Supplementary Material to "SIMPLE: Statistical Inference on Membership Profiles in Large Networks"

Jianqing Fan, Yingying Fan, Xiao Han and Jinchi Lv

This Supplementary Material contains all the proofs and technical details.

A Proofs of main results

To facilitate the technical presentation, we list two definitions below, where n represents the network size and dimensionality of eigenvectors.

Definition 1. *Let ζ and ξ be a pair of random variables that may depend on n. We say that they satijsfy $\xi = O_{\prec}(\zeta)$ if for any pair of positive constants (a, b), there exists some positive integer $n_0(a, b)$ depending only on a and b such that $\mathbb{P}(|\xi| > n^a|\zeta|) \leq n^{-b}$ for all $n \geq n_0(a, b)$.*

Definition 2. *We say that an event $\mathfrak{A}_n$ holds with high probability if for any positive constant a, there exists some positive integer $n_0(a)$ depending only on a such that $\mathbb{P}(\mathfrak{A}_n) \geq 1 - n^{-a}$ for all $n \geq n_0(a)$.*

From Definitions 1 and 2 above, we can see that if $\xi = O_{\prec}(\zeta)$, then it holds that $\xi = O(n^a|\zeta|)$ with high probability for any positive constant a. The strong probabilistic bounds in the statements of Definitions 1 and 2 are in fact consequences of analyzing large binary random matrices given by networks.

Let us introduce some additional notation. Since the eigenvectors are always up to a sign change, for simplicity we fix the orientation of the empirical eigenvector $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k \geq 0$ for each $1 \leq k \leq K$, where $\mathbf{v}_k$ is the kth population eigenvector of the low-rank mean matrix $\mathbf{H}$ in our general network model (2). It is worth mentioning that all the variables are real-valued throughout the paper except that variable z can be complex-valued. For any nonzero complex number z, deterministic matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of appropriate dimensions, $1 \leq k \leq K$, and n-dimensional unit vector $\mathbf{u}$, we define

$$\mathcal{P}(\mathbf{M}_1, \mathbf{M}_2, z) = z\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, z), \quad \widetilde{\mathcal{P}}_{k,z} = \left[z^2(A_{\mathbf{v}_k,k,z}/z)'\right]^{-1}, \tag{A.1}$$

$$\mathbf{b}_{\mathbf{u},k,z} = \mathbf{u} - \mathbf{V}_{-k}\left[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1} \mathcal{R}^T(\mathbf{u}, \mathbf{V}_{-k}, z), \tag{A.2}$$

where $\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, z)$ is defined in (7),

$$A_{\mathbf{u},k,z} = \mathcal{P}(\mathbf{u}, \mathbf{v}_k, z) - \mathcal{P}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[z(\mathbf{D}_{-k})^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1} \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, z), \tag{A.3}$$

$(A_{\mathbf{v}_k,k,z}/z)'$ denotes the derivative of $A_{\mathbf{v}_k,k,z}/z$ with respect to complex variable z, $\mathbf{V}_{-k}$ represents a submatrix of $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_K)$ by removing the kth column, and $\mathbf{D}_{-k}$ stands for a principal submatrix of $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_K)$ by removing the kth diagonal entry.

A.1 Proof of Theorem 1

We first prove the conclusion in the first part of Theorem 1 under the null hypothesis $H_0 : \boldsymbol{\pi}_i = \boldsymbol{\pi}_j$, where (i, j) with $1 \leq i < j \leq n$ represents a given pair of nodes in the network. In particular, Lemma 9 in Section B.8 of Supplementary Material plays a key role in the technical analysis. For the given pair (i, j), let us define a new random matrix $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{x}}_{lm})_{1\leq l,m\leq n}$ based on the original random matrix $\mathbf{X} = (\mathbf{x}_{lm})_{1\leq l,m\leq n}$ by swapping the roles of nodes i and j, namely by setting

$$\widetilde{\mathbf{x}}_{lm} = \begin{cases} \mathbf{x}_{lm}, & l, m \in \{i, j\}^c \\ \mathbf{x}_{im}, & l = j,\, m \in \{i, j\}^c \\ \mathbf{x}_{jm}, & l = i,\, m \in \{i, j\}^c \\ \mathbf{x}_{li}, & m = j,\, l \in \{i, j\}^c \\ \mathbf{x}_{lj}, & m = i,\, l \in \{i, j\}^c \end{cases} \quad \text{and} \quad \widetilde{\mathbf{x}}_{lm} = \begin{cases} \mathbf{x}_{ij}, & (l, m) = (i, j) \text{ or } (j, i) \\ \mathbf{x}_{ii}, & l = m = j \\ \mathbf{x}_{jj}, & l = m = i \end{cases}, \tag{A.4}$$

where $\{i, j\}^c$ stands for the complement of set $\{i, j\}$ in the node set $\{1, \cdots, n\}$. It is easy to see that the new symmetric random matrix $\widetilde{\mathbf{X}}$ defined in (A.4) is simply the adjacency matrix of a network given by the mixed membership model (10) by swapping the ith and jth rows, $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$, of the community membership probability matrix $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_n)^T$.

By the above definition of $\widetilde{\mathbf{X}}$, we can see that under the null hypothesis $H_0 : \boldsymbol{\pi}_i = \boldsymbol{\pi}_j$, it holds that

$$\widetilde{\mathbf{X}} \overset{d}{=} \mathbf{X}, \tag{A.5}$$

where $\overset{d}{=}$ denotes being equal in distribution. The representation in (A.5) entails that for each $1 \leq k \leq K$, the ith and jth components of the kth population eigenvector $\mathbf{v}_k$ are identical; that is,

$$\mathbf{v}_k(i) = \mathbf{v}_k(j).$$

This identity along with the asymptotic expansion of the empirical eigenvector $\widehat{\mathbf{v}}_k$ in (B.25) given in Lemma 9 results in

$$\widehat{\mathbf{v}}_k(i) - \widehat{\mathbf{v}}_k(j) = \frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W} \mathbf{v}_k}{t_k} + O_{\prec}\Big(\frac{\alpha_n^2}{\sqrt{n}|d_k|^2} + \frac{1}{\sqrt{n}|d_k|}\Big). \tag{A.6}$$

Note that although the expectation of $\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k$ can be nonzero, the difference of expectations $\mathbb{E}(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W} \mathbf{v}_k = 0$ under the null hypothesis by (A.5). It follows from Lemma 7 in Section B.6 and Lemma 15 in Section C.6 of Supplementary Material that

$$n^{1-c_2}\theta \lesssim d_k \sim t_k \lesssim n\theta \quad \text{and} \quad \alpha_n = O(\sqrt{n\theta}),$$

where $\sim$ denotes the same asymptotic order. Condition 3 ensures that there exists some positive constant ϵ such that

$$\mathrm{SD}\left((\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_k\right) \sim \sqrt{\theta} \gg n^{\epsilon} n^{c_2 - 1/2} \gtrsim n^{\epsilon}\left(\frac{\alpha_n^2}{\sqrt{n}|d_k|} + \frac{1}{\sqrt{n}}\right), \tag{A.7}$$

which guarantees that $O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}d_k^2} + \frac{1}{\sqrt{n}|d_k|})$ in (A.6) is negligible compared to the first term on the right hand side. Here SD represents the standard deviation of a random variable. Moreover, by Lemma 6 in Section B.5 of Supplementary Material we have $\|\mathbf{V}\|_\infty = O(\frac{1}{\sqrt{n}}) \ll \min_{1\le k\le K} \mathrm{SD}((\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_k) \sim \sqrt{\theta}$, and hence $((\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_k)_{k=1}^K$ satisfies the conditions of Lemma 4 in Section B.3 of Supplementary Material with $h_n = \theta$. Then it holds that

$$\begin{aligned}
&\mathbf{\Sigma}_1^{-1/2}(\widehat{\mathbf{V}}(i) - \widehat{\mathbf{V}}(j)) \\
&= \mathbf{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T + o_p(1) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}),
\end{aligned} \tag{A.8}$$

which proves (12).

We next establish (13) under the condition of $\sqrt{n\theta}\|\boldsymbol{\pi}_i - \boldsymbol{\pi}_j\| \to \infty$. By (B.25) in Lemma 9, we have

$$\begin{aligned}
&\mathbf{D}(\widehat{\mathbf{V}}(i) - \widehat{\mathbf{V}}(j)) \\
&= \mathbf{D}(\mathbf{V}(i) - \mathbf{V}(j)) + \left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}|d_K|}).
\end{aligned} \tag{A.9}$$

In view of (A.7), it holds that

$$\left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{W}\mathbf{v}_K}{t_K/d_K}\right) = O_p(\sqrt{\theta}).$$

Thus it suffices to show that

$$\|\mathbf{D}(\mathbf{V}(i) - \mathbf{V}(j))\| \gg \sqrt{\theta}.$$

In fact, it follows from (B.17) that

$$\mathbf{D}(\mathbf{V}(i) - \mathbf{V}(j)) = \mathbf{D}\mathbf{B}(\boldsymbol{\pi}_i - \boldsymbol{\pi}_j).$$

This along with (B.18) and Condition 2 leads to

$$\begin{aligned}
\|\mathbf{D}(\mathbf{V}(i) - \mathbf{V}(j))\| = \|\mathbf{D}(\boldsymbol{\pi}_i - \boldsymbol{\pi}_j)^T \mathbf{B}\| &\ge |d_K|\sqrt{(\boldsymbol{\pi}_i - \boldsymbol{\pi}_j)^T(\mathbf{\Pi}^T\mathbf{\Pi})^{-1}(\boldsymbol{\pi}_i - \boldsymbol{\pi}_j)} \\
&\gtrsim \|\pi_1 - \pi_2\| n^{1/2 - c_2}\theta \gg \sqrt{\theta},
\end{aligned}$$

which concludes the proof of (13).

Finally, we prove (14). The conclusion follows immediately from (A.9) and $(\mathbf{V}(i) - \mathbf{V}(j))^T \boldsymbol{\Sigma}_1^{-1} (\mathbf{V}(i) - \mathbf{V}(j)) \to \mu$ as $n \to \infty$. This completes the proof of Theorem 1.

A.2 Proof of Theorem 2

As guaranteed by Slutsky's lemma, the asymptotic distributions of test statistics after replacing $\boldsymbol{\Sigma}_1$ with $\widehat{\mathbf{S}}_1$ stay the same. Thus we need only to prove that the asymptotic distributions are the same after replacing K with its estimate $\widehat{K}$ in the test statistics.

To ease the presentation, we write $T_{ij} = T_{ij}(K)$ and $\widehat{T}_{ij} = T_{ij}(\widehat{K})$ to emphasize their dependency on K and $\widehat{K}$, respectively. By (12) of Theorem 1, we have for any $t > 0$,

$$\lim_{n\to\infty} \mathbb{P}(T_{ij}(K) < t) = \mathbb{P}(\chi_K^2 < t). \tag{A.10}$$

By the condition on $\widehat{K}$, it holds that

$$\mathbb{P}(\widehat{K} = K) = 1 - o(1). \tag{A.11}$$

Then by the properties of conditional probability, we deduce

$$\begin{aligned}
&\mathbb{P}(T_{ij}(\widehat{K}) < t) = \mathbb{P}(T_{ij}(\widehat{K}) < t|\widehat{K} = K)\mathbb{P}(\widehat{K} = K) + \mathbb{P}(T_{ij}(\widehat{K}) < t|\widehat{K} \neq K)\mathbb{P}(\widehat{K} \neq K) \\
&= \mathbb{P}(T_{ij}(K) < t|\widehat{K} = K)\mathbb{P}(\widehat{K} = K) + o(1) \\
&= \mathbb{P}(T_{ij}(K) < t|\widehat{K} = K)\mathbb{P}(\widehat{K} = K) + \mathbb{P}(T_{ij}(K) < t|\widehat{K} \neq K)\mathbb{P}(\widehat{K} \neq K) + o(1) \\
&= \mathbb{P}(T_{ij}(K) < t) + o(1).
\end{aligned} \tag{A.12}$$

Observe that the $o(1)$ term comes from (*A*.11) and thus it holds uniformly for any t. Combining (A.12) with (A.10), we can show that

$$\lim_{n\to\infty} \mathbb{P}(T_{ij}(\widehat{K}) < t) = \mathbb{P}(\chi_K^2 < t). \tag{A.13}$$

Therefore, the same conclusion as in (12) of Theorem 1 is proved. Results in (13) and (14) can be shown using similar arguments and are omitted here for simplicity. This concludes the proof of Theorem 2.

A.3 Proof of Corollary 2

Recall that in the proof of Theorem 2, we denote by $T_{ij} = T_{ij}(K)$ and $\widehat{T}_{ij} = T_{ij}(\widehat{K})$ to emphasize their dependency on K and $\widehat{K}$. It suffices to prove that the impact of the use of

$\widehat{K}$ in place of K is asymptotically negligible. In fact, we can deduce that

$$\begin{aligned}
\mathbb{P}(T_{ij}(\widehat{K}) > \chi^2_{\widehat{K},1-\alpha}) &= \mathbb{P}(T_{ij}(\widehat{K}) > \chi^2_{\widehat{K},1-\alpha} | \widehat{K} = K)\mathbb{P}(\widehat{K} = K) \\
&\quad + \mathbb{P}(T_{ij}(\widehat{K}) > \chi^2_{\widehat{K},1-\alpha} | \widehat{K} \neq K)\mathbb{P}(\widehat{K} \neq K) \\
&= \mathbb{P}(T_{ij}(K) > \chi^2_{K,1-\alpha} | \widehat{K} = K)\mathbb{P}(\widehat{K} = K) + o(1) \\
&= \mathbb{P}(T_{ij}(K) > \chi^2_{K,1-\alpha} | \widehat{K} = K)\mathbb{P}(\widehat{K} = K) \\
&\quad + \mathbb{P}(T_{ij}(K) > \chi^2_{K,1-\alpha} | \widehat{K} \neq K)\mathbb{P}(\widehat{K} \neq K) + o(1) \\
&= \mathbb{P}(T_{ij}(K) > \chi^2_{K,1-\alpha}) + o(1). \qquad \text{(A.14)}
\end{aligned}$$

By (A.14), under the null hypothesis we have

$$\lim_{n\to\infty} \mathbb{P}(\widehat{T}_{ij} > \chi^2_{\widehat{K},1-\alpha}) = \lim_{n\to\infty} \mathbb{P}(T_{ij} > \chi^2_{K,1-\alpha}) = \alpha \qquad \text{(A.15)}$$

for any constant $\alpha \in (0, 1)$. Moreover, by (A.12), under the alternative hypothesis, for any arbitrarily large constant $C > 0$ it holds that

$$\lim_{n\to\infty} \mathbb{P}(\widehat{T}_{ij} > C) = \lim_{n\to\infty} \mathbb{P}(T_{ij} > C) = 1. \qquad \text{(A.16)}$$

Therefore, combining (A.15) and (A.16) completes the proof of Corollary 2.

A.4 Proof of Theorem 3

We begin with listing some basic properties of $\mathbf{v}_k$ and d_k:

1). We can choose a direction such that all components of $\mathbf{v}_1$ are nonnegative. Moreover, $\min_{1\leq l\leq n}\{\mathbf{v}_1(l)\} \sim \frac{1}{\sqrt{n}}$.

2). $\max_{1\leq k\leq K} \|\mathbf{v}_k\|_\infty \leq \frac{C}{\sqrt{n}}$ for some positive constant C.

3). $\alpha_n \leq \sqrt{n}\theta_{\max}$.

4). $|d_K| \geq cn^{1-2c_2}\theta^2_{\min}$ and $|d_1| \leq c^{-1}n\theta^2_{\max}$ for some positive constant c.

Here the second statement is ensured by Lemma 6. The third and fourth statements are guaranteed by Lemma 7, and the remaining properties are entailed by Lemma B.2 of Jin et al. (2017). One should notice that the proof of Lemma B.2 of Jin et al. (2017) does not require $\{d_k\}_{k=1}^K$ have the same order.

By Condition 5 and Statement 4 above, we have

$$\frac{1}{n^{1/2-c_2}|t_k|} \ll \frac{\min_{1\leq k\leq K,\, t=i,j} \sqrt{\mathrm{var}(\mathbf{e}_t^T \mathbf{W}\mathbf{v}_k)}}{|t_k|}.$$

By (B.19), there exists some $K \times K$ matrix $\mathbf{B}$ such that

$$\mathbf{V} = \mathbf{\Theta}\mathbf{\Pi}\mathbf{B}. \tag{A.17}$$

Recall that $\mathbf{\Theta}$ is a diagonal matrix. Then it follows from (A.17) that under the null hypothesis, we have

$$\frac{\mathbf{v}_k(i)}{\mathbf{v}_1(i)} = \frac{\mathbf{v}_k(j)}{\mathbf{v}_1(j)}, \quad k = 1, \cdots, K. \tag{A.18}$$

Here we use the exchangeability between rows i and j of matrix $\mathbf{\Pi}\mathbf{B}$ under the null hypothesis as argued under the mixed membership model (see the beginning of the proof of Theorem 1). In light of the asymptotic expansion in Lemma 9, we deduce

$$\widehat{\mathbf{v}}_k(i) = \mathbf{v}_k(i) + \frac{\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k}{t_k} + O_{\prec}(\frac{1}{n^{1/2-c_2}|t_k|}). \tag{A.19}$$

Moreover, it follows from Corollary 3 in Section C.2 of Supplementary Material, Condition 4, and the statements at the beginning of this proof that

$$\frac{\mathbf{e}_s^T\mathbf{W}\mathbf{v}_k}{t_k} = O_{\prec}(\frac{\theta_{\max}}{|t_k|}), \quad s = i, j, \ k = 1, \cdots, K. \tag{A.20}$$

Thus, by (A.18)–(A.20) and Statement 1 above we have under the null hypothesis that

$$\begin{aligned}
\mathbf{Y}(i,k) - \mathbf{Y}(j,k) &= \frac{\widehat{\mathbf{v}}_k(i)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{\mathbf{v}}_k(j)}{\widehat{\mathbf{v}}_1(j)} \\
&= \frac{\mathbf{v}_k(i) + \frac{\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k}{t_k} + O_{\prec}(\frac{1}{n^{1/2-c_2}|t_k|})}{\mathbf{v}_1(i) + \frac{\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k}{t_1} + O_{\prec}(\frac{1}{n^{1/2-c_2}|t_1|})} - \frac{\mathbf{v}_k(j) + \frac{\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k}{t_k} + O_{\prec}(\frac{1}{n^{1/2-c_2}|t_k|})}{\mathbf{v}_1(j) + \frac{\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k}{t_1} + O_{\prec}(\frac{1}{n^{1/2-c_2}|t_1|})} \\
&= \frac{\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k}{t_k\mathbf{v}_1(i)} - \frac{\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k}{t_k\mathbf{v}_1(j)} - \frac{\mathbf{v}_k(i)\mathbf{e}_i^T\mathbf{W}\mathbf{v}_1}{t_1\mathbf{v}_1^2(i)} + \frac{\mathbf{v}_k(j)\mathbf{e}_j^T\mathbf{W}\mathbf{v}_1}{t_1\mathbf{v}_1^2(j)} + O_{\prec}(\frac{n^{c_2}}{|t_k|}) \\
&= \frac{\mathbf{e}_i^T\mathbf{W}[\mathbf{v}_k - \frac{t_k\mathbf{v}_k(i)}{t_1\mathbf{v}_1(i)}\mathbf{v}_1]}{t_k\mathbf{v}_1(i)} - \frac{\mathbf{e}_j^T\mathbf{W}[\mathbf{v}_k - \frac{t_k\mathbf{v}_k(j)}{t_1\mathbf{v}_1(j)}\mathbf{v}_1]}{t_k\mathbf{v}_1(j)} + O_{\prec}(\frac{n^{c_2}}{|t_k|}).
\end{aligned} \tag{A.21}$$

Denote by $\mathbf{y}_k = \frac{\mathbf{v}_k - \frac{t_k\mathbf{v}_k(i)}{t_1\mathbf{v}_1(i)}\mathbf{v}_1}{t_k\mathbf{v}_1(i)}$ and $\mathbf{z}_k = \frac{\mathbf{v}_k - \frac{t_k\mathbf{v}_k(j)}{t_1\mathbf{v}_1(j)}\mathbf{v}_1}{t_k\mathbf{v}_1(j)}$. Then we have $f_k = \mathbf{e}_i^T\mathbf{W}\mathbf{y}_k - \mathbf{e}_j^T\mathbf{W}\mathbf{z}_k$ with f_k defined in (22), and

$$\mathbf{Y}(i,k) - \mathbf{Y}(j,k) = f_k + O_{\prec}(\frac{n^{c_2}}{|t_k|}). \tag{A.22}$$

To establish the central limit theorem, we need to compare the ord4er of the variance of f_k with that of the residual term $O_{\prec}(\frac{n^{2c_2}}{t_k^2})$. The variance of f_k is

$$\mathrm{var}(f_k) = \sum_{l=1}^{n}\mathrm{var}(w_{il})\mathbf{y}_k^2(l) + \sum_{l=1}^{n}\mathrm{var}(w_{jl})\mathbf{z}_k^2(l) - \mathrm{var}(w_{ij})\left[\mathbf{y}_k(i)\mathbf{z}_k(j) + \mathbf{y}_k(j)\mathbf{z}_k(i)\right]. \tag{A.23}$$

By Statements 1 and 2 at the beginning of this proof and (A.18), we can conclude that $\max_{1\le l\le n}\{|\mathbf{y}_k(l)|, |\mathbf{z}_k(l)|\} = O(\frac{1}{|t_k|})$ and $\mathbf{y}_k(l) \sim \mathbf{z}_k(l)$, $l = 1, \cdots, n$. Consequently, we obtain

$$\mathrm{var}(w_{ij})\left[\mathbf{y}_k(i)\mathbf{z}_k(j) + \mathbf{y}_k(j)\mathbf{z}_k(i)\right] = O(\frac{1}{t_k^2}). \tag{A.24}$$

By Condition 6, it holds that $(n\theta_{\max}^2)^{-1}d_k^2\mathrm{var}(f_k) = (n\theta_{\max}^2)^{-1}d_k^2\mathrm{var}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k - \mathbf{e}_j^T\mathbf{W}\mathbf{z}_k) \sim 1$. Combining the previous two results and by Statement 4, the last term on the left hand side of (A.23) is asymptotically negligible compared to the right hand side.

Note that under the null hypothesis $\boldsymbol{\pi}_i = \boldsymbol{\pi}_j$ and model (6), we have $\frac{\mathbf{H}_{il}}{\theta_i} = \frac{\mathbf{H}_{jl}}{\theta_j}$. Since $\mathbf{X} = \mathbf{H} + \mathbf{W}$ with $\mathbf{W}$ a generalized Wigner matrix, it follows from the properties of Bernoulli random variables that $\mathrm{var}(w_{il}) \sim \mathrm{var}(w_{jl})$. Thus the first two terms on the left hand side of (A.23) are comparable and satisfy that

$$\begin{aligned}
&(n\theta_{\max}^2)^{-1}d_k^2\mathrm{var}\left(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k\right) = (n\theta_{\max}^2)^{-1}d_k^2\sum_{l=1}^{n}2\mathrm{var}(w_{il})\mathbf{y}_k^2(l) \\
&\sim (n\theta_{\max}^2)^{-1}d_k^2\sum_{l=1}^{n}2\mathrm{var}(w_{jl})\mathbf{z}_k^2(l) = (n\theta_{\max}^2)^{-1}d_k^2\mathrm{var}\left(\mathbf{e}_j^T\mathbf{W}\mathbf{z}_k\right) \\
&\sim (n\theta_{\max}^2)^{-1}d_k^2\mathrm{var}(f_k) \sim 1.
\end{aligned} \tag{A.25}$$

Consequently, $\mathrm{var}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k) \sim \mathrm{var}(\mathbf{e}_i^T\mathbf{W}\mathbf{z}_k) \sim \mathrm{var}(f_k)$.

Now we are ready to check the conditions of Lemma 4. By $\max_l\{|\mathbf{y}_k(l)|, |\mathbf{z}_k(l)|\} = O(\frac{1}{|t_k|})$ (see (A.24) above) and noticing that the expectations of the off-diagonal entries of $\mathbf{W}$ are zero, we have $|\mathbb{E}(f_k)| = |\mathbb{E}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k - \mathbf{e}_j^T\mathbf{W}\mathbf{z}_k)| = |\mathbb{E}(w_{ii}\mathbf{y}_k(i) - w_{jj}\mathbf{z}_k(j))| \le |\mathbf{y}_k(i)| + |\mathbf{z}_k(j)| = O(\frac{1}{|t_k|})$, which means that the expectation of $\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k - \mathbf{e}_j^T\mathbf{W}\mathbf{z}_k$ is asymptotically negligible compared to its standard deviation. Moreover, by (A.25) it holds that $\max_l\{|\mathbf{y}_k(l)|, |\mathbf{z}_k(l)|\} \ll \min_{1\le k\le K}\min\{\mathrm{SD}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_k), \mathrm{SD}(\mathbf{e}_j^T\mathbf{W}\mathbf{z}_k)\}$ and hence they satisfy the conditions of Lemma 4 with $h_n = n\theta_{\max}^2$. Thus we arrive at

$$\mathrm{cov}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_2, \mathbf{e}_j^T\mathbf{W}\mathbf{z}_2, \cdots, \mathbf{e}_j^T\mathbf{W}\mathbf{z}_K)^{-1/2}(\mathbf{e}_i^T\mathbf{W}\mathbf{y}_2, \mathbf{e}_j^T\mathbf{W}\mathbf{z}_2, \cdots, \mathbf{e}_j^T\mathbf{W}\mathbf{z}_K)^T \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}). \tag{A.26}$$

Using the compact notation, (A.26) can be rewritten as

$$\boldsymbol{\Sigma}_2^{-1/2}(f_2, \cdots, f_K)^T \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}). \tag{A.27}$$

Furthermore, there exists some positive constant ϵ such that $\mathrm{SD}(f_k) \sim \frac{\sqrt{n}\theta_{\max}}{|t_k|} \gg n^{\epsilon}\frac{n^{c_2}}{|t_k|}$ by Condition 4. Hence $O_{\prec}(\frac{n^{c_2}}{|t_k|})$ involved in (A.22) is negligible compared to f_k. Finally, we can obtain from (A.22) and (A.27) that

$$\boldsymbol{\Sigma}_2^{-1/2}(\mathbf{Y}_i - \mathbf{Y}_j) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}),$$

which completes the proof for part i) of Theorem 3.

It remains to prove part ii) of Theorem 3. Under the alternative hypothesis that $\boldsymbol{\pi}_i \neq \boldsymbol{\pi}_j$, we have the generalized asymptotic expansion

$$\mathbf{Y}(i,k) - \mathbf{Y}(j,k) = \frac{\mathbf{v}_k(i)}{\mathbf{v}_1(i)} - \frac{\mathbf{v}_k(j)}{\mathbf{v}_1(j)} + \mathbf{e}_i^T\mathbf{W}\mathbf{y}_k - \mathbf{e}_j^T\mathbf{W}\mathbf{z}_k + O_{\prec}(\frac{n^{c_2}}{|t_k|}). \tag{A.28}$$

In view of (A.26), to complete the proof it suffices to show that

$$\left\|\frac{\mathbf{V}(i)}{\mathbf{v}_1(i)} - \frac{\mathbf{V}(j)}{\mathbf{v}_1(j)}\right\| \gg \frac{1}{n^{1/2-c_2}\theta_{\min}}. \tag{A.29}$$

Denote by $\mathbf{B}(i)$ the ith column of matrix $\mathbf{B}$ in (A.17). It follows from (A.17) that

$$\frac{\mathbf{V}(i)}{\mathbf{v}_1(i)} = \frac{\boldsymbol{\pi}_i^T\mathbf{B}}{\boldsymbol{\pi}_i^T\mathbf{B}(1)} \quad \text{and} \quad \frac{\mathbf{V}(j)}{\mathbf{v}_1(j)} = \frac{\boldsymbol{\pi}_j^T\mathbf{B}}{\boldsymbol{\pi}_j^T\mathbf{B}(1)}.$$

Let $a_i = \boldsymbol{\pi}_i^T\mathbf{B}(1)$ and $a_j = \boldsymbol{\pi}_j^T\mathbf{B}(1)$. Note that by Statements 1 and 2 at the beginning of this proof, we have $\mathbf{v}_1(i) \sim \mathbf{v}_1(j) \sim \frac{1}{\sqrt{n}}$. In light of (A.17), it holds that $\mathbf{v}_1(i) = \theta_i a_i$ and $\mathbf{v}_1(j) = \theta_j a_j$. Combining these two results yields

$$a_i \sim a_j \sim \frac{1}{\sqrt{n}\theta_{\min}}.$$

Moreover, it holds that

$$\frac{\boldsymbol{\pi}_i^T\mathbf{B}}{\boldsymbol{\pi}_i^T\mathbf{B}(1)} - \frac{\boldsymbol{\pi}_j^T\mathbf{B}}{\boldsymbol{\pi}_j^T\mathbf{B}(1)} = (a_i^{-1}, -a_j^{-1})(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)^T\mathbf{B},$$

which entails that

$$\left\|\frac{\mathbf{V}(i)}{\mathbf{v}_1(i)} - \frac{\mathbf{V}(j)}{\mathbf{v}_1(j)}\right\|^2 \geq \|(a_i^{-1}, -a_j^{-1})\|^2\lambda_{\min}((\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)^T(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j))\lambda_{\min}(\mathbf{B}\mathbf{B}^T).$$

Here $\lambda_{\min}(\cdot)$ stands for the smallest eigenvalue. By (A.17), similar to (B.18) we can show that

$$\mathbf{B}\mathbf{B}^T = (\boldsymbol{\Pi}^T\boldsymbol{\Theta}^2\boldsymbol{\Pi})^{-1}.$$

Thus $\lambda_{\min}(\mathbf{B}\mathbf{B}^T) \sim \frac{1}{n\theta_{\min}^2}$. By the condition that $\lambda_2(\boldsymbol{\pi}_i\boldsymbol{\pi}_i^T + \boldsymbol{\pi}_j\boldsymbol{\pi}_j^T) \gg \frac{1}{n^{1-2c_2}\theta_{\min}^2}$ in Theorem 3, it holds that

$$\lambda_{\min}((\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)^T(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)) = \lambda_2(\boldsymbol{\pi}_i\boldsymbol{\pi}_i^T + \boldsymbol{\pi}_j\boldsymbol{\pi}_j^T) \gg \frac{1}{n^{1-2c_2}\theta_{\min}^2}.$$

Therefore, combining the above arguments results in

$$\left\|\frac{\mathbf{V}(i)}{\mathbf{v}_1(i)} - \frac{\mathbf{V}(j)}{\mathbf{v}_1(j)}\right\|^2 \gg \frac{1}{n^{1-2c_2}\theta_{\min}^2},$$

which concludes the proof of Theorem 3.

A.5 Proof of Theorem 4

The arguments for the proof of Theorem 4 are similar to those for the proof of Theorem 2 in Section A.2.

A.6 Proof of Theorem 5

By Lemma 1, (16) holds. Since $\widehat{K}$ is bounded with probability tending to 1, it suffices to show the entrywise convergence of $\widehat{\mathbf{\Sigma}}_1 = \theta^{-1}\mathbf{D}\widehat{\mathbf{S}}_1\mathbf{D}$ and $\widehat{\mathbf{\Sigma}}_2 = (n\theta_{\max})^{-1}\mathbf{D}\widehat{\mathbf{S}}_2\mathbf{D}$. As will be made clear later, the proof relies heavily on the asymptotic expansions of $(\widehat{\mathbf{\Sigma}}_1)_{11}$, $(\widehat{\mathbf{\Sigma}}_1)_{12}$, $(\widehat{\mathbf{\Sigma}}_2)_{11}$, and $(\widehat{\mathbf{\Sigma}}_2)_{12}$. We will provide only the full details on the convergence of $(\widehat{\mathbf{\Sigma}}_1)_{11}$. For the other cases, the asymptotic expansions will be provided and the technical details will be mostly omitted since the arguments of the proof are similar. Throughout the proof, we will use repeatedly the results in Lemma 9, and the node indices i and j are fixed.

We start with considering $(\widehat{\mathbf{\Sigma}}_1)_{11}$. First, by definitions of $\widehat{\mathbf{W}}$ we have the following expansions

$$(\theta^{-1}\mathbf{D}\mathbf{\Sigma}_1\mathbf{D})_{11} = \theta^{-1}\sum_{t=i,j,\ 1\leq l\leq n}\Big[\sigma_{tl}^2\mathbf{v}_1^2(l) - 2\sigma_{ij}^2\mathbf{v}_1(j)\mathbf{v}_1(i)\Big] \tag{A.30}$$

and

$$(\widehat{\mathbf{\Sigma}}_1)_{11} = (\theta^{-1}\mathbf{D}\widehat{\mathbf{S}}_1\mathbf{D})_{11} = \theta^{-1}\sum_{t=i,j,\ 1\leq l\leq n}\Big[\widehat{w}_{tl}^2\mathbf{v}_1^2(l) - 2\widehat{w}_{ij}^2\widehat{\mathbf{v}}_1(l)\widehat{\mathbf{v}}_1(i)\Big]. \tag{A.31}$$

It follows from Lemma 10 in Section B.9 of Supplementary Material that $\widehat{w}_{ij}^2\widehat{\mathbf{v}}_1(j)\widehat{\mathbf{v}}_1(i) = O_{\prec}(\frac{1}{n})$. In addition, by Lemmas 6 and 7 it holds that

$$\mathrm{var}\Big[\sum_{1\leq l\leq n}(w_{il}^2 - \sigma_{il}^2)\mathbf{v}_1^2(l)\Big] \leq \sum_{1\leq l\leq n}\mathbf{v}_1^4(l)\mathbb{E}w_{il}^2 = O(\frac{1}{n^2})(\alpha_n^2+1) = O(\frac{\theta}{n}). \tag{A.32}$$

The same inequality also holds for $\mathrm{var}[\sum_{1\leq l\leq n}(w_{jl}^2 - \sigma_{jl}^2)\mathbf{v}_1^2(l)]$. Thus we have

$$\sum_{t=i,j,\,1\leq l\leq n}(w_{tl}^2 - \sigma_{tl}^2)\mathbf{v}_1^2(l) = O_p(\frac{\sqrt{\theta}}{\sqrt{n}}), \tag{A.33}$$

which implies that

$$\sum_{t=i,j,\,1\leq l\leq n}w_{tl}^2\mathbf{v}_1^2(l) = \sum_{t=i,j,\,1\leq l\leq n}\sigma_{tl}^2\mathbf{v}_1^2(l) + O_p(\frac{\sqrt{\theta}}{\sqrt{n}}). \tag{A.34}$$

By Lemmas 7 and 9, we have

$$\widehat{\mathbf{v}}_k(j) = \mathbf{v}_k(j) + \frac{\mathbf{e}_j^T \mathbf{W} \mathbf{v}_k}{t_k} + O_{\prec}(\frac{1}{n^{3/2-2c_2}\theta}).$$

It follows from Corollary 3 in Section C.2 and Lemma 13 in Section C.4 of Supplementary Material that

$$\begin{aligned}
&\sum_{t=i,j,1\le l\le n} w_{tl}^2[\mathbf{v}_1^2(l) - \widehat{\mathbf{v}}_1^2(l)] = 2\sum_{t=i,j,1\le l\le n} w_{tl}^2 \mathbf{v}_1(j)[\mathbf{v}_1(l) - \widehat{\mathbf{v}}_1(l)] + O_{\prec}(n^{2c_2-1}) \\
&= -\frac{2}{t_1}\sum_{t=i,j,1\le l\le n} w_{tl}^2 \mathbf{v}_1(l)\mathbf{e}_l^T \mathbf{W}\mathbf{v}_1 + O_{\prec}(\frac{1}{n^{1-2c_2}}) \\
&= O_{\prec}(\frac{\sqrt{\theta}}{n^{1/2-c_2}}). \qquad (A.35)
\end{aligned}$$

Similarly, by Lemma 10 we have

$$\sum_{t=i,j,1\le l\le n}^{n} w_{tl}^2 \widehat{\mathbf{v}}_1^2(l) = \sum_{t=i,j,1\le l\le n}^{n} \widehat{w}_{tl}^2 \widehat{\mathbf{v}}_k^2(l) + O_{\prec}(\frac{\sqrt{\theta}}{n^{1/2-c_2}}). \qquad (A.36)$$

Combining the equalities (A.30)–(A.36) yields

$$(\widehat{\mathbf{\Sigma}}_1)_{11} = \theta^{-1}(\mathbf{D}\mathbf{\Sigma}_1\mathbf{D})_{11} + O_{\prec}(\frac{1}{n^{1/2-c_2}\sqrt{\theta}}) + O_p(\frac{1}{\sqrt{n\theta}}) = \theta^{-1}(\mathbf{\Sigma}_1)_{11} + o_p(1), \qquad (A.37)$$

where we have used $O_{\prec}(\frac{1}{n^{1/2-c_2}\sqrt{\theta}}) = o_p(1)$ by Condition 2. This has proved the convergence of $(\widehat{\mathbf{\Sigma}}_1)_{11}$ to $(\mathbf{\Sigma}_1)_{11}$.

We next consider $(\widehat{\mathbf{\Sigma}}_1)_{12}$. By definitions, we have the following expansions

$$(\theta^{-1}\mathbf{D}\mathbf{\Sigma}_1\mathbf{D})_{12} = \theta^{-1}\Big\{\sum_{t=i,j} \sigma_{tl}^2 \mathbf{v}_1(l)\mathbf{v}_2(l) - \sigma_{ij}^2[\mathbf{v}_1(j)\mathbf{v}_2(i) + \mathbf{v}_1(i)\mathbf{v}_2(j)]\Big\} \qquad (A.38)$$

and

$$(\widehat{\mathbf{\Sigma}}_1)_{12} = \theta^{-1}\Big\{\sum_{t=i,j} \widehat{w}_{tl}^2 \widehat{\mathbf{v}}_1(l)\widehat{\mathbf{v}}_2(l) - \widehat{w}_{ij}^2[\widehat{\mathbf{v}}_1(j)\widehat{\mathbf{v}}_2(i) + \widehat{\mathbf{v}}_1(i)\widehat{\mathbf{v}}_2(j)]\Big\}. \qquad (A.39)$$

Based on the above two expansions, using similar arguments to those for proving (A.37) we can show that

$$(\widehat{\mathbf{\Sigma}}_1)_{12} = \theta^{-1}(\mathbf{D}\mathbf{\Sigma}_1\mathbf{D})_{12} + o_p(1). \qquad (A.40)$$

Now let us consider $\widehat{\mathbf{\Sigma}}_2$. Similar as above, we will provide only the asymptotic expansions for $(\widehat{\mathbf{\Sigma}}_2)_{11}$ and $(\widehat{\mathbf{\Sigma}}_2)_{12}$, and the remaining arguments are similar. By definitions, we can deduce that

$$
\begin{aligned}
&((n\theta_{\max}^2)^{-1}\mathbf{D}\boldsymbol{\Sigma}_2\mathbf{D})_{11} = (n\theta_{\max}^2)^{-1} d_2^2 \mathrm{var}(f_2) \\
&= \frac{d_2^2}{t_2^2 n\theta_{\max}^2}\Big\{\sum_{l\neq j}\sigma_{il}^2\Big[\frac{\mathbf{v}_2(l)}{\mathbf{v}_1(i)} - \frac{t_2\mathbf{v}_2(i)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(i)^2}\Big]^2 + \sum_{l\neq i}\sigma_{jl}^2\Big[\frac{\mathbf{v}_2(l)}{\mathbf{v}_1(j)} - \frac{t_2\mathbf{v}_2(j)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(j)^2}\Big]^2 \\
&\quad + \sigma_{ij}^2\Big[\frac{\mathbf{v}_2(j)}{\mathbf{v}_1(i)} - \frac{t_2\mathbf{v}_2(i)\mathbf{v}_1(j)}{t_1\mathbf{v}_1(i)^2} - \frac{\mathbf{v}_2(i)}{\mathbf{v}_1(j)} + \frac{t_2\mathbf{v}_2(j)\mathbf{v}_1(i)}{t_1\mathbf{v}_1(j)^2}\Big]^2\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
(\widehat{\boldsymbol{\Sigma}}_2)_{11} = \frac{d_2^2}{\widehat{d}_2^2 n\theta_{\max}^2}\Big\{&\sum_{l\neq j}\widehat{w}_{il}^2\Big[\frac{\widehat{\mathbf{v}}_2(l)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(i)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2}\Big]^2 + \sum_{l\neq i}\widehat{w}_{jl}^2\Big[\frac{\widehat{\mathbf{v}}_2(l)}{\widehat{\mathbf{v}}_1(j)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(j)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big]^2 \\
&+\widehat{w}_{ij}^2\Big[\frac{\widehat{\mathbf{v}}_2(j)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(i)\widehat{\mathbf{v}}_1(j)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2} - \frac{\widehat{\mathbf{v}}_2(i)}{\widehat{\mathbf{v}}_1(j)} + \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(j)\widehat{\mathbf{v}}_1(i)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big]^2\Big\}.
\end{aligned}
$$

Note that the expression of $(n\theta_{\max}^2\mathbf{D}\boldsymbol{\Sigma}_2\mathbf{D})_{11}$ is essentially the same as (A.30) up to a normalization factor involving $\mathbf{v}_1(i)$ and $\mathbf{v}_1(j)$. Thus applying the similar arguments to those for proving (17), we can establish the desired result.

Finally, the consistency of $(\widehat{\boldsymbol{\Sigma}}_2)_{12}$ can also be shown similarly using the following expansions

$$
\begin{aligned}
&((n\theta_{\max}^2)^{-1}\mathbf{D}\boldsymbol{\Sigma}_2\mathbf{D})_{12} \\
&= \frac{d_2 d_3}{t_2 t_3 n\theta_{\max}^2}\Big\{\sum_{l\neq j}\sigma_{il}^2\Big[\frac{\mathbf{v}_2(l)}{\mathbf{v}_1(i)} - \frac{t_2\mathbf{v}_2(i)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(i)^2}\Big]\Big[\frac{\mathbf{v}_3(l)}{\mathbf{v}_1(i)} - \frac{t_3\mathbf{v}_3(i)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(i)^2}\Big] \\
&\quad + \sum_{l\neq i}\sigma_{jl}^2\Big[\frac{\mathbf{v}_2(l)}{\mathbf{v}_1(j)} - \frac{t_2\mathbf{v}_2(j)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(j)^2}\Big]\Big[\frac{\mathbf{v}_2(l)}{\mathbf{v}_1(j)} - \frac{t_3\mathbf{v}_3(j)\mathbf{v}_1(l)}{t_1\mathbf{v}_1(j)^2}\Big] \\
&\quad + \sigma_{ij}^2\Big[\frac{\mathbf{v}_2(j)}{\mathbf{v}_1(i)} - \frac{t_2\mathbf{v}_2(i)\mathbf{v}_1(j)}{t_1\mathbf{v}_1(i)^2} - \frac{\mathbf{v}_2(i)}{\mathbf{v}_1(j)} + \frac{t_2\mathbf{v}_2(j)\mathbf{v}_1(i)}{t_1\mathbf{v}_1(j)^2}\Big] \\
&\quad \times \Big[\frac{\mathbf{v}_3(j)}{\mathbf{v}_1(i)} - \frac{t_3\mathbf{v}_3(i)\mathbf{v}_1(j)}{t_1\mathbf{v}_1(i)^2} - \frac{\mathbf{v}_3(i)}{\mathbf{v}_1(j)} + \frac{t_3\mathbf{v}_3(j)\mathbf{v}_1(i)}{t_1\mathbf{v}_1(j)^2}\Big]\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
(\widehat{\Sigma}_2)_{12} &= \frac{d_2 d_3}{\widehat{d}_2\widehat{d}_3 n\theta_{\max}^2}\Big\{\sum_{l\neq j}\widehat{w}_{il}^2\Big[\frac{\widehat{\mathbf{v}}_2(l)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(i)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2}\Big]\Big[\frac{\widehat{\mathbf{v}}_3(l)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_3\widehat{\mathbf{v}}_3(i)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2}\Big] \\
&+ \sum_{l\neq i}\widehat{w}_{jl}^2\Big[\frac{\widehat{\mathbf{v}}_2(l)}{\widehat{\mathbf{v}}_1(j)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(j)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big]\Big[\frac{\widehat{\mathbf{v}}_3(l)}{\widehat{\mathbf{v}}_1(j)} - \frac{\widehat{d}_3\widehat{\mathbf{v}}_3(j)\widehat{\mathbf{v}}_1(l)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big] \\
&+ \widehat{w}_{ij}^2\Big[\frac{\widehat{\mathbf{v}}_2(j)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(i)\widehat{\mathbf{v}}_1(j)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2} - \frac{\widehat{\mathbf{v}}_2(i)}{\widehat{\mathbf{v}}_1(j)} + \frac{\widehat{d}_2\widehat{\mathbf{v}}_2(j)\widehat{\mathbf{v}}_1(i)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big] \\
&\times \Big[\frac{\widehat{\mathbf{v}}_3(j)}{\widehat{\mathbf{v}}_1(i)} - \frac{\widehat{d}_3\widehat{\mathbf{v}}_3(i)\widehat{\mathbf{v}}_1(j)}{\widehat{d}_1\widehat{\mathbf{v}}_1(i)^2} - \frac{\widehat{\mathbf{v}}_3(i)}{\widehat{\mathbf{v}}_1(j)} + \frac{\widehat{d}_3\widehat{\mathbf{v}}_3(j)\widehat{\mathbf{v}}_1(i)}{\widehat{d}_1\widehat{\mathbf{v}}_1(j)^2}\Big]\Big\}.
\end{aligned}
$$

This completes the proof of Theorem 5.

B Some key lemmas and their proofs

B.1 Proof of Lemma 1

For each pair (i,j) with $i \neq j$, let us define a matrix $\mathbf{W}(i,j) = w_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)$. For $i = j$, we define a matrix $\mathbf{W}(i,j) = (w_{ii} - \mathbb{E}w_{ii})\mathbf{e}_i\mathbf{e}_j^T$. Then it is easy to see that

$$\|\sum_{1\leq i\leq j\leq n} \mathbf{W}(i,j) - \mathbf{W}\| = \|\text{diag}(\mathbf{W} - \mathbb{E}\mathbf{W})\| \leq 1. \tag{B.1}$$

It is straightforward to show that

$$\|\sum_{1\leq i\leq j\leq n} \mathbb{E}\mathbf{W}(i,j)^2\| = \alpha_n^2.$$

By Theorem 6.2 of Tropp (2012), for any constant $c > \sqrt{2}$ we have

$$\mathbb{P}(\|\sum_{1\leq i\leq j\leq n} \mathbf{W}(i,j)\| \geq c\sqrt{\log n}\alpha_n - 1) \leq n\exp\left[\frac{-(c\sqrt{\log n}\alpha_n - 1)^2}{2\alpha_n^2 + 2(c\sqrt{\log n}\alpha_n - 1)}\right] = o(1). \tag{B.2}$$

This together with (B.1) entails that

$$\mathbb{P}(\|\mathbf{W}\| \leq c\sqrt{\log n}\alpha_n) \geq 1 - o(1). \tag{B.3}$$

Note that this result is weaker than Lemma 14 in Section C.5.

By (B.3) and $|\widehat{d}_K - d_K| \leq \|\mathbf{W}\|$, and using the assumption of $|d_K| \gg \sqrt{\log n}\alpha_n$, it holds that

$$|\widehat{d}_K| \gg \sqrt{\log n}\alpha_n \tag{B.4}$$

with probability tending to one. Finally, by Weyl's inequality we have

$$\lambda_n(\mathbf{W}) = \lambda_n(\mathbf{W}) - \lambda_{K+1}(\mathbf{H}) \leq \lambda_{K+1}(\mathbf{X}) = \lambda_{K+1}(\mathbf{H}+\mathbf{W}) \leq \lambda_1(\mathbf{W}) + \lambda_{K+1}(\mathbf{H}) = \lambda_1(\mathbf{W}),$$

which leads to

$$|\widehat{d}_{K+1}| = |\lambda_{K+1}(\mathbf{X})| \leq \|\mathbf{W}\|. \tag{B.5}$$

Let us choose $c = \sqrt{2.01}$ and define

$$\widetilde{K} = \#\left\{|\widehat{d}_i| > \sqrt{2.01\log n}\alpha_n, i = 1, \cdots, n\right\}. \tag{B.6}$$

Then by (B.4)–(B.5), we can show that

$$\mathbb{P}(\widetilde{K} = K) = 1 - o(1). \tag{B.7}$$

Recall that X_{ij} follows the Bernoulli distribution. Thus it holds that

$$\sum_{j=1} \mathbb{E}w_{ij}^2 \leq \sum_{j=1} \mathbb{E}X_{ij}.$$

By Lemma 11 in Section C.1, choosing $l = 1$, $\mathbf{x} = \mathbf{e}_i$, and $\mathbf{y} = \frac{1}{\sqrt{n}}\mathbf{1}$ yields

$$\sum_{j=1} \mathbb{E}X_{ij} = \sum_{j=1} X_{ij} + O_{\prec}(\alpha_n),$$

where we have used $X_{ij} - \mathbb{E}X_{ij} = w_{ij}$. Thus it holds that

$$\max_i \sum_{j=1} X_{ij} \geq \max_i \sum_{j=1} \mathbb{E}w_{ij}^2 + O_{\prec}(\alpha_n) = \alpha_n^2 + O_{\prec}(\alpha_n).$$

This together with (B.6) and (B.7) results in

$$\mathbb{P}(\widehat{K} = K) = 1 - o(1), \tag{B.8}$$

which completes the proof of Lemma 1.

B.2 Proofs of Lemmas 2 and 3

The proofs of Lemmas 2 and 3 involve standard calculations and thus are omitted for brevity.

B.3 Lemma 4 and its proof

Lemma 4. *Let m be a fixed positive integer, $\mathbf{x}_i$ and $\mathbf{y}_i$ be n-dimensional unit vectors for $1 \leq i \leq m$, and $\mathbf{\Sigma} = (\Sigma_{ij})$ the covariance matrix with $\Sigma_{ij} = \mathrm{cov}(\mathbf{x}_i^T \mathbf{W} \mathbf{y}_i, \mathbf{x}_j^T \mathbf{W} \mathbf{y}_j)$. Assume that there exists some positive sequence (h_n) such that $\|\mathbf{\Sigma}^{-1}\| \sim \|\mathbf{\Sigma}\| \sim h_n$ and $\max_k\{\|\mathbf{x}_k\|_\infty \|\mathbf{y}_k\|_\infty\} \ll \|\mathbf{\Sigma}^{1/2}\|$. Then it holds that*

$$\mathbf{\Sigma}^{-1/2} \left(\mathbf{x}_1^T(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{y}_1, \cdots, \mathbf{x}_m^T(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{y}_m\right)^T \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{I}). \tag{B.9}$$

Proof. Note that it suffices to show that for any unit vector $\mathbf{c} = (c_1, \cdots, c_m)^T$, we have

$$\mathbf{c}^T \mathbf{\Sigma}^{-1/2} \big(\mathbf{x}_1^T(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{y}_1, \cdots, \mathbf{x}_m^T(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{y}_m\big)^T \xrightarrow{\mathscr{D}} N(0, 1). \tag{B.10}$$

Let $\mathbf{x}_i = (x_{1i}, \cdots, x_{ni})^T$ and $\mathbf{y}_i = (y_{1i}, \cdots, y_{ni})^T$, $i = 1, \cdots, m$. Since $\mathbf{W}$ is a symmetric random matrix of independent entries on and above the diagonal, we can deduce

$$\mathbf{x}_i^T\mathbf{W}\mathbf{y}_i - \mathbf{x}_i^T\mathbb{E}\mathbf{W}\mathbf{y}_i = \sum_{1\le s,t\le n,\, s<t} w_{st}(x_{si}y_{ti} + x_{ti}y_{si}) + \sum_{1\le s\le n} (w_{ss} - \mathbb{E}w_{ss})x_{si}y_{si} \tag{B.11}$$

and

$$\begin{aligned} s_n^2 &:= \mathrm{var}\left[\mathbf{c}^T\mathbf{\Sigma}^{-1/2}(\mathbf{x}_1^T(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{y}_1, \cdots, \mathbf{x}_m^T(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{y}_m)^T\right] \\ &= \mathbf{c}^T\mathbf{\Sigma}^{-1/2}\mathrm{cov}\left[(\mathbf{x}_1^T\mathbf{W}\mathbf{y}_1, \cdots, \mathbf{x}_m^T\mathbf{W}\mathbf{y}_m)^T\right]\mathbf{\Sigma}^{-1/2}\mathbf{c} = \mathbf{c}^T\mathbf{c} = 1. \end{aligned} \tag{B.12}$$

Denote by $\widetilde{\mathbf{c}} = \mathbf{\Sigma}^{-1/2}\mathbf{c} = (\widetilde{c}_1, \cdots, \widetilde{c}_m)^T$. Then it holds that

$$\mathbf{c}^T\mathbf{\Sigma}^{-1/2}\big(\mathbf{x}_1^T(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{y}_1, \cdots, \mathbf{x}_m^T(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{y}_m\big)^T = \mathrm{tr}\Big[(\mathbf{W}-\mathbb{E}\mathbf{W})\sum_{s=1}^m \widetilde{c}_s\mathbf{y}_s\mathbf{x}_s^T\Big].$$

Let $\mathbf{M} = (M_{ij}) = \sum_{s=1}^m \widetilde{c}_s\mathbf{y}_s\mathbf{x}_s^T$. By assumption, we have $\max_k \|\mathbf{x}_k\mathbf{y}_k^T\|_\infty \ll \|\mathbf{\Sigma}^{1/2}\| \sim \|\mathbf{\Sigma}^{-1/2}\|$, which entails that

$$\|\mathbf{M}\|_\infty \ll 1. \tag{B.13}$$

Then it follows from the assumption of $\max_{1\le i,j\le n}|w_{ij}| \le 1$ and (B.13) that

$$\begin{aligned} &\frac{1}{|s_n|^3}\Big(\sum_{1\le i,j\le n,\, i<j} \mathbb{E}|w_{ij}|^3|M_{ij}+M_{ji}|^3 + \sum_{1\le i\le n} \mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^3|M_{ii}|^3\Big) \\ &\quad\le \frac{2}{|s_n|^3}\Big(\sum_{1\le i,j\le n,\, i<j} \mathbb{E}|w_{ij}|^2|M_{ij}+M_{ji}|^3 + \sum_{1\le i\le n} \mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^2|M_{ii}|^3\Big) \\ &\quad\ll \frac{2}{|s_n|^3}\Big(\sum_{1\le i,j\le n,\, i<j} \mathbb{E}|w_{ij}|^2|M_{ij}+M_{ji}|^2 + \sum_{1\le i\le n} \mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^2|M_{ii}|^2\Big) \\ &\quad\le 2. \end{aligned} \tag{B.14}$$

Since w_{ij} with $1\le i<j\le n$ and $w_{ii}-\mathbb{E}w_{ii}$ with $1\le i\le n$ are independent random variables with zero mean, by the Lyapunov condition (see, for example, Theorem 27.3 of Billingsley (1995)) we can conclude that (B.10) holds. This concludes the proof of Lemma 4.

B.4 Lemma 5 and its proof

Lemma 5. *Under either model* (10) *and Conditions 1–2, or model (6) and Conditions 1 and 4, it holds that*

$$\|(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\| = O(|z|) \text{ for any } z \in [a_k, b_k], \tag{B.15}$$

where a_k and b_k are defined in (8).

Proof. The conclusion of Lemma 5 has been proved in (A.16) of Fan et al. (2020).

B.5 Lemma 6 and its proof

Lemma 6. *Under model* (10) *and Conditions 1–2, we have*

$$\max_{1 \le k \le K} \|\mathbf{v}_k\|_\infty = O(\frac{1}{\sqrt{n}}). \tag{B.16}$$

The same conclusion also holds under model (6) *and Conditions 1 and 4.*

Proof. We first consider model (10) and prove (B.16) under Conditions 1 and 2. In light of $\theta \mathbf{\Pi} \mathbf{P} \mathbf{\Pi}^T = \mathbf{V}\mathbf{D}\mathbf{V}^T$, we have $\theta \mathbf{\Pi}(\mathbf{P}\mathbf{\Pi}^T \mathbf{V}\mathbf{D}^{-1}) = \mathbf{V}$. This shows that $\mathbf{V}$ belongs to the space expanded by $\mathbf{\Pi}$. Thus there exists some $K \times K$ matrix $\mathbf{B}$ such that

$$\mathbf{V} = \mathbf{\Pi}\mathbf{B}. \tag{B.17}$$

Since $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, it holds that $\mathbf{B}^T\mathbf{\Pi}^T\mathbf{\Pi}\mathbf{B} = \mathbf{I}$, which entails that $\mathbf{B}\mathbf{B}^T\mathbf{\Pi}^T\mathbf{\Pi}\mathbf{B}\mathbf{B}^T = \mathbf{B}\mathbf{B}^T$ and

$$\mathbf{B}\mathbf{B}^T = (\mathbf{\Pi}^T\mathbf{\Pi})^{-1}. \tag{B.18}$$

By Condition 2, we can conclude that $\|(\mathbf{\Pi}^T\mathbf{\Pi})^{-1}\| = O(n^{-1})$ and thus each entry of matrix $\mathbf{B}$ is of order $O(\frac{1}{\sqrt{n}})$. Hence in view of (B.17), the desired result can be established.

Now let us consider model (6) under Conditions 1 and 4. For this model, we also have $\mathbf{\Theta}\mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^T\mathbf{\Theta} = \mathbf{V}\mathbf{D}\mathbf{V}^T$ and thus

$$\mathbf{\Theta}\mathbf{\Pi}(\mathbf{P}\mathbf{\Pi}^T\mathbf{\Theta}\mathbf{V}\mathbf{D}^{-1}) = \mathbf{V}. \tag{B.19}$$

Since $\mathbf{\Theta}$ is a diagonal matrix, we can see that $\mathbf{V}$ belongs to the space expanded by $\mathbf{\Pi}$. Let $\widetilde{\mathbf{\Pi}} = (\widetilde{\boldsymbol{\pi}}_1, \cdots, \widetilde{\boldsymbol{\pi}}_n)^T$ be the submatrix of $\mathbf{\Pi}$ such that

$$\widetilde{\boldsymbol{\pi}}_i = \begin{cases} \boldsymbol{\pi}_i & \text{if there exists some } 1 \le k \le K \text{ such that } \boldsymbol{\pi}_i(k) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

By Condition 4, it holds that $c_2 n^{c_2}\mathbf{I} \le \widetilde{\mathbf{\Pi}}^T\widetilde{\mathbf{\Pi}} = \sum_{i=1}^n \widetilde{\boldsymbol{\pi}}_i\widetilde{\boldsymbol{\pi}}_i^T \le \sum_{i=1}^n \boldsymbol{\pi}_i\boldsymbol{\pi}_i^T = \mathbf{\Pi}^T\mathbf{\Pi}$, which leads to $\|(\mathbf{\Pi}^T\mathbf{\Pi})^{-1}\| = O(n^{-1})$. Therefore, an application of similar arguments to those for (B.17)–(B.18) concludes the proof of Lemma 6.

B.6 Lemma 7 and its proof

Lemma 7. *Under model* (10) *and Condition 2 , it holds that*

$$\alpha_n^2 \le n\theta, \quad d_k \gtrsim n^{1-c_2}\theta, \quad d_1 = O(n\theta), \quad k = 1, \cdots, K. \tag{B.20}$$

Under model (6) and Condition 4, similarly we have

$$\alpha_n^2 \leq n\theta_{\max}^2, \quad d_k \gtrsim n^{1-c_2}\theta_{\min}^2, \quad d_1 = O(n\theta_{\max}^2), \quad k = 1, \cdots, K. \tag{B.21}$$

Proof. We show (B.20) first. It follows from $\sum_{k=1}^{K} \boldsymbol{\pi}_i(k) = 1$ that $\|\mathbf{\Pi}\|_F^2 = \sum_{i=1}^{n}\sum_{k=1}^{K} \boldsymbol{\pi}_i^2(k) \leq n$ and $\lambda_1(\mathbf{\Pi}^T\mathbf{\Pi}) = O(n)$. By Condition 2, we have

$$d_K = \theta\lambda_K(\mathbf{P}\mathbf{\Pi}^T\mathbf{\Pi}) \geq \theta\lambda_K(\mathbf{\Pi}^T\mathbf{\Pi})\lambda_K(\mathbf{P}) \geq c_0^2\theta n^{1-c_2}$$

and

$$d_1 \leq \theta\lambda_1(\mathbf{\Pi}^T\mathbf{\Pi})\lambda_1(\mathbf{P}) = O(\theta n).$$

Thus the second result in (B.20) is proved. Next by model (10), the (i,j)th entry h_{ij} of matrix $\mathbf{H}$ satisfies that

$$h_{ij} = \theta \sum_{s,t=1}^{K} \boldsymbol{\pi}_i(s)\boldsymbol{\pi}_j(t)p_{st} \leq \theta. \tag{B.22}$$

Since the entries of $\mathbf{X}$ follow the Bernoulli distributions, it follows from (B.22) that $\mathrm{var}(w_{ij}) \leq \theta$. Therefore, in view of the definition of α_n, we have

$$\alpha_n^2 = \max_j \sum_{i=1}^{n} \mathrm{var}(w_{ij}) \leq n\theta.$$

The results in (B.21) can also be proved using similar arguments. This completes the proof of Lemma 7.

B.7 Lemma 8

The following 3 Lemmas follow from Lemma 12 and exactly the same proof as Fan et al. (2020)

Lemma 8. *Under either model (10) and Conditions 1–2, or model (6) and Conditions 1 and 4, for $\mathbf{u} = \mathbf{e}_i$ or $\mathbf{v}_k$ we have the following asymptotic expansions*

$$\mathbf{u}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{v}_k = \Big[\widetilde{\mathcal{P}}_{k,t_k} - 2t_k^{-1}\widetilde{\mathcal{P}}_{k,t_k}^2\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}t_k^2})\Big]\Big[A_{\mathbf{u},k,t_k} - t_k^{-1}\mathbf{b}_{\mathbf{u},k,t_k}^T\mathbf{W}\mathbf{v}_k + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}t_k^2})\Big]$$
$$\times \Big[A_{\mathbf{v}_k,k,t_k} - t_k^{-1}\mathbf{b}_{\mathbf{v}_k,k,t_k}^T\mathbf{W}\mathbf{v}_k + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}t_k^2})\Big], \tag{B.23}$$

$$\widehat{d}_k = t_k + \mathbf{v}_k^T\mathbf{W}\mathbf{v}_k + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}|d_k|}). \tag{B.24}$$

B.8 Lemma 9

Lemma 9. *Under model* (10) *and Conditions 1–2, we have*

$$t_k \left[\mathbf{e}_i^T \widehat{\mathbf{v}}_k - \mathbf{v}_k(i)\right] = \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k + O_{\prec}(\frac{\alpha_n^2}{\sqrt{n}|t_k|} + \frac{1}{\sqrt{n}}). \tag{B.25}$$

The same conclusion also holds under model (6) *and Conditions 1 and 4.*

B.9 Lemma 10

Lemma 10. *Assume that $\widehat{K} = K$. Then under the mixed membership model (10) and Conditions 1–2, it holds uniformly over all i, j that*

$$\widehat{w}_{ij} = w_{ij} + O_{\prec}(\frac{\sqrt{\theta}}{\sqrt{n}}). \tag{B.26}$$

Under the degree-corrected mixed membership model (6), if Conditions 1 and 4–5 are satisfied, then it holds uniformly over all i, j that

$$\widehat{w}_{ij} = w_{ij} + O_{\prec}(\frac{\theta_{\max}}{\sqrt{n}}). \tag{B.27}$$

C Additional technical details

C.1 Lemma 11 and its proof

Lemma 11. *For any n-dimensional unit vectors $\mathbf{x}$, $\mathbf{y}$ and any positive integer r, we have*

$$\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^{2r} \le C_r(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})^{2r}, \tag{C.1}$$

where l is any positive integer and C_r is some positive constant determined only by r.

Proof. The main idea of the proof is similar to that for Lemma 4 in Fan et al. (2020), which is to count the number of nonzero terms in the expansion of $\mathbb{E}[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}]^{2r}$. It will be made clear that the nonzero terms in the expansion consist of terms such as w_{ij}^s with $s \ge 2$. In counting the nonzero terms, we will fix one index, say i, and vary the other index j which ranges from 1 to n. Note that for any $i = 1, \cdots, n$ and $s \ge 2$, we have $\sum_{j=1}^n \mathbb{E}|w_{ij}|^s \le \alpha_n^2$ since $|w_{ij}| \le 1$. Thus roughly speaking, counting the maximal moment of α_n is the crucial step in our proof.

Let $\mathbf{x} = (x_1, \cdots, x_n)^T$, $\mathbf{y} = (y_1, \cdots, y_n)^T$, and C_r be a positive constant depending only on r and whose value may change from line to line. Recall that $l, r \ge 1$ are two integers. We can expand $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y}-\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^{2r}$ to obtain the following expression

$$\begin{aligned}
&\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y}-\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^{2r}\\
&=\sum_{\substack{1\le i_1,\cdots,i_{l+1},i_{l+2},\cdots,i_{2l+2},\cdots,\\ i_{(2r-1)(l+1)+1},\cdots,i_{2r(l+1)}\le n}}\mathbb{E}\Big[\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}-\mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right)\times\cdots\\
&\quad\times\big(x_{i_{(2r-1)(l+1)+1}}w_{i_{(2r-1)(l+1)+1}i_{(2r-1)(l+1)+2}}w_{i_{(2r-1)(l+1)+2}i_{(2r-1)(l+1)+3}}\cdots w_{i_{2r(l+1)-1}i_{2r(l+1)}}y_{i_{2r(l+1)}}\\
&-\mathbb{E}x_{i_{(2r-1)(l+1)+1}}w_{i_{(2r-1)(l+1)+1}i_{(2r-1)(l+1)+2}}w_{i_{(2r-1)(l+1)+2}i_{(2r-1)(l+1)+3}}\cdots w_{i_{2r(l+1)-1}i_{2r(l+1)}}y_{i_{2r(l+1)}}\big)\Big].
\end{aligned}\tag{C.2}$$

Let $\mathbf{i}^{(j)}=(i_{(j-1)(l+1)+1},\cdots,i_{j(l+1)})$, $j=1,\cdots,2r$, be $2r$ vectors taking values in $\{1,\cdots,n\}^{l+1}$. Then for each $\mathbf{i}^{(j)}$, we define a graph $\mathcal{G}^{(j)}$ whose vertices represent distinct values of the components of $\mathbf{i}^{(j)}$. Each adjacent component of $\mathbf{i}^{(j)}$ is connected by an undirected edge in $\mathcal{G}^{(j)}$. It can be seen that for each j, $\mathcal{G}^{(j)}$ is a connected graph, which means that there exists some path connecting any two nodes in $\mathcal{G}^{(j)}$. For each fixed $i_1,\cdots,i_{l+1},\cdots,i_{(2r-1)(l+1)+1},\cdots,i_{2r(l+1)}$, consider the following term

$$\begin{aligned}
&\mathbb{E}\Big[\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}-\mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right)\times\cdots\\
&\quad\times\big(x_{i_{(2r-1)(l+1)+1}}w_{i_{(2r-1)(l+1)+1}i_{(2r-1)(l+1)+2}}w_{i_{(2r-1)(l+1)+2}i_{(2r-1)(l+1)+3}}\cdots w_{i_{2r(l+1)-1}i_{2r(l+1)}}y_{i_{2r(l+1)}}\\
&-\mathbb{E}x_{i_{(2r-1)(l+1)+1}}w_{i_{(2r-1)(l+1)+1}i_{(2r-1)(l+1)+2}}w_{i_{(2r-1)(l+1)+2}i_{(2r-1)(l+1)+3}}\cdots w_{i_{2r(l+1)-1}i_{2r(l+1)}}y_{i_{2r(l+1)}}\big)\Big],
\end{aligned}\tag{C.3}$$

which corresponds to graph $\mathcal{G}^{(1)}\cup\cdots\cup\mathcal{G}^{(2r)}$. If there exists one graph $\mathcal{G}^{(s)}$ that is unconnected to the remaining graphs $\mathcal{G}^{(j)}$, $j\neq s$, then the corresponding expectation in (C.3) is equal to zero. This shows that for any graph $\mathcal{G}^{(s)}$, there exists at least one connected $\mathcal{G}^{(s')}$ to ensure the nonzero expectation in (C.3). To analyze each nonzero (C.3), we next calculate how many distinct vertices are contained in the graph $\mathcal{G}^{(1)}\cup\cdots\cup\mathcal{G}^{(2r)}$.

Denote by $\mathfrak{S}(2r)$ the set of partitions of the integers $\{1,2,\cdots,2r\}$ and $\mathfrak{S}_{\ge 2}(2r)$ the subset of $\mathfrak{S}(2r)$ whose block sizes are at least two. To simplify the notation, define

$$\mathfrak{h}_j=x_{i_{(j-1)(l+1)+1}}w_{i_{(j-1)(l+1)+1}i_{(j-1)(l+1)+2}}w_{i_{(j-1)(l+1)+2}i_{(j-1)(l+1)+3}}\cdots w_{i_{j(l+1)-1}i_{j(l+1)}}y_{i_{j(l+1)}}.$$

Let $\mathcal{A}\in\mathfrak{S}_{\ge 2}(2r)$ be a partition of $\{1,2,\cdots,2r\}$ and $|\mathcal{A}|$ the number of groups in $\mathcal{A}$. We can further define $A_j\in\mathcal{A}$ as the jth group in $\mathcal{A}$ and $|A_j|$ as the number of integers in A_j. For example, let us consider $\mathcal{A}=\{\{1,2,3\},\{4,5,\cdots,2r\}\}$. Then we have $|\mathcal{A}|=2$, set $A_1=\{1,2,3\}\in\mathcal{A}$, and $|A_1|=3$. It is easy to see that there is a one-to-one correspondence between the partitions of $\{1,2,\cdots,2r\}$ and the graphs $\mathcal{G}^{(1)},\cdots,\mathcal{G}^{(2r)}$ such that $\mathcal{G}^{(s)}$ and $\mathcal{G}^{(s')}$ are connected if and only if s and s' belong to one group in the partition. For any $A_j\in\mathcal{A}\in\mathfrak{S}_{\ge 2}(2r)$, there are $|A_j|l$ edges in the graph $\bigcup_{w\in A_j}\mathcal{G}^{(j)}$ since for each integer $w\in A_j$, there is a chain containing l edges by $\mathfrak{h}_w$. Since $\mathbb{E}w_{ss'}=0$ for $s\neq s'$, in order to obtain a nonzero value of (C.3) each edge in $\bigcup_{w\in A_j}\mathcal{G}^{(j)}$ should have at least one additional copy. Thus for each nonzero (C.3), we have $[\frac{|A_j|l}{2}]$ distinct edges without self loops in $\bigcup_{w\in A_j}\mathcal{G}^{(j)}$. Since the graph $\bigcup_{w\in A_j}\mathcal{G}^{(j)}$ is connected, we can conclude that there are at most $[\frac{|A_j|l}{2}]+1$ distinct vertices in $\bigcup_{w\in A_j}\mathcal{G}^{(j)}$. Let $\mathcal{S}(\mathcal{A})$ be the collection of all choices of $\bigcup_{s=1}^{2r}\mathbf{i}^{(s)}$ such that

1). $\bigcup_{s=1}^{2r}\mathcal{G}^{(s)}$ has the same partition as $\mathcal{A}$ such that they are connected within the same group and unconnected between groups;

2). Within each group A_j, there are at most $[\frac{|A_j|l}{2}]$ distinct edges without self loops and $[\frac{|A_j|l}{2}]+1$ distinct vertices.

Similarly we can define $\mathcal{S}(A_j)$ since A_j can be regarded as a special partition of A_j with only one group. Summarizing the arguments above, (C.2) can be rewritten as

$$(C.2)=\sum_{\mathcal{A}\in\mathfrak{S}_{\geq 2}(2r)}\sum_{\bigcup_{s=1}^{2r}\mathbf{i}^{(s)}\in\mathcal{S}(\mathcal{A})}\prod_{j=1}^{|\mathcal{A}|}\Big[\mathbb{E}\prod_{\gamma\in A_j}(\mathfrak{h}_\gamma-\mathbb{E}\mathfrak{h}_\gamma)\Big]. \tag{C.4}$$

Let us further simplify $\mathbb{E}\prod_{\gamma\in A_j}(\mathfrak{h}_\gamma-\mathbb{E}\mathfrak{h}_\gamma)$. Let $\mathcal{B}_j$ be the set of partitions of A_j such that each partition contains exactly two groups. Without loss of generality, let $\mathcal{B}_j=\{b_{j_1},b_{j_2}\}$, where for any $w\in A_j$, we have $w\in b_{j_1}$ or $w\in b_{j_2}$. Then it holds that

$$|\mathbb{E}\prod_{\gamma\in A_j}(\mathfrak{h}_\gamma-\mathbb{E}\mathfrak{h}_\gamma)|\leq\sum_{\gamma\in\mathcal{B}_j}\mathbb{E}\Big|\prod_{\gamma\in b_{j_1}}\mathfrak{h}_\gamma\Big|\prod_{\gamma\in b_{j_2}}\Big|\mathbb{E}\mathfrak{h}_\gamma\Big|. \tag{C.5}$$

Observe that by definition, $\mathfrak{h}_\gamma$ is the product of some independent random variables, and $\mathfrak{h}_{\gamma_1}$ and $\mathfrak{h}_{\gamma_2}$ may share some dependency through factors $w_{ab}^{m_1}$ and $w_{ab}^{m_2}$, respectively, for some w_{ab} and nonnegative integers m_1 and m_2. Thus in light of the inequality

$$\mathbb{E}|w_{ab}|^{m_1}\mathbb{E}|w_{ab}|^{m_2}\leq\mathbb{E}|w_{ab}|^{m_1+m_2},$$

(C.5) can be bounded as

$$(C.5)\leq 2^{|A_j|}\mathbb{E}\Big|\prod_{\gamma\in A_j}\mathfrak{h}_\gamma\Big|. \tag{C.6}$$

By (C.6), we can deduce

$$\begin{aligned}(C.4)&\leq 2^{2r}\sum_{\mathcal{A}\in\mathfrak{S}_{\geq 2}(2r)}\sum_{\bigcup_{s=1}^{2r}\mathbf{i}^{(s)}\in\mathcal{S}(\mathcal{A})}\prod_{j=1}^{|\mathcal{A}|}\mathbb{E}\Big|\prod_{\gamma\in A_j}\mathfrak{h}_\gamma\Big|\\&\leq 2^{2r}\sum_{\mathcal{A}\in\mathfrak{S}_{\geq 2}(2r)}\prod_{j=1}^{|\mathcal{A}|}\Big(\sum_{\mathbf{i}^{(s)}\in\mathcal{S}(A_j)}\mathbb{E}\Big|\prod_{\gamma\in A_j}\mathfrak{h}_\gamma\Big|\Big).\end{aligned} \tag{C.7}$$

Thus it suffices to show that

$$\sum_{\mathbf{i}^{(s)}\in\mathcal{S}(A_j)}\mathbb{E}\Big|\prod_{\gamma\in A_j}\mathfrak{h}_\gamma\Big| = C_{|A_j|}(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})^{|A_j|},$$

using the fact that $\sum_{j=1}^{|\mathcal{A}|}|A_j| = 2r$. Without loss of generality, we prove the most difficult case of $|\mathcal{A}| = 1$, that is, there is only one connected chain which is $A = \{1, 2, \cdots, 2r\}$. It has the most components in the chain $\prod_{\gamma\in A}\mathfrak{h}_\gamma$. Other cases with smaller $|A|$ can be shown in the same way. Using the same arguments as those for (C.4), we have the basic property for this chain that there are at most $[\frac{|A|l}{2}] + 1 = rl + 1$ distinct vertices and rl distinct edges without self loops.

To facilitate our technical presentation, let us introduce some additional notation. Denote by $\psi(r, l)$ the set of partitions of the edges $\{(i_s, i_{s+1}), 1 \le s \le 2rl, i_s \ne i_{s+1}\}$ and $\psi_{\ge 2}(r, l)$ the subset of $\psi(r, l)$ whose blocks have size at least two. Let $\widetilde{\mathbf{i}} = \bigcup_{s=1}^{2r}\mathbf{i}^{(s)}$ and $P(\widetilde{\mathbf{i}}) \in \psi_{\ge 2}(2l+2)$ be the partition of $\{(i_s, i_{s+1}), 1 \le s \le 2rl, i_s \ne i_{s+1}\}$ that is associated with the equivalence relation $(i_{s_1}, i_{s_1+1}) \sim (i_{s_2}, i_{s_2+1})$, which is defined as if and only if $(i_{s_1}, i_{s_1+1}) = (i_{s_2}, i_{s_2+1})$ or $(i_{s_1}, i_{s_1+1}) = (i_{s_2+1}, i_{s_2})$. Denote by $|P(\widetilde{\mathbf{i}})| = m$ the number of groups in the partition $P(\widetilde{\mathbf{i}})$ such that the edges are equivalent within each group. We further denote the distinct edges in the partition $P(\widetilde{\mathbf{i}})$ as $(s_1, s_2), (s_3, s_4), \cdots, (s_{2m-1}, s_{2m})$ and the corresponding counts in each group as $r_1, \cdots, r_m$, and define $\widetilde{\mathbf{s}} = (s_1, s_2, \cdots, s_{2m})$. For the vertices, let $\phi(2m)$ be the set of partitions of $\{1, 2, \cdots, 2m\}$ and $Q(\widetilde{\mathbf{s}}) \in \phi(2m)$ the partition that is associated with the equivalence relation $a \sim b$, which is defined as if and only if $s_a = s_b$. Note that $s_{2j-1} \ne s_{2j}$ by the definition of the partition. By $|w_{aa}| \le 1$, we can deduce

$$\begin{aligned}
&\sum_{\mathbf{i}^{(s)}\in\mathcal{S}(A)}\mathbb{E}\Big|\prod_{\gamma\in A}\mathfrak{h}_\gamma\Big| = \sum_{\mathbf{i}^{(s)}\in\mathcal{S}(A)}\mathbb{E}\Big|\prod_{\gamma=1}^{2r}\mathfrak{h}_\gamma\Big| \\
&\le \sum_{\substack{1\le|P(\widetilde{\mathbf{i}})|=m\le rl\\ P(\widetilde{\mathbf{i}})\in\psi_{\ge 2}(2l+2)}}\ \sum_{\substack{\widetilde{\mathbf{i}}\text{ with partition } P(\widetilde{\mathbf{i}})\\ r_1,\cdots,r_m\ge 2}}\ \sum_{Q(\widetilde{\mathbf{s}})\in\phi(2m)}\ \sum_{\substack{\widetilde{\mathbf{s}}\text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}}\ \prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|) \\
&\times\prod_{j=1}^{m}\mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}.
\end{aligned} \tag{C.8}$$

Denote by $\mathcal{F}_{\widetilde{\mathbf{s}}}$ the graph constructed by the edges of $\widetilde{\mathbf{s}}$. Since the edges in $\widetilde{\mathbf{s}}$ are the same as those of the edges in $\bigcup_{s=1}^{2r}\mathcal{G}^{(s)}$ with the structure $\mathcal{S}(A)$, we can see that $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is also a connected graph. In view of (C.8), putting term $|x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}|$ aside we need to analyze the summation

$$\sum_{\substack{\widetilde{\mathbf{s}}\text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}}\ \prod_{j=1}^{m}\mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}.$$

If index s_{2k-1} satisfies that $s_{2k-1} \ne s$ for all $s \in \{s_1, \cdots, s_{2m}\} \setminus \{s_{2k-1}\}$, that is, index s_{2k-1}

appears only in one $w_{s_{2j-1}s_{2j}}$, we call s_{2k-1} a single index (or single vertex). If there exists some single index s_{2k-1}, then it holds that

$$\begin{aligned}&\sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}\\ &\le \sum_{\substack{\widetilde{\mathbf{s}}\backslash\{s_{2k-1}\} \text{ with partition } Q(\widetilde{\mathbf{s}}\backslash\{s_{2k-1}\})\\ 1\le s_1,\cdots,s_{2k-2},s_{2k+2},s_{2m}\le n\\ s_{2k}=s_j \text{ for some } 1\le j\le 2m}} \prod_{j=1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \sum_{s_{2k-1}=1}^{n} \mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}. \qquad \text{(C.9)}\end{aligned}$$

Note that since graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is connected and index s_{2k-1} is single, there exists some j such that $s_j = s_{2k}$, which means that in the summation $\sum_{s_{2k-1}=1}^{n} \mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}$, index s_{2k} is fixed. Then it follows from the definition of α_n, $|w_{ij}| \le 1$, and $r_k \ge 2$ that

$$\sum_{s_{2k-1}=1}^{n} \mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k} \le \alpha_n^2.$$

After taking the summation over index s_{2k-1}, we can see that there is one less edge in $\mathcal{F}(\widetilde{\mathbf{s}})$. That is, by taking the summation above we will have one additional α_n^2 in the upper bound while removing one edge from graph $\mathcal{F}(\widetilde{\mathbf{s}})$. For the single index s_{2k}, we also have the same bound. If s_{2k_1-1} is not a single index, without loss of generality we assume that $s_{2k_1-1} = s_{2k-1}$. Then this vertex s_{2k-1} needs some delicate analysis. By the assumption of $|w_{ij}| \le 1$, we have

$$\mathbb{E}|w_{2k-1,2k}|^{r_k}|w_{2k_1-1,2k_1}|^{r_{k_1}} \le \frac{\mathbb{E}|w_{2k-1,2k}|^{r_k} + \mathbb{E}|w_{2k_1-1,2k_1}|^{r_{k_1}}}{2}.$$

Then it holds that

$$\begin{aligned}&\sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}\\ &\le \frac{1}{2} \sum_{\substack{\widetilde{\mathbf{s}}\backslash(s_{2k-1},s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{s}}\backslash(s_{2k-1},s_{2k_1-1}))\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1,\, j\neq k}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}\\ &\quad + \frac{1}{2} \sum_{\substack{\widetilde{\mathbf{s}}\backslash(s_{2k-1},s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{s}}\backslash(s_{2k-1},s_{2k_1-1}))\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1,\, j\neq k_1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}. \qquad \text{(C.10)}\end{aligned}$$

Note that since $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is a connected graph, if we delete either edge (s_{2k-1}, s_{2k}) or edge (s_{2k_1-1}, s_{2k_1}) from graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$, the resulting graph is also connected. Then the two summations on the right hand side of (C.10) can be reduced to the case in (C.9) for the graph with edge (s_{2k-1}, s_{2k}) or (s_{2k_1-1}, s_{2k_1}) removed, since s_{2k-1} or s_{2k_1-1} is a single index in the subgraph. Similar to (C.9), after taking the summation over index s_{2k-1} or s_{2k_1-1} there are two less edges in graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$ and thus we now obtain $2\alpha_n^2$ in the upper bound.

For the general case when there are m_1 vertices belonging to the same group, without loss of generality we denote them as $w_{ab_1}, \cdots, w_{ab_{m_1}}$. If for any k graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is still connected after deleting edges $(a, b_1), \cdots, (a, b_{k-1}), (a, b_{k+1}), \cdots, (a, b_{m_1})$, then we repeat the process in (C.10) to obtain a new connected graph by deleting $k-1$ edges in $w_{ab_1}, \cdots, w_{ab_{m_1}}$ and thus obtain $k\alpha_n^2$ in the upper bound. Motivated by the key observations above, we carry out an iterative process in calculating the upper bound as follows.

(1) If there exists some single index in $\widetilde{\mathbf{s}}$, using (C.9) we can calculate the summation over such an index and then delete the edge associated with this vertex in $\mathcal{F}_{\widetilde{\mathbf{s}}}$. The corresponding vertices associated with this edge are also deleted. For simplicity, we also denote the new graph as $\mathcal{F}_{\widetilde{\mathbf{s}}}$. In this step, we obtain α_n^2 in the upper bound.

(2) Repeat (1) until there is no single index in graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$.

(3) Suppose there exists some index associated with k edges such that graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is still connected after deleting any $k-1$ edges. Without loss of generality, let us consider the case of $k = 2$. Then we can apply (C.9) to obtain α_n^2 in the upper bound. Moreover, we delete k edges associated with this vertex in $\mathcal{F}_{\widetilde{\mathbf{s}}}$.

(4) Repeat (3) until there is no such index.

(5) If there still exists some single index, go back to (1). Otherwise stop the iteration.

Completing the graph modification process mentioned above, we can obtain a final graph $\mathbf{Q}$ that enjoys the following properties:

i) Each edge does not contain any single index;

ii) Deleting any vertex makes the graph disconnected.

Let $\mathbf{S_Q}$ be the spanning tree of graph $\mathbf{Q}$, which is defined as the subgraph of $\mathbf{Q}$ with the minimum possible number of edges. Since $\mathbf{S_Q}$ is a subgraph of $\mathbf{Q}$, it also satisfies property ii) above. Assume that $\mathbf{S_Q}$ contains p edges. Then the number of vertices in $\mathbf{S_Q}$ is $p+1$. Denote by $q_1, \cdots, q_{p+1}$ the vertices of $\mathbf{S_Q}$ and $\deg(q_i)$ the degree of vertex q_i. Then by the degree sum formula, we have $\sum_{i=1}^{p+1} \deg(q_i) = 2p$. As a result, the spanning tree has at least two vertices with degree one and thus there exists a subgraph of $\mathbf{S_Q}$ without either of the vertices that is connected. This will result in a contradiction with property ii) above unless the number of vertices in graph $\mathbf{Q}$ is exactly one. Since l is a bounded constant, the numbers of partitions $P(\widetilde{\mathbf{i}})$ and $Q(\widetilde{\mathbf{s}})$ are also bounded. It follows that

$$(C.8) \leq C_r d_{\mathbf{x}}^{2r} d_{\mathbf{y}}^{2r} \sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}, \tag{C.11}$$

where $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$, $d_{\mathbf{y}} = \|\mathbf{x}\|_\infty$, and C_r is some positive constant determined by l. Combining these arguments above and noticing that there are at most l distinct edges in graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$, we can obtain

$$\begin{aligned}(C.11) &\leq C_r d_{\mathbf{x}}^{2r} d_{\mathbf{y}}^{2r} \alpha_n^{2rl-2} \sum_{1\leq s_{2k_0-1}, s_{2k_0}\leq n,\, (s_{2k_0-1}, s_{2k_0})=\mathbf{Q}} \mathbb{E}|w_{s_{2k_0-1}s_{2k_0}}|^{r_{k_0}} \\ &\leq C_r d_{\mathbf{x}}^{2r} d_{\mathbf{y}}^{2r} \alpha_n^{2rl} n. \end{aligned} \tag{C.12}$$

Therefore, we have established a simple upper bound of $C_r d_{\mathbf{x}}^{2r} d_{\mathbf{y}}^{2r} \alpha_n^{2rl} n$.

In fact, we can improve the aforementioned upper bound to $C_r \alpha_n^{r(l-1)}$. Note that the process mentioned above did not utilize the condition that both $\mathbf{x}$ and $\mathbf{y}$ are unit vectors, that is, $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. Since term $\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)$ is involved in (C.8), we can analyze them together with random variables w_{ij}. First, we need to deal with some distinct lower indices with low moments in $\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)$. If there are two distinct lower indices, without loss of generality denoted them as i_s and $i_{s'}$ and then the corresponding entries are x_{i_s} (or y_{i_s}) and $y_{i_{s'}}$ (or $x_{i_{s'}}$). Moreover, there are only one x_{i_s} and $y_{i_{s'}}$ involved in $\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)$. Without loss of generality, let us assume that $s = 1$ and $s' = l + 1$. Then it holds that

$$\begin{aligned}\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|) &= |x_{i_1}||y_{i_{l+1}}| \prod_{j=2}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|) \\ &\leq \frac{x_{i_1}^2}{2} \prod_{j=2}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|) + \frac{y_{i_{l+1}}^2}{2} \prod_{j=2}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|). \end{aligned} \tag{C.13}$$

That is, if we have two lower indices and each index appears only once in the product above, we can use (C.13) to increase the moment of x_{i_s}(or $y_{i_{s'}}$) and delete the other one. For (C.13), it is equivalent for us to consider the case when the lower index $i_1 = i_{l+1}$. Repeating the procedure (C.13), finally we can obtain a product $\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)$ with the following properties:

1). Except for one vertex i_{s_0}, for each i_s with $s \neq s_0$ there exists some $i_{s'}$ such that $i_s = i_{s'}$ with $s \neq s'$.

2). Except for one vertex i_{s_0}, for each i_s with $s \neq s_0$ the term $x_{i_s}^{m_1} y_{i_s}^{m_2}$ involved in $\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)$ satisfies the condition that $m_1 + m_2 \geq 2$. Moreover, at least one of m_1 and m_2 is larger than one.

By the properties above, let us denote by $\Upsilon(2r)$ the set of partitions of the vertices $\{i_{(j-1)(l+1)+1}, i_{j(l+1)}, j = 1, \cdots, 2r\}$ such that except for one group, the remaining groups in Υ with $\Upsilon \in \Upsilon(2r)$ have blocks with size at least two. There are three different cases to consider.

Case 1). All the groups in Υ have block size two. Then it follows that

$$|\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)| = \prod_{k=1}^{|\Upsilon|}|x|_{i_s}^{m_{1k}}|y|_{i_k}^{m_{2k}}, \tag{C.14}$$

where $m_{1k}+m_{2k}=2$. In fact, by the second property of Υ above, $m_{1k}=0$ or $m_{2k}=0$. Without loss of generality, we assume that $m_{2k}=0$. Then we need only to consider the equation

$$|\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)| = \prod_{k=1}^{|\Upsilon|}|x|_{i_k}^{2}.$$

Then by (C.8), it remains to bound

$$\sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{k=1}^{|\Upsilon|}|x|_{i_k}^2 \prod_{j=1}^{m}\mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}. \tag{C.15}$$

To simplify the presentation, assume without loss of generality that $i_k = s_k$, $k = 1,\cdots,|\Upsilon|$. Then the summation in (C.15) becomes

$$\sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1}^{|\Upsilon|}|x|_{s_j}^2 \prod_{j=1}^{m}\mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}.$$

By repeating the iterative process (1)–(5) mentioned before, we can bound the summation for fixed $s_2,\cdots,s_{|\Upsilon|}$ and obtain an alternative upper bound

$$\sum_{s_1=1}^{n} x_{s_1}^2\mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j} \le \sum_{s_1=1}^{n} x_{s_1}^2 = 1$$

since $\mathbf{x}$ is a unit vector. Thus for this step of the iteration, we obtain term one instead of α_n^2 in the upper bound. Repeat this step until there is only $x_{s_{|\Upsilon|}}^2$ left. Since the graph is always connected during the iteration process, there exists another vertex b such that $w_{s_{|\Upsilon|}b}$ is involved in (C.15). For index $s_{|\Upsilon|}$, we do not delete the edges containing $s_{|\Upsilon|}$ in the graph during the iterative process (1)–(5). Then after the iteration stops, the final graph $\mathbf{Q}$ satisfies properties i) and ii) defined earlier except for vertex $s_{|\Upsilon|}$. Since there are at least two vertices with degree one in $\mathbf{S_Q}$, we will also reach a contradiction unless the number of vertices in graph $\mathbf{Q}$ is exactly one. By (C.14), it holds that $2|\Upsilon| = 4r$. As a result, we can obtain the upper bound

$$(C.8) \le C_r\alpha_n^{2rl-2|\Upsilon|}\sum_{1\le s_2,b\le n,\,(s_2,b)=\mathbf{Q}}\mathbb{E}x_{s_{|\Upsilon|}}^2|w_{s_{|\Upsilon|}b}|^r \le C_r\alpha_n^{2rl-2r} \tag{C.16}$$

with C_r some positive constant. Therefore, the improved bound $C_r\alpha_n^{2r(l-1)}$ is shown for this case.

Case 2). All the groups in Υ have block size at least two and there is at least one block with size larger than two. Then it follows that

$$|\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)| = \prod_{k=1}^{|\Upsilon|}|x|_{i_s}^{m_{1k}}|y|_{i_k}^{m_{2k}}.$$

Since $m_{1k}+m_{2k}\geq 2$ by the second property of Υ above, define the nonnegative integer $r_1=\sum_{k=1}^{|\Upsilon|}(m_{1k}+m_{2k}-2)$. There are at most $[\frac{2rl+2-r_1}{2}]$ distinct vertices in the graph $\mathcal{F}_{\widetilde{\mathbf{s}}}$ and at most $[\frac{2rl+2-r_1}{2}]-1$ distinct edges. Similar to Case 1 with less distinct edges, we have

$$(C.8)\leq C\alpha_n^{2[\frac{2rl+2-r_1}{2}]-2|\Upsilon|-2}\sum_{1\leq s_1,b\leq n,\,(s_1,b)=\mathbf{Q}}\mathbb{E}x_{s_1}^2|w_{s_1b}|^r\leq C\alpha_n^{2[\frac{2rl+2-r_1}{2}]-2|\Upsilon|}. \tag{C.17}$$

By the definition of r_1 and $\sum_{k=1}^{|\Upsilon|}(m_{1k}+m_{2k})=4r$, it holds that

$$r_1+2|\Upsilon|=4r.$$

Thus r_1 is an even number and $2[\frac{2rl+2-r_1}{2}]-2|\Upsilon|=2rl-r_1-2|\Upsilon|+2\leq 2rl-2r$. The improved bound $C_r\alpha_n^{2r(l-1)}$ is also shown for this case.

Case 3). Except for one index i_{k_0}, the other groups in Υ have block size at least two. Let us define $r_1'=\sum_{k=1,k\neq k_0}^{|\Upsilon|}(m_{1k}+m_{2k}-2)$. There are at most $[\frac{2rl+2-r_1'}{2}]$ distinct vertices and at most $[\frac{2rl+2-r_1'}{2}]-1$ distinct edges. For the parameter $|x_{i_{k_0}}|$ (or $|y_{i_{k_0}}|$), we can bound it by one since $\mathbf{x}$ and $\mathbf{y}$ are unit vectors. Then similar to Case 2, we can deduce

$$(C.8)\leq C\alpha_n^{2[\frac{2rl+2-r_1'}{2}]-2|\Upsilon|}\sum_{1\leq s_1,b\leq n,\,(s_1,b)=\mathbf{Q}}\mathbb{E}x_{s_1}^2|w_{s_1b}|^r\leq C\alpha_n^{2[\frac{2rl+2-r_1'}{2}]-2|\Upsilon|+2}. \tag{C.18}$$

By the definition of r_1' in this case, it holds that

$$r_1'+2|\Upsilon|=4r+1.$$

Then r_1' is an odd number and thus

$$2[\frac{2rl+2-r_1'}{2}]-2|\Upsilon|+2=2rl-r_1-2|\Upsilon|+3\leq 2rl-2r.$$

Summarizing the arguments above, for this case we can also obtain the desired bound $C_r\alpha_n^{2r(l-1)}$.

In addition, we can also improve the upper bound to $C_r(\min\{d_{\mathbf{x}}^{2r}\alpha_n^{2rl},d_{\mathbf{y}}^{2r}\alpha_n^{2rl}\})$. The technical arguments for this refinement are similar to those for the improvement to order

$C_r\alpha_n^{2r(l-1)}$ above. As an example, we can bound the components of $\mathbf{y}$ by $d_{\mathbf{y}} = \|\mathbf{y}\|_\infty$, which leads to $|\prod_{j=1}^{2r}(|x_{i_{(j-1)(l+1)+1}}||y_{i_{j(l+1)}}|)| \le d_{\mathbf{y}}^{2r}|\prod_{j=1}^{2r}|x_{i_{(j-1)(l+1)+1}}|$. Then the analysis becomes similar to the three cases above. The only difference is that $\sum_{k=1}^{|\Upsilon|} m_{1k} = 2r$ instead of $\sum_{k=1}^{|\Upsilon|}(m_{1k}+m_{2k}) = 4r$. For this case, we have

$$(C.8) \le C d_{\mathbf{y}}^{2r}\alpha_n^{2rl-2|\Upsilon|} \sum_{1\le s_2,b\le n,\,(s_2,b)=\mathbf{Q}} \mathbb{E}x_{s_1}^2|w_{s_1b}|^r \le C_r d_{\mathbf{y}}^{2r}\alpha_n^{2rl}. \tag{C.19}$$

Thus we can obtain the claimed upper bound $C_r(\min\{d_{\mathbf{x}}^{2r}\alpha_n^{2rl}, d_{\mathbf{y}}^{2r}\alpha_n^{2rl}\})$. Therefore, combining the two aforementioned improved bounds yields the desired upper bound of

$$C_r(\min\{\alpha_n^{2r(l-1)}, d_{\mathbf{x}}^{2r}\alpha_n^{2rl}, d_{\mathbf{y}}^{2r}\alpha_n^{2rl}\}),$$

which completes the proof of Lemma 11.

C.2 Corollary 3 and its proof

Lemma 11 ensures the following corollary immediately.

Corollary 3. *Under the conditions of Lemma 11, it holds that for any positive constants a and b, there exists some $n_0(a,b) > 0$ such that*

$$\sup_{\|\mathbf{x}\|=\|\mathbf{y}\|=1} \mathbb{P}\left(\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} \ge n^a \min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\}\right) \le n^{-b} \tag{C.20}$$

for any $n \ge n_0(a,b)$ and $l \ge 1$. Moreover, we have

$$\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_{\prec}(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\}). \tag{C.21}$$

Proof. It suffices to show (C.20) because then (C.21) follows from the definition. For any positive constants a and b, there exists some integer r such that $2ar \ge b+1$. By the Chebyshev inequality, it holds that

$$\sup_{\|\mathbf{x}\|=\|\mathbf{y}\|=1} \mathbb{P}(|\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}| \ge n^a \min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$$

$$\le \sup_{\|\mathbf{x}\|=\|\mathbf{y}\|=1} \frac{\mathbb{E}(\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y})^{2r}}{n^{2ar}(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})^{2r}} \le \frac{C_r}{n^{b+1}},$$

which can be further bounded by n^{-b} as long as $n \ge C_r$. It is seen that C_r is determined completely by a and b. This concludes the proof of Corollary 3.

C.3 Lemma 12 and its proof

Lemma 12. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, we have*

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(\alpha_n^l), \tag{C.22}$$

where $l \geq 2$ is a positive integer. Furthermore, if the number of nonzero components of $\mathbf{x}$ is bounded, then it holds that

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(\alpha_n^l d_{\mathbf{y}}), \tag{C.23}$$

where $d_{\mathbf{y}} = \|\mathbf{y}\|_\infty$.

Proof. The result in (C.22) follows directly from Lemma 5 of Fan et al. (2020). Thus it remains to show (C.23). The main idea of the proof is similar to that for the proof of Lemma 11. Denote by $\mathfrak{C}$ the set of positions of the nonzero components of $\mathbf{x}$. Then we have

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = \sum_{\substack{i_1\in\mathfrak{C}, 1\leq i_2,\cdots,i_{l+1}\leq n\\ i_s\neq i_{s+1}}} \mathbb{E}\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right). \tag{C.24}$$

Note that the cardinality of set $\mathfrak{C}$ is bounded. Thus it suffices to show that for fixed i_1, we have

$$\sum_{\substack{1\leq i_2,\cdots,i_{l+1}\leq n\\ i_s\neq i_{s+1}}} \mathbb{E}\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right) = O(d_{\mathbf{y}}\alpha_n^l). \tag{C.25}$$

By the definition of graph $\mathcal{G}^{(1)}$ in the proof of Lemma 11, we can also get a similar expression as (C.6) that

$$\begin{aligned}&|(C.24)|\\ &\leq d_{\mathbf{y}} \sum_{\mathcal{G}^{(1)} \text{ with at most } [l/2] \text{ distinct edges without self loops and } [l/2]+1 \text{ distinct vertices, } i_1 \text{ is fixed}} \mathbb{E}\big|w_{i_1i_2}w_{i_2i_3}\\ &\quad\cdots w_{i_li_{l+1}}\big|.\end{aligned} \tag{C.26}$$

Using similar arguments for bounding the order of the summation through the iterative process as those for (C.11)–(C.12) in the proof of Lemma 11, we can obtain a similar bound

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} \leq Cd_{\mathbf{y}}\alpha_n^{l-2}\sum_{i_{k_0}=1}^n \mathbb{E}|w_{i_1i_{k_0}}|^{r_0} \leq Cd_{\mathbf{y}}\alpha_n^l \tag{C.27}$$

with $r_0 \geq 2$. Here we do not remove the lower index i_1 during the iteration procedure. The additional factor n on the right hand side of (C.12) can be eliminated since i_1 is fixed. This completes the proof of Lemma 12.

C.4 Lemma 13 and its proof

Lemma 13. *Assume that* $\xi_1 = O_\prec(\zeta), \cdots, \xi_m = O_\prec(\zeta)$ *with* $m = \lfloor n^c \rfloor$ *and* c *some positive constant. If*

$$\mathbb{P}\left[|\xi_i| > n^a|\zeta|\right] \le n^{-b} \tag{C.28}$$

uniformly for ξ_i*,* $i = 1, \cdots, m$*, and any positive constants* a*,*b *with* $n \ge n_0(a, b)$*, then for any positive random variables* $X_1, \cdots, X_m$*, we have*

$$\sum_{i=1}^m X_i\xi_i = O_\prec\Big(\sum_{i=1}^m X_i\zeta\Big).$$

Proof. For any positive constants a and b, let $b_1 = c + b$. By (C.28), it holds that

$$\mathbb{P}\left[|\xi_i| > n^a|\zeta|\right] \le n^{-b_1}$$

for all $n \ge n_0(a, b_1)$, where $n_0(a, b_1)$ is determined completely by a and b_1. Then we have

$$\mathbb{P}\left[|\sum_{i=1}^m X_i\xi_i| > n^a|\zeta|\sum_{i=1}^m X_i\right] \le \sum_{i=1}^m \mathbb{P}\left[|\xi_i| > n^a|\zeta|\right] \le n^{-b}$$

for large enough $n \ge n_0(a, b_1)$. Since $b_1 = c + b$ and c is fixed, the constant $n_0(a, b_1)$ is determined essentially by a and b. This concludes the proof of Lemma 13.

C.5 Lemma 14 and its proof

Lemma 14. *For any positive constant* $\mathfrak{L}$*, it holds that*

$$\mathbb{P}(\|\mathbf{W}\| \ge \alpha_n \log n) \le n^{-\mathfrak{L}}$$

for all sufficiently large n*.*

Proof. The conclusion of Lemma 14 follows directly from Theorem 6.2 of Tropp (2012). We can also prove it by (B.1) and the inequality with $c\sqrt{\log n}\alpha_n - 1$ replaced by $\alpha_n \log n$ in (B.2).

C.6 Lemma 15

Lemma 15 (Fan et al. (2020))**.** *There exists a unique solution* $z = t_k$ *to equation* (9) *on the interval* $[a_k, b_k]$*, and thus* t_k*'s are well defined. In addition, for each* $k = 1, \cdots, K$*, we have* $t_k/d_k \to 1$ *as* $n \to \infty$*.*

D Sufficient conditions for Condition 3

D.1 Lemma 16 and its proof

Lemma 16. *Under Conditions 1–2, if $\theta < 1$ and $\min_{1\leq i,j\leq K} \mathbf{P}_{ij} \geq c$ for some positive constant c, then Condition 3 holds.*

Proof. The key step of the proof is to calculate $\mathrm{cov}[(\mathbf{e}_i-\mathbf{e}_j)^T\mathbf{W}\mathbf{V}]$. Without loss of generality, let us assume that $(i,j) = (1,2)$. Note that the main difference between the null and alternative hypotheses is that the mean value of $(\mathbf{e}_1 - \mathbf{e}_2)^T\mathbb{E}\mathbf{W}$ is 0 under the former and is $(\mathbb{E}w_{1,1}, -\mathbb{E}w_{2,2}, 0, \cdots, 0)^T$, which may be nonzero, under the latter. However, since the main idea of the proof applies to both cases, we will provide only the technical details under the null hypothesis.

First, some direct calculations show that

$$\begin{aligned}\theta^{-1}\mathbf{D}\mathbf{\Sigma}_1\mathbf{D} &= \theta^{-1}\mathrm{cov}[(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{V}]\\ &= \theta^{-1}\mathbf{V}^T\mathbb{E}(\mathbf{W}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W})\mathbf{V}\\ &= \theta^{-1}\mathbf{V}^T\mathbf{Q}\mathbf{V},\end{aligned} \tag{D.1}$$

where $\mathbf{Q} = \mathrm{diag}(\mathbb{E}(\mathbf{w}_{i1} - \mathbf{w}_{j1})^2, \cdots, \mathbb{E}(\mathbf{w}_{in} - \mathbf{w}_{jn})^2) + \mathbb{E}\mathbf{w}_{ij}^2\mathbf{e}_i\mathbf{e}_j^T + \mathbb{E}\mathbf{w}_{ij}^2\mathbf{e}_j\mathbf{e}_i^T$. By the assumptions that $\theta < 1$ and $\min_{1\leq i,j\leq K} \mathbf{P}_{ij} \geq c$, we see that the entries of the mean matrix $\mathbf{H} = (h_{ij})$ are bounded from below by $c\theta$ and from above by θ. Since $\mathbb{E}w_{ij}^2 \sim h_{ij}$ and w_{ik} and w_{jk} are independent for $i \neq j$, it holds that

$$\theta\mathbf{I} \lesssim \mathrm{diag}(\mathbb{E}(w_{i1} - w_{j1})^2, \cdots, \mathbb{E}(w_{in} - w_{jn})^2) \lesssim \theta\mathbf{I}. \tag{D.2}$$

Then it follows from (D.2) that

$$\mathbf{I} \lesssim \theta^{-1}\mathbf{V}^T\mathrm{diag}(\mathbb{E}(w_{i1} - w_{j1})^2, \cdots, \mathbb{E}(w_{in} - w_{jn})^2)\mathbf{V} \lesssim \mathbf{I}. \tag{D.3}$$

Since $\mathbf{\Sigma}_1 \in \mathbb{R}^{K\times K}$ with K a finite integer, we can deduce that

$$\|\theta^{-1}\mathbf{V}^T(\mathbb{E}w_{ij}^2\mathbf{e}_i\mathbf{e}_j^T + \mathbb{E}w_{ij}^2\mathbf{e}_j\mathbf{e}_i^T)\mathbf{V}\| \lesssim \frac{1}{n}. \tag{D.4}$$

Therefore, combining (D.1)–(D.4), we can obtain the desired conclusion under the null hypothesis. This completes the proof of Lemma 16.

E Uniform convergence

Theorem 6. *Assume that the null hypotheses $H_{0,ij} : \boldsymbol{\pi}_i = \boldsymbol{\pi}_j$ hold for all $1 \leq i \neq j \leq n$. Then*

1) Under Conditions 1–3 and the mixed membership model (10), we have for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \sup_{1\le i\ne j\le n} |\mathbb{P}(T_{ij} \le x) - \mathbb{P}(\chi^2_K \le x)| = 0. \tag{E.1}$$

2) Under Conditions 1 and 4–7 and DCMM (6), we have for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \sup_{1\le i\ne j\le n} |\mathbb{P}(G_{ij} \le x) - \mathbb{P}(\chi^2_{K-1} \le x)| = 0. \tag{E.2}$$

Proof. We provide the detailed proof only for (E.1) since the proof of (E.2) is almost identical. Recall that $T_{ij} = \|\mathbf{\Sigma}_1^{-1/2}(\widehat{\mathbf{V}}(i) - \widehat{\mathbf{V}}(j))\|^2$. Let us investigate the asymptotic behavior of random vector $\mathbf{\Sigma}_1^{-1/2}(\widehat{\mathbf{V}}(i) - \widehat{\mathbf{V}}(j))$. Checking the proof of Theorem 1 in Section A.1 carefully, we can see that there exists some positive constant ϵ such that

$$\begin{aligned}
&\mathbf{\Sigma}_1^{-1/2}(\widehat{\mathbf{V}}(i) - \widehat{\mathbf{V}}(j)) \\
&= \mathbf{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T + O_{\prec}(n^{-\epsilon}),
\end{aligned} \tag{E.3}$$

where the $o_p(1)$ term in (A.8) is replaced by $O_{\prec}(n^{-\epsilon})$. By (E.3) and the continuity of the standard multivariate Gaussian distribution, it suffices to show that for any convex set $\mathbf{S} \subset \mathbb{R}^K$, we have

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{i\ne j}\left|\mathbb{P}\left(\mathbf{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T \in \mathbf{S}\right) - \mathbb{P}\left(\mathbf{x}_K \in \mathbf{S}\right)\right| \\
&\quad = 0,
\end{aligned} \tag{E.4}$$

where $\mathbf{x}_K \sim N(\mathbf{0}, \mathbf{I}_K)$.

For an application of Theorem 1.1 in Raič (2019), we need to rewrite

$$\mathbf{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_1}{t_1/d_1}, \cdots, \frac{(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T$$

as the sum of independent random vectors. Indeed, some direct calculations yield

$$
\begin{aligned}
&\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(\mathbf{e}_i-\mathbf{e}_j)^T\mathbf{W}\mathbf{v}_1}{t_1/d_1},\cdots,\frac{(\mathbf{e}_i-\mathbf{e}_j)^T\mathbf{W}\mathbf{v}_K}{t_K/d_K}\right)^T \\
&=\sum_{l=1}^n \boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{(w_{il}-w_{jl})\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{(w_{il}-w_{jl})\mathbf{v}_{Kl}}{t_K/d_K}\right)^T \\
&=\sum_{l\neq i,j}\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T \\
&\quad+\sum_{l\in\{i,j\}}\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T, \qquad \text{(E.5)}
\end{aligned}
$$

where the first term in the last step is the sum of independent random vectors. Then it follows from Lemma 6 and Condition 3 that

$$
\sum_{l\in\{i,j\}}\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T = O(\frac{1}{\sqrt{n\theta}}).
$$

Combining this with (E.4) and (E.5), we see that it remains to show that

$$
\begin{aligned}
&\lim_{n\to\infty}\sup_{i\neq j}\left|\mathbb{P}\left(\sum_{l\neq i,j}\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T\in\mathbf{S}\right)-\mathbb{P}(\mathbf{x}_K\in\mathbf{S})\right| \\
&=0. \qquad \text{(E.6)}
\end{aligned}
$$

From Theorem 1.1 in Raič (2019), Condition 3, and Lemma 6, we can deduce that for any fixed i,j, there exists some positive constant C (independent of i,j) such that

$$
\begin{aligned}
&\left|\mathbb{P}\left(\sum_{l\neq i,j}\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T\in\mathbf{S}\right)-\mathbb{P}(\mathbf{x}_K\in\mathbf{S})\right| \\
&\leq C\sum_{l\neq i,j}\mathbb{E}\|\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}(w_{il}-w_{jl})\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T\|_2^3 \\
&=C\sum_{l\neq i,j}\left(\|\boldsymbol{\Sigma}_1^{-1/2}\mathbf{D}^{-1}\left(\frac{\mathbf{v}_{1l}}{t_1/d_1},\cdots,\frac{\mathbf{v}_{Kl}}{t_K/d_K}\right)^T\|_2^3\times\mathbb{E}|w_{il}-w_{jl}|^3\right) \\
&\leq\frac{C^2}{\sqrt{n}}\max_{l\neq i,j}\mathbb{E}|\frac{w_{il}-w_{jl}}{\sqrt{\theta}}|^3 \\
&=O(\frac{1}{\sqrt{n\theta}}),
\end{aligned}
$$

which entails (E.6). Therefore, the desired conclusions of the theorem follow immediately, which concludes the proof of Theorem E.