Statistica Sinica Preprint No: SS-2024-0330	
Title	Weighted Conditional Network Testing for Multiple High-Dimensional Correlated Data Sets
Manuscript ID	SS-2024-0330
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0330
Complete List of Authors	Takwon Kim, Inyoung Kim and Ki-Ahm Lee
Corresponding Authors	Inyoung Kim
E-mails	inyoungk@vt.edu
Notice: Accepted author version.	

Statistica Sinica

Weighted Conditional Network Testing for Multiple High-Dimensional Correlated Data Sets

Takwon Kim[1], Inyoung Kim[2,*], Ki-Ahm Lee[1,3]

[1] *School of Mathematics, Statistics and Data Science, Sungshin Women's University, Seoul, Republic of Korea*

[2] *Department of Statistics, Virginia Tech University, Blacksburg, VA 24061, USA*

[3] *Department of Mathematical Science, Seoul National University, Seoul, Republic of Korea*

Abstract: Gaussian graphical models (GGMs) have been investigated to infer dependence (or network) structure among high-dimensional data by estimating a precision matrix. However, while many estimation methods for GGM have been developed, methods for testing the equality of two precision matrixes are still limited. Because testing the equality of the precision matrix depends on other given precision matrices, we develop a weighted conditional network testing for considering other given precision matrices information and also provide theoretical properties. None of the existing methods can be applied to test conditional differences when other networks are conditionally given and different. We demonstrate the advantage of our approach using a simulation study and genetic pathway analysis.

Key words and phrases: Conditional Difference, Gaussian Graphical Model, Precision Matrix.

1. Introduction

Graphical models (Friedman et al., 2008; Qiao et al., 2019) have become popular for investigating networks in various scientific fields, such as social science, neuroscience, precision medicine, and omics.

Most graphical models have been developed for multiple estimations of precision matrices but not for testing their equality. Precision matrices play a fundamental role in many high-dimensional inference problems. In the Gaussian graphical model (GGM) framework, the difference between two precision matrices characterizes the differential network, which measures the amount of change in the network between two groups. Xia et al. (2015) proposed testing the differential networks of two precision matrices and applied this test to the detection of gene-gene interaction under a GGM. Cai (2017) provided a substantive review paper summarizing recent developments in hypothesis testing for high-dimensional covariance structures, including global testing for the overall pattern of covariance structures and simultaneous testing of a large collection of hypotheses with false discovery proportion (FDP) and false discovery rate (FDR) control. Xia et al. (2018) proposed multiple testing procedures for the sub-matrices of a precision matrix, applicable to the identification of between-pathway interactions. Xia and Li (2019) adopted the matrix GGM framework to develop statistical

inferential procedures for testing the equality of individual entries of partial correlation matrices across multiple groups. Ye et al. (2021) proposed a paired test of matrix graphs to infer brain connectivity networks when groups of samples are correlated. However, none of these existing methods are able to test the difference in networks between two groups when some components of precision matrices are conditionally given and these given structures are different, nor can they conduct corresponding multiple tests in such conditional cases. For example, we are often interested in testing whether a specific gene network or pathway network is different between two groups, given networks from other pathways that are also different between two groups. Multiple high-dimensional correlated data sets, such as pathways, are highly correlated with each other due to shared genes and biological processes. Hence, when we test the equality of pathway networks between two groups, it is important to account for the other pathway network's structures. Our goal is to develop a testing procedure to provide a more accurate statistical inference. We refer to such testing as "weighted conditional differential network testing." We develop global and multiple tests under the weighted conditional differential network testing framework.

1.1 Main Contribution of This Paper

Since our global and multiple test has some similarities to that of Xia et al. (2015, 2018), we first illustrate a main contribution of our conditional testing by comparing it to that model. Figures 1a-1f are simple representations of testing in Xia et al. (2015, 2018) and our testing, where the whole square indicates the precision matrix and each sub-square represents a sub-matrix of the precision matrix. In the global test by Xia et al. (2015), they tested the equality of the two precision matrices, $\mathbf{\Omega}^1$ and $\mathbf{\Omega}^2$, as in Figure 1a; however, as Figure 1e shows, we can test the equality of two precision matrices by taking into account other sub-matrices, represented here by the areas with green checkered patterns and blue stripes. Likewise, in their testing of a given sub-matrix, Xia et al. (2018) considered other given sub-matrices, represented by the area with green checkered patterns. However, their testing is only applicable to determine whether a small off-diagonal sub-matrix is zero. The test by Xia et al. (2018) is not applicable to the equality of sub-matrices when there are two different given sub-matrices. As Figure 1e shows, we can conduct an equality test when the different sub-matrices (indicated by green checkered patterns and blue stripes) are conditionally given. Furthermore, there are significant differences between our multiple testings and the multiple testings by Xia et al. (2015, 2018). The multiple

testings by Xia et al. (2015) and ours are for cases in which components differ between two population groups. The only difference between the testing Xia et al. (2015) conducted and ours is whether it takes into account other given networks (see Figures 1b and 1f). On the other hand, in the multiple testings of Xia et al. (2018), they tested which sub-matrices were zero, as Figure 1d shows.

Therefore, although our procedure is similar to the testing procedures in Xia et al. (2015) and Xia et al. (2018), with the distinguishing feature being "weighted testing", the main difference lies in the following aspect. The existing methods, such as Xia et al. (2015) and Xia et al. (2018), are only applicable when given other networks are the "same" and also to the equal sample size case, while our method can be used when given other networks are "different" and also unequal sample size cases. Since we use the weighting scheme of the estimated precision matrix using all information, including given other networks, this weight affects the test statistics.

1.2 The Purpose of This Paper

We developed our conditional differential network testing of two population groups by incorporating "given network" and "network given other net-

works." We explain the meanings of "given network" and "network given other networks" in Figures 1g-1j. Let us consider the network between the black circle nodes, which is represented by black solid lines in Figures 1g and 1h. If there is no connection between the gray square nodes, as in Figures 1g and 1h, the network test by Xia et al. (2015) can be used to test the networks' equality. However, if the connectivity is unequal between the gray square nodes or between the black circle nodes and the gray square nodes, as in Figures 1i and 1j, the tests by Xia et al. (2015) and by Xia et al. (2018) can lose power. Although the networks between the black circle nodes are the same in Figures 1i and 1j, the blue dotted lines and red dash-dotted lines are different. We refer to the networks represented by blue dotted lines or red dash-dotted lines as "given networks" and to the networks represented by black solid lines as "networks given other networks," which is our main interest. Moreover, we refer to a network that consists of a "given network" and a "network given other networks" as a "conditional network."

We now explain why our testing is "weighted testing." Our main tests of interest are those that determine the conditional difference between two precision matrices when some components of the precisions are conditionally given and different. Although testing procedures often assume that

every diagonal element of precision matrices is the same and that sample size is balanced between two groups, our testing procedure does not require these assumptions. Hence, we consider a weighted scheme for the normalization of diagonal elements under the unbalanced sample sizes of two groups. Furthermore, we use the conditional dependency of the network to determine which edges are different. Although two networks can have the same structures, their conditional dependency can differ due to diagonal elements, which we explain in detail in Section 2.2. Thus, we need to adjust each edge dividing these diagonal elements and refer to testing for adjusted conditional networks as "weighted conditional testing." Therefore, in this paper, we develop conditional differential network testing between two population groups by incorporating a "given network" to provide more accurate statistical inference under the GGM.

1.3 The Outline of This Paper

We organize the rest of the paper as follows. In Section 2, we introduce some notations and definitions. In Section 3, we propose global and multiple testing procedures. In Section 4, we provide the asymptotic properties of testing conditional networks. In Section 5, we summarize the algorithms of our global and multiple testing procedures and present their computa-

tional complexities. In Section 6, we describe simulation studies conducted to compare the performance of our conditional test to that of alternative approaches. In Section 7, we describe the application of our conditional network test to breast cancer genetic pathway analysis. Finally, we provide concluding remarks in Section 8.

2. Testing a Conditional Network

We first define some notations in Section 2.1, describe our test hypothesis in Section 2.2, and provide the conditions for testing in Section 2.3.

2.1 Notations and Definitions

1. Sub-matrices and sub-vectors: Let $\mathbf{A} = (a_{i,j})_{1\le i\le p, 1\le j\le q} \in \mathbb{R}^{p\times q}$ and $\mathbf{v} \in \mathbb{R}^p$ be a matrix and a vector. Then, we define sub-matrices and sub-vectors as follows: $\mathbf{A}_{i,\cdot}$: the i-th row vector, $\mathbf{A}_{\cdot,j}$: the j-th column vector, $\mathbf{A}_{-i,\cdot}$: the sub-matrix of A without the i-th row, $\mathbf{A}_{\cdot,-j}$: the sub-matrix of A without the j-th column, $\mathbf{A}_{-i,-j}$: the sub-matrix of A without the i-th row and j-th column, and $\mathbf{v}_{-i}$: the sub-vector of v without the i-th element. Also, we define averages of matrices as follows: $\bar{\mathbf{A}}_{i,\cdot} = \frac{1}{q}\sum_{j=1}^{q} a_{i,j}$, $\bar{\mathbf{A}}_{\cdot,j} = \frac{1}{p}\sum_{i=1}^{q} a_{i,j}$, $\bar{\mathbf{A}}_{-i,\cdot} = \frac{1}{q}\sum_{j=1}^{q} \mathbf{A}_{-i,j}$, $\bar{\mathbf{A}}_{\cdot,-j} = \frac{1}{p}\sum_{i=1}^{q} \mathbf{A}_{i,-j}$.

2. Norms: For a vector $\mathbf{v} = (v_1, \cdots, v_p) \in \mathbb{R}^p$, we define ℓ_q norm by $\|\mathbf{v}\|_q = (\sum_{i=1}^p |v_i|^q)^{1/q}$ for $1 \leq q < \infty$ or $\|\mathbf{v}\|_\infty = \max_{1\leq i\leq p} |v_i|$. Also, we define $|\mathbf{v}|_{\min}$ by $|\mathbf{v}|_{\min} = \min_{1\leq i\leq q} |v_i|$.

3. Conditional network: In a GGM, the network structure can be represented by its nodes' precision matrix. A marginal network is represented by a precision matrix of a marginal distribution, but a conditional network is represented by a precision matrix of a conditional distribution that is equivalent to a sub-matrix of a precision matrix of joint distribution.

2.2 Test Hypotheses

Let $\mathbf{X}^1, \mathbf{X}^2 \in \mathbb{R}^p$ be two independent random vectors following multivariate normal distributions with mean $\boldsymbol{\mu}^1$, $\boldsymbol{\mu}^2$ and covariance $\boldsymbol{\Sigma}^1$, $\boldsymbol{\Sigma}^2$, respectively. That is, $\mathbf{X}^d \sim N(\boldsymbol{\mu}^d, \boldsymbol{\Sigma}^d)$ for $d = 1, 2$.

A precision matrix $\boldsymbol{\Omega}^d = (\boldsymbol{\Sigma}^d)^{-1} = (\omega^d_{i,j})_{1\leq i,j\leq p}$ represents a conditional dependence network of $\mathbf{X}^d$. Therefore, we test whether the two matrices $\boldsymbol{\Omega}^1$ and $\boldsymbol{\Omega}^2$ have the same structrure.

Let us split $\mathbf{X}^d$ into two parts: one is a variable set that consists of the p_1 variables in which we are interested, and the other is a variable set that consists of the p_2 variables in which we are not. We denote these sets as $\mathbf{X}^d_{(1,\cdots,p_1)}$

and $\mathbf{X}^d_{(p_1+1,\cdots,p_1+p_2)}$, respectively. $\mathbf{X}^d = \left(\mathbf{X}^d_{(1,\cdots,p_1)}{}^T, \mathbf{X}^d_{(p_1+1,\cdots,p_1+p_2)}{}^T\right)^T \in \mathbb{R}^{p_1+p_2}$, $\boldsymbol{\mu}^d = (\mu^d_i)_{1\le i\le p_1+p_2}$, $\boldsymbol{\Sigma}^d = (\sigma^d_{i,j})_{1\le i,j\le p_1+p_2}$.

Let $\mathbf{X}^d_I = (\mathbf{X}^d_i)_{i\in I}$ denote a sub-vector of $\mathbf{X}^d$ for any tuple I in $\{1,\cdots,p_1+p_2\}$. Also, we can break down $\boldsymbol{\Sigma}^d$ into $\boldsymbol{\Sigma}^{d,1,1} = \mathrm{Var}(\mathbf{X}^d_{I_1})$, $\boldsymbol{\Sigma}^{d,2,2} = \mathrm{Var}(\mathbf{X}^d_{I_2})$, and $\boldsymbol{\Sigma}^{d,1,2} = \boldsymbol{\Sigma}^{d,2,1^T} = \mathrm{Cov}(\mathbf{X}^d_{I_1}, \mathbf{X}^d_{I_2})$, where $I_1 = (1,2,\cdots,p_1)$ and $I_2 = (p_1+1, p_1+2, \cdots, p_1+p_2)$.

Our main interest is the network structure of $\mathbf{X}^d_{I_1}$ conditioned by $\mathbf{X}^d_{I_2}$. We denote conditional covariance and precision matrices as $\boldsymbol{\Sigma}^{d,1|2} := \mathrm{Var}(\mathbf{X}^d_{I_1}|\mathbf{X}^d_{I_2}) = \boldsymbol{\Sigma}^{d,1,1} - \boldsymbol{\Sigma}^{d,1,2}(\boldsymbol{\Sigma}^{d,2,2})^{-1}\boldsymbol{\Sigma}^{d,2,1}$, $\boldsymbol{\Sigma}^{d,2|1} := \mathrm{Var}(\mathbf{X}^d_{I_2}|\mathbf{X}^d_{I_1}) = \boldsymbol{\Sigma}^{d,2,2} - \boldsymbol{\Sigma}^{d,2,1}(\boldsymbol{\Sigma}^{d,1,1})^{-1}\boldsymbol{\Sigma}^{d,1,2}$, $\boldsymbol{\Omega}^{d,1|2} := (\boldsymbol{\Sigma}^{d,1|2})^{-1} = (\omega^{d,1|2}_{i,j})_{1\le i,j\le p_1+p_2}$, $\boldsymbol{\Omega}^{d,2|1} = (\boldsymbol{\Sigma}^{d,2|1})^{-1} = (\omega^{d,2|1}_{i,j})_{1\le i,j\le p_1+p_2}$. Then, we can rewrite $\boldsymbol{\Omega}^d$ as follows: $\boldsymbol{\Omega}^d = \begin{pmatrix} \boldsymbol{\Omega}^{d,1|2} & -(\boldsymbol{\Sigma}^{d,1,1})^{-1}\boldsymbol{\Sigma}^{d,1,2}\boldsymbol{\Omega}^{d,2|1} \\ -(\boldsymbol{\Sigma}^{d,2,2})^{-1}\boldsymbol{\Sigma}^{d,2,1}\boldsymbol{\Omega}^{d,1|2} & \boldsymbol{\Omega}^{d,2|1} \end{pmatrix}$ with $(\boldsymbol{\Sigma}^{d,2,2})^{-1}\boldsymbol{\Sigma}^{d,2,1}\boldsymbol{\Omega}^{d,1|2} = \boldsymbol{\Omega}^{d,2|1}\boldsymbol{\Sigma}^{d,2,1}(\boldsymbol{\Sigma}^{d,1,1})^{-1}$. We use $\boldsymbol{\Omega}^{d,1|2}$ to refer to a network given other networks, $d = 1, 2$, and $(-(\boldsymbol{\Sigma}^{d,1,1})^{-1}\boldsymbol{\Sigma}^{d,1,2}\boldsymbol{\Omega}^{d.2|1}, \boldsymbol{\Omega}^{d,2|1})$ to a given network. We denote them as $\boldsymbol{\Omega}^d_{I_1,I_1}$ and $(\boldsymbol{\Omega}^d_{I_1,I_2}, \boldsymbol{\Omega}^d_{I_2,I_2})$, respectively, where $\boldsymbol{\Omega}^d_{I_k,I_l} = (\omega^d_{i,j})_{i\in I_k, j\in I_l}$ is a sub-matrix of $\boldsymbol{\Omega}^d$. And the matrix $\boldsymbol{\Omega}^d$ represents a conditional network. We explained these networks in Section 1.2. We test whether the two matrices $\boldsymbol{\Omega}^1_{I_1,I_1}$ and $\boldsymbol{\Omega}^2_{I_1,I_1}$ (or $\boldsymbol{\Omega}^{1,1|2}$ and $\boldsymbol{\Omega}^{2,1|2}$) have the same structure. In fact, we want to investigate whether the structure of two networks are indistinguishable, so the absolute value of the difference between two $\omega^1_{i,j}$ and $\omega^2_{i,j}$ adjusted by $\omega^d_{i,i}$ and $\omega^d_{j,j}$ is outside of our

interest. Our hypotheses are then $H_0 : \max_{i,j\in I_1} \left| \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} - \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}} \right| = 0$ vs. $H_1 : \max_{i,j\in I_1} \left| \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} - \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}} \right| \neq 0$.

Due to the adjustment of $\omega^d_{i,j}$ by $\omega^d_{i,i}$ and $\omega^d_{j,j}$, we refer to our test as weighted conditional testing. Furthermore, if the null hypothesis H_0 is rejected, we need to determine which edges are significantly different, as well. Therefore, we conduct multiple tests with FDR control using the following hypotheses: $H_{0,i,j} : \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} = \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}}$ vs. $H_{1,i,j} : \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} \neq \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}}$ for $1 \leq i < j \leq p_1$. Let $\{\mathbf{X}^d_k = (\mathbf{X}^d_{k,1}, \cdots, \mathbf{X}^d_{k,p_1+p_2})^T \in \mathbb{R}^{p_1+p_2} : k = 1, \cdots, n_d\}$ be a set of independent n_d copies of $\mathbf{X}^d$ for $d = 1, 2$ and denote a data matrix as $(\mathbf{X}^d_{k,i})_{1\leq k\leq n_d, 1\leq i\leq p_1+p_2} = (\mathbf{X}^d_1, \cdots, \mathbf{X}^d_{n_d})^T \in \mathbb{R}^{n_d\times(p_1+p_2)}$.

Let $\boldsymbol{\beta}^d_i = (\beta^d_{1,i}, \cdots, \beta^d_{p_1+p_2-1,i})^T \in \mathbb{R}^{p_1+p_2-1}$ be the regression coefficients of $\mathbf{X}^d_{k,i}$ to $\mathbf{X}^d_{k,-i}$, then $\mathbf{X}^d_{k,i} = \beta^d_{0,i} + \mathbf{X}^d_{k,-i}\boldsymbol{\beta}^d_i + \varepsilon^d_{k,i}$, $\boldsymbol{\beta}^d_i = (\boldsymbol{\Sigma}^d_{-i,-i})^{-1}\boldsymbol{\Sigma}^d_{-i,i} = -(\omega^d_{i,i})^{-1}\boldsymbol{\Omega}^d_{-i,i}$ and $\beta^d_{0,i} = \mu^d_i - \boldsymbol{\mu}^d_{-i}\cdot\boldsymbol{\beta}^d_i$, where $\varepsilon^d_{k,i} = (\mathbf{X}^d_{k,i} - \mu^d_i) - \left(\mathbf{X}^d_{k,-i} - \boldsymbol{\mu}^d_{-i}\right)\boldsymbol{\beta}^d_i$. Because $\varepsilon^d_{k,i}$ can be represented by $\varepsilon^d_{k,i} = (\omega^d_{i,i})^{-1}\boldsymbol{\Omega}^d_{i,\cdot}(\mathbf{X}^d_k - \boldsymbol{\mu}^d)$, $\varepsilon^d_{k,i} \sim N(0, (\omega^d_{i,i})^{-1})$, and $\varepsilon^d_{k,i} \perp \mathbf{X}^d_{k,-i}$. Denoting $r^d_{i,j} := \mathrm{Cov}(\varepsilon^d_{k,i}, \varepsilon^d_{k,j})$ for $1 \leq i, j \leq p_1$, we then have $r^d_{i,j} = \frac{\omega^d_{i,j}}{\omega^d_{i,i}\omega^d_{j,j}}$, $\omega^d_{i,j} = \frac{r^d_{i,j}}{r^d_{i,i}r^d_{j,j}}$, and $\frac{r^d_{i,j}}{\sqrt{r^d_{i,i}r^d_{j,j}}} = \frac{\omega^d_{i,j}}{\sqrt{\omega^d_{i,i}\omega^d_{j,j}}}$.

2.3 Regularity Condition

We developed our testing procedure under some conditions described in the supplementary materials (Conditions **??**-**??**).

3. Global and Multiple Tests

In this section, we describe the global and multiple weighted conditional testing procedures in Sections 3.1 and 3.2, respectively.

3.1 Global Weighted Conditional Testing Procedure

Let $\hat{\varepsilon}^d_{k,i}$ and $\tilde{r}^d_{i,j}$ be the estimators of $\varepsilon^d_{k,i}$ and $r^d_{i,j}$, which we must obtain from the following forms: $\hat{\varepsilon}^d_{k,i} = \mathbf{X}^d_{k,i} - \bar{\mathbf{X}}^d_{\cdot,i} - (\mathbf{X}^d_{k,-i} - \bar{\mathbf{X}}^d_{\cdot,-i})\hat{\boldsymbol{\beta}}^d_i$ and $\tilde{r}^d_{i,j} = \frac{1}{n_d}\sum_{k=1}^{n_d} \hat{\varepsilon}^d_{k,i}\hat{\varepsilon}^d_{k,j}$. Xia et al. (2015) suggested a bias-corrected estimator $\hat{r}^d_{i,j}$ of $r^d_{i,j}$, $\hat{r}^d_{i,j} = -\tilde{r}^d_{i,j} - \tilde{r}^d_{i,i}\hat{\beta}^d_{i,j} - \tilde{r}^d_{j,j}\hat{\beta}^d_{j-1,i}$ for $1 \le i < j \le p_1$ and $\hat{r}^d_{i,i} = \tilde{r}_{i,i}$ for $1 \le i \le p_1$. Therefore, the intuitive estimator of $\omega^d_{i,j}$ and weighted $\omega^d_{i,j}$ (denoted as $\omega^{d,w}_{i,j}$) are $\hat{\omega}^d_{i,j} = \frac{\hat{r}^d_{i,j}}{\hat{r}^d_{i,i}\hat{r}^d_{j,j}}$ and $\hat{\omega}^{d,w}_{i,j} = \frac{\hat{\omega}^d_{i,j}}{\sqrt{\hat{\omega}^d_{i,i}\hat{\omega}^d_{j,j}}} = \frac{\hat{r}^d_{i,j}}{\sqrt{\hat{r}^d_{i,i}\hat{r}^d_{j,j}}}$ for $1 \le i \le j \le p_1$. To normalize $\hat{\omega}^d_{i,j}$, we need to estimate $\hat{\omega}^d_{i,j}$'s variance. By denoting the following notations, we can estimate $\hat{\omega}^d_{i,j}$'s variance as $\tilde{U}^d_{i,j} = \frac{r^d_{i,j} - U^d_{i,j}}{\sqrt{r^d_{i,i}r^d_{j,j}}}$ and $U^d_{i,j} = \frac{1}{n_d}\sum_{k=1}^{n_d}(\varepsilon^d_{k,i}\varepsilon^d_{k,j} - E[\varepsilon^d_{k,i}\varepsilon^d_{k,j}])$ for $1 \le i \le j \le p_1$. Then, under Conditions **??**, **??**, and **??**, by Lemma **??** of the supplementary materials, we can show the following result:

$$\max_{(i,j)\in\tilde{A}} \left|\hat{\omega}^{d,w}_{i,j} - \tilde{U}^d_{i,j}\right| = O_p\{(\log p_1/n_d)^{1/2}\} \max_{(i,j)\in\tilde{A}} |r^d_{i,j}| + o_p\{(n_d \log p_1)^{-1/2}\} \qquad (3.1)$$

for any subset $\tilde{A} \subset A = \{(i,j) : 1 \le i \le j \le p\}$.

Remark 1. Using equation (3.1), we can approximate the estimator $\hat{\omega}^d_{i,j}$ of $\frac{\omega^d_{i,j}}{\sqrt{\hat{\omega}^d_{i,i}\hat{\omega}^d_{j,j}}}$ by the random variable $\tilde{U}^d_{i,j}$, derived from the multivariate normal distribution $(\varepsilon^d_{k,i})_{i=1,\cdots,n_d}$. However, while this approximation requires a sufficient convergence rate, equation (3.1) only provides a relatively slow convergence rate of $O_p\{(\log p_1/n_d)^{1/2}\}$. This necessitates Condition **??**, which indicates that the set A_τ of indices (i,j), where $|r^d_{i,j}|$ is not small, has a negligible size. As a result, for indices in A^c_τ, we have $\max_{(i,j)\in A^c_\tau}|r^d_{i,j}| \leq (\log p_1)^{-1-\tau}$. By using the equation (3.1), we can achieve a faster convergence rate for $\hat{\omega}^d_{i,j}$, specifically $o_p\{(n_d \log p_1)^{-1/2}\}$, when considering the indices in A^c_τ. For the remaining indices in A_τ, we can disregard the set because A_τ is a small set. Therefore, using (3.1), $\hat{\omega}^d_{i,j}$ can be approximated by the random variable $\tilde{U}^d_{i,j}$ with an appropriate convergence rate. Since the distribution of $\tilde{U}^d_{i,j}$ is much easier to compute compared to that of $\hat{\omega}^d_{i,j}$, we can derive the asymptotic distribution of $\hat{\omega}^d_{i,j}$ through (3.1)

Therefore, our normalized estimator of the difference of weighted precision elements is $\Delta_{i,j} = \frac{\hat{\omega}^{1,w}_{i,j} - \hat{\omega}^{2,w}_{i,j}}{(\hat{\theta}^1_{i,j}+\hat{\theta}^2_{i,j})^{1/2}}$ for $1 \leq i \leq j \leq p_1$, $\hat{\theta}^d_{i,j} = \frac{1}{n_d}\left(1 + (\hat{\beta}^d_{i,j})^2 \frac{\hat{r}^d_{i,i}}{\hat{r}^d_{j,j}}\right)$ for $1 \leq i < j \leq p_1$, and $\hat{\theta}^d_{i,i} = \frac{2}{n_d}$ for $1 \leq i \leq p_1$, where $\hat{\theta}^d_{i,j}$ is an estimator of $\theta^d_{i,j} := \mathrm{Var}(\tilde{U}^d_{i,j})$. Therefore, our test statistic is $M_n = \max_{1\leq i\leq j\leq p_1}(\Delta_{i,j})^2$, where $n = \min(n_1, n_2)$. Let $\tilde{\Delta}_{i,j} := \Delta_{i,j} - \frac{1}{(\hat{\theta}^1_{i,j}+\hat{\theta}^2_{i,j})^{1/2}}\left(\frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} - \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}}\right)$ and $\tilde{M}_n := \max_{1\leq i\leq j\leq p_1}\left(\tilde{\Delta}_{i,j}\right)^2$. Because $\tilde{\Delta}_{i,j}$s are normalized, we can assume

that $\tilde{\Delta}_{i,j}$s follow the standard normal distribution. If $\tilde{\Delta}_{i,j}$s are independent, we can heuristically obtain the following probability: $P(\tilde{M}_n - 4\log p_1 + \log\log p_1 \leq t) \approx \exp(-(8\pi)^{-1/2}e^{-t/2}) =: \Phi_M(t)$, where $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ for any $t \in \mathbb{R}$ and sufficiently large p_1. From the heuristic computation, we can define the rejection region and hypothesis test for global testing as follows: rejection region $R = \{x \in \mathbb{R} : x \geq q_\alpha + 4\log p_1 - \log\log p_1\}$ and hypothesis test $\Psi_\alpha = I_R(M_n)$, where I_R is an indicator function of R and $q_\alpha = -\log(8\pi) - 2\log|\log(1-\alpha)|$ for $0 < \alpha < 1$.

3.2 Multiple Weighted Conditional Testing Procedure

When the global null hypothesis is rejected, we need to determine which edges are significant. If we set the size of each tests ($H_{0,i,j}$ vs $H_{1,i,j}$) as α, the family-wise type I error approaches 1 as the number of tests approaches infinity because Family-wise type I error $= E\left[\bigcup_{1\leq i<j\leq p_1} \text{Reject } H_{0,i,j} \middle| \bigcap_{1\leq i<j\leq p_1} H_{0,i,j}\right]$ $\approx 1-(1-\alpha)^{p_1(p_1-1)/2} \approx 1$. Therefore, we need to consider the FDP and FDR, as well. With the following notations, $\mathcal{I} = \{(i,j) : 1 \leq i < j \leq p_1\}$, $\mathcal{I}_0 = \left\{(i,j) \in \mathcal{I} : \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} = \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}}\right\}$, $N(t) = \{(i,j) \in \mathcal{I} : |\Delta_{i',j'}| \geq t\}$, $N_0(t) = \{(i,j) \in \mathcal{I}_0 : |\Delta_{i',j'}| \geq t\}$, we can define FDP and FDR as $\text{FDP}(t) = \frac{N_0(t)}{\max(N(t),1)}$ and $\text{FDR}(t) = E[\text{FDP}(t)]$, where t is a threshold that determines where $H_{0,i,j}$ is rejected.

For multiple tests, we must choose $t(>0)$ while $\text{FDP}(t) \leq \alpha$ for $0 < \alpha < 1$. We denote this t as t_0 and obtain t_0 using $t_0 = \inf\{0 \leq t \leq 2(\log p_1)^{1/2} : \text{FDP}(t) \leq \alpha\}$. Because $\text{FDP}(t)$ is unknown, we should approximate either $\text{FDP}(t)$ or $N_0(t)$. By using the fact that $\mathbf{\Omega}^d$s are sparse and $|\Delta_{i,j}|$ is similar to $Z \sim N(0,1)$ for $(i,j) \in \mathcal{I}_0$, we can approximate $N_0(t) \approx 2[1-\Phi(t)]|\mathcal{I}_0| \approx [1-\Phi(t)](p_1^2-p_1)$. By using this approximation, we can then estimate t_0 as $\hat{t}_0 = \inf\left\{0 \leq t \leq 2(\log p_1)^{1/2} : \frac{[1-\Phi(t)](p_1^2-p_1)}{\max(N(t),1)} \leq \alpha\right\}$, where $\Phi(t) = \int_{-\infty}^{t} \varphi(x)dx$ is the standard normal cumulative distribution function. Therefore, we can perform the multiple tests with threshold $\hat{t}_0$. For $1 \leq i < j \leq p_1$, our decision rule is Reject $H_{0,i,j}$ if and only if $|\Delta_{i,j}| \geq \hat{t}_0$.

4. Asymptotic Properties of Weighted Conditional Network Test

In this section, we describe our testing procedures' asymptotic properties. We derive the asymptotic distribution of test statistic and asymptotic size under the null hypotheses, asymptotic power, and properties of FDR in Sections 4.1, 4.2, and 4.3, respectively.

4.1 Asymptotic Distribution of Test Statistic

We derive the asymptotic distribution of test statistic M_n and an asymptotic size α of the global test under Conditions 1-4 and describe them in Theorem

1.

Theorem 1. *Under Conditions* **??**, **??**, **??**, *and* **??**, *for any* $t \in \mathbb{R}$,

$$P(\tilde{M}_n - 4\log p_1 + \log\log p_1 \le t) \to \Phi_M(t) \quad \text{as } n, p_1 \to \infty. \tag{4.1}$$

In particular, under Conditions **??**, **??**, **??**, **??**, *and* H_0, *for any* $t \in \mathbb{R}$,

$$P(M_n - 4\log p_1 + \log\log p_1 \le t) \to \Phi_M(t) \quad \text{as } n, p_1 \to \infty.$$

Therefore, the global test Ψ_α *has an asymptotic size* α: $\lim_{n,p_1\to\infty} E[\Psi_\alpha|H_0] = \lim_{n,p_1\to\infty} P(M_n \ge q_\alpha + 4\log p_1 - \log\log p_1|H_0) = 1 - \exp(-(8\pi)^{-1/2}e^{-q_\alpha/2}) = \alpha$, *where* $q_\alpha = -\log(8\pi) - 2\log|\log(1-\alpha)|$ *for* $0 < \alpha < 1$.

4.2 Asymptotic Power of Test Ψ_α

For the power of the test Ψ_α, we need some pairs of distinct precision matrices $(\mathbf{\Omega}^1, \mathbf{\Omega}^2)$. We define a class of precision matrices as $\mathcal{U}(c) := \mathcal{U}(c, (\log p_1)^{1/2})$ $= \left\{ (\mathbf{\Omega}^1, \mathbf{\Omega}^2) : \sup_{1\le i\le j\le p_1} \frac{1}{(\theta^1_{i,j} + \theta^2_{i,j})^{1/2}} \left| \frac{\omega^1_{i,j}}{\sqrt{\omega^1_{i,i}\omega^1_{j,j}}} - \frac{\omega^2_{i,j}}{\sqrt{\omega^2_{i,i}\omega^2_{j,j}}} \right| \ge c(\log p_1)^{1/2} \right\}$. The smaller c is, the closer $(\mathbf{\Omega}^1, \mathbf{\Omega}^2)$ in $\mathcal{U}(c)$ are. Theorem 2 states that $c = 4$ is a sufficient distance to reach the asymptotic power of 1. We demonstrate how the power of the test based on our weighted statistic converges to 1 as p_1 and n approach infinity in Theorem 2.

Theorem 2. *Under Conditions* **??**, **??**, **??**, *and* **??**, *if* $K^{-1} < \frac{n_1}{n_2} < K$ *for some* $K > 0$, *the power of the test* Ψ_α *approaches 1 as* n_1, n_2, *and* p_1 *approach infinity in* $\mathcal{U}(4)$. *That is,* $\inf_{(\boldsymbol{\Omega}^1,\boldsymbol{\Omega}^2)\in\mathcal{U}(4)} P(\Psi_\alpha = 1) \to 1 \quad$ *as* $n, p_1 \to \infty$.

Theorem 2 guarantees that the convergence of the power even when the dimension p_1 is sufficiently larger than the sample sizes n_1 and n_2, with $\log p_1 = o(n^{1/5})$ where $n = \min\{n_1, n_2\}$. Thus we need the scale assumption $\log p_1 = o(n^{1/5})$ for p_1 and n in Condition **??**.

If $f(p_1)$ diverges to infinity faster than $(\log p_1)^{1/2}$, then Theorem 2 holds for any small $c > 0$. That means if $\log p_1 = o((f(p_1))^2)$, then $\inf_{(\boldsymbol{\Omega}^1,\boldsymbol{\Omega}^2)\in\mathcal{U}(c,f(p_1))} P(\Psi_\alpha = 1) \to 1 \quad$ as $n_1, n_2, p_1 \to \infty$. Therefore, in Theorem 3, we can check that $(\log p_1)^{1/2}$ is the optimal rate to reach asymptotic power of 1.

Theorem 3. *Suppose* $\log p_1 = o(n_1)$ *and* $K^{-1} < \frac{n_1}{n_2} < K$ *for some* $K > 0$. *For any* $0 < \alpha, \gamma < 1$ *with* $\alpha + \gamma < 1$, *there exists* $c_0 > 0$ *such that* $\inf_{(\boldsymbol{\Omega}^1,\boldsymbol{\Omega}^2)\in\mathcal{U}(c_0)} \sup_{T_\alpha\in\mathcal{T}_\alpha} P(T_\alpha = 1) \le 1 - \gamma$, *for sufficiently large* n, p_1, *where* $\mathcal{T}_\alpha$ *is the set of all* α*-level tests.*

Theorem 3 says that if $f(p_1) = o((\log p_1)^{1/2})$, then $\inf_{(\boldsymbol{\Omega}^1,\boldsymbol{\Omega}^2)\in\mathcal{U}(c,f(p_1))} P(\Psi_\alpha = 1) \not\to 1$ as $n, p_1 \to \infty$ for any $c > 0$.

4.3 Properties of Asymptotic FDR

We demonstrate the FDP and FDR of the test based on our weighted statistic converge to $\frac{\alpha q_0}{q}$ as p_1 and n approach infinity in Theorem 4.

Theorem 4. *Suppose that* $|A_\tau \cap \mathcal{I}_0| = o(p_1^\nu)$ *for any* $\nu > 0$. *Assume that* $q_0 = |\mathcal{I}_0| \geq cp_1^2$ *for some* $c > 0$. *Let* $q = (p_1^2 - p_1)/2$. *Then, under Conditions* **??**, **??**, *and* **??** *with* $p_1 \leq cn^r$ *for some* $c > 0$ *and* $r > 0$, *we have* $\frac{\mathrm{FDP}(\hat{t}_0)}{\alpha q_0/q} \to 1$, *in probability, and* $\frac{\mathrm{FDR}(\hat{t}_0)}{\alpha q_0/q} \to 1$, *as* $n, p_1 \to \infty$.

According to Theorem 4, the FDP and FDR converge even if the dimension p_1 is much larger than the sample sizes n_1 and n_2, as long as $p_1 \leq cn^r$ for some constants $c > 0$ and $r > 0$, where $n = \min\{n_1, n_2\}$. Unlike Theorem 2, this result ensures convergence even in cases of unbalanced sample sizes. That is, the convergence of the FDP and FDR can be expected even when $n_1 \gg n_2$ or $n_2 \gg n_1$.

Remark 2. We assume the comparability of n_1 and n_2, i.e., $K^{-1} < \frac{n_1}{n_2} < K$ for some $K > 0$ in Theorems 2 and 3, which is also assumed in Xia et al. (2015). However, in Theorems 1 and 4, we admit the $n_1 \gg n_2$ or $n_2 \gg n_1$ case.

Remark 3. Our method and Xia et al. (2018)'s method appear similar in that both consider subnetworks. However, a significant distinction lies

in the conditions related to the size of the subnetwork. Their focus is on identifying which subnetworks differ when the number of subnetworks ($\frac{p}{p_1}$) is very large. In contrast, our model aims to identify which edges differ within a subnetwork when the subnetwork itself is large, leading us to assume a relationship between p_1 and n. Therefore, the primary difference between Xia et al. (2018) and our work lies in whether the focus is on differences between subnetworks or within a subnetwork.

5. Computational Algorithms for Weighted Conditional Testing

We program from estimation to testing using Python code. The summaries of algorithms we used are described in Algorithms **??**-**??** of the supplementary materials. We compare the computational costs by comparing the nonconditional test (NCT) by Xia et al. (2015) with our weighted conditional testing. The weighted conditional test (WCT) has computational costs in terms of $O((p_1+p_2)^4+(p_1+p_2)^3 n+(p_1+p_2)n^2)$, which is the same as the NCT when $p_2 = O(p_1)$. When the number of nodes $p_1 = p_2 = p$ and sample sizes $n_1 = n_2 = 50, 100$, the computational plots are displayed in Figures **??** and **??** of the supplementary materials.

6. Simulation Studies

In this section, we conduct a simulation study to investigate the performance of our weighted conditional network testing. We compare the NCT by Xia et al. (2015) with WCT in terms of size, power, and FDR. Additionally, we include the non-weighted conditional test (NWCT) to distinguish between conditional testing and the effect of weighting. We conducted simulations for various scenarios as follows: 1) same sample size sample size and same dimension of network, 2) different sample sizes and same dimension of network, 3) same sample size and different dimensions of network, and 4) different diagonal elements. We only describe the first scenario, and the results for the scenario are provided in Table 1 of the main document and Tables **??**-**??** of the supplementary materials, with detailed settings and results for scenarios 2), 3), and 4) included in Section **??** and Tables **??**-**??** of the supplementary materials.

We consider four types of precision matrices $\mathbf{\Omega}^{(m)}$, $m = 1, \ldots, 4$. In the first scenario (Table 1 of the main document and Tables **??**-**??** of the supplementary materials), we set $n = n_1 = n_2 = 100$ and $\tilde{p} = p_1 = p_2 = 50, 100, 200$. For each combination of n, $\tilde{p}$, and m, we conduct 1000 simulations and study the performance of our approach when $n \approx \tilde{p}$, $n > \tilde{p}$, and $n < \tilde{p}$. For size, power, and FDR, we consider two situations: one has

the same given networks and the other has different given networks. The simulated precision matrix, size, power, and FDR are described in Sections 6.1-6.4.

6.1 Types of Simulated Precision Matrix

We generate four types of sparse precision matrices, $\mathbf{\Omega}^{(m)}$, $m = 1, \ldots, 4$, described as follows:$(\mathbf{\Omega}_1, \mathbf{\Omega}_2)$

1. $\mathbf{\Omega}^{(1)} = (\omega_{i,j})$ as a $(p_1 + p_2) \times (p_1 + p_2)$ matrix with $\omega_{i,i} = 1$, $\omega_{i+1,i} = \omega_{i,i+1} = 0.6$, $\omega_{i+2,i} = \omega_{i,i+2} = 0.36$, $\omega_{i,j} = \omega_{j,i} = 1$ for $(i, j) \in A$, otherwise $\omega_{i,j} = 0$, where A is a set of randomly chosen p_1 elements in $\{(i, j) : 1 \leq i \leq p_1, p_1 + 1 \leq j \leq p_1 + p_2\}$. This precision matrix is a pentadiagonal matrix.

2. $\mathbf{\Omega}^{(2)} = (\omega_{ij})$ as a scale-free network generated by the Barabasi-Albert algorithm. For $\mathbf{\Omega}^{(2)}$, the number of new edges connected to new nodes is 2 and $\omega_{i,i} = 1$, $\omega_{i,j} \sim \text{Uniform}(-1, 1)$, if i and j are connected.

3. $\mathbf{\Omega}^{(3)} = (\omega_{i,j})$ as $\omega_{i,i} = 1$, $\omega_{i,j} = \omega_{j,i} \sim 0.8 \times \text{Binomial}(1, 0.05)$, for $1 \leq i < j \leq p_1 + p_2$.

4. $\mathbf{\Omega}^{(4)} = (\omega_{i,j})$ as $\omega_{i,i} = 1$, $\omega_{i,j} = \omega_{j,i} = 0.5$ for $1 \leq k \leq \frac{p_1+p_2}{10}, i = 10(k-1)+1$, and $i + 1 \leq j \leq i + 9$.

From $\Omega^{(m)}$ for $m = 1, 2, 3, 4$, we generate pairs of precision matrices (Ω^1, Ω^2) according to the simulation purpose of each. Simple examples of $\Omega^{(m)}$ for $p_1 = p_2 = 10$ and $m = 1, 2, 3, 4$ are displayed in Figure **??** of the supplementary materials. For each $\Omega^{(m)}$, the submatrix $\Omega^{(m)}_{I_1,I_1}$ and $(\Omega^{(m)}_{I_1,I_2}, \Omega^{(m)}_{I_2,I_2})$ represent shaded and unshaded parts in Figure **??** of the supplementary materials.

6.2 Asymptotic Size of Test

To investigate the size of the tests, we consider two situations. One has the same given networks in which $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) = (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$, and the other has a different given networks in which $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) \neq (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$.

6.2.1 Global Test with Same Given Networks

To generate two precision matrices under H_0 and have the same given networks, we set $\mathbf{\Omega}^1 = \mathbf{\Omega}^2 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$ and $m = 1, \ldots, 4$, where D is a diagonal matrix whose diagonal elements are uniformly distributed between 0.5 and 2.5 independently and $\delta = \left|\lambda_{\min}(\mathbf{\Omega}^{(m)})\right| + 0.05$. As a result, we have the same given networks, $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) = (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$, under H_0. To calculate the empirical size of the test, we generate $1,000$ copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$.

6.2.2 Global Test with Different Given Networks

We generate some differences in the given networks by changing some elements. Let $\tilde{\mathbf{V}} = (v_{i,j})$, a symmetric $(p_1 + p_2) \times (p_1 + p_2)$ matrix that $v_{i+p_1,j+p_1} = v_{i+p_1,j+p_1} = v_{k,j+p_1} = v_{j+p_1,k} \sim \text{Uniform}\left(-\frac{\omega_{\max}}{3}, \frac{\omega_{\max}}{3}\right)$, where $\omega_{\max} = \max_i \omega_{i,i}^{(m)}$ and i, j, and k are uniformly distributed for $1 \leq i < j \leq p_2$ and $1 \leq k \leq p_1$. Then, we set new $\mathbf{\Omega}^1$ and $\mathbf{\Omega}^2$: $\mathbf{\Omega}^1 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$ and $\mathbf{\Omega}^2 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{V}} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$, where $\delta = \min(\left|\lambda_{\min}(\mathbf{\Omega}^{(m)})\right|, |\lambda_{\min}(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{V}})|) + 0.05$. As a result, we have the different given networks: $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) \neq (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$ under H_0. To calculate the empirical size of the test, we generate 1000 copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$.

6.3 Asymptotic Power of Test

In the same way, we evaluated the size of the test, and we also consider two situations to investigate the power of the test.

6.3.1 Global Test with Same Given Networks

To generate two precision matrices under H_1 and have the same given networks, we use another sparse symmetric matrix $\tilde{\mathbf{W}}$, of which each element is 0 except eight elements whose location and value are uniformly distributed in $1 \leq i, j \leq p_1$ and $(-2\sqrt{\log p_1/n} \cdot \omega_{\max}, \sqrt{\log p_1/n} \cdot \omega_{\max}) \cup (\sqrt{\log p_1/n} \cdot$

$\omega_{\max}, 2\sqrt{\log p_1/n} \cdot \omega_{\max})$, where $\omega_{\max} = \max_i \omega_{ii}^{(m)}$, respectively. We set $\mathbf{\Omega}^1 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$ and $\mathbf{\Omega}^2 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{W}} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$, where $\delta = \min(\left|\lambda_{\min}(\mathbf{\Omega}^{(m)})\right|, |\lambda_{\min}(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{W}})|) + 0.05$. As a result, we have the same given networks, $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) = (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$, under H_1. We generate 1000 copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$.

6.3.2 Global Test with Different Given Networks

To emphasize the role of given networks, we generate some differences in the given network by changing some elements under H_1. We add $\tilde{\mathbf{V}} = (v_{i,j})$, defined in Section 6.2.2. Then, we have $\mathbf{\Omega}^1$ and $\mathbf{\Omega}^2$ as the following: $\mathbf{\Omega}^1 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$ and $\mathbf{\Omega}^2 = \frac{1}{1+\delta}\mathbf{D}^{1/2}\left(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{W}} + \tilde{\mathbf{V}} + \delta\mathbf{I}\right)\mathbf{D}^{1/2}$, where $m = 1, \ldots, 4$ and $\delta = \min(\left|\lambda_{\min}(\mathbf{\Omega}^{(m)})\right|, |\lambda_{\min}(\mathbf{\Omega}^{(m)} + \tilde{\mathbf{W}} + \tilde{\mathbf{V}})|) + 0.05$. As a result, we have the different given networks, $(\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) \neq (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})$, under H_1. We generate 1000 copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$.

6.4 Empirical FDR of Test

We also investigate FDR for two situations when $\alpha = 0.1$ and 0.2.

6.4.1 Multiple Test with Same Given Networks

We consider a setting similar to that of Section 6.3.1, except that the number of nonzero elements of $\tilde{\mathbf{W}}$ is 40. We generate 100 copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$. For the multiple test, we choose κ_d in (**??**). For $\kappa = 1/20, 2/20, \cdots, 39/20, 40/20$, we choose
$\hat{\kappa}_0 = \arg\min \sum_{l=1}^{10} \left(\frac{\sum_{1\le i<j\le p_1} I\left\{|\Delta_{i,j}| \ge \Phi^{-1}(1-l[1-\Phi\{(\log p_1)^{1/2}\}]/10)\right\}}{l p_1(p_1-1)[1-\Phi\{(\log p_1)^{1/2}\}]/10} - 1 \right)^2$.

6.4.2 Multiple Test with Different Given Networks

We use the same setting as in Section 6.3.2, except that the number of nonzero elements of $\tilde{\mathbf{W}}$ is 40. We generate 100 copies of $\mathbf{X}^1 \sim N(0, (\mathbf{\Omega}^1)^{-1})$ and $\mathbf{X}^2 \sim N(0, (\mathbf{\Omega}^2)^{-1})$.

6.5 Simulation Results

Empirical Size: The empirical size with the same given networks for the first scenario (same sample size and same dimension of network) is summarized in the first part of Table 1. As shown in Table 1, the type I errors of the WCT and the NWCT are smaller than those of the NCT. The difference between the two type I errors of WCT and NCT is larger than 1% except for $(\tilde{p}, m) = (200, 1)$, $(100, 2)$, and $(200, 2)$. Additionally, the difference in type I error between WCT and the NWCT is mostly within

0.5%. We also observe that the WCT and the NWCT have smaller type I errors than NCT does, even in the same given networks case. The empirical size with different given networks for the first scenario is in the second part of Table 1. When we consider the situation of the different given networks, the WCT and the WNCT outperform the NCT. The type I error of the WCT (and the NWCT) remains below 5% except for $(\tilde{p}, m) = (50, 1)$ and $(50, 3)$ ($(\tilde{p}, m) = (100, 2)$ and $(50, 3)$ for the NWCT). In the $(\tilde{p}, m) = (50, 1)$, $(50, 3)$, $(100, 2)$ cases, the type I errors of the WCT and NWCT are still close to the significance level of $\alpha = 5\%$. Conversely, the NCT generally exhibits a type I error exceeding 10%. When $(\tilde{p}, m) = (50, 1)$ and $(50, 3)$, the type I error is much larger than 20%. The slight difference between the given networks makes the NCT work much worse, while the WCT and the NWCT perform well.

Empirical Power: The tests' empirical power with the same given networks for the first scenario (same sample size and same dimension of network) is summarized in the third part of Table 1. As we see in Table 1, the almost power of the WCT is greater than 99%, whereas the powers of the NWCT and the NCT are smaller than 99%. Also, the difference in power between the WCT and the NCT is greater than 1%, while the NWCT is more than 0.5% lower than the NCT. The WCT has more power than the

NWCT and the NCT does, even in the same given networks case. The tests' empirical power with different given networks for the first scenario is in the last part of Table 1. The power of the WCT is greater than 95% except for the $(\tilde{p}, m) = (200, 2)$ case. However, the power values for the NWCT and the NCT is smaller than 95%.

Empirical FDR: The empirical FDR and power with the same given networks for the first scenario (same sample size and same dimension of network) are summarized in Table **??** of the supplementary materials. As observed in Table **??** of the supplementary materials, there is no significant difference in the empirical FDRs among the WCT, the NWCT, and the NCT, but the WCT consistently demonstrates higher power than others. For each α, the FDRs for the WCT, the NWCT, and the NCT are approximately α. Additionally, the power is consistently higher for $\alpha = 0.2$ compared to $\alpha = 0.1$. The power of the WCT is around $45\% \sim 50\%$ when $\alpha = 0.1$ and $53\% \sim 58\%$ when $\alpha = 0.2$. In contrast, the power of the NWCT (resp. NCT) is around $35\% \sim 43\%$ (resp. $38\% \sim 45\%$) when $\alpha = 0.1$ and $40\% \sim 48\%$ (resp. $45\% \sim 50\%$) when $\alpha = 0.2$. Although all of the methods decrease the power compared to the global test, our method performs better than others. The empirical FDR and the power with different given networks for the first scenario are in Table **??** of the supplementary materi-

als. The results with different given networks is similar to that of the given networks case.

Summary: We have developed two approaches (WCT and NWCT) to improve the testing method by considering both conditional testing and adjusting the diagonal elements. Conditional testing effectively reduces Type I error but still lowers the power. Adjusting the diagonal elements increases the power. As a result, NWCT, which applies only conditional testing, shows significantly lower Type I errors than NCT but also lower power. Meanwhile, WCT, which incorporates both conditional testing and adjustment of diagonal elements, achieves both lower Type I error and higher power. We note that although, in Theorem 2, we showed that the power of all three methods converges to 1 as n, $p_1 \to \infty$ in the simulation results, the power of NWCT is lower than that of WCT and NCT because n and p_1 are not sufficiently large compared to the conditions of Theorem 2. Therefore, our simulation results suggest that the WCT outperformed the NCT, in terms of type I error, power of the global test, and power of the multiple test, whereas two methods did not have any clear superiority in terms of FDR. In general, the WCT tends to outperform the NCT as the given networks have different structures.

7. Breast Cancer Genetic Pathways

Our main question of interest concerns the equality of two precision matrices of genetic pathways between White and non-White female racial groups when one or all other pathways are conditionally given. The human breast cancer data set was collected from the University of Texas M.D. Anderson Cancer Center (Shi et al., 2010), which contains $22,283$ gene expression measurements from 176 White patients and 102 non-White patients. We collect 25 pathways (Xu et al., 2019) whose ID, name, and the numbers of genes were listed in Table **??** of the supplementary materials. We then tested all pairs-wise comparisons, so we conducted $\binom{25}{2}$ multiple comparisons. Due to the page limitation, we only provided small sets of results. See other sets of results in Section S5 of the supplementary materials.

We compare our approach with the NCT by Xia et al. (2015). Using our approach, we also consider both the WCT and the non-weighted conditional test (NWCT), and then we compare them with the NCT of a pathway given other pathways. Here, the NWCT is a test in which we apply only the conditional data, without the weighted procedure of the WCT. We compare the network of the White group ($d = 1$) to that of the non-White group ($d = 2$). We have $n_1 = 176$ (the number of White patients) and $n_2 = 92$ (the number of non-White patients). Here, $p_1 =$ the number of

genes in the pathway given other pathways, and p_2 = the number of genes in the given pathway and not in the testing pathway. For $k = 1, \cdots, n_d$, $i = 1, \cdots, (p_1 + p_2)$, and $d = 1, 2$, let $\mathbf{X}^1_{k,i}$ = k-th White patient's i-th gene data in the pathway given other pathways and $\mathbf{X}^2_{k,i}$ = k-th non-White patient's i-th gene data in the pathway given other pathways for $1 \leq i \leq p_1$ and $\mathbf{X}^1_{k,i}$ = k-th White patient's $(i - p_1)$-th gene data in the given pathway and $\mathbf{X}^2_{k,i}$ = k-th non-White patient's $(i - p_1)$-th gene data in the given pathway for $p_1 + 1 \leq i \leq p_1 + p_2$.

We display selected results in Figures **??**-**??** and Tables **??**-**??** of the supplementary materials and Figure 2 of the main document. In the bar-shaped graphs displayed in Figures **??**, **??**, **??**, **??**, **??**, and **??** of the supplementary materials, the x-axis and y-axis represent $|\Delta_{i,j}|$ (or $|\Delta^{NCT}_{i,j}|$) and the index of edges, respectively. We note that $|\Delta_{i,j}| = |\hat{\omega}^{1,w}_{i,j} - \hat{\omega}^{2,w}_{i,j}|/(\hat{\theta}^1_{i,j} + \hat{\theta}^2_{i,j})^{1/2}$ is the statistic for the multiple test in the WCT, and $|\Delta^{NCT}_{i,j}| = |\hat{\omega}^1_{i,j} - \hat{\omega}^2_{i,j}|/(\hat{\theta}^{1,NCT}_{i,j} + \hat{\theta}^{2,NCT}_{i,j})^{1/2}$ is the statistic for the multiple test in the NCT, where $\hat{\theta}^{d,NCT}_{i,j} = \hat{\theta}^d_{i,j}/(\hat{r}^d_{i,i}\hat{r}^d_{j,j})$ for $d = 1, 2$. The red vertical and green vertical lines are $\hat{t}_0$ for $\alpha = 0.1$ and $\alpha = 0.2$, respectively. The insignificant edges for $\alpha = 0.1, 0.2$ are plotted on the right side of the figure, while significant edges are plotted on the left side of the figure with the edge names on the line. If a significant edge is significant under $\alpha = 0.1$, the edge is colored red; otherwise, it is colored

green. In the circle-shaped graphs displayed in Figures **??**, **??**, and **??** of the supplementary materials, the name of the gene in the testing network is displayed in the circle. Two connected genes are significant. The left side of the circle-shaped graph is the result for the WCT, and the right side of the circle-shaped graph is the result for the NCT. In the other circle-shaped graphs displayed in Figure 2 of the main document and Figures **??** and **??** of the supplementary materials, the solid black line represents two genes whose edge is significant in both the WCT and the NCT, the dash-dotted orange line represents an edge that is significant only in the WCT, and the dashed purple line represents an edge that is significant only in the NCT. The left side of the circle-shaped graph is the result with significance level of $\alpha = 0.1$, and the right side of the circle-shaped graph is the result with significance level of $\alpha = 0.2$. In Tables **??**-**??** of the supplementary materials, we summarize the significant edges for each test (NCT, NWCT, and WCT) with significance level of 0.2. In each row, we place the common significant edges.

In the case of testing pathway 113 conditioned by pathway 137, as we display in Figures **??**, **??**, **??**, and **??** and Table **??** of the supplementary materials, the NCT detects more significant edges than the WCT and NWCT do. In fact, the global test for the WCT and NWCT do not reject the global

null hypothesis. Thus, only the NCT detects the three significant edges. The main difference between WCT/NWCT and NCT is that WCT/NWCT can be used even when the given networks are different, whereas NCT is only applicable when the two given networks are identical. Therefore, this outcome can be interpreted as the outcome when the given two networks are different.

For the case of testing pathway 137 conditioned by pathway 717, as we see in Table **??** and Figures **??**, **??**, **??**, and **??** of the supplementary materials, the WCT detects more significant edges than the NCT does. In the result of the WCT, there are eight significant edges, whereas the NCT detects only one significant edge. Although the numbers of significant edges are different, an edge FASLG − MAP2K7 is a common significant edge detected by both the WCT and the NCT. In this case, our method, WCT, also reveals more findings than NWCT. WCT employs a weighted precision matrix adjusting the effect of each gene, while NWCT does not adjust this effect. Therefore, this outcome is likely due to the significant effect of each gene.

In the case of testing pathway 1206 conditioned by pathway 750, as we see in Figure 2 of the main document and Figures **??**, **??**, and **??** and Table **??** of the supplementary materials, the WCT detects 17 significant

edges (under $\alpha = 0.1$: 0, $\alpha = 0.2$: 17), but the NCT detects 11 significant edges (under $\alpha = 0.1$: 7, $\alpha = 0.2$: 11). Although the numbers of significant edges are different, all significant edges under $\alpha = 0.1$ for the NCT are also significant edges under $\alpha = 0.2$ for the WCT. Also, significant edges under $\alpha = 0.2$ for the NCT, except an edge PSMD2 $\times$ CLASP1, are also significant edges under $\alpha = 0.2$ for the WCT.

As we can see in the results, our proposed algorithm, WCT, reveals more and fewer findings compared to other methods (NCT or NWCT), depending on the situation. However, based on the ability of our weighted test, which can detect the difference between two networks when given two networks are different, adjusting individual gene effect, simulation results, and relevant literature on breast cancer, we conclude that the interactions identified by WCT are significantly more reliable than those detected by NCT and NWCT.

Since WCT employs the weighted conditional precision matrices, WCT can be applied to more general situations where two given networks are different, or the effects of each gene are significantly distinct. Moreover, as demonstrated in Tables **??**, **??**, **??**, **??**, **??**, **??**, **??**, and **??** (simulation results for multiple testing) of the supplementary materials, the WCT shows approximately 10% higher power than NWCT and NCT, while all

three tests exhibit comparable false discovery rates (FDR). Furthermore, some results obtained using our method have been linked to breast cancers through literature review, although not all. These reviews are documented in Section S5 of the supplementary materials. Based on this evidence, we conclude that the interactions identified by WCT are substantially more reliable than those detected by NCT and NWCT.

8. Discussion

In this paper, we proposed a WCT for the equality of two conditional networks between two population groups. Because the difference of dependence structures depends on other components of dependence, simultaneous inferences play an important role in having greater power and in avoiding incorrect inferences. We also provided multiple testing procedures and theoretical properties of this testing procedure.

We showed that the power of our WCT is about 1.4 times higher than that of NCT in both global and multiple tests. Especially when we conducted multiple tests, we observed that the power values were not sufficiently large compared with the global test. This result from multiple tests may be affected by the dependency among multiple tests. In future research, we further investigate how dependency structures among multiple

testings affect the power of multiple testings.

Note that our WCT was developed for the equality of two conditional networks two population groups. However, there could be more than two population groups. Thus, it is worthwhile to extend our testing procedures for multiple population groups. We also noted that our approach was developed under the GGM. However, this Gaussian assumption is often not applicable to real application. Hence, it is worthwhile to develop a testing procedure under the non-Gaussian case, such as one using the Copula method. Last but not least, although some significant gene connectivity associated to breast cancer were identified, they need to be further validated biologically.

Supplementary Materials

Technical proofs, additional tables and figures referenced are available in a separate file for Supplementary materials of this paper.

Acknowledgments

The research of Ki-Ahm Lee and Takwon Kim is supported by the National Research Foundation of Korea (NRF-2020R1A2C1A01006256) and (RS-2024-00351151), respectively. We are also grateful to the reviewers for their

valuable suggestions and constructive input.

Bibliography

Cai, T. (2017). Global testing and large scale multiple testing for high dimensional covariance structure. *Annual Rev Stat Appl* (4), 423–426.

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Qiao, X., S. Guo, and G. M. James (2019). Functional graphical models. *Journal of the American Statistical Association 114*, 211–222.

Shi, L. et al. (2010). The microarray quality control (maqc)-ii study of common practices for the development and validation of microarray-based predictive models. *Nature biotechnology 28*, 827–38.

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

Xia, Y., T. Cai, and T. Cai (2018). Multiple testing of submatrices of a precision matrix with application to identification of between pathway interactions. *Journal of the American Statistical Association 113*(521), 328–339.

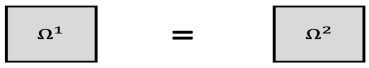
Xia, Y. and L. Li (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statistica Sinica* (1), 303–328.

Xu, Y., I. Kim, and R. Carroll (2019). A hybrid omnibus test for generalized semiparametric single-index models with high dimensional covariate sets. *Biometrics 75*(3), 757–767.

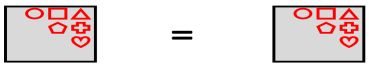
Ye, Y., Y. Xia, and L. Li (2021). Paried test of matrix graphs and brain connectivity analysis. *Biostatistics 22*(2), 402–420.

Table 1: Empirical sizes (%) and powers (%) for testing the equality of two precision matrices $(\mathbf{\Omega}_1, \mathbf{\Omega}_2)$ when same or different given networks using the weighted conditional test (WCT), nonweighted conditional test (NWCT), and nonconditional test (NCT); $(\mathbf{\Omega}_1, \mathbf{\Omega}_2)$ are generated under H_0 using $\mathbf{\Omega}^{(m)}$, $m = 1, \ldots, 4$, with the number of nodes of the network given other networks $p_1 = \tilde{p}$ = 50, 100, 200, the number of nodes of the given network $p_2 = \tilde{p} = 50, 100, 200$, sample size $n_1 = n_2 = 100$, and significance level $\alpha = 0.05$. For $m = 1, 2, 3, 4$, $\mathbf{\Omega}^{(1)}$ is a pentadiagonal matrix, $\mathbf{\Omega}^{(2)}$ is a scale-free network, $\mathbf{\Omega}^{(3)}$ is a symmetric matrix whose upper off-diagonal elements are from *Binomial* distribution, and $\mathbf{\Omega}^{(4)}$ is a symmetric matrix whose randomly assigned off-diagonal elements is 0.5.

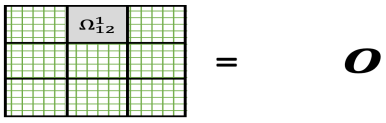
	$\mathbf{\Omega}^{(1)}$			$\mathbf{\Omega}^{(2)}$			$\mathbf{\Omega}^{(3)}$			$\mathbf{\Omega}^{(4)}$		
Empirical sizes (%) with the same given networks $\left((\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) = (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})\right)$												
$\tilde{p}$	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT
50	3.5	3.1	5.7	2.2	3.5	4.2	2.2	3.7	4.6	2.4	2.8	3.6
100	2.7	2.7	3.7	3.0	3.2	3.4	2.3	2.9	3.7	2.9	2.9	4.4
200	3.6	3.7	4.1	2.2	2.2	2.3	1.9	2.4	3.1	2.2	2.9	3.4
Empirical sizes (%) with different given networks $\left((\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) \neq (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})\right)$												
$\tilde{p}$	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT
50	5.5	4.7	21.1	3.5	4.1	13.4	6.2	5.3	21.1	4.5	4.5	18.0
100	3.9	3.9	17.1	4.6	5.2	12.5	3.9	3.7	19.6	4.2	4.1	18.0
200	3.2	3.0	15.3	3.0	3.0	10.2	2.5	2.7	19.5	2.9	3.2	13.1
Empirical powers (%) with the same given networks case $\left((\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) = (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})\right)$												
$\tilde{p}$	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT
50	99.3	97.4	98.2	98.4	94.4	95.4	99.6	97.9	98.4	99.7	98.1	98.7
100	99.9	96.4	97.0	99.0	93.7	94.7	99.9	97.3	97.8	99.3	96.1	97.3
200	99.4	92.7	94.4	98.2	89.6	91.6	99.9	93.1	94.0	98.7	91.6	93.1
Empirical powers (%) with different given networks $\left((\Omega^1_{I_1,I_2}, \Omega^1_{I_2,I_2}) \neq (\Omega^2_{I_1,I_2}, \Omega^2_{I_2,I_2})\right)$												
$\tilde{p}$	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT	WCT	NWCT	NCT
50	97.6	89.7	94.2	96.1	88.5	91.0	98.1	91.6	94.1	97.4	90.2	92.9
100	97.1	88.7	92.0	95.6	85.9	89.6	98.4	87.9	93.5	96.7	85.5	90.1
200	97.8	82.4	86.7	93.7	73.9	80.6	98.3	80.3	86.3	95.2	80.0	85.9



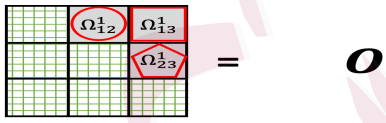
(a) Global test of Xia et al. (2015)



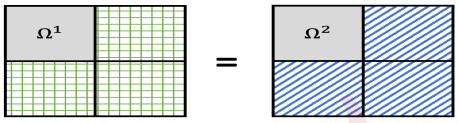
(b) Multiple test of Xia et al. (2015)



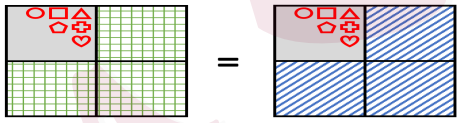
(c) Test a given sub-matrix of Xia et al. (2018)



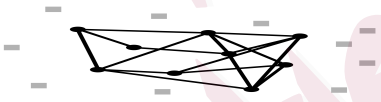
(d) Multiple test in Xia et al. (2018)



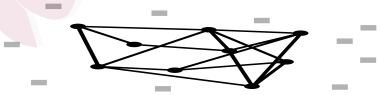
(e) Global test of weighted conditional differential network testing



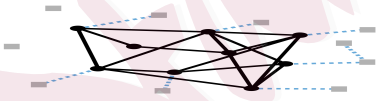
(f) Multiple test of weighted conditional differential network testing



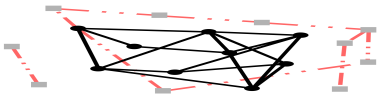
(g)



(h)



(i)



(j)

Figure 1: (a) and (b) are the global and multiple test for Xia et al. (2015). (c) and (d) are the global and multiple test for Xia et al. (2018). (e) and (f) are the global and multiple test for our weighted global and multiple test. (g) and (h) have the same given networks and networks given other networks. (i) and (j) have the same networks given other networks and different given networks.

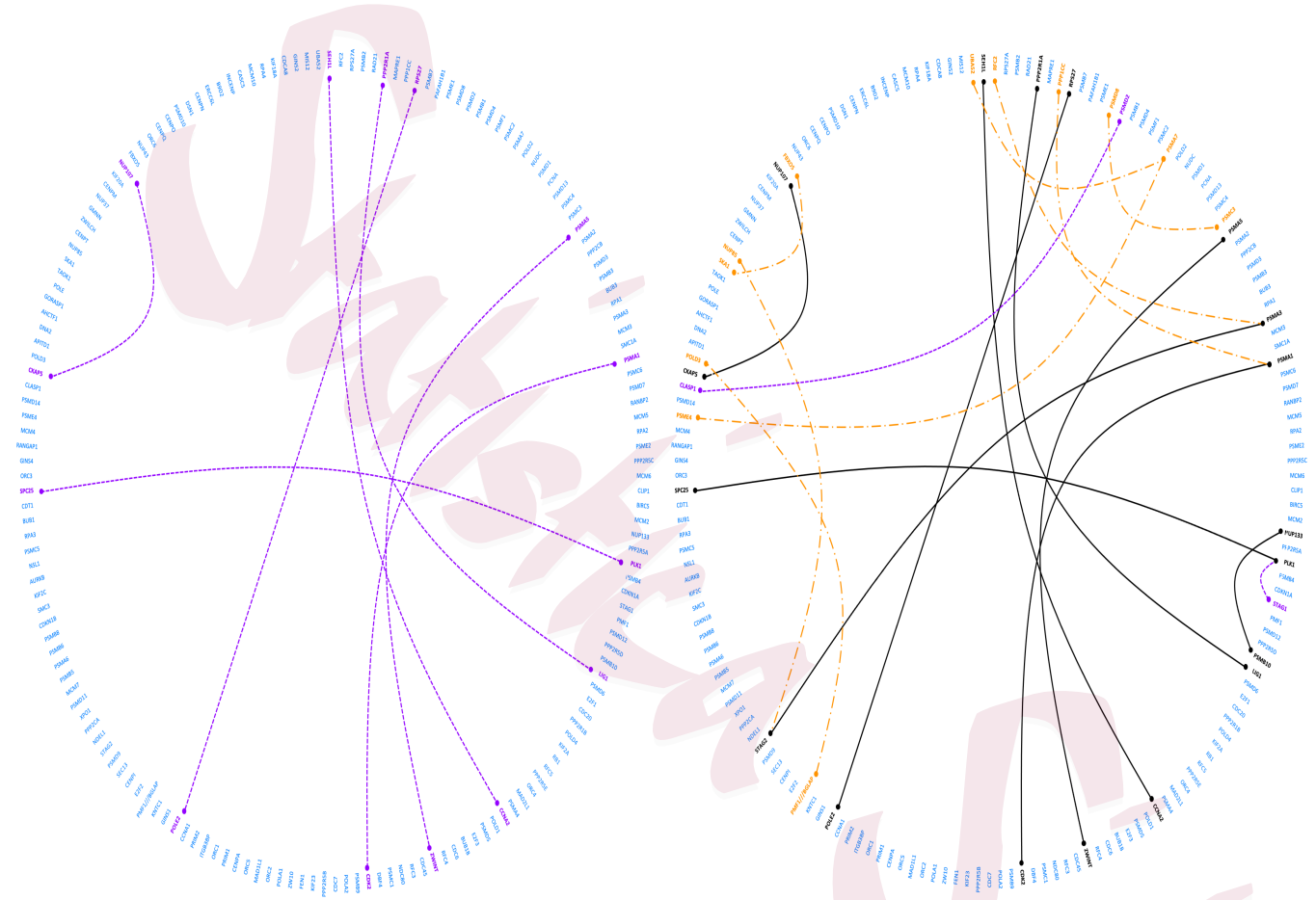


Figure 2: $\Delta_{i,j}$ for breast cancer genetic pathway with the testing pathway: 1206, the conditional pathway: 750