Supplementary Material for "Universally Consistent Tests for the Graph of a Gaussian Graphical Model"

Thien-Minh Le[1], Ping-Shou Zhong[2] and Chenlei Leng[3]

[1] *University of Tennessee at Chattanooga*

[2] *University of Illinois Chicago*

[3] *University of Warwick*

The supplementary materials contain detailed technical conditions, complete proofs for all the theorems in the main text, and examples of precision structures that satisfy condition (C1). Additionally, we include further simulation studies, a proposed data-driven procedure for selecting tuning parameters, and insights into the real data analysis.

1. Technique assumptions

Because $\boldsymbol{S}_n$ is invariant to $\boldsymbol{\mu}$, without loss of generality, we assume that $\boldsymbol{\mu} = 0$ in the rest of the proof. We replace $\boldsymbol{S}_n$ by $\boldsymbol{V}_n = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}}/n$ because the terms related to $\bar{\mathbf{X}}$ in $\boldsymbol{S}_n$ are small order of $\boldsymbol{V}_n$. We assume the following multivariate model (Bai and Saranadasa, 1996; Chen et al. , 2010) for the random variable $\mathbf{X}$, which includes Gaussian distribution as a special case:

Assumption (D1): Assume $\mathbf{X} = \boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is a p-dimensional constant vector, $\boldsymbol{\Gamma}$ is a $m \times p$ constant matrix with $m \geq p$ so that $\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} = \boldsymbol{\Sigma}$, and $\mathbf{Z} = (Z_1, \cdots, Z_m)^{\mathrm{T}}$ satisfies $E(\mathbf{Z}) = 0$, $\mathrm{var}(\mathbf{Z}) = \mathbf{I}_m$ and $E(Z_i^4) = 3 + \kappa$ for a finite constant Δ. Additionally, Z_i has a uniformly bounded 8th moment for $i = 1, \cdots, m$, and for any integers $l_v \geq 0$ such that $\sum_{v=1}^{8} l_v \leq 8$, we have $E(Z_{i_1}^{l_1}, Z_{i_2}^{l_2} \cdots Z_{i_q}^{l_q}) = E(Z_{i_1}^{l_1})E(Z_{i_2}^{l_2}) \cdots E(Z_{i_q}^{l_q})$ whenever $1 \leq i_1, \cdots, i_q \leq m$ are distinct indices.

Assumption (D2): Recall $\mathbf{X}_1 = (X_{11}, \cdots, X_{1p})^{\mathrm{T}}$. Assume $E\{\exp(\eta X_{1j}^2)\} \leq C$ for $j = 1, \cdots, p$ and some finite constants η and C.

2. Examples of precision structures satisfying Condition (C1)

The main idea of the proof of Theorem 1 is to approximate the test statistics $\hat{D}_n$ by a modified version of D_n. Denote $A = \{(i,j), 1 \leq i,j \leq p\}$ be the set of all pairs of indices that $\hat{D}_n$ will be maximized over and write $\hat{D}_n = \max_{(i,j)\in A} \hat{D}_{ij}^2$. Let $A_0 = \{(i,j), \omega_{ij}^* \neq 0\}$ be the set of indices that excluding the sparse set of non-zeros in $\mathbf{\Omega}^*$. Let $A_1 = \cup_{i=1}^{p}\{(i,k) : \lim_{p\to\infty} s_0\sigma_{ik} \neq 0, \forall (i,k) \notin A_0\}$ be the set of indices that variables (i,k) having covariance larger than $1/s_0$. Define $B_0 = A_0 \cup A_1$ as the union of A_0 and A_1. For convenience, denote , $\hat{D}_n^* = \max_{(i,j)\in A} \hat{D}_{ij}^{*2}$, $\hat{D}_{n1}^* = \max_{(i,j)\in A/A_0} \hat{D}_{ij}^{*2}$, $\hat{D}_{n2}^* = \max_{(i,j)\in A/B_0} \hat{D}_{ij}^{*2}$, where $\hat{D}_{ij}^{*2} = (\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)^2/\theta_{ij}$. We will show that the distribution of $\hat{D}_n$ can be approximated by the distribution of $D_{n2} = \max_{(i,j)\in A/B_0} D_{ij}^2$. Notice that for technical details in the proof we need $\mathrm{card}(A/B_0) = p^2\{1 + o(1)\}$, which followed by $||\mathbf{\Sigma}^*||_1 \leq C_1$, for some $C_1 > 0$.

We now provide some examples of classes of precision matrices that satisfying Condition (C1) and their corresponding forms of A/B_0.

Example 1. (Polynomial decay) Let $\mathbf{\Omega}^* = (\omega_{ij}^*)_{p\times p}$ be a banded polynomial precision matrix defined by $\omega_{ij}^* = 1/(1 + |i-j|)^\lambda$, for $|i-j| < s_0$, $s_0 = o(\sqrt{n}), \lambda \geq 2$, and $\omega_{ij}^* = 0$ otherwise. Lemma 4 shows that $|\mathbf{\Sigma}_{i,j}| \leq C*(1+|i-j|)^{-\lambda}$.Therefore, $||\mathbf{\Sigma}^*||_1 < C_1$, for some C_1. Furthermore, we also have, $\sigma_{jk} = O(1/s_0^\lambda) = o(1/s_0)$ for (j,k) such that $|j-k| \geq s_0$. So $A_1 = \cup_{i=1}^{p}\Big\{(i,k) : \max\{|k-(i+s0-1)|, |k-(i-s0+1)|\} \leq h, k \notin [i-s_0+1, i+s_0-1]\Big\}$, where $h = s_0$.

Meanwhile $B_0 = \cup_{i=1}^{p}\Big\{(i,k), \max\{|k-(i+s0-1)|, |k-(i-s0+1)|\} \le h\Big\}$. Therefore, $A/B_0 = \cup_{i=1}^{p}\Big\{(i,k), \min\{|k-(i+s0-1)|, |k-(i-s0+1)|\} \ge h\Big\}$. As a result, $\text{card}(A/B_0) = p^2\{1+o(1)\}$.

Example 2. (Exponential decay) Let $\mathbf{\Omega}^* = (\omega_{ij}^*)_{p\times p}$ be a precision matrix decaying at an exponential rate so that $\omega_{ij}^* = \theta^{|i-j|}$ for $|i-j| < s_0, s_0 = o(\sqrt{n}), 0 < \theta < 1$, and $\omega_{ij}^* = 0$ otherwise. Lemma 5 shows that $\sigma_{jk} = O\{\exp(-\beta|j-k|)\}$, for some $0 < \beta < -\log\theta$. Therefore, $||\mathbf{\Sigma}^*||_1 < C_1$, for some $C_1 > 0$. So $A/B_0 = \cup_{i=1}^{p}\{(i,k), \min(|k-(i+s0-1)|, |k-(i-s0+1)|) \ge h\}$, where $h = s_0^{\gamma}$, for some small $\gamma > 0$. As a result, $\text{card}(A/B_0) = p^2\{1+o(1)\}$.

Example 3. (Banded) Assume that precision matrix $\mathbf{\Omega}^*$ has a banded structure such that $\omega_{ij}^* = 0$, for $|i-j| \ge s_0$ where $s_0 = o(\sqrt{n})$. Then

$$A/B_0 = \cup_{i=1}^{p}\{(i,k), \min(|k-(i+s0-1)|, |k-(i-s0+1)|) \ge h\}$$

where $h = s_0^{1+\gamma}$, for some small $\gamma > 0$. Lemma 6 implies that $|\sigma_{ij}| \le C\lambda_1^{|i-j|}$, for $0 < \lambda_1 = (\sqrt{\text{cond}(\mathbf{\Omega}^*)} - 1)/(\sqrt{\text{cond}(\mathbf{\Omega}^*)} + 1) < 1$, where $\text{cond}(\mathbf{\Omega}^*) = ||\mathbf{\Omega}^*||||\mathbf{\Omega}^{*-1}||$. Therefore $||\mathbf{\Sigma}^*||_1 < C_1$, for some $C_1 > 0$, and $\sigma_{jk} < \lambda_1^{2|j-k|/s_0} = \lambda_1^{2s_0^{\gamma}} = o(1/s_0)$ on A/B_0. We also have that $\text{card}(A/B_0) = p^2\{1+o(1)\}$.

Example 4. (Factor model) Assume that $\mathbf{\Omega}^*$ is generated from a factor model. Specifically, $\mathbf{\Omega}^* = \mathbf{I}_p + \sum_{i=1}^{k}\alpha_i \boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}$ where $\mathbf{I}_p$ is the identity matrix and for each $i = 1,\ldots,k$ ($k \in \mathbb{Z}^+$), $\alpha_i \in \mathbb{R}$, $\boldsymbol{u}_i$ is a p-dimensional vector in $\mathbb{R}^p$ such that $||\mathbf{\Omega}^*||_1 = O(1)$. Lemma 7 shows that $A/B_0 = A/A_0$, since $\sigma_{jk} = 0$ for $(j,k) \in A/B_0$. As a result, $||\mathbf{\Sigma}^*||_1 < C_1$ for some $C_1 > 0$ and $\text{card}(A/B_0) = p^2\{1+o(1)\}$.

3. Proof of Lemmas

Proof of Lemma 1 in the main text: (1) We have

$$\begin{aligned}\mathrm{var}(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i) &= \mathrm{var}(\mathbf{e}_j^{\mathrm{T}}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\mathbf{w}_i^*)/n^2 = \mathrm{var}(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_i^*)/n \\ &= \mathrm{E}(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{e}_j)/n - (\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*)^2/n.\end{aligned}$$

We write $\mathbf{X}_1$ as $\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}$, where $\mathbf{Z}$ is a p-dimensional standard normally distributed random vector and $\boldsymbol{\Sigma}^* = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}$. Then we have

$$\begin{aligned}\mathrm{E}(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{e}_j) &= \mathrm{E}(\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j) \\ &= \mathrm{E}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{T}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}) \\ &= \mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) + 2\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) \\ &= \mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j + 2(\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*) \\ &= \omega_{ii}^*\sigma_{jj}^* + 2(\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*).\end{aligned}$$

Since $\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^* = 0, \mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^* = 0$, for $1 \le i \ne j \le p$. This yields $\mathrm{var}(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i) = \omega_{ii}^*\sigma_{jj}^*/n$. (2) If $1 \le i = j \le p$, we have $\mathbf{e}_j^{T}\boldsymbol{\Sigma}^*\mathbf{w}_i^* = 1$ and $\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^* = 1$. So $\mathrm{var}(\mathbf{e}_i^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* - \mathbf{e}_i^{\mathrm{T}}\mathbf{e}_i) = (\omega_{ii}^*\sigma_{ii}^* + 1)/n$.

If $\mathbf{X}_1$ follows a multivariate model as in Bai and Saranadasa (1996) and Chen et al. (2010) and $\boldsymbol{\Sigma}^* = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}$, then we have

$$\begin{aligned}&\mathrm{E}(\mathbf{e}_j^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{e}_j) \\ &= \mathrm{E}(\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j) \\ &= \mathrm{E}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{T}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}) \\ &= \mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) + 2\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) + \Delta\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_i^*\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}} \circ \boldsymbol{\Gamma}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) \\ &= \mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j + 2(\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^*) + \Delta(\mathbf{w}_i^{*\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{e}_j)^2\end{aligned}$$

$$= \omega_{ii}^* \sigma_{jj}^* + 2(\mathbf{w}_i^{*\mathrm{T}} \mathbf{\Sigma}^* \mathbf{e}_j \mathbf{e}_j^{\mathrm{T}} \mathbf{\Sigma}^* \mathbf{w}_i^*) + \Delta(\mathbf{w}_i^{*\mathrm{T}} \mathbf{\Sigma}^* \mathbf{e}_j)^2.$$

Similar to the normal distribution cases, when $i \neq j$, we have $\mathrm{var}(\mathbf{e}_j^{\mathrm{T}} \mathbf{V}_n \mathbf{w}_i^* - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i) = \omega_{ii}^* \sigma_{jj}^* / n$.

(2) If $1 \leq i = j \leq p$, we have $\mathrm{var}(\mathbf{e}_i^{\mathrm{T}} \mathbf{V}_n \mathbf{w}_i^* - \mathbf{e}_i^{\mathrm{T}} \mathbf{e}_i) = (\omega_{ii}^* \sigma_{ii}^* + 1 + \Delta)/n$. □

4. Technical Lemmas and their proofs

We include the following Lemmas 1-3 and Lemmas 8-13 that are needed for the proof of the main theorems in the main text.

Lemma 1 (Bonferroni Inequality). *Let $B = \cup_{t=1}^{p} B_t$ we have*

$$\sum_{t=1}^{2k} (-1)^{t-1} E_t \leq pr(B) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} E_t$$

where $E_t = \sum_{1 \leq i_1 \leq \ldots \leq i_t \leq p} pr(B_{i_1} \cap \cdots \cap B_{i_t})$ and $k < [p/2]$.

Lemma 2 (Berman (1962)). *If X and Y are bi-variate normally distributed with expectations 0, unit variance and correlation ρ, then*

$$\lim_{c \to \infty} \frac{pr(X > c, Y > c)}{\{2\pi(1-\rho)^{1/2} c^2\}^{-1} \exp\{-c^2/(1+\rho)\}(1+\rho)^{1/2}} = 1,$$

uniformly for all ρ such that $|\rho| < \delta$, for any $0 < \delta < 1$.

Lemma 3 (Zaitsev (1987)). *Let $\tau > 0$, $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n \in \mathbb{R}^k$ are independent random variables such that $\mathcal{L}(\boldsymbol{\xi}_i) \in \mathcal{B}_1(k, \tau)$, for $i = 1, \ldots, n$, where $\mathcal{B}_1(k, \tau) = \{\mathcal{L}(\boldsymbol{\xi}) \in \mathcal{F}_k : E\boldsymbol{\xi} = 0, |E(\boldsymbol{\xi}, \boldsymbol{t})^2 (\boldsymbol{\xi}, \boldsymbol{u})^{m-2}| \leq m! \tau^{m-2} ||\boldsymbol{u}||^{m-2} E(\boldsymbol{\xi}, \boldsymbol{t})^2/2$, for every integer $m \geq 3$ and for all $\boldsymbol{t}, \boldsymbol{u}$ $\}$, $\mathcal{L}(\boldsymbol{\xi})$ is the distribution of random variable $\boldsymbol{\xi}$, $\mathcal{F}_k$ is the class of random distribution on $\mathbb{R}^k$, $(\boldsymbol{\xi}, \boldsymbol{t})$ is the inner product of $\boldsymbol{\xi}$ and $\boldsymbol{t}$. Denote $S = \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 + \cdots + \boldsymbol{\xi}_n, F = \mathcal{L}(S)$. Let Φ be a Gaussian distribution with mean vector 0 and the same covariance matrix with F. Define*

$\pi(F,\Phi;\lambda) = sup_{H\in\mathcal{B}_k} max\{F(H)-\Phi(H^\lambda),\Phi(H)-F(H^\lambda)\}$, *where* $\mathcal{B}_k$ *is the* σ*-field of Borel subsets of* $\mathbb{R}^k$, $H^\lambda = \{y\in\mathbb{R}^k : inf_{x\in H}||y-x|| \le \lambda\}$. *Then*

$$\pi(F,\Phi;\lambda) \le c_1 k^{5/2}\exp(-\frac{\lambda}{\tau c_2 k^{5/2}}),$$

for all $\lambda > 0$.

The following Lemmas 4 - 7 are used in Examples 1-4 for some special classes of precision matrices.

Lemma 4 (Hall & Lin (2010))**.** *For* $\lambda \ge 1, c_0 > 0, M > 0$. *For any sequence of matrices* $\mathbf{\Sigma}_n$ *such that*

$$\mathbf{\Sigma}_n \in \Theta_n^\star(\lambda,c_0,M) = \{\mathbf{\Sigma}_n : |\mathbf{\Sigma}_n(j,k)| \le M*(1+|j-k|)^{-\lambda}, ||\mathbf{\Sigma}_n|| \ge c_0\}.$$

There exists a constant $C = C(\lambda,c_0,M)$ *such that for any* n *and any* $1\le j,k\le n$,

$$|\mathbf{\Sigma}_n^{-1}(j,k)| \le C*(1+|j-k|)^{-\lambda}.$$

Lemma 5 (Gröchenig & Leitner (2006))**.** *Let* $\mathbf{A} = (a_{ij})_{p\times p}, \mathbf{A}^{-1} = (b_{ij})_{p\times p}$, $\lambda_{\max}(\mathbf{A})$ *and* $\lambda_{\max}(\mathbf{A}^{-1})$ *are bounded. If* $a_{ij} = O\big\{\exp(-\alpha|i-j|)\big\}$, *then* $b_{ij} = O\big\{\exp(-\beta|i-j|)\big\}$ *for some* β *such that* $0<\beta<\alpha$.

Lemma 6 (Demko et al. (1984))**.** *Let* $\mathbf{A} = (a_{ij})_{p\times p}$ *and* $\mathbf{A}^{-1} = (b_{ij})_{p\times p}$. *Assume that* $\lambda_{max}(\mathbf{A})$ *and* $\lambda_{max}(\mathbf{A}^{-1})$ *are bounded. If* $\mathbf{A}$ *is positive definite and* m*-banded, then we have* $|b_{ij}| \le C\lambda^{|i-j|}$ *where* $\lambda = \big[\{\sqrt{cond(\mathbf{A})}-1\}/\{\sqrt{cond(\mathbf{A})}+1\}\big]^{2/m}$, $cond(\mathbf{A}) = ||\mathbf{A}||\ ||\mathbf{A}^{-1}||$, $C = ||\mathbf{A}^{-1}||\max[1$ *and* $\{1+\sqrt{cond(\mathbf{A})}\}^2/\{2cond(\mathbf{A})\}]$.

Lemma 7. *Let* $\mathbf{I}_p$ *be an identity matrix and* $\mathbf{A} = \mathbf{I}_p + \sum_{i=1}^k \alpha_i \boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}$ *for any vector* $\boldsymbol{u}_i \in \mathbb{R}^{p\times 1}, \alpha_i\in\mathbb{R}, i=1,\ldots,k$. *Then outside the support of* $\mathbf{A}$, *and* $\mathbf{A}^{-1}$ *have the same zeros pattern.*

Proof: Let us denote $\mathbf{U}_{p\times k} = (\alpha_1 \boldsymbol{u}_1, \dots, \alpha_k \boldsymbol{u}_k), \mathcal{V}_{k\times p} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_k)^{\mathrm{T}}$, then $\sum_{i=1}^{k} \alpha_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{T}} = \mathbf{U}\mathcal{V}$. So $\mathbf{A} = \mathbf{I}_p + \mathbf{U}\mathcal{V}$. Applying Woodbury formula from page 211 in Hager (1989) we have:

$$\mathbf{A}^{-1} = (\mathbf{I}_p + \mathbf{U}\mathcal{V})^{-1} = \mathbf{I}_p + \mathbf{U}(\mathbf{I}_k - \mathcal{V}\mathbf{U})^{-1}\mathcal{V}.$$

Denote $\mathbf{M} = (\mathbf{I}_k - \mathcal{V}\mathbf{U})^{-1}, \mathbf{H} = \mathbf{U}\mathbf{M}\mathcal{V}$, then $\mathbf{A}^{-1} = \mathbf{I}_p + \mathbf{H}$. It can be checked that the zero patterns of $\mathbf{H}$ and $\mathbf{U}\mathcal{V}$ are the same. For easy to understand, let us consider a special case $\mathbf{A} = \mathbf{I}_p + \boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{u}_2^{\mathrm{T}}$ where $\boldsymbol{u}_1 = \mathbf{e}_1 + \mathbf{e}_2 \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{u}_2 = \mathbf{e}_3 + \mathbf{e}_4 \in \mathbb{R}^{p\times 1}$. Then

$$\mathcal{V} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 1 & 0 & \dots & 0 \end{pmatrix}_{2\times p} \quad \text{and} \quad \mathbf{U} = \mathcal{V}^{\mathrm{T}}.$$

For (i,j)th position of $\mathbf{H}$ where $i \notin \{1,2,3,4\}$ or $j \notin \{1,2,3,4\}$, we have $\mathbf{H}(i,j) = \mathbf{U}(i,)\mathbf{M}\mathcal{V}(,j) = 0$. Since the zero patterns on $\mathbf{H}$ and $\mathbf{U}\mathcal{V}$ are the same, using this fact together with $\mathbf{A} = \mathbf{I}_p + \mathbf{U}\mathcal{V}$ and $\mathbf{A}^{-1} = \mathbf{I}_p + \mathbf{H}$ completes the proof of this Lemma. □

Lemma 8. $\max_{(i,j)\in A_0} \hat{D}^*_{ij} = o_p(1)$.

Proof: Recall that $\hat{D}^*_{ij} = |\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i|/\sqrt{\theta_{ij}}$. Consider the numerator

$$\begin{aligned}
\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i &= \frac{n-1}{n}(\mathbf{e}_j^{\mathrm{T}}\boldsymbol{S}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i) - \frac{1}{n}\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i + \mathbf{e}_j^{\mathrm{T}}\bar{\mathbf{X}}\bar{\mathbf{X}}^{\mathrm{T}}\hat{\mathbf{w}}_{i,0} \\
&= \frac{n-1}{n}\{\mathbf{e}_j^{\mathrm{T}}\boldsymbol{S}_n\mathbf{B}_{i,0}(\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{S}_n\mathbf{B}_{i,0})^{-1}\mathbf{B}_{i,0}\mathbf{e}_i - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i\} \\
&\quad - \frac{1}{n}\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i + \mathbf{e}_j^{\mathrm{T}}\bar{\mathbf{X}}\bar{\mathbf{X}}^{\mathrm{T}}\hat{\mathbf{w}}_{i,0}.
\end{aligned}$$

Notice that the first term is indeed 0. For notation convenience, consider $i = 1$ and suppose that

$$\mathbf{w}_{1,0} = (w_{11}, w_{12}, w_{13}, w_{14}, 0, \dots, 0)^{\mathrm{T}} = \mathbf{B}_{1,0}\mathbf{w}_{11,0} \in \mathbb{R}^{p\times 1},$$

where $\mathbf{w}_{11,0} = (w_{11}, w_{12}, w_{13}, w_{14})^{\mathrm{T}} \in \mathbb{R}^{4\times 1}$ is non zero components of $\mathbf{w}_{1,0}$ and

$$\mathbf{B}_{1,0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 \end{pmatrix} \in R^{p\times 4}.$$

Then

$$\mathbf{B}_{1,0}^{\mathrm{T}} \boldsymbol{S}_n \hat{\mathbf{w}}_{1,0} = \mathbf{B}_{1,0}^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{1,0} (\mathbf{B}_{1,0}^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{1,0})^{-1} \mathbf{B}_{1,0} \mathbf{e}_1 = \mathbf{B}_{1,0} \mathbf{e}_1.$$

So

$$\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{1,0} (\mathbf{B}_{1,0}^{\mathrm{T}} \boldsymbol{S}_n \mathbf{B}_{1,0})^{-1} \mathbf{B}_{1,0} \mathbf{e}_1 - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_1 = 0, \text{ for } j = 1, 2, 3, 4.$$

So we have

$$\hat{D}_{ij}^* = \frac{\mathbf{e}_j^{\mathrm{T}} \mathbf{V}_n \hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i}{\sqrt{\theta_{ij}}} = -\frac{\mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i / n}{\sqrt{\theta_{ij}}} + \frac{\mathbf{e}_j^{\mathrm{T}} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{T}} \hat{\mathbf{w}}_{i,0}}{\sqrt{\theta_{ij}}}.$$

So

$$\max_{(i,j)\in A_0} \hat{D}_{ij}^* \le \max_{(i,j)\in A_0} \left|\frac{\mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i / n}{\sqrt{\theta_{ij}}}\right| + \max_{(i,j)\in A_0} \left|\frac{\mathbf{e}_j^{\mathrm{T}} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{T}} \hat{\mathbf{w}}_{i,0}}{\sqrt{\theta_{ij}}}\right|. \tag{4.1}$$

From Lemma 1 in the main text, we have the denominator $\sqrt{\theta_{ij}}$ is at the order of $1/\sqrt{n}$. This gives us

$$\max_{(i,j)\in A_0} \frac{\mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i / n}{\sqrt{(1/n)}} = o(1). \tag{4.2}$$

For the second term in (4.1), we note that $\mathbf{e}_j^{\mathrm{T}} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{T}} \hat{\mathbf{w}}_{i,0} = \sum_{k=1}^{s_0} \bar{X}_j \bar{X}_{i_k} \hat{w}_{i i_k}$ where $1 \le i_1, i_2, \dots, i_{s_0} \le p$ are non zero positions in $\mathbf{w}_{i,0}$. From page 2582 in Bickel & Levina (2008), we have $\max_{1\le i\le p} \bar{X}_i = O_p\{\sqrt{(\log\ p/n)}\}$. This gives us

$$\max_{(i,j)\in A_0} |\mathbf{e}_j^{\mathrm{T}} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{T}} \hat{\mathbf{w}}_{i,0}| \le \max_{1\le i\le p} \bar{X}_i^2 \sum_{k=1}^{s_0} |\hat{w}_{i i_k}| = O_p(\log p/n).$$

This gives us

$$\max_{(i,j)\in A_0} |\mathbf{e}_j^{\mathrm{T}} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{T}} \hat{\mathbf{w}}_{i,0}| / \sqrt{\theta_{ij}} = O_p(\log p / \sqrt{n}) = o_p(1). \tag{4.3}$$

The facts (4.1), (4.2), and (4.3) together verify the Lemma. □

Lemma 9. $\sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}| = \sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* + \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0}| + o_p(\sqrt{\log\ p})$.

Proof: On the one hand, we have

$$
\begin{aligned}
&\sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}(\mathbf{V}_n-\mathbf{\Sigma}^*)\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1}-\mathbf{\Omega}_i^*)\mathbf{f_i}| \\
&\le \sqrt{n}s_0\max_{1\le i,j\le p}|v_{ij}-\sigma_{ij}^*|\max_{1\le i,j\le p}|\hat{w}_{ij,0}-\omega_{ij}^*| \\
&= O_p(s_0 \log p/\sqrt{n}) = o_p(\sqrt{\log p}). \qquad (4.4)
\end{aligned}
$$

On the other hand

$$
\begin{aligned}
&\sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}(\mathbf{V}_n-\mathbf{\Sigma}^*)\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1}-\mathbf{\Omega}_i^*)\mathbf{f_i}| \\
&= \sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}-\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^*-\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0}|. \qquad (4.5)
\end{aligned}
$$

Combining (4.4) and (4.5), we get

$$
\sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}| = \sqrt{n}\max_{(i,j)\in A/A_0}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* + \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\hat{\mathbf{w}}_{i,0}| + o_p(\sqrt{\log\ p}).
$$

This completes the proof the Lemma. □

Lemma 10. *For any* $(i,j)\in A/B_0$, *let* $\mathbf{a}^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}} - \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}$, *then*

$$
Var(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}) = (\omega_{ii}^*\sigma_{jj}^* - \omega_{ii}^*\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{e}_j)/n.
$$

Proof: We first note that

$$
\begin{aligned}
\mathrm{E}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}) &= \mathrm{E}(\mathbf{a}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_{i,0}) = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_{i,0} \\
&= (\mathbf{e}_j^{\mathrm{T}} - \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}(\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0})^{-1}\mathbf{B}_{i,0}^{\mathrm{T}})\mathbf{\Sigma}^*\mathbf{w}_{i,0} \\
&= -\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}(\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0})^{-1}\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{e}_i = -\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_{i,0} = 0, \\
\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}) &= \mathrm{E}\{(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0})^2\} = \frac{1}{n}\mathrm{E}(\mathbf{a}^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{a}),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\mathbf{a}^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\mathbf{a}) &= \mathrm{E}(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{a}) \\
&= \mathrm{E}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{a}\mathbf{a}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{Z}) \\
&= \mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{a}\mathbf{a}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) + 2\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{a}\mathbf{a}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) \\
&\quad +\Delta\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{w}_{i,0}\mathbf{w}_{i,0}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}} \circ \boldsymbol{\Gamma}\mathbf{a}\mathbf{a}^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}) \\
&= \omega_{ii}^*\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{a} + (2+\Delta)\mathbf{e}_i^{\mathrm{T}}\mathbf{a}\mathbf{a}^{\mathrm{T}}\mathbf{e}_i.
\end{aligned}
$$

Thus, we have

$$
\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}) = (\omega_{ii}^*\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{a} + 2\mathbf{e}_i^{\mathrm{T}}\mathbf{a}\mathbf{a}^{\mathrm{T}}\mathbf{e}_i)/n. \tag{4.6}
$$

Recall that $\boldsymbol{\Omega}_i^* = \mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Omega}^*\mathbf{B}_{i,0}$. We note the following

$$
\begin{aligned}
\mathbf{a}^T\boldsymbol{\Sigma}^*\mathbf{a} &= (\mathbf{e}_j^{\mathrm{T}} - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}})\boldsymbol{\Sigma}^*(\mathbf{e}_j - \mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j) \\
&= \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j - 2\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j \\
&+ \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}(\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0})^{-1}\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j \\
&= \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j \\
&= \sigma_{jj}^* - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j, \qquad (4.7) \\
\mathbf{e}_i^{\mathrm{T}}\mathbf{a}\mathbf{a}^{\mathrm{T}}\mathbf{e}_i &= \mathbf{e}_i^{\mathrm{T}}(\mathbf{e}_j - \mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j)(\mathbf{e}_j^{\mathrm{T}} - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}})\mathbf{e}_i \\
&= \mathbf{e}_i^{\mathrm{T}}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i - 2\mathbf{e}_i^{\mathrm{T}}\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i \\
&+ \mathbf{e}_i^{\mathrm{T}}\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{T}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{e}_i \\
&= \mathbf{e}_i^{\mathrm{T}}\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_i^* \\
&= \mathbf{e}_i^{\mathrm{T}}\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i \\
&= 0. \qquad (4.8)
\end{aligned}
$$

Plugging (4.7) and (4.8) into (4.6), we get the variance expression in the Lemma. □

Lemma 11. *$pr(\max_{(i,j)\in A_1} \hat{D}_{ij}^{*2} \geq t_p) = o(1)$, where $t_p = t + 4\log p - \log(\log p)$.*

Proof: Lemma 9 gives us

$$\sqrt{n}\max_{(i,j)\in A_1}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}| = \sqrt{n}\max_{(i,j)\in A_1}|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* + \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0}| + o_p(\sqrt{\log\ p}).$$

We have

$$\begin{aligned} pr(\max_{(i,j)\in A_1}\hat{D}_{ij}^{*2}\geq t_p) &= pr(\max_{(i,j)\in A_1}|\hat{D}_{ij}^*|\geq\sqrt{t_p}) \\ &= pr(\max_{(i,j)\in A_1}|\frac{\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}}{\sqrt{\theta_{ij}}}|\geq\sqrt{t_p}) \\ &= pr(\max_{(i,j)\in A_1}\frac{|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* + \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0}|}{\sqrt{\theta_{ij}}} + o_p(\sqrt{\log p})\geq\sqrt{t_p}) \\ &= pr(\max_{(i,j)\in A_1}\frac{|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^* + \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0}|}{\sqrt{\theta_{ij}}}\geq\sqrt{t_p}). \end{aligned}$$

We have

$$\begin{aligned} \sqrt{1/\theta_{ij}}\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0} &= \sqrt{1/\theta_{ij}}(\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_{i,0}) \\ &= \sqrt{1/\theta_{ij}}\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\Big\{\boldsymbol{S}_i^{-1} - (\mathbf{B}_{i,0}^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0})^{-1}\Big\}\mathbf{f}_i \\ &= \sqrt{1/\theta_{ij}}\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i. \end{aligned} \tag{4.9}$$

Applying Lemma 5 in Le and Zhong (2021), we have

$$\begin{aligned} &\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\Big\{\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*\Big\}\mathbf{f}_i \\ &= -\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*(\boldsymbol{S}_i - \mathbf{\Sigma}_i^*)\mathbf{\Omega}_i^*\mathbf{f}_i \\ &- \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*(\boldsymbol{S}_i - \mathbf{\Sigma}_i^*)(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i. \end{aligned}$$

Let us denote $R = \max_{(i,j)\in A_1}|\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*(\boldsymbol{S}_i - \mathbf{\Sigma}_i^*)(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i|$. Since $||\mathbf{\Sigma}^*||_1$ and $||\mathbf{\Omega}_i^*||_1$ are bounded, so $||\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}\mathbf{\Omega}_i^*||_1 = O(1)$. Then we have

$$R \leq s_0\max_{1\leq i,j\leq p}|s_{ij} - \sigma_{ij}^*|\max_{1\leq i,j\leq p}|\hat{\omega}_{ij,0} - \omega_{ij}^*| = O_p(s_0\log\ p/n).$$

So

$$
\begin{aligned}
&\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}[\boldsymbol{S}_i^{-1}-\boldsymbol{\Omega}_i^*]\mathbf{f}_i \\
&= -\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0}+O_p(s_0\log\ p/n) \\
&= -\mathbf{m}^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0}+O_p(s_0\log\ p/n), \qquad (4.10)
\end{aligned}
$$

where $\mathbf{m}^{\mathrm{T}}=\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}$. Notice that $||\mathbf{m}||_1=O(1)$ and $||\mathbf{w}_{i,0}||_1=1$.

We have

$$
\mathbf{m}^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0}=\mathbf{m}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}+\frac{1}{n-1}\mathbf{m}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}-\frac{n}{n-1}\mathbf{m}^{\mathrm{T}}\bar{\mathbf{X}}\bar{\mathbf{X}}^{\mathrm{T}}\mathbf{w}_{i,0}.
$$

In addition we have,

$$
\begin{aligned}
\frac{1}{n-1}|\mathbf{m}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}| &\le \frac{1}{n-1}|\mathbf{m}^{\mathrm{T}}(\mathbf{V}_n-\boldsymbol{\Sigma}^*)\mathbf{w}_{i,0}|+\frac{1}{n-1}|\mathbf{m}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{w}_{i,0}| \\
&= \frac{1}{n-1}|\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\{\mathbf{B}_{i,0}^{\mathrm{T}}(\mathbf{V}_n-\boldsymbol{\Sigma}^*)\mathbf{B}_{i,0}\}\mathbf{w}_{i1,0}|+O(1/n) \\
&= O_p(s_0/n).
\end{aligned}
$$

In other words, we have

$$
\frac{1}{n-1}|\mathbf{m}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}|=O_p(s_0/n). \qquad (4.11)
$$

Furthermore, we have

$$
\frac{n}{n-1}\mathbf{m}^{\mathrm{T}}\bar{\mathbf{X}}\bar{\mathbf{X}}^{\mathrm{T}}\mathbf{w}_{i,0}\le\frac{n}{n-1}\max_{i=1,\ldots,p}\bar{X}_i^2\ ||\mathbf{m}||_1\ ||\mathbf{w}_{i,0}||_1=O_p(\log p/n). \qquad (4.12)
$$

Applying (4.9), (4.10), (4.11), and (4.12), we get

$$
\begin{aligned}
pr(\max_{(i,j)\in A_1} \hat{D}_{ij}^{*2} \geq t_p) &= pr(\max_{(i,j)\in A_1} \frac{|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0} + \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\hat{\mathbf{w}}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p}) \\
&= pr(\max_{(i,j)\in A_1} \frac{|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0} - \mathbf{m}^{\mathrm{T}}\boldsymbol{S}_n\mathbf{w}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p}) \\
&= pr\{\max_{(i,j)\in A_1} \frac{|(\mathbf{e}_j^{\mathrm{T}} - \mathbf{m}^{\mathrm{T}})\mathbf{V}_n\mathbf{w}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p}\} \\
&= pr(\max_{(i,j)\in A_1} \frac{|\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p})
\end{aligned}
$$

where $\mathbf{a}^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}} - \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}$.

Lemma 10 implies that for $(i,j) \in A_1$,

$$
\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}) = (\omega_{ii}^*\sigma_{jj}^* - \omega_{ii}^*\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_i\boldsymbol{\Omega}_i^*\mathbf{B}_i^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j)/n \leq (\omega_{ii}^*\sigma_{jj}^*)/n = \theta_{ij},
$$

where we notice that $\omega_{ii}^*\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}\boldsymbol{\Omega}_i^*\mathbf{B}_{i,0}^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{e}_j \geq 0$, since $\boldsymbol{\Omega}_i^*$ is positive definite.

By central limit theory, we have $\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0} = (\sum_{k=1}^{n} \mathbf{a}^{\mathrm{T}}\mathbf{X}_k\mathbf{X}_k^{\mathrm{T}}\mathbf{w}_{i,0})/n \to N(0, \mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}))$. In addition $\mathrm{card}(A_1) = o(p^2)$, this gives us

$$
\begin{aligned}
pr(\max_{(i,j)\in A_1} \frac{|\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0} + \mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_i\hat{\mathbf{w}}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p}) &= pr(\max_{(i,j)\in A_1} \frac{\sqrt{n}|\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}|}{\sqrt{\theta_{ij}}} \geq \sqrt{t_p}) \\
&\leq pr\{\max_{(i,j)\in A_1} \frac{|\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}|}{\sqrt{\{\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0})\}}} \geq \sqrt{t_p}\} \\
&\leq \sum_{(i,j)\in A_1} pr\{\frac{|\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0}|}{\sqrt{\{\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_{i,0})\}}} \geq \sqrt{t_p}\} \\
&\leq o(p^2)e^{-t_p/2} = o(p^2)e^{-2\log\ p} = o(1)
\end{aligned}
$$

where the last inequality is due to Gaussian tail inequality. The Lemma is proved. □

Lemma 12. $\max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}_i^*)\mathbf{f}_i| = o_p(\sqrt{(\log\ p/n)})$.

Proof: When the underlying network structure is a factor model, it can be seen that $\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}_i^*)\mathbf{f}_i = 0,$ for all $(i,j) \in A/B_0$. So the Lemma is satisfied.

Now we consider the case for other network structures with their covariance matrix and precision matrix satisfying conditions (C1). Let us denote $\mathbf{b}^{\mathrm{T}} = \mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0} = (\sigma^*_{ji_1}, \dots, \sigma_{ji^*_{s_0}})$. where $i_1, \dots, i_{s_0}$ are nonzero positions at column $\mathbf{w}_{i,0}$ of the precision matrix $\mathbf{\Omega}$. Since $||\mathbf{b}||_1 = O(1)$, applying Theorem 4 in Le and Zhong (2021), we have

$$\sqrt{(n/a_{ij})}\mathbf{b}^{\mathrm{T}}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i \sim N(0,1) \tag{4.13}$$

where $a_{ij} = \mathrm{var}(\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{X}_{1i}\mathbf{X}_{1i}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{f}_i),$ for all $(i,j) \in A/B_0$.

Denote $\mathbf{\Omega}_i^* = (\gamma_{ij})_{s_0 \times s_0}$. By Lemma 7 in Le and Zhong (2021), we get

$$\begin{aligned}
a_{ij} &= \mathrm{var}(\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{X}_{1i}\mathbf{X}_{1i}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{f}_i) = E\{\mathbf{X}_{1i}^T\mathbf{\Omega}_i^*\mathbf{b}\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{X}_{1i}\mathbf{X}_{1i}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{f}_i\mathbf{f}_i^T\mathbf{\Omega}_i^*\mathbf{X}_{1i}\} - (\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{\Sigma}_i\mathbf{\Omega}_i^*\mathbf{f}_i)^2 \\
&= \mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{\Sigma}_i^*\mathbf{\Omega}_i^*\mathbf{b}\mathbf{f}_i^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{\Sigma}_i^*\mathbf{\Omega}_i^*\mathbf{f}_i + (\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{\Sigma}_i\mathbf{\Omega}_i^*\mathbf{f}_i)^2 + \Delta\mathrm{tr}(\Gamma_i^T\mathbf{\Omega}_i^*\mathbf{b}\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\Gamma_i^T \circ \Gamma_i^T\mathbf{\Omega}_i^*\mathbf{f}_i\mathbf{f}_i^T\mathbf{\Omega}_i^*\Gamma_i^T) \\
&= \mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{b}\mathbf{f}_i^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{f}_i + (1+\Delta)(\mathbf{b}^{\mathrm{T}}\mathbf{\Omega}_i^*\mathbf{f}_i)^2 \\
&= \omega_{ii}^* \sum_{k,l \in \{i_1,\dots,i_{s_0}\}} \sigma^*_{jk}\sigma^*_{jl}\gamma_{kl} + (1+\Delta) \sum_{k,l \in \{i_1,\dots,i_{s_0}\}} \sigma^*_{jk}\sigma^*_{jl}\gamma_{ik}\gamma_{il}.
\end{aligned} \tag{4.14}$$

On A/B_0 we have

$$\sigma^*_{jk}\sigma^*_{jl} = o(1/s_0^2), \text{ for all } k,l \in \{i_1,\dots,i_{s_0}\}, j \neq k,l. \tag{4.15}$$

The facts (4.14) and (4.15) give us $a_{ij} = o(1),$ for all $(i,j) \in A/B_0$.

Let us denote $a = \max_{(i,j)\in A/B_0} \sqrt{a_{ij}}$, so $a = o(1)$. Applying (4.13), we have

$$\begin{aligned}
&pr\Big\{ \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i| \geq t\Big\} \\
&\leq pr\Big\{ \max_{(i,j)\in A/B_0} |\sqrt{(n/a_{ij})}\mathbf{e}_j\mathbf{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i| \geq (\sqrt{n}t/a)\Big\} \\
&\leq p^2\exp\{-nt^2/(2a^2)\}.
\end{aligned}$$

Choose $t = Ma\sqrt{\{(\log p)/n\}}$ for $M > 0$ sufficient large, then we have

$$pr\Big\{\max_{(i,j)\in A/B_0} |\mathbf{e}_j \mathbf{\Sigma}^* \mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i| \geq Ma\sqrt{\frac{\log p}{n}}\Big\} \leq p^2 \exp(-\frac{nM^2a^2 \log\ p}{2a^2 n})$$

$$= p^2 \exp(\log\ p^{-M/2})$$

$$= p^{2-M/2} \to 0.$$

Or

$$\max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}} \mathbf{\Sigma}^* \mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \mathbf{\Omega}_i^*)\mathbf{f}_i| = O_p\{a\sqrt{(\log p/n)}\} = o_p\{\sqrt{(\log p/n)}\}.$$

The Lemma is verified. □

Lemma 13.

$$\sum_{1\leq k_1<\ldots<k_d\leq q} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n(\log p)^{-1/2}\} = \frac{1}{d!}\{\frac{1}{\sqrt{(2\pi)}}\exp(-\frac{t}{2})\}^d\{1+o(1)\}, \quad (4.16)$$

where $\mathbf{N}_d = (N_{k_1}, \ldots, N_{k_d})^{\mathrm{T}}$ is a d-dimensional multivariate Gaussian random variable with mean vector 0 and covariance matrix $\mathrm{cov}(\mathbf{N}_d) = \mathrm{cov}(\mathbf{W}_1)$. Here $\mathbf{W}_1$ is the random variable defined as in equation (27) of the proof of Theorem 1 in the main text.

Proof: Notice that for $X \sim N(0,1)$, we have

$$pr(|X| \geq x) = 2\{1+o(1)\}\frac{\exp^{-x^2/2}}{x\sqrt{(2\pi)}}.$$

So when $d = 1$, we get

$$pr\Big\{|\mathbf{N}_1|_{\min} \geq t_p^{1/2} \pm \epsilon_n(\log p)^{-1/2}\Big\} = 2\{1+o(1)\}\frac{\exp^{-t_p/2}}{\sqrt{t_p}\sqrt{(2\pi)}}$$

$$= \{1+o(1)\}\frac{2\exp(-t/2 - 2\mathrm{log}p)(\mathrm{log}p)^{1/2}}{2\sqrt{(\mathrm{log}p)}\sqrt{(2\pi)}}$$

$$= \{1+o(1)\}\frac{p^{-2}\exp^{-t/2}}{\sqrt{(2\pi)}}.$$

This leads

$$\sum_{1\leq k_1\leq q} pr\{|\mathbf{N}_1|_{\min} \geq t_p^{1/2} \pm \epsilon_n(\mathrm{log}p)^{-1/2}\} = \frac{\exp^{-t/2}}{\sqrt{(2\pi)}}\{1+o(1)\}. \quad (4.17)$$

The Lemma is verified for $d = 1$.

Let us consider when $d \geq 2$, we need to show that

$$pr\Big\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n(\log p)^{-1/2}\Big\} = \{1+o(1)\}\Big\{\frac{1}{\sqrt{(2\pi)}}\exp(-\frac{t}{2})^d p^{-2d}\Big\}. \tag{4.18}$$

Let $\mathbf{R} = (\rho_{ij})_{p\times p}$ be the correlation matrix and $\tilde{\mathbf{\Omega}} = (\tilde{\omega}_{ij})_{p\times p}$ is the standardized version of the precision matrix $\mathbf{\Omega}^*$ where $\tilde{\omega}_{ij} = \omega_{ij}^*/\sqrt{(\omega_{ii}^*\omega_{jj}^*)}$. For a fixed constant $\alpha_0 > 0$, for $j = 1, 2, \ldots, p$, define

$$s_j = s_j(\alpha_0) = \text{card}\{i : |\rho_{ij}| \geq (\log p)^{-1-\alpha_0}\}, h_j = h_j(\alpha_0) = \text{card}\{i : |\tilde{\omega}_{ij}| \geq (\log p)^{-1-\alpha_0}\}.$$

We need two following conditions for our proof

$$\max_{j=1,\ldots,p} s_j(\alpha_0) = o(p^\gamma), \max_{j=1,\ldots,p} h_j(\alpha_0) = o(p^\gamma), \forall\gamma > 0. \tag{4.19}$$

$$\text{There exists some } r \in (0,1), \rho_{ij} < r, \tilde{\omega}_{ij} < r, \text{ for all } 1 \leq i \neq j \leq p. \tag{4.20}$$

Notice that the above conditions are mild. Condition (4.19) is met if $\mathbf{R}$, and $\mathbf{\Omega}^*$ has maximum eigenvector bounded from the above. And condition (4.20) met once the off diagonal elements of $\mathbf{R}$ and $\tilde{\mathbf{\Omega}}$ are bounded by r. We have $EZ_{lk_1}Z_{lk_2} = \mathbf{e}_{j_{k_2}}^{\mathrm{T}}\mathbf{e}_{i_{k_1}}\mathbf{e}_{j_{k_1}}^{\mathrm{T}}\mathbf{e}_{i_{k_2}} + \sigma_{j_{k_1}j_{k_2}}^*\omega_{i_{k_1}i_{k_2}}^*$. When either $i_{k_1} \neq j_{k_2}$ or $i_{k_2} \neq j_{k_1}$, then $EZ_{lk_1}Z_{lk_2} = \sigma_{j_{k_1}j_{k_2}}^*\omega_{i_{k_1}i_{k_2}}^*$. Notice that on A/B_0, we have $\omega_{i_{k_1}j_{k_1}}^* = \omega_{i_{k_2}j_{k_2}}^* = 0$, so when $i_{k_1} = j_{k_2}$ and $i_{k_2} = j_{k_1}$, we get $EZ_{lk_1}Z_{lk_2} = \sigma_{i_{k_1}j_{k_1}}^*\omega_{i_{k_1}j_{k_1}}^* + 1 = 1$.

For two different pairs $(i_a, j_a), (i_b, j_b)$, we can establish a graph defined by $G_{i_aj_ai_bj_b} = (V_{i_aj_ai_bj_b}, E_{i_aj_ai_bj_b})$ where $V_{i_aj_ai_bj_b} = \{i_a, j_a, i_b, j_b\}$ is the set of vertices and $E_{i_aj_ai_bj_b}$ is the set of edges. We say there is an edge (connection) between $i \neq j \in \{i_a, j_a, i_b, j_b\}$ if $|\rho_{ij}| \geq (\log p)^{-1-\alpha_0}$ or $|\tilde{\omega}_{ij}| \geq (\log p)^{-1-\alpha_0}$.

We say G_{abcd} is a k-vertices graph (k-G) if the number of different vertices is k, in our case $k \in \{2, 3, 4\}$. For sake of convenient, we denote "3G-1E" for a three vertices graph when either

$\rho_{i_a i_b}$ or $\tilde{\omega}_{j_a j_b}$ form an edge. We denote "4G- 2E" for a four vertices graph when both $\rho_{i_a i_b}$ and $\tilde{\omega}_{j_a j_b}$ form edges. We say a graph $G = G_{i_{m_1} j_{m_1} i_{m_2} j_{m_2}}$ satisfy condition $(\star)$ if

$(\star)$: Either $\tilde{\omega}_{i_{m_1} i_{m_2}} \leq (\log p)^{-1-\alpha_0}$ or $\rho_{j_{m_1} j_{m_2}} \leq (\log p)^{-1-\alpha_0}$.

Remark: Those graphs satisfying $(\star)$ also satisfy

$$\operatorname{cov}(\tilde{Z}_{lm_1}, \tilde{Z}_{lm_2}) \to \rho_{j_{m_1} j_{m_2}} \tilde{\omega}_{i_{m_1} i_{m_2}} = O\{(\log p)^{-1-\alpha_0}\}. \tag{4.21}$$

As shown above for any two different pairs $(i_{k_1}, j_{k_1}), (i_{k_2}, j_{k_2})$ we have

$$\operatorname{cov}(\tilde{Z}_{lk_1}, \tilde{Z}_{lk_2}) \to \sqrt{\{1/(\omega^*_{i_{k_1} i_{k_1}} \omega^*_{i_{k_2} i_{k_2}} \sigma^*_{j_{k_1} j_{k_1}} \sigma^*_{j_{k_2} j_{k_2}})\}} E Z_{lk_1} Z_{lk_2}.$$

For any matrices $\mathbf{A} = (a_{ij})_{p \times p}, \mathbf{B} = (b_{ij})_{p \times p} = \mathbf{A}^{-1}$, page 472 in Robinson & Wahten (1992) tells us

$$b_{ii} \geq a_{jj}/(a_{ii} a_{jj} - a_{ij}^2), \text{ for any } 1 \leq i \neq j \leq p.$$

This gives us

$$\omega^*_{i_{k_1} i_{k_1}} \omega^*_{j_{k_1} j_{k_1}} \sigma^*_{i_{k_1} i_{k_1}} \sigma^*_{j_{k_1} j_{k_1}} \geq \{(\omega^*_{i_{k_1} i_{k_1}} \omega^*_{j_{k_1} j_{k_1}})/(\omega^*_{i_{k_1} i_{k_1}} \omega^*_{j_{k_1} j_{k_1}} - \omega^*_{i_{k_1} j_{k_1}})\}^2 > 1/r,$$

for some $r \in (0, 1)$. So for a 2G- 1E of two pairs $(i_{k_1}, j_{k_1}), (j_{k_1}, i_{k_1})$ we have, for some $r \in (0, 1)$,

$$\operatorname{cov}(\tilde{Z}_{lk_1}, \tilde{Z}_{lk_2}) \to \sqrt{\{1/(\omega^*_{i_{k_1} i_{k_1}} \omega^*_{j_{k_1} j_{k_1}} \sigma^*_{i_{k_1} i_{k_1}} \sigma^*_{j_{k_1} j_{k_1}})\}} < r, \tag{4.22}$$

For "4G-2E" or "3G-1E" of two different pairs $(i_{k_1}, j_{k_1}), (i_{k_2}, j_{k_2})$ we have

$$\begin{aligned} \operatorname{cov}(\tilde{Z}_{lk_1}, \tilde{Z}_{lk_2}) &\to \sqrt{\{1/(\omega_{i_{k_1} i_{k_1}} \omega^*_{i_{k_2} i_{k_2}} \sigma^*_{j_{k_1} j_{k_1}} \sigma^*_{j_{k_2} j_{k_2}})\}} \sigma^*_{j_{k_1} j_{k_2}} \omega^*_{i_{k_1} i_{k_2}} \\ &= \rho_{j_{k_1} j_{k_2}} \tilde{\omega}_{i_{k_1} i_{k_2}} < r, \end{aligned} \tag{4.23}$$

for some $0 < r < 1$.

Now we define the following sets $I = \{1 \leq k_1 < k_2 < \ldots < k_d \leq q\}$, d is a fixed positive integer. $I_0 = \{1 \leq k_1 < k_2 < \ldots < k_d \leq q :$ for some $m_1 \neq m_2 \in k_1, \ldots, k_d, G = G_{i_{m_1} j_{m_1} i_{m_2} j_{m_2}}$

does not satisfy $(\star)\}$. $I_0^c = \{1 \leq k_1 < k_2 < \ldots < k_d \leq q$: for any $m_1 \neq m_2 \in k_1, \ldots, k_d, G = G_{i_{m_1} j_{m_1} i_{m_2} j_{m_2}}$ satisfies $(\star)\}$.

Notice that $I = I_0 \cup I_0^c$. For any subset S of $\{k_1, \ldots, k_d\}$, we say that S satisfies $(\star\star)$ if $(\star\star)$ for any $m_1 \neq m_2 \in S, G_{i_{m_1} j_{m_1} i_{m_2} j_{m_2}}$ satisfies $(\star)$. For $2 \leq l \leq d$, let $I_{0l} = \{1 \leq k_1 < k_2 < \ldots < k_d \leq q : \text{card}(S) = l$, where S is largest subset of $k_1 < \ldots < k_d$, satisfies $(\star\star)\}$.

$I_{01} = \{1 \leq k_1 < k_2 < \ldots < k_d \leq q$: for any $m_1 \neq m_2 \in k_1, \ldots, k_d, G = G_{i_{m_1} j_{m_1} i_{m_2} j_{m_2}}$ does not satisfy $(\star)\}$. So $I_0^c = I_{0d}, I_0 = \cup_{l=1}^{d-1} I_{0l}$.

Claim:

$$\text{card}(I_{0l}) \leq C_d q^{l+2\gamma(d-l)}, \tag{4.24}$$

where C_d is a constant depends only on d. In addition

$$\text{card}(I_0^c) = \{1 + o(1)\} C_q^d. \tag{4.25}$$

Proof: First, we verify (4.24), $\text{card}(I_{0l}) \leq C_d q^{l+2\gamma(d-l)}$. There are at most C_q^l ways of choosing S with cardinality l from $1, 2..., q$. For a fixed element "a" in S, there is at most $p^\gamma p^\gamma = p^{2\gamma}$ choices for "b" which satisfies $G_{i_a j_a i_b j_b}$ not satisfies $(\star)$. So there will be at most $Clp^{2\gamma}$ choices for values "b" not go with l elements of S for properties $(\star)$. So we get $\text{card}(I_{0l}) \leq C_q^l (Clp^{2\gamma})^{d-l} \leq C_d q^{l+2\gamma(d-l)}$.

The claim (4.24) is verified.

Second, we show (4.25), $\text{card}(I_0^c) = \{1 + o(1)\} C_q^d$. We have $\text{card}(I) = C_q^d$, since we are choosing d numbers from q numbers without order.

$$\text{card}(I_0) \leq \sum_{l=1}^{d-1} \text{card}(I_{0l}) \leq \sum_{l=1}^{d-1} C_d q^{l+2\gamma(d-l)} = o(q^d) = o(C_q^d).$$

This gives us

$$\text{card}(I_0^c) = C_q^d - o(C_q^d) = \{1 + o(1)\} C_q^d.$$

This clarifies (4.25).

We claim that *the follows are true:*

$$\sum_{I_0} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} = o(1) \tag{4.26}$$

and

$$\sum_{I_0^c} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} = \frac{1}{d!}\{\frac{1}{\sqrt{(2\pi)}}\exp(-\frac{t}{2})\}^d\{1+o(1)\}, \tag{4.27}$$

Proof: Before verify (4.26), we need to divide our set I_{0l} a bit further. For $1 \leq a \neq b \leq q$, we define $d((i_a, j_a), (i_b, j_b)) = 1$, if $G_{i_a j_a i_b j_b}$ does not satisfies $(\star)$; $d((i_a, j_a), (i_b, j_b)) = 0$ otherwise. We further divide I_{0l} as the following. Let $(k_1, k_2, \ldots, k_d) \in I_{0l}$ and let $S_\star \subset (k_1, \ldots, k_d)$ be the largest cardinality subset satisfying $(\star\star)$ (if there are more than two subsets attain the largest cardinality, then we choose any of them). Define $I_{0l1} = \{(k_1, \ldots, k_d) \in I_{0l}$: there exists an $a \notin S_\star$, such that for some $b_1 \neq b_2 \in S_\star$ with, $d((i_a, j_a), (i_{b_1}, j_{b_1})) = 1$, and $d((i_a, j_a), (i_{b_2}, j_{b_2})) = 1\}$, $I_{0l2} = I_{0l}/I_{0l1}$. We have $I_{011} = \emptyset, I_{012} = I_{01}$. Recall that d fixed and $l \leq d-1$. We can show that

$$\text{card}(I_{0l1}) \leq Cq^{l-1+2\gamma(d-l+1)}. \tag{4.28}$$

$$\text{card}(I_{0l2}) \leq C_d q^{l+2\gamma(d-l)}. \tag{4.29}$$

Write $S_\star = (b_1, b_2, \ldots, b_l)$, for $(k_1, \ldots, k_d) \in I_{0l2}$. Since there exists an $a \notin S_\star$ such that $d((i_a, j_a), (i_{b_1}, j_{b_1})) = 1$ and $d((i_a, j_a), (i_{b_2}, j_{b_2})) = 1$ for some $b_1 \neq b_2 \in S_\star$. We consider b_1 is the first element in $S_\star$, there are at most q ways to choose b_1. There are at most $p^{2\gamma}$ to choose the second element in $S_\star$ not goes with "a" for $\star$. For the other $l-2$ elements in $S_\star$ there are at most C_q^{l-2} ways of choosing. For the rest $d-l$ elements outside $S_\star$, there are at most $p^{2\gamma(d-l)}$ ways of choosing. So on the whole, we have

$$\text{card}(I_{0l1}) \leq qp^{2\gamma} C_q^{l-2} p^{2\gamma(d-l)} \leq Cq^{l-1+2\gamma(d-l+1)},$$

which verifies (4.28). We have

$$\mathrm{card}(I_{0l}) = \mathrm{card}(I_{0l1}) + \mathrm{card}(I_{0l2}) \le C_d q^{l+2\gamma(d-l)}. \tag{4.30}$$

On the other hand

$$\mathrm{card}(I_{0l1}) \le C q^{l-1+2\gamma(d-l+1)} = o(q^l). \tag{4.31}$$

Applying (4.30) and (4.31), we get (4.29).

We go back to check our claim (4.26)

$$\sum_{I_0} pr\{|\mathbf{N}_d|_{min} \ge t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} = o(1).$$

On I_{0l} we have For any $k_1, \ldots, k_d \in I_{0l}$, write $S_\star = (b_1, b_2, \ldots, b_l)$, $\mathbf{U}_l$ is the covariance matrix of $(N_{b_1}, \ldots, N_{b_l})$, then $||\mathbf{U}_l - \mathbf{I}_l|| = O\{(\log p)^{-1-\alpha_0}\}$ (by (4.21)). As a result, we also have $|\mathbf{U}_l| \to 1$ as $p \to \infty$. Let us denote $|\mathbf{y}|_{\max} = \max_{1\le i\le l} |y_i|$, for $\mathbf{y} = (y_1, \ldots, y_l)^{\mathrm{T}}$ and $x_p = t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}$. We claim that

$$\begin{aligned} &\frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\ge(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y} \\ &= O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \end{aligned} \tag{4.32}$$

and

$$\begin{aligned} &\frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\le(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y} \\ &= \frac{1+O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\le(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y}. \end{aligned} \tag{4.33}$$

First, we check (4.32). We have

$$\frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}}\int_{|\mathbf{y}|_{\min}\geq x_p,|\mathbf{y}|_{\max}\geq(\log p)^{1/2+\alpha_0/4}}\exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y}$$
$$= pr\{|\mathbf{N}_d|_{\min}\geq x_p, |\mathbf{N}_d|_{\max}\geq(\log p)^{1/2+\alpha_0/4}\}$$
$$\leq \sum_{i=1}^{l} pr\{|N_i|\geq(\log p)^{1/2+\alpha_0/4}\}$$
$$= O\Big[\exp\{-\frac{1}{2}(\log p)^{1+\alpha_0/2}\}\Big] = O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big],$$

which validates (4.32).

We now verify (4.33). We have

$$||\mathbf{U}_l^{-1}-\mathbf{I}_l|| \leq ||\mathbf{U}_l^{-1}||||\mathbf{U}_l-\mathbf{I}_l|| = O\{(\log\ p)^{-1-\alpha_0}\}.$$

So, on set $\{|\mathbf{y}|_{\min}\geq x_p, |\mathbf{y}|_{\max}\leq(\log\ p)^{1/2+\alpha_0/4}\}$, using Taylor expansion we have:

$$\exp\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{U}_l^{-1}-\mathbf{I}_l)\mathbf{y}\} = 1+O\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{U}_l^{-1}-\mathbf{I}_l)\mathbf{y}\} = 1+O\{(\log p)^{-\alpha_0/2}\}.$$

Therefore,

$$\frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}}\int_{|\mathbf{y}|_{\min}\geq x_p,|\mathbf{y}|_{\max}\leq(\log p)^{1/2+\alpha_0/4}}\exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y}$$
$$= \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}}\int_{|\mathbf{y}|_{\min}\geq x_p,|\mathbf{y}|_{\max}\leq(\log p)^{1/2+\alpha_0/4}}\exp\{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{U}_l^{-1}-\mathbf{I}_l)\mathbf{y}\}exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y}$$
$$= \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}}\int_{|\mathbf{y}|_{\min}\geq x_p,|\mathbf{y}|_{\max}\leq(\log p)^{1/2+\alpha_0/4}}\exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y}$$
$$= \frac{1+O((\log p)^{-\alpha_0/2})}{(2\pi)^{l/2}}\int_{|\mathbf{y}|_{\min}\geq x_p,|\mathbf{y}|_{\max}\leq(\log p)^{1/2+\alpha_0/4}}\exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y}.$$

So we proved (4.33). The two claims are proved, we come back to show (4.26).

$$pr\{|\mathbf{N}_d|_{\min}\geq t_p^{1/2}\pm\epsilon_n(\log p)^{-1/2}\} \leq pr(|N_{b_1}|\geq x_p,\ldots,|N_{b_l}|\geq x_p)$$
$$= \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}}\int_{|\mathbf{y}|_{\min}\geq x_p}\exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y}$$

$$
\begin{aligned}
&= \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\le(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y} \\
&\quad + \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\ge(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y} \\
&= \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\le(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{U}_l^{-1}\mathbf{y})d\mathbf{y} \\
&\quad + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
&= \frac{1+O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min}\ge x_p, |\mathbf{y}|_{\max}\le(\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y} \\
&\quad + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
&= \frac{1+O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min}\ge x_p} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y} + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big]. \qquad (4.34)
\end{aligned}
$$

We have

$$
\begin{aligned}
\int_{|\mathbf{y}|_{\min}\ge x_p} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y} &= \Big(\int_{|u|\ge x_p} \exp(-\frac{1}{2}u^2)du\Big)^l \\
&= \Big(\frac{2\frac{1}{\sqrt{(2\pi)}}\exp[-\frac{1}{2}\{t_p^{1/2}\pm\epsilon_n(\log p)^{-1/2}\}]}{\sqrt{\{t_p^{1/2}\pm\epsilon_n(\log p)^{-1/2}\}}}\Big)^l \\
&= \{1+o(1)\}\{\frac{2}{\sqrt{(8\pi)}}\exp(-\frac{t}{2})\}^l p^{-2l}. \qquad (4.35)
\end{aligned}
$$

In addition,

$$
O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] = o(p^{-2l}). \qquad (4.36)
$$

The facts (4.34), (4.35), and (4.36) together give us

$$
\begin{aligned}
&\frac{1+O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min}\ge x_p} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y})d\mathbf{y} + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
&\qquad = \{1+o(1)\}\{\frac{2}{\sqrt{(8\pi)}}\exp(-\frac{t}{2})\}^l p^{-2l} \\
&\qquad = O(p^{-2l}). \qquad (4.37)
\end{aligned}
$$

So we have

$$\sum_{I_{0l1}} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n (\log\ p)^{-1/2}\} \leq \text{card}(I_{0l1})\ O(p^{-2l})$$

$$= O(p^{2l-2+4\gamma(d-l+1)-2l})$$

$$= o(1). \tag{4.38}$$

Let $\bar{a} = min\{a : a \in (k_1, k_2, \ldots, k_d), a \notin S_\star\}$. WLOG we assume $d((i_{\bar{a}}, j_{\bar{a}}), (i_{b_1}, j_{b_1})) = 1$, then $I_{0l2} = \{(k_1, \ldots, k_d) \in I_{0l2} : G_{i_{\bar{a}} j_{\bar{a}} i_{b_1} j_{b_1}}$ is $2E - 1G$ Or "$3G - 1E$" Or "$4G - 2E$"$\}$. On I_{0l2}, we have

$$\sum_{I_{0l2}} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} \leq \sum_{I_{0l2}} pr\{|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p, \ldots, |N_{b_l}| \geq x_p\}. \tag{4.39}$$

Now covariance matrix of $(N_{\bar{a}}, N_{b_1}, \ldots, N_{b_l})$ is $\mathbf{V}_l$, and the covariance matrix satisfies

$$||\mathbf{V}_l - \text{diag}(\mathbf{D}, \mathbf{I_{l-1}})|| = O\{(\log p)^{-1-\alpha_0}\}$$

where $\mathbf{D}$ is the covariance matrix of $(N_{\bar{a}}, N_{b_1})$.

Applying (4.22), (4.23), and Lemma 2 in Berman (1962), we obtain

$$pr(|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p) \leq C \exp(-\frac{4\log\ p}{1+r}) = Cp^{-4/(1+r)}. \tag{4.40}$$

Combining (4.39) and (4.40), we get

$$\begin{aligned}
&\sum_{I_{0l2}} pr(|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p, \ldots, |N_{b_l}| \geq x_p) \\
&\leq C \sum_{I_{0l2}} \Big[pr(|N_{\bar{a}}| \geq x_p, |N_{b_1}| \geq x_p) \times p^{-2l+2} + \exp\{-(\log p)^{1+\alpha_0/2}/4\} \Big] \\
&\leq C \sum_{I_{0l2}} \Big[p^{-2l-(2-2r)/(1+r)} + \exp\{-(\log p)^{1+\alpha_0/2}/4\} \Big] \\
&\leq C\ \ p^{-(2-2r)/(1+r)+4\gamma(d-l)} = o(1).
\end{aligned} \tag{4.41}$$

The facts (4.38) and (4.41) yield (4.26).

Last but not least, we prove (4.27). Repeat the above argument on I_0^c, and since $I_0^c = I_{0d}$, or $l = d$, we have

$$
\begin{aligned}
& pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} = P(|N_{b_1}| \geq x_p, \ldots, |N_{b_l}| \geq x_p) \\
& = \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min} \geq x_p} \exp(-\frac{1}{2}\mathbf{y}^T \mathbf{U}_l^{-1}\mathbf{y}) d\mathbf{y} \\
& = \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min} \geq x_p, |\mathbf{y}|_{\max} \leq (\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^T \mathbf{U}_l^{-1}\mathbf{y}) d\mathbf{y} \\
& + \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min} \geq x_p, |\mathbf{y}|_{\max} \geq (\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}} \mathbf{U}_l^{-1}\mathbf{y}) d\mathbf{y} \\
& = \frac{1}{(2\pi)^{l/2}|\mathbf{U}_l|^{1/2}} \int_{|\mathbf{y}|_{\min} \geq x_p, |\mathbf{y}|_{\max} \leq (\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}} \mathbf{U}_l^{-1}\mathbf{y}) d\mathbf{y} \\
& + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
& = \frac{1 + O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min} \geq x_p, |\mathbf{y}|_{\max} \leq (\log p)^{1/2+\alpha_0/4}} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y}) d\mathbf{y} \\
& + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
& = \frac{1 + O(\log p)^{-\alpha_0/2}}{(2\pi)^{l/2}} \int_{|\mathbf{y}|_{\min} \geq x_p} \exp(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y}) d\mathbf{y} + O\Big[\exp\{-\frac{1}{4}(\log p)^{1+\alpha_0/2}\}\Big] \\
& = \{1 + o(1)\}\{\frac{2}{\sqrt{(8\pi)}} \exp(-\frac{t}{2})\}^l p^{-2l} \\
& = \{1 + o(1)\}\{\frac{2}{\sqrt{(8\pi)}} \exp(-\frac{t}{2})\}^d p^{-2d}.
\end{aligned}
$$

So

$$
\begin{aligned}
\sum_{I_0^c} pr\{|\mathbf{N}_d|_{min} \geq t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} &= \mathrm{card}(I_0^c)\{1 + o(1)\}\{\frac{2}{\sqrt{(8\pi)}} \exp(-\frac{t}{2})\}^d p^{-2d} \\
&= \{1 + o(1)\} C_q^d \{\frac{2}{\sqrt{(8\pi)}} \exp(-\frac{t}{2})\}^d p^{-2d} \\
&= \{1 + o(1)\} \frac{1}{d!} p^{2d} \{\frac{2}{\sqrt{(8\pi)}} \exp(-\frac{t}{2})\}^d p^{-2d} \\
&= \frac{1}{d!}\{\frac{1}{\sqrt{(2\pi)}} \exp(-\frac{t}{2})\}^d \{1 + o(1)\}, \qquad (4.42)
\end{aligned}
$$

which confirms (4.27).

Using (4.26) and (4.27), we have

$$\sum_{1\le k_1<\ldots<k_d\le q} pr\{|\mathbf{N}_d|_{\min} \ge t_p^{1/2} \pm \epsilon_n(\log p)^{-1/2}\} = \frac{1}{d!}\{\frac{1}{\sqrt{(2\pi)}}\exp(-\frac{t}{2})\}^d\{1+o(1)\}, \quad (4.43)$$

for any $d \ge 2$ and $t \in \mathbb{R}$. Lemma 13 now is verified due to (4.17) and (4.43). □

5. Proof of main theorems

Proof of Theorem 1: Let us first assume $\beta = 0$ and then $\gamma = 1$. The proof of the general case is given at the end of this proof. Denote $\mathbf{V}_n = \sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}/n$, the leading order term of quantity $\mathbf{e}_j^{\mathrm{T}}\mathbf{S}_n\hat{\mathbf{w}}_{i,0}$ is also equivalent with $\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}$. Therefore, it is sufficient to prove the theorem under the leading order term $\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0}$.

We first approximate $\hat{D}_n$ by its counter part $\hat{D}_n^*$ defined by $\hat{D}_n^* = \max_{1\le i,j\le p}\hat{D}_{ij}^{*2}$ and $\hat{D}_{ij}^* = (\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)/\sqrt{\theta_{ij}}$. Based on Theorem 3 in Le & Zhong (2021), we have $\max_{1\le i,j\le p}|\hat{\omega}_{ij,0} - \omega_{ij}^*| = O_p\{\sqrt{(\log p/n)}\}$. Moreover, by Lemma A.3 in Bickel & Levina (2008), we have $\max_{1\le i,j\le p}|v_{ij} - \sigma_{ij}^*| = O_p\{\sqrt{(\log p/n)}\}$, and $|\hat{\kappa} - \kappa| = O_p(1/\sqrt{np})$. Then we have

$$\begin{aligned}|\hat{D}_n/\hat{D}_n^* - 1| &\le \max_{1\le i,j\le p}|\hat{\theta}_{ij,0}/\theta_{ij} - 1| \\ &= \max_{1\le i,j\le p}|\hat{\omega}_{ii,0}v_{jj} - \omega_{ii}^*\sigma_{jj}^*|/(\omega_{ii}^*\sigma_{jj}^*) + |\hat{\kappa} - \kappa| = o_p\{\sqrt{(\log p/n)}\}.\end{aligned}$$

Since $\log p/n \to 0$, we have $|\hat{D}_n - \hat{D}_n^*| = o_p(\hat{D}_n^*)$. Therefore, it is sufficient to prove that

$$pr\{\hat{D}_n^* - 4\log(p) + \log(\log p) \le t\} \to \exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}.$$

Define $t_p = t + 4\log(p) - \log(\log p)$ and $\hat{D}_{n1}^* = \max_{(i,j)\in A/A_0}\hat{D}_{ij}^{*2}$ where $A = \{(i,j) : 1 \le i,j \le p\}$ and $A_0 = \{(i,j) : \omega_{ij}^* \ne 0\}$. Applying Lemma 8, it is enough to show that

$$pr(\hat{D}_{n1}^* \le t_p) \to \exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}. \quad (5.44)$$

Define $\hat{D}^*_{n2} = \max_{(i,j)\in A/B_0} \hat{D}^{*2}_{ij}$ where $B_0 = A_0 \cup A_1$ and $A_1 = \cup_{i=1}^p \{(i,k) : \lim_{p\to\infty} s_0\sigma_{ik} \neq 0,\ \text{for all } (i,k) \notin A_0\}$. Using Lemma 11, we have

$$|pr(\hat{D}^*_{n2} \geq t_p) - pr(\hat{D}^*_{n1} \geq t_p)| \leq pr(\max_{(i,j)\in A_1} \hat{D}^{*2}_{ij} \geq t_p) = o(1).$$

It is then sufficient to show that

$$pr(\hat{D}^*_{n2} \leq t_p) \to \exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}. \tag{5.45}$$

Recall that $D_{ij} = (\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)/\sqrt{\theta_{ij}}$ and $D_{n2} = \max_{(i,j)\in A/B_0} D^2_{ij}$. It then follows that

$$\begin{aligned}
|\hat{D}^{*1/2}_{n2} - D^{1/2}_{n2}| &= |\max_{(i,j)\in A/B_0} |\hat{D}^*_{ij}| - \max_{(i,j)\in A/B_0} |D_{ij}|| \leq \max_{(i,j)\in A/B_0} |\hat{D}^*_{ij} - D_{ij}| \\
&\leq C\sqrt{n} \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n(\hat{\mathbf{w}}_{i,0} - \mathbf{w}^*_i)| \\
&= C\sqrt{n} \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}^*_i)\mathbf{f}_i| \\
&\leq C\sqrt{n} \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}(\mathbf{V}_n - \boldsymbol{\Sigma}^*)\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}^*_i)\mathbf{f}_i| \\
&\quad + C\sqrt{n} \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Sigma}^*\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}^*_i)\mathbf{f}_i|
\end{aligned} \tag{5.46}$$

for some positive constant C where $\boldsymbol{\Omega}^*_i = \mathbf{B}^{\mathrm{T}}_{i,0}\boldsymbol{\Omega}^*\mathbf{B}_{i,0}$ and $\boldsymbol{S}_i = \mathbf{B}^{\mathrm{T}}_{i,0}\boldsymbol{S}^*_n\mathbf{B}_{i,0}$.

For the first term on the right-hand side of (5.46), we have

$$\begin{aligned}
&\sqrt{n} \max_{(i,j)\in A/B_0} |\mathbf{e}_j^{\mathrm{T}}(\mathbf{V}_n - \boldsymbol{\Sigma}^*)\mathbf{B}_{i,0}(\boldsymbol{S}_i^{-1} - \boldsymbol{\Omega}^*_i)\mathbf{f_i}| \\
&\qquad\qquad \leq \sqrt{n}s_0 \max_{1\leq i,j\leq p} |v_{ij} - \sigma^*_{ij}| \max_{1\leq i,j\leq p} |\hat{w}_{ij,0} - \omega^*_{ij}| \\
&\qquad\qquad = O_p(s_0 \log p/\sqrt{n}) = o_p(\sqrt{\log p}).
\end{aligned} \tag{5.47}$$

Applying Lemma 12, the second term on the right-hand side of (5.46) is at the order of $o_p(\sqrt{\log p})$. Then we have $|\hat{D}^{*1/2}_{n2} - D^{1/2}_{n2}| = o_p(\sqrt{\log p})$. Because of the inequality $|\hat{D}^*_{n2} - D_{n2}| \leq 2|D^{1/2}_{n2}|\ |\hat{D}^{*1/2}_{n2} - D^{1/2}_{n2}| + |\hat{D}^{*1/2}_{n2} - D^{1/2}_{n2}|^2$, to verify (5.45) is sufficient to show

$$pr(D_{n2} \leq t_p) \to \exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}. \tag{5.48}$$

Suppose there are k isolated nodes in the true network, for any two nodes i and j belong to this isolated nodes set, we have $(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_i^*)^2/(n\omega_{ii}^*\sigma_{jj}^*) = (\mathbf{e}_i^{\mathrm{T}}\mathbf{V}_n\mathbf{w}_j^*)^2/(n\omega_{jj}^*\sigma_{ii}^*)$. On the set of isolated nodes, we are only maximizing over $[k^2/2]$ components. Thus, D_{n2} involves the maximization of $p^2 - k^2/2$ components D_{ij}^2. For convenience, denote the set that D_{n2} is maximizing over as A/B_0^*. For any $(i,j) \in A/B_0^*, D_{ij}^2 \neq D_{ji}^2$. It is clear that $A/B_0^* \subset A/B_0$.

Re-enumerate the index pairs (i,j) in A/B_0^* to (i_k, j_k), where $k = 1, \dots, q$, for $q = \mathrm{card}(A/B_0^*)$. Since $k = o(p)$ and $\mathrm{card}(A/B_0^*) = p^2\{1 + o(1)\}$, we have $q = p^2\{1 + o(1)\}$. Then, (5.48) is rewritten as

$$pr(\max_{1\le k\le q} {V_k}^2 \le t_p) \to \exp\big\{ - \exp(-t/2)/\sqrt{(2\pi)}\big\} \tag{5.49}$$

where $V_k = \sum_{l=1}^n \mathbf{e}_{j_k}^{\mathrm{T}}\mathbf{X}_l\mathbf{X}_l^{\mathrm{T}}\mathbf{w}_{i_k}^*/\sqrt{n^2\theta_{i_kj_k}}$.

Define $Z_{lk} = \mathbf{e}_{j_k}^{\mathrm{T}}\mathbf{X}_l\mathbf{X}_l^{\mathrm{T}}\mathbf{w}_{i_k}^*$ and $\tau_n = 8C\eta^{-1}\log(p+n)$ where C is some positive constant and η is a constant as in (5.53). Define $\hat{Z}_{lk} = Z_{lk}I\{|Z_{lk}| \le \tau_n\} - E[Z_{lk}I\{|Z_{lk}| \le \tau_n\}]$ as the centralized truncated version of Z_{lk}, and $\hat{V}_k = \sum_{l=1}^n \hat{Z}_{lk}/\sqrt{n^2\theta_{i_kj_k}}$. To show (5.49), it is sufficient to show

$$pr(\max_{1\le k\le q} \hat{V}_k^2 \le t_p) \to \exp\big\{ - \exp(-t/2)/\sqrt{(2\pi)}\big\}. \tag{5.50}$$

The above claim is true if (5.50) implies (5.49). First, note that

$$\begin{aligned}\max_{1\le k\le q}\frac{1}{\sqrt{n}}\sum_{l=1}^n E|Z_{lk}|\ I\{|Z_{lk}| \ge \tau_n\} &= \max_{1\le k\le q}\frac{1}{\sqrt{n}}\sum_{l=1}^n E|Z_{lk}|\ I\{|\eta Z_{lk}| \ge 2C\ \log(p+n)^4\}\\ &\le \max_{1\le k\le q}\max_{1\le l\le n}\sqrt{n}\ E|Z_{lk}|\ I\{|\eta Z_{lk}| \ge 2C\ \log(p+n)^4\}\\ &\le \max_{1\le k\le q}\max_{1\le l\le n}\sqrt{n}(p+n)^{-4}\ E|Z_{lk}|\ \exp\big(\eta|Z_{lk}/(2C)|\big).\end{aligned} \tag{5.51}$$

The last inequality is due to $\exp\{|\eta Z_{lk}/(2C)|\}(p+n)^{-4} \ge 1$ if $|\eta Z_{lk}/(2C)| \ge \log(p+n)^4$.

Assume that the first s_0 components $\mathbf{w}^*_{i_k}$ are non-zeros to simplify notations, that is $\mathbf{w}^*_{i_k} = (\omega^*_{i_k 1}, \ldots, \omega^*_{i_k s_0}, 0, \ldots, 0)^{\mathrm{T}}$, then

$$|Z_{lk}| = |\mathbf{e}^{\mathrm{T}}_{j_k}\mathbf{X}_l\mathbf{X}^{\mathrm{T}}_l\mathbf{w}^*_{i_k}| = |\omega^*_{i_k 1}X_{lj_k}X_{l1} + \cdots + \omega^*_{i_k s_0}X_{lj_k}X_{ls_0}|$$
$$\leq \frac{1}{2}(X^2_{lj_k} + \max_{a=1,\ldots,s_0} X^2_{la})\sum_{a=1}^{s_0}|\omega^*_{i_k a}| \leq C\max_{h\in U_0} X^2_{lh}, \tag{5.52}$$

where $U_0 = \{j_k, l_1, \ldots, l_{s_0}\}$. If $\mathbf{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}}$ is multivariate Gaussian or X_i has a sub-Gaussian tail, then for some $\eta > 0$ we have

$$E\{\exp(\eta\ X^2_i)\} \leq C, \tag{5.53}$$

for $i = 1, 2, \ldots, p$. Applying (5.53), we get

$$E[|Z_{lk}|\exp\{\eta|Z_{lk}/(2C)|\}] \leq CE\{\exp(\eta|Z_{lk}/C|)\} \leq CE[\exp\{\eta\max_{h\in U_0}(X^2_{lh})\}]$$
$$= CE\max_{h\in U_0}\exp(\eta X^2_{lh}) = O(s_0). \tag{5.54}$$

Combining (5.51) and (5.54), we obtain

$$\max_{1\leq k\leq q}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}E|Z_{lk}|\ I\{|Z_{lk}| \geq \tau_n\} = O\{s_0\sqrt{n}nq(p+n)^{-4}\} = o\big\{(\log p)^{-1}\big\}. \tag{5.55}$$

Because $\max_{1\leq k\leq q}|\sum_{l=1}^{n}EZ_{lk}I\{|Z_{lk}| \geq \tau_n\}| \leq \max_{1\leq k\leq q}\sum_{l=1}^{n}E|Z_{lk}|I\{|Z_{lk}| \geq \tau_n\}$, equation (5.55) gives us

$$\max_{1\leq k\leq q}\frac{1}{\sqrt{n}}|\sum_{l=1}^{n}EZ_{lk}\ I\{|Z_{lk}| \geq \tau_n\}| = o\big\{(\log p)^{-1}\big\}. \tag{5.56}$$

In addition, on the set A/B^*_0, we have $EZ_{lk} = E(\mathbf{e}^{\mathrm{T}}_{j_k}\mathbf{X}_l\mathbf{X}^{\mathrm{T}}_l\mathbf{w}_{i_k}) = \mathbf{e}^{\mathrm{T}}_{j_k}\mathbf{\Sigma}^*\mathbf{w}^*_{i_k} = 0$, therefore

$$\max_{1\leq k\leq q}\frac{1}{\sqrt{n}}|\sum_{l=1}^{n}EZ_{lk}| = 0. \tag{5.57}$$

Using (5.56) and (5.57), we get

$$\max_{1\leq k\leq q}\frac{1}{\sqrt{n}}|\sum_{l=1}^{n}EZ_{lk}\ I\{|Z_{lk}| \leq \tau_n\}| = o\big\{(\log p)^{-1}\big\}. \tag{5.58}$$

Hence

$$pr\{\max_{1\le k\le q}|V_k-\hat{V}_k|\ge(\log p)^{-1}\}=pr\{\max_{1\le k\le q}|\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(Z_{lk}-\hat{Z}_{lk})|\ge(\log p)^{-1}\}$$

$$=pr\Big[\max_{1\le k\le q}|\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(Z_{lk}I\{|Z_{lk}|\ge\tau_n\}+EZ_{lk}\ I\{|Z_{lk}|\le\tau_n\})|\ge(\log p)^{-1}\Big]$$

$$=pr\Big[\max_{1\le k\le q}|\frac{1}{\sqrt{n}}\sum_{l=1}^{n}Z_{lk}I\{|Z_{lk}|\ge\tau_n\}|\ge(\log p)^{-1}\Big].$$

It follows that

$$pr\{\max_{1\le k\le q}|V_k-\hat{V}_k|\ge(\log p)^{-1}\}\le pr(\max_{1\le k\le q}\max_{1\le l\le n}|Z_{lk}|\ge\tau_n)$$

$$\le pr\Big[\max_{1\le k\le q}\max_{1\le l\le n}C\{X_{lj_k}^2+\max(X_{l1}^2,\ldots,X_{ls_0}^2)\}\ge\tau_n\Big]$$

$$\le q\cdot pr\{\max_{1\le l\le n}X_{lj_k}^2\ge\tau_n/(2C)\}+q\cdot pr\{\max_{1\le l\le n}\max(X_{l1}^2,\ldots,X_{ls_0}^2)\ge\tau_n/(2C)\}. \tag{5.59}$$

For $j=1,\ldots,p$, we have

$$pr(X_j^2\ge\tau_n/2C)=pr\big\{\exp(\eta X_j^2)\ge\exp(\eta\tau_n/2C)\big\}$$

$$\le \mathrm{E}\Big\{\exp(\eta X_j^2)\Big\}\exp(-\eta\tau_n/2C)\le 2(p+n)^{-4}.$$

This gives us

$$q\cdot pr(\max_{1\le l\le n}X_{lj_k}^2\ge\tau_n/2C)\le nq\max_{1\le j\le p}P(X_j^2\ge\tau_n/2C)\le 2nq/(n+p)^4=o(1) \tag{5.60}$$

and

$$q\cdot pr\{\max_{1\le l\le n}\max(X_{l1}^2,\ldots,X_{ls_0}^2)\ge\tau_n/(2C)\}\le nqs_0\max_{1\le j\le p}pr(X_j^2\ge\tau_n/2C)$$

$$\le 2nqs_0/(n+p)^4=o(1). \tag{5.61}$$

Combining (5.59), (5.60), and (5.61), we obtain $pr\{\max_{1\le k\le q}|V_k-\hat{V}_k|\ge(\log p)^{-1}\}=o(1)$. This means

$$\max_{1\le k\le q}|V_k-\hat{V}_k|=O_P\big\{(\log p)^{-1}\big\}. \tag{5.62}$$

We have

$$|\max_{1\le k\le q} V_k^2 - \max_{1\le k\le q} \hat{V}_k^2| \le 2\max_{1\le k\le q} |\hat{V}_k| \max_{1\le k\le q} |V_k - \hat{V}_k| + \max_{1\le k\le q} |V_k - \hat{V}_k|^2. \tag{5.63}$$

If (5.50) holds, then $\max_{1\le k\le q} \hat{V}_k = O_p(\sqrt{\log p})$. In addition, (5.62) and (5.63) implies that $|\max_{1\le k\le q} V_k^2 - \max_{1\le k\le q} \hat{V}_k^2| = o_p(1)$. As a result, to prove (5.49), it is sufficient to prove (5.50).

Finally, we prove (5.50). Let us denote

$$\tilde{Z}_{lk} = \hat{Z}_{lk}/\sqrt{\omega^*_{i_k i_k}\sigma^*_{j_k j_k}}, \quad \mathbf{W}_l = (\tilde{Z}_{lk_1}, \tilde{Z}_{lk_2}, \ldots, \tilde{Z}_{lk_d}), \tag{5.64}$$

for $l = 1, \ldots, n$, and denote $E_{kj} = \{\hat{V}_{kj}^2 \ge t_p\}$ for any integer $1 \le k_j \le q$. Applying Bonferroni inequality in Lemma 1 for $pr(\max_{1\le k\le q} \hat{V}_k^2 \ge t_p)$, we have

$$\begin{aligned}\sum_{d=1}^{2m} (-1)^{d-1} \sum_{1\le k_1<\cdots<k_d\le q} pr(\bigcap_{j=1}^{d} E_{k_j}) &\le pr(\max_{1\le k\le q} \hat{V}_k^2 \ge t_p)\\ &\le \sum_{d=1}^{2m-1} (-1)^{d-1} \sum_{1\le k_1<\cdots<k_d\le q} pr(\bigcap_{j=1}^{d} E_{k_j}), \end{aligned}\tag{5.65}$$

for any fixed integer $m < [q/2]$.

Rewrite $pr(\bigcap_{j=1}^d E_{k_j})$ as $pr(\bigcap_{j=1}^d E_{k_j}) = pr(|n^{-1/2}\sum_{l=1}^n \mathbf{W}_l|_{\min} \ge t_p^{1/2})$. We will apply Zaitsev approximation to approximate this probability. To this end, we first check the conditions for Zaitsev approximation in Lemma 3. Define

$$\begin{aligned}\boldsymbol{\xi}_i &= n^{-1/2}\mathbf{w}_i^* = n^{-1/2}(\tilde{Z}_{ik_1}, \tilde{Z}_{ik_2}, \ldots, \tilde{Z}_{ik_d})\\ &= n^{-1/2}\{\hat{Z}_{ik_1}/(\omega^*_{i_{k_1}i_{k_1}}\sigma^*_{j_{k_1}j_{k_1}})^{1/2}, \ldots, \hat{Z}_{ik_d}/(\omega^*_{i_{k_d}i_{k_d}}\sigma^*_{j_{k_d}j_{k_d}})^{1/2}\}.\end{aligned}$$

We have $E\boldsymbol{\xi}_i = 0$, for $i = 1, \ldots, n$, and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are independent. We also have

$$\begin{aligned}|(\boldsymbol{\xi}_i, \boldsymbol{u})|^{m-2} &\le ||\boldsymbol{\xi}_i||^{m-2}||\boldsymbol{u}||^{m-2} \le ||\boldsymbol{u}||^{m-2}(2\sqrt{(d/n)}\tau_n)^{m-2} = 2^{m-2}\tau^{m-2}||\boldsymbol{u}||^{m-2}\\ &\le \frac{1}{2}m!\tau^{m-2}||\boldsymbol{u}||^{m-2}\end{aligned}$$

where $m \geq 3$ and $\tau = \sqrt{(d/n)}\tau_n = 8C\eta^{-1}\sqrt{(d/n)}\log(p{+}n)$. It follows that $|\mathrm{E}(\boldsymbol{\xi}_i, \boldsymbol{t})^2(\boldsymbol{\xi}_i, \boldsymbol{u})^{m-2}| \leq 1/2m!\tau^{m-2}||\boldsymbol{u}||^{m-2}\mathrm{E}(\boldsymbol{\xi}_i, \boldsymbol{t})^2$, for $i = 1, \ldots, n$.

Applying Lemma 3, we have

$$pr(|\mathbf{N}_d|_{\min} \geq \sqrt{t_p} + \epsilon_n/\sqrt{\log p}) - pr(|\sum_{l=1}^{n}\mathbf{W}_l|_{\min} \geq \sqrt{nt_p})$$
$$\leq c_1 d^{5/2}\exp\Big\{-\epsilon_n(d^5\log p)^{-1/2}/(\tau c_2)\Big\} \qquad (5.66)$$

where $\mathbf{N}_d = (N_{k_1}, N_{k_2}, \cdots, N_{k_d})^{\mathrm{T}}$ is a d-dimensional multivariate normal distributed random vector with mean vector $\mathrm{E}\mathbf{N}_d = 0$ and covariance matrix $\mathrm{cov}(\mathbf{N}_d) = \mathrm{cov}(\mathbf{W}_1)$. Notice that d is fixed and does not depend on n, p, and

$$c_1 d^{5/2}\exp\{-\frac{\epsilon_n(\log\ p)^{-1/2}}{\tau c_2 d^{5/2}}\} = c_1 d^{5/2}\exp\{-\frac{\epsilon_n(\log\ p)^{-1/2}}{8C\eta^{-1}\sqrt{(d/n)}\log(p+n)c_2 d^{5/2}}\}$$
$$= O\Big[\exp\{-\epsilon_n\sqrt{n}/(\log p)^{3/2}\}\Big] = O(p^{-M}), \qquad (5.67)$$

for some $M > 0$ and $\epsilon_n \to 0$ sufficient slow. The facts (5.66) and (5.67) give us

$$pr(|\mathbf{N}_d|_{\min} \geq \sqrt{t_p} + \epsilon_n/\sqrt{\log p}) - pr\{|\sum_{l=1}^{n}\mathbf{W}_l|_{\min} \geq \sqrt{(nt_p)}\} = O(p^{-M}), \qquad (5.68)$$

for some $M > 0$. Similarly, we can prove

$$pr\{|\sum_{l=1}^{n}\mathbf{W}_l|_{\min} \geq \sqrt{(nt_p)}\} - pr(|\mathbf{N}_d|_{\min} \geq \sqrt{t_p} - \epsilon_n/\sqrt{\log p}) = O(p^{-M}), \qquad (5.69)$$

for some $M > 0$.

Applying (5.65) and (5.69), we get

$$\begin{aligned} pr(\max_{1\leq k\leq q}\hat{V}_k^2 \geq t_p) &\leq \sum_{d=1}^{2m-1}(-1)^{d-1}\sum_{1\leq k_1<\cdots<k_d\leq q} pr(\bigcap_{j=1}^{d}E_{k_j}) \\ &= \sum_{d=1}^{2m-1}(-1)^{d-1}\sum_{1\leq k_1<\cdots<k_d\leq q} pr(|n^{-1/2}\sum_{l=1}^{n}\mathbf{W}_l|_{\min} \geq t_p^{1/2}) \\ &= \sum_{d=1}^{2m-1}(-1)^{d-1}\sum_{1\leq k_1<\cdots<k_d\leq q} pr\{|\mathbf{N}_d|_{\min} \geq t_p^{1/2} - \epsilon_n(\log p)^{-1/2}\} + o(1). \end{aligned} \qquad (5.70)$$

Similarly, applying (5.65) and (5.68), we get

$$pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p) \ge \sum_{d=1}^{2m-1}(-1)^{d-1}\sum_{1\le k_1<\cdots<k_d\le q} pr\{|\mathbf{N}_d|_{\min} \ge t_p^{1/2} + \epsilon_n(\log p)^{-1/2}\} - o(1). \tag{5.71}$$

Combining Lemma 13, (5.65), (5.70), and (5.71), we obtain

$$\sum_{d=1}^{2m}(-1)^{d-1}\frac{1}{d!}\{\frac{1}{\sqrt{2\pi}}\exp(-\frac{t}{2})\}^d\{1+o(1)\} \le pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p)$$
$$\le \sum_{d=1}^{2m-1}(-1)^{d-1}\frac{1}{d!}\{\frac{1}{\sqrt{2\pi}}\exp(-\frac{t}{2})\}^d\{1+o(1)\}.$$

It follows that

$$\limsup_{n\to\infty} pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p) \le \sum_{d=1}^{2m-1}(-1)^{d-1}\frac{1}{d!}\Big\{\frac{1}{\sqrt{(2\pi)}}\exp(-t/2)\Big\}^d.$$

Let $m\to\infty$ then

$$\limsup_{n\to\infty} pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p) \le 1-\exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}. \tag{5.72}$$

Similarly, we get

$$\liminf_{n\to\infty} pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p) \ge 1-\exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}. \tag{5.73}$$

The facts (5.72) and (5.73) give us

$$\lim_{n\to\infty} pr(\max_{1\le k\le q}\hat{V}_k^2 \ge t_p) = 1-\exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}.$$

In other words,

$$\lim_{n\to\infty} pr(\max_{1\le k\le q}\hat{V}_k^2 \le t_p) = \exp\big\{-\exp(-t/2)/\sqrt{(2\pi)}\big\}.$$

This finishes the proof of equation (5.50) and then the result in Theorem 1 holds.

The proof of the general case with $\beta \neq 0$ is similar to the above proof with $\beta = 0$ but we are maximizing over $q = p^2 - k^2/2 = p^2(1 - \beta^2/2)$ components which changes Lemma 13 to

$$\sum_{1 \le k_1 < \cdots < k_d \le q} pr\{|\mathbf{N}_d|_{\min} \ge t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} = \frac{1}{d!}\{\frac{1}{\sqrt{(2\gamma\pi)}} \exp(-t/2)\}^d \{1 + o(1)\}.$$

To verify this, we can repeat the proof of Lemma 13 where equation (4.42) in the lemma is replaced by

$$\begin{aligned}\sum_{I_0^c} pr\{|\mathbf{N}_d|_{\min} \ge t_p^{1/2} \pm \epsilon_n (\log p)^{-1/2}\} &= \{1 + o(1)\} C_q^l \{\frac{2}{\sqrt{(8\pi)}} \exp(-t/2)\}^l p^{-2l} \\ &= p^{2l}\{(1 - \frac{\beta^2}{2})\frac{2}{\sqrt{(8\pi)}} \exp(-t/2)\}^l p^{-2l}\{1 + o(1)\} \\ &= \frac{1}{d!}\{\frac{1}{\sqrt{(2\gamma\pi)}} \exp(-t/2)\}^d \{1 + o(1)\}. \qquad \square\end{aligned}$$

Proof of Theorem 2: Let $\hat{\omega}_{i1,0}^{(j)}$ and $\hat{\omega}_{i1}^{*,(j)}$ be the jth components of non-zeros parts estimators $\hat{\mathbf{w}}_{i1,0}$ and $\hat{\mathbf{w}}_{i1}^*$ that are constructed under the hypothesis $H_2 : \mathcal{E}^* \subsetneq \mathcal{E}_0$ and the true underlying structure $\mathcal{E}^*$, respectively. Denote $\hat{\mathbf{w}}_{i1,0} = (\hat{\omega}_{i1,0}^{(1)}, \ldots, \hat{\omega}_{i1,0}^{(g_i)})^{\mathrm{T}}$ and $\hat{\mathbf{w}}_{i1}^* = (\hat{\omega}_{i1}^{*,(1)}, \ldots, \hat{\omega}_{i1}^{*,(s_i)})^{\mathrm{T}}$, where $g_i \ge s_i$. The asymptotic normality result in (2.2) in the main text gives $\hat{\omega}_{i1,0}^{(j)} = \omega_{i1,0}^{(j)} + O_p(1/\sqrt{n})$. Therefore, under the hypothesis $H_2 : \mathcal{E}^* \subsetneq \mathcal{E}_0$, an estimator for the position j of column i of the precision matrix $\mathbf{w}_i$ that belongs to $\mathcal{E}_0 \cap \mathcal{E}^{*,c}$ is a consistent estimator of 0. Here $\mathcal{E}^{*,c}$is the complement set of $\mathcal{E}^*$. So both $\hat{\mathbf{w}}_{i,0} = \mathbf{B}_{i,0}\hat{\mathbf{w}}_{i1,0}$ and $\hat{\mathbf{w}}_i^* = \mathbf{B}_i^*\hat{\mathbf{w}}_{i1}^*$ are consistent estimators of $\mathbf{w}_i^*$, column i of the underlying precision matrix $\mathbf{\Omega}^*$. Here, $\mathbf{B}_{i,0}$, and $\mathbf{B}_i^*$ are 0, 1 matrices corresponding the hypothesis H_2 and the underlying true structure.

Rewrite $\hat{\mathbf{w}}_{i1,0} = (\hat{\omega}_{i1,0}^{(1)}, \ldots, \hat{\omega}_{i1,0}^{(g_i)})^{\mathrm{T}} = \hat{\mathbf{w}}_{i1,01} + \hat{\mathbf{w}}_{i1,02}$, where $\hat{\mathbf{w}}_{i1,01} = (\hat{\omega}_{i1}^{*,(1)}, \ldots, \hat{\omega}_{i1}^{*,(s_i)}, 0, \ldots, 0)^{\mathrm{T}}$ and $\hat{\mathbf{w}}_{i1,02} = (\hat{\omega}_{i1,0}^{(1)} - \hat{\omega}_{i1}^{*,(1)}, \ldots, \hat{\omega}_{i1,0}^{(s_i)} - \hat{\omega}_{i1}^{*,(s_i)}, \hat{\omega}_{i1,0}^{(s_i+1)}, \ldots, \hat{\omega}_{i1,0}^{(g_i)}))^{\mathrm{T}}$. Notice that, $\mathbf{B}_{i,0}\hat{\mathbf{w}}_{i1,01} = \hat{\mathbf{w}}_i^*$. In addition, we also have $\hat{\omega}_{i1,0}^{(k)} - \hat{\omega}_{i1}^{*,(k)} = (\hat{\omega}_{i1,0}^{(k)} - \omega_{i1}^{(k)}) - (\hat{\omega}_{i1}^{*,(k)} - \omega_{i1}^{(k)}) = O_p(1/\sqrt{n})$, for all $1 \le k \le s_i$ and $\hat{\omega}_{i1,0}^{(m)} = O_p(1/\sqrt{n})$ for all $s_i + 1 \le m \le g_i$. Thus componentwise, all elements of $\hat{\mathbf{w}}_{i1,02}$ are at the order $O_p(1/\sqrt{n})$.

The test statistic under H_2 is $\hat{D}_n = \max_{1\le i,j\le p}(\mathbf{e}_j^{\mathrm{T}}\boldsymbol{S}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)^2/\hat{\theta}_{ij,0}$, so its leading order terms for the numerator and denominator are the same as the test statistic constructed under the underlying true structure, $\mathcal{E}_*$. In other words, the limiting distribution of the test statistic constructed under the hypothesis H_2 is the same as the test statistic constructed under the underlying true structure, $\mathcal{E}^*$. □

Proof of Theorem 3: We will show that, under H_0, the modified test statistic $\tilde{D}_n$ converges to the same distributions as $\hat{D}_n$ as in Theorem 1. We first note that $\tilde{D}_{ij}^2 = \hat{D}_{ij}^2$ if $\Delta_i = 0$ for all $i = 1, \dots, p$. Then,

$$\begin{aligned} pr(\tilde{D}_n \le t_p) &= pr(\tilde{D}_n \le t_p, \Delta_i = 0) + pr(\tilde{D}_n \le t_p, \Delta_i \ne 0) \\ &= pr(\hat{D}_n \le t_p) + pr(\tilde{D}_n \le t_p, \Delta_i \ne 0). \end{aligned}$$

Since $pr(\tilde{D}_n \le t_p, \Delta_i \ne 0) \le pr(\Delta_i \ne 0)$, it is sufficient to show that $pr(\Delta_i \ne 0) = 0$ under H_0. For any $(i,j) \in \mathcal{E}$ but $(i,j) \notin \mathcal{E}_0$, under H_0, $\Delta_{ij} = 0$ according to the definition of $\mathbf{B}_{i,0}$. Thus, it is enough to show:

$$pr(\max_{(i,j)\in\mathcal{E}_0} \Delta_{ij} = 0) = 1. \tag{5.74}$$

To this end, we note

$$\begin{aligned} pr(\max_{(i,j)\in\mathcal{E}_0} \Delta_{ij} = 0) &= pr(\min_{i=1,\dots,p,j=1,\dots,s_i} |\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} > \delta_n) \\ &= 1 - \cup_{i=1,\dots,p,j=1,\dots,s_i} pr(|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \le \delta_n) \\ &\ge 1 - \sum_{i=1}^{p}\sum_{j=1}^{s_i} pr(|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \le \delta_n). \end{aligned}$$

Under H_0, $\omega_{i1,0}^{(j)} \ne 0$, and hence $w_{i1,0}^{(j)}/\sigma_{i1,0}^{(j)} = C_{ij}\sqrt{n}$ for some constants C_{ij}. Then, for $\delta_n \asymp \sqrt{\log(n)}$, we have

$$\begin{aligned} pr\Big(|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \le \delta_n\Big) &= pr\Big(- \delta_n - \frac{\omega_{i1,0}^{(j)}}{\sigma_{i1,0}^{(j)}} \le \frac{\hat{\omega}_{i1,0}^{(j)} - \omega_{i1,0}^{(j)}}{\sigma_{i1,0}^{(j)}} \le \delta_n - \frac{\omega_{i1,0}^{(j)}}{\sigma_{i1,0}^{(j)}}\Big) \\ &\le \Phi(\delta_n - C_{ij}\sqrt{n}) \asymp \exp(-C_{ij}^2 n/2)/\sqrt{n}, \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard normal. Under Condition (C2) and $s_0 \asymp o(\sqrt{n})$, we have $\sum_{i=1}^{p}\sum_{j=1}^{s_i} pr(|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \leq \delta_n) \to 0$. Therefore, under the null hypothesis H_0, (5.74) holds and the asymptotic distributions of $\hat{D}_n$ and $\tilde{D}_n$ are the same when $\delta_n \asymp \sqrt{\log(n)}$ and $C_n > 0$.

□

Proof of Theorem 4: If $\mathcal{E}_0$ specified under the null hypothesis H_0 includes the true network structure $\mathcal{E}^*$, then there exist some $\omega_{i1,0}^{(j)} = 0$, say $\omega_{i_0 1,0}^{(j_0)} = 0$ and the corresponding $\hat{\omega}_{i_0 1,0}^{(j_0)}$ are consistent estimators of $\omega_{i_0 1,0}^{(j_0)} = 0$ for some $i_0 \in \{1,\ldots,p\}$ and $j_0 \in \{1,\ldots,s_i\}$. This event happens with probability one because

$$
\begin{aligned}
&pr(\Delta_{i1} \neq 0 \text{ for some } i = 1,\ldots,p) = pr\Big(\cup_{i=1}^{p}\cup_{j=1}^{s_i}\{|\hat{\omega}_{i1,0}^{(j)}|/\hat{\sigma}_{i1,0}^{(j)} \leq \delta_n\}\Big)\\
&\geq pr\Big(\{|\hat{\omega}_{i_0 1,0}^{(j_0)}|/\hat{\sigma}_{i_0 1,0}^{(j_0)} \leq \delta_n\}\Big) = pr\Big(-\delta_n \leq \hat{\omega}_{i_0 1,0}^{(j_0)}/\hat{\sigma}_{i_0 1,0}^{(j_0)} \leq \delta_n\Big)\\
&= \Phi(\delta_n) - \Phi(-\delta_n) \to 1.
\end{aligned}
$$

This implies that $pr(\Delta_i = 0 \text{ for all } i) = 0$. It follows that

$$
\begin{aligned}
pr(\tilde{D}_n > t_p) &= pr(\tilde{D}_n > t_p, \Delta_i = 0 \text{ for all } i) + pr(\tilde{D}_n > t_p, \Delta_i \neq 0 \text{ for some } i)\\
&= pr(\tilde{D}_n > t_p, \Delta_i \neq 0 \text{ for some } i).
\end{aligned}
$$

When the event $\{\Delta_i \neq 0 \text{ for some } i\}$ happens, there exists at least one $\Delta_{ij} = C_n \asymp \sqrt{\log(p)} \neq 0$. Without loss of generality, assume that there exists one $\Delta_{ij^*} = C_n = C\sqrt{\log(p)} \neq 0$ and $\sigma_{jj^*} \neq 0$ for some j, then, in probability, we have $(\mathbf{e}_j^T\mathbf{V}_n\Delta_i)^2/\hat{\theta}_{ij,0} \geq C_n^2(\sum_{l=1}^{n} X_{lj}X_{lj^*})^2/n^2\hat{\theta}_{ij,0} \to C^2\log(p)(\sigma_{jj}^*\sigma_{j^*j^*}^* + 2\sigma_{jj^*}^{*2})/(\omega_{ii}^*\sigma_{jj}^* + 1)$, for some positive constant C.

Applying Theorem 1, for a small $\epsilon > 0$, $pr\{\max_{i,j}\hat{D}_{ij}^2 \leq (4+\epsilon)\log(p)\} \to 1$. Using the definition of $\tilde{D}_{ij}^2$, we have the following decomposition of $\tilde{D}_{ij}^2$,

$$
\tilde{D}_{ij}^2 = \hat{D}_{ij}^2 + \{(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\Delta_i)^2 + 2(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\hat{\mathbf{w}}_{i,0} - \mathbf{e}_j^{\mathrm{T}}\mathbf{e}_i)\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\Delta_i\}/\hat{\theta}_{ij,0}.
$$

If $C > 4\max_{i,j}(\omega_{ii}^*\sigma_{jj}^* + 1)/(\sigma_{ii}^*\sigma_{jj}^* + 2\sigma_{ij}^{*2})$, then $\max_{i,j}\tilde{D}_{ij}^2 \asymp \max_{i,j}(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\Delta_i)^2/\hat{\theta}_{ij,0}$ with probability one and hence

$$pr(\tilde{D}_n > t_p, \Delta_i \neq 0 \text{ for some } i) = pr\big\{\max_{1\le i,j\le p}(\mathbf{e}_j^{\mathrm{T}}\mathbf{V}_n\Delta_i)^2/\hat{\theta}_{ij,0} > t_p, \Delta_i \neq 0 \text{ for some } i\big\}$$

$$\geq pr\big\{C_n^2(\sum_{i=1}^n X_{ij}X_{ij^*})^2/n^2\hat{\theta}_{ij,0} > t_p\big\} \to 1.$$

So, $pr(\tilde{D}_n > t_p) \to 1$ for all the alternatives in H_2 where $\mathcal{E}_0$ includes $\mathcal{E}^*$.

If $\mathcal{E}_0 \neq \mathcal{E}^*$ in H_1 but not in H_2, then there exist pairs of $k \neq l$ such that $\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_{l,0} - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l \neq 0$.

By the construction of the estimator $\hat{\mathbf{w}}_{l,0}$, it could be shown that there exist $\hat{\mathbf{w}}_l^*$ and constants c_{kl} such that $\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_{l,0} - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l = \mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\hat{\mathbf{w}}_l^* - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l + c_{kl}$, where $\hat{\mathbf{w}}_l^*$ is a consistent estimator of $\mathbf{w}_l^*$ such that $\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_l^* - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l = 0$. It follows that we can decompose the test statistic $\hat{D}_n$ as following. Using Theorem 1 and condition (C2), we may find the leading order term as $\hat{D}_n = \max_{1\le k,l\le p}\big\{(\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_l^* - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l)^2 + 2(\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}^*\mathbf{w}_l^* - \mathbf{e}_k^{\mathrm{T}}\mathbf{e}_l)c_{kl} + c_{kl}^2\big\}/\hat{\theta}_{kl,0} \asymp \max_{1\le k,l\le p} c_{kl}^2/\theta_{kl,0} \asymp n$. Then, we have $P(\hat{D}_n - 4\log p + \log(\log p) \to \infty) = 1$ as $n \to \infty$ under condition (C2). Hence, $\tilde{D}_n$ is also consistent for any fixed alternatives in H_1 but not in H_2. In summary, Theorem 4 is proved. □

6. Additional simulation results

6.1 Simulation with non-Gaussian random vectors

We investigate the performance of the proposed test statistics under model misspecification. The simulation settings are the same as that in Section 4.2 of the main text for $s_0 = 4$, except that data $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are not drawn from a multivariate normal distribution. Instead, they are generated from the multivariate model specified in Assumption (D1) using the following three steps.

(1) Generate np independent observations $(h_{ij})_{p\times n}$ from a Gamma$(\alpha = 2, \beta = 1)$ distribution,

where α is the shape parameter and β is the scale parameter. Normalize the dataset using the transformation $z_{ij} = (h_{ij} - 2)/\sqrt{2}$. The standardized values z_{ij} are then assigned to a $p \times n$ matrix $Z = (z_{ij})_{p\times n}$.

(2) Perform eigenvalue decomposition on the underlying covariance matrix under the null hypothesis, $\mathbf{\Sigma}^* = Q\Lambda Q^T$. Denote $\Gamma = Q\Lambda^{1/2}$. It follows that $\Gamma\Gamma^T = \mathbf{\Sigma}^*$, where Γ is a $p \times p$ matrix.

(3) Use $\mathbf{X} = (\Gamma Z)^T$ as the $n \times p$ observed data matrix, where $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$.

We applied the test statistics $\hat{D}_n$ and $\tilde{D}_n$ from the main paper, using the estimated κ given by $\hat{\kappa} = \frac{1}{np}\sum_{i=1}^{n} s_{jj}^{-4}(X_{ij} - \bar{X}_i)^4$. Table 1 reports the empirical sizes and powers of the proposed test statistics $\hat{D}_n$ and $\tilde{D}_n$. We observe that the proposed tests perform reasonably well, with empirical sizes close to the nominal 5% level under the null hypothesis and empirical powers close to one, indicating the consistency of the proposed tests. Overall, the proposed tests are robust with respect to the Gaussian distribution assumption.

6.2 Simulation with an underlying sparse graph and small signals in Ω^*

In this subsection, we report simulation studies for examining the performance of the proposed tests when the underlying graph is sparse. The simulation settings are the same as that in Section 4.2 except that the underlying graph structure is a sparse matrix generated randomly. More specifically, we construct the random sparse structure, its nested structure, and included structure as follows. We first choose an identical matrix $\mathbf{I}_p$, then for every column of $\mathbf{I}_p$, we randomly assign the weight 0.8 to one of its elements. The obtained matrix is denoted as $\mathbf{B}$. Then the underlying precision matrix is chosen as $\mathbf{B}\mathbf{B}^{\mathrm{T}}$. For the nested structure, we replace

Table 1: Type 1 error and empirical power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for nested and included structures with non-Gaussian data.

			$\hat{D}_n$			$\tilde{D}_n$		
			Empirical	Power of $\hat{D}_n$		Empirical	Power of $\tilde{D}_n$	
s_0	n	p/n	Size	Nested	Included	Size	Nested	Included
4	500	0.50	0.080	1.000	0.080	0.080	1.000	0.210
		1.00	0.010	1.000	0.010	0.010	1.000	0.070
		2.00	0.020	1.000	0.040	0.020	1.000	0.120
	1000	0.50	0.030	1.000	0.030	0.030	1.000	1.000
		1.00	0.020	1.000	0.020	0.020	1.000	1.000
		2.00	0.030	1.000	0.030	0.030	1.000	1.000

the first column of $\mathbf{B}$ by the first column of $\mathbf{I}_p$ and obtain matrix $\mathbf{B}_1$. Then we use $\mathbf{B}_1\mathbf{B}_1^{\mathrm{T}}$ as the nested structure. For the included structure, we randomly assign 0.8 to one element of the first column of $\mathbf{B}$, and denote it as the matrix $\mathbf{B}_2$. We then use $\mathbf{B}_2\mathbf{B}_2^{\mathrm{T}}$ as the included structure.

Table 2 reports the empirical size, power, and running time (in seconds) of the two proposed test statistics, $\hat{D}_n$ and $\tilde{D}_n$. For the test statistic $\tilde{D}_n$, we choose $C_n = 0.3$ and $\delta_n = \sqrt{\log(n)}$. We observe that Table 2 has a similar pattern with that in Table 3 of the main text. Both tests maintain the type I error at the nominal level and exhibit comparable power in detecting the alternative with the nested structure. The modified test statistic $\tilde{D}_n$ outperforms $\hat{D}_n$ when dealing with the included structure. We also include the average computation time for each simulation replication in Table 2. Both tests demonstrate similar computational complexity in terms of running time.

Table 3 examines the performance of our proposed tests when the signal size in $\mathbf{\Omega}^*$ is

Table 2: Empirical size and power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for both nested and included structures when the true network structure is a random sparse matrix.

			$\hat{D}_n$				$\tilde{D}_n$			
				Power		Running		Power		Running
s_0	n	p/n	Size	Nested	Included	Time	Size	Nested	Included	Time
4	500	0.50	0.030	1.000	0.030	0.178	0.030	1.000	0.700	0.18
		1.00	0.040	1.000	0.040	1.09	0.040	1.000	0.970	1.10
		2.00	0.030	1.000	0.030	6.93	0.030	1.000	1.000	6.96
	1000	0.50	0.010	1.000	0.010	1.14	0.010	1.000	1.000	1.14
		1.00	0.060	1.000	0.060	6.54	0.060	1.000	1.000	6.54
		2.00	0.020	1.000	0.020	56.21	0.020	1.000	0.990	56.26
6	500	0.50	0.030	1.000	0.030	0.17	0.030	1.000	1.000	0.17
		1.00	0.020	1.000	0.020	1.11	0.020	1.000	0.890	1.11
		2.00	0.040	1.000	0.040	7.11	0.040	1.000	1.000	7.15
	1000	0.50	0.020	1.000	0.020	1.13	0.020	1.000	1.000	1.14
		1.00	0.060	1.000	0.060	6.54	0.060	1.000	1.000	6.54
		2.00	0.000	1.000	0.000	55.25	0.000	1.000	0.990	55.27

Table 3: Type I error and empirical power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for nested and included structures as affected by sample size at the sparsity level $s_0 = 10$

		$\hat{D}_n$				$\tilde{D}_n$			
			Power		Running		Power		Running
p	n	Size	Nested	Included	Time	Size	Nested	Included	Time
500	500	0.020	0.020	0.020	1.8	0.020	0.020	0.020	1.8
	5000	0.020	0.020	0.020	1.7	0.020	0.020	0.040	1.7
	20000	0.030	0.020	0.020	1.7	0.030	0.130	0.520	1.7
1000	500	0.010	0.020	0.020	11.1	0.030	0.050	0.040	11.1
	5000	0.030	0.030	0.030	11.1	0.030	0.040	0.040	11.1
	20000	0.040	0.050	0.040	10.9	0.040	0.130	0.460	11.1

small. The simulation settings mirror those described in Section 4.2 of the main text, with the exception that the bandwidth s_0 is increased to $s_0 = 10$ so that the smallest signal approaches zero, presenting a more challenging scenario. We use dimensions $p = 500$ and 1000, and sample sizes $n = 500$, 1000, and 20000. As shown in the table, increasing the sample size does not significantly impact the performance of the naive test statistic $\hat{D}_n$ or the modified test statistic $\tilde{D}_n$.

6.3 Simulation for test statistics with different estimators for $\mathbf{\Omega}_0$

To evaluate the performance of the test statistics using different estimators for the precision matrix $\mathbf{\Omega}_0$, we conducted a simulation study. We compared test statistics constructed similarly to $\hat{D}_n$, but with alternative estimators for $\mathbf{\Omega}_0$. Specifically, we used the GLASSO estimator

(Friedman, 2019) with a known graphical structure (denoted as $\hat{D}_{n,G}$) and a modified positive definite and symmetric estimator described in Section 2 of the main text (denoted as $\hat{D}_{n,PSD}$).

The positive definite and symmetric estimator was obtained by symmetrizing $\hat{\boldsymbol{\Omega}}$ using

$$\hat{\boldsymbol{\Omega}}_1 = \frac{\hat{\boldsymbol{\Omega}} + \hat{\boldsymbol{\Omega}}^T}{2}.$$

To ensure that $\hat{\boldsymbol{\Omega}}_1$ is positive definite, we applied a small perturbation to its eigenvalues:

$$\hat{\boldsymbol{\Omega}}_\tau = \hat{\boldsymbol{\Omega}}_1 + \tau \boldsymbol{I}_p,$$

where

$$\tau = \left(\left| \Lambda_{\min}(\hat{\boldsymbol{\Omega}}_1) \right| + n^{-1/2} \right) \cdot 1 \left\{ \Lambda_{\min}(\hat{\boldsymbol{\Omega}}_1) \leq 0 \right\}.$$

For further details, see Remark 1 in Liu (2015).

Tables 4 and 5 compare the performance of the proposed test statistic $\hat{D}_n$ with $\hat{D}_{n,G}$ and $\hat{D}_{n,PSD}$, respectively.

Table 4 demonstrates that $\hat{D}_n$ performs slightly better in terms of power compared to $\hat{D}_{n,G}$, which is based on the GLASSO estimator. Additionally, $\hat{D}_n$ consistently shows greater efficiency than $\hat{D}_{n,G}$ with respect to computational time.

Results from Table 5 reveal that $\hat{D}_n$ and $\hat{D}_{n,PSD}$ exhibit similar performance in terms of empirical size and computational time. Both statistics have comparable power for detecting included structure alternatives. However, $\hat{D}_{n,PSD}$ slightly outperforms $\hat{D}_n$ for nested structure alternatives, likely due to the symmetric information of $\boldsymbol{\Omega}_0$ being utilized in $\hat{D}_{n,PSD}$ but not in $\hat{D}_n$.

6.4 Tuning parameter selection and computational time

In this subsection, we investigate the sensitivity of the proposed test $\tilde{D}_n$ to the choices of tuning parameters C_n and δ_n. Tables 6 and 7 report the performance of $\tilde{D}_n$ for various choices of tuning

Table 4: Type I error and empirical power of the test statistics $\hat{D}_n$ and $\hat{D}_{n,G}$ (with $\mathbf{\Omega}_0$ estimated by the GLASSO) for nested and included structures

			$\hat{D}_n$				$\hat{D}_{n,G}$			
				Power		Running		Power		Running
s_0	n	p/n	Size	Nested	Included	Time	Size	Nested	Included	Time
4	500	0.50	0.020	1.000	0.010	0.20	0.030	1.000	0.020	0.49
		1.00	0.030	1.000	0.040	1.19	0.040	1.000	0.040	3.50
		2.00	0.030	1.000	0.020	6.39	0.030	1.000	0.020	22.30
	1000	0.50	0.040	1.000	0.040	1.19	0.050	1.000	0.030	3.39
		1.00	0.030	1.000	0.030	6.33	0.030	1.000	0.030	21.62
		2.00	0.020	1.000	0.010	54.70	0.010	1.000	0.010	187.26
6	500	0.50	0.030	0.100	0.020	0.19	0.020	0.050	0.020	0.42
		1.00	0.050	0.060	0.050	1.23	0.050	0.040	0.040	2.90
		2.00	0.020	0.040	0.030	6.37	0.020	0.030	0.020	17.80
	1000	0.50	0.030	0.720	0.040	1.21	0.030	0.190	0.040	2.87
		1.00	0.070	0.700	0.060	6.35	0.060	0.150	0.060	17.41
		2.00	0.040	0.480	0.040	54.71	0.040	0.050	0.040	148.24

parameters with sample sizes $n = 500$ and $n = 1000$, respectively. We find that the performance of the proposed test is influenced by the choices of both δ_n and C_n, but it is more sensitive to the selection of δ_n, as the empirical size and power remain similar across different C_n values when δ_n is fixed. The test performs well when δ_n is of the order $\{\log(n)\}^{1/k}$ for $k \geq 2$, particularly for $k = 4$ and a sample size of $n = 1000$. This finding is consistent with the recommended choice in Theorems 3 and 4 of the main text. Moreover, in Section 7 of the supplemental material, a

Table 5: Type I error and empirical power of the test statistics $\hat{D}_n$ and $\hat{D}_{n,PSD}$ (with $\mathbf{\Omega}_0$ estimated by the positive definite and symmetric estimator defined in Section 2 of the main paper) for nested and included structures

			$\hat{D}_n$				$\hat{D}_{n,PSD}$			
				Power		Running		Power		Running
s_0	n	p/n	Size	Nested	Included	Time	Size	Nested	Included	Time
4	500	0.50	0.070	1.000	0.080	0.20	0.100	1.000	0.090	0.20
		1.00	0.060	1.000	0.050	0.65	0.070	1.000	0.060	0.70
		2.00	0.020	1.000	0.030	7.93	0.030	1.000	0.030	8.13
	1000	0.50	0.030	1.000	0.040	1.16	0.030	1.000	0.030	1.19
		1.00	0.060	1.000	0.060	7.85	0.050	1.000	0.040	8.04
		2.00	0.060	1.000	0.060	67.03	0.060	1.000	0.050	68.28
6	500	0.50	0.020	0.080	0.020	0.20	0.020	0.320	0.010	0.21
		1.00	0.020	0.040	0.020	1.36	0.020	0.280	0.020	1.40
		2.00	0.030	0.050	0.030	8.36	0.030	0.170	0.030	8.54
	1000	0.50	0.020	0.740	0.010	1.24	0.020	1.000	0.010	1.27
		1.00	0.040	0.640	0.040	8.23	0.040	0.960	0.040	8.44
		2.00	0.040	0.510	0.040	58.89	0.040	0.970	0.040	58.92

data-driven procedure is developed to choose δ_n and C_n.

We illustrate the running time, in seconds (s), for the modified test statistic $\tilde{D}_n$ as a function of the data dimension in Figure 1. This figure is based on the simulation study described in Section 4.2 of the main text. We plot the data dimension p against the square root of the running time for four scenarios: (S1) $n = 500$, $s_0 = 4$; (S2) $n = 1000$, $s_0 = 4$; (S3) $n = 500$, $s_0 = 6$; and

Table 6: Type I error and empirical power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for nested and included structures as affected by tuning parameters, $n = 500, p = 1000, s_0 = 4, \delta_n = (\log\, n)^{1/k}$

		$\hat{D}_n$				$\tilde{D}_n$			
			Power		Running		Power		Running
C_n	k	Size	Nested	Included	Time	Size	Nested	Included	Time
.05	1	0.050	1.000	0.040	6.47	1.000	1.000	1.000	6.48
	2	0.020	1.000	0.020	6.38	0.020	1.000	1.000	6.40
	4	0.030	1.000	0.050	6.36	0.030	1.000	0.990	6.39
.2	1	0.070	1.000	0.070	6.37	0.160	1.000	1.000	6.42
	2	0.040	1.000	0.050	6.39	0.040	1.000	1.000	6.42
	4	0.020	1.000	0.020	6.34	0.020	1.000	1.000	6.37
.5	1	0.030	1.000	0.040	6.36	1.000	1.000	1.000	6.38
	2	0.010	1.000	0.010	6.36	0.820	1.000	1.000	6.40
	4	0.030	1.000	0.040	6.34	0.040	1.000	1.000	6.37
1	1	0.040	1.000	0.040	6.39	1.000	1.000	1.000	6.42
	2	0.020	1.000	0.000	6.36	0.850	1.000	1.000	6.37
	4	0.050	1.000	0.040	6.34	0.070	1.000	1.000	6.38

(S4) $n = 1000$, $s_0 = 6$.

In general, we observe a linear relationship between the data dimension p and the square root of the running time. This indicates that the computational time grows quadratically with respect to p, i.e., the computational time is on the order of p^2 with respect to the data dimension.

Table 7: Type I error and empirical power of the test statistics $\hat{D}_n$ and $\tilde{D}_n$ for nested and included structures as affected tuning parameters, $n = 1000, p = 1000, s_0 = 4, \delta_n = (\log n)^{1/k}$

		$\hat{D}_n$				$\tilde{D}_n$			
			Power		Running		Power		Running
C_n	k	Size	Nested	Included	Time	Size	Nested	Included	Time
.05	1	0.030	1.000	0.030	6.40	1.000	1.000	1.000	6.41
	2	0.040	1.000	0.040	6.30	0.040	1.000	1.000	6.32
	4	0.050	1.000	0.030	6.30	0.050	1.000	1.000	6.32
.2	1	0.100	1.000	0.090	6.31	0.090	1.000	1.000	6.32
	2	0.030	1.000	0.040	6.31	0.030	1.000	1.000	6.32
	4	0.050	1.000	0.050	6.30	0.050	1.000	1.000	6.32
.5	1	0.040	1.000	0.030	6.31	1.000	1.000	1.000	6.32
	2	0.070	1.000	0.050	6.31	0.070	1.000	1.000	6.33
	4	0.050	1.000	0.050	6.29	0.050	1.000	1.000	6.30
1	1	0.070	1.000	0.070	6.31	1.000	1.000	1.000	6.32
	2	0.020	1.000	0.020	6.31	0.020	1.000	1.000	6.32
	4	0.070	1.000	0.070	6.28	0.070	1.000	1.000	6.30

7. A data-driven procedure for choosing tuning parameters in the proposed test

The key idea behind the consistency-enhanced test is to introduce pseudo signals to edges whose underlying weights are essentially zero but are included in the specified null structure $\mathcal{E}_0$. Specifically, for all edges with $\omega^*_{ij} = 0$, the estimated weights $\hat{\omega}_{ij}$ should remain close to zero.

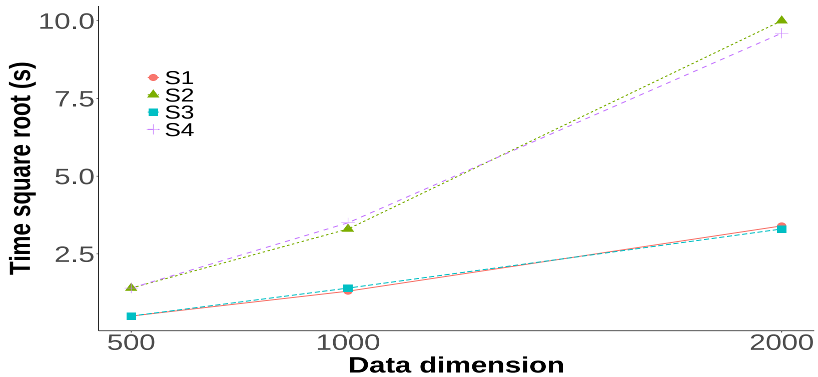


Figure 1: The square root of the running time for the test statistic $\tilde{D}_n$ versus the data dimension is shown for different combinations of n, p, and s_0 in the following scenarios. $S_1 : (n, s_0) = (500, 4)$, $S_2 : (n, s_0) = (1000, 4)$, $S_3 : (n, s_0) = (500, 6)$, $S_4 : (n, s_0) = (1000, 6)$.

When pseudo signals are added to an edge (i, j) where $\omega^*_{ij} = 0$, replacing

$$\tilde{\mathbf{w}}^{(j)}_{i1,0} = \hat{\mathbf{w}}^{(j)}_{i1,0} + \Delta^{(j)}_{i1}$$

with $\tilde{\mathbf{w}}^{(j)}_{i1,0} = \Delta^{(j)}_{i1}$ in the test statistic $\tilde{D}_n$ should not significantly alter its value. Here, the pseudo signal is defined as

$$\Delta^{(j)}_{i1} = C_n I \left\{ \frac{|\hat{\omega}^{(j)}_{i1,0}|}{\hat{\sigma}^{(j)}_{i1,0}} \leq \delta_n \right\}.$$

This motivates us to compare the test statistic $\tilde{D}_n$ with a modified version:

$$\tilde{D}^*_n = \max_{1\leq i,j\leq p} \frac{\left(\mathbf{e}_j^{\mathrm{T}} \boldsymbol{S}_n \tilde{\mathbf{w}}^*_{i,0} - \mathbf{e}_j^{\mathrm{T}} \mathbf{e}_i\right)^2}{\hat{\theta}_{ij,0}} = \max_{1\leq i,j\leq p} (\tilde{D}^*_{ij})^2,$$

where $\tilde{\mathbf{w}}^*_{i1,0} = \left(\tilde{\omega}^{*(1)}_{i1,0}, \ldots, \tilde{\omega}^{*(s_i)}_{i1,0}\right)^{\mathrm{T}}$, with

$$\tilde{\mathbf{w}}^{*(j)}_{i1,0} = \hat{\mathbf{w}}^{(j)}_{i1,0} I \left(\frac{|\hat{\mathbf{w}}^{(j)}_{i1,0}|}{\hat{\sigma}^{(j)}_{i1,0}} > \delta_n \right) + \Delta^{(j)}_{i1}.$$

It is worth noting that if δ_n is chosen appropriately, the distributions of $\tilde{D}_n$ and $\tilde{D}_n^*$ should be approximately the same. However, if δ_n is chosen incorrectly, their distributions will differ.

More specifically, the proposed data-driven procedure for selecting the tuning parameters C_n and δ_n consists of the following steps:

1. Choose C_n based on Theorem 4, where

$$C_n = \sqrt{\log(p)} \max_{i,j} 4 \frac{\omega_{ii}^* \sigma_{jj}^* + 1}{\sigma_{ii}^* \sigma_{jj}^* + 2\sigma_{ij}^{*2}}.$$

   Estimate $\omega_{ii}^*, \sigma_{jj}^*, \sigma_{ij}^*$ using their sample versions. Order the candidate values for δ_n from smallest to largest, given by

$$\mathcal{S}_{\delta_n} := \left\{ \frac{|\hat{\omega}_{i1,0}^{(j)}|}{\hat{\sigma}_{i1,0}^{(j)}}, (i,j) \in \text{Supp}(\mathbf{\Omega}_0) \right\} = \{\delta_n^{(k)} : k = 1, \ldots, |\text{Supp}(\mathbf{\Omega}_0)|\}.$$

   Initialize $k = 1$.

2. Randomly split the data into two equal-sized parts.

3. For the k-th candidate value $\delta_n^{(k)}$ in $\mathcal{S}_{\delta_n}$, compute the test statistic $\tilde{D}_n$ using the first half of the data, denoted as $\tilde{D}_n^k$. Compute the modified test statistic $\tilde{D}_n^*$ using the second half of the data, denoted as $\tilde{D}_n^{*k}$.

4. Repeat Steps 2-3 for B iterations, obtaining two sets of test statistics:

$$\{\tilde{D}_n^{k,b}\}_{b=1}^B \quad \text{and} \quad \{\tilde{D}_n^{*k,b}\}_{b=1}^B.$$

5. Perform a t-test to compare the means of $\{\tilde{D}_n^{k,b}\}_{b=1}^B$ and $\{\tilde{D}_n^{*k,b}\}_{b=1}^B$.

6. If the p-value from the t-test is large (e.g., > 0.05), increment k by 1, set $\delta_n = \delta_n^{(k)}$, and repeat Steps 2-6. If the p-value is small (e.g., < 0.05), terminate the algorithm.

We conducted a small simulation study to evaluate the performance of the proposed data-driven procedure for selecting the tuning parameter δ_n. The data were generated using the same simulation settings as in Table 2 of Section 6.2 of the supplemental material, where the underlying graph $\mathcal{E}^*$ follows a random sparse structure matrix.

The hypothesis test is defined as $H_0 : \mathcal{E} = \mathcal{E}_0$ versus $H_1 : \mathcal{E} \neq \mathcal{E}_0$, where $\mathcal{E}_0 = \mathcal{E}^* \cup \mathcal{E}_1$, and $\mathcal{E}_1$ is a banded structure with bandwidth 2, i.e., $\mathcal{E}_1 = \{(i, j) : |i - j| < 3\}$.

All simulation results are based on 100 replications. For each replication, we apply the above data-driven procedure to select the tuning parameters C_n and δ_n, with $B = 20$. Table 8 summarizes the empirical size and power of the proposed test statistic, with the tuning parameters chosen using the described procedure.

Table 8: Empirical size and power of the proposed test statistic $\tilde{D}_n$ using the tuning parameters selected from the proposed data-driven procedure.

n	p	Empirical Size	Empirical Power
500	250	0.020	1.000
500	500	0.040	1.000
500	1000	0.070	1.000

Figure 2 displays the histograms of the selected values of the constant C_n and threshold δ_n for all 100 replications when $n = 500$ and $p = 250$.

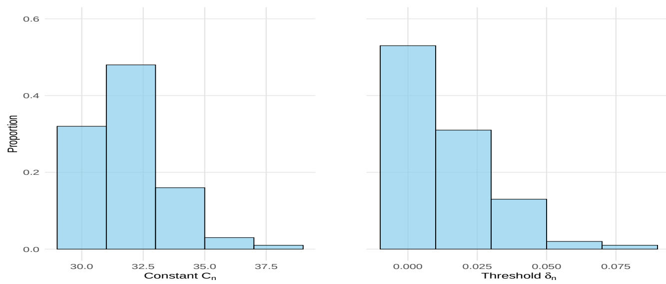


Figure 2: Histograms of the selected tuning parameters C_n and δ_n for $n = 500$ and $p = 250$.

8. Extension: a goodness-of-fit test of graphical structure families

In this section, we outline the generalization of our test for the goodness-of-fit for a family of graphical structures. Specifically, we aim to test the goodness-of-fit for a family of graphical structures indexed by some parameters. We can generalize our test to:

$$H_0 : \mathcal{E}^* \in \mathcal{E}_0(\gamma) \quad \text{vs.} \quad H_1 : \mathcal{E}^* \notin \mathcal{E}_0(\gamma), \tag{8.75}$$

where $\mathcal{E}_0(\gamma)$ represents a family of graphical structures indexed by parameters γ, and γ is unknown. For example, $\mathcal{E}_0(\gamma)$ could represent a banded structure with an unknown bandwidth γ. For this goodness-of-fit test, there is no need to specify a single particular graph; instead, one only needs to specify a family of graphical structures.

To test the hypothesis in (8.75), we propose the following algorithm:

1. Split the sample $\mathcal{S} = \{1, \cdots, n\}$ into two non-overlapping parts, $\mathcal{S}_1$ and $\mathcal{S}_2$, such that $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$.

2. Use the first part of the sample, $\mathcal{S}_1$, to estimate the unknown parameters γ in $\mathcal{E}_0(\gamma)$. Denote the estimated parameters as $\hat{\gamma}$.

3. Apply the proposed test statistic $\tilde{D}_n$ from Section 3 of the paper to test:

$$H_0 : \mathcal{E}^* = \mathcal{E}_0(\hat{\gamma}) \quad \text{vs.} \quad H_1 : \mathcal{E}^* \neq \mathcal{E}_0(\hat{\gamma}), \tag{8.76}$$

   using the second part of the data, $\mathcal{S}_2$. Reject the null hypothesis H_0 in (8.76) if the test statistic $\tilde{D}_n$ exceeds the given critical values.

4. Repeat Steps 1-3 for B iterations and reject the null hypothesis (8.75) if the null hypothesis (8.76) is rejected in more than qB cases (for some $q > 0.5$).

We conducted a small simulation study to illustrate the performance of the proposed algorithm for testing if $\mathcal{E}^*$ belongs to a banded graphical structure $\mathcal{E}_0(\gamma)$, given by the following hypothesis:

$$H_0 : \mathcal{E}^* \in \mathcal{E}_0(\gamma) \quad \text{vs.} \quad H_1 : \mathcal{E}^* \notin \mathcal{E}_0(\gamma), \tag{8.77}$$

where $\mathcal{E}_0(\gamma) = \{(i, j) : |i - j| < \gamma\} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges consisting of node pairs whose corresponding entries in $\mathbf{\Omega}^*$ are non-zero among the nodes $\mathcal{V} = \{1, \cdots, p\}$. Here, γ represents an unknown bandwidth.

We generated $n = 1,000$ independent and identically distributed $p = 1,000$-dimensional random vectors from a multivariate normal distribution with mean zero and precision matrix $\mathbf{\Omega}^*$ under the following two scenarios:

(a) $\mathbf{\Omega}^* = (\omega^*_{ij})_{p\times p}$ where $\omega^*_{ij} = 0.6^{-|i-j|}$ for $|i - j| < 4$ and $\omega^*_{ij} = 0$ otherwise. In this example, the underlying true γ is 4.

(b) $\mathbf{\Omega}^*$ has the same sparse structure as specified in Table 2.

We aim to test the hypothesis in (8.77) to determine if the underlying graphical structure belongs to a banded graphical structure family. For data generated under scenario (a), we evaluate the type I error of the proposed algorithm. For data generated under scenario (b), we assess the empirical power of the proposed algorithm.

The simulation results are based on 100 replications. For each simulated dataset, we used 40% of the data as the training set $\mathcal{S}_1$ and 60% as the test set $\mathcal{S}_2$. Let $\tilde{D}_n(\gamma)$ be the proposed test statistic for testing:

$$H_0 : \mathcal{E}^* = \mathcal{E}_0(\gamma) \quad \text{vs.} \quad H_1 : \mathcal{E}^* \neq \mathcal{E}_0(\gamma),$$

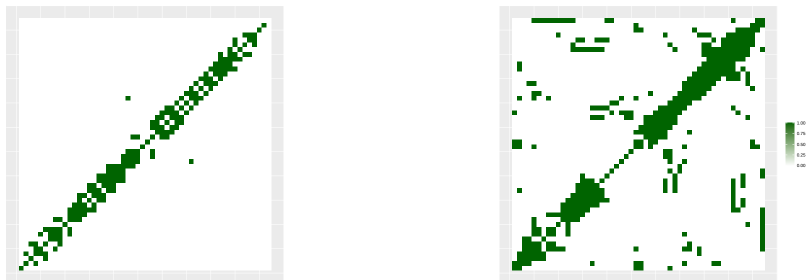
where $\mathcal{E}_0(\gamma)$ is specified in (8.77). We estimate the unknown bandwidth γ by choosing the value that minimizes the test statistic $\tilde{D}_n(\gamma)$. Specifically, we use:

$$\hat{\gamma} = \arg\min_{\gamma} \tilde{D}_n(\gamma).$$

The empirical size of the proposed algorithm was 0.07 for data generated under scenario (a). The empirical power of the proposed algorithm was 1.00 for data generated under scenario (b).

9. Additional information on real data analysis

Figure 3 presents heatmaps of the estimated graphical structures from the real data analyzed in Section 5 of the main paper. The left panel displays the graph estimated using the TIGER approach Liu & Wang (2017), while the right panel shows the graph estimated using the GLASSO method Friedman (2019). For the TIGER method, we applied the default settings to estimate the precision matrix and derived the corresponding graphical structure. For the GLASSO approach, we used a tuning parameter $\rho = 10$. Both estimated network structures suggest that a banded structure is reasonable for this dataset, motivating us to test whether a banded structure adequately models the underlying graph in the real data discussed in the main text.



(a) TIGER Estimation (b) GLASSO Estimation

Figure 3: Heatmaps of estimated graphical structures obtained by (a) TIGER estimation and (b) GLASSO estimation.

References

Bai, Z. D. & Saranadasa, H. (1996). Effect of High Dimension: By an Example of a two sample problem. *The Annals of Mathematical Statistics* **2**, 311-329.

Berman, S. M. (1962). A law of large number for the maximum in a stationary Gaussian sequence. *The Annals of Mathematical Statistics* **33**, 93-97.

Bickel, P. J. & Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199-227.

Cai, T., Liu, W. & Xia, Y. (2013). Two sample covariance testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108**, 265-277.

Chen, S. X., Zhang, L. X., & Zhong, P. S. (2010). Tests for High-Dimensional Covariance Matrices. *Journal of the American Statistical Association*, **105**(490), 810–819.

Demko, S., Moss, W. S. & Smith, P. W. (1984). Decay rates for inverses of band matrices. *Mathematics of Computation* **43**, 491-499.

Gröchenig, K. & Leinert, M. (2006). Symmetry and inverse closedness of matrix algebras and functional calculus for infinite matrices. *Transactions of the American Mathematical Society* **358**, 2695-2711.

Hager, W. W. (1989). Updating the inverse of a matrix. *Society for Industrial and Applied Mathematics* **31**, 221-239.

Hall, P. & Lin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* **38**, 1686-1732.

Liu, H. & Wang, L. (2017). TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *lectronic Journal of Statistics* **11**, 241-294.

Le, T. M. & Zhong, P. S. (2021). High-dimensional precision matrix estimation with a known graphical structure. *Stat*.

Friedman, J., Hastie, T. & Tibshirani, R. (2019). glasso: Graphical Lasso: Estimation of Gaussian Graphical Models. *R package version 1.11.* https://CRAN.R-project.org/package=glasso.

Robinson,P. D. & Wahten, A. J. (1992). Variational bounds on the entries of the inverse of a matrix. *IMA Journal of Numerical Analysis* **12**, 463-486.

Zaitsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Berstein's inequality conditions. *Theory and Related Fields* **74**, 535-566.

Liu, H., Lafferty, J. & Wasserman, L. (2009). The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research.*

Liu, W. & Lou, X. (2015). Fast and adaptive sparse precision matrix estimation in high dimensions. *Journal of Multivariate Analysis.*