (EBIC) (Barber and Drton, 2015). On the other hand, since the proposed multiple testing procedure relies on the approximation of $2\{1 - \Phi(\tau)\}|\mathcal{H}_0|$ to $R_0(\tau)$, thus the tuning parameters needed in Algorithm 2 were chosen with the principle of ensuring $2\{1 - \Phi(\tau)\}|\mathcal{H}_0|$ to be as close as possible to $\sum_{(r,t)\in\mathcal{H}_0} I(|W_{rt}| \geq \tau)$. That is, to minimize the error

$$\int_c^1 \left(\frac{\sum_{(r,t)\in\mathcal{H}_0} I\{|W_{rt}| \geq \Phi^{-1}(1 - z/2)\}}{p(p-1)\cdot z} - 1 \right)^2 dz,$$

where $c > 0$ is a fixed number bounded away from zero. To be more specific, we selected the tuning parameters as follows.

- Let $\lambda_{r,k} = (b/20)\sqrt{\hat{\Sigma}_{rr,k} \log p / n_k}$ for $r = 1, \ldots, p, k = 1, 2$ and $b = 1, \ldots, 40$. For each b, calculate

Table 1: Empirical sizes (type I error) of the global testing with $\alpha = 5\%, n_1 = n_2 = n$ over 1000 replications

	n	G_{ER}	G_{WS}	G_{BA}
		type I error		
p=100	100	2.8	3.1	3.0
	200	2.0	2.5	2.6
p=200	100	3.0	3.1	2.9
	200	2.4	2.3	2.9
		power		
p=100	100	73.4	79.7	72.4
	200	100.0	99.9	99.9
p=200	100	84.0	84.3	86.5
	200	100.0	100.0	100.0

$\hat{\theta}_{r,k}$ via the penalized logistic regression for each k and construct the test statistic $W_{rt}^{(b)}$.

- The optimal $\hat{b}$ was chosen as

$$\hat{b} = \arg\min_{b} \sum_{l=1}^{10} \left(\frac{\sum_{1 \leq r < t \leq p} I\{|W_{rt}^{(b)}| \geq \Phi^{-1}(1 - l[1 - \Phi(\sqrt{\log p})]/20)\}}{p(p-1) \cdot l[1 - \Phi(\sqrt{\log p})]/10} \right)^2 .$$

4.2 Results

Table 1 presents the empirical sizes (i.e. type I errors) of the global test. In all settings, the global testing procedure in Algorithm 1 controls the type I errors well.

To evaluate the power of the global test, we constructed the difference matrix $\Delta = (\Delta_{ij})$ such that five entries in the upper triangular part of Δ were randomly drawn from $[-2\sqrt{\log p/n_k}, -\sqrt{\log p/n_k}] \cup [\sqrt{\log p/n_k}, 2\sqrt{\log p/n_k}]$. Let Θ_0 be generated from one of the three aforementioned models, $\Theta_1 = \Theta_0 - \Delta$ and $\Theta_2 = \Theta_0 + \Delta$. The empirical powers (i.e. the empirical rates of rejecting the null hypothesis) for testing $\Theta_1 = \Theta_2$ are shown in Table 1. One can see that the proposed global testing procedure yields

Finally, using the above design, we examined multiple testing of individual entries in the differential network $\Theta_2 - \Theta_1$ while controlling the false discovery rate at $\alpha = 10\%$. The true differential network Δ was constructed to be sparse such that the number of edges is approximately $0.02 \cdot p(p-1)/2$, with nonzero entries drawn uniformly from $[-0.5, -0.1] \cup [0.1, 0.5]$. The empirical false discovery rates were evaluated by

$$\text{Average} \left\{ \frac{\sum_{(r,t) \in \mathcal{H}_0} I(|W_{rt}| \geq \hat{\tau})}{\max\{R(\hat{\tau}), 1\}} \right\},$$

and the empirical true positive rates by

$$\text{Average} \left\{ \frac{\sum_{(i,j) \in \mathcal{H}_1} I(|W_{rt}| \geq \hat{\tau})}{\sum_{1 \leq r < t \leq p} I(\theta_{rt,1} \neq \theta_{rt,2})} \right\},$$

where $\mathcal{H}_1$ denotes the set of nonzero locations. The ROC curves for recovering the true different networks with different dimensions and sample sizes are shown in Figure 2. Table 2 presents the empirical false

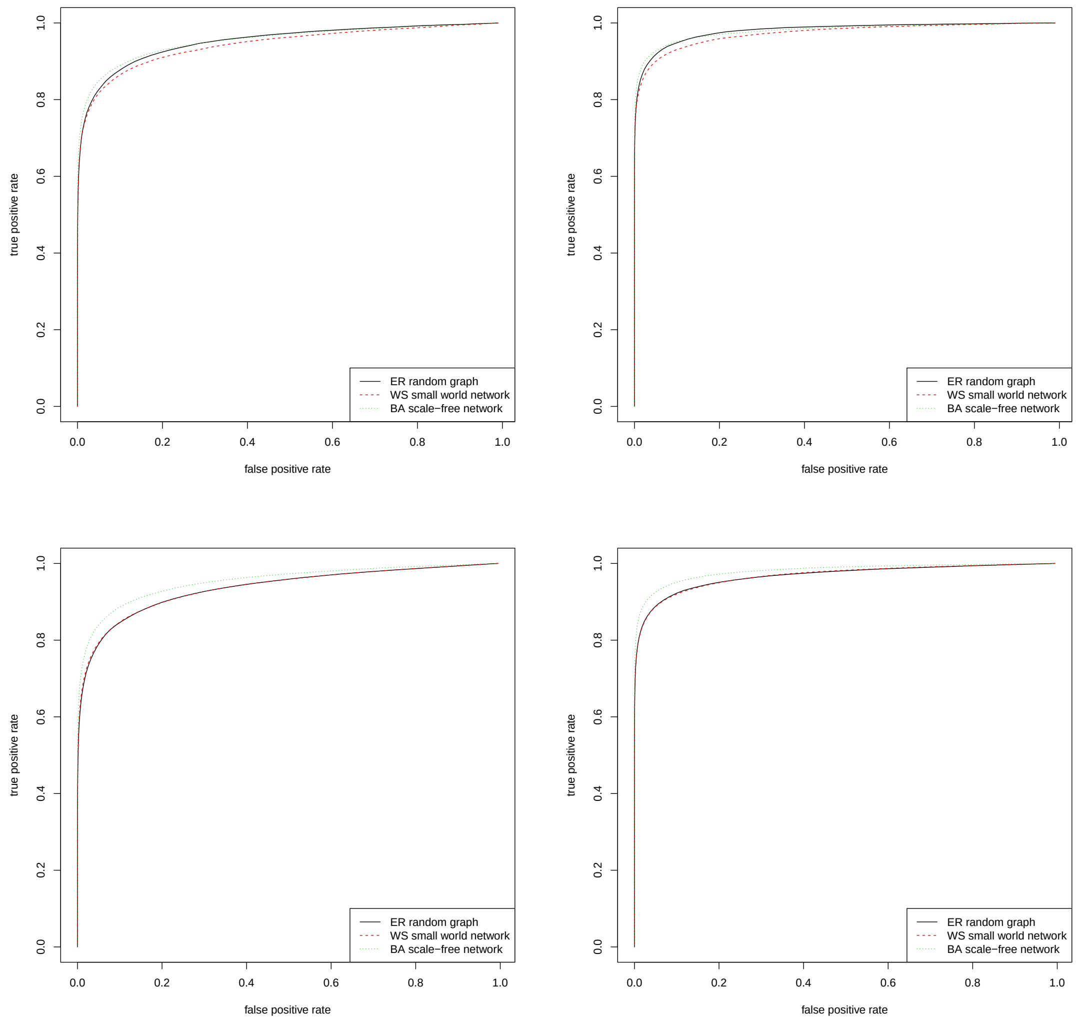

Figure 2: ROC curves for multiple testing of Δ evaluated over 200 replications. Curves in the first row correspond to dimension $p = 100$, with the second row to $p = 200$, whereas those in the first column correspond to sample size $n = 100$ and the second column to $n = 200$.

discovery rates and true positive rates evaluated over 200 replications based on $n_1 = n_2 = n$ observations. In addition, we compared the performance of direct inference of the differential network via two-sample testing with the results obtained via two one-sample inference of the individual networks. This is shown ...

5 Application to Gut Microbiome in UK Twins

We apply the proposed testing procedures to a gut microbiome study in UK twins (Goodrich et al., 2016). The original study in Goodrich et al. (2016) investigates whether host genotype shapes the gut microbiome, using 16S rRNA sequencing data collected from 3,261 fecal samples of 2,731 individuals (530 individuals sampled at a second visit). The data are available at `http://www.ebi.ac.uk/ena/data/view/PRJEB13747`. In the present paper, we are interested in whether the underlying microbial co-

Table 2: Empirical false discovery rates and powers (%) for multiple testing with $\alpha = 10\%, n_1 = n_2 = n$ over 200 replications

	n	G_{ER}	G_{WS}	G_{BA}
		false discovery rate		
p=100	100	10.0	9.1	9.0
	200	9.4	10.7	9.8
p=200	100	9.8	11.6	8.6
	200	8.5	9.4	9.1
		true positive rate		
p=100	100	56.0	56.8	62.7
	200	73.8	74.3	79.1
p=200	100	49.2	53.1	58.2
	200	69.1	70.2	77.0

occurrence relationships are associated with the host age. To this end, we focused only on measurements from the first visit consisting of 2,731 samples across 294 bacterial taxa at the genus level. Taxa with more than 90% zero counts were removed, leaving a matrix of 82 bacterial taxa collected for 2,731 samples. To examine the association of microbial interactions with host age, we selected two groups of subjects, one with individuals no more than 51 years old and the other with individuals at least 69 years old, for further investigation into differential network analysis. The number of subjects in each group is $n_1 = 609$ and $n_2 = 603$.

References

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, 1999.

Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

Rina Foygel Barber and Mathias Drton. High-dimensional ising model selection with bayesian information criteria. *Electronic Journal of Statistics*, 9(1):567–607, 2015.

Albert Barberán, Scott T Bates, Emilio O Casamayor, and Noah Fierer. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal*, 6(2):343–351, 2012.

T Tony Cai and Weidong Liu. Large-scale multiple testing of correlations. *Journal of the American Statistical Association*, 111(513):229–240, 2016.

T Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501): 265–277, 2013.

Yuanpei Cao, Wei Lin, and Hongzhe Li. Large covariance estimation for compositional data via composition-adjusted thresholding. *arXiv preprint arXiv:1601.04397*, 2016.

Paul Erdős and A Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5: 17–61, 1960.

Karoline Faust, J Fah Sathirapongsasuti, Jacques Izard, Nicola Segata, Dirk Gevers, Jeroen Raes, and Curtis Huttenhower. Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, 8(7):e1002606, 2012.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

Julia K Goodrich, Emily R Davenport, Michelle Beaumont, Matthew A Jackson, Rob Knight, Carole Ober, Tim D Spector, Jordana T Bell, Andrew G Clark, and Ruth E Ley. Genetic determinants of the gut microbiome in uk twins. *Cell Host & Microbe*, 19(5):731–743, 2016.

Bing-Yi Jing, Qi-Man Shao, and Qiying Wang. Self-normalized cramér-type large deviations for independent random variables. *The Annals of Probability*, 31(4):2167–2215, 2003.

Pradeep Ravikumar, Martin J Wainwright, and John D Lafferty. High-dimensional ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

Zhao Ren, Cun hui Zhang, and Harrison H. Zhou. Asymptotic normality in estimation of large ising graphical model. *Preprint*, 2016.

Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393 (6684):440–442, 1998.

Yin Xia, Tianxi Cai, and T Tony Cai. Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika*, 102(2):247–266, 2015.

A Yu Zaitsev. On the gaussian approximation of convolutions under multidimensional analogues of sn bernstein's inequality conditions. *Probability Theory and Related Fields*, 74(4):535–566, 1987.

Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.

A Technical Details and Proofs

A.1 Auxiliary Lemmas

In this section, we list a few technical lemmas that are necessary for establishing the theoretical properties of the proposed testing procedures.

Our first lemma is adapted from Lemma 3 of Ren et al. (2016). The concentration bound in Lemma A.1 will be used in proving Lemma A.2.

Lemma A.1. *We have uniformly for all* $1 \le r < t \le p$,

$$\left| \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i),o} \ddot{f}(u_{r,1}^{(i)}) X_t^{(i)} - \frac{1}{2} F_{rt,1} \right| = O_{\mathrm{p}} \left(\{\log p / n_1\}^{-1/2} \right), \tag{31}$$

$$\left| \frac{1}{n_2} \sum_{i=1}^{n_2} v_{rt,2}^{(i),o} \ddot{f}(u_{r,2}^{(i)}) Y_t^{(i)} - \frac{1}{2} F_{rt,2} \right| = O_{\mathrm{p}} \left(\{\log p / n_2\}^{-1/2} \right), \tag{32}$$

and

$$\max_{1 \le r < t \le p} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i),o} \ddot{f}(u_{r,1}^{(i)}) X_{-\{r,t\}}^{(i)} \right| = O_{\mathrm{p}} \left(\{\log p / n_1\}^{-1/2} \right), \tag{33}$$

$$\max_{1 \le r < t \le p} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i),o} \ddot{f}(u_{r,1}^{(i)}) Y_{-\{r,t\}}^{(i)} \right| = O_{\mathrm{p}} \left(\{\log p / n_2\}^{-1/2} \right). \tag{34}$$

Here $v_{rt,k}^{(i),o}$ *is the oracle score vector obtained with* $\theta_{r,k}$ *and* $\theta_{t,k}$, $u_{r,1}^{(i)} = X_{-r}^{(i)} \theta_{r,1}$, $u_{r,2}^{(i)} = Y_{-r}^{(i)} \theta_{r,2}$ *and* $F_{rt,k}$*'s are defined in* (16).

The proof of Lemma A.1 uses the fact that $|v_{rt,1}^{(i),o}| \le 1, |\ddot{f}(u_{r,1}^{(i)})| \le 1$, $|X_t^{(i)}| = 1$ and $|Y_t^{(i)}| = 1$ for all $1 \le r < t \le p$. Thus Hoeffding's inequalities can be applied to independent bounded random variables $v_{rt,1}^{(i),o} \ddot{f}(u_{r,1}^{(i)}) X_{-\{r,t\}}^{(i)}$ and $v_{rt,1}^{(i),o} \ddot{f}(u_{r,1}^{(i)}) X_t^{(i)} - F_{rt,1}/2$ to obtain the inequalities (33) and (31). Similarly one can show (34) and (32). More details are available in the proof of Lemma 3 of Ren et al. (2016).

Lemma A.2. *Under assumptions (A1) and (A2), uniformly for all* $1 \le r < t \le p$,

$$|\check{\theta}_{rt,k} - \tilde{\theta}_{rt,k}| = o_{\mathrm{p}} \left(\{n_k \log p\}^{-1/2} \right), \quad k = 1, 2. \tag{35}$$

Proof of Lemma A.2. Let $\Delta_{r,k} = \theta_{r,k} - \hat{\theta}_{r,k}$ for $r = 1, \ldots, p$ and $k = 1, 2$. We only prove (35) for the case where $k = 1$. The case where $k = 2$ can be shown similarly.

We first remind the readers that the definition of $\check{\theta}_{rt,1}$ in (15) can be written more specifically as

$$\check{\theta}_{rt,1} = \theta_{rt,1} + \frac{n_1^{-1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \varepsilon_{r,1}^{(i)}}{n_1^{-1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \ddot{f}(\hat{u}_{r,1}^{(i)}) X_t^{(i)}} + \frac{REM_1 + REM_2}{n_1^{-1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \ddot{f}(\hat{u}_{r,1}^{(i)}) X_t^{(i)}}, \tag{36}$$

where the two remainder terms are, respectively,

$$\begin{aligned}
REM_1 =& \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \cdot Re_i \\
=& \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \cdot \ddot{f}(\hat{u}_{r,1}^{(i)}) h_{\{i\}} - \frac{1}{n_1} \sum_{i=1}^{n_1} \int_0^1 v_{rt,1}^{(i)} \cdot \dddot{f}(\hat{u}_{r,1}^{(i)} - z h_{\{i\}}) h_{\{i\}} dz, \\
REM_2 =& \frac{1}{n_1} \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \cdot \ddot{f}(\hat{u}_{r,1}^{(i)}) X_{-\{r,t\}}^{(i)} \Delta_{\{r,-t\},1}.
\end{aligned}$$

Using the definition of $\check{\theta}_{rt,1}$ in (36), we must have

$$
\begin{aligned}
\check{\theta}_{rt,1} - \tilde{\theta}_{rt,1} &= \frac{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}}{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} + \frac{REM_1 + REM_2}{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} - \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{2v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}}{F_{rt,1}} \\
&= \underbrace{\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}}{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}} - \frac{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}}{F_{rt,1}/2}}_{(i)} + \\
&\quad \underbrace{\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}}{F_{rt,1}/2} - \frac{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}}{F_{rt,1}/2}}_{(ii)} + \underbrace{\frac{REM_1 + REM_2}{n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}}}_{(iii)} \\
&= a_1 + a_2 + a_3.
\end{aligned}
$$

To bound $|\check{\theta}_{rt,1} - \tilde{\theta}_{rt,1}|$, it suffices to bound $|a_k|$ $(k = 1, 2, 3)$ separately. In the following, we first quantify the deviance $\ddot{f}(\hat{u}_{r,1}^{(i)}) - \ddot{f}(u_{r,1}^{(i)})$ and $v_{rt,1}^{(i)} - v_{rt,1}^{(i),o}$, which are then used to bound $|a_k|$ for $k = 1, 2, 3$.

Denote by $h_{(i)} = X_{-r}^{(i)}\Delta_{r,1}, h_{(i),1} = X_{-\{r,t\}}^{(i)}\Delta_{\{r,-t\},1}$ and $h_{(i),2} = X_{-\{r,t\}}^{(i)}\Delta_{\{t,-r\},1}$. Under Assumption (A1), and conditions on the initial estimation errors in (8) and (9), we can show via Hölder's inequality that

$$\max\left\{h_{(i)}, |h_{(i),1}|, |h_{(i),2}|\right\} = o\left(\{\log p\}^{-1}\right), \tag{37}$$

and

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\ddot{f}(u_{r,1}^{(i)})(h_{(i),1}^2 + h_{(i),2}^2 + \Delta_{\{r,t\},1}^2) = o\left(\{n_1 \log p\}^{-1/2}\right). \tag{38}$$

Using the fact that

$$\exp(-2h)\ddot{f}(x) \leq \ddot{f}(x+h) \leq \exp(2h)\ddot{f}(x), \quad h \geq 0, \tag{39}$$

we have

$$|\ddot{f}(u_{r,1}^{(i)}) - \ddot{f}(\hat{u}_{r,1}^{(i)})| \leq |\exp(2|h_{(i)}|) - 1|\ddot{f}(u_{r,1}^{(i)}) = o\left(\{\log p\}^{-1}\right). \tag{40}$$

In the meanwhile,

$$
\begin{aligned}
\max_{1\leq i\leq n_1}|v_{rt,1}^{(i)} - v_{rt,1}^{(i),o}| &= \max_{1\leq i\leq n_1}\left|g(X_{-\{r,t\}}^{(i)}, \hat{\theta}_{r,1}, \hat{\theta}_t, 1) - g(X_{-\{r,t\}}^{(i)}, \theta_{r,1}, \theta_{t,1})\right| \\
&\leq \max_{1\leq i\leq n_1}\left|\exp(|h_{(i),1}| + |h_{(i),2}| + |\Delta_{\{r,t\},1}|) - 1\right| = o\left(\{\log p\}^{-1}\right),
\end{aligned} \tag{41}
$$

where the inequality follows from Lemma 1 of Ren et al. (2016).

Consequently, by (31), (40) and (41), we have

$$
\begin{aligned}
&\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)} - F_{rt,1}/2\right| \leq \left|\frac{1}{n_1}\sum_{i=1}^{n_1}\left(v_{rt,1}^{(i)} - v_{rt,1}^{(i),o}\right)\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)}\right| \\
&+ \left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\left(\ddot{f}(\hat{u}_{r,1}^{(i)}) - \ddot{f}(u_{r,1}^{(i)})\right)X_t^{(i)}\right| + \left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\ddot{f}(u_{r,1}^{(i)})X_t^{(i)} - F_{rt,1}/2\right| \leq o_{\mathrm{p}}\left(\{\log p\}^{-1}\right).
\end{aligned} \tag{42}
$$

Note $\mathbb{E}[v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}] = 0$ and $\mathbb{E}[v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}]^2 = F_{rt,1}/4$ by definition of the score vector $v_{rt,1}^{(i),o}$ and $F_{rt,1}$. Therefore by Hoeffding's inequality, there exists $C_1 > 0$ such that for any $\xi > 0$

$$\mathrm{pr}\left(\left|n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}\right| \geq C_1\sqrt{\log p/n_1}\right) \leq O(p^{-\xi}).$$

Similarly there exists $C_2 > 0$ such that

$$\mathrm{pr}\left(\left|n_1^{-1}\sum_{i=1}^{n_1} \varepsilon_{r,1}^{(i)}\right| \geq C_2\sqrt{\log p/n_1}\right) \leq O(p^{-\xi}).$$

Hence we have $|n_1^{-1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\varepsilon_{r,1}^{(i)}| = O_{\mathrm{p}}(\sqrt{\log p/n_1})$, and

$$|a_1| = O_{\mathrm{p}}\left(\sqrt{\frac{\log p}{n_1}}\right)\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)} - F_{rt,1}/2\right| = o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right).$$

For term a_2, noting that $F_{rt,1} > c^* > 0$ by Assumption (A4), we must have

$$\begin{aligned}|a_2| =& \frac{2}{F_{rt,1}}\left|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{rt,1}^{(i)} - v_{rt,1}^{(i),o})\varepsilon_{r,1}^{(i)}\right| \leq \max_{r,t}|v_{rt,1}^{(i)} - v_{rt,1}^{(i),o}|\frac{2}{F_{rt,1}}\left|\frac{1}{n_1}\sum_{i=1}^{n_1}\varepsilon_{r,1}^{(i)}\right| \\ =& o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right).\end{aligned}$$

Finally, to bound a_3, since its denominator is of constant order by (42), it suffices to bound the numerators REM_1 and REM_2 separately. To bound REM_2, by (41), (38), and equation (68) in Ren et al. (2016), we have

$$\begin{aligned}\left|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{rt,1}^{(i)} - v_{rt,1}^{(i),o})\ddot{f}(\hat{u}_{r,1}^{(i)})h_{(i),1}\right| &\leq \frac{C}{n_1}\sum_{i=1}^{n_1}\left(|h_{(i),1}| + |h_{(i),2}| + |\Delta_{\{r,t\},1}|\right)\ddot{f}(\hat{u}_{r,1}^{(i)})|h_{(i),1}| \\ &= o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right).\end{aligned}$$

Thus by (33),

$$\begin{aligned}|REM_2| &\leq o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right) + \left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{rt,1}^{(i),o}\ddot{f}(\hat{u}_{r,1}^{(i)})h_{(i),1}\right| \\ &\leq o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right) + \sqrt{\log p/n_1}\|\Delta_{r,1}\|_1 = o_{\mathrm{p}}\left(\{n_1\log p\}^{-1/2}\right).\end{aligned}$$

For REM_1, by equations (37), (38) and (39) and the fact that $\max_i |v_{rt,1}^{(i)}| \leq 1$, we have

$$
\begin{aligned}
|REM_1| =& \left| \sum_{i=1}^{n_1} v_{rt,1}^{(i)} \ddot{f}(\hat{u}_{r,1}^{(i)}) h_{(i)} - \frac{1}{n_1} \sum_{i=1}^{n_1} \int_0^1 v_{rt,1}^{(i)} \cdot \ddot{f}(\hat{u}_{r,1}^{(i)} - z h_{\{i\}}) h_{\{i\}} dz \right| \\
=& \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \int_0^1 v_{rt,1}^{(i)} \cdot \left(\ddot{f}(u_{r,1}^{(i)} + h_{\{i\}}) - \ddot{f}(u_{r,1}^{(i)} + z h_{\{i\}}) \right) h_{\{i\}} dz \right| \\
\leq& \frac{1}{n_1} \sum_{i=1}^{n_1} |v_{rt,1}^{(i)}| \cdot \ddot{f}(u_{r,1}^{(i)}) |\exp(2h_{\{i\}}) - 1| \cdot |h_{\{i\}}| \\
\leq& \frac{C}{n_1} \sum_{i=1}^{n_1} |v_{rt,1}^{(i)}| \cdot \ddot{f}(u_{r,1}^{(i)}) |h_{\{i\}}^2| = o_{\mathrm{p}} \left(\{n_1 \log p\}^{-1/2} \right).
\end{aligned}
$$

It follows immediately that $|a_3| = o_{\mathrm{p}} \left(\{n_1 \log p\}^{-1/2} \right)$.

In summary, we have shown that

$$
|\tilde{\theta}_{rt,1} - \check{\theta}_{rt,1}| = o_{\mathrm{p}} \left(\{n_1 \log p\}^{-1/2} \right).
$$

□

The following lemma states the large deviation bound for the variances of $\check{\theta}_{rt,k}$ for all $1 \leq r < t \leq p$ and $k = 1, 2$.

Lemma A.3. *Under the conditions of Theorem 2.1, we have*

$$
\max_{1 \leq r < t \leq p} |\check{s}_{rt,k} - s_{rt,k}| = o_{\mathrm{p}}(\{\log p\}^{-1}). \tag{43}
$$

Proof. We only need to prove it for $k = 1$ since the derivation for $k = 2$ is similar. By definition,

$$
\check{s}_{rt,1} = \left\{ 4 n_1^{-1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i)})^2 \ddot{f}\big(\hat{u}_{r,1}^{(i)}\big) \right\}^{-1}.
$$

It is easy to check that $4n_1^{-1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i)})^2 \ddot{f}\big(\hat{u}_{r,1}^{(i)}\big)$ is bounded. We first look at the difference

$$
\frac{1}{n_1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i)})^2 \ddot{f}\big(X_{-r}^{(i)} \hat{\theta}_{r,1}\big) - F_{rt,1}/4.
$$

Following the proof of Lemma A.2, one can show similarly that

$$
\begin{aligned}
\left| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i)})^2 \ddot{f}\big(\hat{u}_{r,1}\big) - F_{rt,1}/4 \right| \leq& \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \big\{ (v_{rt,1}^{(i)})^2 - (v_{rt,1}^{(i),o})^2 \big\} \ddot{f}\big(\hat{u}_{r,1}\big) \right| + \\
& \left| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i),o})^2 \left\{ \ddot{f}\big(\hat{u}_{r,1}\big) - \ddot{f}\big(u_{r,1}\big) \right\} \right| + \\
& \left| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{rt,1}^{(i),o})^2 \ddot{f}\big(u_{r,1}\big) - F_{rt,1}/4 \right| = b_1 + b_2 + b_3,
\end{aligned}
$$

where $b_1 = o_{\mathrm{p}}(\{\log p\}^{-1}), b_2 = o_{\mathrm{p}}(\{\log p\}^{-1})$, and $b_3 = O_{\mathrm{p}}(\{\log p/n_k\}^{1/2}) = o_{\mathrm{p}}(\{\log p\}^{-1})$. We thus obtain

$$\Big|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{rt,1}^{(i)})^2\ddot{f}(\hat{u}_{r,1}) - F_{rt,1}/4\Big| = o_{\mathrm{p}}(\{\log p\}^{-1}). \tag{44}$$

The result in (43) follows immediately as the denominator of $\check{s}_{rt,1}$ is bounded below by a constant. This completes the proof.

□

The following lemma, adapted from Lemma 4 in Cai et al. (2013), is key to our proof.

Lemma A.4. *Under the conditions of Theorem 2.1, we have some constant $C > 0$ such that for any $\epsilon > 0$*

$$\mathrm{pr}\left(\max_{(r,t)\in\Lambda}\frac{(\tilde{\theta}_{rt,1} - \tilde{\theta}_{rt,2} - \theta_{rt,1} + \theta_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2} \geq x^2\right) \leq C|\Lambda|\{1 - \Phi(x)\} + O(p^{-\epsilon}),$$

uniformly for $0 \leq x \leq (8\log p)^{1/2}$ and $\Lambda \subset \{(r,t) : 1 \leq r < t \leq p\}$.

Proof. Let

$$Z_{rt,i} = \frac{n_2}{n_1}\frac{2v_{rt,1}^{(i),o}\varepsilon_{r,1}^{(i)}}{F_{rt,1}}, \quad 1 \leq i \leq n_1,$$
$$Z_{rt,i} = -\frac{2v_{rt,2}^{(i),o}\varepsilon_{r,2}^{(i)}}{F_{rt,2}}, \quad n_1 + 1 \leq i \leq n_1 + n_2,$$

Then

$$\frac{(\tilde{\theta}_{rt,1} - \tilde{\theta}_{rt,2} - \theta_{rt,1} + \theta_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2} = \frac{\left(\sum_{i=1}^{n_1+n_2} Z_{rt,i}\right)^2}{\sum_{i=1}^{n_1+n_2} Z_{rt,i}^2}\frac{\sum_{i=1}^{n_1+n_2} Z_{rt,i}^2}{n_2^2 s_{rt,1}/n_1 + n_2 s_{rt,2}}. \tag{45}$$

It is easy to check that $n_1^{-1}\sum_{i=1}^{n_1} Z_{rt,i}^2 \to (n_2/n_1)^2 s_{rt,1}$ and $n_2^{-1}\sum_{i=1}^{n_2} Z_{rt,i}^2 \to s_{rt,2}$. In particular, since $Z_{rt,i}^2$ are bounded for all i, we apply Hotelling's inequality to obtain

$$\mathrm{pr}\left(\frac{1}{n_1}\sum_{i=1}^{n_1} Z_{rt,i}^2 - \frac{n_2^2}{n_1^2}s_{rt,1} \geq C\sqrt{\frac{\log p}{n_1}}\right) \leq p^{-2C^2}, \tag{46}$$

$$\mathrm{pr}\left(\frac{1}{n_2}\sum_{i=1+n_1}^{n_1+n_2} Z_{rt,i}^2 - s_{rt,2} \geq C\sqrt{\frac{\log p}{n_2}}\right) \leq p^{-2C^2}. \tag{47}$$

Thus the second term on the RHS of (45) can be treated as a constant, up to a small deviation in the order of $\sqrt{\log p/n_k}$.

It suffices to check the large deviation bound of the term

$$\frac{\left(\sum_{i=1}^{n_1+n_2} Z_{rt,i}\right)^2}{\sum_{i=1}^{n_1+n_2} Z_{rt,i}^2}.$$

Since $Z_{rt,i}$ $(i = 1, \ldots, n_1 + n_2)$ are all bounded and thus have finite $(2+\delta)$-th moment $(0 < \delta \leq 1)$, we can apply the self-normalized large deviation theorem for independent random variables in Jing et al.

(2003) to obtain

$$\max_{1\le r<t\le p} \operatorname{pr}\left(\frac{\left(\sum_{i=1}^{n_1+n_2} Z_{rt,i}\right)^2}{\sum_{i=1}^{n_1+n_2} Z_{rt,i}^2} \ge x^2\right) \le C\{1-\Phi(x)\} \tag{48}$$

uniformly for $0 \le x \le (8\log p)^{1/2}$. □

A.2 Global Testing

Proof of Theorem 2.1. Recall the data generating model in (6) assumes that

$$X_r = \mathbb{E}_\theta[X_r \mid X_{-r}] + \varepsilon_{r,1} = \dot f(X_{-r}\theta_r) + \varepsilon_{r,1}, \tag{49}$$

where

$$\dot f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} = \frac{\sinh(u)}{\cosh(u)}.$$

The variance of X_r given X_{-r} is

$$\eta_r(X,\theta_r) = \frac{4\exp(2X_{-r}\theta_r)}{\{\exp(2X_{-r}\theta_r)+1\}^2} = \ddot f(X_{-r}\theta_r). \tag{50}$$

Recall the test statistic is

$$M_{n,p} = \max_{r<t} W_{rt}^2 = \max_{1\le r<t\le p} \frac{(\check\theta_{rt,1}-\check\theta_{rt,2})^2}{\check s_{rt,1}/n_1 + \check s_{rt,2}/n_2}. \tag{51}$$

Define

$$\widehat M_{n,p} = \max_{1\le r<t\le p} \widehat W_{rt}^2 = \max_{1\le r<t\le p} \frac{(\check\theta_{rt,1}-\check\theta_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2},$$
$$\widetilde M_{n,p} = \max_{1\le r<t\le p} \widetilde W_{rt}^2 = \max_{1\le r<t\le p} \frac{(\tilde\theta_{rt,1}-\tilde\theta_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2}.$$

From $M_{n,p}$ to $\widehat M_{n,p}$, we have replaced the estimated variances $\check s_{rt,k}$ in $M_{n,p}$ by their population counterparts $s_{rt,k} = 1/F_{rt,k}$. From $\widehat M_{n,p}$ to $\widetilde M_{n,p}$, we have replaced $\check\theta_{rt,k}$ by $\tilde\theta_{rt,k}$ since $\tilde\theta_{rt,k}$ is a good approximation to $\check\theta_{rt,k}$. Under the null, let $\theta_{rt} = \theta_{rt,1} = \theta_{rt,2}$. Define

$$\mathcal{A} = \{(r,t) : 1 \le r < t \le p\},$$
$$\mathcal{A}_t(\xi) = \{r : |\sinh(2\theta_{rt})| \ge (\log p)^{-2-\xi}\}.$$
$$\mathcal{D}_0 = \{(r,t) : 1 \le t \le p, r \in \mathcal{A}_t(\xi)\}.$$

To prove Theorem 2.1, we first show that (1) $M_{n,p}$ is close to $\widetilde M_{n,p}$, and (2) terms in $\mathcal{D}_0$ are negligible for defining the limiting distribution of $\widetilde M_{n,p}$. Consequently it suffices to study the limiting distribution of $\widetilde{M}_{n,p}$ on $\mathcal{A}\backslash\mathcal{D}_0$. Note

$$\begin{aligned}W_{rt}-\widetilde{W}_{rt} =&W_{rt}-\widehat{W}_{rt}+\widehat{W}_{rt}-\widetilde{W}_{rt}\\ =&W_{rt}\cdot o_{\mathrm{p}}(\{\log p\}^{-1})+\frac{(\check{\theta}_{rt,1}-\check{\theta}_{rt,2})-(\tilde{\theta}_{rt,1}-\tilde{\theta}_{rt,2})}{\sqrt{s_{rt,1}/n_1+s_{rt,2}/n_2}}\\ =&W_{rt}\cdot o_{\mathrm{p}}(\{\log p\}^{-1})+o_{\mathrm{p}}(\{\log p\}^{-1/2})=o_{\mathrm{p}}(\{\log p\}^{-1/2}),\end{aligned}$$

where we have used Lemma A.2, Lemma A.3 and the fact that $s_{rt,k}=1/F_{rt,k}\geq 1/4$. Thus it suffices to show

$$\mathrm{pr}(\widetilde{M}_{n,p}-4\log p+\log\log p\leq z)\to\exp\{-(8\pi)^{-1/2}e^{-z/2}\},\quad n_1,n_2,p\to\infty.$$

Set $y_p=z+4\log p-\log\log p$ and denote by

$$\widetilde{M}_{\mathcal{D}_0}=\max_{(r,t)\in\mathcal{D}_0}\frac{(\tilde{\theta}_{rt,1}-\tilde{\theta}_{rt,2})^2}{s_{rt,1}/n_1+s_{rt,2}/n_2},\quad \widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0}=\max_{(r,t)\in\mathcal{A}\backslash\mathcal{D}_0}\frac{(\tilde{\theta}_{rt,1}-\tilde{\theta}_{rt,2})^2}{s_{rt,1}/n_1+s_{rt,2}/n_2}.$$

Then $\mathrm{pr}(\widetilde{M}_{n,p}\geq y_p)-\mathrm{pr}(\widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0}\geq y_p)\leq\mathrm{pr}(\widetilde{M}_{\mathcal{D}_0}\geq y_p)$. Note Assumption (A2) indicates that $|\mathcal{D}_0|=o(p)$. Thus by Lemma A.4, for any fixed $x\in\mathbb{R}$, we must have

$$\mathrm{pr}(\widetilde{M}_{\mathcal{D}_0}\geq y_p)\leq|\mathcal{D}_0|\cdot Cp^{-2}+o(1)=o(1).$$

Hence it suffices to show that

$$\mathrm{pr}(\widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0}-4\log p+\log\log p\leq z)\to\exp\{-(8\pi)^{-1/2}e^{-z/2}\},\quad n_1,n_2,p\to\infty.$$

Now let us rearrange the two-dimensional indices in the set $\{(r,t):(r,t)\in\mathcal{A}\backslash\mathcal{D}_0\}$ as $\{(r_j,t_j):1\leq j\leq q\}$, where $q=|\mathcal{A}\backslash\mathcal{D}_0|$. We show that for suitably defined random variables with index spanning over $\mathcal{A}\backslash\mathcal{D}_0$ will satisfy (58). For $1\leq j\leq q$, let

$$\begin{aligned}Z_{ji}&=\frac{n_2}{n_1}\frac{2v^{(i),o}_{r_jt_j,1}\varepsilon^{(i)}_{r_j,1}}{F_{r_jt_j,1}}=\frac{n_2}{n_1}\Big(\frac{2v^{(i),o}_{r_jt_j,1}\varepsilon^{(i)}_{r_j,1}}{F_{r_jt_j,1}}+\theta_{rt,1}-\theta_{rt,1}\Big),\quad 1\leq i\leq n_1,\\ Z_{ji}&=-\frac{2v^{(i),o}_{r_jt_j,2}\varepsilon^{(i)}_{r_j,2}}{F_{r_jt_j,2}}=-\Big(\frac{2v^{(i),o}_{r_jt_j,2}\varepsilon^{(i)}_{r_j,2}}{F_{r_jt_j,2}}+\theta_{rt,2}-\theta_{rt,2}\Big),\quad n_1+1\leq i\leq n_1+n_2,\end{aligned}$$

Note Z_{ji}'s are all bounded and thus there is no need to truncate them. Let

$$V_j=\frac{1}{\sqrt{n_2^2/n_1s_{j1}+n_2s_{j2}}}\sum_{i=1}^{n_1+n_2}Z_{ji},\quad 1\leq j\leq q.$$

It suffices to show that for any $z\in\mathbb{R}$, as $n,p\to\infty$,

$$\mathrm{pr}\left(\max_{1\leq j\leq q}V_j^2-4\log p+\log\log p\leq z\right)\to\exp\left\{-\frac{1}{\sqrt{8\pi}}\exp(-\frac{z}{2})\right\}. \tag{52}$$

Let $\mathcal{E}_{j_g} = \{V_{j_g}^2 \geq y_p\}$. By Bonferroni inequality (Cai et al., 2013, Lemma 1), for any integer m with $0 < m < q/2$,

$$\sum_{\ell=1}^{2m}(-1)^{\ell-1}\sum_{1\leq j_1<\ldots<j_\ell\leq q}\text{pr}\left(\bigcap_{g=1}^{\ell}\mathcal{E}_{j_g}\right) \leq \text{pr}\left(\max_{1\leq j\leq q} V_j^2 \geq y_p\right)$$
$$\leq \sum_{\ell=1}^{2m-1}(-1)^{\ell-1}\sum_{1\leq j_1<\ldots<j_\ell\leq q}\text{pr}\left(\bigcap_{g=1}^{\ell}\mathcal{E}_{j_g}\right). \tag{53}$$

Let

$$\tilde{Z}_{ji} = Z_{ji}/(n_2 s_{j1}/n_1 + s_{j2})^{1/2}, \quad 1 \leq j \leq q$$

and $\mathbf{W}_i = (\tilde{Z}_{j_1 i}, \ldots, \tilde{Z}_{j_\ell i})$ for all $1 \leq i \leq n_1 + n_2$. Denote $|\boldsymbol{a}|_{\min} = \min_{1\leq i\leq \ell}|a_i|$ for any vector $\boldsymbol{a} \in \mathbb{R}^\ell$. Then we have

$$\text{pr}\left(\bigcap_{g=1}^{\ell}\mathcal{E}_{j_g}\right) = \text{pr}\left(\left|\frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_1+n_2}\mathbf{W}_i\right|_{\min} \geq {y_p}^{1/2}\right).$$

By Theorem 1 in Zaitsev (1987), we have

$$\text{pr}\left(\left|\frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_1+n_2}\mathbf{W}_i\right|_{\min} \geq {y_p}^{1/2}\right) \leq \text{pr}\left(|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n(\log p)^{-1/2}\right)$$
$$+ c_1\ell^{5/2}\exp\left(-\frac{n_2^{1/2}\epsilon_n}{c_2\ell^{5/2}\tau_n(\log p)^{1/2}}\right). \tag{54}$$

where $c_1, c_2 > 0$ are absolute constants, $\epsilon_n \to 0$ which will be specified later, $\tau_n > 0$ are constants, and $N_\ell := (N_{j_1}, \ldots, N_{j_\ell})$ is an ℓ-dimensional normal vector with mean $\mathbf{0}$ and covariance $n_1\text{Cov}(\mathbf{W}_1)/n_2 + \text{Cov}(\mathbf{W}_{n_1+1})$. Recall that ℓ is a fixed integer that is independent of n_k and p. Since $\log p = o(n_k^{1/5})$ by Assumption (A1), we can choose $\epsilon_n \to 0$ sufficiently slow and

$$c_1\ell^{5/2}\exp\left(-\frac{n_2^{1/2}\epsilon_n}{c_2\ell^{5/2}\tau_n(\log p)^{1/2}}\right) \leq c_1\ell^{5/2}\exp\left(-\frac{\epsilon_n\log^3 p}{c_2\ell^{5/2}\tau_n}\right) = o(p^{-2}). \tag{55}$$

Thus by (53), (54), (55),

$$\text{pr}\left(\max_{1\leq j\leq q} V_j^2 \geq y_p\right) \leq \sum_{\ell=1}^{2m-1}(-1)^{\ell-1}\sum_{1\leq j_1<\ldots<j_\ell\leq q}\text{pr}\left(|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n(\log p)^{-1/2}\right) + o(1). \tag{56}$$

Similarly, by Theorem 1 in Zaitsev (1987), one can also get

$$\text{pr}\left(\max_{1\leq j\leq q} V_j^2 \geq y_p\right) \geq \sum_{\ell=1}^{2m}(-1)^{\ell-1}\sum_{1\leq j_1<\ldots<j_\ell\leq q}\text{pr}\left(|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n(\log p)^{-1/2}\right) + o(1). \tag{57}$$

To complete our proof, the next lemma (Cai et al., 2013, Lemma 5) is crucial.

Lemma A.5. *For any fixed integer $\ell \geq 1$ and real number $x \in \mathbb{R}$,*

$$\sum_{1 \leq j_1 < \cdots < j_\ell \leq q} \mathrm{pr}\left(|N_\ell|_{\min} \geq y_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\right) = \frac{1}{\ell!}\left\{\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{x}{2}\right)\right\}^\ell \{1 + o(1)\}. \tag{58}$$

Under the conditions of Theorem 2.1,

$$|\mathrm{Cov}(v^{(i),o}_{r_j t_j,k} \varepsilon^{(i)}_{r_j,k}, v^{(i),o}_{r_{j'} t_{j'},k} \varepsilon^{(i)}_{r_{j'},k})| \lesssim \{|\sinh(2\theta_{r_j r_{j'},k})| + |\sinh(2\theta_{r_j t_{j'},k})|\}\{|\sinh(2\theta_{t_j t_{j'},k})| + |\sinh(2\theta_{t_j r_{j'},k})|\},$$

for $1 \leq j \neq j' \leq q$ and $k = 1, 2$. Following the proof in the supplementary materials of Cai et al. (2013), one can show that Lemma A.5 also holds in the current context. Now applying Lemma A.5 to (56) and (57), one has for any positive integer $m > 0$

$$\limsup_{n_k, p \to \infty} \mathrm{pr}\left(\max_{1 \leq j \leq q} V_j^2 \geq y_p\right) \leq \sum_{\ell=1}^{2m-1} (-1)^{\ell-1} \frac{1}{\ell!}\left\{\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{z}{2}\right)\right\}^\ell \{1 + o(1)\},$$

$$\limsup_{n_k, p \to \infty} \mathrm{pr}\left(\max_{1 \leq j \leq q} V_j^2 \geq y_p\right) \geq \sum_{\ell=1}^{2m} (-1)^{\ell-1} \frac{1}{\ell!}\left\{\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{z}{2}\right)\right\}^\ell \{1 + o(1)\}.$$

Letting $m \to \infty$, we obtain the desired convergence in (52). □

Proof of Theorem 2.2. Define

$$M_n^1 = \max_{1 \leq r < t \leq p} \frac{(\check{\theta}_{rt,1} - \check{\theta}_{rt,2} - \theta_{rt,1} + \theta_{rt,2})^2}{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2}.$$

By Lemma (A.2), (A.3), (A.4) and the tail bound for standard normal distribution,

$$\mathrm{pr}\Big(M_n^1 \leq 4 \log p - \frac{1}{2} \log\log p\Big) \to 1, \quad n_1, n_2, p \to \infty.$$

By Lemma (A.3), the inequalities

$$\max_{1 \leq r < t \leq p} \frac{(\theta_{rt,1} + \theta_{rt,2})^2}{\check{s}_{rt,1}/n_1 + \check{s}_{rt,2}/n_2} \leq 2M_n^1 + 2M_{n,p},$$

and

$$\max_{1 \leq r < t \leq p} \frac{(\theta_{rt,1} + \theta_{rt,2})^2}{s_{rt,1}/n_1 + s_{rt,2}/n_2} \geq 16 \log p,$$

we immediately obtain that $\mathrm{pr}(M_{n,p} \geq q_\alpha + 4 \log p - \log\log p) \to 1$ as $_1, n_2, p \to \infty$. This completes the proof. □

Proof of Theorem 2.3. To prove the lower bound result, we first construct the worst case scenario to test between Θ_1 and Θ_2.

Let $\mathcal{M}(a_p, p-1)$ be the set of all subsets of $\{1, \ldots, p-1\}$ with cardinality $a_p = p^\gamma$ for $\gamma < 1/2$. Let $\hat{m}$ be a random variable uniformly distributed over $\mathcal{M}(a_p, p-1)$. We construct a class $\mathcal{N} = \{\Theta_{\hat{m}}, \hat{m} \in \mathcal{M}(a_p, p-1)\}$, such that $\theta_{rt} = 0$ for $|r - t| > 1$, $\theta_{r,r+1} = \theta_{r+1,r} = \rho \cdot I(r \in \hat{m})$ and $\rho = c(\log p/n)^{1/2}$. Let

Θ_1 be uniformly distributed over $\mathcal{N}$ and $\Theta_2 = \mathbf{0}$ be the zero matrix. Denote by μ_ρ the distribution of $\Theta_1 - \Theta_2 = \Theta_1$. Note that μ_ρ is a probability measure on $\{\Delta \in \mathcal{S}(2a_p) : \|\Delta\|_F^2 = 2a_p\rho^2\}$, where $\mathcal{S}(2a_p)$ is the class of matrices with $2a_p$ nonzero entries. Let $d\mathrm{pr}_1(\{X^{(1)}, \ldots, X^{(n)}\})$ and $d\mathrm{pr}_2(\{Y^{(1)}, \ldots, Y^{(n)}\})$ be the likelihood functions with partial correlation matrices Θ_1 and Θ_2, respectively. Then we have

$$L_{\mu_\rho} = \mathbb{E}_{\mu_\rho}\left[\frac{d\mathrm{pr}_1(\{X^{(1)}, \ldots, X^{(n)}\})}{d\mathrm{pr}_2(\{Y^{(1)}, \ldots, Y^{(n)}\})}\right],$$

where $\mathbb{E}_{\mu_\rho}[\cdot]$ is the expectation over the distribution of Θ_1. By the arguments in Baraud (2002), it suffices to show that $\mathbb{E}[L_{\mu_\rho}^2] \leq 1 + o(1)$.

It is easy to check that

$$L_{\mu_\rho} = \mathbb{E}_{Y,\hat{m}}\left[\prod_{i=1}^n \frac{2^p}{Z_{\hat{m}}(\Theta_{\hat{m}})} \exp\left(\rho \sum_{r\in\hat{m}} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right],$$

where $Y_1^{(i)}, \ldots, Y_p^{(i)}$ are independent Rademacher random variables (i.e. for each r, the $Y_r^{(i)}$'s are i.i.d. random variables taking values ± 1 with probability 0.5, corresponding to a zero partial correlation matrix) and $Z_{\hat{m}}(\Theta_{\hat{m}})$ is the normalizing constant corresponding to $\mathrm{P}_{\Theta_{\hat{m}}}$. Since $\Theta_{\hat{m}} \in \mathcal{N}$ and $|\hat{m}| = a_p$, we can express $Z_{\hat{m}}(\Theta_{\hat{m}})$ analytically as

$$\sum_{X_1,\ldots,X_p} \exp\left(\rho \sum_{r\in\hat{m}} X_r X_{r+1}\right) = 2^p\{\cosh(\rho)\}^{a_p}.$$

Consequently, one can rewrite L_{μ_ρ} as

$$\begin{aligned} L_{\mu_\rho} =& \mathbb{E}_{Y,\hat{m}}\left[\prod_{i=1}^n \frac{2^p}{2^p\{\cosh(\rho)\}^{a_p}} \exp\left(\rho \sum_{r\in\hat{m}} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right] \\ =& \frac{1}{C_{p-1}^{a_p}} \sum_{m\in\mathcal{M}(a_p,p-1)} \frac{1}{\{\cosh(\rho)\}^{na_p}} \mathbb{E}_Y\left[\prod_{i=1}^n \exp\left(\rho \sum_{r\in m} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right], \end{aligned}$$

and therefore

$$\begin{aligned} \mathbb{E}[L_{\mu_\rho}^2] =& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \mathbb{E}_Y\left[\prod_{i=1}^n \exp\left(\rho \sum_{r\in m} Y_r^{(i)} Y_{r+1}^{(i)} + \rho \sum_{r\in m'} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right] \\ =& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \mathbb{E}_Y\left[\prod_{i=1}^n \exp\left(2\rho \sum_{r\in m\cap m'} Y_r^{(i)} Y_{r+1}^{(i)} + \rho \sum_{r\in m\Delta m'} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right] \\ =& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \times \\ & \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \prod_{i=1}^n \mathbb{E}_Y\left[\exp\left(2\rho \sum_{r\in m\cap m'} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right] \mathbb{E}_Y\left[\exp\left(\rho \sum_{r\in m\Delta m'} Y_r^{(i)} Y_{r+1}^{(i)}\right)\right], \end{aligned}$$

by mutual independence of Y_r's, where $m\Delta m' = (m\backslash m') \cup (m'\backslash m)$. Further, it can be shown that

$$\mathbb{E}_Y\Big[\exp\Big(\rho \sum_{r\in m} Y_r^{(i)} Y_{r+1}^{(i)}\Big)\Big] = \cosh^m(\rho).$$

Hence

$$\begin{aligned}
\mathbb{E}[L_{\mu_\rho}^2] =& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \prod_{i=1}^n \Big\{\cosh(2\rho)\Big\}^{|m\cap m'|} \Big\{\cosh(\rho)\Big\}^{|m\Delta m'|} \\
=& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \Big\{\cosh(2\rho)\Big\}^{n|m\cap m'|} \Big\{\cosh(\rho)\Big\}^{n|m\Delta m'|} \\
=& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \frac{1}{\{\cosh(\rho)\}^{2na_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \left\{\frac{\cosh(2\rho)}{\cosh^2(\rho)}\right\}^{n|m\cap m'|} \Big\{\cosh(\rho)\Big\}^{2n|m\cap m'|+n|m\Delta m'|} \\
=& \frac{1}{C_{p-1}^{a_p} C_{p-1}^{a_p}} \sum_{m,m'\in\mathcal{M}(a_p,p-1)} \left\{\frac{2\cosh(2\rho)}{1+\cosh(2\rho)}\right\}^{n|m\cap m'|},
\end{aligned}$$

where the last equation follows from the fact that $|m \cap m'| + |m\Delta m'|/2 = a_p$. In view of the random variables $m, m' \in \mathcal{M}(a_p, p-1)$, we can further write

$$\begin{aligned}
\mathbb{E}[L_{\mu_\rho}^2] =& \binom{p-1}{a_p}^{-2} \sum_{j=1}^{a_p} \binom{a_p}{j} \binom{p-1-a_p}{a_p-j} \left\{\frac{2\cosh(2\rho)}{1+\cosh(2\rho)}\right\}^{nj} \\
=& \{1+o(1)\} \left(1 + \frac{a_p}{p-1} \left\{\frac{2\cosh(2\rho)}{1+\cosh(2\rho)}\right\}^n\right)^{a_p}.
\end{aligned}$$

Noting that $\rho = c\sqrt{\log p/n}$, the term $\frac{2\cosh(2\rho)}{1+\cosh(2\rho)} \sim 1 + 2\rho^2$ by local Taylor approximation. Therefore

$$\begin{aligned}
\mathbb{E}[L_{\mu_\rho}^2] \le& \{1+o(1)\} \exp\left(a_p \log\left\{1 + \frac{a_p}{p-1}(1+2\rho^2)^n\right\}\right) \\
\le& \{1+o(1)\} \exp\left(a_p^2 p^{2c^2-1}\right) = 1 + o(1),
\end{aligned}$$

for sufficiently small $c > 0$, where the last inequality follows from local Taylor approximation $\log(1+x) \sim x$ for x in a small neighborhood of $x_0 = 0$.

□

A.3 Multiple Testing

Proof of Theorem 3.1. We first show that $\hat{\tau}$ defined in Algorithm 2 must lie in $[0, b_p]$ where $b_p = \sqrt{4\log p - 2\log(\log p)}$. Under the conditions of Theorem 3.1, we have with high probability

$$\sum_{1\le r<t\le p} I(|W_{rt}| \ge 2\sqrt{\log p}) \ge \left\{\frac{1}{(8\pi)^{1/2}\alpha} + \zeta\right\} (\log\log p)^{1/2}.$$

Therefore with high probability,

$$\frac{(p^2-p)/2}{\max\{\sum_{1\le r<t\le p} I(|W_{rt}|\ge 2\sqrt{\log p}),1\}} \le \frac{p^2-p}{2}\left\{\frac{1}{(8\pi)^{1/2}\alpha}+\zeta\right\}^{-1}(\log\log p)^{-1/2}.$$

By the tail bound of standard normal random variable, $1-\Phi(b_p)\sim(\sqrt{2\pi}b_p)^{-1}\exp(-b_p^2/2)$, and we thus must have $\mathrm{pr}(0\le\hat\tau\le b_p)\to 1$ by definition of $\hat\tau$ in Algorithm 2.

Denote by $G(\tau)=2\{1-\Phi(\tau)\}$. For $\hat\tau\in[0,b_p]$ obtained in Algorithm 2, it holds that

$$\frac{G(\hat\tau)(p^2-p)/2}{\max\{\sum_{1\le r<t\le p} I(|W_{rt}|\ge\hat\tau),1\}}\le\alpha.$$

One key step in Algorithm 2 is to approximate the number of true nulls $R_0(\tau)$ by $q_0G(\tau)$ for $\tau\in[0,b_p]$. Thus to prove Theorem 3.1, it suffices to show that

$$\left|\frac{R_0(\tau)}{q_0G(\tau)}-1\right|=\left|\frac{\sum_{(r,t)\in\mathcal{H}_0}\{I(|W_{rt}|\ge\tau)-G(\tau)\}}{q_0G(\tau)}\right|\to 0,\quad 0\le\tau\le b_p,$$

in probability. Recall that $W_{rt}-\widetilde{W}_{rt}=o_{\mathrm{p}}(\{\log p\}^{-1/2})$ for all $r<t$. Thus the problem boils down to showing that

$$\frac{\left|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau)-G(\tau)\}\right|}{q_0G(\tau)}\to 0, \tag{59}$$

in probability.

Let $0=\tau_0<\tau_1<\cdots<\tau_m=b_p$ such that $\tau_\ell-\tau_{\ell-1}=v_p$ for $\ell=1,\ldots,m-1$ and $\tau_m-\tau_{m-1}\le v_p$, where $v_p=(c_p^2\log p)^{-1/2}$ and $c_p=\log\log\log\log p$. Thus we have $m\sim b_p/v_p$. For any τ such that $\tau_{\ell-1}\le\tau\le\tau_\ell$, it holds that

$$\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau_\ell)}{q_0G(\tau_{\ell-1})}\le\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau)}{q_0G(\tau)}\le\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau_{\ell-1})}{q_0G(\tau_\ell)},$$

or equivalently

$$\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau_\ell)}{q_0G(\tau_\ell)}\frac{G(\tau_\ell)}{G(\tau_{\ell-1})}\le\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau)}{q_0G(\tau)}\le\frac{\sum_{(r,t)\in\mathcal{H}_0} I(|\widetilde{W}_{rt}|\ge\tau_{\ell-1})}{q_0G(\tau_{\ell-1})}\frac{G(\tau_{\ell-1})}{G(\tau_\ell)}.$$

Therefore it suffices to show that $\max_{0\le\ell\le m}|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-G(\tau_\ell)\}|/\{q_0G(\tau_\ell)\}\to 0$ in probability. Note that

$$\begin{aligned}
&\mathrm{pr}\left(\max_{0\le\ell\le m}\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-G(\tau_\ell)\}|}{q_0G(\tau_\ell)}\ge\epsilon\right)\\
&\le\sum_{\ell=0}^{m}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-G(\tau_\ell)\}|}{q_0G(\tau_\ell)}\ge\epsilon\right)\\
&\le\frac{1}{v_p}\int_0^{b_p}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau)-G(\tau)\}|}{q_0G(\tau)}\ge\epsilon\right)d\tau+\\
&\sum_{\ell=m-1}^{m}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-G(\tau_\ell)\}|}{q_0G(\tau_\ell)}\ge\epsilon\right).
\end{aligned}$$

Hence one only needs to prove that

$$\int_0^{b_p}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau)-G(\tau)\}|}{q_0G(\tau)}\ge\epsilon\right)d\tau=o(v_p),$$

and

$$\sum_{\ell=m-1}^{m}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-G(\tau_\ell)\}|}{q_0G(\tau_\ell)}\ge\epsilon\right)=o(1).$$

The conclusion in (59) follows immediately. Further, by (27) and the fact that $W_{rt}-\widetilde{W}_{rt}=o_{\mathrm{p}}(\{\log p\}^{-1/2})$ for all $r<t$, it suffices to show

$$\int_0^{b_p}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau)-\mathrm{pr}(|\widetilde{W}_{rt}|\ge\tau)\}|}{q_0G(\tau)}\ge\epsilon\right)d\tau=o(v_p),\tag{60}$$

and

$$\sum_{\ell=m-1}^{m}\mathrm{pr}\left(\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{rt}|\ge\tau_\ell)-\mathrm{pr}(|\widetilde{W}_{rt}|\ge\tau)\}|}{q_0G(\tau_\ell)}\ge\epsilon\right)=o(1).\tag{61}$$

To this end, define

$$\begin{aligned}
\mathcal{A}_t(\xi)&=\{r:|\sinh(2\theta_{rt,1})|+|\sinh(2\theta_{rt,2})|\ge2(\log p)^{-2-\xi}\},\\
\mathcal{D}_0&=\{(r,t):1\le t\le p,r\in\mathcal{A}_t(\xi)\},\\
\mathcal{H}_{01}&=\mathcal{H}_0\cap\mathcal{D}_0,\quad \mathcal{H}_{02}=\mathcal{H}_0\cap\mathcal{D}_0^c.
\end{aligned}$$

By Assumption (A2), we have $|\mathcal{H}_{01}|=o(p^{1+\gamma})$ for $0<\gamma<1$. By (27) and the condition that $q_0\ge c_1p^2$, we have

$$\mathbb{E}\left[\frac{|\sum_{(r,t)\in\mathcal{H}_{01}}\{I(|\widetilde{W}_{rt}|\ge\tau)-\mathrm{pr}(|\widetilde{W}_{rt}|\ge\tau)\}|}{q_0G(\tau)}\right]\lesssim\frac{p^{1+\gamma}G(\tau)}{q_0G(\tau)}=O(p^{-1+\gamma}).\tag{62}$$

Note that for $(r,t)\in\mathcal{H}_{02}$ and $(l,m)\in\mathcal{H}_{02}$,

$$|\mathrm{Cov}(v_{rt,1}\varepsilon_{r,1},v_{lm,1}\varepsilon_{l,1})|\lesssim\{\sinh(2\theta_{rl,1})+\sinh(2\theta_{rm,1})\}\{\sinh(2\theta_{tl,1})+\sinh(2\theta_{tm,1})\}.$$

Next we split the set $\mathcal{H}_{02}$ into several subsets as done in Cai and Liu (2016). For some large constant $C > 0$, define the following events

$$\mathcal{H}_{3X} = \{(r,t,l,m) : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}, |\text{Cov}(v_{rt,1}\varepsilon_{r,1}, v_{lm,1}\varepsilon_{l,1})| \leq C\{\log p\}^{-2-\xi}\},$$
$$\mathcal{H}_{4X} = \{(r,t,l,m) \notin \mathcal{H}_{3X} : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}, |\text{Cov}(v_{rt,1}\varepsilon_{r,1}, v_{lm,1}\varepsilon_{l,1})| \leq \zeta + C\{\log p\}^{-2-\xi}\},$$
$$\mathcal{H}_{5X} = \{(r,t,l,m) \notin \mathcal{H}_{3X} \cup \mathcal{H}_{4X} : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}\}.$$

The events $\mathcal{H}_{3Y}, \mathcal{H}_{4Y}, \mathcal{H}_{5Y}$ are defined similarly. Let $\mathcal{H}_3 = \mathcal{H}_{3X} \cap \mathcal{H}_{3Y}$, $\mathcal{H}_5 = \mathcal{H}_{5X} \cup \mathcal{H}_{5Y}$, and $\mathcal{H}_4 = \mathcal{H}_{4X} \cup \mathcal{H}_{4Y} \backslash \mathcal{H}_5$. One can show that $|\mathcal{H}_4| = O(p^{2+2\gamma})$ and $|\mathcal{H}_5| = O(p^2 + p^{1+3\gamma})$.

Set $f_{rtlm}(\tau) = \text{pr}(|\widetilde{W}_{rt}| \geq \tau, |\widetilde{W}_{lm}| \geq \tau) - \text{pr}(|\widetilde{W}_{rt}| \geq \tau)\text{pr}(|\widetilde{W}_{lm}| \geq \tau)$. Then

$$\mathbb{E}\left[\frac{|\sum_{(r,t)\in\mathcal{H}_{02}}\{I(|\widetilde{W}_{rt}| \geq \tau) - \text{pr}(|\widetilde{W}_{rt}| \geq \tau)\}|}{q_0 G(\tau)}\right]^2 = \frac{\sum_{(r,t)\in\mathcal{H}_{02}}\sum_{(l,m)\in\mathcal{H}_{02}} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}. \tag{63}$$

By Lemma 4 in the supplementary materials of Cai and Liu (2016), we have

$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_3} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \leq C A_n, \tag{64}$$

where $A_n \leq C(\log p)^{-1-\gamma_2}$ for some $\gamma_2 > 0$. Further, we also have

$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_4} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \leq \frac{C(\tau+1)^{2/(1+\zeta)-2}}{p^{2-2\gamma}\{G(\tau)\}^{2\zeta/(1+\zeta)}}, \tag{65}$$
$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_5} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \leq \frac{C}{p^2 G(\tau)} + \frac{C}{p^{3-3\gamma}G(\tau)}. \tag{66}$$

Combining (62), (63), (64), (65), (66) and the inequality

$$\int_0^{b_p}\left\{p^{-1+\gamma} + A_n + \frac{C(\tau+1)^{2/(1+\zeta)-2}}{p^{2-2\gamma}\{G(\tau)\}^{2\zeta/(1+\zeta)}} + \frac{C}{p^2 G(\tau)} + \frac{C}{p^{3-3\gamma}G(\tau)}\right\} d\tau = o(v_p),$$

we prove (60). Proof of (61) follows similarly.

□