Supplementary Material for "Precision Matrix Estimation for Multiple Compositional Vectors"

Shen Zhang[1], Huaying Fang[2,] and Tao Hu[2]

[1] *Qufu Normal University and* [2] *Capital Normal University*

S1 Proof of Proposition 1

We show that

$$\mathbf{\Omega}_c = \mathbf{\Omega}_0 - \mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\mathbf{\Omega}_0$$

satisfies the four conditions of the Moore-Penrose inverse,

$$\begin{aligned}
&(1) \quad \mathbf{\Sigma}_c\mathbf{\Omega}_c\mathbf{\Sigma}_c = \mathbf{\Sigma}_c;\\
&(2) \quad \mathbf{\Omega}_c\mathbf{\Sigma}_c\mathbf{\Omega}_c = \mathbf{\Omega}_c;\\
&(3) \quad (\mathbf{\Omega}_c\mathbf{\Sigma}_c)^\top = \mathbf{\Omega}_c\mathbf{\Sigma}_c;\\
&(4) \quad (\mathbf{\Sigma}_c\mathbf{\Omega}_c)^\top = \mathbf{\Sigma}_c\mathbf{\Omega}_c.
\end{aligned}$$

Since $\mathbf{\Sigma}_c = \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}$ with $\mathbf{G} = \mathbf{I}_p - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top$, we have

$$\mathbf{\Omega}_c\mathbf{\Sigma}_c = \mathbf{\Sigma}_c\mathbf{\Omega}_c = \mathbf{G}, \tag{S1.1}$$

which immediately verifies (3) and (4). Consequently,

$$\mathbf{\Sigma}_c\mathbf{\Omega}_c\mathbf{\Sigma}_c = \mathbf{G}\mathbf{\Sigma}_c = \mathbf{G}(\mathbf{G}\mathbf{\Sigma}_0\mathbf{G}) = \mathbf{G}\mathbf{\Sigma}_0\mathbf{G} = \mathbf{\Sigma}_c,$$

and

$$\mathbf{\Omega}_c\mathbf{\Sigma}_c\mathbf{\Omega}_c = \left\{\mathbf{\Omega}_0 - \mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\mathbf{\Omega}_0\right\}\left\{\mathbf{I}_p - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\} = \mathbf{\Omega}_c,$$

thereby establishing (1) and (2).

Property (a). This follows directly from (S1.1).

Property (b). We have

$$\begin{aligned}\mathbf{G}\mathbf{\Omega}_c &= \mathbf{\Omega}_c - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{\Omega}_0 + \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\mathbf{\Omega}_0 \\ &= \mathbf{\Omega}_c,\end{aligned}$$

and hence

$$\mathbf{\Omega}_c\mathbf{R} = \mathbf{\Omega}_c\mathbf{G}\mathbf{R} = \mathbf{\Omega}_c\{\mathbf{R} - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{R}\} = \mathbf{0}.$$

Property (c). Since $\mathbf{\Omega}_0 = \mathbf{\Sigma}_0^{-1}$ and $\mathbf{\Omega}_c$ is the Moore-Penrose inverse of $\mathbf{\Sigma}_c$, the eigenvalue inequalities

$$\lambda_{\min}\left(\mathbf{\Omega}_0\right) \le \lambda_{\min}^{+}\left(\mathbf{\Omega}_c\right) \le \lambda_{\max}\left(\mathbf{\Omega}_c\right) \le \lambda_{\max}\left(\mathbf{\Omega}_0\right)$$

are equivalent to

$$\lambda_{\max}\left(\mathbf{\Sigma}_0\right) \ge \lambda_{\max}\left(\mathbf{\Sigma}_c\right) \ge \lambda_{\min}^{+}\left(\mathbf{\Sigma}_c\right) \ge \lambda_{\min}\left(\mathbf{\Sigma}_0\right). \tag{S1.2}$$

The first inequality in (S1.2) follows from

$$\lambda_{\max}\left(\mathbf{\Sigma}_c\right) = \max_{\|\boldsymbol{s}\|_2=1}\boldsymbol{s}^\top(\mathbf{G}\mathbf{\Sigma}_0\mathbf{G})\boldsymbol{s} \le \lambda_{\max}\left(\mathbf{\Sigma}_0\right)\max_{\|\boldsymbol{s}\|_2=1}\boldsymbol{s}^\top\mathbf{G}\mathbf{G}\boldsymbol{s} = \lambda_{\max}\left(\mathbf{\Sigma}_0\right),$$

and the second inequality in (S1.2) is immediate. The third inequality in (S1.2) follows from the Courant-Fischer min-max theorem,

$$\lambda_{\min}^{+}(\boldsymbol{\Sigma}_c) = \min_{\substack{\mathcal{U}\subseteq\mathbb{R}^p \\ \dim(\mathcal{U})=d+1}} \max_{\substack{\boldsymbol{s}\in\mathcal{U} \\ \|\boldsymbol{s}\|_2=1}} \boldsymbol{s}^\top(\mathbf{G}\boldsymbol{\Sigma}_0\mathbf{G})\boldsymbol{s}$$

$$\geq \lambda_{\min}(\boldsymbol{\Sigma}_0) \min_{\substack{\mathcal{U}\subseteq\mathbb{R}^p \\ \dim(\mathcal{U})=d+1}} \max_{\substack{\boldsymbol{s}\in\mathcal{U} \\ \|\boldsymbol{s}\|_2=1}} \boldsymbol{s}^\top\mathbf{G}\boldsymbol{s} = \lambda_{\min}(\boldsymbol{\Sigma}_0).$$

Property (d). By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\min_{1\leq k\leq p}(\omega_{kk}^c\sigma_{kk}^c) &= \min_{1\leq k\leq p}\left(\boldsymbol{e}_k^\top\boldsymbol{\Omega}_c\boldsymbol{e}_k\boldsymbol{e}_k^\top\boldsymbol{\Sigma}_c\boldsymbol{e}_k\right) \\
&\geq \min_{1\leq k\leq p}\left(\boldsymbol{e}_k^\top\boldsymbol{\Sigma}_c^{1/2}\boldsymbol{\Omega}_c^{1/2}\boldsymbol{e}_k\right)^2 \\
&= \min_{1\leq k\leq p}\left(\boldsymbol{e}_k^\top\mathbf{G}\boldsymbol{e}_k\right)^2 = \left(1-\frac{1}{p_d}\right)^2.
\end{aligned}$$

S2 Proof of Proposition 2

Part (a). Let $\mathbf{tr}(\cdot)$ denote the trace operator. Then

$$\begin{aligned}
\|\boldsymbol{\Omega}_c-\boldsymbol{\Omega}_0\|_{\max} &= \left\|\boldsymbol{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\boldsymbol{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\boldsymbol{\Omega}_0\right\|_{\max} \\
&\geq \frac{\mathbf{tr}\left(\mathbf{R}^\top\boldsymbol{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\boldsymbol{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\boldsymbol{\Omega}_0\mathbf{R}\right)}{\sum_{i=1}^d p_i^2} = \frac{\mathbf{tr}\left(\mathbf{R}^\top\boldsymbol{\Omega}_0\mathbf{R}\right)}{\sum_{i=1}^d p_i^2} \\
&\geq \frac{\mathbf{tr}\left(\mathbf{R}^\top\mathbf{R}\right)\lambda_{\min}(\boldsymbol{\Omega}_0)}{\sum_{i=1}^d p_i^2} = \frac{\sum_{i=1}^d p_i}{\sum_{i=1}^d p_i^2}\lambda_{\min}(\boldsymbol{\Omega}_0) \geq \frac{p_d}{rp_1^2}.
\end{aligned}$$

These inequalities use the block-diagonal structure $\mathbf{R} = \operatorname{diag}(\mathbf{1}_{p_1},\ldots,\mathbf{1}_{p_d})$.

In particular, the first inequality follows from

$$\mathbf{tr}\left(\mathbf{R}^\top A\mathbf{R}\right) = \sum_{i=1}^d \mathbf{1}_{p_i}^\top A_{ii}\mathbf{1}_{p_i} \leq \sum_{i=1}^d \|A\|_{\max}p_i^2 = \|A\|_{\max}\sum_{i=1}^d p_i^2,$$

where A_{ii} denotes the (i,i)th block of $A = \mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\mathbf{\Omega}_0$. The second inequality follows by expanding $\mathbf{tr}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)$ and lower bounding it by $\lambda_{\min}(\mathbf{\Omega}_0)$.

Since the largest-magnitude entry of a positive semidefinite matrix lies on its diagonal, we also have

$$\begin{aligned}
\|\mathbf{\Omega}_c - \mathbf{\Omega}_0\|_{\max} &= \max_{1\le k\le p} \boldsymbol{e}_k^\top \mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\mathbf{R}^\top\mathbf{\Omega}_0\boldsymbol{e}_k \\
&\le \lambda_{\max}\left(\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\right)\max_{1\le k\le p} \boldsymbol{e}_k^\top\mathbf{\Omega}_0\mathbf{R}\mathbf{R}^\top\mathbf{\Omega}_0\boldsymbol{e}_k \\
&\le \frac{\|\mathbf{\Omega}_0\|_1^2}{\lambda_{\min}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right) &= \min_{\|\boldsymbol{s}\|_2=1}\left(\boldsymbol{s}^\top\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\boldsymbol{s}\right) \\
&\ge \lambda_{\min}(\mathbf{\Omega}_0)\min_{\|\boldsymbol{s}\|_2=1}\left(\boldsymbol{s}^\top\mathbf{R}^\top\mathbf{R}\boldsymbol{s}\right) = p_d\lambda_{\min}(\mathbf{\Omega}_0).
\end{aligned}$$

Therefore, if $\mathbf{\Omega}_0 \in \mathcal{U}_{q,r}(s_0(p), M_p)$, we have

$$\|\mathbf{\Omega}_c - \mathbf{\Omega}_0\|_{\max} \le \frac{\|\mathbf{\Omega}_0\|_1^2}{p_d\lambda_{\min}(\mathbf{\Omega}_0)} \le \frac{rM_p^2}{p_d}.$$

Part (b). Since $\|\mathbf{R}\|_2 = \sqrt{p_1}$ and $p \leq dp_1$, we obtain

$$\begin{aligned}
\|\mathbf{\Omega}_c - \mathbf{\Omega}_0\|_1 &\leq \left\|\mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\right\|_1 \left\|\mathbf{R}^\top\mathbf{\Omega}_0\right\|_1 \\
&\leq \sqrt{p}\left\|\mathbf{\Omega}_0\mathbf{R}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\right\|_2 \|\mathbf{\Omega}_0\|_1 \\
&\leq \sqrt{p}\,\|\mathbf{\Omega}_0\|_2 \|\mathbf{R}\|_2 \left\|\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)^{-1}\right\|_2 \|\mathbf{\Omega}_0\|_1 \\
&\leq \frac{\sqrt{pp_1}\lambda_{\max}\left(\mathbf{\Omega}_0\right)}{\lambda_{\min}\left(\mathbf{R}^\top\mathbf{\Omega}_0\mathbf{R}\right)} \|\mathbf{\Omega}_0\|_1 \\
&\leq \frac{\sqrt{d}p_1\lambda_{\max}\left(\mathbf{\Omega}_0\right)}{p_d\lambda_{\min}\left(\mathbf{\Omega}_0\right)} \|\mathbf{\Omega}_0\|_1 .
\end{aligned}$$

Consequently,

$$\begin{aligned}
\|\mathbf{\Omega}_c\|_1 &\leq \|\mathbf{\Omega}_c - \mathbf{\Omega}_0\|_1 + \|\mathbf{\Omega}_0\|_1 \\
&\leq \left(1 + \frac{\sqrt{d}p_1\lambda_{\max}\left(\mathbf{\Omega}_0\right)}{p_d\lambda_{\min}\left(\mathbf{\Omega}_0\right)}\right)\|\mathbf{\Omega}_0\|_1 \\
&\leq \left(1 + \frac{r^2\sqrt{d}p_1}{p_d}\right)M_p.
\end{aligned}$$

Part (c). This follows directly from Proposition 1(c). ∎

S3 Proof of Theorem 1 and theoretical properties under weak tail conditions

The tail conditions imposed in the main text can be relaxed while maintaining essentially the same theoretical guarantees. In this section, we first state analogous results under weaker tail assumptions and then prove Theorem 1 together with the corresponding error bounds under these relaxed conditions.

Condition S1.

(i) (*Weak sub-Gaussian tails*). There exist constants $\eta > 0$ and $K \geq 1$ such that

$$\max_{1 \leq j \leq p} E\left\{\exp\left(\eta Y_j^2\right)\right\} \leq K.$$

(ii) (*Weak polynomial-type tails*). There exist constants $\tau > 4$ and $K \geq 0$ such that

$$\max_{1 \leq j \leq p} E\left(|Y_j|^\tau\right) \leq K.$$

Condition S1 is a weaker version of Condition 1. Under these relaxed assumptions, one can still obtain an upper bound on the estimation error of the Amcoda estimator by choosing an appropriate tuning parameter λ.

Theorem S1. *Suppose* $\mathbf{\Omega}_0 \in \mathcal{U}_{q,r}\left(s_0(p), M_p\right)$ *and* $M_p = o\left(\sqrt{p}\right)$.

(a) If $\boldsymbol{Y}$ *satisfies Condition S1(i), then for any* $\theta > 0$, *there exist constants* $C_1 > 0$, $\eta_1 > 0$ *and* $\eta_2 > 0$ *such that, if* $\frac{\log p}{n} \leq \eta_2$ *and*

$$\lambda = \left(\frac{\sqrt{p}+1}{p_d \lambda_{\min}\left(\mathbf{\Omega}_0\right)} + 4\eta_1\sqrt{\frac{\log p}{n}}\right)\|\mathbf{\Omega}_0\|_1,$$

then the Amcoda estimator $\widehat{\mathbf{\Omega}}$ *satisfies*

$$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_{\max} \leq C_1 M_p^2\left(\sqrt{\frac{\log p}{n}} + \frac{\sqrt{p}}{p_d}\right)$$

with probability as least $1 - O\left(p^{-\theta}\right)$.

(b) If $\boldsymbol{Y}$ satisfies Condition S1(ii) and $p \leq Cn^{\tau_0}$ for constants $0 < \tau_0 < \tau/4 - 1$ and $C > 0$, then for any $\theta > 0$, there exist constants $C_1 > 0$, $\eta_1 > 0$ and $\eta_2 > 0$ such that, if $\frac{\log p}{n} \leq \eta_2$ and λ is chosen as in part (a), then

$$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max} \leq C_1 M_p^2 \left(\sqrt{\frac{\log p}{n}} + \frac{\sqrt{p}}{p_d} \right)$$

with probability as least $1 - O\left(p^{-\theta} + n^{-\tau/4+1+\tau_0} (\log p)^{\tau/4}\right)$.

To prove Theorem 1 and Theorem S1, we require several auxiliary lemmas.

Lemma S1. *If random variables U and V are sub-Gaussian, $U, V \sim \mathrm{subG}(\sigma^2)$, then there exists a constant $\eta_0 > 0$ such that $UV - E(UV)$ is sub-exponential, $UV - E(UV) \sim \mathrm{subE}(\eta_0 \sigma^2)$, i.e.,*

$$E\left[\exp\left\{s\left(UV - E(UV)\right)\right\}\right] \leq \exp\left\{\frac{s^2}{2}(\eta_0\sigma^2)^2\right\}$$

for $|s| \leq 1/(\eta_0\sigma^2)$.

Proof Since $U, V \sim \mathrm{subG}(\sigma^2)$, we have

$$\begin{aligned} \max\left\{E\left(|V|^{2k}\right), E\left(|U|^{2k}\right)\right\} &\leq 2\int_0^{+\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) d(u^{2k}) \\ &= 2\left(2\sigma^2\right)^k (k!),\ k = 1, 2, \ldots \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$E\left(|UV|^k\right) \leqslant \left\{E\left(|U|^{2k}\right) E\left(|V|^{2k}\right)\right\}^{1/2} \leq 2\left(2\sigma^2\right)^k (k!).$$

Let $\eta_0 = 2 + 2\sqrt{17}$. For $|s| \leq 1/(\eta_0 \sigma^2) < 1/(4\sigma^2)$,

$$\frac{8\sigma^2}{\sqrt{1 - 4\sigma^2|s|}} \leq \frac{8\sigma^2}{\sqrt{1 - 4/\eta_0}} = \eta_0 \sigma^2.$$

A series expansion together with standard sub-Gaussian arguments yields

$$\begin{aligned}
&E\left[\exp\left\{s\left(UV - E(UV)\right)\right\}\right] \leq 1 + \sum_{k=2}^{\infty} \frac{|s|^k}{k!} E[|UV - E(UV)|^k] \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{|s|^k}{k!} \left[\left\{E\left(|UV|^k\right)\right\}^{1/k} + E|UV|\right]^k \quad \text{(Minkowski's inequality)} \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{|s|^k}{k!} 2^k E|UV|^k \qquad \text{(Jensen's inequality)} \\
&\leq 1 + \sum_{k=2}^{\infty} \frac{|s|^k}{k!} 2^k 2(2\sigma^2)^k (k!) = 1 + \frac{s^2}{2}\left\{\frac{8\sigma^2}{(1 - 4\sigma^2|s|)^{1/2}}\right\}^2 \\
&\leq 1 + \frac{s^2}{2}(\eta_0\sigma^2)^2 \leq \exp\left\{\frac{s^2}{2}\left(\eta_0\sigma^2\right)^2\right\}.
\end{aligned}$$

■

Lemma S2. *Assume $Y_1, \ldots, Y_p$ are mean-zero sub-Gaussian random variables with $Y_j \sim \mathrm{subG}(\sigma^2)$ for $j = 1, \ldots, p$. Let $\widetilde{\boldsymbol{Z}} = (\widetilde{Z}_j) = \mathbf{G}(Y_1, \ldots, Y_p)^\top$. Then $\widetilde{Z}_j \sim subG(4\sigma^2)$ for $j = 1, \ldots, p$.*

Proof Let $\mathbf{G} = (G_{jl})$. For any $\gamma_1, \ldots, \gamma_p > 1$ such that $1/\gamma_1 + \cdots + 1/\gamma_p = 1$, Holder's inequality and the moment generating function bound for sub-Gaussian random variables yield

$$\begin{aligned}
E\left\{\exp\left(s\widetilde{Z}_j\right)\right\} &= E\left\{\exp\left(s\sum_{l=1}^{p} G_{jl}Y_l\right)\right\} \\
&\leq \inf_{\gamma_1,\ldots,\gamma_p} \prod_{l=1}^{p} \left[E\left\{\exp\left(\gamma_l s G_{jl} Y_l\right)\right\}\right]^{\frac{1}{\gamma_l}} \\
&\leq \inf_{\gamma_1,\ldots,\gamma_p} \exp\left(\frac{\sigma^2 s^2 \sum_{l=1}^{p} \gamma_l G_{jl}^2}{2}\right) \\
&= \exp\left\{\frac{s^2\sigma^2}{2}\left(\sum_{l=1}^{p} |G_{jl}|\right)^2\right\},
\end{aligned}$$

where the last equality follows from

$$\inf_{\gamma_1,\ldots,\gamma_p} \sum_{l=1}^{p} \gamma_l G_{jl}^2 = \left(\sum_{l=1}^{p} |G_{jl}|\right)^2.$$

Using $\|\mathbf{G}\|_\infty^2 = \left(2 - \frac{2}{p_1}\right)^2 \leq 4$, we obtain

$$E\left\{\exp\left(s\widetilde{Z}_j\right)\right\} \leq \exp\left(\frac{s^2\sigma^2}{2}\|\mathbf{G}\|_\infty^2\right) \leq \exp\left(\frac{s^2 4\sigma^2}{2}\right),$$

which shows that $\widetilde{Z}_j \sim \mathrm{subG}(4\sigma^2)$. ■

Lemma S3. *Assume $Y_1, \ldots, Y_p$ are mean-zero sub-Gaussian random variables with $Y_j \sim \mathrm{subG}(\sigma^2)$ for $1 \leq j \leq p$. Let $\mathbf{\Sigma}_0 = (\sigma_{kl}^0)$ be the covariance matrix of $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top$ where $\sigma_{kl}^0 = E(Y_k Y_l)$. For independent and identical distributed observations $Y_1, \ldots, Y_n$ of Y where $\boldsymbol{Y}_s = (Y_{s1}, \ldots, Y_{sp})^\top$ for $s = 1, \ldots, n$, define the sample covariance matrix $\widehat{\mathbf{\Sigma}}_0 = (\widehat{\sigma}_{kl}^0)$ by*

$$\widehat{\sigma}_{kl}^0 = \frac{1}{n}\sum_{s=1}^{n} Y_{sk}Y_{sl} - \overline{Y}_k\overline{Y}_l, \quad \overline{Y}_l = \frac{1}{n}\sum_{s=1}^{n} Y_{sl}, \ k, l = 1, \ldots, p.$$

Then, for any $\theta > 0$, there exist constants $\eta_1 > 0$ and $\eta_2 > 0$, depending only on σ^2 and θ, such that if $(\log p)/n \leq \eta_2$, then

$$P\left(\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} > \eta_1 \sqrt{\frac{\log p}{n}}\right) \leq 4p^{-\theta}.$$

Proof By Lemma S1, there exists $\eta_0 > 0$ such that $Y_k Y_l - \sigma_{kl}^0 \sim$ subE$(\eta_0 \sigma^2)$. Let

$$\delta_0 = \frac{2}{1 + \sqrt{1 + 4/(\eta_0^2 \sigma^2)}}, \quad \lambda_0 = \eta_0 \sigma^2, \quad C_0 = \frac{\lambda_0}{\delta_0}.$$

Then $\delta_0 < 1$ and $\sigma^2/(1-\delta_0) = C_0^2$. For any $t > 0$,

$$\begin{aligned}
&P\left(\left\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\right\|_{\max} > t\right) = P\left(\max_{1\leq k,l\leq p}\left|\frac{1}{n}\sum_{s=1}^n Y_{sk}Y_{sl} - \overline{Y}_k\overline{Y}_l - \sigma_{kl}^0\right| > t\right)\\
&\leq P\left(\max_{1\leq k,l\leq p}\left|\frac{1}{n}\sum_{s=1}^n Y_{sk}Y_{sl} - \sigma_{kl}^0\right| + \max_{1\leq k,l\leq p}\left|\overline{Y}_k\overline{Y}_l\right| > t\right)\\
&\leq P\left(\max_{1\leq k,l\leq p}\left|\frac{1}{n}\sum_{s=1}^n Y_{sk}Y_{sl} - \sigma_{kl}^0\right| > \delta_0 t\right) + P\left(\max_{1\leq k,l\leq p}\left|\overline{Y}_k\overline{Y}_l\right| > (1-\delta_0)t\right)\\
&\leq p^2 \max_{1\leq k,l\leq p} P\left(\left|\frac{1}{n}\sum_{s=1}^n Y_{sk}Y_{sl} - \sigma_{kl}^0\right| > \delta_0 t\right) + p\max_{1\leq k,l\leq p} P\left(\left|\overline{Y}_k\right| > \sqrt{(1-\delta_0)t}\right).
\end{aligned}$$

For the sub-exponential term, the Bernstein's inequality yields

$$P\left(\left|\frac{1}{n}\sum_{s=1}^n Y_{sk}Y_{sl} - \sigma_{kl}^0\right| > \delta_0 t\right) \leq 2\exp\left\{-\frac{n\delta_0 t}{2\lambda_0}\min\left(1, \frac{\delta_0 t}{\lambda_0}\right)\right\}.$$

For the sample mean, the Hoeffding's inequality gives

$$P\left(\left|\overline{Y}_k\right| > \sqrt{(1-\delta_0)t}\right) \leq 2\exp\left\{-\frac{n(1-\delta_0)t}{2\sigma^2}\right\}.$$

For $t \leq \min(1, C_0)$, we have $\delta_0 t/\lambda_0 \leq 1$, and hence

$$
\begin{aligned}
&P\left(\left\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\right\|_{\max} > t\right) \\
\leq & 2p^2 \exp\left\{-\frac{n\delta_0 t}{2\lambda_0} \min\left(1, \frac{\delta_0 t}{\lambda_0}\right)\right\} + 2p \exp\left(-\frac{n(1-\delta_0)t}{2\sigma^2}\right) \\
\leq & 2\left(p^2 + p\right) \exp\left\{-\frac{nt^2}{2C_0^2}\right\}.
\end{aligned}
$$

For any $\theta > 0$, let

$$
\eta_1 = C_0\sqrt{4+2\theta},\ \eta_2 = \{\min(1, C_0)/\eta_1\}^2,\ t = \eta_1\sqrt{(\log p)/n}.
$$

If $\frac{\log p}{n} \leq \eta_2$, then

$$
\begin{aligned}
P\left(\left\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) &\leq 2(p^2+p)\exp\{-(2+\theta)\log p\} \\
&\leq 4p^{-\theta}.
\end{aligned}
$$

■

Lemma S4. *Suppose $\boldsymbol{Y}$ satisfies Condition 1(i) and $\lambda_{\max}(\boldsymbol{\Omega}_0) \leq C$ for a constant $C > 0$. Then, for any $\theta > 0$, there exist constants $\eta_1 > 0$ and $\eta_2 > 0$, depending only on η, K and C, such that if $\frac{\log p}{n} \leq \eta_2$, then*

$$
P\left(\left\|\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log(p)}{n}}\right) \leq 4p^{-\theta}.
$$

Proof Under Condition 1(i), each Y_j is sub-Gaussian with $Y_j \sim \mathrm{subG}((2K+$

$1)/\eta)$. Indeed,

$$
\begin{aligned}
E\left\{\exp\left(sY_j\right)\right\} &= 1 + s^2 \int_0^1 (1-t) E\left\{Y_j^2 \exp\left(stY_j\right)\right\} dt \\
&\leq 1 + \frac{s^2}{2} E\left\{Y_j^2 \exp\left(|s|\,|Y_j|\right)\right\} \\
&\leq 1 + \frac{s^2}{2} E\left\{Y_j^2 \exp\left(\frac{s^2/\eta + \eta Y_j^2}{2}\right)\right\} \\
&\leq 1 + \frac{s^2}{\eta} \exp\left(\frac{s^2}{2\eta}\right) E\left\{\exp\left(\eta Y_j^2\right)\right\} \\
&\leq 1 + \frac{s^2 K}{\eta} \exp\left(\frac{s^2}{2\eta}\right) \qquad \text{(Condition 1(i))} \\
&\leq \left(1 + \frac{s^2 K}{\eta}\right) \exp\left(\frac{s^2}{2\eta}\right) \\
&\leq \exp\left\{\frac{s^2}{2}\left(\frac{2K+1}{\eta}\right)\right\}.
\end{aligned}
$$

Let $\widetilde{\boldsymbol{Z}} = (\mathbf{I}_p, \boldsymbol{\Omega}_0)^\top \mathbf{G} \boldsymbol{Y} = (\widetilde{Z}_j)$. By Lemma S2, $\widetilde{Z}_j \sim \text{subG}(4(2K+1)/\eta)$ for $j = 1, \ldots, p$. For $j = p+1, \ldots, 2p$, applying the same argument yields

$$
E\left\{\exp\left(s\widetilde{Z}_j\right)\right\} \leq \exp\left\{\frac{s^2}{2}\left(\frac{2K+1}{\breve{\eta}}\right)\right\}
$$

where $\breve{\eta} = \eta/(\max_{1\leq k\leq p} \|\boldsymbol{v}_k\|_2^2)$. Moreover,

$$
\max_{1\leq k\leq p} \|\boldsymbol{v}_k\|_2^2 = \max_{1\leq k\leq p} \boldsymbol{e}_k^\top \boldsymbol{\Omega}_0 \mathbf{G} \boldsymbol{\Omega}_0 \boldsymbol{e}_k \leq \lambda_{\max}^2\left(\boldsymbol{\Omega}_0\right) \leq C^2.
$$

Consequently,

$$
\widetilde{Z}_j \sim \text{subG}(\widetilde{\sigma}^2), \quad \widetilde{\sigma}^2 = \frac{2K+1}{\eta} \max\left(4, C^2\right), \quad j = 1, \ldots, 2p.
$$

Applying Lemma S3, for any $\theta > 0$, there exist constants $\widetilde{\eta}_1 > 0$ and $\widetilde{\eta}_2 > 0$

such that if $\frac{\log(2p)}{n} \le \widetilde{\eta}_2$, then

$$P\left(\left\|\widehat{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{Z}}} - \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Z}}}\right\|_{\max} > \widetilde{\eta}_1\sqrt{\frac{\log(2p)}{n}}\right) \le 4p^{-\theta}.$$

Let $\eta_1 = \sqrt{2}\widetilde{\eta}_1$ and $\eta_2 = \widetilde{\eta}_2/2$. Since $[\log(2p)]/n \le 2(\log p)/n$ and

$$\widehat{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{Z}}} - \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Z}}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c & \left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0 \\ \boldsymbol{\Omega}_0\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right) & \boldsymbol{\Omega}_0\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0, \end{pmatrix},$$

it follows that

$$\begin{aligned} &P\left(\left\|\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) \\ \le &P\left(\left\|\widehat{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{Z}}} - \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Z}}}\right\|_{\max} > \widetilde{\eta}_1\sqrt{\frac{\log(2p)}{n}}\right) \le 4p^{-\theta}. \end{aligned}$$

■

Lemma S5. *Suppose there exist constants $K > 0$ and $\tau \ge 2$ such that $E(|U_1|^\tau) \le K$ for independently and identically distributed random variables $U_1, \ldots, U_n$. Then, for any $M > 0$, if $C_2 \ge 16KM^{2-\tau}/3$ and $2KM^{1-\tau} \le t \le 3C_2M^{-1}/8$, then*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n \{U_i - E(U_i)\}\right| > t\right) \le 2\exp\left(-\frac{nt^2}{8K^{2/\tau} + C_2}\right) + KnM^{-\tau}.$$

Proof Decompose U_i as $U_i = H_i + \widetilde{H}_i$ where $H_i = U_i I(|U_i| \le M)$ and $\widetilde{H}_i = U_i I(|U_i| > M)$. Then

$$\begin{aligned} &P\left(\left|\frac{1}{n}\sum_{i=1}^n \{U_i - E(U_i)\}\right| > t\right) \\ &\le P\left(\left|\frac{1}{n}\sum_{i=1}^n \{H_i - E(H_i)\}\right| > \frac{t}{2}\right) + P\left(\left|\frac{1}{n}\sum_{i=1}^n \left\{\widetilde{H}_i - E(\widetilde{H}_i)\right\}\right| > \frac{t}{2}\right). \end{aligned}$$

For H_i, note that $|H_i - E(H_i)| \le |H_i| + |E(H_i)| \le 2M$ and

$$\mathrm{Var}\,(H_i) \le E(U_i^2) \le \{E\,(|U_i|^\tau)\}^{2/\tau} \le K^{2/\tau}.$$

If $C_2 \ge 8Mt/3$, Bernstein's inequality for bounded random variables yields

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\{H_i - E(H_i)\}\right| > \frac{t}{2}\right) &\le 2\exp\left\{-\frac{nt^2}{8\mathrm{Var}\,(H_i) + 8Mt/3}\right\} \\ &\le 2\exp\left\{-\frac{nt^2}{8K^{2/\tau} + C_2}\right\}. \end{aligned}$$

For $\widetilde{H}_i$, we have

$$E(\widetilde{H}_i) \le E\left\{|U_i|^\tau\,|U_i|^{1-\tau}\,I\,(|U_i| > M)\right\} \le KM^{1-\tau}.$$

For $t \ge 2KM^{1-\tau}$, Markov's inequality implies

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{H}_i - E\left(\widetilde{H}_i\right)\right\}\right| > \frac{t}{2}\right) &\le P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{H}_i\right| > \frac{t}{2} - \left|E\left(\widetilde{H}_1\right)\right|\right) \\ &\le P\left(\max_{1\le i\le n}|U_i| > M\right) \\ &\le n\max_{1\le i\le n}P\,(|U_i| > M) \\ &\le nM^{-\tau}E\,(|U_i|^\tau) \le KnM^{-\tau}. \end{aligned}$$

Finally, the conditions $C_2 \ge 8Mt/3$ and $t \ge 2KM^{1-\tau}$ imply $C_2 \ge 16KM^{2-\tau}/3$ and $2KM^{1-\tau} \le t \le 3C_2M^{-1}/8$. ∎

Lemma S6. *Suppose $\boldsymbol{Y}$ satisfies Condition S1(ii). Let $\widehat{\boldsymbol{\Sigma}}_0$ and $\boldsymbol{\Sigma}_0$ denote the population and sample covariance matrix of $\boldsymbol{Y}$, respectively. Then, for* *any $\theta > 0$, there exist constants $\eta_1 > 0$, $\eta_2 > 0$ and $\eta_3 > 0$, depending only on τ, K and θ, such that if $(\log p)/n \le \eta_2$, then*

$$P\left(\left\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) \le 4p^{-\theta} + \eta_3 n^{1-\tau/4} p\left(\log p\right)^{\tau/4}.$$

Proof Let

$$C_0 = \max\left(1, (2K)^{1/\tau}, K^{1/2}, \left[\max\left(4, (8K)^{-1}\right)\right]^{1/(\tau-4)}\right)$$

and

$$C_1 = \max(32K^{4/\tau} + 16K/3, 16K^{2/\tau} + 64K/3).$$

For any $t > 0$ and $M_1 \ge C_0$, we have

$$\begin{aligned}
P\left(\left\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\right\|_{\max} > t\right) &= P\left(\max_{1\le k,l\le p}\left|\frac{1}{n}\sum_{m=1}^{n} Y_{mk}Y_{ml} - \overline{Y}_k\overline{Y}_l - E\left(Y_kY_l\right)\right| > t\right)\\
&\le P\left(\max_{1\le k,l\le p}\left|\frac{1}{n}\sum_{m=1}^{n} Y_{mk}Y_{ml} - E\left(Y_kY_l\right)\right| > \frac{t}{2}\right) + P\left(\max_{1\le k,l\le p}\left|\overline{Y}_k\overline{Y}_l\right| > \frac{t}{2}\right)\\
&\le \underbrace{P\left(\max_{1\le k,l\le p}\left|\frac{1}{n}\sum_{m=1}^{n} Y_{mk}Y_{ml}I\left(|Y_{mk}Y_{ml}| > M_1^2\right) - E\left\{Y_kY_lI\left(|Y_kY_l| > M_1^2\right)\right\}\right| > \frac{t}{4}\right)}_{T_1}\\
&\quad + \underbrace{P\left(\max_{1\le k,l\le p}\left|\frac{1}{n}\sum_{m=1}^{n} Y_{mk}Y_{ml}I\left(|Y_{mk}Y_{ml}| \le M_1^2\right) - E\left\{Y_kY_lI\left(|Y_kY_l| \le M_1^2\right)\right\}\right| > \frac{t}{4}\right)}_{T_2}\\
&\quad + \underbrace{P\left(\max_{1\le k,l\le p}\left|\overline{Y}_k\overline{Y}_l\right| > \frac{t}{2}\right)}_{T_3}.
\end{aligned}$$

Since

$$E\left\{|Y_kY_l|\,I\left(|Y_kY_l|>M_1^2\right)\right\}=E\left\{|Y_kY_l|^{\tau/2}\,|Y_kY_l|^{1-\tau/2}\,I\left(|Y_kY_l|>M_1^2\right)\right\}$$

$$\leq E\,|Y_kY_l|^{\tau/2}\,M_1^{2-\tau}\leq\{E\left(|Y_k|^\tau\right)E\,|Y_k|^\tau\}^{1/2}\,M_1^{2-\tau}\leq KM_1^{2-\tau},$$

then for $t\geq 4KM_1^{2-\tau}$,

$$\begin{aligned}
T_1 &\leq P\left(\max_{1\leq k,l\leq p}\left|\frac{1}{n}\sum_{m=1}^{n}Y_{mk}Y_{ml}I\left(|Y_{mk}Y_{ml}|>M_1^2\right)\right|>\frac{t}{4}-\max_{1\leq k,l\leq p}E\left\{|Y_kY_l|\,I\left(|Y_kY_l|>M_1^2\right)\right\}\right)\\
&\leq P\left(\max_{1\leq k,l\leq p}\left|\frac{1}{n}\sum_{m=1}^{n}Y_{mk}Y_{ml}I\left(|Y_{mk}Y_{ml}|>M_1^2\right)\right|>\frac{t}{4}-KM_1^{2-\tau}\right)\\
&\leq P\left(\max_{\substack{1\leq m\leq n\\1\leq k,l\leq p}}|Y_{mk}Y_{ml}|>M_1^2\right)\leq np\max_{\substack{1\leq m\leq n\\1\leq k,l\leq p}}P\left(|Y_{mk}|>M_1\right)\\
&\leq np\max_{\substack{1\leq m\leq n\\1\leq k,l\leq p}}M_1^{-\tau}E\,|Y_{mk}|^\tau\leq npKM_1^{-\tau}.
\end{aligned}$$

Let $H_m=Y_{mk}Y_{ml}I(|Y_{mk}Y_{ml}|\leq M_1^2)-E\{Y_kY_lI(|Y_kY_l|\leq M_1^2)\}$ for $m=1,\ldots,n$, then $|H_m|\leq 2M_1^2$ and

$$\begin{aligned}
\mathrm{Var}\left(H_m\right)&\leq E\left\{\left(Y_{mk}Y_{ml}\right)^2\right\}\leq\left\{E\left(|Y_{mk}|^{\tau/2}\,|Y_{ml}|^{\tau/2}\right)\right\}^{4/\tau}\\
&\leq\{E\left(|Y_{mk}|^\tau\right)E\left(|Y_{ml}|^\tau\right)\}^{2/\tau}\leq K^{4/\tau}.
\end{aligned}$$

Therefore, for $t\leq KM_1^{-2}$, Bernstein's inequality of bounded random variables yields

$$\begin{aligned}
T_2 &\leq p^2\max_{1\leq k,l\leq p}P\left(\left|\frac{1}{n}\sum_{m=1}^{n}H_m\right|>\frac{t}{4}\right)\leq 2p^2\exp\left(-\frac{nt^2}{32\mathrm{Var}\left(H_m\right)+16M_1^2t/3}\right)\\
&\leq 2p^2\exp\left(-\frac{nt^2}{32K^{4/\tau}+16K/3}\right)\leq 2p^2\exp\left(-\frac{nt^2}{C_1}\right).
\end{aligned}$$

For $8K^2M_1^{2-2\tau} \le t \le \min(1, 32K^2M_1^{2-2\tau})$, Lemma S5 yields

$$
\begin{aligned}
T_3 &\le p \max_{1\le k\le p} P\left(\left|\overline{Y}_k\right| > \left(\frac{t}{2}\right)^{1/2}\right) \\
&\le 2p \exp\left(-\frac{nt}{16K^{2/\tau} + 64K/3}\right) + npKM_1^{-\tau} \\
&\le 2p \exp\left(-\frac{nt^2}{16K^{2/\tau} + 64K/3}\right) + npKM_1^{-\tau} \\
&\le 2p \exp\left(-\frac{nt^2}{C_1}\right) + npKM_1^{-\tau}.
\end{aligned}
$$

Combining bounds for $T_1,\ T_2,\ T_3$ and

$$
\max(4KM_1^{2-\tau}, 8K^2M_1^{2-2\tau}) = 4KM_1^{2-\tau},
$$

$$
\min(KM_1^{-2}, 1, 32K^2M_1^{2-2\tau}) = KM_1^{-2}\min(1, 32K),
$$

for $4KM_1^{2-\tau} \le t \le KM_1^{-2}\min(1, 32K)$, we have

$$
P\left(\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\right\|_{\max} > t\right) \le 2(p^2 + p)\exp\left(-\frac{nt^2}{C_1}\right) + 2npKM_1^{-\tau}.
$$

For $\theta > 0$, let

$$
\eta_1 = \{C_1(2+\theta)\}^{1/2},\ \eta_3 = 2(K\eta_1)^{1/2}/\{\min(1, 32K)\}^{1/2},\ \eta_2 = 2K/(c_0\eta_3)^4
$$

and

$$
M_1 = (2K/\eta_3)(n/\log p)^{1/4},\ t = \eta_1\{(\log p)/n\}^{1/2}.
$$

Then, for $(\log p)/n \le \eta_2$, we have

$$
\begin{aligned}
&P\left(\left\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) \\
\le& 2\left(p^2+p\right)p^{-(2+\theta)} + 2npKn^{-\tau/4}\left(\log p\right)^{\tau/4}\eta_3/(2K) \\
\le& 4p^{-\theta} + \eta_3 n^{1-\tau/4} p\left(\log p\right)^{\tau/4}.
\end{aligned}
$$

■

Lemma S7. *Suppose* $\boldsymbol{Y}$ *satisfies Condition 1(ii),* $\lambda_{\max}(\boldsymbol{\Omega}_0) \le C_0$ *and* $p \le C_1 n^{\tau_0}$ *for* $0 < \tau_0 < \tau/4 - 1$ *and constants* $C_0 > 0$ *and* $C_1 > 0$*. Then, for any* $\theta > 0$*, there exist constants* $\eta_1 > 0$*,* $\eta_2 > 0$ *and* $\eta_3 > 0$*, depending only on* $\tau,\ K,\ \theta,$ C_0 *and* C_1*, such that if* $(\log p)/n \le \eta_2$*, then*

$$
P\left(\left\|\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) \le 4p^{-\theta} + \eta_3 n^{1-\tau/4+\tau_0}\left(\log p\right)^{\tau/4}.
$$

Proof Define $\widetilde{\boldsymbol{Z}} = (\mathbf{I}_p, \boldsymbol{\Omega}_0)^\top \mathbf{G}\boldsymbol{Y} = (\widetilde{Z}_j)$ and write $\mathbf{G} = (G_{jl})$. Since

$$
\max_{1\le l\le p} E(|Y_l|^\tau) \le K, \qquad \sum_{l=1}^{p}|G_{jl}| \le 2(1-1/p_1) \le 2,
$$

Minkowski's inequality yields

$$
\begin{aligned}
\max_{1\le j\le p} E\left(\left|\widetilde{Z}_j\right|^\tau\right) &= \max_{1\le j\le p} E\left(\left|\sum_{l=1}^{p} G_{jl}Y_l\right|^\tau\right) \\
&\le \max_{1\le j\le p}\left[\sum_{l=1}^{p}\left\{E\left(|G_{jl}Y_l|^\tau\right)\right\}^{1/\tau}\right]^\tau \le 2^\tau K.
\end{aligned}
$$

Similarly, for $j = 1, \dots, 2p$,

$$\begin{aligned}\max_{1\le j\le 2p} E\left(\left|\widetilde{Z}_j\right|^\tau\right) &\le \max\left\{2^\tau K, \max_{1\le j\le p}\left\{\|\boldsymbol{v}_j\|_2^\tau E\left(\left|\frac{\boldsymbol{v}_j^\top \boldsymbol{Y}}{\|\boldsymbol{v}_j\|_2}\right|^\tau\right)\right\}\right\}\\ &\le K\max\left(2^\tau, \max_{1\le j\le p}\|\boldsymbol{v}_j\|_2^\tau\right)\\ &\le K\max\left(2^\tau, \lambda_{\max}^\tau(\boldsymbol{\Omega}_0)\right) \le K\{\max(2, C_0)\}^\tau.\end{aligned}$$

Applying Lemma S6, for any $\theta > 0$, there exist constants $\widetilde{\eta}_1 > 0$, $\eta_2 > 0$ and $\widetilde{\eta}_3 > 0$, such that if $[\log(2p)/n] \le \eta_2$, then

$$\begin{aligned}P\left(\left\|\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0\right\|_{\max} > \widetilde{\eta}_1\sqrt{\frac{\log(2p)}{n}}\right) &\le P\left(\left\|\widehat{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{Z}}} - \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Z}}}\right\|_{\max} > \widetilde{\eta}_1\sqrt{\frac{\log(2p)}{n}}\right)\\ &\le 4p^{-\theta} + 2\widetilde{\eta}_3 n^{1-\tau/4} p\left\{\log(2p)\right\}^{\tau/4}.\end{aligned}$$

Setting $\eta_1 = 2^{1/2}\widetilde{\eta}_1$ and $\eta_3 = 2C_1\widetilde{\eta}_3 2^{\tau/4}$ yields

$$P\left(\left\|\left(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c\right)\boldsymbol{\Omega}_0\right\|_{\max} > \eta_1\sqrt{\frac{\log p}{n}}\right) \le 4p^{-\theta} + \eta_3 n^{1-\tau/4+\tau_0}(\log p)^{\tau/4}.$$

■

Lemma S8. *For the Amcoda estimator $\widehat{\boldsymbol{\Omega}}$, if*

$$\lambda \ge \|\boldsymbol{\Omega}_0\|_1 \frac{p^{1/2}+1}{p_d \lambda_{\min}(\boldsymbol{\Omega}_0)} + \|(\widehat{\boldsymbol{\Sigma}}_c - \boldsymbol{\Sigma}_c)\boldsymbol{\Omega}_0\|_{\max},$$

or

$$\lambda \ge \left(\frac{p^{1/2}+1}{p_d \lambda_{\min}(\boldsymbol{\Omega}_0)} + 4\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\|_{\max}\right)\|\boldsymbol{\Omega}_0\|_1,$$

then

$$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max} \le 2\|\boldsymbol{\Omega}_0\|_1\left(\lambda + \frac{1}{p_d}\right).$$

Proof First, note that

$$
\begin{aligned}
&\|\mathbf{\Sigma}_c(\mathbf{\Omega}_c - \mathbf{\Omega}_0)\|_{\max} = \left\|(\mathbf{I}_p - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top)\mathbf{\Sigma}_0\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{\Omega}_0\right\|_{\max} \\
&\leq \left\|(\mathbf{I}_p - \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top)\mathbf{\Sigma}_0\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\|\mathbf{\Omega}_0\|_1 \\
&\leq \left(\left\|\mathbf{\Sigma}_0\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max} + \left\|\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{\Sigma}_0\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\right)\|\mathbf{\Omega}_0\|_1 \\
&\leq \left(\left\|\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\|\mathbf{\Sigma}_0\|_1 + \lambda_{\max}(\mathbf{\Sigma}_0)\left\|\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\right)\|\mathbf{\Omega}_0\|_1 \\
&\leq \frac{1}{p_d}\left(p^{1/2}\lambda_{\max}(\mathbf{\Sigma}_0) + \lambda_{\max}(\mathbf{\Sigma}_0)\right)\|\mathbf{\Omega}_0\|_1 = \frac{p^{1/2}+1}{p_d\lambda_{\min}(\mathbf{\Omega}_0)}\|\mathbf{\Omega}_0\|_1 .
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\left\|(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c)\mathbf{\Omega}_0\right\|_{\max} &\leq \left\|\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c\right\|_{\max}\|\mathbf{\Omega}_0\|_1 \leq \left\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\right\|_{\max}\|\mathbf{G}\|_1^2\|\mathbf{\Omega}_0\|_1 \\
&\leq 4\left\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\right\|_{\max}\|\mathbf{\Omega}_0\|_1.
\end{aligned}
$$

where basic inequalities of matrix norms are used.

Hence,

$$
\begin{aligned}
&\left\|\widehat{\mathbf{\Sigma}}_c\mathbf{\Omega}_0 - \mathbf{G}\right\|_{\max} = \left\|(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c)\mathbf{\Omega}_0 - \mathbf{\Sigma}_c(\mathbf{\Omega}_c - \mathbf{\Omega}_0)\right\|_{\max} \\
&\leq \|\mathbf{\Sigma}_c(\mathbf{\Omega}_c - \mathbf{\Omega}_0)\|_{\max} + \left\|(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c)\mathbf{\Omega}_0\right\|_{\max}, \\
&\leq \frac{p^{1/2}+1}{p_d\lambda_{\min}(\mathbf{\Omega}_0)}\|\mathbf{\Omega}_0\|_1 + \left\|\left(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c\right)\mathbf{\Omega}_0\right\|_{\max} \\
&\leq \frac{p^{1/2}+1}{p_d\lambda_{\min}(\mathbf{\Omega}_0)}\|\mathbf{\Omega}_0\|_1 + 4\left\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\right\|_{\max}\|\mathbf{\Omega}_0\|_1.
\end{aligned}
$$

Under either condition on λ, $\mathbf{\Omega}_0$ is feasible for the Amcoda constraints and therefore $\|\widehat{\mathbf{\Omega}}\|_1 \le \|\mathbf{\Omega}_0\|_1$. Consequently,

$$
\begin{aligned}
&\left\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\right\|_{\max} \\
&= \left\|\left(\mathbf{G} - \widehat{\mathbf{\Sigma}}_c\mathbf{\Omega}_0\right)\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\left(\mathbf{G} - \widehat{\mathbf{\Sigma}}_c\mathbf{\Omega}_0\right) + \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max} \\
&\le \left\|\widehat{\mathbf{\Omega}}\right\|_1\left(\left\|\widehat{\mathbf{\Sigma}}_c\mathbf{\Omega}_0 - \mathbf{G}\right\|_{\max} + \left\|\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\right) \\
&\quad + \|\mathbf{\Omega}_0\|_1\left(\left\|\widehat{\mathbf{\Sigma}}_c\mathbf{\Omega}_0 - \mathbf{G}\right\|_{\max} + \left\|\mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\right\|_{\max}\right) \\
&\le \left(\lambda + \frac{1}{p_d}\right)\left\|\widehat{\mathbf{\Omega}}\right\|_1 + \left(\lambda + \frac{1}{p_d}\right)\|\mathbf{\Omega}_0\|_1 \le 2\|\mathbf{\Omega}_0\|_1\left(\lambda + \frac{1}{p_d}\right).
\end{aligned}
$$

■

Proof of Theorem 1. If $\boldsymbol{Y}$ satisfies Condition 1(i) (respectively, Condition 1(ii)), then Lemma S4 (respectively, Lemma S7) implies that there exists a constant $\eta_1 > 0$ such that

$$
\left\|\left(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c\right)\mathbf{\Omega}_0\right\|_{\max} \le \eta_1\sqrt{\frac{\log p}{n}}
$$

with probability at least $1 - 4p^{-\theta}$ (respectively, at least $1 - \{4p^{-\theta} + \eta_3 n^{1-\tau/4+\tau_0}(\log p)^{\tau/4}\}$ for some constant $\eta_3 > 0$). Define

$$
\lambda = \frac{p^{1/2}+1}{p_d\lambda_{\min}(\mathbf{\Omega}_0)}\|\mathbf{\Omega}_0\|_1 + \eta_1\sqrt{\frac{\log p}{n}}.
$$

On the event $\|(\widehat{\mathbf{\Sigma}}_c - \mathbf{\Sigma}_c)\mathbf{\Omega}_0\|_{\max} \le \eta_1\sqrt{(\log p)/n}$, Lemma S8 implies

$$
\left\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\right\|_{\max} \le 2\|\mathbf{\Omega}_0\|_1\left(\lambda + \frac{1}{p_d}\right) = 2\|\mathbf{\Omega}_0\|_1\left(\frac{\sqrt{p}+1}{p_d\lambda_{\min}(\mathbf{\Omega}_0)}\|\mathbf{\Omega}_0\|_1 + \eta_1\sqrt{\frac{\log p}{n}} + \frac{1}{p_d}\right).
$$

Since $\mathbf{\Omega}_0 \in \mathcal{U}_{q,r}(s_0(p), M_p)$, we further obtain

$$
\left\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\right\|_{\max} \le 2M_p\left\{\eta_1\sqrt{\frac{\log p}{n}} + \frac{3rM_p\sqrt{p}}{p_d}\right\} \le \max(2\eta_1, 6r)M_p\left\{\sqrt{\frac{\log p}{n}} + M_p\frac{\sqrt{p}}{p_d}\right\}.
$$

■

Proof of Theorem S1. The argument is analogous to the proof of Theorem 1, with the only difference being the choice of λ in Lemma S8. We omit the details for brevity. ■

S4 Proof of Lemma 1

Let $\boldsymbol{x} \odot \boldsymbol{y}$ denote the Hadamard product of the column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. For any $\widetilde{C} > 0$, note that

$$\begin{aligned}
&\left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right) - a\right\|_1 - \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right)\right\|_1 \\
\geq\ & \|\boldsymbol{a}\|_1 - \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right)\right\|_1 - \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right)\right\|_1 = \|\boldsymbol{a}\|_1 - \|\boldsymbol{b}\|_1 \geq 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{b} - \boldsymbol{a}\|_1 &= \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right) + \boldsymbol{b} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right) - \boldsymbol{a}\right\|_1 \\
&\leq \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right) - a\right\|_1 + \left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right)\right\|_1 \\
&\leq 2\left\|\boldsymbol{b} \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right) - \boldsymbol{a}\right\|_1 \\
&= 2\left\|(\boldsymbol{b} - \boldsymbol{a}) \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right) - \boldsymbol{a} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right)\right\|_1 \\
&\leq 2\left(\left\|\boldsymbol{a} \odot I\left(|\boldsymbol{a}| \leq \widetilde{C}\right)\right\|_1 + \left\|(\boldsymbol{b} - \boldsymbol{a}) \odot I\left(|\boldsymbol{a}| > \widetilde{C}\right)\right\|_1\right).
\end{aligned}$$

Let $\widetilde{C} = C\|\boldsymbol{b}-\boldsymbol{a}\|_\infty$. Then

$$\left\|\boldsymbol{a}\odot I\left(|\boldsymbol{a}|\le \widetilde{C}\right)\right\|_1 \le \left\|\boldsymbol{a}\odot I\left(|\boldsymbol{a}|\le \widetilde{C}\right)\right\|_\infty^{1-\alpha}\left\|a\odot I\left(|\boldsymbol{a}|\le \widetilde{C}\right)\right\|_\alpha^{\alpha}$$
$$\le C^{1-\alpha}\|\boldsymbol{a}\|_\alpha^\alpha\|\boldsymbol{b}-\boldsymbol{a}\|_\infty^{1-\alpha},$$

and

$$\left\|(\boldsymbol{b}-\boldsymbol{a})\odot I\left(|\boldsymbol{a}|> \widetilde{C}\right)\right\|_1 \le \|\boldsymbol{b}-\boldsymbol{a}\|_\infty^{1-\alpha}\left\|(\boldsymbol{b}-\boldsymbol{a})\odot I\left(|\boldsymbol{a}|> \widetilde{C}\right)\right\|_\alpha^{\alpha}$$
$$\le C^{-\alpha}\|\boldsymbol{a}\|_\alpha^\alpha\|\boldsymbol{b}-\boldsymbol{a}\|_\infty^{1-\alpha}.$$

Combining these bounds yields

$$\|\boldsymbol{b}-\boldsymbol{a}\|_1 \le 2\left(C^{-\alpha}+C^{1-\alpha}\right)\|\boldsymbol{a}\|_\alpha^\alpha\|\boldsymbol{b}-\boldsymbol{a}\|_\infty^{1-\alpha}.$$

Taking $C=1$ gives the desired result,

$$\|\boldsymbol{b}-\boldsymbol{a}\|_1 \le 4\|\boldsymbol{a}\|_\alpha^\alpha\|\boldsymbol{b}-\boldsymbol{a}\|_\infty^{1-\alpha}.$$

■

S5 Proof of Theorem 2

By Lemma 1, for $0 \leqslant q < 1$,

$$\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_1 = \max_{1\le l\le p}\|\widehat{\boldsymbol{\omega}}_l-\boldsymbol{\omega}_l^0\|$$
$$\le 4\max_{1\le l\le p}\left(\|\boldsymbol{\omega}_l^0\|_q^q\|\widehat{\boldsymbol{\omega}}_l-\boldsymbol{\omega}_l^0\|_\infty^{1-q}\right) \le 4s_0(p)\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_{\max}^{1-q},$$

where $\widehat{\boldsymbol{\omega}}_l$ and $\boldsymbol{\omega}_l^0$ denote the lth columns of $\widehat{\boldsymbol{\Omega}}$ and $\boldsymbol{\Omega}_0$, respectively. By Theorem 1, and letting $C_2 = 4C_1^{1-q}$ and $C_3 = 4C_1^{2-q}$, we have

$$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2 \leq \sqrt{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_1 \leq 4s_0(p)\sqrt{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max}^{1-q},$$

and

$$\frac{1}{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\mathrm{F}}^2 \leq \|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_1\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max} \leq 4s_0(p)\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max}^{2-q}.$$

Therefore,

$$\begin{aligned}
\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_1 &\leq C_2 s_0(p) M_p^{1-q}\left(\sqrt{\frac{\log p}{n}} + M_p\frac{\sqrt{p}}{p_d}\right)^{1-q}, \\
\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2 &\leq C_2 s_0(p)\sqrt{p} M_p^{1-q}\left((\sqrt{\frac{\log p}{n}} + M_p\frac{\sqrt{p}}{p_d}\right)^{1-q}, \\
\frac{1}{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\mathrm{F}}^2 &\leq C_3 s_0(p) M_p^{2-q}\left(\sqrt{\frac{\log p}{n}} + M_p\frac{\sqrt{p}}{p_d}\right)^{2-q}.
\end{aligned}$$

Finally, since $M_p = o\big(p_d[(\log p)/(np)]^{1/2}\big)$, the rates simplify to

$$\begin{aligned}
\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\max} &= O\left(M_p\left(\frac{\log p}{n}\right)^{1/2}\right), \\
\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_1 &= O\left(s_0(p) M_p^{1-q}\left(\frac{\log p}{n}\right)^{(1-q)/2}\right), \\
\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2 &= O\left(s_0(p) p^{1/2} M_p^{1-q}\left(\frac{\log p}{n}\right)^{(1-q)/2}\right), \\
\frac{1}{p}\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_{\mathrm{F}}^2 &= O\left(s_0(p) M_p^{2-q}\left(\frac{\log p}{n}\right)^{(2-q)/2}\right).
\end{aligned}$$

■

S6 Proof of Corollary 1

By Theorem 1, $\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_{\max} \le \tau_{np}$ with probability at least $1 - O(p^{-\theta})$ (or $1 - O(p^{-\theta} + n^{-\tau/4+1+\tau_0}(\log p)^{\tau/4})$ under Condition 1(ii)). For $k, l = 1, \ldots, p$, if $\omega_{kl}^0 = 0$, then

$$|\widehat{\omega}_{kl}^{\tau}| \le \left\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\right\|_{\max} \le \tau_{np},$$

and hence $\widehat{\omega}_{kl}^{\tau} = 0$ by the construction of $\widehat{\mathbf{\Omega}}^{\tau}$.

If $\omega_{kl}^0 \ne 0$, then

$$\omega^* - |\widehat{\omega}_{kl}^{\tau}| \le \left|\omega_{kl}^0\right| - |\widehat{\omega}_{kl}^{\tau}| \le \left|\widehat{\omega}_{kl}^{\tau} - \omega_{kl}^0\right| \le \left\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\right\|_{\max} \le \tau_{np} < \omega^*/2.$$

Therefore, $|\widehat{\omega}_{kl}^{\tau}| > \omega^*/2 > \tau_{np}$ and $\widehat{\omega}_{kl}^{\tau}$ has the same sign as ω_{kl}^0. ■

S7 Choice of tuning parameters in Amcoda

In practice, we allow tuning parameters to be column-specific. Specifically, we choose λ_j adaptively for the jth column $(1 \le j \le p)$. The tuning parameter λ_j is selected as follows.

First, we define the loss function,

$$L(\boldsymbol{\omega}_j, \mathbf{\Sigma}_c) = \frac{1}{2}\boldsymbol{\omega}_j^{\top}\mathbf{\Sigma}_c\,\boldsymbol{\omega}_j - \boldsymbol{g}_j^{\top}\boldsymbol{\omega}_j,$$

where $\boldsymbol{\omega}_j$ and $\boldsymbol{g}_j$ denote the jth columns of $\mathbf{\Omega}$ and $\boldsymbol{G}$, respectively. The gradient of $L(\boldsymbol{\omega}_j, \mathbf{\Sigma}_c)$ with respect to $\boldsymbol{\omega}_j$ is $\mathbf{\Sigma}_c\boldsymbol{\omega}_j - \boldsymbol{g}_j$, which matches the constraint in the optimization problem (2.4).

Next, we randomly split the observations into a training set of size n_1 and a test set of size $n_2 = n - n_1$, and repeat this sample-splitting procedure K times. For the kth split, let $\tilde{\boldsymbol{\omega}}_j^{(1k)}(\lambda_j)$ denote the estimator computed from the training set with the tuning parameter λ_j, and let $\hat{\boldsymbol{\Sigma}}_c^{(2k)}$ denote the sample covariance matrix computed from the test set. We then select λ_j by minimizing the average test loss,

$$\lambda_j^* = \arg\min_{\lambda_j} \frac{1}{K} \sum_{k=1}^{K} L\left(\tilde{\boldsymbol{\omega}}_j^{(1k)}(\lambda_j),\, \hat{\boldsymbol{\Sigma}}_c^{(2k)}\right).$$

Table S1: Running time (seconds) of Amcoda (parallel implementation using 20 threads) and competing methods (gmcoda, glasso and SPIEC-EASI) for the band network with $p = 100,\ 200,\ 300,\ 400,\ 500$. A dash ("–") indicates that the running time exceeded 30 minutes and the job was terminated. Results are averaged over 20 simulation replicates.

Method	$p = 100$	$p = 200$	$p = 300$	$p = 400$	$p = 500$
Amcoda	8	36	83	146	199
gmcoda	1	6	8	20	44
glasso	98	903	–	–	–
SPIEC-EASI	39	184	577	1360	–

Table S2: Performance comparison of Amcoda and competing methods (gmcoda, glasso and SPIEC-EASI) for $p = 300$ and $d = 3, 4, 5$. A dash ("–") indicates that the running time exceeded 30 minutes and the job was terminated. Results are averaged over 20 simulation replicates.

d	Method	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_F$	FPR	Recall	Precision	F1	Time (s)
Band									
3	Amcoda	3.75	3.25	23.43	0.01	0.67	0.62	0.64	317
	gmcoda	3.96	3.56	27.42	0.03	0.66	0.25	0.36	9
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.96	3.30	24.05	0.01	0.82	0.49	0.61	602
4	Amcoda	3.78	3.27	23.59	0.01	0.65	0.60	0.62	338
	gmcoda	3.97	3.57	27.48	0.03	0.65	0.25	0.36	11
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.98	3.30	24.01	0.01	0.81	0.49	0.61	633
5	Amcoda	3.77	3.29	23.76	0.01	0.64	0.59	0.61	343
	gmcoda	3.99	3.57	27.48	0.03	0.65	0.24	0.35	12
	glasso	4.84	4.19	26.33	0.05	0.19	0.05	0.08	1448
	SPIEC-EASI	4.00	3.30	23.99	0.01	0.81	0.49	0.61	590
Hub									
3	Amcoda	4.99	2.28	13.60	0.01	0.86	0.51	0.64	285
	gmcoda	6.71	3.26	23.05	0.01	0.59	0.41	0.48	7
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.30	3.45	28.75	0.01	0.91	0.33	0.49	765
4	Amcoda	4.83	2.26	13.63	0.01	0.85	0.50	0.63	328
	gmcoda	6.71	3.26	23.03	0.01	0.58	0.41	0.48	7
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.35	3.47	28.78	0.01	0.90	0.34	0.49	743
5	Amcoda	4.92	2.30	13.79	0.01	0.85	0.50	0.63	306
	gmcoda	6.58	3.21	22.65	0.01	0.61	0.41	0.49	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.58	3.54	29.46	0.01	0.85	0.37	0.52	817
Block									
3	Amcoda	6.44	3.63	20.80	0.01	0.73	0.65	0.69	312
	gmcoda	7.66	4.47	29.48	0.02	0.63	0.35	0.45	9
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.43	4.67	36.89	0.02	0.81	0.34	0.48	851
4	Amcoda	6.58	3.68	21.16	0.01	0.71	0.64	0.68	353
	gmcoda	7.44	4.40	28.84	0.02	0.66	0.33	0.44	18
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.79	4.86	37.90	0.02	0.73	0.31	0.43	948
5	Amcoda	6.79	3.75	21.20	0.01	0.71	0.63	0.67	348

Continued on next page

Table S2 (continued)

d	Method	$\|\widehat{\mathbf{\Omega}}-\mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}}-\mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}}-\mathbf{\Omega}_0\|_F$	FPR	Recall	Precision	F1	Time (s)
	gmcoda	7.73	4.48	29.31	0.02	0.64	0.33	0.44	12
	glasso	8.99	5.74	37.21	0.03	0.52	0.17	0.25	1298
	SPIEC-EASI	7.40	4.66	36.78	0.02	0.82	0.34	0.48	869
Random									
3	Amcoda	6.46	3.43	21.78	0.01	0.73	0.63	0.68	351
	gmcoda	8.27	3.98	27.42	0.02	0.76	0.32	0.45	26
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–
4	Amcoda	6.35	3.47	21.94	0.01	0.72	0.64	0.68	368
	gmcoda	8.16	3.96	27.21	0.02	0.78	0.31	0.45	37
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–
5	Amcoda	6.26	3.46	22.02	0.01	0.72	0.63	0.67	330
	gmcoda	8.35	4.02	27.60	0.02	0.75	0.32	0.44	28
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–

Table S3: Performance comparison of Amcoda and gmcoda for $p = 1000$ and $d = 3$. Amcoda is implemented in parallel using 20 threads. Results are averaged over 20 simulation replicates.

Network	Method	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_\mathrm{F}$	FPR	Recall	Precision	F1	Time (s)
Band	Amcoda	3.93	3.53	49.45	0.00	0.34	0.67	0.45	975
	gmcoda	4.16	3.68	52.85	0.01	0.47	0.21	0.29	113
Hub	Amcoda	7.69	3.04	30.05	0.00	0.71	0.65	0.68	981
	gmcoda	8.46	3.41	43.45	0.00	0.45	0.36	0.40	45
Block	Amcoda	9.06	4.23	46.08	0.00	0.46	0.60	0.52	981
	gmcoda	16.34	5.27	62.68	0.00	0.11	0.25	0.15	39
Random	Amcoda	7.84	3.95	46.28	0.00	0.52	0.66	0.58	977
	gmcoda	22.18	5.05	63.79	0.00	0.17	0.29	0.21	62

Table S4: Performance comparison of Amcoda and competing methods (gmcoda, glasso and SPIEC-EASI) for $n = 200$, $p = 300$ and $p_1 : p_2 = 2, 4, 8, 10$. A dash ("–") indicates that the running time exceeded 30 minutes and the job was terminated. Results are averaged over 20 simulation replicates.

$p_1 : p_2$	Method	$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_1$	$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_2$	$\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0\|_F$	FPR	Recall	Precision	F1	Time (s)
Band									
2	Amcoda	3.73	3.23	23.19	0.01	0.69	0.63	0.66	342
	gmcoda	3.95	3.55	27.28	0.03	0.68	0.25	0.36	11
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.99	3.30	24.07	0.01	0.81	0.48	0.61	634
4	Amcoda	3.74	3.25	23.16	0.01	0.69	0.63	0.66	278
	gmcoda	3.95	3.56	27.29	0.03	0.68	0.25	0.37	9
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.96	3.31	24.06	0.01	0.81	0.48	0.61	660
8	Amcoda	3.72	3.27	23.11	0.01	0.69	0.63	0.66	270
	gmcoda	3.96	3.56	27.36	0.03	0.67	0.25	0.36	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.95	3.31	24.05	0.01	0.81	0.49	0.61	642
10	Amcoda	3.75	3.28	23.13	0.01	0.70	0.63	0.66	262
	gmcoda	3.97	3.55	27.29	0.03	0.68	0.25	0.36	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.96	3.32	24.05	0.01	0.81	0.48	0.61	614
Hub									
2	Amcoda	4.78	2.23	13.52	0.01	0.86	0.51	0.64	270
	gmcoda	6.69	3.24	22.97	0.01	0.59	0.42	0.49	5
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.24	3.44	28.73	0.01	0.91	0.37	0.53	749
4	Amcoda	4.87	2.25	13.62	0.01	0.85	0.50	0.63	249
	gmcoda	6.72	3.26	23.03	0.01	0.59	0.42	0.49	5
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.36	3.45	28.74	0.01	0.90	0.34	0.49	748
8	Amcoda	4.99	2.26	13.63	0.01	0.85	0.51	0.64	325
	gmcoda	6.75	3.27	23.21	0.01	0.57	0.41	0.48	7
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.29	3.43	28.76	0.01	0.89	0.31	0.46	799
10	Amcoda	4.85	2.24	13.53	0.01	0.85	0.51	0.64	268
	gmcoda	6.70	3.25	23.01	0.01	0.59	0.41	0.48	6
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.24	3.45	28.73	0.01	0.90	0.34	0.49	781
Block									
2	Amcoda	6.52	3.66	20.61	0.01	0.74	0.66	0.70	261

Continued on next page

Table S4 (continued)

$p_1 : p_2$	Method	$\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_1$	$\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_2$	$\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0\|_\mathrm{F}$	FPR	Recall	Precision	F1	Time (s)
	gmcoda	7.53	4.50	29.76	0.01	0.63	0.36	0.45	6
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.45	4.66	36.83	0.02	0.81	0.34	0.48	912
4	Amcoda	6.67	3.71	20.58	0.01	0.74	0.65	0.69	297
	gmcoda	7.39	4.46	28.89	0.02	0.66	0.35	0.45	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.42	4.66	36.95	0.02	0.81	0.35	0.49	910
8	Amcoda	6.27	3.72	20.75	0.01	0.73	0.65	0.69	270
	gmcoda	7.47	4.53	29.69	0.01	0.63	0.36	0.46	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.40	4.74	36.95	0.02	0.80	0.34	0.47	905
10	Amcoda	6.39	3.74	20.73	0.00	0.73	0.66	0.69	275
	gmcoda	7.42	4.46	29.07	0.02	0.65	0.35	0.45	11
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.42	4.75	37.06	0.02	0.79	0.34	0.47	896
Random									
2	Amcoda	6.26	3.43	21.78	0.01	0.73	0.64	0.68	326
	gmcoda	8.26	3.97	27.35	0.02	0.77	0.32	0.45	21
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–
4	Amcoda	6.31	3.42	21.74	0.01	0.72	0.64	0.68	302
	gmcoda	8.33	3.96	27.31	0.02	0.77	0.32	0.45	15
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–
8	Amcoda	6.17	3.43	21.74	0.01	0.73	0.64	0.68	290
	gmcoda	8.24	3.94	27.10	0.02	0.78	0.32	0.45	15
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–
10	Amcoda	6.37	3.44	21.76	0.01	0.73	0.64	0.68	337
	gmcoda	8.37	3.99	27.52	0.02	0.76	0.32	0.45	14
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–

Table S5: Performance comparison of Amcoda and CARE in simulations with $p = 100,\ 300$. Results are averaged over 20 simulation replicates.

$p_1 = p_2$	Method	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_{\mathrm{F}}$	FPR	Recall	Precision	F1	Time (s)
Band									
50	Amcoda	3.55	2.81	11.00	0.03	0.85	0.52	0.64	16
	CARE	4.32	3.87	16.50	0.05	0.44	0.26	0.33	16
150	Amcoda	3.71	3.21	23.10	0.01	0.69	0.64	0.66	244
	CARE	4.30	3.87	29.83	0.01	0.30	0.30	0.30	274
Hub									
50	Amcoda	4.47	2.08	7.84	0.01	0.86	0.61	0.72	16
	CARE	6.96	3.18	11.93	0.04	0.68	0.27	0.39	17
150	Amcoda	5.00	2.29	13.49	0.01	0.86	0.51	0.64	244
	CARE	6.64	3.15	20.13	0.01	0.61	0.38	0.47	272
Block									
50	Amcoda	4.80	2.98	10.09	0.02	0.76	0.60	0.67	17
	CARE	8.06	4.73	16.93	0.05	0.39	0.21	0.28	17
150	Amcoda	6.51	3.64	20.71	0.00	0.73	0.66	0.69	248
	CARE	8.59	5.22	29.55	0.01	0.29	0.26	0.27	271
Random									
50	Amcoda	5.41	3.27	11.54	0.02	0.75	0.66	0.70	15
	CARE	6.20	3.67	12.97	0.04	0.69	0.38	0.49	18
150	Amcoda	6.27	3.42	21.69	0.01	0.73	0.64	0.68	243
	CARE	7.30	3.82	24.05	0.01	0.66	0.53	0.59	272

Table S6: Performance of Amcoda under varying proportions of zero counts for the band network with $(n, p_1, p_2) = (200, 50, 50)$ and $(200, 150, 150)$. Results are averaged over 20 simulation replicates.

$p_1 = p_2$	Proportion of 0s	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_{\mathrm{F}}$	FPR	Recall	Precision	F1	Time (s)
50	5%	4.01	3.51	13.38	0.03	0.70	0.46	0.55	24
	25%	4.00	3.54	14.85	0.03	0.50	0.43	0.46	12
	50%	3.93	3.44	14.43	0.02	0.30	0.37	0.33	21
150	5%	3.92	3.53	24.61	0.01	0.70	0.55	0.61	296
	25%	3.94	3.51	25.69	0.01	0.52	0.42	0.46	280
	50%	3.87	3.35	24.49	0.01	0.30	0.31	0.31	280

Table S7: Performance comparison of Amcoda and competing methods (gmcoda, glasso and SPIEC-EASI) for band, hub, block and random networks with $\mu \sim \text{Unif}([0,5]^p)$. A dash ("–") indicates that the running time exceeded 30 minutes and the job was terminated. Results are averaged over 20 simulation replicates.

$p_1 = p_2$	Method	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_1$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2$	$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_F$	FPR	Recall	Precision	F1	Time (s)
Band									
50	Amcoda	3.52	2.82	10.94	0.03	0.86	0.51	0.64	14
	gmcoda	3.81	3.40	14.89	0.07	0.80	0.32	0.45	1
	glasso	4.15	3.82	17.36	0.01	0.01	0.03	0.01	97
	SPIEC-EASI	3.84	3.35	14.40	0.03	0.76	0.54	0.63	37
150	Amcoda	3.74	3.20	23.14	0.01	0.69	0.63	0.66	269
	gmcoda	3.95	3.56	27.42	0.03	0.66	0.25	0.37	8
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	3.96	3.31	24.06	0.01	0.81	0.48	0.60	575
Hub									
50	Amcoda	4.29	1.99	7.78	0.01	0.86	0.58	0.69	12
	gmcoda	6.57	3.11	13.96	0.02	0.46	0.37	0.41	1
	glasso	9.37	4.34	20.31	0.06	0.04	0.00	0.01	182
	SPIEC-EASI	7.21	3.57	17.83	0.05	0.63	0.21	0.31	58
150	Amcoda	4.93	2.27	13.44	0.01	0.86	0.51	0.64	245
	gmcoda	6.66	3.24	22.78	0.01	0.61	0.42	0.50	5
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	6.27	3.44	28.72	0.01	0.91	0.36	0.52	715
Block									
50	Amcoda	4.98	3.02	10.12	0.02	0.77	0.60	0.67	13
	gmcoda	6.19	4.16	15.92	0.03	0.56	0.38	0.45	1
	glasso	7.61	4.89	21.16	0.13	0.18	0.04	0.06	114
	SPIEC-EASI	6.33	4.09	17.69	0.04	0.80	0.39	0.53	52
150	Amcoda	6.46	3.63	20.71	0.01	0.73	0.65	0.69	295
	gmcoda	7.44	4.37	28.88	0.02	0.67	0.34	0.45	10
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	7.74	4.87	37.88	0.02	0.74	0.34	0.47	906
Random									
50	Amcoda	5.26	3.22	11.44	0.02	0.76	0.64	0.69	13
	gmcoda	8.18	4.46	17.56	0.04	0.57	0.36	0.44	2
	glasso	9.62	5.89	27.11	0.01	0.01	0.02	0.01	116
	SPIEC-EASI	9.40	5.64	25.88	0.09	0.39	0.15	0.21	88
150	Amcoda	6.35	3.43	21.73	0.01	0.73	0.64	0.68	301
	gmcoda	8.19	3.96	27.21	0.02	0.77	0.32	0.45	25
	glasso	–	–	–	–	–	–	–	–
	SPIEC-EASI	–	–	–	–	–	–	–	–

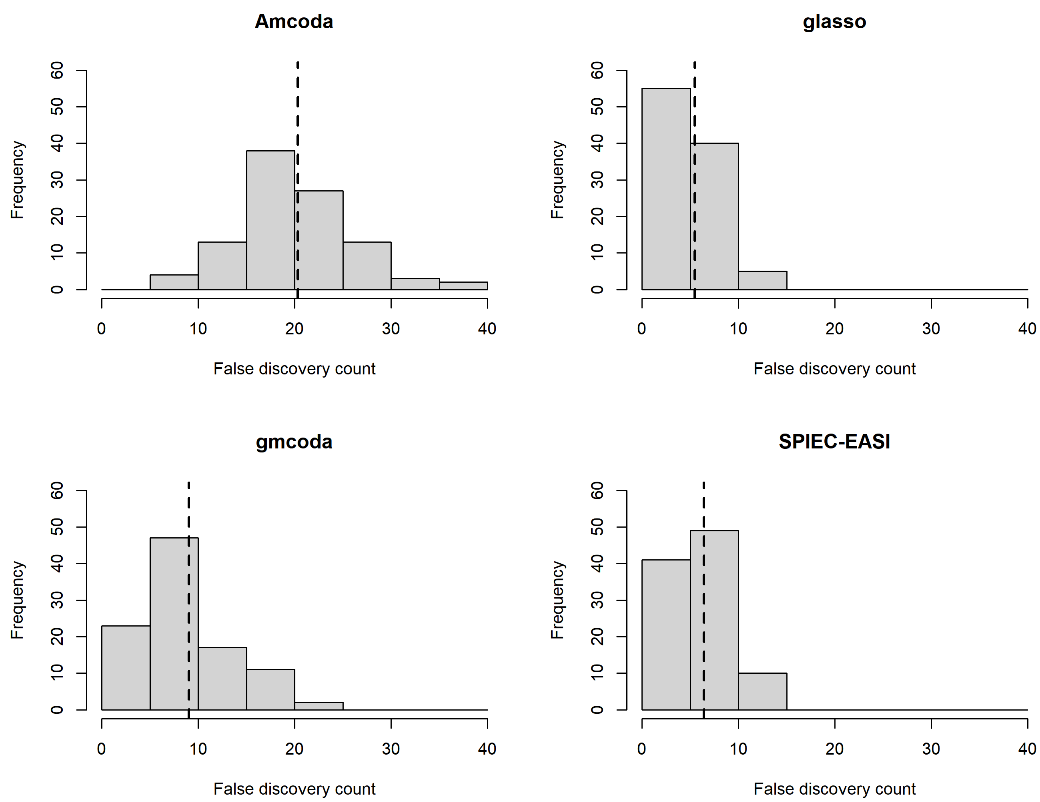

Figure S1: Histogram of false discovery counts under the null network (no edges) with $(n, p_1, p_2) = (47, 33, 15)$. The histogram is based on 100 replicates; the dashed line indicates the mean across replicates.