Statistica Sinica Preprint No: SS-2023-0298	
Title	Inferring Hub Nodes on Differential Gaussian Graphical Models
Manuscript ID	SS-2023-0298
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0298
Complete List of Authors	Xin Zhou, Kean Ming Tan and Junwei Lu
Corresponding Authors	Junwei Lu
E-mails	junweilu@hsph.harvard.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

Statistica Sinica

Inferring Hub Nodes on Differential Gaussian Graphical Models

Xin Zhou[1], Kean Ming Tan[2], and Junwei Lu[3]

[1]*Division of Biostatistics, University of California, Berkeley*

[2]*Department of Statistics, University of Michigan*

[3]*Department of Biostatistics, Harvard Chan School of Public Health*

Abstract:

Identifying changes between two networks, also referred to as differential network analysis, has brought new insights to many biological applications. A lot of progress has been made in the development of statistical inference tools for detecting changes between two networks, with most work focused on testing whether two networks are exactly the same, or whether there is an edge that is missing in one network but present in another. However, in many scientific settings, it is often more interesting to identify nodes that have different conditional dependency structures between two networks, which we refer to as differential hub nodes. In this paper, we propose an inferential framework to test whether there is at least one differential hub node in a differential Gaussian graphical model. As a by-product, our proposed test statistic can also be used to test the hypothesis on whether there is a differential edge and construct a confidence interval for the corresponding differential edge. Theoretically, we show that the proposed method yields an asymptotic valid test and that the type II error decreases to zero asymptotically. The proposed method is applied to both simulated data and the Genotype-Tissue Expression (GTEx) data to evaluate whether gene regulatory networks between males and females for different tissues are different.

Key words and phrases: Differential network; Gaussian multiplier bootstrap; hypothesis testing; maximum degree.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

---

1. Introduction

Undirected graphical models have been used extensively for modeling conditional dependence relationships among a set of random variables in many scientific domains (Markowetz and Spang, 2007; Rubinov and Sporns, 2010). An undirected graph consists of d nodes and a set of edges: each node represents a random variable, and an edge between two nodes indicates that the corresponding two random variables are conditionally dependent, conditioned on all the other variables. Given a number of independent and identically distributed random samples, many methods were proposed to estimate a sparse undirected graphical model under various assumptions on the random variables (Meinshausen and Bühlmann, 2006; Friedman et al., 2008; Cai et al., 2011; Lee and Hastie, 2015; Tan et al., 2016; Yang et al., 2018). We refer the reader to Drton and Maathuis (2017) for a review on recent developments of estimating an undirected graph.

In many applications, the primary interest is not in estimating a particular network, but in assessing whether there are any differences between two undirected graphs or networks. The set of changes or differences between two networks is often referred to as the differential network (Shojaie, 2020). Differential network analysis has been considered in many scientific disciplines such as genomics and neuroscience (Ideker and Krogan, 2012; Jackson et al., 2016). For instance, in the context of genomics, one may collect gene expression measurements for a set of normal tissue samples and a set of cancer tissue samples. Locating differentially connected nodes between gene regulatory networks of healthy individuals and cancer patients can help researchers focus on potential genes that will lead to an understanding of the underlying disease mechanism (Ideker and Krogan, 2012; Ha et al., 2015).

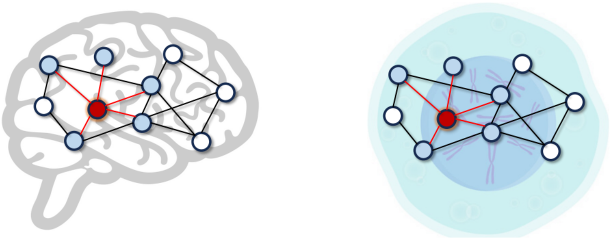
Existing work on estimating differential networks can be grouped into three different approaches. The first approach is to estimate each network separately using methods for

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

constructing sparse undirected graphical models and construct the differential network by taking the difference between the two estimated networks. However, such an approach relies on the assumption that each of the true underlying networks must be sparse, which may be restrictive for many applications (Meinshausen and Bühlmann, 2006; Cai et al., 2011). The second approach is to jointly estimate the two networks by assuming that the two networks share some similar conditional dependencies structure. This can be done by incorporating a group lasso or fused lasso type penalties to encourage the two networks to have similar edge sets (Mohan et al., 2014; Danaher et al., 2014; Ma and Michailidis, 2016). The third approach estimates the differential network directly and imposes the sparsity assumption on the differential network, i.e., such an approach allows each of the networks to be dense (Zhao et al., 2014; Liu et al., 2014; Kim et al., 2021). We refer the reader to Shojaie (2020) for a review of various approaches for estimating differential networks.

While estimation procedures are well-developed for estimating differential networks, statistical inference methods for differential networks are relatively lacking. In the context of a single undirected graph, various methods were developed for inferring whether there is an edge between two nodes, i.e., whether two random variables are conditionally independent given the others (Jankova and Van De Geer, 2015; Janková and van de Geer, 2017; Ren et al., 2015; Tan et al., 2016; Xia and Li, 2017; Yang et al., 2018; Yu et al., 2020). Related ideas have been generalized to the context of differential networks, but most methods rely on separate estimations of each network (Xia et al., 2015; Liu et al., 2017; Cai et al., 2019). Moreover, the aforementioned work mainly focuses on testing whether two networks are exactly the same, or whether an edge is present between two nodes in both networks.

We instead consider the problem of testing whether there exists a hub node in a differential network, i.e., nodes that exhibit substantial variations in their connectivity between two networks.



Brain network with
a voxel hub

Genomic network with
a gene hub

Figure 1: An illustration of a 5-hub (red node) of a brain network and a genomic network.

In specific, given a graph $G = (V, E)$, the degree of a node $j \in V$ is the number of edges connected to j, i.e., $\deg(j) = |\{j \in V \mid (i, j) \in E\}|$ and a k-hub is a node with degree larger than k. Identifying hub nodes in a differential network is crucial to many biological scientific applications. For instance, in the context of genomics, there may be mutated genes that lead to uncontrolled cell growth that causes cancer, and such genes tend to have significantly different interactions with other genes between networks for healthy and cancer subjects (Chalhoub and Baker, 2009; Mohan et al., 2014). In the context of neuroscience, the brain connectivity network for patients with Alzheimer's disease was shown to be significantly different from that of healthy subjects, and such differences between the two brain networks are hypothesized to be potentially due to disruptions of several hub brain regions (Li et al., 2016; Kundu et al., 2019). Various methods were developed for estimating a network or a differential network with hubs (Mohan et al., 2014; Tan et al., 2014; Sulaimanov et al., 2019; Kim et al., 2019). On the other hand, statistical inference procedure for testing the existence of hub nodes is relatively lacking, except the work of Neykov et al. (2019), Lu et al. (2017), Neykov and Liu (2019), and Tan et al. (2021) in the context of Gaussian graphical models, Ising models, and time-varying Gaussian graphical models, respectively.

Let $\mathbf{\Delta} \in \mathbb{R}^{d\times d}$ be the parameter matrix that encodes the difference between two networks. That is, $\mathbf{\Delta}_{jk} = 0$ if and only if both networks have the same conditional dependency structure between the jth and kth random variables, and vice versa. Let $\mathbf{\Delta}_{\bullet j}$ be the jth column of $\mathbf{\Delta}$. In this paper, we develop a hypothesis testing framework for testing whether there are any hub nodes in a differential graph based on the method proposed in Zhao et al. (2014) for two sets of d-dimensional multivariate Gaussian random variables. In particular, we are interested in testing

$$H_0 : \max_j \; \|\mathbf{\Delta}_{\bullet j}\|_0 \le k \qquad \text{versus} \qquad H_1 : \max_j \; \|\mathbf{\Delta}_{\bullet j}\|_0 > k, \tag{1.1}$$

where k is a user-specified constant. Hypothesis testing problem (1.1) is a challenging problem due to its combinatorial nature, and existing approaches for testing a single edge for graphical models cannot be directly adapted to testing (1.1). Our proposed method generalizes existing work on combinatorial inference to the context of the differential graph (Neykov et al., 2019; Lu et al., 2017; Neykov and Liu, 2019; Tan et al., 2021). As a by-product, hypothesis test for entrywise changes in the differential network is also developed for the method of Zhao et al. (2014):

$$H_{0,jk} : \mathbf{\Delta}_{jk} = 0 \qquad \text{versus} \qquad H_{1,jk} : \mathbf{\Delta}_{jk} \neq 0. \tag{1.2}$$

Our method complements existing procedures on hypothesis testing for differential networks (Xia et al., 2015; Liu et al., 2017; Cai et al., 2019).

Specifically, given the estimation framework in Zhao et al. (2014), we provide a statistical inference framework and theory for testing whether there is a differential hub between two networks. The main idea includes constructing a debiased estimator, proving that the debiased estimator is asymptotically normal, and proposing an algorithm that is statistically valid for testing (1.1). In Section 2, we provide a brief review of Gaussian differential graph in Zhao

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

et al. (2014). Our proposed inferential framework is detailed in Section 3. Section 4 contains the theoretical results for the proposed framework. We conclude with some numerical studies and data application in Section 5, and a discussion section.

Notation: For $n \in \mathbb{N}$, we use the shorthand notation $[n] = \{1, 2, \ldots, n\}$. For a matrix $\mathbf{A}$, we use $\mathbf{A}_{j\bullet}$ and $\mathbf{A}_{\bullet k}$ to denote the jth row and kth column of $\mathbf{A}$, respectively. Let $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ be the maximum and minimum eigenvalue of the matrix $\mathbf{A}$, respectively. Let $\|\mathbf{A}\|_{\max} = \max_{jk} |\mathbf{A}_{jk}|$, $\|\mathbf{A}\|_{1,1} = \sum_{j,k} |\mathbf{A}_{jk}|$, and $\|\mathbf{A}\|_q = \sup_{\|x\|_q=1} \|\mathbf{A}x\|_q$. In addition, we denote the induced ℓ_q-norm of a matrix as $\|\mathbf{A}\|_1 = \max_k \sum_j |\mathbf{A}_{jk}|$, $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$, and $\|\mathbf{A}\|_\infty = \max_j \sum_k |\mathbf{A}_{jk}|$, where $\sigma_{\max}(\mathbf{A})$ is the largest singular value of $\mathbf{A}$. For a sub-Gaussian variable X, we define the sub-Gaussian norm of X as $\|X\|_{\psi_1} = \sup_{p\geq 1} p^{-\frac{1}{2}}(\mathbb{E}|X|^p)^{\frac{1}{p}}$. And for a sub-exponential variable X, we define the sub-exponential norm of X as $\|X\|_{\psi_2} = \sup_{p\geq 1} p^{-1}(\mathbb{E}|X|^p)^{\frac{1}{p}}$. We denote $\Phi(x)$ as the cumulative distribution function of a standard normal random variable. For a sequence of random variables X_n and a random variable X, we write $X_n \rightsquigarrow X$ if X_n converges in distribution to X.

2. A Brief Review on Estimating a Gaussian Differential Network

In this section, we provide a brief review of previous approaches to estimating Gaussian differential networks. Let $\boldsymbol{X} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_X)$ and $\boldsymbol{Y} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_Y)$ be two d-dimensional random variables, where $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ are the $d \times d$ covariance matrices for the two d-dimensional vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. For notational convenience, we denote $\boldsymbol{\Theta}_X = (\boldsymbol{\Sigma}_X)^{-1}$ and $\boldsymbol{\Theta}_Y = (\boldsymbol{\Sigma}_Y)^{-1}$ as the corresponding inverse covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Under the Gaussian assumption, $\boldsymbol{\Theta}_{X,jk} = 0$ if and only if there is an edge between X_j and X_k. In other words, the inverse covariance matrices $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$ encode the conditional dependence relationships among

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

the random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Thus, under the Gaussian assumption on $\boldsymbol{X}$ and $\boldsymbol{Y}$, the differential network is defined as the difference between the two precision matrices, i.e., $\boldsymbol{\Delta} = \boldsymbol{\Theta}_Y - \boldsymbol{\Theta}_X$.

A natural way to estimate a differential network under the Gaussian assumption on $\boldsymbol{X}$ and $\boldsymbol{Y}$ is by minimizing the negative log-likelihood with some form of regularizations on the parameter of interest to encourage sparsity or similarity between two networks. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_X}$ and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_Y}$ be n_X and n_Y random samples of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Moreover, let $\widehat{\boldsymbol{\Sigma}}_X = n_X^{-1} \sum_{i=1}^{n_X} \boldsymbol{X}_i \boldsymbol{X}_i^T$ and $\widehat{\boldsymbol{\Sigma}}_Y = n_Y^{-1} \sum_{i=1}^{n_Y} \boldsymbol{Y}_i \boldsymbol{Y}_i^T$ be the sample covariance matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Then, a differential graph can be estimated by solving the optimization problem:

$$\underset{\boldsymbol{\Theta}_X, \boldsymbol{\Theta}_Y}{\text{minimize}} \; -\log\det(\boldsymbol{\Theta}_X) - \log\det(\boldsymbol{\Theta}_Y) + \text{tr}(\widehat{\boldsymbol{\Sigma}}_X \boldsymbol{\Theta}_X) + \text{tr}(\widehat{\boldsymbol{\Sigma}}_Y \boldsymbol{\Theta}_Y) + P_{\boldsymbol{\lambda}}(\boldsymbol{\Theta}_X, \boldsymbol{\Theta}_Y),$$

where $P_{\boldsymbol{\lambda}}(\boldsymbol{\Theta}_X, \boldsymbol{\Theta}_Y)$ is a penalty function on the parameters $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$. One naive approach is to estimate each network separately by adding the lasso penalty to encourage sparsity, i.e., by setting $P_{\boldsymbol{\lambda}}(\boldsymbol{\Theta}_X, \boldsymbol{\Theta}_Y) = \lambda_1 \|\boldsymbol{\Theta}_X\|_{1,1} + \lambda_2 \|\boldsymbol{\Theta}_Y\|_{1,1}$ (Meinshausen and Bühlmann, 2006; Yuan and Lin, 2007; Friedman et al., 2008). Another approach is to employ a fused lasso type penalty to encourage the two networks to be similar (Mohan et al., 2014; Danaher et al., 2014; Ma and Michailidis, 2016). However, the aforementioned approaches assume sparsity on both $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$. Such approaches limit the application of differential network analysis for problems with potentially dense networks such as brain connectivity networks.

In this paper, we focus on the proposed method in Zhao et al. (2014) in which the differential network is estimated directly, and thus allowing $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$ to be dense as long as the differential network $\boldsymbol{\Delta}$ is sparse. The main crux of their proposed method relies on the observation that

---

$\mathbf{\Sigma}_X\mathbf{\Delta}\mathbf{\Sigma}_Y - (\mathbf{\Sigma}_X - \mathbf{\Sigma}_Y) = \mathbf{0}$. Thus, to obtain an estimate of $\mathbf{\Delta}$, Zhao et al. (2014) proposed to solve the following optimization problem:

$$\widehat{\mathbf{\Delta}} = \arg\min \|\mathbf{\Delta}\|_{1,1}, \text{ such that } \|\widehat{\mathbf{\Sigma}}_X\mathbf{\Delta}\widehat{\mathbf{\Sigma}}_Y - (\widehat{\mathbf{\Sigma}}_X - \widehat{\mathbf{\Sigma}}_Y)\|_{\max} \le \lambda, \tag{2.1}$$

where λ is a sparsity tuning parameter that encourages the estimated $\mathbf{\Delta}$ to be sparse. Optimization problem (2.1) can be solved using a linear programming method, and we refer the reader to Zhao et al. (2014) for details.

3. Statistical Inference

3.1 Inference on a Single Edge

In this section, we consider testing a prespecified component in $\mathbf{\Delta}$ as in hypothesis testing problem (1.2):

$$H_{0,jk} : \mathbf{\Delta}_{jk} = 0 \qquad \text{versus} \qquad H_{1,jk} : \mathbf{\Delta}_{jk} \neq 0.$$

Due to the max-norm constrained in (2.1), the estimator is no longer asymptotically normal, and test statistics based on $\widehat{\mathbf{\Delta}}$ are no longer valid. To address this issue, we construct a debiased estimator for $\widehat{\mathbf{\Delta}}$ that can be shown to be asymptotically normal. Similar ideas have been considered in the context of graphical models for testing whether there is an edge between two nodes (Jankova and Van De Geer, 2015; Janková and van de Geer, 2017; Ren et al., 2015; Tan et al., 2016; Yang et al., 2018; Xia and Li, 2017; Xia et al., 2015; Liu et al., 2017; Cai et al., 2019).

Motivated by Neykov et al. (2018), we construct a debiased estimator for $\widehat{\mathbf{\Delta}}$ obtained from solving (2.1):

$$\mathrm{vec}(\widehat{\boldsymbol{\Delta}}^d) = \mathrm{vec}(\widehat{\boldsymbol{\Delta}}) - \mathbf{V}\{(\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y)\mathrm{vec}(\widehat{\boldsymbol{\Delta}}) - \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_X - \widehat{\boldsymbol{\Sigma}}_Y)\}$$

where $\mathbf{V} = \mathbf{V}_X \otimes \mathbf{V}_Y \in \mathbb{R}^{d^2 \times d^2}$ can be interpreted as a bias correction matrix. The above equation can be rewritten as

$$\widehat{\boldsymbol{\Delta}}^d = \widehat{\boldsymbol{\Delta}} - \mathbf{V}_Y\{\widehat{\boldsymbol{\Sigma}}_Y \widehat{\boldsymbol{\Delta}} \widehat{\boldsymbol{\Sigma}}_X - (\widehat{\boldsymbol{\Sigma}}_X - \widehat{\boldsymbol{\Sigma}}_Y)\}\mathbf{V}_X^T, \tag{3.1}$$

where $\mathbf{V}_X \in \mathbb{R}^{d \times d}$ and $\mathbf{V}_Y \in \mathbb{R}^{d \times d}$ are to be defined in (3.4) later in this section. We now provide an intuition for establishing asymptotic normality of the debiased estimator $\widehat{\boldsymbol{\Delta}}^d$ in (3.1).

Through some algebraic manipulations, $\widehat{\boldsymbol{\Delta}}^d - \boldsymbol{\Delta}$ can be partitioned into two terms, the leading term and the remainder term, i.e., $\widehat{\boldsymbol{\Delta}}^d - \boldsymbol{\Delta} = \text{Leading} + \text{Remainder}$, where

$$\text{Leading} = (\mathbf{V}_Y - \boldsymbol{\Delta})\big(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X\big)\mathbf{V}_X^T - \mathbf{V}_Y\big(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})^T; \tag{3.2}$$

$$\begin{aligned}\text{Remainder} =& \{(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}) - \mathbf{V}_Y\widehat{\boldsymbol{\Sigma}}_Y(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})(\mathbf{V}_X\widehat{\boldsymbol{\Sigma}}_X)^T\} - \mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\boldsymbol{\Delta}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X)\mathbf{V}_X^T \\ & - \mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\boldsymbol{\Delta}(\mathbf{V}_X\boldsymbol{\Sigma}_X - \mathbf{I})^T - (\mathbf{V}_Y\boldsymbol{\Sigma}_Y - \mathbf{I})\boldsymbol{\Delta}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X)\mathbf{V}_X^T.\end{aligned} \tag{3.3}$$

The key idea is to show that the leading term (3.2) is asymptotically normal and the remainder term (3.3) converges to zero in probability. From (3.3), we see that the remainder term converges to zero only if $\mathbf{V}_X\widehat{\boldsymbol{\Sigma}}_X \approx \mathbf{I}$ and $\mathbf{V}_Y\widehat{\boldsymbol{\Sigma}}_Y \approx \mathbf{I}$. Intuitively, we can construct the bias correction matrices $\mathbf{V}_X$ and $\mathbf{V}_Y$ such that the difference between $\mathbf{V}_X\widehat{\boldsymbol{\Sigma}}_X$ and $\mathbf{I}$, and $\mathbf{V}_Y\widehat{\boldsymbol{\Sigma}}_Y$ and $\mathbf{I}$ to be small. However, the leading term depends also on $\mathbf{V}_X$ and $\mathbf{V}_Y$, and is challenging to analyze due to the dependency between $\mathbf{V}_X$ and $\widehat{\boldsymbol{\Sigma}}_X$, and $\mathbf{V}_Y$ and $\widehat{\boldsymbol{\Sigma}}_Y$, respectively.

The aforementioned challenge motivates the following sample splitting procedure for constructing $\mathbf{V}_X$ and $\mathbf{V}_Y$. Such a sample splitting procedure was also considered in Ma et al. (2021) in the context of performing inference on a single Gaussian graphical model. For notational simplicity, we assume that $n_X = n_Y = 2n$, where n is a positive integer. The main crux is to split the data into two parts $\mathcal{D}_1$ and $\mathcal{D}_2$, each of which consists of n independent samples for both random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$. We use the first data set $\mathcal{D}_1$ to estimate the sample covariance matrices $\widehat{\boldsymbol{\Sigma}}_X$ and $\widehat{\boldsymbol{\Sigma}}_Y$, which are then used to estimate $\widehat{\boldsymbol{\Delta}}$ by solving the optimization problem in (2.1). Then, we use the second data set $\mathcal{D}_2$ to estimate $\widehat{\boldsymbol{\Sigma}}'_X$ and $\widehat{\boldsymbol{\Sigma}}'_Y$, which are then used to estimate $\mathbf{V}_X$ and $\mathbf{V}_Y$ by solving the following optimization problems:

$$\begin{aligned} \mathbf{V}_X &= \arg\min \|\mathbf{U}\|_\infty, \text{such that } \ \|\mathbf{U}\widehat{\boldsymbol{\Sigma}}'_X - \mathbf{I}\|_{\max} \le \lambda', \\ \mathbf{V}_Y &= \arg\min \|\mathbf{Z}\|_\infty, \text{such that } \|\mathbf{Z}\widehat{\boldsymbol{\Sigma}}'_Y - \mathbf{I}\|_{\max} \le \lambda', \end{aligned} \tag{3.4}$$

where λ' is a tuning parameter that gives the approximation errors of $\mathbf{V}_X\widehat{\boldsymbol{\Sigma}}'_X$ and $\mathbf{V}_Y\widehat{\boldsymbol{\Sigma}}'_Y$ for estimating $\mathbf{I}$. In other words, the correction matrices $\mathbf{V}_X$ and $\mathbf{V}_Y$ depend only on the covariance matrices $\widehat{\boldsymbol{\Sigma}}'_X$ and $\widehat{\boldsymbol{\Sigma}}'_Y$ constructed using $\mathcal{D}_2$, and thus $\mathbf{V}_X$ and $\mathbf{V}_Y$ are independent of $\widehat{\boldsymbol{\Sigma}}_X$ and $\widehat{\boldsymbol{\Sigma}}_Y$. With such a choice for $\mathbf{V}_X$ and $\mathbf{V}_Y$, the leading term can then be viewed as an empirical process type quantity that can be shown to be asymptotically normal. In particular, we will show in Section 4.1 that $\sqrt{n}(\widehat{\boldsymbol{\Delta}}^d_{jk} - \boldsymbol{\Delta}_{jk}) \rightsquigarrow N(0, \xi^2_{jk})$, where ξ^2_{jk} is the asymptotic variance defined in (4.3) in Section 4.1.

Let $\widehat{\xi}_{jk}$ be an estimator for ξ_{jk} defined in (4.5). Based on the above result on the asymptotic normality of $\widehat{\boldsymbol{\Delta}}^d_{jk}$, for a given significance level $0 < \alpha < 1$, we construct the following test statistic

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

to test the hypothesis testing problem (1.2):

$$T_{jk}(\alpha) = \begin{cases} 1, & |\sqrt{n}\widehat{\mathbf{\Delta}}^d_{jk}/\widehat{\xi}_{jk}| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right), \\ 0, & \text{otherwise}, \end{cases} \tag{3.5}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Moreover, we construct a $(1-\alpha)\%$ confidence interval as

$$I_{jk}(\alpha) = [\widehat{\mathbf{\Delta}}^d_{jk} - \eta(\alpha, n), \widehat{\mathbf{\Delta}}^d_{jk} + \eta(\alpha, n)], \text{ where } \eta(\alpha, n) = \frac{\widehat{\xi}_{jk}}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right). \tag{3.6}$$

We will show in Section 4.1 that under the null hypothesis (1.2), the type I error of $T_{jk}(\alpha)$ converges to α, and the confidence interval in (3.6) is asymptotically valid. The overall procedure for conducting the hypothesis test for (1.2) is summarized in Algorithm 1.

Algorithm 1 Proposed method for testing $H_0 : \mathbf{\Delta}_{jk} = 0$.

```
    Input: significance level $\alpha$, tuning parameters $\lambda$ and $\lambda'$, and the data $\mathcal{D}_1$ and $\mathcal{D}_2$.
    Step 1: Construct two sets of covariance matrices: $\widehat{\mathbf{\Sigma}}_X$ and $\widehat{\mathbf{\Sigma}}_Y$ using data from $\mathcal{D}_1$; and $\widehat{\mathbf{\Sigma}}'_X$ and $\widehat{\mathbf{\Sigma}}'_Y$ using data from $\mathcal{D}_2$.
    Step 2: Obtain $\widehat{\mathbf{\Delta}}$ from solving (2.1) using data from $\mathcal{D}_1$.
    Step 3: Obtain the bias correction matrices $\mathbf{V}_X$ and $\mathbf{V}_Y$ in (3.4) using data from $\mathcal{D}_2$.
    Step 4: Construct the debiased estimator $\widehat{\mathbf{\Delta}}^d_{jk}$ in (3.1).
    Step 5: Calculate $\widehat{\xi}_{jk}$ defined in (4.5).
    Step 6: Reject $H_0 : \mathbf{\Delta}_{jk} = 0$ if $|\sqrt{n}\widehat{\mathbf{\Delta}}^d_{jk}/\widehat{\xi}_{jk}| > \Phi^{-1}(1-\alpha/2)$.
```

3.2 Inference on Maximum Degree

In this section, we turn to testing hypothesis problem (1.1):

$$H_0 : \max_j \|\mathbf{\Delta}_{\bullet j}\|_0 \leq k \qquad \text{versus} \qquad H_1 : \max_j \|\mathbf{\Delta}_{\bullet j}\|_0 > k,$$

where k is a user-specified constant, usually based on the scientific context. Due to the combinatorial nature of the testing problem, the debiased estimator (3.1) cannot be directly applied to test the aforementioned problem. To this end, we propose an inferential method for testing problem (1.1). Our method generalizes existing work on testing combinatorial graph structure in the context of a single graphical model (Neykov et al., 2019; Lu et al., 2017; Neykov and Liu, 2019; Tan et al., 2021).

Let $E \subseteq V \times V$ be an edge set of the differential graph defined by $\mathbf{\Delta}$. We first construct the following test statistic:

$$T_E = \max_{(j,k)\in E} \sqrt{n}(\widehat{\mathbf{\Delta}}^d_{jk} - \mathbf{\Delta}_{jk}). \tag{3.7}$$

We then approximate the distribution of T_E using the Gaussian multiplier bootstrap (Chernozhukov et al., 2013). Recall from (3.2) and (3.3) that $\widehat{\mathbf{\Delta}}^d - \mathbf{\Delta}$ can be decomposed as the sum of a leading term in (3.2) and a remainder term in (3.3). Since the remainder term converges to zero in probability, it suffices to obtain a good approximation to the leading term. To approximate the leading term, for $i = 1, \ldots, n$ in $\mathcal{D}_1$, we propose the following bootstrap statistic

$$T_E^B = \max_{(j,k)\in E} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{(\mathbf{V}_Y - \widehat{\mathbf{\Delta}})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T - \widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T - \mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \widehat{\mathbf{\Sigma}}_Y)(\mathbf{V}_X + \widehat{\mathbf{\Delta}})_{k\bullet}^T\}\xi_i, \tag{3.8}$$

where $\xi_1, \ldots, \xi_n \overset{\text{iid}}{\sim} N(0,1)$, $\widehat{\mathbf{\Delta}}$ is an estimator obtained from solving (2.1) using data from $\mathcal{D}_1$, and $\mathbf{V}_X$ and $\mathbf{V}_Y$ are the bias correction matrices obtained from solving (3.4) using data from $\mathcal{D}_2$. Let

$$c(\alpha, E) = \inf\{t \in \mathbb{R} \big| \mathbb{P}(|T_E^B| > t|\mathcal{D}_1) \leq \alpha\} \tag{3.9}$$

be the α-quantile of the bootstrap statistic T_E^B given the data from $\mathcal{D}_1$, which can be calculated using Monte-Carlo. In Section 4.2, we will show theoretically that, conditioned on the data $\mathcal{D}_2$, the quantile of T_E can be estimated accurately by the quantile of the bootstrap statistic T_E^B.

We now propose a method for testing (1.1) in Algorithm 2, i.e., whether the maximum degree of the differential graph is less than or equal to a pre-specified number k. The main crux of our proposed method is to compute the conditional quantile $c(\alpha, E)$ using the bootstrap statistic, and construct a rejected edge set for edges with test statistics that are larger than the conditional quantile. Then, the maximum degree of the rejected edge set is calculated and the null hypothesis is rejected if the maximum degree of the rejected edge set is larger than k. We will show that the aforementioned method leads to a valid test for hypothesis problem (1.1) in Section 4.2. Besides the maximum degree, our proposed method can also be used to test other combinatorial graph structures such as the number of connected subgraphs, the size of the longest chain, and the number of isolated nodes. We refer the reader to Neykov et al. (2019) and Lu et al. (2017) for details on other graph structures.

4. Theoretical Results

In Section 4.1, we establish the asymptotic normality of the debiased estimator $\widehat{\mathbf{\Delta}}^d$ in (3.1). We then show that Algorithm 1 is a valid test. In particular, under the null hypothesis (1.2), the type I error of the test statistic $T_{jk}(\alpha)$ in (3.5) converges to α, and the confidence interval in (3.6) is asymptotically valid. In Section 4.2, we show that the conditional quantile $c(\alpha, E)$ of the Gaussian multiplier bootstrap T_E^B in (3.7) is a consistent estimator of the quantile of T_E. We

Algorithm 2 Method for testing maximum degree $H_0 : \max_j \ \|\boldsymbol{\Delta}_{\bullet j}\|_0 \leq k$.

```
Input: type I error $\alpha$, pre-specified degree $k$, debiased estimator $\widehat{\boldsymbol{\Delta}}^d$, edge set $E = \{(i,j) \in V \times V \mid i \neq j\}$
Step I: Compute the conditional quantile $c(\alpha, E) = \inf\{t \in \mathbb{R} \mid \mathbb{P}(|T_E^B| > t|\mathcal{D}_1) \leq \alpha\}$.
Step II: Construct the rejected edge set $\mathcal{R} = \{e \in E \mid \sqrt{n}(\widehat{\boldsymbol{\Delta}}_e^d - \boldsymbol{\Delta}_e) > c(\alpha, E)\}$.
Step III: Compute the maximum degree $d_{\text{rej}}$ of the graph based on the rejected edge set.
Output: Reject the null hypothesis if $d_{\text{rej}} > k$.
```

then show that the proposed method in Algorithm 2 for testing hypothesis testing problem (1.1) is a valid test. Throughout the section, we study the asymptotic regime in which n, d, and s are allowed to increase.

4.1 Asymptotic Normality of the Debiased Estimator (3.1)

We start with the asymptotic normality of $\widehat{\boldsymbol{\Delta}}^d$. Motivated by Zhao et al. (2014), we consider the following family of true differential graphs:

$$\mathcal{U}(M, s) = \Big\{\boldsymbol{\Delta} \in \mathbb{R}^{d \times d} \mid \sum_{j \in [d]} \|\boldsymbol{\Delta}_{\bullet j}\|_0 \leq s, \|\boldsymbol{\Delta}\|_1 \leq M\Big\}, \tag{4.1}$$

where $s < n$ and M is a constant that does not depend on n, s, and d. The family of differential graphs we consider in (4.1) requires only that the differential graph is sparse and that each of the inverse covariance matrices can be dense. This is particularly useful in applications such as brain connectivity networks where each network is usually very dense. As suggested in Zhao et al. (2014), the aforementioned is the primary advantage over estimating the differential graph by estimating the inverse covariance matrices separately and taking the difference, in which

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

sparsity on each inverse covariance matrix is required.

Next, we impose a condition on the population covariance and inverse covariance matrices adapted from Zhao et al. (2014). We refer the reader to Zhao et al. (2014) for further details on the interpretation of Assumption 1.

Assumption 1. Assume that $\|\mathbf{\Sigma}_X\|_2 = O(1)$, $\|\mathbf{\Sigma}_Y\|_2 = O(1)$, $\|\mathbf{\Theta}_X\|_1 = O(1)$, and $\|\mathbf{\Theta}_Y\|_1 = O(1)$. Moreover, assume that

$$4 \max\{\max_j \Sigma_{X,jj}, \max_{j\neq k} |\Sigma_{X,jk}|, \max_j \Sigma_{Y,jj}, \max_{j\neq k} |\Sigma_{Y,jk}|\} \leq \sigma^S_{\min}(2s)^{-1},$$

where $\sigma^S_{\min} = \min_{j,k}\{\Sigma_{X,jj}\Sigma_{Y,jj}, \Sigma_{X,jj}\Sigma_{Y,kk} + 2\Sigma_{Y,kj}\Sigma_{X,jk} + \Sigma_{X,kk}\Sigma_{Y,jj}\}$.

Under Assumption 1, the following theorem establishes the asymptotic normality of the debiased estimator $\widehat{\mathbf{\Delta}}^d$ in Theorem 4.1.

Theorem 4.1. Assume that Condition 1 holds and that $\mathbf{\Delta} \in \mathcal{U}(M, s)$. Moreover, assume the scaling conditions $(s \log d)/\sqrt{n} = o(1)$ and $\log^6 d/n = o(1)$. Let $\lambda = C\sqrt{\log d/n}$ and $\lambda' = C'\sqrt{\log d/n}$, where C and C' are some sufficiently large constants. Then, conditional on $\mathcal{D}_2$,

$$\sqrt{n}(\widehat{\mathbf{\Delta}}^d_{jk} - \mathbf{\Delta}_{jk})/\xi_{jk} \rightsquigarrow N(0,1), \quad \forall j,k \in [d], \tag{4.2}$$

where

$$\begin{aligned}\xi^2_{jk} = {} & \{(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X(\mathbf{V}_Y - \mathbf{\Delta})^T_{j\bullet}\}(\mathbf{V}_{X,k\bullet}\mathbf{\Sigma}_X\mathbf{V}^T_{X,k\bullet}) + \{\}(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X\mathbf{V}^T_{X,k\bullet}\}^2 \\ & + (\mathbf{V}_{Y,j\bullet}\mathbf{\Sigma}_Y\mathbf{V}^T_{Y,j\bullet})\{(\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}\mathbf{\Sigma}_Y(\mathbf{V}_X + \mathbf{\Delta})^T_{k\bullet}\} + \{\mathbf{V}_{Y,j\bullet}\mathbf{\Sigma}_Y(\mathbf{V}_X + \mathbf{\Delta})^T_{k\bullet}\}^2\end{aligned} \tag{4.3}$$

is the asymptotic variance of $\widehat{\mathbf{\Delta}}^d_{jk}$.

4.1 Asymptotic Normality of the Debiased Estimator (3.1)

By the Edgeworth expansion (Hall, 2013), we further have the convergence rate of the asymptotic normality

$$\sup_t \left|\mathbb{P}\{\sqrt{n}(\widehat{\boldsymbol{\Delta}}^d_{jk} - \boldsymbol{\Delta}_{jk})/\xi_{jk} \le t\} - \Phi(t)\right| = O(1/\sqrt{n}). \tag{4.4}$$

The asymptotic variance ξ^2_{jk} in (4.3) depends on population quantities of $\boldsymbol{\Delta}$, $\boldsymbol{\Sigma}_X$, and $\boldsymbol{\Sigma}_Y$ and need to be estimated. Let

$$\begin{aligned}\widehat{\xi}^2_{jk} = &\{(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}\widehat{\boldsymbol{\Sigma}}_X(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})^T_{j\bullet}\}(\mathbf{V}_{X,k\bullet}\widehat{\boldsymbol{\Sigma}}_X\mathbf{V}^T_{X,k\bullet}) + \{(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}\widehat{\boldsymbol{\Sigma}}_X\mathbf{V}^T_{X,k\bullet}\}^2 \\ &+ (\mathbf{V}_{Y,j\bullet}\widehat{\boldsymbol{\Sigma}}_Y\mathbf{V}^T_{Y,j\bullet})\{(\mathbf{V}_X + \widehat{\boldsymbol{\Delta}})_{k\bullet}\widehat{\boldsymbol{\Sigma}}_Y(\mathbf{V}_X + \widehat{\boldsymbol{\Delta}})^T_{k\bullet}\} + \{\mathbf{V}_{Y,j\bullet}\widehat{\boldsymbol{\Sigma}}_Y(\mathbf{V}_X + \widehat{\boldsymbol{\Delta}})^T_{k\bullet}\}^2\end{aligned} \tag{4.5}$$

be the estimated variance by substituting $\boldsymbol{\Sigma}_X$, $\boldsymbol{\Sigma}_Y$ and $\boldsymbol{\Delta}$ with $\widehat{\boldsymbol{\Sigma}}_X$, $\widehat{\boldsymbol{\Sigma}}_Y$ and $\widehat{\boldsymbol{\Delta}}$, respectively. In the following lemma, we show that (4.5) is a consistent estimator of (4.3).

Lemma 4.2. Under the same conditions as in Theorem 4.1, we have $\widehat{\xi}_{jk}/\xi_{jk} \xrightarrow{p} 1$.

Combining the results in Lemma 4.2 and Theorem 4.1 with Slutsky's theorem, we have the following corollary on the validity of the test statistic (3.5) and confidence interval in (3.6).

Corollary 4.3. Under the same conditions as in Theorem 4.1, we have

$$\sqrt{n}(\widehat{\boldsymbol{\Delta}}^d_{jk} - \boldsymbol{\Delta}_{jk})/\widehat{\xi}_{jk} \rightsquigarrow N(0,1), \quad \forall j,k \in [d]. \tag{4.6}$$

Moreover, under the null hypothesis (1.2), the type I error of (3.5) is asymptotically α and the confidence interval in (3.6) is asymptotically valid, i.e.,

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\Delta}_{jk}=0}\{T_{jk}(\alpha) = 1\} = \alpha \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}\{\boldsymbol{\Delta}_{jk} \in I_{jk}(\alpha)\} = 1 - \alpha.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

4.2 Asymptotic Validity of the Inferential Method in Algorithm 2

In this section, we show that the conditional α-quantile of the Gaussian multiplier bootstrap statistic T_E^B provides a good approximation to the distribution of T_E. We then show that the proposed inferential method in Algorithm 2 is valid for testing (1.1):

$$H_0 : \max_j \|\mathbf{\Delta}_{\bullet j}\|_0 \leq k \qquad \text{versus} \qquad H_1 : \max_j \|\mathbf{\Delta}_{\bullet j}\|_0 > k.$$

That is, the proposed method in Algorithm 2 yields a type I error that is controlled at a pre-specified level α.

We start with the following theorem that establishes the validity of the proposed Gaussian multiplier bootstrap, conditional on the data $\mathcal{D}_2$ used for obtaining $\mathbf{V}_X$ and $\mathbf{V}_Y$.

Theorem 4.4. Assume that the conditions in Theorem 4.1 hold. Conditional on $\mathcal{D}_2$ and under the scaling conditions $s^2\{\log(dn)\}^4/n = o(1)$ and $\{\log(dn)\}^7/n = o(1)$, for any edge set $E \subseteq V \times V$, we have

$$\lim_{n\to\infty} \sup_{\mathbf{\Delta}\in\mathcal{U}(M,s)} \mathbb{P}\{\max_{(j,k)\in E} \sqrt{n}(\widehat{\mathbf{\Delta}}_{jk}^{d} - \mathbf{\Delta}_{jk}) > c(\alpha, E)\} \leq \alpha. \tag{4.7}$$

The scaling conditions in the above theorem is stronger than that of the conditions in Theorem 4.1, and is primarily needed to show the consistency of the Gaussian multiplier bootstrap. The conditions are similar to that of Neykov et al. (2019), which arises from the high-dimensional central limit theorem in Chernozhukov et al. (2013). We now show that our testing framework in Algorithm 2 is asymptotic valid.

To this end, we denote the range of the maximum degree $\mathcal{I}$ as $[I_L, I_U]$, the default of which is $[0, d-1]$. Furthermore, let $G(\mathbf{\Delta})$ be the graph induced by the support of $\mathbf{\Delta}$, and let

$\mathcal{I}(\mathbf{\Delta}) = \mathcal{I}(G(\mathbf{\Delta}))$ be the maximum degree of the induced graph. Then, we define the parameter space as

$$\mathcal{U}_{M,s}(I_L, I_U) = \{\mathbf{\Delta} \in \mathcal{U}(M, s) | \mathcal{I}(\mathbf{\Delta}) \in [I_L, I_U]\}.$$

Moreover, denote the edge set as $E_{\text{Sig}}(\mathbf{\Delta}, \mu) = \{(j,k) | |\mathbf{\Delta}_{jk}| \geq \mu\sqrt{\log d/n}\}$, where μ is a cut-off level parameter. Coupling Theorem 4.4 with Theorems 4.1 and 4.6 in Lu et al. (2017), we have the following theorem that shows that the family-wise error is controlled at level α and that the type II error is asymptotically zero.

Corollary 4.5. Assume the conditions in Theorems 4.1 and 4.4. Given $\mathcal{I}(\mathbf{\Delta})$, the proposed inferential framework in Algorithm 2 for testing hypothesis problem (1.1) has the following property:

$$\limsup_{n\to\infty} \sup_{\mathbf{\Delta}\in\mathcal{U}_{M,s}(I_L,I_U)} \mathbb{P}_{\mathbf{\Delta}}\big(\exists k \geq \mathcal{I}(\mathbf{\Delta}) \text{ such that } H_0 \text{ is rejected}\big) \leq \alpha.$$

Moreover, for any sufficiently large μ, the expected number of type II error satisfies

$$\lim_{n\to\infty} \mathbb{E}_{\mathbf{\Delta}}\{\mathcal{I}(\mathbf{\Delta}) - d_{\text{rej}}\} \leq \mathcal{I}(\mathbf{\Delta}) - \mathcal{I}\{E_{\text{Sig}}(\mathbf{\Delta}, \mu)\},$$

where d_{rej} is as defined in Algorithm 2.

---

5. Simulation Studies

5.1 Inference on a Single Edge

In this section, we perform numerical studies to assess the performance of our proposed inferential method for testing the hypothesis:

$$H_{0,jk} : \boldsymbol{\Delta}_{jk} = 0 \qquad \text{versus} \qquad H_{0,jk} : \boldsymbol{\Delta}_{jk} \neq 0.$$

In particular, we construct the test statistic and confidence interval for $\boldsymbol{\Delta}_{jk}$ as defined in (3.5) and (3.6), respectively. We compare our proposed method with the naive method and the cross-fitting method. For the naive method, we construct the debiased estimators $\widehat{\boldsymbol{\Theta}}^d_X$ and $\widehat{\boldsymbol{\Theta}}^d_Y$ separately using (S.34) in Lu et al. (2017), and take the difference $(\widehat{\boldsymbol{\Theta}}^d_Y - \widehat{\boldsymbol{\Theta}}^d_X)$. We refer the reader to Lu et al. (2017) for details. For the cross-fitting method, in addition to obtain estimation $\widehat{\boldsymbol{\Delta}}^{(1)}$ using data from $\mathcal{D}_1$ and bias correction matrices $\mathbf{V}^{(1)}_X$ and $\mathbf{V}^{(1)}_Y$ using data from $\mathcal{D}_2$ to construct the debiased estimator $\widehat{\boldsymbol{\Delta}}^{d,(1)}$ from Algorithm 1, we also get the estimation $\widehat{\boldsymbol{\Delta}}^{(2)}$ using data from $\mathcal{D}_2$ and bias correction matrices $\mathbf{V}^{(2)}_X$ and $\mathbf{V}^{(2)}_Y$ using data from $\mathcal{D}_1$ to construct the debiased estimator $\widehat{\boldsymbol{\Delta}}^{d,(2)}$. The final debiased estimator $\widehat{\boldsymbol{\Delta}}^d$ is defined as $\widehat{\boldsymbol{\Delta}}^d = (\widehat{\boldsymbol{\Delta}}^{d,(1)} + \widehat{\boldsymbol{\Delta}}^{d,(2)})/2$.

Let $\mathcal{S} = \text{supp}(\boldsymbol{\Delta})$ be the support of the true differential graph $\boldsymbol{\Delta}$ and let $\mathcal{S}^c$ be the complement of $\mathcal{S}$. To evaluate our proposed test statistic, we report the type I error and type II error: the type I error is estimated by calculating the proportion of the constructed confidence intervals I_{jk} for $(j,k) \in \mathcal{S}^c$ that do not contain zero, and type II error is estimated as the proportion of constructed confidence intervals I_{jk} for $(j,k) \in \mathcal{S}$ that contain zero. In addition, we calculate the average length of the confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ and denote them as

$\text{Avglen}_{\mathcal{S}}$ and $\text{Avglen}_{\mathcal{S}^c}$, respectively.

To generate the data, we first construct $\mathbf{\Theta}_X$ and $\mathbf{\Delta}$. We consider two scenarios for sparse and dense $\mathbf{\Theta}_X$. For sparse and dense $\mathbf{\Theta}_X$, we generate $\mathbf{\Theta}_X$ by setting the diagonal elements equal to one and each off-diagonal element to -0.5 with probability 0.15 and 0.3, respectively. For $\mathbf{\Delta}$, to evaluate the influence of its sparsity on the performance, we define SP as the sparsity probability and the off-diagonal element of $\mathbf{\Delta}$ is set to be -1 with probability SP. Finally, we add $\rho\mathbf{I}$ to $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$ so that they are positive definite, i.e., $\rho = 0.5 + \max(0, -\lambda_{\min}(\mathbf{\Theta}_X), -\lambda_{\min}(\mathbf{\Theta}_Y))$, where $\lambda_{\min}(\mathbf{\Theta}_X)$ and $\lambda_{\min}(\mathbf{\Theta}_Y)$ are the minimum eigenvalues of $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$, respectively.

For both the sparse and dense models, we generate n training samples from the multivariate normal distribution with mean zero and covariance matrices $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$, respectively. Note that the total number of parameters to be estimated is $O(d^2)$, and throughout this section, we let $p = d^2$ for notational convenience. To implement our proposed method in Algorithm 1, we use $(n/2)$ samples to estimate $\widehat{\mathbf{\Delta}}$ in (2.1), and use the rest of the $(n/2)$ samples to estimate $\mathbf{V}_X$ and $\mathbf{V}_Y$ as in (3.4). The tuning parameters λ and λ' are set to equal $\sqrt{\log d/(n/2)}$ and $\sqrt{\log d/(n/2)}$, respectively. We then construct the test statistic and confidence interval based on (3.5) and (3.6).

Figures 2 and 3 report the type I error, type II error, and average length of the confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ for the sparse model when $n = 800$ and $p = 3600$. We can find that even in the sparse model setting, only our proposed estimator is able to control the type I error at around $\alpha = 0.05$ with minimal type II error. The Type I errors for the naive method and cross-fitting are around 0.08 and 0.06, which are larger than 0.05. In addition, as SP increases from 0.005 to 0.025, the type I error, type II error, and the average length of confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ also increase on average. Furthermore, the additional numerical results for $n = \{400, 600, 800\}$, $p = \{1600, 3600\}$, and $\text{SP} = \{0.005, 0.01, 0.015, 0.02, 0.025\}$ under the sparse

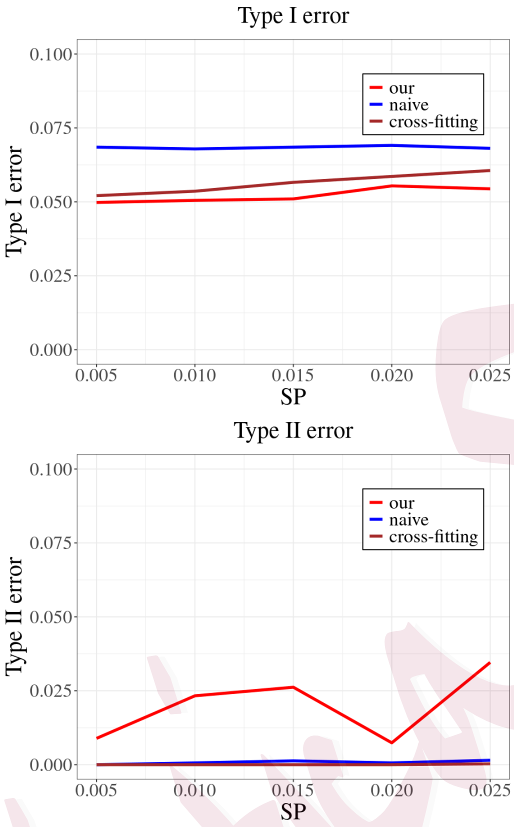


Figure 2: Type I errors and Type II errors for the sparse model when $n = 800$ and $p = 3600$, averaged over 100 replications.

and dense models are presented in the supplement.

5.2 Inference on Maximum Degree

In this section, we conduct numerical studies to assess the performance of the proposed testing procedure in Algorithm 2 for testing hypothesis problem (1.1). We consider the scenario in which $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$ are generated similar to that of Section 5.1. We generate multiple differential networks $\mathbf{\Delta}$ such that the maximum degree, m, is larger than or not larger than k, where k is a pre-specified constant in (1.1). Specifically, we construct $\mathbf{\Delta}$ with maximum degree m by

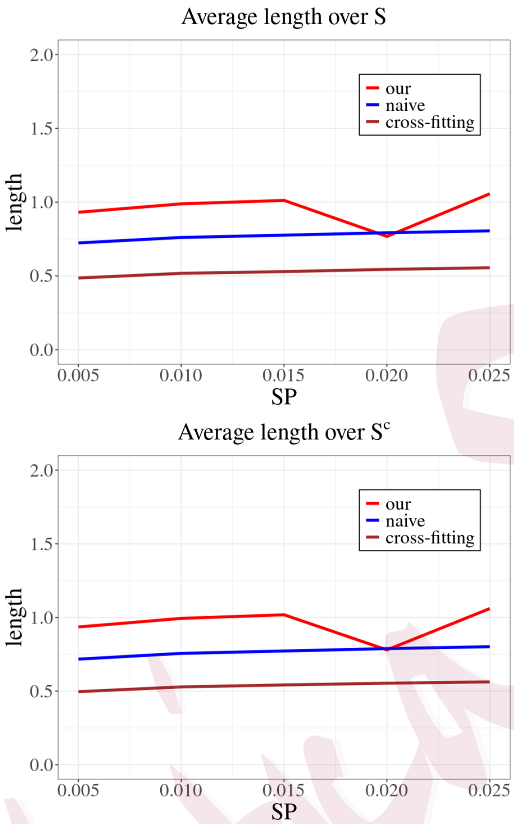


Figure 3: The average length of the confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ for the sparse model when $n = 800$ and $p = 3600$, averaged over 100 replications.

generating an m-star graph as depicted in Figure 4. The values of the corresponding edges are set to equal $\boldsymbol{\Delta}_{jk} = -1$ if there is an edge between the jth and kth variables. To calculate the type I error, we calculate the proportion of the rejected null hypothesis (1.1) when the maximum degree of $\boldsymbol{\Delta}$ is not larger than k. On the other hand, we estimate the type II error by calculating the proportion of the null hypothesis that is not rejected when the maximum degree m is in fact larger than k.

The results for $n = \{400, 600, 800\}$, $p = 3600$, $k = 5$, and $m \in \{3, 4, \ldots, 8\}$ are presented in Table 1. We see from Table 1 that our proposed method is able to control the type I error at

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0298

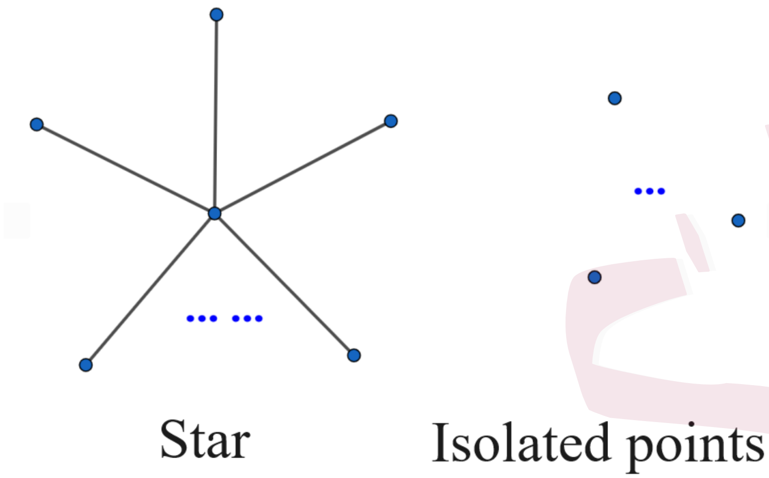


Figure 4: An m-star graph with maximum degree m and $(d-m-1)$ isolated nodes.

less than $\alpha = 0.05$, and that the type II error decreases to zero as we increase the sample size n for both scenarios when the model is sparse and dense.

Table 1: The type I and type II errors for our proposed method, calculated over 500 replications.

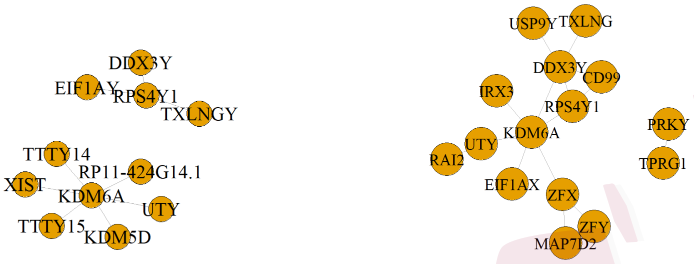
Model	Estimator	p	n	Type I Error	Type II Error
Sparse	proposed method	3600	400	0.000	0.567
Sparse	proposed method	3600	600	0.030	0.086
Sparse	proposed method	3600	800	0.019	0.004
Dense	proposed method	3600	400	0.000	0.630
Dense	proposed method	3600	600	0.016	0.198
Dense	proposed method	3600	800	0.016	0.039

5.3 Data Application

We now apply our proposed method to the Genotype-Tissue Expression (GTEx) data studied in Lonsdale et al. (2013). The GTEx project began with a 2.5-year pilot phase to study tissue-specific gene expression and regulation. The data was collected from 54 non-diseased tissue sites across 549 research subjects. We focus mainly on analyzing the breast mammary and muscle-skeletal tissues. To begin, we transform the expression counts by taking a log-transformation, and the resulting log-transformed counts for each gene is standardized to have mean zero and unit variance. We then employ the method in Gershoni and Pietrokovski (2017) to compute the sex-differential expression (SDE) protein-coding genes for each tissue.

For each tissue, we choose $d = 40$ SDE protein-coding genes with the lowest SDE test p-values for further analysis. The goal is to perform inference on the estimated sexual dimorphism differential graphs between males and females for the two aforementioned breast and muscle tissues. We are interested in testing whether there exist nodes whose interactions with other nodes vary significantly between female and male networks for each tissue. In particular, we test whether the maximum degree of the differential graph is larger than 10% of the total number of nodes, i.e., $k = 4$ since $d = 40$. Our proposed method involves selecting a tuning parameter, which we select using cross-validation. This yields $\lambda = 1.5$ for the breast mammary tissue and $\lambda = 5.0$ for the muscle skeletal tissue. We then construct the debiased estimator in (3.1) and employ Algorithm 2 to test the aforementioned hypothesis problem.

In Figure 5, we present the rejected edge set resulting from Algorithm 2 for the sexual dimorphism differential graphs for each of the two tissues. For the breast mammary tissue, there are a total of nine rejected edges in the sexual dimorphism differential graph. The maximum degree of the rejected edge set is six, and thus the null hypothesis is rejected. In Figure 5(a), we see that the X-linked lysine demethylase 6A (KDM6A) has six rejected edges, indicating that the conditional dependence relationships between KDM6A and the other nodes vary significantly between female and male. Our results coincide with that of Berletch et al. (2013) that KDM6A, a histone demethylase with female-biased expression, can be involved in



(a) Breast mammary tissue (b) Muscle skeletal tissue

Figure 5: The rejected edge set from Algorithm 2 for the sexual dimorphism differential graphs for breast mammary and muscle-skeletal tissues.

the etiology of developmental and reproduction-related effects of X chromosome anomalies.

For the muscle skeletal tissue, we see from Figure 5(b) that there are a total of 16 edges in the rejected edge set from Algorithm 2. We see that the maximum degree is six, and thus the aforementioned null hypothesis is rejected. Specifically, KDM6A has six rejected differential edges and DEAD-Box Helicase 3 Y-Linked (DDX3Y) has five rejected edge set. Our results match that of Sekiguchi et al. (2004), suggesting that both genes KDM6A and DDX3Y vary between male and female networks, where DDX3Y is in the Y chromosome region encoding a putative DEAD-box RNA helicase protein, a mutation that can result in male infertility and Sertoli cell-only syndrome.

6. Discussion

In this manuscript, we have developed a statistical inference framework and theory for testing whether there exist differential hub nodes that exhibit substantial variations in their connectivity between two networks under the assumption that the data are multivariate normal. One future direction is to explore whether the proposed method can be generalized to the non-Gaussian setting such as the exponential family graphical models (Yang et al., 2015) or a class of graphical models under the score-matching loss (Yu et al., 2020). As one of the reviewers suggested, the proposed method may also be applicable in the streaming data set-up in which data arrives sequentially, and we leave it for future work.

Supplementary Material

The Supplementary Material contains technical proofs of all the theoretical results and additional numeric results.

Acknowledgement

We thank the editor, associate editor, and two reviewers for their valuable comments that substantially improved the manuscript.

References

Berletch, J. B., Deng, X. and Disteche, C. M. (2013). Female bias in Rhox6 and 9 regulation by the histone demethylase KDM6A. *PLoS Genetics* **9** e1003489.

Cai, T., Li, H., Ma, J. and Xia, Y. (2019). Differential Markov random field analysis with an application to detecting differential microbial community networks. *Biometrika* **106** 401–416.

Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106** 594–607.

Chalhoub, N. and Baker, S. J. (2009). PTEN and the PI3-kinase pathway in cancer. *Annual Review of Pathology: Mechanisms of Disease* **4** 127–150.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41** 2786–2819.

Danaher, P., Wang, P. and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 373–397.

Drton, M. and Maathuis, M. H. (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application* **4** 365–393.

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

Gershoni, M. and Pietrokovski, S. (2017). The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC biology* **15** 1–15.

Ha, M. J., Baladandayuthapani, V. and Do, K.-A. (2015). DINGO: differential network analysis in genomics. *Bioinformatics* **31** 3413–3420.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.

Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Molecular systems biology* **8** 565.

Jackson, R. L., Hoffman, P., Pobric, G. and Ralph, M. A. L. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience* **36** 1490–1501.

Jankova, J. and Van De Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* **9** 1205–1229.

Janková, J. and van de Geer, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test* **26** 143–162.

Kim, B., Liu, S. and Kolar, M. (2021). Two-sample inference for high-dimensional markov networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **83** 939–962.

Kim, J., Do, K.-A., Ha, M. J. and Peterson, C. B. (2019). Bayesian inference of hub nodes across multiple networks. *Biometrics* **75** 172–182.

Kundu, S., Lukemire, J., Wang, Y. and Guo, Y. (2019). A novel joint brain network analysis using longitudinal Alzheimer's disease data. *Scientific Reports* **9** 1–18.

Lee, J. D. and Hastie, T. J. (2015). Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics* **25** 230–253.

Li, W., Wang, M., Zhu, W., Qin, Y., Huang, Y. and Chen, X. (2016). Simulating the evolution of functional brain networks in alzheimer's disease: exploring disease dynamics from the perspective of global activity. *Scientific reports* **6** 34156.

Liu, S., Quinn, J. A., Gutmann, M. U., Suzuki, T. and Sugiyama, M. (2014). Direct learning of sparse changes in Markov networks by density ratio estimation. *Neural computation* **26** 1169–1197.

Liu, W. et al. (2017). Structural similarity and difference testing on multiple sparse gaussian graphical models. *The Annals of Statistics* **45** 2680–2707.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. et al. (2013). The genotype-tissue expression (gtex) project. *Nature genetics* **45** 580–585.

Lu, J., Neykov, M. and Liu, H. (2017). Adaptive inferential method for monotone graph invariants. *arXiv preprint arXiv:1707.09114* .

Ma, C., Lu, J. and Liu, H. (2021). Inter-subject analysis: A partial gaussian graphical model approach. *Journal of the American Statistical Association* **116** 746–755.

Ma, J. and Michailidis, G. (2016). Joint structural estimation of multiple graphical models. *The Journal of Machine Learning Research* **17** 5777–5824.

Markowetz, F. and Spang, R. (2007). Inferring cellular networks–a review. *BMC Bioinformatics* **8** S5.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34** 1436–1462.

Mohan, K., London, P., Fazel, M., Witten, D. and Lee, S.-I. (2014). Node-based learning of multiple Gaussian graphical models. *The Journal of Machine Learning Research* **15** 445–488.

Neykov, M. and Liu, H. (2019). Property testing in high-dimensional Ising models. *The Annals of Statistics* **47** 2472–2503.

Neykov, M., Lu, J. and Liu, H. (2019). Combinatorial inference for graphical models. *The Annals of Statistics* **47** 795–827.

Neykov, M., Ning, Y., Liu, J. S. and Liu, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statistical Science* **33** 427–443.

Ren, Z., Sun, T., Zhang, C.-H. and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics* **43** 991–1026.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52** 1059–1069.

Sekiguchi, T., Iida, H., Fukumura, J. and Nishimoto, T. (2004). Human ddx3y, the y-encoded isoform of rna helicase ddx3, rescues a hamster temperature-sensitive et24 mutant cell line with a ddx3x mutation. *Experimental cell research* **300** 213–222.

Shojaie, A. (2020). Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics* e1508.

Sulaimanov, N., Kumar, S., Burdet, F., Ibberson, M., Pagni, M. and Koeppl, H. (2019). Inferring gene expression networks with hubs using a degree weighted Lasso approach. *Bioinformatics* **35** 987–994.

Tan, K. M., London, P., Mohan, K., Lee, S.-I., Fazel, M. and Witten, D. (2014). Learning graphical models with hubs. *The Journal of Machine Learning Research* **15** 3297–3331.

Tan, K. M., Lu, J., Zhang, T. and Liu, H. (2021). Estimating and inferring the maximum degree of stimulus-locked time-varying brain connectivity networks. *Biometrics* **77** 379–390.

Tan, K. M., Ning, Y., Witten, D. M. and Liu, H. (2016). Replicates in high dimensions, with applications to latent variable graphical models. *Biometrika* **103** 761–777.

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

Xia, Y. and Li, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics* **73** 780–791.

Yang, E., Ravikumar, P., Allen, G. I. and Liu, Z. (2015). Graphical models via univariate exponential family distributions. *The Journal of Machine Learning Research* **16** 3813–3847.

Yang, Z., Ning, Y. and Liu, H. (2018). On semiparametric exponential family graphical models. *The Journal of Machine Learning Research* **19** 2314–2372.

Yu, M., Gupta, V. and Kolar, M. (2020). Simultaneous inference for pairwise graphical models with generalized score matching. *Journal of Machine Learning Research* **21** 1–51.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35.

Zhao, S. D., Cai, T. T. and Li, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268.

Xin Zhou, Division of Biostatistics, University of California, Berkeley, Berkeley, California, U.S.A.

E-mail: xinzhou@berkeley.edu

Kean Ming Tan, Department of Statistics, University of Michigan, Ann Arbor, Michigan, U.S.A.

E-mail: keanming@umich.edu

Junwei Lu, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, U.S.A.

E-mail: junweilu@hsph.harvard.edu