Supplementary Material to "Asymptotic Theory of Eigenvectors for Random Matrices with Diverging Spikes"

Jianqing Fan, Yingying Fan, Xiao Han and Jinchi Lv

This Supplementary Material contains additional technical details. In particular, we present in Section B the proofs of all the lemmas and provide in Section C some further technical details on under what regularity conditions the asymptotic normality can hold for the asymptotic expansion in Theorem 5. Section D contains the technical details on relaxing the spike strength condition when considering scenario ii) of Condition 2 in place of scenario i), as well as the proof sketch for results in Section 4.2.

B Proofs of technical lemmas

B.1 Proof of Lemma 1

Let $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$ be two arbitrary n-dimensional unit vectors. Since $\mathbf{W}$ is a symmetric random matrix of independent entries above the diagonal, it is easy to show that

$$\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y} = \sum_{1\le i,j\le n,\, i<j} w_{ij}(x_iy_j + x_jy_i) + \sum_{1\le i\le n}(w_{ii} - \mathbb{E}w_{ii})(x_iy_i) \tag{A.1}$$

and

$$s_n^2 \equiv \mathbb{E}(\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y})^2 = \sum_{1\le i,j\le n,\, i<j} \mathbb{E}w_{ij}^2(x_iy_j + x_jy_i)^2 + \sum_{1\le i\le n}\mathbb{E}(w_{ii} - \mathbb{E}w_{ii})^2 x_i^2y_i^2. \tag{A.2}$$

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 see that if

$$\frac{1}{s_n^3}\left[\sum_{1\le i,j\le n,\, i<j}\mathbb{E}|w_{ij}|^3|x_iy_j + x_jy_i|^3 + \sum_{1\le i\le n}\mathbb{E}|w_{ii} - \mathbb{E}w_{ii}|^3|x_iy_i|^3\right] \to 0,$$

then it holds that

$$\frac{\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y}}{s_n} \xrightarrow{\mathscr{D}} N(0,1).$$

Since by assumption $\max_{1\le i,j\le n}|w_{ij}|\le 1$ and $\|\mathbf{x}\|_\infty\|\mathbf{y}\|_\infty \ll s_n$, we have

$$\begin{aligned}
&\frac{1}{s_n^3}\left[\sum_{1\le i,j\le n,\, i<j}\mathbb{E}|w_{ij}|^3|x_iy_j+x_jy_i|^3+\sum_{1\le i\le n}\mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^3|x_iy_i|^3\right]\\
&\quad\le \frac{2}{s_n^3}\left[\sum_{1\le i,j\le n,\, i<j}\mathbb{E}|w_{ij}|^2|x_iy_j+x_jy_i|^3+\sum_{1\le i\le n}\mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^2|x_iy_i|^3\right]\\
&\quad\ll \frac{2s_n}{s_n^3}\left[\sum_{1\le i,j\le n,\, i<j}\mathbb{E}|w_{ij}|^2|x_iy_j+x_jy_i|^2+\sum_{1\le i\le n}\mathbb{E}|w_{ii}-\mathbb{E}w_{ii}|^2|x_iy_i|^2\right]\le 2, \qquad \text{(A.3)}
\end{aligned}$$

which completes the proof of Lemma 1.

B.2 Proof of Lemma 2

The technical arguments for the proof of Lemma 2 are similar to those for the proof of Lemma 1 in Section B.1. For the case of $\mathbf{x}^T(\mathbf{W}^2-\mathbb{E}\mathbf{W}^2)\mathbf{y}$, let us first consider the term $\mathbf{x}^T\mathbf{W}^2\mathbf{y}$. Such a term can be written as

$$\begin{aligned}
&\sum_{1\le k,i,l\le n} w_{ki}w_{il}x_ky_l = \sum_{1\le k,i,l\le n,\, k>l} w_{ki}w_{il}(x_ky_l+x_ly_k)+\sum_{1\le k,i\le n} w_{ki}^2x_ky_k\\
&=\sum_{1\le k,i,l\le n,\, k>l,\, k<i} w_{ki}w_{il}(x_ky_l+x_ly_k)+\sum_{1\le k,i,l\le n,\, k>l,\, k>i} w_{ki}w_{il}(x_ky_l+x_ly_k)\\
&\quad+\sum_{1\le l<k\le n} w_{kk}w_{kl}(x_ky_l+x_ly_k)+\sum_{1\le k,i\le n} w_{ki}^2x_ky_k\\
&=\sum_{1\le k,i,l\le n,\, k>l,\, k<i} w_{ki}w_{il}(x_ky_l+x_ly_k)+\sum_{1\le k,i,l\le n,\, i>l,\, i>k} w_{ik}w_{kl}(x_iy_l+x_ly_i)\\
&\quad+\sum_{1\le l<k\le n} w_{kk}w_{kl}(x_ky_l+x_ly_k)+\sum_{1\le k,i\le n} w_{ki}^2x_ky_k\\
&=\sum_{1\le k<i\le n} w_{ki}\Big(x_k\sum_{1\le l<k\le n} w_{il}y_l+y_k\sum_{1\le l<k\le n} w_{il}x_l+x_i\sum_{1\le l<i\le n} w_{kl}y_l+y_i\sum_{1\le l<i\le n} w_{kl}x_l\Big)\\
&\quad+\sum_{1\le l<k\le n} w_{kk}w_{kl}(x_ky_l+x_ly_k)+\sum_{1\le k<i\le n} w_{ki}^2(x_ky_k+x_iy_i)+\sum_{1\le k\le n} w_{kk}^2x_ky_k. \qquad \text{(A.4)}
\end{aligned}$$

Then it follows from (A.4) and the independence of entries w_{ki} with $1\le k\le i\le n$ that

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^2\mathbf{y}=\sum_{1\le k,i\le n,\, k<i}\mathbb{E}w_{ki}^2(x_ky_k+x_iy_i)+\sum_{1\le k\le n}\mathbb{E}w_{kk}^2x_ky_k.$$

To ease the technical presentation, let us define some new notation $\omega_{kk}=2^{-1}w_{kk}$ and

$\sigma_{kk}^2 = \mathbb{E}\omega_{kk}^2$. We can further show that

$$\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y} = \sum_{1\le k,i\le n,\, k<i} w_{ki}\Big[x_k \sum_{1\le l<k\le n} w_{il}y_l + y_k \sum_{1\le l<k\le n} w_{il}x_l + x_i \sum_{1\le l<i\le n} w_{kl}y_l$$
$$+ y_i \sum_{1\le l<i\le n} w_{kl}x_l + \mathbb{E}w_{ii}(x_iy_k + x_ky_i)\Big] + \sum_{1\le k,i\le n,\, k<i} \Big[(w_{ki}^2 - \sigma_{ki}^2)(x_ky_k + x_iy_i)$$
$$+ 2(\omega_{kk}^2 - \sigma_{kk}^2)(x_ky_k + x_iy_i)\Big] + \sum_{1\le k\le n} 2(\omega_{kk} - \mathbb{E}\omega_{kk})\Big(x_k \sum_{1\le l<k\le n} w_{kl}y_l$$
$$+ y_k \sum_{1\le l<k\le n} w_{kl}x_l\Big), \tag{A.5}$$

where $\sigma_{ki}^2 = \mathbb{E}w_{ki}^2$ denotes the variance of entry w_{ki} as defined before.

We next define a σ-algebra $\mathcal{F}_t = \sigma\{\mathfrak{w}_1, \cdots, \mathfrak{w}_t\}$, where $\mathfrak{w}_t = w_{kl}$ with $t = k + 2^{-1}l(l-1)$ and $1 \le k \le l \le n$. Clearly we have $t \le 2^{-1}n(n+1)$. In fact, there is a one to one correspondence between $t \le 2^{-1}n(n+1)$ and (k,l) with $k \le l$. Suppose that such a statement is not true. Then there exist two different pairs (k_1, l_1) and (k_2, l_2) with $1 \le k_1 \le l_1 \le n$ and $1 \le k_2 \le l_2 \le n$ such that

$$k_1 + \frac{l_1(l_1-1)}{2} = k_2 + \frac{l_2(l_2-1)}{2}. \tag{A.6}$$

It is easy to see that we must have $k_1 \ne k_2$ and $l_1 \ne l_2$. Without loss of generality, let us assume that $l_1 < l_2$. Then by (A.6), it holds that

$$\frac{l_2(l_2-1)}{2} - \frac{l_1(l_1-1)}{2} = k_1 - k_2 \le k_1 - 1.$$

On the other hand, since $l_1 < l_2$ we have

$$\frac{l_2(l_2-1)}{2} - \frac{l_1(l_1-1)}{2} \ge \frac{l_1(l_1+1)}{2} - \frac{l_1(l_1-1)}{2} \ge l_1 \ge k_1,$$

which contradicts the previous inequality. Thus we have shown that there is indeed a one to one correspondence between $t \le 2^{-1}n(n+1)$ and (k,l) with $k \le l$.

Assume that $t_1 \le t_2$ with $t_1 = k_1 + 2^{-1}l_1(l_1-1)$ and $t_2 = k_2 + 2^{-1}l_2(l_2-1)$. Then using the similar arguments we can show that $l_1 \le l_2$ and further $k_1 \le k_2$ when $l_1 = l_2$. This means that for $t = k + 2^{-1}l(l-1)$ with $1 \le k \le l \le n$, we have $\mathcal{F}_t = \sigma\{\mathfrak{w}_1, \cdots, \mathfrak{w}_t\} = \sigma\{w_{ij} : 1 \le i \le j < l \text{ or } 1 \le i \le k \le j = l\}$. With such a representation, we can see that the expression in (A.5) is in fact a sum of martingale differences with respect to the σ-algebra $\mathcal{F}_{k+2^{-1}i(i-1)}$. This fact entails that for $1 \le k \le i \le n$,

$$\mathbb{E}\left[(w_{ki} - \mathbb{E}w_{ki})b_{ki} + (w_{ki}^2 - \mathbb{E}w_{ki}^2)c_{ki}|\mathcal{F}_{k+2^{-1}i(i-1)-1}\right] = 0,$$

where $b_{ki} = x_k \sum_{1\le l<k\le n} w_{il}y_l + y_k \sum_{1\le l<k\le n} w_{il}x_l + x_i \sum_{1\le l<i\le n} w_{kl}y_l + y_i \sum_{1\le l<i\le n} w_{kl}x_l +$

$(1-\delta_{ki})\mathbb{E}w_{ii}(x_iy_k+x_ky_i)$ with $\delta_{ki}=1$ when $k=i$ and 0 otherwise, and $c_{ki}=x_ky_k+x_iy_i$. The conditional variance is given by

$$\begin{aligned}
&\sum_{1\le k,i\le n,\, k<i} \mathbb{E}\left\{\left[w_{ki}b_{ki}+(w_{ki}^2-\sigma_{ki}^2)c_{ki}\right]^2|\mathcal{F}_{k+2^{-1}i(i-1)-1}\right\} \\
&\quad+\sum_{1\le k\le n}\mathbb{E}\left\{\left[(\omega_{kk}-\mathbb{E}\omega_{kk})b_{kk}+2(\omega_{kk}^2-\sigma_{kk}^2)c_{kk}\right]^2|\mathcal{F}_{2^{-1}k(k+1)-1}\right\} \\
&=\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2b_{ki}^2+2\sum_{1\le k,i\le n,\, k\le i}\gamma_{ki}b_{ki}c_{ki}+\sum_{1\le k,i\le n,\, k\le i}\kappa_{ki}c_{ki}^2, \qquad \text{(A.7)}
\end{aligned}$$

where $\gamma_{ki}=\mathbb{E}w_{ki}^3$ and $\kappa_{ki}=\mathbb{E}(w_{ki}^2-\sigma_{ki}^2)^2$ for $k\neq i$, and $\gamma_{kk}=2(\mathbb{E}\omega_{kk}^3-\sigma_{kk}^2\mathbb{E}\omega_{kk})$ and $\kappa_{kk}=4\mathbb{E}(\omega_{kk}^2-\sigma_{kk}^2)^2$.

The mean of the random variable in (A.7) can be calculated as

$$\begin{aligned}
s_{\mathbf{x},\mathbf{y}}^2=\mathbb{E}(A.7)=&\sum_{1\le k,i\le n,\, k\le i}\Big[\kappa_{ki}(x_ky_k+x_iy_i)^2+\sigma_{ki}^2\sum_{1\le l<k\le n}\sigma_{il}^2(x_ky_l+y_kx_l)^2 \\
&+\sigma_{ki}^2\sum_{1\le l<i\le n}\sigma_{kl}^2(x_iy_l+y_ix_l)^2\Big]+\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2(1-\delta_{ki})\left[\mathbb{E}(w_{ii}+w_{kk})\right]^2 \\
&\times(x_ky_i+x_iy_k)^2. \qquad \text{(A.8)}
\end{aligned}$$

Moreover, the variance of the random variable in (A.7) is given by

$$\begin{aligned}
\kappa_{\mathbf{x},\mathbf{y}}=\mathrm{var}(A.7)=&\sum_{1\le k_1,i_1,k_2,i_2\le n,\, k_1\le i_1,\, k_2\le i_2}\mathbb{E}\Big\{\big[\sigma_{k_1i_1}^2(z_{k_1i_1}^2-\mathbb{E}z_{k_1i_1}^2) \\
&+2\gamma_{k_1i_1}(x_{k_1}y_{k_1}+x_iy_i)z_{k_1i_1}\big]\big[\sigma_{k_2i_2}^2(z_{k_2i_2}^2-\mathbb{E}z_{k_2i_2}^2) \\
&+2\gamma_{k_2i_2}(x_{k_2}y_{k_2}+x_iy_i)z_{k_2i_2}\big]\Big\}, \qquad \text{(A.9)}
\end{aligned}$$

where $z_{ki}=\sum_{1\le l<k\le n}w_{il}(x_ky_l+y_kx_l)+\sum_{1\le l<i\le n}w_{kl}(x_iy_l+y_ix_l)+(1-\delta_{ki})\mathbb{E}w_{ii}(x_iy_k+x_ky_i)$.

Let us recall the classical martingale CLT; see, for example, Lemma 9.12 of Bai and Silverstein (2006). If a martingale difference sequence (Y_t) with respect to a σ-algebra $\mathcal{F}_t$ satisfies the following conditions:

a) $\frac{\sum_{t=1}^T\mathbb{E}(Y_t^2|\mathcal{F}_{t-1})}{\sum_{t=1}^T\mathbb{E}Y_t^2}\xrightarrow{\mathrm{P}}1$,

b) $\frac{\sum_{t=1}^T\mathbb{E}[Y_t^2I(|Y_t|/\sqrt{\sum_{t=1}^T\mathbb{E}Y_t^2}|\ge\epsilon)]}{\sum_{t=1}^T\mathbb{E}Y_t^2}\le\frac{\sum_{t=1}^T\mathbb{E}Y_t^4}{\epsilon^2(\sum_{t=1}^T\mathbb{E}Y_t^2)^2}\to0$ for any $\epsilon>0$,

then we have $\frac{\sum_{t=1}^TY_t}{\sqrt{\sum_{t=1}^T\mathbb{E}Y_t^2}}\xrightarrow{\mathscr{D}}N(0,1)$ as $T\to\infty$, where $I(\cdot)$ denotes the indicator function. It follows from the assumption of $\kappa_{\mathbf{x},\mathbf{y}}^{1/4}\ll s_{\mathbf{x},\mathbf{y}}$ that

$$\frac{(A.7)}{\mathbb{E}(A.7)}\xrightarrow{\mathrm{P}}1,$$

which shows that condition a) above is satisfied. Moreover, by the simple fact that for any fixed i, $\mathbb{E}w_{li}^2 y_l^2 \le 1$, and the assumptions that $s_{\mathbf{x},\mathbf{y}} \to \infty$ and $\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \to 0$, we have

$$\sum_{1\le k,i\le n,\, k<i} \mathbb{E}\Big\{ w_{ki}\Big[x_k \sum_{1\le l<k\le n} w_{il} y_l + y_k \sum_{1\le l<k\le n} w_{il} x_l + x_i \sum_{1\le l<i\le n} w_{kl} y_l$$
$$+ y_i \sum_{1\le l<i\le n} w_{kl} x_l + \mathbb{E}w_{ii}(x_i y_k + x_k y_i)\Big]\Big\}^4 + \sum_{1\le k\le n} \mathbb{E}\Big[2(\omega_{kk} - \mathbb{E}\omega_{kk})$$
$$\times \Big(x_k \sum_{1\le l<k\le n} w_{kl} y_l + y_k \sum_{1\le l<k\le n} w_{kl} x_l \Big)\Big]^4$$
$$+ \sum_{1\le k,i\le n,\, k<i} \Big\{ \mathbb{E}\big[(w_{ki}^2 - \sigma_{ki}^2)(x_k y_k + x_i y_i)\big]^4 + \mathbb{E}\big[(\omega_{kk}^2 - \sigma_{ki}^2)(x_k y_k + x_i y_i)\big]^4 \Big\} \ll s_{\mathbf{x},\mathbf{y}}^4,$$

which entails that condition b) above is also satisfied. Therefore, an application of the martingale CLT concludes the proof of Lemma 2.

B.3 Further technical details on conditions of Lemma 2

Let us gain some further insights into the technical conditions in Lemma 2. Define $a_{kl} = x_k y_l + y_k x_l$ and note that $\kappa_{ij} = \mathbb{E}(w_{ij}^2 - \sigma_{ij}^2)^2 = \mathbb{E}w_{ij}^4 - \sigma_{ij}^4$. By the assumption of $|w_{ij}| \le 1$, it is easy to see that $0 \le \kappa_{ij} \le \mathbb{E}w_{ij}^4 \le \mathbb{E}w_{ij}^2 = \sigma_{ij}^2$. Then we can show that the random variable in (A.7) subtracted by its mean $s_{\mathbf{x},\mathbf{y}}^2$ can be represented as

$$(A.7) - s_{\mathbf{x},\mathbf{y}}^2 = \sum_{1\le k,i\le n,\, k\le i} \sigma_{ki}^2 \Big[\sum_{1\le l<k\le n} (w_{il}^2 - \sigma_{il}^2) a_{kl}^2 + \sum_{1\le l<i\le n} (w_{kl}^2 - \sigma_{kl}^2) a_{il}^2$$
$$+ \sum_{1\le l_1,l_2<k\le n,\, l_1\ne l_2} w_{il_1} w_{il_2} a_{kl_1} a_{kl_2} + \sum_{1\le l_1,l_2<i\le n,\, l_1\ne l_2} w_{kl_1} w_{kl_2} a_{il_1} a_{il_2} \Big]$$
$$+ 2 \sum_{1\le k,i\le n,\, k\le i,\, 1\le l_1<k\le n,\, 1\le l_2<i\le n} \sigma_{ki}^2 w_{il_1} w_{kl_2} a_{kl_1} a_{il_2}$$
$$+ 2 \sum_{1\le k,i\le n,\, k\le i} \big[\gamma_{ki} a_{kk} + \sigma_{ki}^2 a_{ki}(1-\delta_{ki})\mathbb{E}w_{ii}\big] \Big(\sum_{1\le l<k\le n} w_{il} a_{kl} + \sum_{1\le l<i\le n} w_{kl} a_{il} \Big). \quad \text{(A.10)}$$

By (A.10) and (A.31), we have

$$\kappa_{\mathbf{x},\mathbf{y}} = \mathbb{E}\left[(A.7) - s_{\mathbf{x},\mathbf{y}}^2\right]^2 \le C\Big\{\mathbb{E}\Big[\sum_{1\le k,i\le n,\,k\le i}\sigma_{ki}^2\sum_{1\le l<k\le n}(w_{il}^2-\sigma_{il}^2)a_{kl}^2\Big]^2$$
$$+\mathbb{E}\Big[\sum_{1\le k,i\le n,\,k\le i}\sigma_{ki}^2\sum_{1\le l<i\le n}(w_{kl}^2-\sigma_{kl}^2)a_{il}^2\Big]^2+\mathbb{E}\Big(\sum_{1\le k,i\le n,\,k\le i}\sigma_{ki}^2\sum_{1\le l_1,l_2<k\le n,\,l_1\ne l_2}w_{il_1}w_{il_2}$$
$$\times a_{kl_1}a_{kl_2}\Big)^2+\mathbb{E}\Big(\sum_{1\le k,i\le n,\,k\le i}\sigma_{ki}^2\sum_{1\le l_1,l_2<i\le n,\,l_1\ne l_2}w_{kl_1}w_{kl_2}a_{il_1}a_{il_2}\Big)^2$$
$$+\mathbb{E}\Big(\sum_{1\le k,i\le n,\,k\le i,\,1\le l_1<k\le n,\,1\le l_2<i\le n}\sigma_{ki}^2 w_{il_1}w_{kl_2}a_{kl_1}a_{il_2}\Big)^2+\mathbb{E}\sum_{1\le k,i\le n,\,k\le i}\big[\gamma_{ki}a_{kk}$$
$$+\sigma_{ki}^2a_{ki}(1-\delta_{ki})\mathbb{E}w_{ii}\big]\Big(\sum_{1\le l<k\le n}w_{il}a_{kl}+\sum_{1\le l<i\le n}w_{kl}a_{il}\Big)\Big\}$$
$$\le C\Big\{\Big(\sum_{1\le k,i\le n,\,k\le i}\sigma_{ki}^2\Big)^2\Big(\sum_{1\le l<k\le n}\kappa_{il}a_{kl}^4+\sum_{1\le l<i\le n}\kappa_{kl}a_{il}^4\Big)$$
$$+\sum_{1\le k_1,k_2,l_1,l_2\le n,\,l_1\ne l_2,\,l_1<k_1,\,l_2<k_2}\sigma_{k_1i}^2\sigma_{k_2i}^2\sigma_{il_1}^2\sigma_{il_2}^2a_{k_1l_1}a_{k_1l_2}a_{k_2l_1}a_{k_2l_2}$$
$$+\sum_{1\le k,i_1,i_2,l_1,l_2\le n,\,l_1\ne l_2<\min\{i_1,i_2\}}\sigma_{ki_1}^2\sigma_{ki_2}^2\sigma_{kl_1}^2\sigma_{kl_2}^2a_{i_1l_1}a_{i_1l_2}a_{i_2l_1}a_{i_2l_2}$$
$$+\sum_{1\le k,i,l_1,l_2\le n,\,k<i,\,l_1<k,\,l_2<i}\sigma_{ki}^2\sigma_{il_1}^2\sigma_{kl_2}^2a_{kl_1}^2a_{il_2}^2$$
$$+\sum_{1\le k,i\le n,\,k<i}\big\{\gamma_{ki}^2a_{kk}^2+\sigma_{ki}^4a_{ki}^2(1-\delta_{ki})[\mathbb{E}(w_{ii}+w_{kk})]^2\big\}\Big(\sum_{1\le l<k\le n}\sigma_{il}^2a_{kl}^2+\sum_{1\le l<i\le n}\sigma_{kl}^2a_{il}^2\Big)\Big\}$$
$$=O\left\{n\sigma_n^8\|\mathbf{x}\|_\infty^4\|\mathbf{y}\|_\infty^4\right\}, \tag{A.11}$$

where C is some positive constant.

Given $\|\mathbf{x}\|_\infty\|\mathbf{y}\|_\infty\to 0$, it follows from (A.8) that

$$s_{\mathbf{x},\mathbf{y}}^2=\sum_{1\le k,i\le n,\,k<i}\Big[\kappa_{ki}(x_ky_k+x_iy_i)^2+\sum_{1\le l<k\le n}\sigma_{il}^2(x_ky_l+y_kx_l)^2$$
$$+\sum_{1\le l<i\le n}\sigma_{kl}^2(x_iy_l+y_ix_l)^2\Big]$$
$$\ge\sigma_{\min}^2\sum_{1\le k,i\le n,\,k<i}\Big[\sum_{1\le l<k\le n}(x_ky_l+y_kx_l)^2+\sum_{1\le l<i\le n}(x_iy_l+y_ix_l)^2\Big]$$
$$\ge c\sigma_{\min}^2 n, \tag{A.12}$$

where $\sigma_{\min}^2$ is defined in Condition 3. Then we can exploit the upper bound on $\kappa_{\mathbf{x},\mathbf{y}}$ in (A.11) and the lower bound on $s_{\mathbf{x},\mathbf{y}}^2$ in (A.12) to simplify the conditions of Lemma 2, which can be reduced to

$$\|\mathbf{x}\|_\infty\|\mathbf{y}\|_\infty\to 0,\ \frac{\alpha_n^4\|\mathbf{x}\|_\infty^2\|\mathbf{y}\|_\infty^2}{n^{1/2}\sigma_{\min}^2}\to 0,\ \text{and}\ \sigma_{\min}^2 n\to\infty. \tag{A.13}$$

Therefore, the conclusions of Lemma 2 hold as long as condition (A.13) is satisfied.

B.4 Proof of Lemma 3

In view of the definition of the function $f_k(z)$ defined in (10), we have

$$f_k'(z) = d_k\Big\{\mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1} \times \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)\Big\}'. \tag{A.14}$$

For $z \in [a_k, b_k]$, it follows from Lemma 5, Condition 2, and the definition of $\mathcal{R}$ in (6) that

$$\begin{aligned}\left\|\mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) + z^{-1}\mathbf{I}\right\| &= \left\|-\sum_{l=2}^{L} z^{-(l+1)}\mathbf{V}_{-k}^T \mathbb{E}\mathbf{W}^l \mathbf{V}_{-k}\right\| \\ &\leq \sum_{l=2}^{L} z^{-(l+1)}\left\|\mathbf{V}_{-k}^T \mathbb{E}\mathbf{W}^l \mathbf{V}_{-k}\right\| = O(\alpha_n^2 |z|^{-3}).\end{aligned} \tag{A.15}$$

Without loss of generality, we assume that $k \neq 1$. For l such that $|d_l| > |d_k|$, by (A.15) the diagonal entry of $\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)$ corresponding to d_l is given by

$$d_l^{-1} - z^{-1} + O(\alpha_n^2|z|^{-3}) = (z - d_l)/(zd_l) + O(\alpha_n^2|z|^{-3}).$$

By Condition 2, there exists some positive constant c such that $\max\{|a_k|, |b_k|\} \leq (1-c)|d_l|$. It follows that $|(z - d_l)/(zd_l)| \geq c/|z|$ and thus $|(z-d_l)/(zd_l) + O(\alpha_n^2|z|^{-3})|^{-1} = O(|z|)$. For the remaining diagonal entry with $|d_l| < |d_k|$, there exists some positive constant c_1 such that $\min\{|a_k|, |b_k|\} \geq (1 + c_1)|d_l|$ and similarly we have $|(z - d_l)/(zd_l) + O(\alpha_n^2|z|^{-3})|^{-1} = O(|z|)$. Thus it follows from (A.15) that the off diagonal entries of $\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)$ are dominated by the diagonal ones, leading to

$$\left\|\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\right\| = O(|z|) \tag{A.16}$$

for all $z \in [a_k, b_k]$.

Next an application of Lemma 5 gives

$$\mathcal{R}'(\mathbf{v}_k, \mathbf{v}_k, z) = \sum_{l=0,\, l\neq 1}^{L} \frac{l+1}{z^{l+2}} \mathbf{v}_k^T \mathbb{E}\mathbf{W}^l \mathbf{v}_k = \frac{1}{z^2} + O(\alpha_n^2|z|^{-4}).$$

By (A.14) and Condition 2, we have

$$\Big\{\mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)\Big\}' = O(\alpha_n^4|z|^{-6}) = o(\alpha_n^2|z|^{-4}).$$

Thus in view of (A.14), it holds that

$$f_k'(z) = d_k z^{-2}\left[1 + o(1)\right] \tag{A.17}$$

for $z \in [a_k, b_k]$. We can see from (A.17) that $f_k(z)$ is a monotone function over $z \in [a_k, b_k]$ when matrix size n is large enough.

Now recall that

$$f_k(d_k) = 1+d_k \left\{\mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, d_k) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, d_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, d_k)\right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, d_k)\right\}.$$

By Lemma 5, we have

$$1 + d_k\mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, d_k) = 1 - \sum_{l=0,\, l\neq 1}^{L} \frac{1}{d_k^l}\mathbf{v}_k^T\mathbb{E}\mathbf{W}^l\mathbf{v}_k = O(\alpha_n^2 d_k^{-2})$$

and

$$d_k\mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, d_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, d_k)\right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, d_k) = O(\alpha_n^2 d_k^{-2}).$$

Thus it holds that $f_k(d_k) = O(\alpha_n^2 d_k^{-2}) = o(1)$. Noticing that the derivative $f_k'(z) = d_k z^{-2}\left[1 + o(1)\right] \sim d_k z^{-2} \sim |d_k|^{-1}$ and by the mean value theorem, we have $f_k(a_k) \sim o(1) + |d_k|^{-1}(a_k - d_k)$ and $f_k(b_k) \sim o(1) + |d_k|^{-1}(b_k - d_k)$, where $\sim$ represents the asymptotic order. Therefore, we see that $f_k(a_k)f_k(b_k) < 0$ and consequently the equation $f_k(z) = 0$ has a unique solution for $z \in [a_k, b_k]$, which solution satisfies that $t_k = d_k + o(d_k)$. This completes the proof of Lemma 3.

B.5 Proof of Lemma 4

The asymptotic bounds characterized in Lemma 4 play a key role in establishing the more general asymptotic theory in Theorems 4 and 5. We first assume that all the diagonal entries of $\mathbf{W} = (w_{ij})_{1\leq i,j\leq n}$ are zero, that is, $w_{ii} = 0$. The general case of possibly $w_{ii} \neq 0$ will be dealt with later. The main idea of the proof is to calculate the moments by counting the number of nonzero terms involved in $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$, which is a frequently used idea in random matrix theory; see, for example, Chapter 2 of Bai and Silverstein (2006). An important difference is that bounding the order of $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$ by simply counting the number of nonzero terms inside is too rough for our setting since the variances of the entries of $\mathbf{W}$ can be very different from each other. Observe that the nonzero terms of the variance involve the product of w_{ij}^m with $m \geq 2$. We thus collect all such terms with the same index i but different index j, which means that we will bound $\sum_{j=1}^n \mathbb{E}|w_{ij}|^m \leq \alpha_n^2$ instead of using $\mathbb{E}|w_{ij}|^m \leq 1$. Then we can obtain a more accurate order since α_n^2 can be much smaller than n in general. Our technical arguments here provide useful refinements to the classical idea of counting the number of nonzero terms from the random matrix theory.

Let $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$ be two arbitrary n-dimensional unit vectors, and $l \geq 1$ an integer. Expanding $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$ yields

$$\begin{aligned}
&\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2 \\
&= \sum_{\substack{1\le i_1,\cdots,i_{l+1},j_1,\cdots,j_{l+1}\le n,\\ i_s\neq i_{s+1},\, j_s\neq j_{s+1},\, 1\le s\le l}} \mathbb{E}\Big[\big(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\big) \\
&\quad \times \big(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\big)\Big]. \qquad \text{(A.18)}
\end{aligned}$$

Let $\mathbf{i} = (i_1, \cdots, i_{l+1})$ and $\mathbf{j} = (j_1, \cdots, j_{l+1})$ be two vectors taking values in $\{1, \cdots, n\}^{l+1}$. For any given vector $\mathbf{i}$, we define a graph $\mathcal{G}_{\mathbf{i}}$ whose vertices represent distinct values of the components of $\mathbf{i}$. Vertices i_s and i_{s+1} of $\mathcal{G}_{\mathbf{i}}$ are connected by undirected edges for $1 \leq s \leq l$. Similarly we can also define graph $\mathcal{G}_{\mathbf{j}}$ corresponding to $\mathbf{j}$. It can be seen that $\mathcal{G}_{\mathbf{i}}$ is a connected graph, which means that there exists some path from i_s to $i_{s'}$ for any $1 \leq s \neq s' \leq n$. Thus for each product

$$\begin{aligned}
&\mathbb{E}\Big[\big(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\big) \\
&\quad \times \big(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\big)\Big], \qquad \text{(A.19)}
\end{aligned}$$

there exists a corresponding graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. If $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is not a connected graph, then the corresponding expectation

$$\begin{aligned}
&\mathbb{E}\Big[\big(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\big) \\
&\quad \times \big(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\big)\Big] = 0.
\end{aligned}$$

This shows that in order to calculate the order of $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$, it suffices to consider the scenario of connected graphs $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$.

To analyze the term in (A.59), let us calculate how many distinct vertices are contained in the connected graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since there are $2l$ edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ and $\mathbb{E}w_{ss'} = 0$ for $s \neq s'$, in order to get a nonzero value of (A.59) each edge in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ has at least one copy. Thus for each nonzero (A.59), we have l distinct edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is connected, there are at most $l+1$ distinct vertices in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Denote by $\mathcal{S}$ the set of all such pairs $(\mathbf{i}, \mathbf{j})$. Combining the above arguments, we can conclude that

$$\begin{aligned}
(A.18) = &\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}} \mathbb{E}\Big[\big(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\big) \\
&\quad \times \big(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\big)\Big]. \qquad \text{(A.20)}
\end{aligned}$$

For notational simplicity, we denote $j_1, \cdots, j_{l+1}$ by $i_{l+2}, \cdots, i_{2l+2}$ and define $\widetilde{\mathbf{i}} = (i_1, \cdots, i_{l+1}, j_1, \cdots, j_{l+1}) = (i_1, \cdots, i_{2l+2})$. We also denote $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ by $\mathcal{F}_{\widetilde{\mathbf{i}}}$ which has at most $l+1$ distinct vertices and l distinct edges, with each edge having at least two copies. Then it holds that

$$
\begin{aligned}
|(A.60)| = \Big| & \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big[(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}) \\
& \times (x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}} - \mathbb{E}x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}})\big]\Big| \\
\le & \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big|x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}\big| \\
& + \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big|x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\big|\mathbb{E}\big|x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}\big|. \quad \text{(A.21)}
\end{aligned}
$$

Observe that each expectation in (A.61) involves the product of some independent random variables, and $x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}$ and $x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}$ 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

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

we can further bound (A.61) as

$$
\begin{aligned}
(A.61) \le 2 & \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big|x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots \\
& \times w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}\big|. \quad \text{(A.22)}
\end{aligned}
$$

To facilitate our technical presentation, let us introduce some additional notation. Denote by $\psi(2l+2)$ the set of partitions of the edges $\{(i_1,i_2),(i_2,i_3),\cdots,(i_{2l+1},i_{2l+2})\}$ and $\psi_{\ge 2}(2l+2)$ the subset of $\psi(2l+2)$ whose blocks have size at least two. Let $P(\widetilde{\mathbf{i}}) \in \psi_{\ge 2}(2l+2)$ be the partition of $\{(i_1,i_2),(i_2,i_3),\cdots,(i_{2l+1},i_{2l+2})\}$ 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 $s_{2j-1} \neq s_{2j}$

since the diagonal entries of $\mathbf{W}$ are assumed to be zero for the moment. Then we have

$$\begin{aligned}
&\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big|x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}\big| \\
&\leq \sum_{\substack{1\leq|P(\widetilde{\mathbf{i}})|=m\leq l\\ P(\widetilde{\mathbf{i}})\in\psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}})\\ r_1,\cdots,r_m\geq 2}} \sum_{Q(\widetilde{\mathbf{s}})\in\phi(2m)} \sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\leq s_1,\cdots,s_{2m}\leq n}} |x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}| \\
&\quad\times\prod_{j=1}^m \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}. \qquad\qquad (A.23)
\end{aligned}$$

We 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 graph $\mathcal{F}_{\widetilde{\mathbf{i}}}$, we see that $\mathcal{F}_{\widetilde{\mathbf{s}}}$ is also a connected graph. In view of (A.63), putting term $|x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}|$ aside we need to analyze the summation

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

If index s_{2k-1} satisfies that $s_{2k-1}\neq s$ 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 we have

$$\begin{aligned}
&\sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})\\ 1\leq s_1,\cdots,s_{2m}\leq n}} \prod_{j=1}^m \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \\
&\leq \sum_{\substack{\widetilde{\mathbf{s}}\setminus\{s_{2k-1}\} \text{ with partition } Q(\widetilde{\mathbf{s}}\setminus\{s_{2k-1}\})\\ 1\leq s_1,\cdots,s_{2k-2},s_{2k+2},s_{2m}\leq n\\ s_{2k}=s_j \text{ for some } 1\leq j\leq 2m}} \prod_{j=1}^m \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \sum_{s_{2k-1}=1}^n \mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}. \qquad (A.24)
\end{aligned}$$

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. It follows from the definition of α_n, $|w_{ij}|\leq 1$, and $r_k\geq 2$ that

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

After taking the summation over index s_{2k-1}, we 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} need to deal with carefully. By the assumption of

$|w_{ij}| \le 1$, we have

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

Then it holds that

$$\begin{aligned}
&\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \\
&\le \sum_{\substack{\widetilde{\mathbf{S}}\backslash(s_{2k-1},s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{S}}\backslash(s_{2k-1},s_{2k_1-1})) \\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1,\, j\neq k}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \\
&\quad + \sum_{\substack{\widetilde{\mathbf{S}}\backslash(s_{2k-1},s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{S}}\backslash(s_{2k-1},s_{2k_1-1})) \\ 1\le s_1,\cdots,s_{2m}\le n}} \prod_{j=1,\, j\neq k_1}^{m} \mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j} \qquad \text{(A.25)}
\end{aligned}$$

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 (A.25) can be reduced to the case in (A.24) 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 (A.24), 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 by $w_{ab_1}, \cdots, w_{ab_{m_1}}$. 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 (A.25) 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.

(i) If there exists some single index in $\widetilde{\mathbf{s}}$, using (A.24) 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.

(ii) Repeat (i) until there is no single index in graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$.

(iii) If 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 (A.24) to obtain α_n^2 in the upper bound. Moreover, we delete k edges associated with this vertex in $\mathcal{F}_{\widetilde{\mathbf{S}}}$.

(iv) Repeat (iii) until there is no such index.

(v) If there still exists some single index, turn back to (i). 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

$$(A.63) \leq C d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}, \tag{A.26}$$

where $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$, $d_{\mathbf{y}} = \|\mathbf{x}\|_\infty$, and C 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

$$\begin{aligned}(A.26) &\leq C d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \alpha_n^{2l-2} \sum_{1 \leq s_{2k_0-1}, s_{2k_0} \leq n, \, (s_{2k_0-1}, s_{2k_0}) = \mathbf{Q}} \mathbb{E}|w_{s_{2k_0-1}s_{2k_0}}|^{r_{k_0}} \\ &\leq C d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \alpha_n^{2l} n.\end{aligned}$$

Therefore, we have established a simple upper bound of $O\{d_{\mathbf{x}} d_{\mathbf{y}} \alpha_n^l n^{1/2}\}$.

In fact, we can improve the aforementioned upper bound to $O(\alpha_n^{l-1})$. 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 $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|$ is involved in (A.63), we can analyze them together with random variables w_{ij}. There are four different cases to consider.

1). Two pairs of indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are equal. Without loss of generality, let us assume that $i_1 = i_{l+1} \neq i_{l+2} = i_{2l+2}$. Then it holds that $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = |x_{i_1} y_{i_1} x_{i_{l+2}} y_{i_{l+2}}| \leq 4^{-1}(x_{i_1}^2 + y_{i_1}^2)(x_{i_{l+2}}^2 + y_{i_{l+2}}^2)$. Let us consider the bound for

$$\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} x_{i_1}^2 x_{i_{l+2}}^2 \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}. \tag{A.27}$$

We assume without loss of generality that $i_1 = s_1$ and $i_{l+2} = s_2$ for this partition. Then the summation in (A.27) becomes

$$\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} x_{s_1}^2 x_{s_2}^2 \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}.$$

By repeating the iterative process (i)–(v) mentioned before, we can bound the summation for fixed s_2 and obtain an alternative upper bound

$$\sum_{s_1=1}^{n} x_{s_1}^2 \mathbb{E}|w_{s_1 s_2}|^{r_j} \leq \sum_{s_1=1}^{n} x_{s_1}^2 = 1$$

since $\mathbf{x}$ is a unit vector. Thus for this step of the iteration, we obtain 1 instead of α_n^2 in the upper bound. Since the graph is always connected during the iteration process, there exists another vertex b such that $w_{s_2 b}$ is involved in (A.27). For index s_2, we do not delete the edges containing s_2 in the graph during the iterative process (i)–(v). Then after the iteration stops, the final graph $\mathbf{Q}$ satisfies properties i) and ii) defined earlier except for vertex s_2. Since there are at least two vertices with degree one in $\mathbf{S}_{\mathbf{Q}}$, we will also reach a contradiction unless the number of vertices in graph $\mathbf{Q}$ is exactly one. As a result, we can obtain the upper bound

$$(A.63) \leq C\alpha_n^{2l-4} \sum_{1 \leq s_2, b \leq n, \, (s_2, b) = \mathbf{Q}} \mathbb{E} x_{s_2}^2 |w_{s_2 b}|^r \leq C\alpha_n^{2l-2} \tag{A.28}$$

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

2). Indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are all distinct. Then by the triangle inequality, we have $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| \leq 4^{-1}(x_{i_1}^2 + x_{i_{l+2}}^2)(y_{i_{l+2}}^2 + y_{i_{2l+2}}^2)$. Thus this case reduces to case 1 above.

3). Indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are all equal. Then it holds that $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = x_{i_1}^2 y_{i_1}^2 \leq x_{i_1}^2$. We see that there are at most $[(2l+2-2)/2] = l$ distinct vertices in the chain $\prod_{s=1}^{2l-1} w_{i_s i_{s+1}}$ and for this case there are at most $l-1$ distinct edges in $\mathcal{F}_{\widetilde{\mathbf{i}}}$, where $[\cdot]$ denotes the integer part of a number. Compared to case 1, the maximum number of edges in the graph becomes smaller. Therefore, for this case we have

$$(A.63) \leq C\alpha_n^{2l-4} \sum_{1 \leq s_1, b \leq n, \, (s_1, b) = \mathbf{Q}} \mathbb{E} x_{s_1}^2 |w_{s_1 b}|^r \leq C\alpha_n^{2l-2}, \tag{A.29}$$

where C is some positive constant and we have assumed that $i_1 = s_1$ without loss of generality.

4). Three of the indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are equal. For such a case, without loss of generality let us write $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = |x_{i_1}^2 y_{i_1} y_{i_{2l+2}}|$. Then there are at most

$[(2l+2-1)/2] = l$ distinct vertices in the chain $\prod_{s=1}^{2l-1} w_{i_s i_{s+1}}$ and thus for this case there are at most $l-1$ distinct edges in $\mathcal{F}_{\tilde{\mathbf{i}}}$. Therefore, this case reduces to case 3 above.

In addition, we can also improve the upper bound to $O(\min\{d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$. The technical arguments for this refinement are similar to those for the improvement to order $O(\alpha_n^{l-1})$ above. As an example, we can bound the components of $\mathbf{y}$ by $d_{\mathbf{y}} = \|\mathbf{y}\|_\infty$, which leads to $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| \le d_{\mathbf{y}}^2 (x_{i_1}^2 + x_{i_{l+1}}^2)/2$. Then the analysis becomes similar to that for case 3 above. The only difference is that the length of graph $\mathcal{F}_{\tilde{\mathbf{i}}}$ is at most l instead of $l-1$. Thus similar to (A.29), for this case we have

$$(A.63) \le C d_{\mathbf{y}}^2 \alpha_n^{2l-2} \sum_{1\le s_2, b \le n,\, (s_2,b)=\mathbf{Q}} \mathbb{E} x_{s_1}^2 |w_{s_1 b}|^r \le C d_{\mathbf{y}}^2 \alpha_n^{2l}, \tag{A.30}$$

where C is some positive constant and we have assumed that $i_1 = s_1$ or $x_{i_{l+1}} = s_1$ without loss of generality. The other one can then be used to remove a factor of α_n. Thus we can obtain the claimed upper bound $O(\min\{d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$. Therefore, combining the two aforementioned improved bounds yields the desired upper bound of $O_p(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$.

We finally return to the general case of possibly $w_{ii} \ne 0$. Let us rewrite $\mathbf{W}$ as $\mathbf{W} = \mathbf{W}_0 + \mathbf{W}_1$ with $\mathbf{W}_1 = \mathrm{diag}(w_{11}, \cdots, w_{nn})$. Then it holds that

$$\mathbf{x}^T \mathbf{W}^l \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y} = \mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} - \mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y}.$$

Recall the classical inequality

$$\mathbb{E}(X_1 + \cdots + X_m)^2 \le m(\mathbb{E}X_1^2 + \cdots + \mathbb{E}X_m^2), \tag{A.31}$$

where $X_1, \cdots, X_m$ are m random variables with finite second moments. Define a function

$$f(\mathbf{h}) = \prod_{i=1}^{l} \mathbf{W}_{h_i}, \tag{A.32}$$

where the vector $\mathbf{h} = (h_1, \cdots, h_l)$ with $h_i = 0$ or 1. Then we have

$$\begin{aligned} &\mathbb{E}\left[\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} - \mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y}\right]^2 = \mathbb{E}\Big\{\sum_{\mathbf{h}} \mathbf{x}^T [f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\Big\}^2 \\ &\le 2^l \sum_{\mathbf{h}} \mathbb{E}\Big\{\mathbf{x}^T [f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\Big\}^2. \end{aligned} \tag{A.33}$$

This shows that we need only to consider terms of form $\mathbb{E}\{\mathbf{x}^T [f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$, each of which is a polynomial of $\mathbf{W}_0$ and $\mathbf{W}_1$.

As an example, let us analyze the term $\mathbb{E}(\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y})^2$. Similar to

(A.18), it can be shown that

$$\begin{aligned}
&\mathbb{E}(\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y})^2 \\
&= \sum_{\substack{1\le i_1,\cdots,i_l,j_1,\cdots,j_l\le n,\\ i_s\ne i_{s+1},\, j_s\ne j_{s+1},\, 1\le s\le l}} \mathbb{E}\Big[\left(x_{i_1}w_{i_1i_1}w_{i_1i_2}\cdots w_{i_{l-1}i_l}y_{i_l} - \mathbb{E}x_{i_1}w_{i_1i_1}w_{i_1i_2}\cdots w_{i_{l-1}i_l}y_{i_l}\right) \\
&\quad \times \left(x_{j_1}w_{j_1j_1}w_{j_1j_2}\cdots w_{j_{l-1}j_l}y_{j_l} - \mathbb{E}x_{j_1}w_{j_1j_1}w_{j_1j_2}\cdots w_{j_{l-1}j_l}y_{j_l}\right)\Big]. \qquad (A.34)
\end{aligned}$$

Repeating the arguments from (A.18)–(A.62), we can obtain

$$\begin{aligned}
(A.34) &\le 2\sum_{\mathcal{F}_{\tilde{\mathbf{i}}}} \mathbb{E}\big|x_{i_1}w_{i_1i_1}w_{i_1i_2}\cdots w_{i_{l-1}i_l}y_{i_l}x_{i_{l+1}}w_{i_{l+1}i_{l+1}}w_{i_{l+1}i_{l+2}}\cdots w_{i_{2l-1}i_{2l}}y_{i_{2l}}\big| \\
&\le 2\sum_{\mathcal{F}_{\tilde{\mathbf{i}}}} \mathbb{E}\big|x_{i_1}w_{i_1i_2}\cdots w_{i_{l-1}i_l}y_{i_l}x_{i_{l+1}}w_{i_{l+1}i_{l+2}}\cdots w_{i_{2l-1}i_{2l}}y_{i_{2l}}\big|.
\end{aligned}$$

Comparing to (A.62), we can see that by replacing the diagonal entries with 1 in the expectations, the number of edges in this graph is no more than the original one in (A.62). Thus repeating all the steps before (A.34), we can deduce the bound

$$\mathbb{E}(\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y})^2 = O(\min\{\alpha_n^{2(l-1)}, d_{\mathbf{x}}^2\alpha_n^{2l}, d_{\mathbf{y}}^2\alpha_n^{2l}\}).$$

For the other expectations $\mathbb{E}\{\mathbf{x}^T[f(\mathbf{h})-\mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$, by the same reason that $\mathbf{W}_1$ is a diagonal matrix we can obtain a similar expression as (A.34) with the number of edges no larger than the original one for $\mathbb{E}(\mathbf{x}^T\mathbf{W}_0^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_0^l\mathbf{y})^2$. Thus all the technical arguments above can be applied to $\mathbb{E}\{\mathbf{x}^T[f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$ so we can have the same order for the upper bound as before. This shows that all the previous arguments can indeed be extended to the general case of possibly $w_{ii} \ne 0$, which concludes the proof of Lemma 4.

B.6 Proof of Lemma 5

The main idea of the proof is similar to that for the proof of Lemma 4 in Section B.5. We first consider the case when all the diagonal entries of $\mathbf{W} = (w_{ij})_{1\le i,j\le n}$ are zero, that is, $w_{ii} = 0$. Then we can derive a similar expression as (A.18)

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = \sum_{\substack{1\le i_1,\cdots,i_{l+1}\le n\\ i_s\ne i_{s+1},\, 1\le s\le l}} \mathbb{E}\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right). \qquad (A.35)$$

By the definition of graph $\mathcal{G}_{\mathbf{i}}$ in the proof of Lemma 4, we can obtain a similar expression as (A.62)

$$|(A.35)| \leq \sum_{\mathcal{G}_{\mathbf{i}} \text{ with at most } [l/2] \text{ distinct edges and } [l/2]+1 \text{ distinct vertices}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots \times w_{i_l i_{l+1}} y_{i_{l+1}}\big|. \tag{A.36}$$

Using similar arguments for bounding the order of the summation through the iterative process as for case 3 in the proof of Lemma 4 and noticing that $|x_{i_1} y_{i_{l+1}}| \leq 2^{-1}(x_{i_1}^2 + y_{i_{l+1}}^2)$, we can deduce the desired bound

$$\mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y} = O(\alpha_n^{l-1}), \tag{A.37}$$

where the diagonal entries of $\mathbf{W}$ have been assumed to be zero.

For the general case of $\mathbf{W}$ with possibly nonzero diagonal entries, we can apply the similar expansion as in the proof of Lemma 4 to get

$$\mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} = \sum_{\mathbf{h}} \mathbb{E}\mathbf{x}^T f(\mathbf{h}) \mathbf{y}, \tag{A.38}$$

where $\mathbf{W} = \mathbf{W}_0 + \mathbf{W}_1$ with $\mathbf{W}_1 = \operatorname{diag}(w_{11}, \cdots, w_{nn})$, and vector $\mathbf{h}$ and function $f(\mathbf{h})$ are as defined in (A.32). Since by assumption $\mathbf{W}_1$ is a diagonal matrix with bounded entries, an application of similar arguments as in the proof of Lemma 4 gives

$$\mathbb{E}\mathbf{x}^T f(\mathbf{h}) \mathbf{y} = O(\alpha_n^{l-1}).$$

To see this, with similar arguments as below (A.33) let us analyze the term $\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y}$ as an example. Similar to (A.35), it holds that

$$\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} = \sum_{\substack{1 \leq i_1, \cdots, i_l \leq n \\ i_s \neq i_{s+1},\, 1 \leq s \leq l-1}} \mathbb{E}\left(x_{i_1} w_{i_1 i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l}\right). \tag{A.39}$$

By the assumption of $\max_{1 \leq i \leq n} |w_{ii}| \leq 1$, we can derive a similar bound as (A.36)

$$|(A.39)| \leq \sum_{\mathcal{G}_{\mathbf{i}} \text{ with at most } [(l-1)/2] \text{ distinct edges and } [(l-1)/2]+1 \text{ distinct vertices}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots \times w_{i_{l-1} i_l} y_{i_l}\big|. \tag{A.40}$$

Since the number of edges is no more than that in (A.36), we can obtain the same bound

$$\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} = O(\alpha_n^{l-1}).$$

For the other terms in (A.38), by the same reason that $\mathbf{W}_1$ is a diagonal matrix with bounded entries we can derive similar expression as (A.40) with the number of edges no more than that in (A.36). Therefore, since l is a bounded constant we can show that $\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(\alpha_n^{l-1})$ for the general case of $\mathbf{W}$ with possibly nonzero diagonal entries. This completes the proof of Lemma 5.

B.7 Lemma 6 and its proof

Lemma 6. *The random matrix* $\mathbf{W}$ *given in (1) satisfies that for any positive constant* L*, there exist some positive constants* C_L *and* σ *such that*

$$\mathbb{P}\left\{\|\mathbf{W}\| \geq C_L(\log n)^{1/2}\alpha_n\right\} \leq n^{-L}, \tag{A.41}$$

where $\|\cdot\|$ *denotes the matrix spectral norm and* $\alpha_n = \|\mathbb{E}(\mathbf{W}-\mathbb{E}\mathbf{W})^2\|^{1/2}$.

Proof. The conclusion of Lemma 6 follows directly from Theorem 6.2 of Tropp (2012).

C Further technical details on when asymptotic normality holds for Theorem 5

We now consider the joint distribution of the three random variables specified in expression (116) in the proof of Theorem 5 in Section A.6. To establish the joint asymptotic normality under some regularity conditions, it suffices to show that the random vector $(\mathrm{tr}[(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - (\mathbf{W}^2-\mathbb{E}\mathbf{W}^2)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}], \mathrm{tr}((\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{v}_k\mathbf{v}_k^T), \mathrm{tr}((\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}))$ tends to some multivariate normal distribution as matrix size n increases, where we consider the de-meaned version of this random vector for simplicity. Consequently, we need to show that for any constants c_1, c_2, and c_3 such that $c_1^2+c_2^2+c_3^2=1$, the linear combination

$$\begin{aligned} &c_1\mathrm{tr}[(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - (\mathbf{W}^2-\mathbb{E}\mathbf{W}^2)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}] + c_2\mathrm{tr}((\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{v}_k\mathbf{v}_k^T) \\ &\quad + c_3\mathrm{tr}((\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}) \end{aligned} \tag{A.42}$$

converges to a normal distribution asymptotically. Define $\mathbf{S} = \mathbf{v}_k\mathbf{v}_k^T$ and let $\mathbf{J}$, $\mathbf{L}$, and $\mathbf{Q}$ be the rescaled versions of $\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k}$, $\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}$, and $\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}$, respectively, such that the asymptotic variance of each of the above three terms is equal to one. Then it remains to analyze the asymptotic behavior of the random variable

$$\begin{aligned} &\sum_{1\leq k,i\leq n,\, k\leq i} w_{ki}\Big\{c_1\Big[\sum_{1\leq l<k\leq n} w_{il}\mathbf{L}_{kl} + \sum_{1\leq l<i\leq n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii}\Big] \\ &\quad + (1-\delta_{ki})(c_2\mathbf{S}_{ki}+c_3\mathbf{Q}_{ki})\Big\} + c_1\sum_{1\leq k,i\leq n,\, k\leq i}(w_{ki}^2-\sigma_{ki}^2)(\mathbf{L}_{kk}+\mathbf{L}_{ii}), \end{aligned} \tag{A.43}$$

where $\mathbf{A}_{ij}$ indicates the (i,j)th entry of a matrix $\mathbf{A}$ and $\delta_{ki} = 1$ when $k = i$ and 0 otherwise.

Using similar arguments as in (A.7), we can show that (A.43) is in fact a sum of martingale differences with respect to the σ-algebra $\mathcal{F}_{k+2^{-1}i(i-1)-1}$. The conditional variance of the random variable given in (A.43) can be calculated as

$$\begin{aligned}
&\sum_{1\le k,i\le n,\, k\le i} \sigma_{ki}^2 \Big\{ c_1 \Big[\sum_{1\le l<k\le n} w_{il}\mathbf{L}_{kl} + \sum_{1\le l<i\le n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] \\
&\quad + (1-\delta_{ki})(c_2\mathbf{S}_{ki} + c_3\mathbf{Q}_{ki}) \Big\}^2 + c_1^2 \sum_{1\le k,i\le n,\, k\le i} \kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})^2 \\
&\quad + 2c_1 \sum_{1\le k,i\le n,\, k\le i} \gamma_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) \Big\{ c_1 \Big[\sum_{1\le l<k\le n} w_{il}\mathbf{L}_{kl} + \sum_{1\le l<i\le n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} \\
&\quad + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] + (1-\delta_{ki})(c_2\mathbf{S}_{ki} + c_3\mathbf{Q}_{ki}) \Big\}. \qquad \text{(A.44)}
\end{aligned}$$

Moreover, the expectation of the random variable given in (A.44) can be shown to take the form

$$\begin{aligned}
&c_1^2 \sum_{1\le k,i\le n,\, k\le i} \Big\{ \sigma_{ki}^2 \Big[\sum_{1\le l<k\le n} \sigma_{il}^2\mathbf{L}_{kl}^2 + \sum_{1\le l<i\le n} \sigma_{kl}^2\mathbf{L}_{il}^2 + \mathbf{J}_{ki}^2 + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})^2(\mathbb{E}w_{ii})^2 \Big] \\
&\quad + \kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})^2 \Big\} + c_2^2 \sum_{1\le k,i\le n,\, k\le i} \sigma_{ki}^2\mathbf{S}_{ki}^2 + c_3^2 \sum_{1\le k,i\le n,\, k\le i} \sigma_{ki}^2\mathbf{Q}_{ki}^2 \\
&\quad + 2 \sum_{1\le k,i\le n,\, k\le i} \Big[\sigma_{ki}^2(c_2\mathbf{S}_{ki} + c_3\mathbf{Q}_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] \\
&\quad + 2c_1c_2 \sum_{1\le k,i\le n,\, k\le i} \gamma_{ki}\mathbf{S}_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) + 2c_1c_3 \sum_{1\le k,i\le n,\, k\le i} \gamma_{ki}\mathbf{Q}_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) \\
&\quad + 2c_2c_3 \sum_{1\le k,i\le n,\, k\le i} \sigma_{ki}^2\mathbf{S}_{ki}\mathbf{Q}_{ki} + 2c_1^2 \sum_{1\le k,i\le n,\, k\le i} \Big[\kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big]. \qquad \text{(A.45)}
\end{aligned}$$

Let us consider the following three regularity conditions.

i) Assume that the six individual summation terms in (A.45) tend to some constants asymptotically. Then (A.45) tends to some constant C asymptotically. Without loss of generality, we assume that $C \neq 0$; otherwise (A.43) tends to zero in probability.

ii) Assume that SD(*A.44*) $\ll$ (*A.45*), where SD stands for the standard deviation of a random variable.

iii) Assume that

$$\begin{aligned}
&\sum_{1\le k,i\le n,\, k<i} \kappa_{ki} \Big\{ \mathbb{E} \Big[\sum_{1\le l<k\le n} w_{il}\mathbf{L}_{kl} + \sum_{1\le l<i\le n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big]^4 \\
&\quad + (1-\delta_{ki})(\mathbf{S}_{ki}^4 + \mathbf{Q}_{ki}^4) \Big\} + \sum_{1\le k,i\le n,\, k<i} \mathbb{E}(w_{ki}^2 - \sigma_{ki}^2)^4(\mathbf{L}_{kk}+\mathbf{L}_{ii})^4 \ll 1. \qquad \text{(A.46)}
\end{aligned}$$

We can see that conditions i) and ii) entail condition a) in the proof of Lemma 2 in Section

B.2 below (A.9), while condition iii) entails condition b). Therefore, (A.43) converges to a normal distribution asymptotically.

D Relaxing the spike strength condition and proof sketch for results in Section 4.2

The main goal of this section is to show that all the results continue to hold when Condition 2i) is replaced with Condition 2ii), which is a weaker assumption on the spike strength. Thus from now on, we will assume Condition 2ii) instead of Condition 2i). Moreover, we provide the proof sketch for results in Section 4.2.

D.1 Replacing Condition 2i) with Condition 2ii)

Checking the proofs of our theorems, we can see that it is sufficient to show that the asymptotic expansion of $\mathbf{x}^T\mathbf{G}(z)\mathbf{y}$ remains to hold under Condition 2ii). In other words, we need to prove (76) and (108) under Condition 2ii). To accommodate the smaller magnitude of d_K in Condition 2ii), the key idea is to carefully examine the asymptotic expansions (76) and (108) as $L\to\infty$. To this end, we choose $L=\log n$ and define $c'=c/(1+2^{-1}c_0)$. Since $\alpha_n\le n^{1/2}$, we have the following improved version of inequality (66)

$$\frac{\alpha_n^{L+1}(C\log n)^{(L+1)/2}}{\min\{|a_K|,|b_K|\}^{L-2}}\le\frac{\alpha_n^3(C\log n)^{(L+1)/2}}{(c'\log n)^{L-2}}\le\frac{C^{(\log n+1)/2}n^{3/2}}{(\log n)^{(\log n-5)/2}c'^{\log n-2}}\to 0 \tag{A.47}$$

for any positive constant C.

We first show that (76) holds with the choice of $L=\log n$. In view of (75), it is sufficient to establish the following two equations

$$\sum_{l=L+1}^{\infty} z^{-(2l+2)}\mathbf{x}^T\mathbf{W}^l\mathbf{y}=O_p(\frac{1}{|z|^4}) \tag{A.48}$$

and

$$\sum_{l=2}^{L} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l-\mathbb{E}\mathbf{W}^l)\mathbf{y}=O_p(\frac{\alpha_n}{|z|^3}) \tag{A.49}$$

for $z\in\Omega_k$. In fact, (A.48) is a direct consequence of Lemma 6 and (A.47). In light of the definitions of a_k and b_k below (10), we can conclude that for any $z\in\Omega_k$, $|z|>4c_1\alpha_n\log n$. Thus we see that

$$\left\{\frac{\alpha_n^{2l}(4c_1\log n)^{2l}}{|z|^{2l}}\right\}\ \text{is a decreasing sequence when } l \text{ is increasing for } z\in\Omega_k. \tag{A.50}$$

Then it follows from Lemma 7 and (A.50) that

$$\sum_{l=2}^{9} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p(\frac{\alpha_n}{|z|^3}), \tag{A.51}$$

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{l=\sqrt{L}}^{L} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2\right| \le L\sum_{l=\sqrt{L}}^{L}|z|^{-(2l+2)}\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2 \\
&\le CL\sum_{l=\sqrt{L}}^{L}\frac{(4c_1 l)^{2l}\alpha_n^{2l-2}}{|z|^{2l+2}} \le CL\sum_{l=\sqrt{L}}^{L}\frac{(4c_1\log n)^{2l}\alpha_n^{2l-2}}{|z|^{2l+2}} \\
&\le C(\log n)^2\frac{\alpha_n^4(4c_1\log n)^{2\sqrt{\log n}}}{|z|^8(c'\log n)^{2\sqrt{\log n}-4}} \le \frac{C(4c_1)^6\alpha_n^4(\log n)^6}{|z|^8(c'/(4c_1))^{2\sqrt{\log n}-4}} \ll \frac{\alpha_n^4}{|z|^8},
\end{aligned} \tag{A.52}$$

and

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{l=10}^{\sqrt{L}} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2\right| \le \sqrt{L}\sum_{l=10}^{\sqrt{L}}|z|^{-(2l+2)}\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2 \\
&\le C\sqrt{L}\sum_{l=10}^{\sqrt{L}}\frac{(4c_1 l)^{2l}\alpha_n^{2l-2}}{|z|^{2l+2}} \le C\sqrt{L}\sum_{l=10}^{\sqrt{L}}\frac{(4c_1\sqrt{\log n})^{2l}\alpha_n^{2l-2}}{|z|^{2l+2}} \\
&\le C\log n\frac{\alpha_n^4((4c_1)^2\log n)^{10}}{|z|^8(c'\log n)^{20-4}} \ll \frac{\alpha_n^4}{|z|^8}.
\end{aligned} \tag{A.53}$$

Therefore, combining (A.51)–(A.53) yields (A.49).

To establish (108) with $L = \log n$, we need only to prove (A.48) and

$$\sum_{l=3}^{L} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p(\frac{\alpha_n^2}{|z|^4}), \tag{A.54}$$

where the former has been shown before. By Lemma 7, we can deduce

$$\sum_{l=3}^{9} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p(\frac{\alpha_n^2}{|z|^4}). \tag{A.55}$$

Thus (A.54) holds by combining (A.52), (A.53), and (A.55). This concludes the proofs of the desired results.

D.2 Improvement of Lemmas 4 and 5 under Condition 2ii)

Lemma 7. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, there exists some positive constant C independent of l such that*

$$\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2 \le C(4c_1 l)^{2l}(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})^2 \tag{A.56}$$

with $l \geq 1$ some positive integer and $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$.

Lemma 8. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, there exists some positive constant C independent of l such that*

$$\left|\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y}\right| \leq C(2c_1 l)^l \alpha_n^l \tag{A.57}$$

with $l \geq 2$ some bounded positive integer.

D.3 Proof of Lemma 7

The proof of Lemma 7 is a modification of that for Lemma 4. Thus we highlight only the differences of the technical arguments here. We work directly on the general case allowing for $\mathbb{E}w_{ii} \neq 0$. In view of (A.18), we have

$$\begin{aligned}
&\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2 \\
&= \sum_{1\leq i_1,\cdots,i_{l+1},j_1,\cdots,j_{l+1}\leq n} \mathbb{E}\Big[\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right) \\
&\quad \times \left(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\right)\Big].
\end{aligned} \tag{A.58}$$

Let $\mathbf{i} = (i_1, \cdots, i_{l+1})$ and $\mathbf{j} = (j_1, \cdots, j_{l+1})$ be two vectors taking values in $\{1, \cdots, n\}^{l+1}$. For any given vector $\mathbf{i}$, we define a graph $\mathcal{G}_{\mathbf{i}}$ whose vertices represent the components of $\mathbf{i}$. Vertices i_s and i_{s+1} of $\mathcal{G}_{\mathbf{i}}$ are connected by undirected edges for $1 \leq s \leq l$. Similarly we can also define graph $\mathcal{G}_{\mathbf{j}}$ corresponding to $\mathbf{j}$. It can be seen that $\mathcal{G}_{\mathbf{i}}$ is a connected graph, which means that there exists some path from i_s to $i_{s'}$ for any $1 \leq s \neq s' \leq n$. One should notice that here we allow for $i_s = i_{s+1}$ or $j_s = j_{s+1}$. Such relaxation will affect only the number of pairs $(\mathbf{i}, \mathbf{j})$, but will not affect the main arguments of the proof which are similar to the graph arguments for proving Lemma 4. Thus for each product

$$\begin{aligned}
&\mathbb{E}\Big[\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right) \\
&\quad \times \left(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\right)\Big],
\end{aligned} \tag{A.59}$$

there exists a corresponding graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. If $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is not a connected graph, then the corresponding expectation

$$\begin{aligned}
&\mathbb{E}\Big[\left(x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}} - \mathbb{E}x_{i_1}w_{i_1i_2}w_{i_2i_3}\cdots w_{i_li_{l+1}}y_{i_{l+1}}\right) \\
&\quad \times \left(x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}} - \mathbb{E}x_{j_1}w_{j_1j_2}w_{j_2j_3}\cdots w_{j_lj_{l+1}}y_{j_{l+1}}\right)\Big] = 0.
\end{aligned}$$

This shows that in order to calculate the order of $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$, it suffices to consider the scenario of connected graphs $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$.

To analyze the term in (A.59), let us calculate how many distinct vertices are contained in the connected graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. We say that $(i_s, i_{s+1}) \in \mathcal{G}_{\mathbf{i}}$ is an *efficient edge* if $i_s \neq i_{s+1}$.

Since there are at most $2l$ efficient edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ and $\mathbb{E}w_{ss'} = 0$ for $s \neq s'$, in order to get a nonzero value of (A.59) each efficient edge in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ has at least one copy. Thus for each nonzero (A.59), we have at most l distinct efficient edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is connected, there are at most $l+1$ distinct vertices in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Denote by $\mathcal{S}$ the set of all such pairs $(\mathbf{i}, \mathbf{j})$. Combining the above arguments, we can conclude that

$$
\begin{aligned}
(A.18) = \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}} \mathbb{E}\Big[& \big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E} x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big) \\
& \times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E} x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}}\big) \Big]. \qquad (A.60)
\end{aligned}
$$

For notational simplicity, we denote $j_1, \cdots, j_{l+1}$ as $i_{l+2}, \cdots, i_{2l+2}$ and define $\widetilde{\mathbf{i}} = (i_1, \cdots, i_{l+1}, j_1, \cdots, j_{l+1}) = (i_1, \cdots, i_{2l+2})$. We also denote $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ as $\mathcal{F}_{\widetilde{\mathbf{i}}}$ which has at most $l+1$ distinct vertices and l distinct efficient edges, with each edge having at least two copies. Then it holds that

$$
\begin{aligned}
|(A.60)| = \Big| & \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big[(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E} x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}) \\
& \times (x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} - \mathbb{E} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}})\big]\Big| \\
\leq & \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big| \\
& + \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big| \mathbb{E}\big| x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|. \qquad (A.61)
\end{aligned}
$$

Observe that each expectation in (A.61) involves the product of some independent random variables, and $x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}$ and $x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}$ 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 with the aid of the inequality

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

we can further bound (A.61) as

$$
\begin{aligned}
(A.61) \leq 2 \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\, \widetilde{\mathbf{i}} \in \mathcal{S}} & \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots \\
& \times w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|. \qquad (A.62)
\end{aligned}
$$

To facilitate our technical presentation, let us introduce some additional notation. Denote by $\psi(2l+2)$ the set of partitions of the edges $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2}), i_s \neq i_{s+1}, s =$

$1, \cdots, 2l+1\}$ and $\psi_{\geq 2}(2l+2)$ the subset of $\psi(2l+2)$ whose blocks have size at least two. Let $P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)$ be the partition of $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2}), i_s \neq i_{s+1}, s = 1, \cdots, 2l+1\}$ 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 $s_{2j-1} \neq s_{2j}$ since in the partition, we consider only the off-diagonal entries (efficient edges) and for diagonal entries, we use the simple inequality $|w_{ii}| \leq 1$. Then it holds that

$$\begin{aligned}
&\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}\big| \\
&\leq \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| \\
&\quad \times \prod_{j=1}^{m} \mathbb{E}\big|w_{s_{2j-1} s_{2j}}\big|^{r_j} \\
&\leq \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \Big(\frac{c_1^2 \alpha_n^2}{n}\Big)^m \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|.
\end{aligned} \tag{A.63}$$

It suffices to bound the number of graphs in the above summation. In fact, since the graph is connected there are at most $m+1$ different vertices in the graph. Moreover, there are $2l$ edges in the original graph with at most l efficient edges and the partitions corresponding to the edges have at most $(4l)^{2l}$ cases. Thus combining these arguments together we can deduce

$$\begin{aligned}
&\sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \Big(\frac{c_1^2 \alpha_n^2}{n}\Big)^m \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| \\
&\leq d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \Big(\frac{c_1^2 \alpha_n^2}{n}\Big)^l \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\widetilde{\mathbf{s}} \text{ with partition } Q(\widetilde{\mathbf{s}})} 1 \\
&\leq d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \Big(\frac{c_1^2 \alpha_n^2}{n}\Big)^l (4l)^{2l} n^{l+1} \\
&\leq (4c_1 l)^{2l} n \alpha_n^{2l} d_{\mathbf{x}}^2 d_{\mathbf{y}}^2.
\end{aligned} \tag{A.64}$$

Therefore, we can establish the simple upper bound that

$$\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2 \le C(4c_1 l)^{2l} n\alpha_n^{2l} d_{\mathbf{x}}^2 d_{\mathbf{y}}^2. \tag{A.65}$$

For the other upper bounds $C(4c_1l)^{2l}d_{\mathbf{x}}^2\alpha_n^{2l}$, $C(4c_1l)^{2l}d_{\mathbf{y}}^2\alpha_n^{2l}$, and $C(4c_1l)^{2l}\alpha_n^{2l-2}$, the arguments are similar to those for the proof of Lemma 4. The crucial steps are considering the impact of $|x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}|$ from (A.27) to (A.30). For our case, we can directly prove the desired bounds $C(4c_1l)^{2l}d_{\mathbf{x}}^2\alpha_n^{2l}$, $C(4c_1l)^{2l}d_{\mathbf{y}}^2\alpha_n^{2l}$, and $C(4c_1l)^{2l}\alpha_n^{2l-2}$ by combining the left hand side of (A.64) with the arguments from (A.27) to (A.30). This completes the proof of Lemma 7.

D.4 Proof of Lemma 8

Similar to the proof of Lemma 5, the proof of Lemma 8 is a direct modification of that of Lemma 7. Thus we omit it for brevity.

D.5 Proof sketch for results in Section 4.2

By calculating the variance of $\widehat{p}$, we have

$$\widehat{p} = p + O_p\Big(\frac{\sqrt{p(1-p)}}{n}\Big) = p\left[1 + O_p\Big(\frac{\sqrt{1-p}}{n\sqrt{p}}\Big)\right]. \tag{A.66}$$

Then the mean and variance of $\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1$ in (26) can be estimated as

$$\mathbf{v}_1^T\widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1 = n\widehat{p}(1-\widehat{p}) \quad \text{and} \quad \widehat{\mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1)} = \widehat{p}(1-\widehat{p})\left[2(n-1) + \widehat{p}^3 + (1-\widehat{p})^3\right], \tag{A.67}$$

receptively. By Theorem 1, (A.66), and (A.67), direct calculations show that if $n^{-1} \ll p < 1$, then it holds that

$$\lambda_1 - t_1 = O_p\Big(\frac{1}{\sqrt{np}} + \sqrt{p}\Big),$$

$$\frac{\mathbf{v}_1^T\widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1}{\sqrt{\widehat{\mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1)}}} = \frac{\mathbf{v}_1^T\mathbb{E}\mathbf{W}^2\mathbf{v}_1}{\sqrt{\mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1)}} + o_p(1).$$

Thus if the conditions of Corollary 1 hold, by (24) we can obtain

$$\frac{2\lambda_1^2\left(\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1\right) + \mathbf{v}_1^T\widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1}{\left[\widehat{\mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1)}\right]^{1/2}} \quad \xrightarrow{\mathscr{D}} \quad N(0,1). \tag{A.68}$$

Since $\mathbf{v}_1 = n^{-1/2}\mathbf{1}$ under the null hypothesis, the above results together with (A.68) ensure that under the null hypothesis, statistic T_n is asymptotically standard normal.

Next we consider the case of alternative hypothesis. It can be derived that the leading eigenvalue and eigenvector take the following forms

$$d_1 = \frac{1}{2}\left[np + n_1(q-p) + \left(n^2p^2 + 2n_1(2n_1-n)p(q-p) + n_1^2(q-p)^2\right)^{1/2}\right]$$

and $\mathbf{v}_1 = (\mathbf{v}_{1,1}^T, \mathbf{v}_{1,2}^T)^T$, where $\mathbf{v}_{1,1}$ is an n_1-dimensional vector with all entries being

$$\frac{(n-n_1)p}{\sqrt{(n-n_1)(d_1-n_1q)^2 + n_1(n-n_1)^2p^2}}$$

and $\mathbf{v}_{1,2}$ is an $(n-n_1)$-dimensional vector with all entries being

$$\frac{d_1-n_1q}{\sqrt{(n-n_1)(d_1-n_1q)^2 + n_1(n-n_1)^2p^2}}.$$

With some direct calculations, we can show that under the alternative hypothesis,

$$n^{-1/2}\mathbf{1}^T\mathbf{v}_1 = \frac{(n-n_1)(d_1-n_1(q-p))}{\sqrt{n((n-n_1)(d_1-n_1q)^2 + n_1(n-n_1)^2p^2)}}. \tag{A.69}$$

Since $n_1 = o(n)$, $n^{-1} \ll p < q$, and $p \sim q$, by the Taylor expansion we can deduce

$$d_1 = np + n_1^2(q-p)\frac{4p+5(q-p)}{4np} + O(\frac{n_1^3(q-p)^2}{n^2p})$$

and

$$\begin{aligned}
&\sqrt{n((n-n_1)(d_1-n_1q)^2 + n_1(n-n_1)^2p^2)} \\
&= \sqrt{n(n-n_1)}(d_1-n_1q) + \frac{n_1\sqrt{n}(n-n_1)^2p^2}{2\sqrt{(n-n_1)}(d_1-n_1q)} + \frac{n_1^2p}{4} + O(\frac{n_1^3p}{n}) \\
&= \sqrt{n(n-n_1)}\left[np - n_1(q-p) - \frac{n_1p}{2} - \frac{n_1^2p}{4n} + n_1^2(q-p)\frac{4p+5(q-p)}{4np} + O(\frac{n_1^3p}{n^2})\right].
\end{aligned}$$

Substituting the above two equations into (A.69) yields

$$\begin{aligned}
n^{-1/2}\mathbf{1}^T\mathbf{v}_1 - 1 = &\sqrt{\frac{n-n_1}{n}}\Big[1 + \frac{n_1}{2n} - \frac{n_1^2}{4n^2} - \frac{n_1^2(q-p)^2}{n^2p^2} - n_1^2(q-p)\frac{4p+5(q-p)}{4n^2p^2} \\
&+ O(\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2})\Big] - 1 \\
= &-\frac{n_1^2(q-p)^2}{n^2p^2} - n_1^2(q-p)\frac{4p+5(q-p)}{4n^2p^2} + O(\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2}).
\end{aligned} \tag{A.70}$$

If the conditions of Corollary 1 hold, by (24) we have

$$\frac{2\lambda_1^2\left(\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1\right) + \mathbf{v}_1{}^T\mathbb{E}\mathbf{W}^2\mathbf{v}_1}{\left[\mathrm{var}(\mathbf{v}_1{}^T\mathbf{W}^2\mathbf{v}_1)\right]^{1/2}} \overset{\mathscr{D}}{\longrightarrow} N(0,1). \tag{A.71}$$

This entails that

$$\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1 = O_p(\frac{\mathbf{v}_1{}^T\mathbb{E}\mathbf{W}^2\mathbf{v}_1 + \left[\mathrm{var}(\mathbf{v}_1{}^T\mathbf{W}^2\mathbf{v}_1)\right]^{1/2}}{t_1^2}) = O_p(\frac{1}{np}), \tag{A.72}$$

where the last step is obtained by directly calculating the mean and variance of $\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1$ and noting that $t_1 \sim np$. Since $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form an orthonormal basis, it follows from (A.72) that

$$\sum_{j=2}^{n}(\mathbf{v}_j^T\widehat{\mathbf{v}}_1)^2 = 1 - (\mathbf{v}_1^T\widehat{\mathbf{v}}_1)^2 = O_p(\frac{1}{np}). \tag{A.73}$$

Similarly, by (A.69) and the assumptions of $n_1 = o(n)$ and $q \sim p$, we can deduce

$$\sum_{j=2}^{n}(n^{-1/2}\mathbf{1}^T\mathbf{v}_j)^2 = O(\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2}). \tag{A.74}$$

Then it follows from (A.69), (A.72), and (A.74) that

$$\begin{aligned} n^{-1/2}\mathbf{1}^T\widehat{\mathbf{v}}_1 - 1 &= n^{-1/2}\mathbf{1}^T\mathbf{v}_1\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1 + n^{-1/2}\mathbf{1}^T\sum_{j=2}^{n}\mathbf{v}_j\mathbf{v}_j^T\widehat{\mathbf{v}}_1 \\ &= -\Big[\frac{n_1^2(q-p)^2}{n^2p^2} + n_1^2(q-p)\frac{4p+5(q-p)}{4n^2p^2}\Big] + O_p\Big[\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2} + \frac{1}{np}\Big]. \end{aligned} \tag{A.75}$$

Under the alternative hypothesis, it can be shown that the estimators in (27) are of orders $n\widehat{p}(1-\widehat{p}) = O_p(np)$ and $\widehat{p}(1-\widehat{p})\left[2(n-1)+\widehat{p}^3+(1-\widehat{p})^3\right] = O_p(np)$, respectively, and in addition, $t_1 \sim np$. Therefore, if the conditions of Corollary 1 holds and $\frac{n_1^2(q-p)^2}{np} + \frac{n_1^2(q-p)}{n} \gg 1$, with probability tending to one we have

$$T_n \to -\infty,$$

which means that the power can tend to one asymptotically. This concludes the proof sketch for the results in Section 4.2.