Supplementary Material for "Power Enhancement and Phase Transitions for Global Testing of the Mixed Membership Stochastic Block Model"

LOUIS CAMMARATA[1] and ZHENG TRACY KE[2,*]

[1]*Statistics Department, Harvard University, Cambridge, MA 02138, USA.*
E-mail: louis_cammarata@g.harvard.edu
[2]*Statistics Department, Harvard University, Cambridge, MA 02138, USA.*
E-mail: [*]zke@fas.harvard.edu

This supplemental material provides computations for examples and remarks, as well as proofs of theorems, corollaries and propositions. Appendix A covers the computations of τ_n and δ_n in Example 1, while Appendix B contains the calculation of the Intrinsic Number of Communities of the rank-1 model of Example 2, along with computations of τ_n and δ_n for that model. Appendix C shows the signal-to-noise ratios of the order-m Signed Path and Signed Cycle statistics, for m arbitrary. In Appendix D, we derive the asymptotic joint null distribution of Theorem 2.1. Appendix E shows the proof of Theorem 2.2, which consists in providing a lower bound for the expectation of the χ^2 test statistic and an upper bound for its variance under the alternative hypothesis. Likewise, Appendix F derives the lower bound for the expectation of the oSQ test statistic and the upper bound for its variance under the alternative hypothesis, presented in Theorem 2.3. Appendix G and Appendix H respectively report the proofs of Corollary 2.2 and Corollary 2.3 about the level and the power of the χ^2 test and the oSQ test. The proof of Theorem 2.4 about the power and the level of the PE test is provided in Appendix I. Appendix J shows the proof of the lower bound, which corresponds to Theorem 2.5, and Appendix K contains the proof of the minimax result of Theorem 2.6. Finally, Appendix L shows the proof of Proposition 3.1 and Proposition 3.2 which examine the identifiability of MMSBM and give an alternative definition of the Intrinsic Number of Communities.

Contents

A Calculations in Example 1 . . . 3
B Calculations in Example 2 . . . 5
C Calculations in Remark 2 . . . 6
  C.1 SNR of Signed Path statistics . . . 6
  C.2 SNR of Signed Cycle statistics . . . 8
D Proof of Theorem 3.1 . . . 10
  D.1 Proof of the null distribution of ψ_n^{DC} . . . 10
  D.2 Proof of the null distribution of ψ_n^{SQ} . . . 18
  D.3 Proof of the joint null distribution . . . 28
E Proof of Theorem 3.2 . . . 32
F Proof of Theorem 3.3 . . . 37
G Proof of Corollary 3.2 . . . 51
H Proof of Corollary 3.3 . . . 51
I Proof of Theorem 3.4 . . . 52

J Proof of Theorem 3.5 53
K Proof of Theorem 3.6 59
L Proof of Propositions 4.1-4.2 65
L.1 Proof of Proposition 4.1 65
L.2 Proof of Proposition 4.2 66

Appendix A: Calculations in Example 1

Introduce

$$y_n = 1 - 2\epsilon_n, \qquad z_n = (d_n - a_n)/2.$$

Recall that $\bar{a}_n = (a_n + d_n)/2$, and $z_n = (d_n - a_n)/2$. Then,

$$P = (\bar{a}_n - b_n)I_2 - z_n e_1 e_1' + z_n e_2 e_2' + b_n \mathbf{1}_2 \mathbf{1}_2', \qquad h = \frac{1}{2}(1 - y_n, 1 + y_n)'.$$

We calculate α_0, $\|Mh\|^2$ and $\|M\|^2$ in general cases.

First, consider α_0. Note that $\|h\|^2 = h_1^2 + h_2^2 = (1 + y_n^2)/2$ and $h_2^2 - h_1^2 = y_n$. We have

$$\begin{aligned} \alpha_0 = h'Ph &= h'\Big[(\bar{a}_n - b_n)I_2 - z_n e_1 e_1' + z_n e_2 e_2' + b_n \mathbf{1}_2 \mathbf{1}_2'\Big]h \\ &= (\bar{a}_n - b_n)\|h\|^2 + z_n(h_2^2 - h_1^2) + b_n \\ &= \bar{a}_n(1 + y_n^2)/2 + z_n y_n + b_n(1 - y_n^2)/2. \end{aligned} \tag{A.1}$$

Next, we calculate $\|Mh\|^2$. It follows from (A.1) that

$$\alpha_0 - b_n = (\bar{a}_n - b_n)(1 + y_n^2)/2 + z_n y_n. \tag{A.2}$$

We plug it into the expression of Mh to get

$$\begin{aligned} Mh = Ph - \alpha_0 \mathbf{1}_2 &= \Big[(\bar{a}_n - b_n)I_2 - z_n e_1 e_1' + z_n e_2 e_2' + b_n \mathbf{1}_2 \mathbf{1}_2'\Big]h - \alpha_0 \mathbf{1}_2 \\ &= (\bar{a}_n - b_n)\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} - z_n \begin{bmatrix} h_1 \\ 0 \end{bmatrix} + z_n \begin{bmatrix} 0 \\ h_2 \end{bmatrix} + b_n \mathbf{1}_2 - \alpha_0 \mathbf{1}_2 \\ &= \begin{bmatrix} (\bar{a}_n - b_n - z_n)h_1 \\ (\bar{a}_n - b_n + z_n)h_2 \end{bmatrix} - (\alpha_0 - b_n)\mathbf{1}_2 \\ &= \frac{1}{2}\begin{bmatrix} (\bar{a}_n - b_n - z_n)(1 - y_n) \\ (\bar{a}_n - b_n + z_n)(1 + y_n) \end{bmatrix} - (\alpha_0 - b_n)\mathbf{1}_2 \\ &= \frac{1}{2}\begin{bmatrix} \bar{a}_n - b_n - z_n \\ \bar{a}_n - b_n + z_n \end{bmatrix} + \frac{y_n}{2}\begin{bmatrix} -(\bar{a}_n - b_n - z_n) \\ \bar{a}_n - b_n + z_n \end{bmatrix} - (\alpha_0 - b_n)\mathbf{1}_2 \\ &= \frac{1}{2}\begin{bmatrix} \bar{a}_n - b_n - z_n \\ \bar{a}_n - b_n + z_n \end{bmatrix} + \frac{y_n}{2}\begin{bmatrix} -(\bar{a}_n - b_n - z_n) \\ \bar{a}_n - b_n + z_n \end{bmatrix} \\ &\quad - \frac{1}{2}\begin{bmatrix} 2z_n y_n \\ 2z_n y_n \end{bmatrix} - \frac{1}{2}\begin{bmatrix} \bar{a}_n - b_n \\ \bar{a}_n - b_n \end{bmatrix} - \frac{1}{2}\begin{bmatrix} y_n^2(\bar{a}_n - b_n) \\ y_n^2(\bar{a}_n - b_n) \end{bmatrix} \\ &= \frac{z_n + y_n(\bar{a}_n - b_n)}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix} - \frac{y_n z_n + y_n^2(\bar{a}_n - b_n)}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

The two vectors, $\mathbf{1}_2$ and $(1, -1)'$, are orthogonal to each other. It follows that

$$\|Mh\|^2 = \frac{1}{2}\Big[z_n + y_n(\bar{a}_n - b_n)\Big]^2 + \frac{y_n^2}{2}\Big[z_n + y_n(\bar{a}_n - b_n)\Big]^2$$

$$= \frac{1}{2}(1+y_n^2)\Big[z_n + y_n(\bar{a}_n - b_n)\Big]^2. \tag{A.3}$$

Last, we calculate $\|M\|^2$. We have seen that

$$M = (\bar{a}_n - b_n)I_2 - z_n e_1 e_1' + z_n e_2 e_2' - (\alpha_0 - b_n)\mathbf{1}_2\mathbf{1}_2'.$$

Introduce $M_0 = (\bar{a}_n - b_n)I_2 - (\alpha_0 - b_n)\mathbf{1}_2\mathbf{1}_2'$. Then,

$$M = M_0 - z_n \mathrm{diag}(1, -1). \tag{A.4}$$

We compute the two eigenvalues of M_0. Write $v = (1, -1)'$. It is seen that v is orthogonal to $\mathbf{1}_2$; furthermore,

$$M_0 v = \Big[(\bar{a}_n - b_n)I_2 - (\alpha_0 - b_n)\mathbf{1}_2\mathbf{1}_2'\Big]v = (\bar{a}_n - b_n)v \quad \propto \quad v,$$
$$M_0\mathbf{1}_2 = \Big[(\bar{a}_n - b_n)I_2 - (\alpha_0 - b_n)\mathbf{1}_2\mathbf{1}_2'\Big]\mathbf{1}_2 = [(\bar{a}_n - b_n) - 2(\alpha_0 - b_n)]\mathbf{1}_2 \quad \propto \quad \mathbf{1}_2.$$

It follows that $\mathbf{1}_2$ and v are two eigenvectors of M^*, with the associated eigenvalues as

$$\begin{aligned}
\lambda_1(M_0) &= (\bar{a}_n - b_n),\\
\lambda_2(M_0) &= (\bar{a}_n - b_n) - 2(\alpha_0 - b_n)\\
&= (\bar{a}_n - b_n) - [(\bar{a}_n - b_n)(1+y_n^2) + 2z_n y_n]\\
&= -(\bar{a}_n - b_n)y_n^2 - 2z_n y_n,
\end{aligned} \tag{A.5}$$

where we have applied (A.2) in the last equality. Combining (A.4)-(A.5), we have

$$\|M\| \sim \begin{cases} |z_n|, & \text{if } |z_n| \gg |\bar{a}_n - b_n|,\\ |\bar{a}_n - b_n|, & \text{if } |z_n| \ll |\bar{a}_n - b_n|. \end{cases} \tag{A.6}$$

We now combine (A.1), (A.3) and (A.6). In Case (S), $z_n = 0$ and $y_n = 0$. It follows that

$$\alpha_0 = \frac{a_n + b_n}{2}, \qquad \|Mh\|^2 = 0, \qquad \|M\|^2 = (\bar{a}_n - b_n)^2.$$

Plugging them into the definitions of δ_n and τ_n and noting that $\bar{a}_n = a_n$ in this case, we immediately get the claims for Case (S). In Case (AS1), $\bar{a}_n = a_n$ and $z_n = 0$ but y_n may be nonzero. It follows that

$$\alpha_0 = \frac{(1+y_n^2)a_n + (1-y_n^2)b_n}{2}, \qquad \|Mh\|^2 = \frac{1}{2}(1+y_n^2)y_n^2(a_n - b_n)^2, \qquad \|M\|^2 = (a_n - b_n)^2.$$

Assuming that $|a_n - b_n| = O(a_n + b_n)$, it follows that $(1+y_n^2)a_n + (1-y_n^2)b_n = (1+Cy_n^2)(a_n+b_n)$ for some constant $C > 0$. We obtain

$$\alpha_0 \asymp \frac{a_n + b_n}{2}, \qquad \|Mh\|^2 \asymp \frac{1}{2}y_n^2(a_n - b_n)^2, \qquad \|M\|^2 = (a_n - b_n)^2.$$

In Case (AS2), $y_n = 0$ and $z_n \gg |\bar{a}_n - b_n|$. It follows that

$$\alpha_0 = \frac{\bar{a}_n + b_n}{2}, \qquad \|Mh\|^2 = z_n^2/2, \qquad \|M\|^2 \sim z_n^2.$$

In Case (AS3), $y_n = 0$ and $z_n \ll |\bar{a}_n - b_n|$. It follows that

$$\alpha_0 = \frac{\bar{a}_n + b_n}{2}, \qquad \|Mh\|^2 = z_n^2/2, \qquad \|M\|^2 \sim (\bar{a}_n - b_n)^2.$$

The claims follow directly. □

Appendix B: Calculations in Example 2

We start by showing that the rank-1 model of Example 2 has Intrinsic Number of Communities (INC) equal to 2, regardless of K. We first recognize that the INC must be at least greater or equal to 2. Indeed, suppose that the INC is equal to 1, then we can find $\eta^* \in [0,1]$ such that $\Omega = (\eta^*)^2 \mathbf{1}_n \mathbf{1}_n'$. From the original model formulation we had $\Omega = \Pi\eta\eta'\Pi'$, and we assumed that $\eta \not\propto \mathbf{1}_K$. Thus, it is impossible for Ω to have all equal entries if Π is eligible, which contradicts the earlier fact that $\Omega = (\eta^*)^2 \mathbf{1}_n \mathbf{1}_n'$, *QEA*!

We now show that the INC is equal to 2. Define

$$\eta^* = (\eta_1^*, \eta_2^*)' \in [0,1]^2, \quad \text{where} \quad \begin{cases} \eta_1^* = \max_{k \in [\![1,K]\!]} \eta_k \\ \eta_2^* = \min_{k \in [\![1,K]\!]} \eta_k. \end{cases}$$

We also define the matrix $H \in [0,1]^{K \times 2}$ such that

$$H = \frac{1}{\eta_1^* - \eta_2^*} \begin{pmatrix} \eta_1 - \eta_2^* & \eta_1^* - \eta_1 \\ \vdots & \vdots \\ \eta_K - \eta_2^* & \eta_1^* - \eta_K \end{pmatrix}.$$

It is straightforward to check that $H\eta^* = \eta$ and that $\Pi^* := \Pi H$ is an eligible mixed membership matrix. It follows that

$$\Omega = \Pi P \Pi' = \Pi\eta\eta'\Pi' = \Pi H \eta^* (\eta^*)' H' \Pi' = \Pi^* P^* (\Pi^*)', \tag{B.1}$$

where we have defined the matrix $P^* = \eta^*(\eta^*)' \in [0,1]^{2\times 2}$. This shows that the INC of this rank-1 model is equal to 2, regardless of $K \geq 2$.

Next, we compute the Signal-to-Noise Ratios (SNR) of both tests for the rank-1 model introduced in Example 2. We start by computing the SNR of the degree test statistic, δ_n. Recall that

$$\delta_n := n^{3/2} \alpha_0^{-1} \|Ph - \alpha_0 \mathbf{1}_K\|^2.$$

Direct calculations show that

$$P := \eta\eta' = \frac{c_n}{a_n^2 + b_n^2} \begin{pmatrix} a_n^2 & a_n b_n \\ a_n b_n & b_n^2 \end{pmatrix}, \quad \text{and} \quad \alpha_0 := h' P h = \frac{c_n (a_n + b_n)^2}{4(a_n^2 + b_n^2)}.$$

This allows computing

$$Ph - \alpha_0 \mathbf{1}_K = \frac{c_n (a_n + b_n)(a_n - b_n)}{4(a_n^2 + b_n^2)} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Together, the results for α_0 and $Ph - \alpha_0 \mathbf{1}_K$ yield the following expression of the SNR:

$$\delta_n = \frac{1}{2} n^{3/2} c_n \frac{(a_n - b_n)^2}{(a_n^2 + b_n^2)} \propto n^{3/2} c_n \frac{(a_n - b_n)^2}{(a_n^2 + b_n^2)}. \tag{B.2}$$

Then, we compute the SNR of the oSQ test statistic, τ_n. Recall that

$$\tau_n := n^2 \alpha_0^{-2} \| P - \alpha_0 \mathbf{1}_K \mathbf{1}_K' \|^4.$$

We only need to compute $\|P - \alpha_0 \mathbf{1}_K \mathbf{1}_K'\|$. Straightforward calculations reveal that

$$P - \alpha_0 \mathbf{1}_K \mathbf{1}_K' = \frac{c_n(a_n - b_n)}{4(a_n^2 + b_n^2)} \begin{pmatrix} 3a_n + b_n & b_n - a_n \\ b_n - a_n & 3b_n + a_n \end{pmatrix} =: \frac{c_n(a_n - b_n)}{4(a_n^2 + b_n^2)} Q,$$

where we introduced the matrix Q for notational convenience. The eigenvalues λ_+, λ_- of Q are the solutions to the following equation in the x-variable

$$x^2 - \text{Tr}(Q)x + \det(Q) = 0, \quad \text{where} \quad \begin{cases} \text{Tr}(Q) = 4(a_n + b_n) \\ \det(Q) = 2(a_n + b_n)^2 + 8a_n b_n. \end{cases}$$

We thus obtain that

$$\lambda_\pm = 2(a_n + b_n) \pm |a_n - b_n|, \quad \text{so} \quad \lambda_+ \asymp a_n + b_n,$$

where the last equivalence follows from our assumption that $|a_n - b_n| = O(a_n + b_n)$. It follows that

$$\|P - \alpha_0 \mathbf{1}_K \mathbf{1}_K'\| \asymp \frac{c_n(a_n - b_n)(a_n + b_n)}{4(a_n^2 + b_n^2)}.$$

As a consequence,

$$\tau_n \asymp n^2 c_n^2 \frac{(a_n - b_n)^4}{(a_n^2 + b_n^2)^2} \asymp n^{-1} \delta_n. \tag{B.3}$$

Appendix C: Calculations in Remark 2

C.1. SNR of Signed Path statistics

We consider the *length-m Signed Path statistic* $V_n^{(m)}$ defined as

$$V_n^{(m)} = \sum_{i_1, \ldots, i_{m+1} \text{(distinct)}} (A_{i_1 i_2} - \hat{\alpha}_n)(A_{i_2 i_3} - \hat{\alpha}_n) \ldots (A_{i_m i_{m+1}} - \hat{\alpha}_n), \quad \text{for } m \geq 2,$$

where we recall that

$$\hat{\alpha}_n = \frac{1}{n(n-1)} \sum_{i \neq j} A_{ij}.$$

For simplicity, we study the corresponding ideal statistic $\bar{V}_n^{(m)}$, where we replace $\hat{\alpha}_n$ by the population null edge probability α_n:

$$\bar{V}_n^{(m)} = \sum_{i_1,\dots,i_{m+1}(\text{distinct})} (A_{i_1i_2} - \alpha_n)(A_{i_2i_3} - \alpha_n)\dots(A_{i_mi_{m+1}} - \alpha_n).$$

The following lemma derives the null mean and variance as well as the alternative mean of the ideal length-m Signed Path statistic. It uses the following quantities, which are defined in the main text:

$$h = \frac{1}{n}\sum_{i=1}^{n} \pi_i, \qquad \alpha_0 = h'Ph, \qquad \text{and} \qquad G = \frac{1}{n}\Pi'\Pi.$$

In addition, we denote by $\mathbb{E}_1[\cdot]$ the expectation under the alternative distribution and by $\mathbb{E}_0[\cdot]$, $\text{Var}_0(\cdot)$ the expectation and variance under the null distribution, respectively.

Lemma C.1 (Moments of the ideal length-m Signed Path statistic**).** *Suppose that conditions* (3.4) *and* (3.5) *hold. In addition, let $M = P - \alpha_0 \mathbf{1}_K \mathbf{1}_K'$ and suppose that $n^{-1}\|Mh\|^{-1}\|M\|^2 = o(1)$. Then,*

$$\mathbb{E}_0\left[\bar{V}_n^{(m)}\right] = 0, \quad \mathit{Var}_0\left(\bar{V}_n^{(m)}\right) \asymp n^{m+1}\alpha_n^m, \quad \textit{and} \quad \mathbb{E}_1\left[\bar{V}_n^{(m)}\right] \asymp n^{m+1}\|Mh\|^2\|M\|^{m-2}.$$

Proof
Under the null hypothesis, we can write

$$\bar{V}_n^{(m)} = \sum_{i_1,\dots,i_{m+1}(\text{distinct})} W_{i_1i_2}W_{i_2i_3}\dots W_{i_mi_{m+1}},$$

where $W_{ij} = A_{ij} - \alpha_n$ for all $i \neq j$. It is straightforward to obtain that $\mathbb{E}_0\left[\bar{V}_n^{(m)}\right] = 0$. Next, we compute the null variance of the ideal Signed Path statistic. We have, by direct calculations:

$$\begin{aligned}\text{Var}_0\left(\bar{V}_n^{(m)}\right) &= \text{Var}_0\left(\sum_{i_1,\dots,i_{m+1}(\text{distinct})} W_{i_1i_2}W_{i_2i_3}\dots W_{i_mi_{m+1}}\right) \\ &= \mathbb{E}_0\left[\sum_{\substack{i_1,\dots,i_{m+1}(\text{distinct}) \\ j_1,\dots,j_{m+1}(\text{distinct})}} W_{i_1i_2}\dots W_{i_mi_{m+1}}W_{j_1j_2}\dots W_{j_mj_{m+1}}\right] \asymp n^{m+1}\alpha_n^m. \qquad \text{(C.1)}\end{aligned}$$

Under the alternative hypothesis, we choose P and h such that $\alpha_0 := h'Ph = \alpha_n$. This choice ensures that the network will have the same average degree under the null and alternative hypotheses, thus making the testing problem harder. As a result, we can write:

$$\bar{V}_n^{(m)} = \sum_{\substack{i_1,\dots,i_{m+1} \\ (\text{distinct})}} (W_{i_1i_2} + \bar{\Omega}_{i_1i_2})(W_{i_2i_3} + \bar{\Omega}_{i_2i_3})\dots(W_{i_mi_{m+1}} + \bar{\Omega}_{i_mi_{m+1}}),$$

where $W_{ij} = A_{ij} - \Omega_{ij}$ and $\bar{\Omega}_{ij} = \Omega_{ij} - \alpha_0 = \pi_i'M\pi_j$ for all $i \neq j$. It follows that

$$\mathbb{E}_1\left[\bar{V}_n^{(m)}\right] = \sum_{\substack{i_1,\dots,i_{m+1} \\ (\text{distinct})}} \bar{\Omega}_{i_1i_2}\bar{\Omega}_{i_2i_3}\dots\bar{\Omega}_{i_mi_{m+1}}$$

$$= \sum_{i_1,\dots,i_{m+1}} \bar{\Omega}_{i_1 i_2}\bar{\Omega}_{i_2 i_3}\dots\bar{\Omega}_{i_m i_{m+1}} - \sum_{\substack{i_1,\dots,i_{m+1} \\ \text{(not distinct)}}} \bar{\Omega}_{i_1 i_2}\bar{\Omega}_{i_2 i_3}\dots\bar{\Omega}_{i_m i_{m+1}}$$

$$= \mathbf{1}_n'\bar{\Omega}^m\mathbf{1}_n - O(n^m\|M\|^m) = \mathbf{1}_n(\Pi M\Pi')^m\mathbf{1}_n' - O(n^m\|M\|^m)$$

$$= n^{m-1}\mathbf{1}_n(\Pi MGM\dots MGM\Pi')\mathbf{1}_n - O(n^m\|M\|^m)$$

$$= n^{m+1}h'M(GM\dots MG)Mh - O(n^m\|M\|^m)$$

Since we have assumed that $\|G\|, \|G^{-1}\| < c$ and $n^{-1}\|Mh\|^{-1}\|M\|^2 = o(1)$, we obtain that

$$\mathbb{E}_1\left[\bar{V}_n^{(m)}\right] \asymp n^{m+1}\|Ph - \alpha_0\mathbf{1}_K\|^2\|P - \alpha_0\mathbf{1}_K\mathbf{1}_K\|^{m-2}. \tag{C.2}$$

□

The results in Lemma C.1 allow us to compute the SNR for the length-m Signed Path statistic. We derive the SNR assuming that the null variance dominates the alternative variance. Thus,

$$SNR\left(\bar{V}_n^{(m)}\right) = \frac{\left|\mathbb{E}_1\left[\bar{V}_n^{(m)}\right] - \mathbb{E}_0\left[\bar{V}_n^{(m)}\right]\right|}{\sqrt{\max\left\{\text{Var}_0\left(\bar{V}_n^{(m)}\right), \text{Var}_1\left(\bar{V}_n^{(m)}\right)\right\}}} \asymp \frac{\left|\mathbb{E}_1\left[\bar{V}_n^{(m)}\right]\right|}{\sqrt{\text{Var}_0\left(\bar{V}_n^{(m)}\right)}}$$

$$\asymp \frac{n^{m+1}\|Mh\|^2\|M\|^{m-2}}{n^{(m+1)/2}\alpha_0^{m/2}} = \delta_n\tau_n^{(m-2)/4}.$$

Similar to our results in Theorem 3.2, there may be instances in which the alternative variance dominates the null variance. In these cases, the SNR still depends on powers of δ_n and τ_n, and the detection boundary is unchanged; details are omitted.

C.2. SNR of Signed Cycle statistics

We consider the *length-m Signed Cycle statistic* $U_n^{(m)}$ defined as

$$U_n^{(m)} = \sum_{i_1,\dots,i_m \text{(distinct)}} (A_{i_1 i_2} - \hat{\alpha}_n)(A_{i_2 i_3} - \hat{\alpha}_n)\dots(A_{i_m i_1} - \hat{\alpha}_n), \quad \text{for } m \geq 3.$$

For simplicity, we study the corresponding ideal statistic $\bar{U}_n^{(m)}$, where we replace $\hat{\alpha}_n$ by the population null edge probability α_n:

$$\bar{U}_n^{(m)} = \sum_{i_1,\dots,i_m \text{(distinct)}} (A_{i_1 i_2} - \alpha_n)(A_{i_2 i_3} - \alpha_n)\dots(A_{i_m i_1} - \alpha_n), \quad \text{for } m \geq 3.$$

Lemma C.2 (Moments of the ideal length-m Signed Cycle statistic). *Suppose that conditions (3.4) and (3.5) hold. In addition, let $M = P - \alpha_0\mathbf{1}_K\mathbf{1}_K'$ and assume that $|Tr(MG)| \asymp \|MG\|$. Then,*

$$\mathbb{E}_0\left[\bar{U}_n^{(m)}\right] = 0, \quad Var_0\left(\bar{U}_n^{(m)}\right) \asymp n^m\alpha_n^m, \quad \textit{and} \quad \left|\mathbb{E}_1\left[\bar{U}_n^{(m)}\right]\right| \asymp n^m\|M\|^m.$$

Proof
Under the null hypothesis, we can write

$$\bar{U}_n^{(m)} = \sum_{i_1,\ldots,i_m (\text{distinct})} W_{i_1 i_2} W_{i_2 i_3} \ldots W_{i_m i_1},$$

where $W_{ij} = A_{ij} - \alpha_n$ for all $i \neq j$. It is straightforward to obtain that $\mathbb{E}_0 \left[\bar{U}_n^{(m)}\right] = 0$. Next, we compute the null variance of the ideal Signed Cycle statistic. We have, by direct calculations:

$$\text{Var}_0 \left(\bar{U}_n^{(m)}\right) = \text{Var}_0 \left(\sum_{i_1,\ldots,i_m (\text{distinct})} W_{i_1 i_2} W_{i_2 i_3} \ldots W_{i_m i_1} \right).$$

Similar to Equation (D.27), we can decompose the sum into a sum over uncorrelated cycles. It results that

$$\text{Var}_0 \left(\bar{U}_n^{(m)}\right) = C_m \binom{n}{m} \alpha_n^m (1-\alpha_n)^m \asymp n^m \alpha_n^m,$$

where C_m is a constant that depends on m.

Under the alternative hypothesis, we can write

$$\bar{U}_n^{(m)} = \sum_{\substack{i_1,\ldots,i_m \\ (\text{distinct})}} (W_{i_1 i_2} + \bar{\Omega}_{i_1 i_2})(W_{i_2 i_3} + \bar{\Omega}_{i_2 i_3}) \ldots (W_{i_m i_1} + \bar{\Omega}_{i_m i_1}),$$

where $W_{ij} = A_{ij} - \Omega_{ij}$ and $\bar{\Omega}_{ij} = \Omega_{ij} - \alpha_0$ for all $i \neq j$. Then, direct calculations show that:

$$\begin{aligned}
\mathbb{E}_1 \left[\bar{U}_n^{(m)}\right] &= \sum_{\substack{i_1,\ldots,i_m \\ (\text{distinct})}} \bar{\Omega}_{i_1 i_2} \bar{\Omega}_{i_2 i_3} \ldots \bar{\Omega}_{i_m i_1} \\
&= \sum_{i_1,\ldots,i_m} \bar{\Omega}_{i_1 i_2} \bar{\Omega}_{i_2 i_3} \ldots \bar{\Omega}_{i_m i_1} - \sum_{\substack{i_1,\ldots,i_m \\ (\text{not distinct})}} \bar{\Omega}_{i_1 i_2} \bar{\Omega}_{i_2 i_3} \ldots \bar{\Omega}_{i_m i_1} \\
&= \text{Tr}(\bar{\Omega}^m) - O\left(n^{m-1} \|M\|^m\right) = \text{Tr}\left((\Pi M \Pi')^m\right) - O\left(n^{m-1} \|M\|^m\right) \\
&= n^m \text{Tr}\left((MG)^m\right) - O\left(n^{m-1} \|M\|^m\right) \asymp n^m \|MG\|^m - O\left(n^{m-1} \|M\|^m\right).
\end{aligned}$$

Since we have assumed that $\|G\|, \|G^{-1}\| < c$ by condition (3.4), we obtain that

$$\left|\mathbb{E}_1 \left[\bar{U}_n^{(m)}\right]\right| \asymp n^m \|P - \alpha_0 \mathbf{1}_K \mathbf{1}_K'\|^m. \tag{C.3}$$

□

The results in Lemma C.2 allow us to compute the SNR for the length-m Signed Cycle statistic. We derive the SNR assuming that the null variance dominates the alternative variance. Thus,

$$SNR\left(\bar{U}_n^{(m)}\right) = \frac{\left|\mathbb{E}_1 \left[\bar{U}_n^{(m)}\right] - \mathbb{E}_0 \left[\bar{U}_n^{(m)}\right]\right|}{\sqrt{\max\left\{\text{Var}_0\left(\bar{U}_n^{(m)}\right), \text{Var}_1\left(\bar{U}_n^{(m)}\right)\right\}}} \asymp \frac{\left|\mathbb{E}_1 \left[\bar{U}_n^{(m)}\right]\right|}{\sqrt{\text{Var}_0\left(\bar{U}_n^{(m)}\right)}}$$

$$\asymp \frac{n^m \|M\|^m}{n^{m/2}\alpha_n^{m/2}} = \tau_n^{m/4}.$$

Similar to our results in Theorem 3.3, there may be instances in which the alternative variance dominates the null variance. In these cases, the SNR still depends on powers of τ_n, and the detection boundary is unchanged; details are omitted.

Appendix D: Proof of Theorem 3.1

Write $\varphi_n^{DC} = (X_n - n)/\sqrt{2n}$ and $\psi_n^{SQ} = Q_n/(2\sqrt{2}n^2\hat{\alpha}_n^2)$. We aim to show that $(\psi_n^{DC}, \psi_n^{SQ})$ converges to $\mathcal{N}(0, I_2)$ in distribution. By the Cramér-Wold theorem, it suffices to show that

$$u \cdot \psi_n^{DC} + v \cdot \psi_n^{SQ} \quad \xrightarrow[n\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1), \qquad \text{for any } u, v \in \mathbb{R} \text{ with } u^2 + v^2 = 1. \tag{D.1}$$

Below, we first study the null distribution of ψ_n^{DC} and ψ_n^{SQ} respectively. These analyses produce useful intermediate results. We then use them to show the desirable claim in (D.1).

D.1. Proof of the null distribution of ψ_n^{DC}

We aim to show that

$$\varphi_n^{DC} = \frac{X_n - n}{\sqrt{2n}} \quad \xrightarrow{d} \quad \mathcal{N}(0,1). \tag{D.2}$$

First, we derive an equivalent expression of X_n. Let $\hat{T}_n = \sum_{i,j,k \text{ dist.}} (A_{ik} - \hat{\alpha}_n)(A_{jk} - \hat{\alpha}_n)$, where $\hat{\alpha}_n$ is the same as in the definition of X_n. We claim that

$$X_n = n + \frac{\hat{T}_n}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)}. \tag{D.3}$$

We now show (D.3). By definition,

$$X_n = \sum_{i=1}^n \frac{(d_i - \bar{d})^2}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)},$$

where

$$\hat{\alpha}_n = \frac{1}{n(n-1)} \mathbf{1}_n' A \mathbf{1}_n, \qquad d = A\mathbf{1}_n, \qquad \bar{d} = \frac{1}{n}\mathbf{1}_n' A \mathbf{1}_n = (n-1)\hat{\alpha}_n.$$

We expand X_n into a sum of two terms that can be easily studied:

$$X_n = \frac{\|d\|_2^2 - n\bar{d}^2}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)} = \frac{\mathbf{1}_n' A^2 \mathbf{1}_n}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)} - \frac{n(n-1)\hat{\alpha}_n}{1-\hat{\alpha}_n}.$$

We can compute $\mathbf{1}_n' A^2 \mathbf{1}_n$ as follows:

$$\mathbf{1}_n' A^2 \mathbf{1}_n = \sum_{i,j} (A^2)_{ij} = \mathbf{1}_n' A \mathbf{1}_n + \sum_{i,j,k \text{ dist.}} A_{ik}A_{jk} = n(n-1)\hat{\alpha}_n + \sum_{i,j,k \text{ dist.}} A_{ik}A_{jk}.$$

Hence we further reexpress X_n as

$$X_n = \frac{\sum_{i,j,k \text{ dist.}} A_{ik}A_{jk}}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)} + \frac{n - n(n-1)\hat{\alpha}_n}{1-\hat{\alpha}_n}.$$

Recalling that $\hat{T}_n = \sum_{i,j,k \text{ dist.}}(A_{ik} - \hat{\alpha}_n)(A_{jk} - \hat{\alpha}_n)$, we have

$$\begin{aligned}\sum_{i,j,k \text{ dist.}} A_{ik}A_{jk} &= \hat{T}_n + 2(n-2)\hat{\alpha}_n \mathbf{1}_n' A \mathbf{1}_n - n(n-1)(n-2)\hat{\alpha}_n^2 \\ &= \hat{T}_n + n(n-1)(n-2)\hat{\alpha}_n^2.\end{aligned}$$

It follows that

$$X_n - n = \frac{\hat{T}_n + n(n-1)(n-2)\hat{\alpha}_n^2}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)} + \frac{n - n(n-1)\hat{\alpha}_n}{1-\hat{\alpha}_n} - n = \frac{\hat{T}_n}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)}.$$

This proves (D.3).

Next, we introduce an ideal counterpart to $\hat{T}_n$, $T_n = \sum_{i,j,k \text{ dist.}}(A_{ik} - \alpha_n)(A_{jk} - \alpha_n)$. Direct computations show that

$$\mathbb{E}[T_n] = 0, \qquad \text{Var}(T_n) = 2n(n-1)(n-2)\alpha_n^2(1-\alpha_n)^2.$$

Thus

$$\text{Var}\left(\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}\right) = \frac{2n(n-2)}{n-1}.$$

Combining it with (D.3), we obtain

$$\frac{X_n - n}{\sqrt{2n}} = \left(\frac{\alpha_n(1-\alpha_n)}{\hat{\alpha}_n(1-\hat{\alpha}_n}\right)\left(\frac{\hat{T}_n}{T_n}\right)\left(\frac{n-2}{n-1}\right)^{1/2}\left(\frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}}\right).$$

Define

$$U_n = \frac{\alpha_n(1-\alpha_n)}{\hat{\alpha}_n(1-\hat{\alpha}_n)}, \qquad V_n = \frac{\hat{T}_n}{T_n}, \qquad Z_n = \frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}}.$$

We have the following decomposition:

$$\frac{X_n - n}{\sqrt{2n}} = \left(\frac{n-2}{n-1}\right)^{1/2} U_n V_n Z_n. \tag{D.4}$$

Below, we study U_n, V_n, and Z_n, separately.

Consider U_n. Note that

$$\hat{\alpha}_n = \frac{1}{n(n-1)}\mathbf{1}_n' A \mathbf{1}_n = \frac{2}{n(n-1)}\sum_{i<j} A_{ij},$$

where $(A_{ij})_{i<j}$ are i.i.d. Bernoulli random variables with mean α_n. By the Weak Law of Large Numbers we obtain that

$$\frac{\hat{\alpha}_n}{\alpha_n} = \frac{2}{n(n-1)} \sum_{i<j} \frac{A_{ij}}{\alpha_n} \xrightarrow{\mathbb{P}} 1, \tag{D.5}$$

from which we conclude that $U_n \xrightarrow{\mathbb{P}} 1$.

Consider V_n. Note that

$$\begin{aligned}
\hat{T}_n - T_n &= \sum_{i,j,k \text{ dist.}} (A_{ik} - \hat{\alpha}_n)(A_{jk} - \hat{\alpha}_n) - \sum_{i,j,k \text{ dist.}} (A_{ik} - \alpha_n)(A_{jk} - \alpha_n) \\
&= \sum_{i,j,k \text{ dist.}} (\alpha_n - \hat{\alpha}_n)(A_{ik} + A_{jk} - \alpha_n - \hat{\alpha}_n) \\
&= (\alpha_n - \hat{\alpha}_n) \left[2 \left(\sum_{i,j,k \text{ dist.}} A_{ik} \right) - n(n-1)(n-2)(\alpha_n + \hat{\alpha}_n) \right] \\
&= (\alpha_n - \hat{\alpha}_n) \left[2(n-2) \mathbf{1}_n' A \mathbf{1}_n - n(n-1)(n-2)(\alpha_n + \hat{\alpha}_n) \right] \\
&= (\alpha_n - \hat{\alpha}_n) \left[2n(n-1)(n-2)\hat{\alpha}_n - n(n-1)(n-2)(\alpha_n + \hat{\alpha}_n) \right] \\
&= -n(n-1)(n-2)(\alpha_n - \hat{\alpha}_n)^2.
\end{aligned}$$

It follows that

$$\begin{aligned}
\left| \frac{\hat{T}_n - T_n}{T_n} \right| &= \left| \frac{n(n-1)(n-2)(\alpha_n - \hat{\alpha}_n)^2}{T_n} \right| \\
&= \sqrt{\frac{2(n-2)}{n(n-1)}} \left(\sqrt{\frac{n(n-1)}{2}} \frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}} \right)^2 \left| \frac{\sqrt{\frac{2n(n-2)}{n-1}}}{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}} \right| \\
&= \sqrt{\frac{n-2}{2(n-1)}} \left(\sqrt{\frac{n(n-1)}{2}} \frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}} \right)^2 \frac{1}{\sqrt{n}|Z_n|}.
\end{aligned}$$

Note that $\hat{\alpha}_n = \frac{2}{n(n-1)} \sum_{i<j} A(i,j)$, where A_{ij} are i.i.d. Bernoulli random variables with mean α_n. By the Central Limit Theorem,

$$\sqrt{\frac{n(n-1)}{2}} \frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

We will show later that $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. It follows that $(\sqrt{n}|Z_n|)^{-1} \xrightarrow{\mathbb{P}} 0$ (by Slutsky's theorem) and we conclude by Slutsky's theorem again that

$$\left| \frac{\hat{T}_n - T_n}{T_n} \right| \xrightarrow{\mathbb{P}} 0, \tag{D.6}$$

which shows that $V_n \xrightarrow{\mathbb{P}} 1$.

Consider Z_n. We define

$$I_m = \{(i,j,k) \in [\![1,m]\!]^3 \text{ s.t. } i,j,k \text{ are distinct}\},$$

and the following quantities for $m \in [\![1,n]\!]$

$$T_{n,m} = \sum_{(i,j,k)\in I_m} W_{ik}W_{jk}, \qquad \text{and} \qquad T_{n,0} = 0,$$

$$Z_{n,m} = \sqrt{\frac{n-1}{2n(n-2)}}\frac{T_{n,m}}{(n-1)\alpha_n(1-\alpha_n)}, \qquad \text{and} \qquad Z_{n,0} = 0.$$

Consider the filtration $\{\mathcal{F}_{n,m}\}_{1\le m\le n}$ with $\mathcal{F}_{n,m} = \sigma\{W_{ij}, (i,j) \in [\![1,m]\!]^2\}$ for all $m \in [\![1,n]\!]$, $\mathcal{F}_{n,0} = \{\Omega, \emptyset\}$ (where Ω denotes the sample space). It is straightforward to see that for all $0 \le m \le n$, $Z_{n,m}$ is $\mathcal{F}_{n,m}$-measurable, $\mathbb{E}[|Z_{n,m}|] < \infty$ and $\mathbb{E}\left[T_{n,m+1}|\mathcal{F}_{n,m}\right] = T_{n,m}$. This shows that $\{Z_{n,m}\}_{1\le m\le n}$ is a martingale with respect to $\{\mathcal{F}_{n,m}\}_{1\le m\le n}$. Define the martingale difference sequence, for all $m = 1, ..., n$

$$X_{n,m} = Z_{n,m} - Z_{n,m-1}.$$

With these notations we have $Z_n \equiv Z_{n,n} = \sum_{m=1}^n X_{n,m}$. Provided the following two conditions are met

$$\text{(a)} \sum_{m=1}^n \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 1, \tag{D.7}$$

$$\text{(b)} \ \forall \epsilon > 0, \sum_{m=1}^n \mathbb{E}[X_{n,m}^2 \mathbf{1}\{|X_{n,m} > \epsilon|\}|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 0, \tag{D.8}$$

we conclude using the Martingale Central Limit Theorem that $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

So far, we have shown that $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, $U_n \xrightarrow{\mathbb{P}} 1$ and $V_n \xrightarrow{\mathbb{P}} 1$. We plug them into (D.4). Then, (D.2) follows immediately from Slutsky's theorem.

The only remaining steps are to verify that (D.7) and (D.8) are indeed satisfied.

Proof of Equation (D.7)**:** It suffices to show that

$$\mathbb{E}\left[\sum_{m=1}^n \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right] = 1, \tag{D.9}$$

and

$$\text{Var}\left(\sum_{m=1}^n \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right) \xrightarrow{n\to\infty} 0. \tag{D.10}$$

First, we prove Equation (D.9). For notational convenience we write

$$C_n := (n-1)\alpha_n(1-\alpha_n)\sqrt{\frac{2n(n-2)}{n-1}}.$$

It follows that for all $n \in \mathbb{N}^*$ and $m \in [\![1, n]\!]$

$$C_n X_{n,m} = C_n (Z_{n,m} - Z_{n,m-1}) = T_{n,m} - T_{n,m-1} = \sum_{(i,j,k) \in I_m \setminus I_{m-1}} W_{ik} W_{jk}.$$

Triplets in $I_m \setminus I_{m-1}$ are such that one of the nodes is m: either one of the wingnodes $\{i, j\}$, or the centernode k. Hence,

$$C_n X_{n,m} = 2 \sum_{\substack{1 \le j,k \le m-1 \\ j \ne k}} W_{mk} W_{jk} + \sum_{\substack{1 \le i,j \le m-1 \\ i \ne j}} W_{im} W_{jm}. \tag{D.11}$$

As a result (in the following, summations are all up to $m - 1$)

$$C_n^2 X_{n,m}^2 = 4 \sum_{\substack{k \ne j \\ i \ne l}} W_{mk} W_{jk} W_{mi} W_{il} + 4 \sum_{\substack{k \ne j \\ i \ne l}} W_{mk} W_{jk} W_{im} W_{lm} + \sum_{\substack{i \ne j \\ k \ne l}} W_{im} W_{jm} W_{km} W_{lm}.$$

It follows that

$$\begin{aligned}
\mathbb{E}[C_n^2 X_{n,m}^2 | \mathcal{F}_{n,m-1}] = & 4 \sum_{k \ne j;\, i \ne l} W_{jk} W_{il} \mathbb{E}\left[W_{mk} W_{mi}\right] + 4 \sum_{k \ne j;\, i \ne l} W_{jk} \mathbb{E}\left[W_{im} W_{km} W_{lm}\right] \\
& + \sum_{i \ne j;\, k \ne l} \mathbb{E}\left[W_{im} W_{jm} W_{km} W_{lm}\right] \\
= & 4\alpha_n (1 - \alpha_n) \sum_i \sum_{j \ne i, l \ne i} W_{ij} W_{il} + 2(m-1)(m-2)\alpha_n^2 (1 - \alpha_n)^2 \\
= & 4\alpha_n (1 - \alpha_n) \sum_{(i,j,l) \in I_{m-1}} W_{ij} W_{il} + 4\alpha_n (1 - \alpha_n) \sum_{i \ne j} W_{ij}^2 \\
& + 2(m-1)(m-2)\alpha_n^2 (1 - \alpha_n)^2 \\
= & 4\alpha_n (1 - \alpha_n) \Big(T_{n,m-1} + \sum_{i \ne j} W_{ij}^2 \Big) + 2(m-1)(m-2)\alpha_n^2 (1 - \alpha_n)^2.
\end{aligned} \tag{D.12}$$

Let $\mathbf{1}_{n,m} \in \mathbb{R}^n$ be a vector whose m first entries are 1, and whose remaining entries are 0. Define

$$\hat{\alpha}_{n,m} := \frac{\mathbf{1}_{n,m}' A \mathbf{1}_{n,m}}{m(m-1)}.$$

By direct calculations,

$$\begin{aligned}
\sum_{i \ne j} W_{ij}^2 = \sum_{i \ne j} (A_{ij} - \alpha_n)^2 = \sum_{i \ne j} \Big[A_{ij} (1 - 2\alpha_n) + \alpha_n^2 \Big] \\
= (m-1)(m-2)\alpha_n^2 + (1 - 2\alpha_n) \sum_{i,j} A_{ij} \\
= (m-1)(m-2)\alpha_n^2 + (1 - 2\alpha_n)(m-1)(m-2)\hat{\alpha}_{n,m-1}.
\end{aligned}$$

We plug the above equation into (D.12) to get

$$\begin{aligned}\mathbb{E}[C_n^2 X_{n,m}^2|\mathcal{F}_{n,m-1}] = 4\alpha_n(1-\alpha_n)T_{n,m-1} + 2(m-1)(m-2)\alpha_n^2(1-\alpha_n)^2 \\ + 4(m-1)(m-2)\alpha_n^3(1-\alpha_n) \\ + 4(m-1)(m-2)\alpha_n\hat{\alpha}_{n,m-1}(1-\alpha_n)(1-2\alpha_n). \end{aligned} \tag{D.13}$$

It follows that

$$\begin{aligned} C_n^2 \sum_{m=1}^n \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] &= 4\alpha_n(1-\alpha_n)\sum_{m=1}^n T_{n,m-1} \\ &+ \Big[2\alpha_n^2(1-\alpha_n)^2 + 4\alpha_n^3(1-\alpha_n)\Big]\sum_{m=1}^n (m-1)(m-2) \\ &+ 4\alpha_n(1-\alpha_n)(1-2\alpha_n)\sum_{m=1}^n (m-1)(m-2)\hat{\alpha}_{n,m-1}. \end{aligned}$$

Recall that $\mathbb{E}[T_{n,m-1}] = \sum_{(i,j,k)\in I_{m-1}} \mathbb{E}[W_{ik}W_{jk}] = \sum_{(i,k)\in I_{m-1}} \mathbb{E}[W_{ik}^2] = \frac{(m-1)(m-2)}{2}\alpha_n(1-\alpha_n)$. Additionally, $\mathbb{E}[\hat{\alpha}_{n,m-1}] = \alpha_n$. We thus have

$$\begin{aligned} C_n^2\mathbb{E}\left[\sum_{m=1}^n \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right] &= 2\alpha_n^2(1-\alpha_n)^2\sum_{m=1}^n (m-1)(m-2) \\ &+ \Big[2\alpha_n^2(1-\alpha_n)^2 + 4\alpha_n^3(1-\alpha_n)\Big]\sum_{m=1}^n (m-1)(m-2) \\ &+ 4\alpha_n(1-\alpha_n)(1-2\alpha_n)\sum_{m=1}^n \alpha_n(m-1)(m-2) \\ &= 6\alpha_n^2(1-\alpha_n)^2\sum_{m=1}^n (m-1)(m-2) \\ &= 2\alpha_n^2(1-\alpha_n)^2 n(n-1)(n-2) = C_n^2. \end{aligned}$$

This proves (D.9).

Second, we prove Equation (D.10). In the second line of (D.12), we have seen that

$$\begin{aligned} C_n^2\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] &= 4\alpha_n(1-\alpha_n)\sum_k \sum_{\substack{1\le i\ne j\le m-1\\ i\ne k, j\ne k}} W_{ki}W_{kj} + 2(m-1)(m-2)\alpha_n^2(1-\alpha_n)^2 \\ &= 8\alpha_n(1-\alpha_n)\sum_k \sum_{\substack{1\le i< j\le m-1\\ i\ne k, j\ne k}} W_{ki}W_{kj} + 2(m-1)(m-2)\alpha_n^2(1-\alpha_n)^2. \end{aligned}$$

As a result,

$$\mathrm{Var}\left(C_n^2\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right)\leq 64\alpha_n^2\mathrm{Var}\left(\sum_{m=1}^{n}\sum_{k}\sum_{\substack{1\leq i<j\leq m-1\\ i\neq k,j\neq k}}W_{ki}W_{kj}\right).$$

Recall that in the previous sums, summation over the indices i,j,k ranges from 1 to $m-1$. We rearrange the terms of the sums in order to facilitate the computation of the variance. Instead of summing over the order m, then over centernodes k ranging from 1 to $m-1$, and finally over wingnodes i,j also ranging from 1 to $m-1$, we now sum over centernodes k ranging from 1 to $n-1$, wingnodes ranging from 1 to $n-1$, and finally over orders $m>\max(i,j,k)$.

$$\mathrm{Var}\left(C_n^2\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right)\leq 64\alpha_n^2\mathrm{Var}\left(\sum_{k=1}^{n-1}\sum_{\substack{1\leq i<j\leq n-1\\ i\neq k,j\neq k}}\sum_{m>\max(i,j,k)}W_{ki}W_{kj}\right)$$

$$\leq 64\alpha_n^2n^2\mathrm{Var}\left(\sum_{k=1}^{n-1}\sum_{\substack{1\leq i<j\leq n-1\\ i\neq k,j\neq k}}W_{ki}W_{kj}\right)=64\alpha_n^2n^2\sum_{k=1}^{n-1}\mathrm{Var}\left(\sum_{\substack{1\leq i<j\leq n-1\\ i\neq k,j\neq k}}W_{ki}W_{kj}\right),$$

where the last equality comes from the fact that in the above sum, terms corresponding to different values of the index k are uncorrelated. As a result

$$\mathrm{Var}\left(C_n^2\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right)\leq 64\alpha_n^2n^2\sum_{k=1}^{n-1}\sum_{\substack{1\leq i<j\leq n-1\\ 1\leq u<v\leq n-1\\ i,j,u,v\neq k}}\mathrm{Cov}(W_{ki}W_{kj},W_{ku}W_{kv}). \tag{D.14}$$

We examine the possible cases for $\mathrm{Cov}(W_{ki}W_{kj},W_{ku}W_{kv})$.

- Case 1: $(i,j)=(u,v)$, then $\mathrm{Cov}(W_{ki}W_{kj},W_{ku}W_{kv})=\mathrm{Var}(W_{ki}W_{kj})=\alpha_n^2(1-\alpha_n)^2$.
- Case 2: $i=u,j\neq v$ or $i\neq u,j=v$, then $\mathrm{Cov}(W_{ki}W_{kj},W_{ku}W_{kv})=0$.
- All other cases: $\mathrm{Cov}(W_{ki}W_{kj},W_{ku}W_{kv})=0$.

It follows that

$$\mathrm{Var}\left(C_n^2\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right)\leq 64\alpha_n^2n^2\sum_{k=1}^{n-1}\sum_{\substack{1\leq i<j\leq n-1\\ i,j\neq k}}\mathrm{Var}(W_{ki}W_{kj})$$

$$=64\alpha_n^2n^2\sum_{k=1}^{n-1}\sum_{\substack{1\leq i<j\leq n-1\\ i,j\neq k}}\alpha_n^2(1-\alpha_n)^2\leq 32\alpha_n^4n^5.$$

Hence

$$\mathrm{Var}\left(\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right)\leq\frac{32\alpha_n^4n^5}{C_n^4}=\frac{1}{n}\left(\frac{n^4}{(n-1)^2(n-2)^2}\right)\left(\frac{8}{(1-\alpha_n)^4}\right)\xrightarrow[n\to\infty]{}0.$$

This proves (D.10).

Proof of Equation (D.8)**:** Notice that by the Cauchy-Schwarz and Markov inequalities we obtain the following upper bound

$$\begin{aligned}\left|\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2\mathbf{1}\{|X_{n,m}>\epsilon|\}|\mathcal{F}_{n,m-1}]\right| &\le \sum_{m=1}^{n}\sqrt{\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]}\sqrt{\mathbb{P}(|X_{n,m}|>\epsilon|\mathcal{F}_{n,m-1})}\\ &\le \frac{1}{\epsilon^2}\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}].\end{aligned}$$

Thus it suffices to show that $\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]\xrightarrow{\mathbb{P}}0$. Since these random variables are all non-negative, we will equivalently show that

$$\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^4]=\mathbb{E}\left[\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]\right]\xrightarrow{n\to\infty}0. \tag{D.15}$$

We now show (D.15). Recall that (see (D.11))

$$\begin{aligned}C_nX_{n,m}&=2\sum_{1\le i<j\le m-1}W_{im}W_{jm}+2\sum_{1\le i<j\le m-1}W_{ij}W_{jm}\\ &=2\sum_{1\le i<j\le m-1}W_{jm}(W_{ij}+W_{im}).\end{aligned}$$

Then (with summations ranging from 1 to $m-1$)

$$C_n^4X_{n,m}^4=16\sum_{\substack{i<j\\u<v\\k<l\\r<s}}W_{jm}(W_{ij}+W_{im})W_{vm}(W_{uv}+W_{um})W_{lm}(W_{kl}+W_{km})W_{sm}(W_{rs}+W_{rm}).$$

Taking expectations, we consider 4 types of cases in which the expectation is non-zero:

- Case 1: $i=u=k=r$ and $j=v=l=s$ (1 instance),
- Case 2: $i=k,u=r$ with $i\neq u$ and $j=l,v=s$ with $j\neq v$ (3 instances),
- Case 3: $i=u=k=r$ and $j=l,v=s$ with $j\neq v$ (3 instances),
- Case 4: $i=k,u=r$ with $i\neq u$ and $j=v=l=s$ (3 instances),
- Other cases: $\mathbb{E}[W_{jm}(W_{ij}+W_{im})W_{vm}(W_{uv}+W_{um})W_{lm}(W_{kl}+W_{km})W_{sm}(W_{rs}+W_{rm}]=0$.

It follows that

$$\begin{aligned}\mathbb{E}[C_n^4X_{n,m}^4]=16\Bigg[&\sum_{i<j}\mathbb{E}[W_{jm}^4]\mathbb{E}[(W_{ij}+W_{im})^4]\\ &+3\sum_{\substack{i<j,u<v\\i\neq u,j\neq v}}\mathbb{E}[W_{jm}^2]\,\mathbb{E}[(W_{ij}+W_{im})^2]\,\mathbb{E}[W_{vm}^2]\,\mathbb{E}[(W_{ij}+W_{im})^2]\end{aligned}$$

$$+3\sum_{\substack{i<j,v\\ j\neq v}}\mathbb{E}[W_{jm}^2]\,\mathbb{E}[W_{vm}^2]\,\mathbb{E}[(W_{ij}+W_{im})^2(W_{iv}+W_{im})^2]$$

$$+3\sum_{\substack{i,u<j\\ i\neq u}}\mathbb{E}[(W_{ij}+W_{im})^2]\,\mathbb{E}[(W_{uj}+W_{um})^2]\,\mathbb{E}[W_{jm}^4]\Bigg].$$

We provide upper bounds for the above expectations. Indeed for all $(a,b)\in[\![1,n]\!]^2$

$$\mathbb{E}[W_{ab}^4]=(1-\alpha_n)^4\alpha_n+\alpha_n^4(1-\alpha_n)=\alpha_n(1-\alpha_n)(\alpha_n^3+(1-\alpha_n)^3).$$

It is then straightforward to show, taking $c_*>0$ to be a high enough constant, that

$$\mathbb{E}[W_{jm}^4]\le c_*\alpha_n,\qquad \mathbb{E}[(W_{ij}+W_{im})^4]\le c_*\alpha_n,\qquad \mathbb{E}[W_{jm}^2]^2\le c_*\alpha_n^2,$$
$$\mathbb{E}[(W_{ij}+W_{im})^2]^2\le c_*\alpha_n^2,\qquad \mathbb{E}[(W_{ij}+W_{im})^2(W_{iv}+W_{im})^2]\le c_*\alpha_n.$$

It follows that

$$\begin{aligned}\mathbb{E}[C_n^4X_{n,m}^4]&\le 16\left(\sum_{i<j}c_*^2\alpha_n^2+3\sum_{\substack{i<j,u<v\\ i\neq u,j\neq v}}c_*^2\alpha_n^4+3\sum_{\substack{i<j,v\\ j\neq v}}c_*^2\alpha_n^3+3\sum_{\substack{i,u<j\\ i\neq u}}c_*^2\alpha_n^3\right)\\ &\le 16c_*^2n^2\alpha_n^2\left(1+3n^2\alpha_n^2+6n\alpha_n\right)\\ &=O(n^4\alpha_n^4),\end{aligned}\tag{D.16}$$

where in the last line we have used the assumption of $n\alpha_n\to\infty$ to identify the dominating term. Note that C_n is at the order of $n\sqrt{n}\alpha_n$. We thus obtain

$$\mathbb{E}\left[\sum_{m=1}^n\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]\right]=n\cdot O\left(\frac{n^4\alpha_n^4}{(n\sqrt{n}\alpha_n)^4}\right)=O\left(n^{-1}\right).$$

This proves (D.15).

D.2. Proof of the null distribution of ψ_n^{SQ}

We aim to show that

$$\varphi_n^{SQ}=\frac{Q_n}{2\sqrt{2}n^2\hat{\alpha}_n^2}\quad\xrightarrow[n\to\infty]{\mathcal{L}}\quad\mathcal{N}(0,1).\tag{D.17}$$

Let $\hat{\delta}_n=\alpha_n-\hat{\alpha}_n$. We then have $A_{ij}-\hat{\alpha}_n=W_{ij}+\hat{\delta}_n$. It follows that $Q_n=\sum_{(i_1,i_2,i_3,i_4)\text{ dist.}}(W_{i_1i_2}+\hat{\delta}_n)(W_{i_2i_3}+\hat{\delta}_n)(W_{i_3i_4}+\hat{\delta}_n)(W_{i_4i_1}+\hat{\delta}_n)$. We introduce an ideal version of Q_n,

$$\tilde{Q}_n=\sum_{(i_1,i_2,i_3,i_4)\text{ dist.}}W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1},$$

and re-express

$$\psi_n^{SQ} = \frac{Q_n}{2\sqrt{2}n^2\hat{\alpha}_n^2} = \frac{Q_n - \tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}\left(\frac{\alpha_n}{\hat{\alpha}_n}\right)^2 + \frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}\left(\frac{\alpha_n}{\hat{\alpha}_n}\right)^2. \tag{D.18}$$

If we can show that

$$\text{(a) } \frac{Q_n - \tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2} \xrightarrow{\mathbb{P}} 0, \tag{D.19}$$

$$\text{(b) } \frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2} \to_d \mathcal{N}(0,1), \tag{D.20}$$

then (D.17) follows from Slutsky's theorem and the fact that $\hat{\alpha}_n/\alpha_n \xrightarrow{\mathbb{P}} 1$.
What remains is to prove (D.19) and (D.20).

Proof of Equation (D.19): Expanding Q_n, we obtain:

$$\begin{aligned} Q_n - \tilde{Q}_n =& n(n-1)(n-2)(n-3)\hat{\delta}_n^4 + 4(n-2)(n-3)\hat{\delta}_n^3 \sum_{i\neq j} W_{ij} \\ &+ 4(n-3)\hat{\delta}_n^2 \sum_{i,j,k \text{ dist.}} W_{ij}W_{jk} + 2\hat{\delta}_n^2 \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl} \\ &+ 4\hat{\delta}_n \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}. \end{aligned}$$

It follows that

$$\begin{aligned} \left|\frac{Q_n - \tilde{Q}_n}{n^2\alpha_n^2}\right| \leq & n^2\frac{\hat{\delta}_n^4}{\alpha_n^2} + 4\frac{|\hat{\delta}_n|^3}{\alpha_n^2}\left|\sum_{i\neq j} W_{ij}\right| + \frac{4|\hat{\delta}_n|}{n^2\alpha_n^2}\left|\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}\right| \\ &+ \frac{\hat{\delta}_n^2}{\alpha_n^2}\left(\frac{4}{n}\left|\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\right| + \frac{2}{n^2}\left|\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\right|\right). \end{aligned} \tag{D.21}$$

We will bound each of the terms on the right hand side of (D.21).
Consider the first term in (D.21). Note that

$$n^2\frac{\hat{\delta}_n^4}{\alpha_n^2} = 4\frac{(1-\alpha_n)^2}{(n-1)^2}\left(\sqrt{\frac{n(n-1)}{2}}\frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}}\right)^4.$$

By Central Limit Theorem, $\sqrt{\frac{n(n-1)}{2}}\frac{\hat{\alpha}_n-\alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}} \to \mathcal{N}(0,1)$. It follows from Slutsky's theorem that

$$n^2\hat{\delta}_n^4/\alpha_n^2 \xrightarrow{\mathbb{P}} 0. \tag{D.22}$$

Consider the second term in (D.21). Since $\hat{\delta}_n = \hat{\alpha}_n - \alpha_n$, using the definition of $\hat{\alpha}_n$, we immediately have $\sum_{i<j} W_{ij} = \frac{n(n-1)}{2}\hat{\delta}_n$. As a result,

$$\frac{|\hat{\delta}_n^3|}{\alpha_n^2}\Big|\sum_{i\neq j} W_{ij}\Big| = n(n-1)\frac{\hat{\delta}_n^4}{\alpha_n^2} \leq n^2\frac{\hat{\delta}_n^4}{\alpha_n^2} \xrightarrow{\mathbb{P}} 0. \tag{D.23}$$

Consider the fourth term in (D.21). First, let $A_n = \frac{1}{n^3\alpha_n}\sum_{i,j,k \text{ dist.}} W_{ij}W_{ik}$. Applying Chebyshev's inequality, we have that for any $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(|A_n| > \lambda) &\leq \frac{\mathbb{E}[A_n^2]}{\lambda^2} \leq \frac{6^2}{n^6\alpha_n^2\lambda^2}\sum_{\substack{i<j<k\\u<v<w}} \mathbb{E}[W_{ij}W_{jk}W_{uv}W_{vw}] \\ &= \frac{36}{n^6\alpha_n^2\lambda^2}\sum_{i<j<k}\mathbb{E}[W_{ij}^2W_{jk}^2] \leq \frac{36}{n^3\lambda^2} \xrightarrow{n\to\infty} 0,\end{aligned}$$

which shows that $A_n \xrightarrow{\mathbb{P}} 0$. Furthermore,

$$\frac{\hat{\delta}_n^2}{n\alpha_n^2}\Big|\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\Big| = 2(1-\alpha_n)\Big(\frac{n}{n-1}\Big)\left[\sqrt{\frac{n(n-1)}{2}}\frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}}\right]^2 |A_n|.$$

By Slutsky's theorem, we have

$$\frac{\hat{\delta}_n^2}{n\alpha_n^2}\Big|\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\Big| \xrightarrow{\mathbb{P}} 0. \tag{D.24}$$

Second, let $B_n = \frac{1}{\alpha_n n^4}\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}$. We apply Chebyshev's inequality: For any $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(|B_n| > \lambda) &\leq \frac{\mathbb{E}[B_n^2]}{\lambda^2} \leq \frac{1}{\alpha_n^2 n^8\lambda^2}\sum_{\substack{i,j,k,l \text{ dist.}\\ s,t,u,v \text{ dist.}}} \mathbb{E}[W_{ij}W_{kl}W_{st}W_{uv}] \\ &= \frac{24^2}{\alpha_n^2 n^8\lambda^2}\sum_{\substack{i<j<k<l\\s<t<u<v}} \mathbb{E}[W_{ij}W_{kl}W_{st}W_{uv}] = \frac{24^2}{\alpha_n^2 n^8\lambda^2}\sum_{i<j<k<l}\mathbb{E}[W_{ij}^2W_{kl}^2] \\ &= \frac{24^2}{\alpha_n^2 n^8\lambda^2}\sum_{i<j<k<l}\mathbb{E}[W_{ij}^2]\mathbb{E}[W_{kl}^2] \leq \frac{24^2}{n^4\lambda^2} \xrightarrow{n\to\infty} 0,\end{aligned}$$

which shows that $B_n \xrightarrow{\mathbb{P}} 0$. Furthermore,

$$\frac{\hat{\delta}_n^2}{n^2\alpha_n^2}\Big|\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\Big| = \frac{2(1-\alpha_n)n}{n-1}\left[\sqrt{\frac{n(n-1)}{2}}\frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}}\right]^2 |B_n|.$$

We conclude by Slutsky's theorem that

$$\frac{\hat{\delta}_n^2}{n^2\alpha_n^2}\Big|\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\Big| \xrightarrow{\mathbb{P}} 0. \tag{D.25}$$

Consider the third term in (D.21). Write $D_n = \frac{1}{\alpha_n^{3/2} n^3} \sum_{i,j,k,l \text{ dist.}} W_{ij} W_{jk} W_{kl}$. By Chebyshev's inequality, for any $\lambda > 0$,

$$\mathbb{P}(|D_n| > \lambda) \leq \frac{\mathbb{E}[D_n^2]}{\lambda^2} = \frac{1}{\alpha_n^3 n^6 \lambda^2} \mathbb{E}\Bigg[\sum_{\substack{i,j,k,l \text{ dist.} \\ u,v,w,z \text{ dist.}}} W_{ij} W_{jk} W_{kl} W_{uv} W_{vw} W_{wz} \Bigg]$$

$$= \frac{2}{\alpha_n^3 n^6 \lambda^2} \mathbb{E}\Bigg[\sum_{i,j,k,l \text{ dist.}} W_{ij}^2 W_{jk}^2 W_{kl}^2 \Bigg] \leq \frac{2}{n^2 \lambda^2} \xrightarrow[n \to \infty]{} 0,$$

which implies that $D_n \xrightarrow{\mathbb{P}} 0$. Furthermore,

$$\frac{\hat{\delta}_n}{n^2 \alpha_n^2} \Bigg| \sum_{i,j,k,l \text{ dist.}} W_{ij} W_{jk} W_{kl} \Bigg| = \sqrt{\frac{2n(1-\alpha_n)}{n-1}} \Bigg[\sqrt{\frac{n(n-1)}{2}} \frac{\hat{\alpha}_n - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}} \Bigg] |D_n|.$$

We conclude by Slutsky's theorem that

$$\frac{\hat{\delta}_n}{n^2 \alpha_n^2} \Bigg| \sum_{i,j,k,l \text{ dist.}} W_{ij} W_{jk} W_{kl} \Bigg| \xrightarrow{\mathbb{P}} 0. \tag{D.26}$$

We plug (D.22)-(D.26) into (D.21) to get (D.19).

Proof of Equation (D.20)**:** We introduce some notation to simplify the computations. Given 4 distinct nodes, there are 3 different possible cycles, denoted as

$$CC(i_1, i_2, i_3, i_4) = \{(i_1, i_2, i_3, i_4), (i_1, i_2, i_4, i_3), (i_1, i_3, i_2, i_4)\}.$$

Moreover, for $B \subset \{1, 2, ..., n\}^4$, let $CC(B) = \cup_{(i_1,i_2,i_3,i_4) \in B} CC(i_1, i_2, i_3, i_4)$. For $1 \leq m \leq n$, let I_m be the collection of (i_1, i_2, i_3, i_4) such that $1 \leq i_1 < i_2 < i_3 < i_4 \leq m$. We thus have

$$\tilde{Q}_n = 8 \sum_{CC(I_n)} W_{i_1 i_2} W_{i_2 i_3} W_{i_3 i_4} W_{i_4 i_1}. \tag{D.27}$$

It is straightforward to see that $\mathbb{E}[\tilde{Q}_n] = 0$. In addition, notice that the terms in the sum are uncorrelated, since they all correspond to different cycles: to obtain a non-zero correlation between $W_{i_1 i_2} W_{i_2 i_3} W_{i_3 i_4} W_{i_4 i_1}$ and $W_{i'_1 i'_2} W_{i'_2 i'_3} W_{i'_3 i'_4} W_{i'_4 i'_1}$, we would need to uniquely match each factor in $W_{i_1 i_2} W_{i_2 i_3} W_{i_3 i_4} W_{i_4 i_1}$ with a factor in $W_{i'_1 i'_2} W_{i'_2 i'_3} W_{i'_3 i'_4} W_{i'_4 i'_1}$, which is equivalent to overlaying the two cycles $[i_1 i_2 i_3 i_4]$ and $[i'_1 i'_2 i'_3 i'_4]$. Let's compute the variance

$$\text{Var}(\tilde{Q}_n) = 64 \text{Var}\Bigg(\sum_{CC(I_n)} W_{i_1 i_2} W_{i_2 i_3} W_{i_3 i_4} W_{i_4 i_1} \Bigg)$$

$$= 64 \alpha_n^4 (1-\alpha_n)^4 \times 3 \binom{n}{4} = 8 \alpha_n^4 (1-\alpha_n)^4 n(n-1)(n-2)(n-3).$$

Let $Z_n := 2\sqrt{2n(n-1)(n-2)(n-3)}\alpha_n^2(1-\alpha_n)^2$. It is easy to see that $n^2\alpha_n^2/Z_n \xrightarrow{n\to\infty} 1$. By Slutsky's theorem, to show (D.20), it suffices to show that

$$\frac{\tilde{Q}_n}{Z_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1). \tag{D.28}$$

We now prove (D.28). For each $1 \le m \le n$, we define

$$X_{n,m} = \frac{\tilde{Q}_{n,m} - \tilde{Q}_{n,m-1}}{Z_n}, \qquad \text{where} \quad \tilde{Q}_{n,m} = \sum_{CC(I_m)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}.$$

By default, we let $\tilde{Q}_{n,0} = 1$. Recall that we previously defined the filtration $\{\mathcal{F}_{n,m} : 0 \le m \le n\}$ such that $\mathcal{F}_{n,m} = \sigma\{W_{ij} : (i,j) \in [\![1,m]\!]^2\}$ for $m \ge 1$ and $\mathcal{F}_{n,0} = \{\Omega, \emptyset\}$ (where Ω denotes the sample space). It is easy to see that $\mathbb{E}[|\tilde{Q}_{n,m}|] < \infty$. Hence, $\tilde{Q}_{n,m}$ is $\mathcal{F}_{n,m}$-measurable. It is also straightforward to show that $\mathbb{E}[\tilde{Q}_{n,m+1}|\mathcal{F}_{n,m}] = \tilde{Q}_{n,m}$. Therefore, the sequence $\{Q_{n,m} : m \in [\![1,n]\!]\}$ is a martingale with respect to $\{\mathcal{F}_{n,m} : m \in [\![1,n]\!]\}$. It follows that the sequence $\{X_{n,m} : m \in [\![1,n]\!]\}$ is a martingale difference sequence. Note that

$$\tilde{Q}_n/Z_n = \tilde{Q}_{n,n}/Z_n = \sum_{m=1}^{n} X_{n,m}.$$

By the martingale Central Limit Theorem, to show (D.28), it suffices to show:

$$\text{(b1)} \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 1, \tag{D.29}$$

$$\text{(b2)}\ \forall\epsilon > 0, \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2\mathbf{1}\{|X_{n,m} > \epsilon|\}|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 0. \tag{D.30}$$

Below, we show (D.29) and (D.30) separately.

In the first part, we prove (D.29). It suffices to show:

$$\mathbb{E}\left[\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right] = 1, \tag{D.31}$$

and

$$\text{Var}\left(\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right) \xrightarrow{n\to\infty} 0. \tag{D.32}$$

Consider (D.31) first. Recall that by definition

$$X_{n,m} = \frac{\tilde{Q}_{n,m} - \tilde{Q}_{n,m-1}}{Z_n} = \frac{8}{Z_n} \sum_{CC(I_m)\backslash CC(I_{m-1})} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}.$$

An alternative way to enumerate all cycles in $CC(I_m)\backslash CC(I_{m-1})$ is to first select a set of two indices $\{i,j\}$ (we take, *wlog*, $i < j$) from $\{1,...,m-1\}$ and use them as the neighboring nodes of m in the cycle. Then select $k \in \{1, ..., m-1\} \setminus \{i, j\}$ as the last node of the cycle.

$$X_{n,m} = \frac{8}{Z_n} \sum_{1 \le i < j \le m-1} W_{mi} W_{mj} Y_{m-1,ij}, \qquad \text{where} \quad Y_{m-1,ij} = \sum_{\substack{1 \le k \le m-1 \\ k \notin \{i,j\}}} W_{ki} W_{kj}.$$

It follows that

$$\begin{aligned}
\mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] &= \frac{64}{Z_n^2} \sum_{\substack{1 \le i<j \le m-1 \\ 1 \le u < v \le m-1}} \mathbb{E}[W_{mi} W_{mj} Y_{m-1,ij} W_{mu} W_{mv} Y_{m-1,uv} | \mathcal{F}_{n,m-1}] \\
&= \frac{64}{Z_n^2} \sum_{\substack{1 \le i<j \le m-1 \\ 1 \le u < v \le m-1}} Y_{m-1,ij} Y_{m-1,uv} \mathbb{E}[W_{mi} W_{mj} W_{mu} W_{mv}] \\
&= \frac{64}{Z_n^2} \sum_{1 \le i<j \le m-1} Y_{m-1,ij}^2 \mathbb{E}[W_{mi}^2 W_{mj}^2] = \frac{64 \alpha_n^2 (1-\alpha_n)^2}{Z_n^2} \sum_{1 \le i<j \le m-1} Y_{m-1,ij}^2.
\end{aligned}$$

Hence

$$\mathbb{E}\left[\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}]\right] = \frac{64 \alpha_n^2 (1-\alpha_n)^2}{Z_n^2} \sum_{m=1}^{n} \sum_{1 \le i<j \le m-1} \mathbb{E}[Y_{m-1,ij}^2],$$

where

$$\begin{aligned}
\mathbb{E}[Y_{m-1,ij}^2] &= \sum_{\substack{1 \le k,l \le m-1 \\ k,l \notin \{i,j\}}} \mathbb{E}\left[W_{ki} W_{kj} W_{li} W_{lj}\right] = \sum_{\substack{1 \le k \le m-1 \\ k \notin \{i,j\}}} \mathbb{E}\left[W_{ki}^2 W_{kj}^2\right] \\
&= (m-3) \alpha_n^2 (1-\alpha_n)^2.
\end{aligned}$$

It follows that

$$\mathbb{E}\left[\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}]\right] = \frac{64 \alpha_n^2 (1-\alpha_n)^2}{Z_n^2} \sum_{m=1}^{n} \frac{(m-1)(m-2)(m-3)}{2} \alpha_n^2 (1-\alpha_n)^2 = 1.$$

This proves (D.31).

Consider (D.32) next. We decompose $\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}]$ into a sum of two components, then calculate its variance. Note that

$$Y_{m-1,ij}^2 = \left(\sum_{\substack{1 \le k \le m-1 \\ k \notin \{i,j\}}} W_{ki} W_{kj} \right)^2 = \sum_{\substack{1 \le k \le m-1 \\ k \notin \{i,j\}}} W_{ki}^2 W_{kj}^2 + 2 \sum_{\substack{1 \le k < l \le m-1 \\ k,l \notin \{i,j\}}} W_{ki} W_{kj} W_{li} W_{lj}.$$

Hence

$$\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] = \frac{64 \alpha_n^2 (1-\alpha_n)^2}{Z_n^2} \sum_{m=1}^{n} \sum_{1 \le i<j \le m-1} Y_{m-1,ij}^2 = \frac{16 n^4 \alpha_n^4}{Z_n^2} (I_a + I_b),$$

where we denote

$$I_a = \frac{4(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{m=1}^{n} \sum_{1\le i<j\le m-1} \sum_{\substack{1\le k\le m-1\\ k\notin\{i,j\}}} W_{ki}^2 W_{kj}^2,$$

$$I_b = \frac{8(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{m=1}^{n} \sum_{1\le i<j\le m-1} \sum_{\substack{1\le k<l\le m-1\\ k,l\notin\{i,j\}}} W_{ki} W_{kj} W_{li} W_{lj}.$$

Using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}\mathrm{Var}\left(\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right) &= \frac{256n^8\alpha_n^8}{Z_n^4}(\mathrm{Var}(I_a)+\mathrm{Var}(I_b)+2\mathrm{Cov}(I_a,I_b)) \\ &\le \frac{256n^8\alpha_n^8}{Z_n^4}(\sqrt{\mathrm{Var}(I_a)}+\sqrt{\mathrm{Var}(I_b)})^2.\end{aligned}$$

Hence, it suffices to show that $\mathrm{Var}(I_a) \xrightarrow{n\to\infty} 0$ and $\mathrm{Var}(I_b) \xrightarrow{n\to\infty} 0$ separately. For $\mathrm{Var}(I_a)$, we first rearrange the sums in the expression of I_a

$$\begin{aligned}I_a &= \frac{4(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{k=1}^{n-1} \sum_{\substack{1\le i<j\le n-1\\ i,j\ne k}} \sum_{m>\max\{i,j,k\}} W_{ki}^2 W_{kj}^2 \\ &= \frac{4(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{k=1}^{n-1} \sum_{\substack{1\le i<j\le n-1\\ i,j\ne k}} (n-\max\{i,j,k\}+1) W_{ki}^2 W_{kj}^2.\end{aligned}$$

Note that the terms of the first sum over $k=1,...,n$ are pairwise independent, which will facilitate variance computations. Hence

$$\begin{aligned}\mathrm{Var}(I_a) &= \frac{16(1-\alpha_n)^4}{n^8\alpha_n^4} \sum_{k=1}^{n-1} \mathrm{Var}\left(\sum_{\substack{1\le i<j\le n-1\\ i,j\ne k}} (n-\max\{i,j,k\}+1) W_{ki}^2 W_{kj}^2\right) \\ &\le \frac{16(1-\alpha_n)^4}{n^6\alpha_n^4} \sum_{k=1}^{n-1} \sum_{\substack{1\le i<j\le n-1\\ i,j\ne k}} \sum_{\substack{1\le u<v\le n-1\\ u,v\ne k}} \mathrm{Cov}(W_{ki}^2 W_{kj}^2, W_{ku}^2 W_{kv}^2).\end{aligned}$$

We can consider four cases for $\mathrm{Cov}(W_{ki}^2 W_{kj}^2, W_{ku}^2 W_{kv}^2)$:

1. $(i,j)=(u,v)$, then $\mathrm{Var}(W_{ki}^2 W_{kj}^2) \le \mathbb{E}[W_{ki}^4 W_{kj}^4] = \mathbb{E}[W_{ki}^4]^2 \le c\alpha_n^2$,
2. $i=u, j\ne v$, then $\mathrm{Cov}(W_{ki}^2 W_{kj}^2, W_{ki}^2 W_{kv}^2) \le \mathbb{E}[W_{ki}^4 W_{kj}^2 W_{kv}^2] = \mathbb{E}[W_{ki}^4]\mathbb{E}[W_{kj}^2]^2 \le c\alpha_n^3$,
3. The previous bound will also hold for the case $i\ne u, j=v$, the case $i=v$, and the case $j=u$,
4. For any other case, $\mathrm{Cov}(W_{ki}^2 W_{kj}^2, W_{ki}^2 W_{kv}^2) = 0$.

Here, $c > 0$ is a high enough constant. It follows that

$$\text{Var}(I_a) = \frac{16(1-\alpha_n)^4}{n^8\alpha_n^4} \sum_{k=1}^{n-1} \sum_{\substack{1\le i<j\le n-1 \\ i,j\neq k}} \left\{ \text{Var}(W_{ki}^2 W_{kj}^2) + \sum_{\substack{v=i+1 \\ v\notin\{k,j\}}}^{n-1} \text{Cov}(W_{ki}^2 W_{kj}^2, W_{ki}^2 W_{kv}^2) \right.$$

$$\left. + \sum_{\substack{u=1 \\ u\notin\{k,i\}}}^{j-1} \text{Cov}(W_{ki}^2 W_{kj}^2, W_{ku}^2 W_{kj}^2) + \sum_{\substack{u=1 \\ u\neq k}}^{i-1} \text{Cov}(W_{ki}^2 W_{kj}^2, W_{ku}^2 W_{ki}^2) + \sum_{\substack{v=j+1 \\ v\neq k}}^{n-1} \text{Cov}(W_{ki}^2 W_{kj}^2, W_{kj}^2 W_{kv}^2) \right\}$$

$$\le \frac{16c(1-\alpha_n)^4}{n^8\alpha_n^4} \sum_{k=1}^{n-1} \sum_{\substack{1\le i<j\le n-1 \\ i,j\neq k}} \left\{ \alpha_n^2 + 4n\alpha_n^3 \right\} \le \frac{8c}{n^3(n\alpha_n)} \left(4 + \frac{1}{n\alpha_n}\right) \xrightarrow{n\to\infty} 0.$$

Let's now show that $\text{Var}(I_b) \xrightarrow{n\to\infty} 0$. Recall that

$$\begin{aligned}
I_b &= \frac{8(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{m=1}^{n} \sum_{1\le i<j\le m-1} \sum_{\substack{1\le k<l\le m-1 \\ k,l\notin\{i,j\}}} W_{ki}W_{kj}W_{li}W_{lj} \\
&= \frac{2(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{m=1}^{n} \sum_{\substack{1\le i,j,k,l\le m-1 \\ i,j,k,l \text{ dist.}}} W_{ki}W_{kj}W_{li}W_{lj} \\
&= \frac{2(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{\substack{1\le i,j,k,l\le n-1 \\ i,j,k,l \text{ dist.}}} \sum_{m>\max\{i,j,k,l\}} W_{ki}W_{kj}W_{li}W_{lj} \\
&= \frac{2(1-\alpha_n)^2}{n^4\alpha_n^2} \sum_{\substack{1\le i,j,k,l\le n-1 \\ i,j,k,l \text{ dist.}}} (n+1-\max\{i,j,k,l\}) W_{ki}W_{kj}W_{li}W_{lj}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Var}(I_b) &= \frac{4(1-\alpha_n)^4}{n^8\alpha_n^4} \text{Var}\left(\sum_{\substack{1\le i,j,k,l\le n-1 \\ i,j,k,l \text{ dist.}}} (n+1-\max\{i,j,k,l\}) W_{ik}W_{kj}W_{jl}W_{li} \right) \\
&= \frac{4(1-\alpha_n)^4}{n^8\alpha_n^4} \text{Var}\left(8 \sum_{CC(I_{n-1})} (n+1-\max\{i,j,k,l\}) W_{ik}W_{kj}W_{jl}W_{li} \right) \\
&= \frac{32(1-\alpha_n)^4}{n^8\alpha_n^4} \sum_{\substack{1\le i,j,k,l\le n-1 \\ i,j,k,l \text{ dist.}}} (n+1-\max\{i,j,k,l\})^2 \text{Var}(W_{ik}W_{kj}W_{jl}W_{li})
\end{aligned}$$

$$\leq \frac{32(1-\alpha_n)^4}{n^6\alpha_n^4} \sum_{\substack{1\leq i,j,k,l\leq n-1\\ i,j,k,l \text{ dist.}}} \alpha_n^4(1-\alpha_n)^4 \leq \frac{32}{n^2} \xrightarrow{n\to\infty} 0.$$

This gives $\mathrm{Var}(I_b) \xrightarrow{n\to\infty} 0$. Recall that we had:

$$\mathrm{Var}\left(\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\right) \leq \frac{256n^8\alpha_n^8}{Z_n^4}\left(\sqrt{\mathrm{Var}(I_a)}+\sqrt{\mathrm{Var}(I_b)}\right)^2.$$

Since $\frac{256n^8\alpha_n^8}{Z_n^4} \xrightarrow{n\to\infty} 4$, we obtain (D.32). In combination with (D.31), this proves (D.29).

In the second part, we prove (D.30). We have, using the Cauchy-Schwarz and Markov inequalities

$$\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 \mathbf{1}\{|X_{n,m}>\epsilon|\}|\mathcal{F}_{n,m-1}] \leq \sum_{m=1}^{n} \sqrt{\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]}\sqrt{\mathbb{P}(|X_{n,m}|\geq \epsilon|\mathcal{F}_{n,m-1}])}$$
$$\leq \frac{1}{\epsilon^2}\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}].$$

Hence it suffices to show that

$$\mathbb{E}\left[\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]\right] = \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^4] \xrightarrow{n\to\infty} 0. \tag{D.33}$$

Recall that for all $n \in \mathbb{N}^*$, for all $m \in [\![1,n]\!]$

$$X_{n,m} = \frac{2}{Z_n}\sum_{1\leq i<j\leq m-1} W_{mi}W_{mj}Y_{m-1,ij} \quad \text{with} \quad Y_{m-1,ij} = \sum_{\substack{1\leq k\leq m-1\\ k\notin\{i,j\}}} W_{ki}W_{kj}.$$

It follows that

$$\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]$$
$$=\frac{16}{Z_n^4}\sum_{\substack{i<j,u<v\\ k<l,r<s}} Y_{m-1,ij}Y_{m-1,uv}Y_{m-1,kl}Y_{m-1,rs} \times \mathbb{E}[W_{mi}W_{mj}W_{mu}W_{mv}W_{mk}W_{ml}W_{mr}W_{ms}]$$
$$=\frac{16}{Z_n^4}\Bigg\{\sum_{i<j} Y_{m-1,ij}^4\mathbb{E}[W_{mi}^4W_{mj}^4] + 3\sum_{i}\sum_{\substack{j,v\\ j,v>i \text{ and } j\neq v}} Y_{m-1,ij}^2Y_{m-1,iv}^2\mathbb{E}[W_{mi}^4W_{mj}^2W_{mv}^2]$$
$$+3\sum_{j}\sum_{\substack{i,u\\ i,u<j \text{ and } i\neq u}} Y_{m-1,ij}^2Y_{m-1,uj}^2\mathbb{E}[W_{mj}^4W_{mi}^2W_{mu}^2] + 9\sum_{i<j,u<v} Y_{m-1,ij}^2Y_{m-1,uv}^2\mathbb{E}[W_{mi}^2W_{mj}^2W_{mu}^2W_{mv}^2]\Bigg\}$$

$$\leq \frac{16}{Z_n^4}\left\{\sum_{i<j} Y_{m-1,ij}^4 c\alpha_n^2 + 3\sum_i \sum_{\substack{j,v \\ j,v>i \text{ and } j\neq v}} Y_{m-1,ij}^2 Y_{m-1,iv}^2 c\alpha_n^3\right.$$

$$\left. +3\sum_j \sum_{\substack{i,u \\ i,u<j \text{ and } i\neq u}} Y_{m-1,ij}^2 Y_{m-1,uj}^2 c\alpha_n^3 + 9\sum_{i<j,u<v} Y_{m-1,ij}^2 Y_{m-1,uv}^2 c\alpha_n^4\right\},$$

where $c > 0$ is a high enough constant. Hence,

$$\mathbb{E}[X_{n,m}^4] \leq \frac{16c}{Z_n^4}\left\{\alpha_n^2 \sum_{i<j}\mathbb{E}[Y_{m-1,ij}^4] + 3\alpha_n^3\sum_i \sum_{\substack{j,v \\ j,v>i \text{ and } j\neq v}} \mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,iv}^2]\right.$$

$$\left. +3\alpha_n^3\sum_j \sum_{\substack{i,u \\ i,u<j \text{ and } i\neq u}} \mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,uj}^2] + 9\alpha_n^4 \sum_{i<j,u<v}\mathbb{E}[Y_{m-1,ij}^2]\mathbb{E}[Y_{m-1,uv}^2]\right\}.$$

We will now compute upper bounds on $\mathbb{E}[Y_{m-1,ij}^4]$, $\mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,iv}^2]$ and $\mathbb{E}[Y_{m-1,ij}^2]$. We have

$$\begin{aligned}
\mathbb{E}[Y_{m-1,ij}^4] &= \mathbb{E}\left[\sum_{k,l,u,v\notin\{i,j\}} W_{ki}W_{kj}W_{li}W_{lj}W_{ui}W_{uj}W_{vi}W_{vj}\right] \\
&= 3\sum_{k,u\notin\{i,j\}} \mathbb{E}[W_{ki}^2 W_{kj}^2 W_{ui}^2 W_{uj}^2] \\
&= 3\left(\sum_{k\notin\{i,j\}}\mathbb{E}[W_{ki}^4 W_{kj}^4] + \sum_{k\neq u;\, k,u\notin\{i,j\}}\mathbb{E}[W_{ki}^2 W_{kj}^2 W_{ui}^2 W_{uj}^2]\right) \\
&\leq 12m\alpha_n^2 + 3m^2\alpha_n^4 \leq c_1(m\alpha_n^2 + m^2\alpha_n^4),
\end{aligned}$$

where $c_1 > 0$ is a constant. Similarly

$$\mathbb{E}[Y_{m-1,ij}^2] = \mathbb{E}\left[\sum_{k,l\notin\{i,j\}} W_{ki}W_{kj}W_{li}W_{lj}\right] = \sum_{k\notin\{i,j\}}\mathbb{E}[W_{ki}^2 W_{kj}^2] \leq m\alpha_n^2,$$

and

$$\mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,iv}^2] = \mathbb{E}\left[\sum_{k,l,r,s:k,l\notin\{i,j\},r,s\notin\{i,v\}} W_{ki}W_{kj}W_{li}W_{lj}W_{ri}W_{rv}W_{si}W_{sv}\right]$$

$$= \mathbb{E}\left[\sum_{k,r:k\notin\{i,j\},r\notin\{i,v\}} W_{ki}^2 W_{kj}^2 W_{ri}^2 W_{rv}^2\right] = \sum_{k,r:k\notin\{i,j\},r\notin\{i,v\}} \mathbb{E}[W_{ki}^2 W_{kj}^2 W_{ri}^2 W_{rv}^2]$$

$$
\begin{aligned}
&= \sum_{k\notin\{i,j,v\}} \mathbb{E}[W_{ki}^4 W_{kj}^2 W_{kv}^2] + \sum_{k\neq r; k\notin\{i,j\}, r\notin\{i,v\}} \mathbb{E}[W_{ki}^2 W_{kj}^2 W_{ri}^2 W_{rv}^2] \\
&\leq 2m\alpha_n^3 + m^2\alpha_n^4 \leq c_2 m^2 \alpha_n^3,
\end{aligned}
$$

for n big enough (since $\alpha_n \xrightarrow{n\to\infty} 0$), where $c_2 > 0$ is a constant. It follows that, for some constant $\gamma > \max\{1, c, c_1, c_2\}$, we have

$$
\begin{aligned}
\mathbb{E}[X_{n,m}^4] \leq \frac{16c}{Z_n^4} \Bigg\{ &\alpha_n^2 \sum_{i<j} \mathbb{E}[Y_{m-1,ij}^4] + 3\alpha_n^3 \sum_i \sum_{\substack{j,v \\ j,v>i \text{ and } j\neq v}} \mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,iv}^2] \\
&+ 3\alpha_n^3 \sum_j \sum_{\substack{i,u \\ i,u<j \text{ and } i\neq u}} \mathbb{E}[Y_{m-1,ij}^2 Y_{m-1,uj}^2] + 9\alpha_n^4 \sum_{i<j, u<v} \mathbb{E}[Y_{m-1,ij}^2]\mathbb{E}[Y_{m-1,uv}^2] \Bigg\} \\
&\leq \frac{16\gamma^2}{Z_n^4}(m^3\alpha_n^4 + m^4\alpha_n^6 + 6m^5\alpha_n^6 + 9\alpha_n^8 m^6) \\
&\leq \frac{16\gamma^2}{Z_n^4}(n^3\alpha_n^4 + n^4\alpha_n^6 + 6n^5\alpha_n^6 + 9\alpha_n^8 n^6).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\sum_{m=1}^n \mathbb{E}[X_{n,m}^4] &\leq \frac{16\gamma^2}{n^2(n-1)^2(n-2)^2(n-3)^2\alpha_n^8(1-\alpha_n)^8}(n^4\alpha_n^4 + n^5\alpha_n^6 + 6n^6\alpha_n^6 + 9\alpha_n^8 n^7) \\
&= \left(\frac{144\gamma^2 n^6}{(n-1)^2(n-2)^2(n-3)^2(1-\alpha_n)^8}\right)\left(\frac{1}{n^4\alpha_n^4} + \frac{1}{n^3\alpha_n^2} + \frac{1}{n^2\alpha_n^2} + \frac{1}{n}\right) \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

This gives (D.33). Then, (D.30) follows immediately.

D.3. Proof of the joint null distribution

We now show the desirable claim (D.1). We shall use the previously defined notations:

$$
\begin{aligned}
T_n &= \sum_{i,j,k \text{ dist.}} (A_{ik} - \alpha_n)(A_{jk} - \alpha_n), \qquad \hat{T}_n = \sum_{i,j,k \text{ dist.}} (A_{ik} - \hat{\alpha}_n)(A_{jk} - \hat{\alpha}_n), \\
\tilde{Q}_n &= \sum_{(i_1,i_2,i_3,i_4) \text{ dist.}} W_{i_1 i_2} W_{i_2 i_3} W_{i_3 i_4} W_{i_4 i_1}.
\end{aligned}
$$

We have seen the decomposition of ψ_n^{DC} in (D.3) and the decomposition of ψ_n^{SQ} in (D.18). We plug them into the definition of S_n to get:

$$
S_n = u\left[\frac{\hat{T}_n}{(n-1)\hat{\alpha}_n(1-\hat{\alpha}_n)}\right] + v\left[\frac{Q_n - \tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}\left(\frac{\alpha_n}{\hat{\alpha}_n}\right)^2 + \frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}\left(\frac{\alpha_n}{\hat{\alpha}_n}\right)^2\right]
$$

$$= \epsilon_n + u\frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}} + v\frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}, \tag{D.34}$$

where

$$\epsilon_n = u\frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}}\left[\frac{\sqrt{n-1}\alpha_n(1-\alpha_n)\hat{T}_n}{\sqrt{n-2}\hat{\alpha}_n(1-\hat{\alpha}_n)T_n} - 1\right] + v\left[\frac{\alpha_n^2}{\hat{\alpha}_n^2}\frac{(Q_n - \tilde{Q}_n)}{2\sqrt{2}n^2\alpha_n^2} + \left(\frac{\alpha_n^2}{\hat{\alpha}_n^2} - 1\right)\frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2}\right].$$

In Sections D.1-D.2, we have shown that

$$\frac{\hat{\alpha}_n}{\alpha_n} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\hat{T}_n}{T_n} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}} \xrightarrow{d} \mathcal{N}(0,1), \qquad \frac{Q_n - \tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2} \xrightarrow{d} \mathcal{N}(0,1). \tag{D.35}$$

It follows immediately that $\epsilon_n \xrightarrow{\mathbb{P}} 0$. By Slutsky's theorem, it suffices to show that

$$C_n \triangleq u\frac{\frac{T_n}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}} + v\frac{\tilde{Q}_n}{2\sqrt{2}n^2\alpha_n^2} \quad \xrightarrow[n\to\infty]{\mathcal{L}} \quad \mathcal{N}(0,1). \tag{D.36}$$

Below, we show (D.36). In Section D.1, we have defined I_m as the collection of all distinct $\{(i,j,k)$ such that $1 \le i,j,k \le m$; in Section D.2, we have defined $CC(I_m)$. For each $1 \le m \le n$, let

$$T_{n,m} = \sum_{(j_1,j_2,j_3)\in I_m} W_{j_1j_3}W_{j_2j_3}, \qquad \tilde{Q}_{n,m} = \sum_{CC(I_m)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1},$$

where $T_{n,0} = \tilde{Q}_{n,0} = 0$ by default. Introduce

$$C_{n,m} = u\frac{\frac{T_{n,m}}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}} + v\frac{\tilde{Q}_{n,m}}{2\sqrt{2}n^2\alpha_n^2}, \qquad \text{for all } 0 \le m \le n.$$

We have seen that $\{T_{n,m}\}_{0\le m\le n}$ and $\{\tilde{Q}_{n,m}\}_{0\le m\le n}$ are both martingales with respect to the filtration $\{\mathcal{F}_{n,m}\}_{0\le m\le n}$ defined before. It is easy to see that $\{C_{n,m}\}_{0\le m\le n}$ is also a martingale. Write

$$C_n = \sum_{m=1}^{n} D_{n,m}, \qquad \text{where} \quad D_{n,m} \equiv C_{n,m} - C_{n,m-1}.$$

To show $C_n \xrightarrow{d} \mathcal{N}(0,1)$, we apply the martingale Central Limit Theorem. It suffices to show:

$$\text{(a)} \sum_{m=1}^{n} \mathbb{E}[D_{n,m}^2|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 1, \tag{D.37}$$

$$\text{(b)} \ \forall\epsilon > 0, \sum_{m=1}^{n} \mathbb{E}[D_{n,m}^2\mathbf{1}\{|D_{n,m} > \epsilon|\}|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 0. \tag{D.38}$$

It remains to show (D.37)-(D.38). Consider (D.38). Write

$$D^{(1)}_{n,m} = \frac{\frac{T_{n,m}-T_{n,m-1}}{(n-1)\alpha_n(1-\alpha_n)}}{\sqrt{\frac{2n(n-2)}{(n-1)}}}, \qquad \text{and} \qquad D^{(1)}_{n,m} = \frac{\tilde{Q}_{n,m} - \tilde{Q}_{n,m-1}}{2\sqrt{2}n^2\alpha_n^2}.$$

Then, $D_{n,m} = uD^{(1)}_{n,m} + vD^{(2)}_{n,m}$. It follows that $D^4_{n,m} \le 8u^4(D^{(1)}_{n,m})^4 + 8v^4(D^{(2)}_{n,m})^4$. As a result, for any $\epsilon > 0$, by the Cauchy-Schwarz inequality and the Markov inequality, we have

$$\begin{aligned}\left(\sum_{m=1}^n \mathbb{E}[D^2_{n,m}\mathbf{1}\{|D_{n,m} > \epsilon|\}|\mathcal{F}_{n,m-1}]\right)^2 &\le \left(\sum_{m=1}^n \mathbb{E}[D^4_{n,m}|\mathcal{F}_{n,m-1}]\right)\cdot \mathbb{P}\big(|D_{n,m}| > \epsilon|\mathcal{F}_{n,m-1}\big) \\ &\le \sum_{m=1}^n \mathbb{E}[D^4_{n,m}|\mathcal{F}_{n,m-1}] \\ &\le 8u^4 \sum_{m=1}^n \mathbb{E}[(D^{(1)}_{n,m})^4|\mathcal{F}_{n,m-1}] + 8v^4 \sum_{m=1}^n \mathbb{E}[(D^{(2)}_{n,m})^4|\mathcal{F}_{n,m-1}].\end{aligned}$$

With significant efforts, we have shown $\sum_{m=1}^n \mathbb{E}[(D^{(1)}_{n,m})^4|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} 0$ in Section D.1, and we have shown $\sum_{m=1}^n \mathbb{E}[(D^{(2)}_{n,m})^4|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}}$ in Section D.2. Plugging them into the above inequality, we immediately obtain (D.38).

Consider (D.37). Write

$$A_n = \sum_{m=1}^n \mathbb{E}[(D^{(1)}_{n,m})^2|\mathcal{F}_{n,m-1}], \qquad B_n = \sum_{m=1}^n \mathbb{E}[(D^{(2)}_{n,m})^2|\mathcal{F}_{n,m-1}],$$
$$M_n = \sum_{m=1}^n \mathbb{E}[(D^{(1)}_{n,m})D^{(2)}_{n,m}|\mathcal{F}_{n,m-1}].$$

Then,

$$\sum_{m=1}^n \mathbb{E}[D^2_{n,m}|\mathcal{F}_{n,m-1}] = u^2A_n + v^2B_n + 2uvM_n,$$

In Sections D.1-D.2, we have shown that $A_n \xrightarrow{\mathbb{P}} 1$ and $B_n \xrightarrow{\mathbb{P}} 1$. We claim that

$$M_n \xrightarrow{\mathbb{P}} 0. \tag{D.39}$$

Then, it follows that $\sum_{m=1}^n \mathbb{E}[D^2_{n,m}|\mathcal{F}_{n,m-1}] \xrightarrow{\mathbb{P}} u^2\cdot 1 + v^2 \cdot 1 + 2uv\cdot 0 = 1$. This gives (D.37).

It remains to show (D.39). Using the expressions of $D^{(1)}_{n,m}$ and $D^{(2)}_{n,m}$, we have

$$M_n = \frac{\tilde{M}_n}{n^2\alpha_n^3(1-\alpha_n)\sqrt{n(n-1)(n-2)}},$$

where $\tilde{M}_n = \sum_{m=1}^n \mathbb{E}[(T_{n,m} - T_{n,m-1})(\tilde{Q}_{n,m} - \tilde{Q}_{n,m-1})|\mathcal{F}_{n,m-1}]$. We plug in the definitions of $T_{n,m}$ and $\tilde{Q}_{n,m}$ to get

$$\tilde{M}_n = \sum_{m=1}^n \left(\sum_{\substack{(j_1,j_2,j_3)\in \\ I_m\setminus I_{m-1}}} \sum_{\substack{(i_1,i_2,i_3,i_4)\in \\ CC(I_m)\setminus CC(I_{m-1})}} \mathbb{E}\left[W_{j_1j_3}W_{j_2j_3}\cdot W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}\middle|\mathcal{F}_{n,m-1}\right] \right).$$

Let's see when $\mathbb{E}[W_{j_1j_3}W_{j_2j_3}W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}|\mathcal{F}_{n,m-1}] \neq 0$. Since $(i_1,i_2,i_3,i_4)\in CC(I_m)\setminus CC(I_{m-1})$, exactly one of the four indices must be m. We assume $i_1 = m$ without loss of generality. Since $(j_1,j_2,j_3)\in I_m\setminus I_{m-1}$, exactly one of the three indices must be m. Without loss of generality, we assume either $j_1 = m$ or $j_3 = m$. If $j_1 = m$ (and recall that we have assumed $i_1 = m$), then

$$\begin{aligned} & \mathbb{E}[W_{j_1j_3}W_{j_2j_3}W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}|\mathcal{F}_{n,m-1}] \\ = \;& W_{j_2j_3}W_{i_2i_3}W_{i_3i_4}\cdot\mathbb{E}[W_{mj_3}W_{mi_2}W_{i_4m}|\mathcal{F}_{n,m-1}]. \end{aligned}$$

It is nonzero only if $j_3 = i_2 = i_4$. However, this is impossible, because i_2 and i_4 need to be distinct. If $j_3 = m$ (and recall that we have assumed $i_1 = m$), we have

$$\begin{aligned} & \mathbb{E}[W_{j_1j_3}W_{j_2j_3}W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}|\mathcal{F}_{n,m-1}] \\ = \;& W_{i_2i_3}W_{i_3i_4}\cdot\mathbb{E}[W_{j_1m}W_{j_2m}W_{mi_2}W_{i_4m}|\mathcal{F}_{n,m-1}]. \end{aligned}$$

Note that $j_1 \neq j_2$ and $i_2 \neq i_4$. For the above to be nonzero, we must have $\{i_2,i_4\} = \{j_1,j_2\}$. It follows that

$$\begin{aligned} \tilde{M}_n &= 8\sum_{m=1}^n \sum_{\substack{1\leq i_2,i_3,i_4\leq m-1 \\ \text{(distinct)}}} W_{i_2i_3}W_{i_3i_4}\cdot\mathbb{E}[W_{mi_2}^2 W_{mi_4}^2|\mathcal{F}_{n,m-1}] \\ &= 8\alpha_n^2(1-\alpha_n)^2 \sum_{m=1}^n \sum_{(i_2,i_3,i_4)\in I_{m-1}} W_{i_2i_3}W_{i_3i_4} \\ &= 8\alpha_n^2(1-\alpha_n)^2 \sum_{(i_2,i_3,i_4)\in I_{n-1}} (n - \max\{i_2,i_3,i_4\})W_{i_2i_3}W_{i_3i_4}. \end{aligned} \tag{D.40}$$

As a result,

$$\begin{aligned} \mathbb{E}[M_n^2] &= \frac{\mathbb{E}[\tilde{M}_n^2]}{n^5(n-1)(n-2)\alpha_n^6(1-\alpha_n)^2} \\ &= \frac{64\alpha_n^4(1-\alpha_n)^4}{n^5(n-1)(n-2)\alpha_n^6(1-\alpha_n)^2}\times\mathbb{E}\left[\left(\sum_{(i_2,i_3,i_4)\in I_{n-1}}(n-\max\{i_2,i_3,i_4\})W_{i_2i_3}W_{i_3i_4}\right)^2\right] \\ &\leq \frac{C}{n^7\alpha_n^2}\sum_{i_2,i_3,i_4} n^2\cdot\mathbb{E}[W_{i_2i_3}^2W_{i_3i_4}^2] \\ &\leq \frac{C}{n^7\alpha_n^2}\times n^5\alpha_n^2 \quad \xrightarrow{n\to\infty} \quad 0. \end{aligned}$$

Then, (D.39) follows directly. This completes the proof of Theorem 3.1. □

Appendix E: Proof of Theorem 3.2

Define

$$U_n = \frac{\hat{\alpha}_n(1-\hat{\alpha}_n)}{\alpha_0(1-\alpha_0)} - 1, \qquad \text{and} \qquad Z_n^* = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{(n-1)\alpha_0(1-\alpha_0)} - n.$$

By definition, $X_n = (1+U_n)^{-1}(n + Z_n^*)$. It follows that

$$\psi_n^{DC} \equiv \frac{X_n - n}{\sqrt{n}} = \frac{1}{\sqrt{n}(1+U_n)}(Z_n^* - nU_n). \tag{E.1}$$

The asymptotic behavior of ψ_n^{DC} is mainly determined by Z_n^*. Below, we first calculate the mean and variance of Z_n^*; then, we use these results to study the mean and variance of ψ_n^{DC}.

The mean and variance of Z_n^*. We introduce a matrix

$$\tilde{\Omega} = \Omega - \alpha_0 \mathbf{1}_n \mathbf{1}_n', \qquad \text{where} \quad \alpha_0 = h'Ph.$$

Then, $A_{ij} = W_{ij} + \tilde{\Omega}_{ij} + \alpha_0$, for all $i \neq j$. Write $\tilde{\Omega}^* = \tilde{\Omega} - \text{diag}(\tilde{\Omega})$. It follows that

$$\begin{aligned}
\sum_{i=1}^n (d_i - \bar{d})^2 &= \sum_{i=1}^n \left(\sum_{j:j\neq i} (W_{ij} + \tilde{\Omega}_{ij} + \alpha_0) - \frac{1}{n} \sum_{(k,\ell):k\neq \ell} (W_{k\ell} + \tilde{\Omega}_{k\ell} + \alpha_0) \right)^2 \\
&= \sum_{i=1}^n \left(e_i' W \mathbf{1}_n + e_i'\tilde{\Omega}^* \mathbf{1}_n - \frac{1}{n}\mathbf{1}_n' W \mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)^2 \\
&= \sum_{i=1}^n \left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)^2 + 2\sum_{i=1}^n \left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)(e_i'W\mathbf{1}_n) \\
&\quad - 2\sum_{i=1}^n \left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)\left(\frac{1}{n}\mathbf{1}_n'W\mathbf{1}_n \right) + \sum_{i=1}^n (e_i'W\mathbf{1}_n)^2 \\
&\quad + \frac{1}{n}(\mathbf{1}_n'W\mathbf{1}_n)^2 - 2\sum_{i=1}^n (e_i'W\mathbf{1}_n)\left(\frac{1}{n}\mathbf{1}_n'W\mathbf{1}_n \right) \\
&= \sum_{i=1}^n \left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)^2 + 2\sum_{i=1}^n \left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \right)(e_i'W\mathbf{1}_n) \\
&\quad + \sum_{i=1}^n (e_i'W\mathbf{1}_n)^2 - \frac{1}{n}(\mathbf{1}_n'W\mathbf{1}_n)^2.
\end{aligned} \tag{E.2}$$

We further combine the last two terms of (E.2):

$$\sum_{i=1}^n (e_i'W\mathbf{1}_n)^2 - \frac{1}{n}(\mathbf{1}_n'W\mathbf{1}_n)^2$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \Big(\sum_{j:j\neq i} W_{ij} \Big)^2 - \frac{1}{n} \Big(\sum_{i\neq j} W_{ij} \Big)^2 \\
&= \sum_{i\neq j} W_{ij}^2 + \sum_{i,j,k \text{ dist}} W_{ij}W_{ik} - \frac{2}{n}\sum_{i\neq j} W_{ij}^2 - \frac{1}{n}\sum_{i\neq j} \sum_{\substack{k\neq l \\ \{k,l\}\neq\{i,j\}}} W_{ij}W_{kl} \\
&= \frac{n-2}{n} \sum_{i\neq j} W_{ij}^2 + \sum_{i,j,k, \text{ dist}} W_{ij}W_{ik} - \frac{1}{n}\sum_{i\neq j} \sum_{\substack{k\neq l \\ \{k,l\}\neq\{i,j\}}} W_{ij}W_{kl}.
\end{aligned}
$$

We plug it into (E.2) to get

$$
\begin{aligned}
(n-1)\alpha_0(1-\alpha_0)Z_n^* &\equiv \sum_{i=1}^{n} \big(d_i - \bar{d}\big)^2 - n(n-1)\alpha_0(1-\alpha_0) \\
&= Y_1 + 2Y_2 + Y_3 + Y_4 - Y_5, \qquad \text{(E.3)}
\end{aligned}
$$

where

$$
\begin{aligned}
Y_1 &= \sum_{i=1}^{n} \Big(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \Big)^2, \\
Y_2 &= \sum_{i=1}^{n}\sum_{j\neq i} \Big(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n \Big) W_{ij}, \\
Y_3 &= \Big(\frac{n-2}{n} \sum_{i\neq j} W_{ij}^2 \Big) - n(n-1)\alpha_0(1-\alpha_0), \\
Y_4 &= \sum_{i,j,k, \text{ dist}} W_{ij}W_{ik}, \\
Y_5 &= \frac{1}{n}\sum_{i\neq j} \sum_{\substack{k\neq l \\ \{k,l\}\neq\{i,j\}}} W_{ij}W_{kl}.
\end{aligned}
$$

We now compute the mean of Z_n^*. It is easy to see that

$$
\mathbb{E}[Z_n^*] = \frac{Y_1 + \mathbb{E}[Y_3]}{(n-1)\alpha_0(1-\alpha_0)}. \qquad \text{(E.4)}
$$

For Y_1, note that $\tilde{\Omega}^* = \tilde{\Omega} - \text{diag}(\tilde{\Omega})$. Since $\Pi\mathbf{1}_K = \mathbf{1}_n$, we can re-write

$$
\tilde{\Omega} = \Omega - \alpha_0\Pi\mathbf{1}_K\mathbf{1}_K'\Pi' = \Pi\left(P - \alpha_0\mathbf{1}_K\mathbf{1}_K'\right)\Pi' = \Pi M\Pi'.
$$

As a result, $\tilde{\Omega}\mathbf{1}_n = n\Pi Mh$, and $\mathbf{1}_n'\tilde{\Omega}\mathbf{1}_n = 0$. We plug them into the expression of Y_1 and note that $(a+b)^2 \geq \frac{a^2}{2} - b^2$, for any $a, b \in \mathbb{R}$. It follows that

$$
Y_1 = \|\tilde{\Omega}^*\mathbf{1}_n\|^2 - \frac{1}{n}(\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n)^2
$$

$$
\begin{aligned}
&= \|\tilde{\Omega}\mathbf{1}_n - \mathrm{diag}(\tilde{\Omega})\mathbf{1}_n\|^2 - \frac{1}{n}(\mathbf{1}_n'\mathrm{diag}(\tilde{\Omega})\mathbf{1}_n)^2 \\
&\geq \frac{1}{2}\|\tilde{\Omega}\mathbf{1}_n\|^2 - \|\mathrm{diag}(\tilde{\Omega})\mathbf{1}_n\|^2 - \frac{1}{n}(\mathbf{1}_n'\mathrm{diag}(\tilde{\Omega})\mathbf{1}_n)^2 \\
&= \frac{n^2}{2}\|\Pi M h\|^2 - \sum_{i=1}^n \tilde{\Omega}_{ii}^2 - \frac{1}{n}\left(\sum_{i=1}^n \tilde{\Omega}_{ii}\right)^2.
\end{aligned}
$$

Note that $\max_i |\tilde{\Omega}_{ii}| \leq \max_{k,l}|M_{kl}| = C\|M\|$. Moreover, since $G = n^{-1}\Pi'\Pi$ and $\lambda_{\min}(G) \geq c$, we have $\|\Pi M h\|^2 = n(h'MGMh) \geq Cn\|Mh\|^2$, and $\|\Pi M h\|^2 \leq \|\Pi\|^2\|Mh\|^2 \leq Cn\|Mh\|^2$. It follows that

$$
Y_1 = \frac{n^2}{2}\|\Pi M h\|^2 - O(n\|M\|^2) \asymp n^3\|Mh\|^2. \tag{E.5}
$$

For Y_3, we have

$$
\mathbb{E}[Y_3] = \frac{n-2}{n}\sum_{i\neq j}\Omega_{ij}(1-\Omega_{ij}) - n(n-1)\alpha_0(1-\alpha_0).
$$

Write $\Omega_{ij}(1-\Omega_{ij}) = \alpha_0(1-\alpha_0) + (1-2\alpha_0)(\Omega_{ij}-\alpha_0) - (\Omega_{ij}-\alpha_0)^2$. Recalling that $\Omega_{ij} - \alpha_0 = \tilde{\Omega}_{ij}$, we plug these results into $\mathbb{E}[Y_3]$ to get

$$
\begin{aligned}
\mathbb{E}[Y_3] &= \frac{n-2}{n}\sum_{i\neq j}\left[\alpha_0(1-\alpha_0) + (1-2\alpha_0)\tilde{\Omega}_{ij} - \tilde{\Omega}_{ij}^2\right] - n(n-1)\alpha_0(1-\alpha_0) \\
&= -2(n-1)\alpha_0(1-\alpha_0) + \frac{n-2}{n}\left[(1-2\alpha_0)\left(\mathbf{1}_n'\tilde{\Omega}\mathbf{1}_n - \sum_i \tilde{\Omega}_{ii}\right) - \sum_{i\neq j}\tilde{\Omega}_{ij}^2\right] \\
&= -2(n-1)\alpha_0(1-\alpha_0) - \frac{n-2}{n}\left[(1-2\alpha_0)\sum_i \tilde{\Omega}_{ii} + \sum_{i\neq j}\tilde{\Omega}_{ij}^2\right].
\end{aligned}
$$

Then, $|\mathbb{E}[Y_3]| \leq Cn\alpha_0 + Cn\|M\| + Cn^2\|M\|^2$. Recall that by assumption, $\|M\| \leq C\|Mh\|$, $n\alpha_0 \to \infty$ and $\delta_n = n^{-3/2}\alpha_0^{-1}\|Mh\|^2 \to \infty$. It follows that

$$
\frac{|\mathbb{E}[Y_3]|}{n^3\|Mh\|^2} \leq \frac{C}{\sqrt{n}\delta_n} + \frac{C}{n^{3/4}\sqrt{n\alpha_0\delta_n}} + \frac{C}{n} \to 0.
$$

It yields that

$$
\mathbb{E}[Y_3] = o(n^3\|Mh\|^2). \tag{E.6}
$$

We plug (E.5)-(E.6) into (E.4) to get

$$
\mathbb{E}[Z_n^*] = \frac{(n/2)\|\Pi M h\|^2 - o(n^3\|Mh\|^2)}{(n-1)\alpha_0(1-\alpha_0)} \asymp n^2\alpha_0^{-1}\|Mh\|^2. \tag{E.7}
$$

We then compute the variance of Z_n^*, it is easy to see that

$$
\mathrm{Var}(Z_n^*) \leq \frac{C\mathrm{Var}(Y_2) + C\mathrm{Var}(Y_3) + C\mathrm{Var}(Y_4) + C\mathrm{Var}(Y_5)}{(n-1)^2\alpha_0^2(1-\alpha_0)^2}.
$$

By direct calculations, we know that

$$\begin{aligned}\mathrm{Var}(Y_3) &\le C\sum_{i<j}\mathbb{E}[W_{ij}^4] \le C\sum_{i\ne j}\Omega_{ij} \le Cn^2\alpha_0,\\ \mathrm{Var}(Y_4) &\le C\sum_{i,j,k \text{ dist}}\mathbb{E}[W_{ij}^2]\mathbb{E}[W_{ik}^2] \le Cn^3\alpha_0^2,\\ \mathrm{Var}(Y_5) &\le \frac{C}{n^2}\sum_{\substack{i\ne j,k\ne l\\ \{i,j\}\ne\{k,l\}}}\mathbb{E}[W_{ij}^2]\mathbb{E}[W_{k,l}^2] \le Cn^2\alpha_0^2.\end{aligned}$$

In the previous steps, we have seen that $\tilde{\Omega}^* = \tilde{\Omega} - \mathrm{diag}(\tilde{\Omega})$, $\mathbf{1}_n'\tilde{\Omega}\mathbf{1}_n = 0$, $\|\tilde{\Omega}\mathbf{1}_n\|^2 = n^3h'MGMh$, $\Omega_{ij} \le C\alpha_0$, and $|\tilde{\Omega}_{ii}| \le C\|M\|$. It follows that

$$\begin{aligned}\mathrm{Var}(Y_2) &\le C\sum_{i\ne j}\left(e_i'\tilde{\Omega}^*\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n'\tilde{\Omega}^*\mathbf{1}_n\right)^2 \times \Omega_{ij}(1-\Omega_{ij})\\ &= C\sum_{i\ne j}\left[e_i'\tilde{\Omega}\mathbf{1}_n + \tilde{\Omega}_{ii} - \frac{1}{n}\left(\mathbf{1}_n'\mathrm{diag}(\tilde{\Omega})\mathbf{1}_n\right)\right]^2 \times \Omega_{ij}(1-\Omega_{ij})\\ &\le C\left[n\|\tilde{\Omega}\mathbf{1}_n\|^2 + n\sum_i\tilde{\Omega}_{ii}^2 + \left(\mathbf{1}_n'\mathrm{diag}(\tilde{\Omega})\mathbf{1}_n\right)^2\right] \times C\alpha_0\\ &\le Cn^4\alpha_0\|Mh\|^2 + Cn^2\alpha_0\|\mathrm{diag}(M)\|^2\\ &\le Cn^4\alpha_0\|Mh\|^2.\end{aligned}$$

We combine the above results and note that for n big enough, $n\alpha_0 \ge c$. It gives

$$\begin{aligned}\mathrm{Var}(Z_n^*) &\le \frac{C}{n^2\alpha_0^2}\left(n^4\alpha_0\|Mh\|^2 + n^3\alpha_0^2\right)\\ &\le Cn^2\alpha_0^{-1}\|Mh\|^2 + Cn. \qquad \text{(E.8)}\end{aligned}$$

In conclusion, the mean and variance of Z_n^* are characterized by (E.7) and (E.8), respectively.

The mean and variance of ψ_n^{DC}. We now show the claims of this theorem. First, consider the mean of ψ_n^{DC}. Recalling (E.1) and letting $\Delta_n = (1+U_n)^{-1}U_n$, we have

$$\begin{aligned}\sqrt{n}\,\mathbb{E}[\psi_n^{DC}] &\ge \mathbb{E}[Z_n^*] - \mathbb{E}\left|\frac{U_n}{1+U_n}Z_n^*\right| - n\,\mathbb{E}\left|\frac{U_n}{1+U_n}\right|\\ &\ge \mathbb{E}[Z_n^*] - \sqrt{\mathbb{E}[\Delta_n^2]}\sqrt{\mathbb{E}[(Z_n^*)^2]} - n\sqrt{\mathbb{E}[\Delta_n^2]}. \qquad \text{(E.9)}\end{aligned}$$

The mean and variance of Z_n^* have been analyzed above. We now study Δ_n, which is a function of $\hat{\alpha}_n$ and α_0. Note that

$$\max_{i,j}\Omega_{ij} \le \max_{k,\ell}P_{k\ell} \le \mathbf{1}_K'P\mathbf{1}_K \le Ch'Ph = C\alpha_0,$$

where $\mathbf{1}_K' P \mathbf{1}_K \leq C h' P h$ is because $\min_k h_k \geq c$. Since $\hat{\alpha}_n = \frac{1}{n(n-1)} \mathbf{1}_n' A \mathbf{1}_n$ and $\alpha_0 = h' P h = n^{-2} \mathbf{1}_n' \Omega \mathbf{1}_n$, we have

$$\begin{aligned}
|\mathbb{E}[\hat{\alpha}_n] - \alpha_0| &= \frac{1}{n(n-1)} \left| \mathbf{1}_n' \Omega \mathbf{1}_n - \mathbf{1}_n' \mathrm{diag}(\Omega) \mathbf{1}_n - n(n-1)\alpha_0 \right| \\
&= \frac{1}{n(n-1)} \left| n^2 \alpha_0 - \mathbf{1}_n' \mathrm{diag}(\Omega) \mathbf{1}_n - n(n-1)\alpha_0 \right| \ \leq C n^{-1} \alpha_0, \\
\mathrm{Var}(\hat{\alpha}_n) &= \frac{4}{n^2(n-1)^2} \sum_{i<j} \Omega_{ij}(1-\Omega_{ij}) \ \leq C n^{-2} \alpha_0. \qquad \text{(E.10)}
\end{aligned}$$

Furthermore, we write $\hat{\alpha}_n - \mathbb{E}[\hat{\alpha}_n] = \frac{2}{n(n-1)} \sum_{i<j} W_{ij}$, where $\{W_{ij}\}_{i<j}$ is a collection of independent, bounded, zero-mean variables. We apply Bernstein's inequality and use (E.10) to get

$$\mathbb{P}\big(\big|\hat{\alpha}_n - \mathbb{E}[\alpha_n]\big| > t\big) \leq \exp\left(-\frac{t^2/2}{Cn^{-2}\alpha_0 + Cn^{-2}t}\right), \qquad \text{for all } t > 0. \qquad \text{(E.11)}$$

Consider the event $E = \{|\hat{\alpha}_n - \alpha_0| < \delta \cdot \alpha_0\}$, for a sufficiently small constant $\delta > 0$ to be determined. Using the above inequality, $\mathbb{P}(E^c) \leq \exp(-C\delta \cdot n^2 \alpha_0)$ for big enough n. On the event E, we can derive a bound for $|\Delta_n|$. Recalling that $U_n = \frac{\hat{\alpha}_n(1-\hat{\alpha}_n)}{\alpha_0(1-\alpha_0)}$, we have

$$\Delta_n = \frac{U_n}{1+U_n} = \frac{(\hat{\alpha}_n - \alpha_0)(1 - \hat{\alpha}_n - \alpha_0)}{\hat{\alpha}_n(1-\hat{\alpha}_n)}.$$

Since $\alpha_0 \leq 1 - c$ for a constant $c \in (0,1)$, when δ is chosen properly small, $|\Delta_n| \leq C\alpha_0^{-1}|\hat{\alpha}_n - \alpha_0|$ on the event E, where the constant $C > 0$ here does not depend on δ. On the event E^c, according to the footnote on Page 3, $|\Delta_n| \leq Cn^2$. It follows that

$$\begin{aligned}
\mathbb{E}[\Delta_n^2] &\leq Cn^4 \cdot \mathbb{P}(E^c) + C\alpha_0^{-2} \mathbb{E}[(\hat{\alpha}_n - \alpha_0)^2] \\
&\leq Cn^4 \cdot \mathbb{P}(E^c) + C\alpha_0^{-2} \left[(\mathbb{E}[\hat{\alpha}_n] - \alpha_0)^2 + \mathrm{Var}(\hat{\alpha}_n) \right] \\
&\leq Cn^4 \exp(-C\delta n^2 \alpha_0) + C\alpha_0^{-2}(n^{-2}\alpha_0^2 + n^{-2}\alpha_0) \\
&\leq Cn^{-2}\alpha_0^{-1}. \qquad \text{(E.12)}
\end{aligned}$$

We plug (E.12) into (E.9) and then utilize (E.7)-(E.8). Recalling that we have defined $\delta_n = n^{3/2}\alpha_0^{-1}\|Mh\|^2$, it yields that

$$\begin{aligned}
\mathbb{E}[\psi_n^{DC}] &\geq \frac{C}{\sqrt{n}} \Bigg(n^2 \alpha_0^{-1} \|Mh\|^2 - \sqrt{Cn^{-2}\alpha_0^{-1}} \times \\
&\sqrt{(n^2\alpha_0^{-1}\|Mh\|^2)^2 + (n^2\alpha_0^{-1}\|Mh\|^2 + n)} - n\sqrt{Cn^{-2}\alpha_0^{-1}} \Bigg) \\
&= \frac{C}{\sqrt{n}} \left(\sqrt{n}\delta_n - \sqrt{Cn^{-2}\alpha_0^{-1}} \sqrt{n\delta_n^2 + \sqrt{n}\delta_n + n} - n\sqrt{Cn^{-2}\alpha_0^{-1}} \right) \\
&\geq C\delta_n \left(1 - \frac{C}{\sqrt{n^2\alpha_0}} - \frac{C}{\sqrt{n^{5/2}\alpha_0\delta_n}} - \frac{C}{\sqrt{n^2\alpha_0\delta_n^2}} \right) - \frac{C}{\sqrt{n\alpha_0}}
\end{aligned}$$

$$\geq C\delta_n\Big[1 - O\big(n^{-5/4}\alpha_0^{-1/2}\delta_n^{-1/2} + n^{-1}\alpha_0^{-1/2}\delta_n^{-1}\big)\Big] - O\big(n^{-1/2}\alpha_0^{-1/2}\big).$$

Now, assume that $\delta_n \geq C$. Then, there exists a constant $c_1 > 0$ such that

$$\mathbb{E}[\psi_n^{DC}] \geq c_1\delta_n - O\big(n^{-1/2}\alpha_0^{-1/2}\big). \tag{E.13}$$

This gives the first claim.

Next, consider the variance of ψ_n^{DC}. Note that $(1+U_n)^{-1} = 1 - \Delta_n$ and $(1+U_n)^{-1}U_n = \Delta_n$. It follows from (E.1) that $\sqrt{n}\,\psi_n^{DC} = Z_n^* - \Delta_n Z_n^* - n\Delta_n$. Therefore,

$$\begin{aligned}
\mathrm{Var}(\psi_n^{DC}) &\leq Cn^{-1}[\mathrm{Var}(Z_n^*) + \mathrm{Var}(\Delta_n Z_n^*) + n^2\mathrm{Var}(\Delta_n)] \\
&\leq Cn^{-1}\big(\mathrm{Var}(Z_n^*) + \mathbb{E}[\Delta_n^2(Z_n^*)^2] + n^2\mathbb{E}[\Delta_n^2]\big) \\
&\leq Cn^{-1}\Big(n^2\alpha_0^{-1}\|Mh\|^2 + n + \mathbb{E}[\Delta_n^2(Z_n^*)^2] + \alpha_0^{-1}\Big),
\end{aligned} \tag{E.14}$$

where we have used (E.8) and (E.12) in the last inequality.

We calculate $\mathbb{E}[\Delta_n^2(Z_n^*)^2]$. For a large enough constant $B_0 > 0$, we define an event

$$E_1 = \big\{|\hat{\alpha}_n - \mathbb{E}[\hat{\alpha}_n]| \leq B_0 n^{-1}\sqrt{\alpha_0\log(n)}\big\}.$$

By (E.11), $\mathbb{P}(E_1^c) \leq \exp(-B\log(n))$, where the constant $B > 0$ is a monotone increasing function of B_0. With a properly large B_0, we can make $\exp(-B\log(n)) = o(n^8\alpha_0^{-2})$. Now, on the event E_1, we have $|\Delta_n| \leq C\alpha_0^{-1}|\hat{\alpha}_n - \alpha_0| \leq Cn^{-1}\alpha_0^{-1/2}\sqrt{\log(n)}$. On the event E^c, we note that $|\Delta_n| \leq Cn^2$ and $|Z_n^*| \leq Cn^2\alpha_0^{-1}$ hold uniformly. It follows that

$$\begin{aligned}
\mathbb{E}[\Delta_n^2(Z_n^*)^2] &= \mathbb{E}[\Delta_n^2(Z_n^*)^2 \cdot I_{E_1^c}] + \mathbb{E}[\Delta_n^2(Z_n^*)^2 \cdot I_{E_1}] \\
&\leq Cn^8\alpha_0^{-2}\cdot\exp(-B\log(n)) + Cn^{-2}\alpha_0^{-1}\log(n)\mathbb{E}\big[(Z_n^*)^2 \cdot I_{E^c}\big] \\
&\leq o(1) + Cn^{-2}\alpha_0^{-1}\log(n)\big[(\mathbb{E}[Z_n^*])^2 + \mathrm{Var}(Z_n^*)\big] \\
&\leq o(1) + \frac{C\log(n)}{n^2\alpha_0}\Big[(n^2\alpha_0^{-1}\|Mh\|^2)^2 + n^2\alpha_0^{-1}\|Mh\|^2 + n\Big],
\end{aligned} \tag{E.15}$$

where in the last inequality we have used (E.7)-(E.8). We plug (E.15) into (E.14) to get

$$\begin{aligned}
\mathrm{Var}(\psi_n^{DC}) &\leq C\Big(1 + n^{-1}\alpha_0^{-1} + n^{-1/2}\delta_n + \frac{\log(n)}{n^3\alpha_0}(n\delta_n^2 + \sqrt{n}\delta_n + n)\Big) \\
&\leq C\Big[1 + n^{-1/2}\delta_n + n^{-2}\alpha_0^{-1}\delta_n^2\log(n)\Big].
\end{aligned} \tag{E.16}$$

This gives the second claim. □

Appendix F: Proof of Theorem 3.3

Write $\alpha_1 = \mathbb{E}[\hat{\alpha}_n]$, $\overline{\Omega} = \Omega - \alpha_1 \mathbb{1}_n\mathbb{1}_n'$ and $\Delta_n = \alpha_1 - \hat{\alpha}_n$. It follows that

$$Q_n = \sum_{i,j,k,l \text{ dist.}} (A_{ij} - \hat{\alpha}_n)(A_{jk} - \hat{\alpha}_n)(A_{kl} - \hat{\alpha}_n)(A_{li} - \hat{\alpha}_n)$$

Table 1. The 21 different types of the 81 post-expansion sums of Q_n. The order of the mean and variance of each term will be derived in the proofs.

Type	#	$(N_W, N_{\widetilde{\Omega}}, N_\Delta)$	Representative	Mean	Variance
X_1	1	$(4,0,0)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}W_{li}$	0	$O(n^4\alpha_0^4)$
X_2	4	$(3,1,0)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}\overline{\Omega}_{li}$	0	$O(n^4\alpha_0^3\|M\|^2)$
X_3	4	$(3,0,1)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}\Delta_n$	0	$O(n^2\alpha_0^4)$
X_4	4	$(2,2,0)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$	0	$O(n^4\alpha_0^2\|M\|^4)$
X_5	2	$(2,2,0)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}W_{kl}\overline{\Omega}_{li}$	0	$O(n^4\alpha_0^2\|M\|^4)$
X_6	8	$(2,1,1)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\overline{\Omega}_{kl}\Delta_n$	0	$O(n^3\alpha_0^3\|M\|^2)$
X_7	4	$(2,1,1)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}W_{kl}\Delta_n$	0	$O(n^2\alpha_0^3\|M\|^2)$
X_8	4	$(2,0,2)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\Delta_n^2$	$O(n^{1/2}\alpha_0^2)$	$O(n\alpha_0^4)$
X_9	2	$(2,0,2)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\Delta_n^2$	$O(\alpha_0^2)$	$O(\alpha_0^4)$
X_{10}	4	$(1,3,0)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$	0	$O(n^6\alpha_0\|M\|^6)$
X_{11}	8	$(1,2,1)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\Delta_n$	$O(n^2\alpha_0\|M\|^2)$	$O(n^4\alpha_0^2\|M\|^4)$
X_{12}	4	$(1,2,1)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\Delta_n\overline{\Omega}_{li}$	$O(n^2\alpha_0\|M\|^2)$	$O(n^4\alpha_0^2\|M\|^4)$
X_{13}	8	$(1,1,2)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\Delta_n^2$	$O(n^2\alpha_0^{3/2}\|M\|)$	$O(n^4\alpha_0^3\|M\|^2)$
X_{14}	4	$(1,1,2)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{kl}\Delta_n^2$	$O(n^2\alpha_0^{3/2}\|M\|)$	$O(n^4\alpha_0^3\|M\|^2)$
X_{15}	4	$(1,0,3)$	$\sum_{i,j,k,l \text{ dist.}} W_{ij}\Delta_n^3$	$O(\alpha_0^2)$	$O(\alpha_0^4)$
X_{16}	1	$(0,4,0)$	$\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$	$n^4\|M\|^4$	0
X_{17}	4	$(0,3,1)$	$\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\Delta_n$	0	$O(n^6\alpha_0\|M\|^6)$
X_{18}	4	$(0,2,2)$	$\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\Delta_n^2$	$O(n^2\alpha_0\|M\|^2)$	$O(n^4\alpha_0^2\|M\|^4)$
X_{19}	2	$(0,2,2)$	$\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{kl}\Delta_n^2$	$O(n^2\alpha_0\|M\|^2)$	$O(n^4\alpha_0^2\|M\|^4)$
X_{20}	4	$(0,1,3)$	$\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\Delta_n^3$	0	0
X_{21}	1	$(0,0,4)$	$\sum_{i,j,k,l \text{ dist.}} \Delta_n^4$	$O(\alpha_0^2)$	$O(\alpha_0^4)$

$$= \sum_{i,j,k,l \text{ dist.}} (W_{ij} + \overline{\Omega}_{ij} + \Delta_n)(W_{jk} + \overline{\Omega}_{jk} + \Delta_n)(W_{kl} + \overline{\Omega}_{kl} + \Delta_n)(W_{li} + \overline{\Omega}_{li} + \Delta_n).$$

Expanding the sum gives $3^4 = 81$ terms. Combining equal-valued terms, we have the following decomposition:

$$\begin{aligned} Q_n = & X_1 + 4X_2 + 4X_3 + 4X_4 + 2X_5 + 8X_6 + 4X_7 + 4X_8 + 2X_9 + 4X_{10} + 8X_{11} \\ & + 4X_{12} + 8X_{13} + 4X_{14} + 4X_{15} + X_{16} + 4X_{17} + 4X_{18} + 2X_{19} + 4X_{20} + X_{21}, \end{aligned} \tag{F.1}$$

where the expressions of X_1-X_{21} are presented in Column 4 of Table 1. In this table, we also list other information of each term, such as the degree in W (N_W), in $\overline{\Omega}$ ($N_{\overline{\Omega}}$) and in Δ_n (N_Δ). We plan to study the mean and variance of each of X_1-X_{21} and then combine them to show the claims.

In preparation, we derive some useful results. First, we study $|\overline{\Omega}_{ij}|$. Write $M = P - \alpha_0 \mathbf{1}_K \mathbf{1}_K'$. Then,

$$\begin{aligned} |\alpha_1 - \alpha_0| = |\mathbb{E}[\hat{\alpha}_n] - \alpha_0| &= \left| \frac{1}{n(n-1)} \sum_{i \neq j} \pi_i' M \pi_j \right| \\ &= \left| \frac{1}{n(n-1)} \sum_{i,j} \pi_i' M \pi_j - \frac{1}{n(n-1)} \sum_i \pi_i' M \pi_i \right| \\ &\le \frac{n}{n-1} \left| h' M h \right| + \frac{\|M\|}{n-1} \le \frac{C\|M\|}{n}, \end{aligned} \tag{F.2}$$

where we have used in the last line that $h'Mh = h'Ph - \alpha_0 h' \mathbf{1}_K \mathbf{1}_K' h = 0$.

Note that $\overline{\Omega}_{ij} = \pi_i' P \pi_j - \alpha_1 = \pi_i' M \pi_j + \alpha_0 - \alpha_1$. It follows that

$$|\overline{\Omega}_{ij}| \leq |\pi_i' M \pi_j| + |\alpha_0 - \alpha_1| \leq C\|M\|. \tag{F.3}$$

Next, we study Δ_n. By definition,

$$\Delta_n = \mathbb{E}[\hat{\alpha}_n] - \hat{\alpha}_n = -\frac{1}{n(n-1)} \sum_{i \neq j} (A_{ij} - \Omega_{ij}) = -\frac{1}{n(n-1)} \sum_{i \neq j} W_{ij}.$$

Using properties of Bernoulli variables, we have $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1 - \Omega_{ij}) \leq \Omega_{ij}$ and $|\mathbb{E}[W_{ij}^m]| \leq C\Omega_{ij}$, for any fixed $m \geq 3$ (the constant C may depend on m). Note that

$$\Omega_{ij} = \pi_i' P \pi_j \leq \mathbf{1}_K' P \mathbf{1}_K \leq C\alpha_0,$$

where we have used that $\min_k h_k > C/K$, which is a consequence of (3.4). Additionally,

$$\sum_{i,j} \Omega_{ij} = \sum_{i,j} \pi_i' P \pi_j = n^2 h' P h = n^2 \alpha_0.$$

It follows that

$$\begin{aligned}
\mathbb{E}[\Delta_n^2] &= \frac{4}{n^2(n-1)^2} \sum_{i<j} \mathbb{E}(W_{ij}^2) \leq Cn^{-4} \sum_{i \neq j} \Omega_{ij} \leq Cn^{-2}\alpha_0, \\
|\mathbb{E}[\Delta_n^3]| &= \frac{8}{n^3(n-1)^3} \left| \mathbb{E}\left[\sum_{i<j, k<l, u<v} W_{ij} W_{kl} W_{uv} \right] \right| = \frac{8}{n^3(n-1)^3} \left| \mathbb{E}\left[\sum_{i<j} W_{ij}^3 \right] \right| \\
&\leq Cn^{-6} \sum_{i<j} \Omega_{ij} \ \leq \ Cn^{-4}\alpha_0, \\
\mathbb{E}[\Delta_n^4] &= \frac{16}{n^4(n-1)^4} \left(\sum_{i<j} \mathbb{E}[W_{ij}^4] + 3 \sum_{\substack{i<j, k<l \\ (i,j) \neq (k,l)}} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{kl}^2] \right) \\
&\leq Cn^{-8} \left[\sum_{i<j} \Omega_{ij} + \Big(\sum_{i<j} \Omega_{ij} \Big) \Big(\sum_{k<\ell} \Omega_{k\ell} \Big) \right] \leq Cn^{-4}\alpha_0^2, \\
\mathbb{E}[\Delta_n^8] &\leq Cn^{-16} \left(\sum_{i<j, k<l, m<s, q<t} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{kl}^2]\mathbb{E}[W_{ms}^2]\mathbb{E}[W_{qt}^2] \right) \\
&\leq Cn^{-16} \Big(\sum_{i<j} \Omega_{ij} \Big)^4 \leq Cn^{-8}\alpha_0^4.
\end{aligned} \tag{F.4}$$

We shall frequently use (F.3) and (F.4) in the proof below.

Mean and variance of Q_n. We study the mean and variance of each of X_1-X_{21}, and combine them to get the mean and variance of Q_n.

Consider $X_1 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}W_{li}$. It is easy to see that

$$\mathbb{E}[X_1] = 0. \tag{F.5}$$

Furthermore, let $CC(I_n)$ be collection of equivalent classes of 4-tuples (i,j,k,l) (see the proof of (D.20) for details). By elementary probability,

$$\begin{aligned}\text{Var}(X_1) &= \text{Var}\left(8 \sum_{CC(I_n)} W_{ij}W_{jk}W_{kl}W_{li}\right) \\ &= 64 \sum_{CC(I_n)} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{jk}^2]\mathbb{E}[W_{kl}^2]\mathbb{E}[W_{li}^2] \\ &\le C \sum_{i,j,k,l} \Omega_{ij}\Omega_{jk}\Omega_{kl}\Omega_{li} \ \le C\text{Tr}(\Omega^4).\end{aligned}$$

Note that $\Omega = \Pi P \Pi'$ and $\Pi \mathbf{1}_n = \mathbf{1}_K$. Also, we have defined $G = n^{-1}\Pi'\Pi$ in Section 3.2. It follows that

$$\text{Tr}(\Omega^4) = n^4 \text{Tr}(PGPGPGPG) = n^4 \text{Tr}\left((G^{1/2}PG^{1/2})^4\right) \le Kn^4 \left\| G^{1/2}PG^{1/2} \right\|^4.$$

From the definition of G, we have $G_{kl} = n^{-1}\sum_{i,j} \pi_i(k)\pi_j(l) \le 1$ for all $1 \le k,l \le K$. Hence $\|G\| \le K^2$. In addition, recall that $\alpha_0 = h'Ph$. By our assumption (3.4), all the entries of h are lower bounded by a constant $C > 0$. It follows that $\alpha_0 \ge C\mathbf{1}_K' P \mathbf{1}_K$. We immediately have

$$\text{Tr}(\Omega^4) \le K^9 n^4 \|P\|^4 \le K^9 n^4 (\mathbf{1}_K' P \mathbf{1}_K)^4 \le Cn^4 \alpha_0^4,$$

where we have used that $\|P\| \le \mathbf{1}_K' P \mathbf{1}_K$ since P is a nonnegative matrix. Combining the above gives

$$\text{Var}(X_1) \le Cn^4 \alpha_0^4. \tag{F.6}$$

Next, consider $X_2 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}\overline{\Omega}_{li}$. It is easy to see that

$$\mathbb{E}[X_2] = 0. \tag{F.7}$$

Furthermore,

$$\text{Var}\,(X_2) = \text{Var}\left(2 \sum_{\substack{i,j,k,l \text{ dist.} \\ i<l}} W_{ij}W_{jk}W_{kl}\overline{\Omega}_{li}\right) \le C \sum_{\substack{i,j,k,l \text{ dist.} \\ i<l}} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{jk}^2]\mathbb{E}[W_{kl}^2]\overline{\Omega}_{li},$$

where we have used that summands in the expression above are pairwise independent. It follows that

$$\text{Var}(X_2) \le Cn^4 \alpha_0^3 \|M\|^2. \tag{F.8}$$

Next, consider $X_3 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}W_{kl}\Delta_n$. Recall that

$$\Delta_n = \alpha_1 - \hat{\alpha}_n = -\frac{2}{n(n-1)} \sum_{i<j} W_{ij}.$$

It follows that

$$\mathbb{E}[X_3] = -\frac{2}{n(n-1)}\mathbb{E}\left[\sum_{i,j,k,l \text{ dist.}}\sum_{s<t} W_{ij}W_{jk}W_{kl}W_{st}\right] = 0. \tag{F.9}$$

Furthermore,

$$\begin{aligned}
\text{Var}(X_3) &= \frac{1}{n^2(n-1)^2}\text{Var}\left(\sum_{\substack{i,j,k,l \text{ dist.}\\ s\neq t}} W_{ij}W_{jk}W_{kl}W_{st}\right)\\
&\le \frac{C}{n^4}\mathbb{E}\left[\sum_{\substack{i,j,k,l \text{ dist.}\\ a,b,c,d \text{ dist.}\\ s\neq t, u\neq v}} W_{ij}W_{jk}W_{kl}W_{st}W_{ab}W_{bc}W_{cd}W_{uv}\right]\\
&\le \frac{C}{n^4}\left(\sum_{i,j,k,l,s,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2W_{st}^2] + \sum_{i,j,k,l,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2W_{lt}^2]\right.\\
&\sum_{i,j,k,l,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2W_{kt}^2] + \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2W_{lj}^2] +\\
&\sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2W_{li}^2] + \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^4W_{jk}^2W_{kl}^2] +\\
&\left. + \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^4W_{kl}^2]\right).
\end{aligned}$$

It follows that

$$\text{Var}(X_3) \le \frac{C}{n^4}(n^6\alpha_0^4 + n^5\alpha_0^4 + n^4\alpha_0^4 + n^4\alpha_0^3) \le Cn^2\alpha_0^4. \tag{F.10}$$

Next, consider $X_4 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$. It is straightforward to see that

$$\mathbb{E}[X_4] = 0. \tag{F.11}$$

Furthermore,

$$\begin{aligned}
\text{Var}(X_4) &= \mathbb{E}\left[\sum_{\substack{i,j,k,l \text{ dist.}\\ u,v,s,t \text{ dist.}}} W_{ij}W_{jk}W_{uv}W_{vs}\overline{\Omega}_{kl}\overline{\Omega}_{li}\overline{\Omega}_{st}\overline{\Omega}_{tu}\right]\\
&\le C\|M\|^4 \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{jk}^2],
\end{aligned}$$

from which we obtain that

$$\text{Var}(X_4) \leq Cn^4\alpha_0^2\|M\|^4. \tag{F.12}$$

Next, consider $X_5 = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}W_{kl}\overline{\Omega}_{li}$. It is straightforward to see that

$$\mathbb{E}[X_5] = 0. \tag{F.13}$$

Furthermore,

$$\begin{aligned}\text{Var}(X_5) &= \mathbb{E}\left[\sum_{\substack{i,j,k,l \text{ dist.}\\ u,v,s,t \text{ dist.}}} W_{ij}W_{kl}W_{uv}W_{st}\overline{\Omega}_{jk}\overline{\Omega}_{li}\overline{\Omega}_{vs}\overline{\Omega}_{tu}\right]\\ &\leq C\|M\|^4 \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2]\mathbb{E}[W_{kl}^2],\end{aligned}$$

from which we obtain that

$$\text{Var}(X_5) \leq Cn^4\alpha_0^2\|M\|^4. \tag{F.14}$$

Next, consider $X_6 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\overline{\Omega}_{kl}\Delta_n$. Using the definition of Δ_n, we have

$$X_6 = -\frac{1}{n(n-1)} \sum_{i,j,k,l \text{ dist.}} \sum_{s \neq t} W_{ij}W_{jk}\overline{\Omega}_{kl}W_{st}.$$

It follows that

$$\mathbb{E}[X_6] = 0. \tag{F.15}$$

Furthermore,

$$\begin{aligned}\text{Var}(X_6) &= \frac{1}{n^2(n-1)^2} \sum_{\substack{i,j,k,l \text{ dist.}\\ a,b,c,d \text{ dist.}\\ s\neq t, u\neq v}} \overline{\Omega}_{kl}\overline{\Omega}_{cd}\mathbb{E}[W_{ij}W_{jk}W_{st}W_{ab}W_{bc}W_{uv}]\\ &\leq \frac{C\|M\|^2}{n^2} \sum_{\substack{i,j,k \text{ dist.}\\ a,b,c \text{ dist.}\\ s\neq t, u\neq v}} \mathbb{E}[W_{ij}W_{jk}W_{st}W_{ab}W_{bc}W_{uv}]\\ &\leq \frac{C\|M\|^2}{n^2}\Bigg(4\sum_{i,j,k,s,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{st}^2] + 4\sum_{i,j,k,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kt}^2] +\\ &\quad 2\sum_{i,j,k,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{jt}^2] + \sum_{i,j,k \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{ki}^2] + 4\sum_{i,j,k \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^4] +\\ &\quad 4\sum_{i,j,k \text{ dist.}} \mathbb{E}[W_{ij}^3W_{jk}^3]\Bigg).\end{aligned}$$

As a result, we obtain

$$\mathrm{Var}(X_6) \le \frac{C\|M\|^2}{n^2}(n^5\alpha_0^3 + n^4\alpha_0^3 + n^3\alpha_0^3 + n^3\alpha_0^2) \le Cn^3\alpha_0^3\|M\|^2. \tag{F.16}$$

Next, consider $X_7 = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}W_{kl}\Delta_n$. Similarly to X_6, it is easy to see that

$$\mathbb{E}[X_7] = 0. \tag{F.17}$$

Furthermore,

$$\begin{aligned}
\mathrm{Var}(X_7) &= \frac{1}{n^2(n-1)^2} \sum_{\substack{i,j,k,l \text{ dist.} \\ a,b,c,d \text{ dist.} \\ s\neq t, u\neq v}} \overline{\Omega}_{jk}\overline{\Omega}_{bc}\mathbb{E}[W_{ij}W_{kl}W_{st}W_{ab}W_{cd}W_{uv}] \\
&\le \frac{C\|M\|^2}{n^4}\Bigg(\sum_{i,j,k,l,s,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{kl}^2W_{st}^2] + \sum_{i,j,k,l,t \text{ dist.}} \mathbb{E}[W_{ij}^2W_{kl}^2W_{lt}^2] + \\
&\quad \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2W_{jk}^2W_{kl}^2] + \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^2W_{kl}^4] + \sum_{i,j,k,l \text{ dist.}} \mathbb{E}[W_{ij}^3W_{kl}^3]\Bigg).
\end{aligned}$$

As a result, we obtain

$$\mathrm{Var}(X_7) \le \frac{C\|M\|^2}{n^4}(n^6\alpha_0^3 + n^5\alpha_0^3 + n^4\alpha_0^3 + n^4\alpha_0^2) \le Cn^2\alpha_0^3\|M\|^2. \tag{F.18}$$

Next, consider $X_8 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{jk}\Delta_n^2$. We have

$$|\mathbb{E}[X_8]| = (n-3)\left|\mathbb{E}\left[\Delta_n^2 \sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\right]\right| \le n\mathbb{E}[\Delta_n^4]^{1/2}\mathbb{E}\left[\left(\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\right)^2\right]^{1/2}.$$

It follows that

$$|\mathbb{E}[X_8]| \le Cn^{-1}\alpha_0 n^{3/2}\alpha_0 \le Cn^{1/2}\alpha_0^2. \tag{F.19}$$

Furthermore,

$$\begin{aligned}
\mathrm{Var}(X_8) &\le Cn^2\mathbb{E}\left[\Delta_n^4\left(\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\right)^2\right] \\
&\le Cn^2\mathbb{E}[\Delta_n^8]^{1/2}\mathbb{E}\left[\left(\sum_{i,j,k \text{ dist.}} W_{ij}W_{jk}\right)^4\right]^{1/2}.
\end{aligned}$$

The summands above can be grouped into 6 categories, where each category corresponds to a specific upper bound in terms of n and α_0. We obtain

$$\mathrm{Var}(X_8) \le Cn^{-2}\alpha_0^2(n^6\alpha_0^4 + n^5\alpha_0^4 + n^4\alpha_0^4 + n^4\alpha_0^3 + n^3\alpha_0^3 + n^3\alpha_0^2)^{1/2} \le Cn\alpha_0^4. \tag{F.20}$$

Next, consider $X_9 = \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\Delta_n^2$. We have

$$|\mathbb{E}[X_9]| = \left|\mathbb{E}\left[\Delta_n^2 \sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\right]\right| \leq \mathbb{E}[\Delta_n^4]^{1/2}\mathbb{E}\left[\left(\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\right)^2\right]^{1/2}.$$

It follows that

$$|\mathbb{E}[X_9]| \leq Cn^{-2}\alpha_0 n^2\alpha_0 \leq C\alpha_0^2. \tag{F.21}$$

Furthermore,

$$\begin{aligned}\text{Var}(X_9) &\leq C\mathbb{E}\left[\Delta_n^4\left(\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\right)^2\right] \\ &\leq C\mathbb{E}[\Delta_n^8]^{1/2}\mathbb{E}\left[\left(\sum_{i,j,k,l \text{ dist.}} W_{ij}W_{kl}\right)^4\right]^{1/2}.\end{aligned}$$

As for X_8, the summands above can be grouped into 6 categories, where each category corresponds to a specific upper bound in terms of n and α_0. We obtain

$$\text{Var}(X_9) \leq Cn^{-4}\alpha_0^2(n^8\alpha_0^4 + n^7\alpha_0^4 + n^6\alpha_0^4 + n^5\alpha_0^4 + n^4\alpha_0^4 + n^4\alpha_0^2)^{1/2} \leq C\alpha_0^4. \tag{F.22}$$

Next, consider $X_{10} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$. It is straightforward to see that

$$|\mathbb{E}[X_{10}]| = 0. \tag{F.23}$$

Furthermore,

$$\begin{aligned}\text{Var}(X_{10}) &= \sum_{\substack{i,j,k,l \text{ dist.} \\ a,b,c,d \text{ dist.}}} \overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}\overline{\Omega}_{bc}\overline{\Omega}_{cd}\overline{\Omega}_{da}\mathbb{E}[W_{ij}W_{ab}] \\ &\leq C\alpha_0 \sum_{\substack{i,j,k,l \text{ dist.} \\ c\neq d, c,d\notin\{i,j\}}} |\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}\overline{\Omega}_{jc}\overline{\Omega}_{cd}\overline{\Omega}_{di}|.\end{aligned}$$

As a result,

$$\text{Var}(X_{10}) \leq C\alpha_0 n^6\|M\|^6. \tag{F.24}$$

Next, consider $X_{11} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\Delta_n$. Using the definition of Δ_n, we obtain

$$|\mathbb{E}[X_{11}]| = \left|\frac{1}{n(n-1)}\sum_{\substack{i,j,k,l \text{ dist.} \\ u\neq v}} \overline{\Omega}_{jk}\overline{\Omega}_{kl}\mathbb{E}[W_{ij}W_{uv}]\right| \leq C\|M\|^2\sum_{i\neq j, u\neq v}|\mathbb{E}[W_{ij}W_{uv}]|.$$

As a result,

$$|\mathbb{E}[X_{11}]| \leq Cn^2\alpha_0\|M\|^2. \tag{F.25}$$

Furthermore,

$$\begin{aligned}
\mathrm{Var}(X_{11}) &\leq C\mathbb{E}\left[\left(\frac{1}{n(n-1)}\sum_{\substack{i,j,k,l \text{ dist.}\\ u\neq v}} \overline{\Omega}_{jk}\overline{\Omega}_{kl}W_{ij}W_{uv}\right)^2\right] \\
&\leq \frac{C}{n^4}\sum_{\substack{i,j,k,l \text{ dist.}\\ a,b,c,d \text{ dist.}\\ u\neq v, r\neq s}} |\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{bc}\overline{\Omega}_{cd}||\mathbb{E}[W_{ij}W_{uv}W_{ab}W_{rs}]| \\
&\leq C\|M\|^4\sum_{\substack{i\neq j, a\neq b\\ u\neq v, r\neq s}} |\mathbb{E}[W_{ij}W_{uv}W_{ab}W_{rs}]| \\
&\leq C\|M\|^4\left(\sum_{i,j,a,b \text{ dist.}}\mathbb{E}[W_{ij}^2W_{ab}^2] + \sum_{i,j,b \text{ dist.}}\mathbb{E}[W_{ij}^2W_{jb}^2] + \sum_{i,j \text{ dist.}}\mathbb{E}[W_{ij}^4]\right).
\end{aligned}$$

As a result,

$$\mathrm{Var}(X_{11}) \leq C\|M\|^4(n^4\alpha_0^2 + n^3\alpha_0^2 + n^2\alpha_0) \leq Cn^4\alpha_0^2\|M\|^4. \tag{F.26}$$

Next, consider $X_{12} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\Delta_n\overline{\Omega}_{li}$. Computations in this case are exactly equivalent to those for X_{11}, so we obtain:

$$|\mathbb{E}[X_{12}]| \leq Cn^2\alpha_0\|M\|^2. \tag{F.27}$$

and

$$\mathrm{Var}(X_{12}) \leq C\|M\|^4(n^4\alpha_0^2 + n^3\alpha_0^2 + n^2\alpha_0) \leq Cn^4\alpha_0^2\|M\|^4. \tag{F.28}$$

Next, consider $X_{13} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{jk}\Delta_n^2$. We have for the mean:

$$\begin{aligned}
|\mathbb{E}[X_{13}]| &\leq \sum_{i,j,k,l \text{ dist.}} |\overline{\Omega}_{jk}|\mathbb{E}[W_{ij}\Delta_n^2] \leq \sum_{i,j,k,l \text{ dist.}} |\overline{\Omega}_{jk}|\mathbb{E}[W_{ij}^2]^{1/2}E[\Delta_n^4]^{1/2} \\
&\leq Cn^4\|M\|\alpha_0^{1/2}E[\Delta_n^4]^{1/2}.
\end{aligned}$$

It follows that

$$|\mathbb{E}[X_{13}]| \leq Cn^2\alpha_0^{3/2}\|M\|. \tag{F.29}$$

Furthermore,

$$\mathrm{Var}(X_{13}) \leq \mathbb{E}\left[\sum_{\substack{i,j,k,l \text{ dist.}\\ a,b,c,d \text{ dist.}}} W_{ij}W_{ab}\overline{\Omega}_{jk}\overline{\Omega}_{bc}\Delta_n^4\right] \leq Cn^4\|M\|^2\sum_{i\neq j, a\neq b}\mathbb{E}[W_{ij}W_{ab}\Delta_n^4]$$

$$\le Cn^4\|M\|^2 \sum_{i\neq j, a\neq b} \mathbb{E}[W_{ij}^2 W_{ab}^2]^{1/2}\mathbb{E}[\Delta_n^8]^{1/2} \le C\alpha_0^2\|M\|^2 \sum_{i\neq j, a\neq b} \mathbb{E}[W_{ij}^2 W_{ab}^2]^{1/2}$$

$$\le C\alpha_0^2\|M\|^2 \left(\sum_{i,j,a,b \text{ dist.}} \mathbb{E}[W_{ij}^2]^{1/2}\mathbb{E}[W_{ab}^2]^{1/2} + \sum_{i,j,b \text{ dist.}} \mathbb{E}[W_{ij}^2]^{1/2}\mathbb{E}[W_{jb}^2]^{1/2} + \right.$$

$$\left. \sum_{i,j \text{ dist.}} \mathbb{E}[W_{ij}^4]^{1/2} \right).$$

As a result,

$$\text{Var}(X_{13}) \le C\alpha_0^2\|M\|^2(n^4\alpha_0 + n^3\alpha_0 + n^2\alpha_0^{1/2}) \le Cn^4\alpha_0^3\|M\|^2. \tag{F.30}$$

Next, consider $X_{14} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\overline{\Omega}_{kl}\Delta_n^2$. Computations in this case are exactly equivalent to those for X_{13}, so we obtain:

$$|\mathbb{E}[X_{14}]| \le Cn^2\alpha_0^{3/2}\|M\|. \tag{F.31}$$

and

$$\text{Var}(X_{14}) \le C\alpha_0^2\|M\|^2(n^4\alpha_0 + n^3\alpha_0 + n^2\alpha_0^{1/2}) \le Cn^4\alpha_0^3\|M\|^2. \tag{F.32}$$

Next, consider $X_{15} = \sum_{i,j,k,l \text{ dist.}} W_{ij}\Delta_n^3$. Using the definition of Δ_n, note that

$$X_{15} = (n-2)(n-3)\Delta_n^3 \sum_{i\neq j} W_{ij} = -n(n-1)(n-2)(n-3)\Delta_n^4.$$

It follows that

$$|\mathbb{E}[X_{15}]| \le n^4\mathbb{E}[\Delta_n^4] \le C\alpha_0^2. \tag{F.33}$$

and

$$\text{Var}(X_{15}) \le n^8\mathbb{E}[\Delta_n^8] \le C\alpha_0^4. \tag{F.34}$$

Next, consider $X_{16} = \sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li}$. This is a non-stochastic term, whose variance is zero. We the focus on deriving a lower bound for $\mathbb{E}[X_{16}] = X_{16}$. Note that

$$\begin{aligned} X_{16} &= \sum_{i,j,k,l} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li} - \sum_{i,j,k,l \text{ not dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li} \\ &= \text{Tr}(\overline{\Omega}^4) - \sum_{i,j,k,l \text{ not dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\overline{\Omega}_{li} \\ &= \text{Tr}(\overline{\Omega}^4) - O\big(n^3\|M\|^4\big), \end{aligned} \tag{F.35}$$

where the last equality comes from (F.3) and the observation that (i,j,k,l) has at most 3 distinct values in this sum. In the derivation of (F.3), we have seen that $\overline{\Omega}_{ij} = \pi_i' P\pi_j - \alpha_1 = \pi_i'\overline{M}\pi_j$, where $\overline{M} = P - \alpha_1\mathbf{1}_K\mathbf{1}_K' = M + (\alpha_0 - \alpha_1)\mathbf{1}_K\mathbf{1}_K'$. This implies that

$$\overline{\Omega} = \Pi\overline{M}\Pi'.$$

Recall that $G = n^{-1}\Pi'\Pi$. We have

$$\begin{aligned}\text{Tr}(\overline{\Omega}^4) &= \text{Tr}((\Pi\overline{M}\Pi')^4) = n^4\text{Tr}((G^{1/2}\overline{M}G^{1/2})^4) \\ &= n^4\|(G^{1/2}\overline{M}G^{1/2})^2\|_F^2 \\ &\asymp n^4\|(G^{1/2}\overline{M}G^{1/2})^2\|^2 \\ &\asymp n^4\|G^{1/2}\overline{M}G^{1/2}\|^4.\end{aligned}$$

Note that $\|G^{1/2}\overline{M}G^{1/2}\| \leq \|\overline{M}\|\|G\|$. Additionally, $\|\overline{M}\| \leq \|G^{-1}\|\|G^{1/2}\overline{M}G^{1/2}\|$. By the definition of G and our assumption (3.4), $\|G\| \leq C$ and $\|G^{-1}\| \leq C$. It follows that $\|G^{1/2}\overline{M}G^{1/2}\| \asymp \|\overline{M}\|$. We thus have

$$\text{Tr}(\overline{\Omega}^4) \asymp n^4\|\overline{M}\|^4 = n^4\|M + (\alpha_0 - \alpha_1)\mathbf{1}_K\mathbf{1}_K'\|^4.$$

Recall now from (F.2) that $|\alpha_0 - \alpha_1| = O(n^{-1}\|M\|)$. Hence, by Weyl's inequality

$$\big|\|\overline{M}\| - \|M\|\big| \leq K|\alpha_0 - \alpha_1| \leq \frac{CK\|M\|}{n},$$

which implies that $\|\overline{M}\| \asymp \|M\|$, so $\text{Tr}(\overline{\Omega}^4) \asymp n^4\|M\|^4$. Plugging it into (F.35) gives

$$X_{16} = \mathbb{E}[X_{16}] \asymp n^4\|M\|^4. \tag{F.36}$$

Next, consider $X_{17} = \sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\Delta_n$. It is straightforward to see that

$$\mathbb{E}[X_{17}] = 0. \tag{F.37}$$

Furthermore,

$$\text{Var}(X_{17}) \leq \left(\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\overline{\Omega}_{kl}\right)^2 \mathbb{E}[\Delta_n^2] \leq C\alpha_0 n^6\|M\|^6. \tag{F.38}$$

Next, consider $X_{18} = \sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\Delta_n^2$. We first note that $X_{18} = (n-3)\Delta_n^2 \sum_{i,j,k \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}$. Hence,

$$|\mathbb{E}[X_{18}]| \leq \frac{C\alpha_0}{n}\left|\sum_{i,j,k \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{jk}\right| \leq C\alpha_0 n^2\|M\|^2. \tag{F.39}$$

Furthermore,

$$\text{Var}(X_{18}) \leq n^2\left(\sum_{i,j,k \text{ dist.}} \tilde{\Omega}_{ij}\tilde{\Omega}_{jk}\right)^2 \mathbb{E}[\Delta_n^4] \leq C\alpha_0^2 n^4\|M\|^4. \tag{F.40}$$

Next, consider $X_{19} = \sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{kl}\Delta_n^2$. We have

$$|\mathbb{E}[X_{19}]| \leq \frac{C\alpha_0}{n^2}\left|\sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\overline{\Omega}_{kl}\right| \leq C\alpha_0 n^2\|M\|^2. \tag{F.41}$$

Furthermore,

$$\mathrm{Var}(X_{19}) \le \left(\sum_{i,j,k,l \text{ dist.}} \tilde{\Omega}_{ij}\tilde{\Omega}_{kl} \right)^2 \mathbb{E}[\Delta_n^4] \le C\alpha_0^2 n^4 \|M\|^4. \tag{F.42}$$

Next, consider $X_{20} = \sum_{i,j,k,l \text{ dist.}} \overline{\Omega}_{ij}\Delta_n^3$. Notice that

$$X_{20} = \Delta_n^3(n-2)(n-3)\sum_{i\neq j}\overline{\Omega}_{ij} = \Delta_n^3(n-2)(n-3)\left(\sum_{i\neq j}\Omega_{ij} - n(n-1)\alpha_1\right) = 0.$$

It follows that

$$\mathbb{E}[X_{20}] = 0, \tag{F.43}$$

and

$$\mathrm{Var}(X_{20}) = 0. \tag{F.44}$$

Next, consider $X_{21} = \sum_{i,j,k,l \text{ dist.}} \Delta_n^4$. Note that $X_{21} = n(n-1)(n-2)(n-3)\Delta_n^4$. As a result,

$$\mathbb{E}[X_{21}] \le C\alpha_0^2, \tag{F.45}$$

and

$$\mathrm{Var}(X_{21}) \le C\alpha_0^4. \tag{F.46}$$

Mean and variance of $Q_n/(2\sqrt{2}n^2\alpha_0^2)$. We use the results stored in Table 1 in order to provide a lower bound for $\mathbb{E}[Q_n/(2\sqrt{2}n^2\alpha_0^2)]$ and an upper bound for $\mathrm{Var}(Q_n/(2\sqrt{2}n^2\alpha_0^2))$. Recall that we defined

$$\tau_n = \left(\frac{n\|M\|^2}{\alpha_0}\right)^2.$$

We obtain that

$$\begin{aligned}\mathbb{E}\left[\frac{Q_n}{2\sqrt{2}n^2\alpha_0^2}\right] &\asymp n^4\|M\|^4 + O(n^{1/2}\alpha_0^2 + n^2\alpha_0\|M\|^2 + n^2\alpha_0^{3/2}\|M\|) \\ &\asymp \tau_n\left(1 + O\left(\frac{1}{n^{3/2}\tau_n} + \frac{1}{n\tau_n^{1/2}} + \frac{1}{n^{1/2}\tau_n^{3/4}}\right)\right).\end{aligned} \tag{F.47}$$

Similarly, we observe that

$$\begin{aligned}\mathrm{Var}\left(\frac{Q_n}{2\sqrt{2}n^2\alpha_0^2}\right) &= O\left(\frac{n^4\alpha_0^4 + n^4\alpha_0^3\|M\|^2 + n^4\alpha_0^2\|M\|^4 + n^6\alpha_0\|M\|^6}{n^4\alpha_0^4}\right) \\ &= O\left(1 + \frac{\tau_n^{1/2}}{n} + \frac{\tau_n}{n^2} + \frac{\tau_n^{3/2}}{n}\right).\end{aligned} \tag{F.48}$$

Assuming that $\tau_n \geq C$, then we can write

$$\mathbb{E}\left[\frac{Q_n}{2\sqrt{2}n^2\alpha_0^2}\right] \asymp \tau_n, \qquad \text{and} \qquad \text{Var}\left(\frac{Q_n}{2\sqrt{2}n^2\alpha_0^2}\right) = O\left(1 + \frac{\tau_n^{3/2}}{n}\right). \tag{F.49}$$

Mean and variance of ψ_n^{SQ}. Recall that

$$\psi_n^{SQ} = \frac{Q_n}{2\sqrt{2}n^2\hat{\alpha}_n^2}.$$

In the sequel, we let $Z_n^* = Q_n/(2\sqrt{2}n^2\alpha_0^2)$ for ease of notation. First, we compute a lower bound on the mean of ψ_n^{SQ}. Note that

$$\begin{aligned}
\mathbb{E}[\psi_n^{SQ}] = \mathbb{E}\left[\left(\frac{\alpha_0}{\hat{\alpha}_n}\right)^2 Z_n^*\right] &\geq \mathbb{E}[Z_n^*] + 2\mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right) Z_n^*\right] + \mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right)^2 Z_n^*\right] \\
&\geq \mathbb{E}[Z_n^*] - C\sqrt{\mathbb{E}\left[(Z_n^*)^2\right]}\left\{\sqrt{\mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right)^2\right]} + \sqrt{\mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right)^4\right]}\right\}.
\end{aligned}$$

Under the event E defined in Appendix E, it holds that $|\hat{\alpha}_n - \alpha_0| < \delta\alpha_0$, so we can derive the following upper bound:

$$\left|\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right| \leq \frac{|\alpha_0 - \hat{\alpha}_n|}{(1-\delta)\alpha_0}.$$

Under E^c, it holds that

$$\left|\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right| \leq Cn^2.$$

We thus have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right)^2\right] &\leq Cn^4\mathbb{P}(E^c) + C\alpha_0^{-2}\mathbb{E}[(\alpha_0 - \hat{\alpha}_n)^2] \\
&\leq Cn^4\mathbb{P}(E^c) + C\alpha_0^{-2}(\alpha_0 - \mathbb{E}[\hat{\alpha}_n])^2 + C\alpha_0^{-2}\text{Var}(\hat{\alpha}_n) \\
&\leq Cn^4\mathbb{P}(E^c) + \frac{C}{(n-1)^2} + \frac{C}{n^2\alpha_0} \leq \frac{C}{n^2\alpha_0} = o(1).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\alpha_0 - \hat{\alpha}_n}{\hat{\alpha}_n}\right)^4\right] &\leq Cn^8\mathbb{P}(E^c) + C\alpha_0^{-4}\mathbb{E}[(\alpha_0 - \hat{\alpha}_n)^4] \\
&\leq Cn^8\mathbb{P}(E^c) + C\alpha_0^{-4}\mathbb{E}[(\hat{\alpha}_n - \mathbb{E}[\hat{\alpha}_n])^4] + C\alpha_0^{-4}(\mathbb{E}[\hat{\alpha}_n] - \alpha_0)^4 \\
&\leq Cn^8\mathbb{P}(E^c) + \frac{C}{n^4}
\end{aligned}$$

$$+\frac{C\alpha_0^{-4}}{n^8}\mathbb{E}\left[\sum_{\substack{i<j,k<l\\u<v,r<t}}(A_{ij}-\Omega_{ij})(A_{kl}-\Omega_{kl})(A_{uv}-\Omega_{uv})(A_{rs}-\Omega_{rs})\right]$$

$$\leq Cn^8\mathbb{P}(E^c)+\frac{C}{n^4}+\frac{C\alpha_0^{-4}}{n^8}(n^4\alpha_0^2+n^2\alpha_0)\leq\frac{C}{\alpha_0^2n^4}=o(1).$$

It follows that, for n big enough,

$$\begin{aligned}\mathbb{E}[\psi_n^{SQ}]&\geq\mathbb{E}[Z_n^*]-o\left(\sqrt{\mathbb{E}\left[(Z_n^*)^2\right]}\right)=\mathbb{E}[Z_n^*]-o\left(\sqrt{\text{Var}(Z_n^*)+\mathbb{E}[Z_n^*]^2}\right)\\&=\mathbb{E}[Z_n^*]-o\left(\sqrt{1+n^{-1}\tau_n^{1/2}+n^{-2}\tau_n+n^{-1}\tau_n^{3/2}+\mathbb{E}[Z_n^*]^2}\right)\\&\geq\mathbb{E}[Z_n^*](1-o(1))-o\left(1+n^{-1/2}\tau_n^{1/4}+n^{-1}\tau_n^{1/2}+n^{-1/2}\tau_n^{3/4}\right).\end{aligned}$$

Assuming that $\tau_n\geq C$, we know from (F.49) that there exists a constant $c_2>0$ such that

$$\mathbb{E}[\psi_n^{SQ}]\geq c_2\tau_n-o\left(1+n^{-1/2}\tau_n^{3/4}\right).\tag{F.50}$$

Next, we compute an upper bound on the variance of ψ_n^{SQ}. We have

$$\begin{aligned}\text{Var}(\psi_n^{SQ})&=\text{Var}\left(\left(\frac{\alpha_0}{\hat{\alpha}_n}\right)^2Z_n^*\right)=\text{Var}\left(\left(\frac{\alpha_0-\hat{\alpha}_n}{\hat{\alpha}_n}+1\right)^2Z_n^*\right)\\&\leq C\text{Var}(Z_n^*)+C\mathbb{E}\left[\left(\frac{\alpha_0-\hat{\alpha}_n}{\hat{\alpha}_n}\right)^2(Z_n^*)^2\right]+C\mathbb{E}\left[\left(\frac{\alpha_0-\hat{\alpha}_n}{\hat{\alpha}_n}\right)^4(Z_n^*)^2\right].\end{aligned}$$

Recall the event E_1 defined in Appendix E. We had that $\mathbb{P}(E_1^c)\leq\exp(-B\log(n))$, where B is a constant chosen large enough. Then, on the event E_1, we have that $|(\alpha_0-\hat{\alpha}_n)/\hat{\alpha}_n|\leq Cn^{-1}\alpha_0^{-1/2}\sqrt{\log(n)}$. On the event E_1^c, it holds uniformly that $|(\alpha_0-\hat{\alpha}_n)/\hat{\alpha}_n|\leq Cn^2$ and $|Z_n^*|\leq n^2\alpha_0^{-2}$. It follows that

$$\mathbb{E}\left[\left(\frac{\alpha_0-\hat{\alpha}_n}{\hat{\alpha}_n}\right)^2(Z_n^*)^2\right]\leq Cn^8\alpha_0^{-4}\mathbb{P}(E_1^c)+Cn^{-2}\alpha_0^{-1}\log(n)\mathbb{E}[(Z_n^*)^2],$$

$$\mathbb{E}\left[\left(\frac{\alpha_0-\hat{\alpha}_n}{\hat{\alpha}_n}\right)^4(Z_n^*)^2\right]\leq Cn^{12}\alpha_0^{-4}\mathbb{P}(E_1^c)+Cn^{-4}\alpha_0^{-2}\log(n)^2\mathbb{E}[(Z_n^*)^2].$$

So we obtain that

$$\begin{aligned}\text{Var}(\psi_n^{SQ})&\leq C\text{Var}(Z_n^*)+Cn^{-2}\alpha_0^{-1}\log(n)\mathbb{E}[(Z_n^*)^2]+o(1)\\&\leq C\text{Var}(Z_n^*)+Cn^{-2}\alpha_0^{-1}\log(n)\mathbb{E}[(Z_n^*)]^2+o(1).\end{aligned}\tag{F.51}$$

Recall from (F.49) that when $\tau_n \geq C$, $\mathbb{E}[Z_n^*] \asymp \tau_n$ and $\text{Var}(Z_n^*) = O(1 + n^{-1}\tau_n^{3/2})$. It follows that

$$\text{Var}(\psi_n^{SQ}) = O\left(1 + \frac{\tau_n^{3/2}}{n} + \frac{\log(n)\tau_n^2}{n^2\alpha_0}\right). \tag{F.52}$$

□

Appendix G: Proof of Corollary 3.2

Let ψ_n^{DC} denote the degree test statistic as in the proof of Theorem 3.2. Let $\epsilon \in (0,1)$ and q_ϵ be the $(1-\epsilon)$-quantile of the standard normal distribution.

Under the alternative hypothesis, we suppose that $\delta_n \to \infty$. It follows from Theorem 3.2 that

$$\mathbb{E}[\psi_n^{DC}] \geq c_1\delta_n, \quad \text{and} \quad \text{Var}(\psi_n^{DC}) = O(1 + n^{-1/2}\delta_n + n^{-2}\alpha_0^{-1}\delta_n^2\log(n)).$$

We have, for n big enough,

$$\mathbb{P}\left(\psi_n^{DC} < q_\epsilon\right) = \mathbb{P}\left(\mathbb{E}[\psi_n^{DC}] - \psi_n^{DC} > \mathbb{E}[\psi_n^{DC}] - q_\epsilon\right) \leq \frac{C\text{Var}(\psi_n^{DC})}{\mathbb{E}[\psi_n^{DC}]^2} \asymp \frac{1}{\text{SNR}(\psi_n^{DC})^2},$$

where we have seen that $\text{SNR}(\psi_n^{DC}) \to \infty$ if $\delta_n \to \infty$ under the alternative (see the paragraph before the statement of Corollary 3.2). It follows that under the alternative, the power of the test

$$\mathbb{P}\left(\psi_n^{DC} > q_\epsilon\right) \xrightarrow[n\to\infty]{} 1. \tag{G.1}$$

Furthermore, under the null hypothesis, we know from Corollary 3.1 that $\psi_n^{DC} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, hence the level of the test tends to ϵ as $n \to \infty$. □

Appendix H: Proof of Corollary 3.3

Let ψ_n^{SQ} denote the degree test statistic as in the proof of Theorem 3.3. Let $\epsilon \in (0,1)$ and q_ϵ be the $(1-\epsilon)$-quantile of the standard normal distribution.

Under the alternative hypothesis, we suppose that $\tau_n \to \infty$. It follows from Theorem 3.2 that

$$\mathbb{E}[\psi_n^{SQ}] \geq c_2\tau_n, \quad \text{and} \quad \text{Var}(\psi_n^{SQ}) = O(1 + n^{-1}\tau_n^{3/2} + n^{-2}\alpha_0^{-1}\tau_n^2\log(n)).$$

We have, for n big enough,

$$\mathbb{P}\left(\psi_n^{SQ} < q_\epsilon\right) = \mathbb{P}\left(\mathbb{E}[\psi_n^{SQ}] - \psi_n^{SQ} > \mathbb{E}[\psi_n^{SQ}] - q_\epsilon\right) \leq \frac{C\text{Var}(\psi_n^{SQ})}{\mathbb{E}[\psi_n^{SQ}]^2} \asymp \frac{1}{\text{SNR}(\psi_n^{SQ})^2},$$

where we have seen that $\text{SNR}(\psi_n^{SQ}) \to \infty$ if $\tau_n \to \infty$ under the alternative (see the paragraph before the statement of Corollary 3.3). It follows that under the alternative, the power of the test

$$\mathbb{P}\left(\psi_n^{SQ} > q_\epsilon\right) \xrightarrow[n\to\infty]{} 1. \tag{H.1}$$

Furthermore, under the null hypothesis, we know from Corollary 3.1 that $\psi_n^{SQ} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, hence the level of the test tends to ϵ as $n \to \infty$. □

Appendix I: Proof of Theorem 3.4

As in the proofs of Theorem 3.2 and Theorem 3.3, we let ψ_n^{DC} denote the degree chi-squared test statistic and ψ_n^{SQ} denote the oSQ statistic. Recall that the PET statistic is

$$S_n = \left(\psi_n^{DC}\right)^2 + \left(\psi_n^{SQ}\right)^2.$$

Let $A > 0, \epsilon > 0$ be arbitrary constants. Then,

$$\begin{aligned}\mathbb{P}(S_n < A) &\leq \min\left\{\mathbb{P}\left(\psi_n^{DC} < \sqrt{A}\right), \mathbb{P}\left(\psi_n^{SQ} < \sqrt{A}\right)\right\}\\ &\leq \min\left\{\mathbb{P}\left(\mathbb{E}[\psi_n^{DC}] - \psi_n^{DC} > \mathbb{E}[\psi_n^{DC}] - \sqrt{A}\right), \mathbb{P}\left(\mathbb{E}[\psi_n^{SQ}] - \psi_n^{SQ} > \mathbb{E}[\psi_n^{SQ}] - \sqrt{A}\right)\right\}.\end{aligned}$$

In the regime where $\max\{\delta_n, \tau_n\} \to \infty$, for any constant $B > 0$, there exists $N > 0$ such that for all $n > N$, $\delta_n > B$ or $\tau_n > B$. We will denote by $N(B)$ the smallest such constant. We choose $B \gg \sqrt{A}$ and $N > N(B)$ such that for all $n > N$,

$$\frac{1}{B^2} + \frac{1}{n^{1/2}B} + \frac{\log(n)}{n^2\alpha_0} < \frac{\epsilon}{C}$$
$$\frac{1}{B^2} + \frac{1}{nB^{1/2}} + \frac{\log(n)}{n^2\alpha_0} < \frac{\epsilon}{C}$$

Now, suppose that we are in the case $\delta_n > B$. Then from Theorem 3.2, we know that

$$\mathbb{E}[\psi_n^{DC}] > c\delta_n > cB \qquad \text{and} \qquad \text{Var}(\psi_n^{DC}) < C\left(1 + \frac{\delta_n}{n^{1/2}} + \frac{\log(n)}{n^2\alpha_0}\delta_n^2\right).$$

Then,

$$\mathbb{P}\left(\mathbb{E}[\psi_n^{DC}] - \psi_n^{DC} > \mathbb{E}[\psi_n^{DC}] - \sqrt{A}\right) \leq \frac{\text{Var}(\psi_n^{DC})}{\mathbb{E}[\psi_n^{DC}]^2} \leq C\left(\frac{1}{\delta_n^2} + \frac{1}{n^{1/2}\delta_n} + \frac{\log(n)}{n^2\alpha_0}\right) < \epsilon,$$

which implies that $\mathbb{P}(S_n < A) < \epsilon$.

Now, suppose that we are in the case $\tau_n > B$. By Theorem 3.3, we have

$$\mathbb{E}[\psi_n^{SQ}] \geq c\tau_n > cB \quad \text{and} \quad \text{Var}(\psi_n^{SQ}) < C\left(1 + \frac{\tau_n^{3/2}}{n} + \frac{\log(n)\tau_n^2}{n^2\alpha_0}\right).$$

Then

$$\mathbb{P}\left(\mathbb{E}[\psi_n^{SQ}] - \psi_n^{SQ} > \mathbb{E}[\psi_n^{SQ}] - \sqrt{A}\right) \leq \frac{\text{Var}(\psi_n^{SQ})}{\mathbb{E}[\psi_n^{SQ}]^2} \leq C\left(\frac{1}{\tau_n^2} + \frac{1}{n\tau_n^{1/2} + \frac{\log(n)}{n^2\alpha_0}}\right) < \epsilon,$$

which implies that $\mathbb{P}(S_n < A) < \epsilon$.

It follows that for all $n > N$, it holds that $\mathbb{P}(S_n < A) < \epsilon$. We have just shown that

$$S_n \xrightarrow[n\to\infty]{\mathbb{P}} \infty. \tag{I.1}$$

Now, fix $\epsilon \in (0,1)$ and let q denote the $(1-\epsilon)$-quantile of the $\chi_2^2(0)$ distribution. From Corollary 3.1, we know that as $n \to \infty$, the level of the test tends to ϵ. From (I.1), we know that under the alternative

$$\mathbb{P}(S_n > q) \xrightarrow[n\to\infty]{} 1,$$

so the power of the test tends to 1 as $n \to \infty$. □

Appendix J: Proof of Theorem 3.5

Denote by $D_{\chi^2}(P_0\|P_1)$ the chi-square divergence between two hypotheses, where P_0 and P_1 denote the probability measures under two model, respectively. and then study the symmetric alternative and the asymmetric alternative separately. By definition,

$$1 + D_{\chi^2}(P_0\|P_1) = \int \left(\frac{dP_1}{dP_0}\right)^2 dP_0.$$

Letting $q_{ij}(\Pi) = \pi_i' P \pi_j$, we can write

$$dP_0 = \prod_{i<j} \alpha^{A_{ij}}(1-\alpha)^{1-A_{ij}}, \qquad dP_1 = \mathbb{E}_{\Pi}\left[\prod_{i<j} q_{ij}(\Pi)^{A_{ij}}(1-q_{ij}(\Pi))^{1-A_{ij}}\right].$$

Let $\tilde{\Pi}$ be an independent copy of Π. Then it follows that

$$\left(\frac{dP_1}{dP_0}\right)^2 = \mathbb{E}_{\Pi,\tilde{\Pi}}\left[\prod_{i<j}\left(\frac{q_{ij}(\Pi)q_{ij}(\tilde{\Pi})}{\alpha^2}\right)^{A_{ij}}\left(\frac{(1-q_{ij}(\Pi))(1-q_{ij}(\tilde{\Pi}))}{(1-\alpha)^2}\right)^{1-A_{ij}}\right].$$

We denote

$$\Sigma(A,\Pi,\tilde{\Pi}) = \prod_{i<j}\left(\frac{q_{ij}(\Pi)q_{ij}(\tilde{\Pi})}{\alpha^2}\right)^{A_{ij}}\left(\frac{(1-q_{ij}(\Pi))(1-q_{ij}(\tilde{\Pi}))}{(1-\alpha)^2}\right)^{1-A_{ij}},$$

and further obtain using the Tonelli theorem that

$$1 + D_{\chi^2}(P_0\|P_1) = \mathbb{E}_0[\mathbb{E}_{\Pi,\tilde{\Pi}}[\Sigma(A,\Pi,\tilde{\Pi})]] = \mathbb{E}_{\Pi,\tilde{\Pi}}[\mathbb{E}_0[\Sigma(A,\Pi,\tilde{\Pi})]].$$

Recalling that for all $i<j$ the A_{ij}'s are mutually independent, we can calculate $\mathbb{E}_0[\Sigma(A,\Pi,\tilde{\Pi})]$ easily. The calculations yield that

$$1 + D_{\chi^2}(P_0\|P_1) = \mathbb{E}_{\Pi,\tilde{\Pi}}\left[\prod_{i<j}\left(1 + \frac{\Delta_{ij}\tilde{\Delta}_{ij}}{\alpha(1-\alpha)}\right)\right],$$

where for $i<j$ we have $\Delta_{ij} = \pi_i' P \pi_j - \alpha$ and $\tilde{\Delta}_{ij} = \tilde{\pi}_i' P \tilde{\pi}_j - \alpha$. Since for all x in $\mathbb{R}$ it holds that $1 + x \le e^x$, we can bound the above by

$$1 + D_{\chi^2}(P_0\|P_1) \le \mathbb{E}_{\Pi,\tilde{\Pi}}\left[\exp\left(\sum_{i<j}\frac{\Delta_{ij}\tilde{\Delta}_{ij}}{\alpha(1-\alpha)}\right)\right]$$

$$= \mathbb{E}_{\Pi,\tilde{\Pi}}\left[\exp\left(\frac{S}{2(1-\alpha)}\right)\right], \qquad \text{where} \quad S \equiv \alpha^{-1}\sum_{i\neq j}\Delta_{ij}\tilde{\Delta}_{ij}. \tag{J.1}$$

Recall that we chose $\alpha = h'Ph$ for the null model. Let $y_i = \pi_i - h$ for $i = 1, ..., n$, hence $\mathbb{E}[y_i] = 0$. We obtain, for all $i \neq j$

$$\begin{aligned}\Delta_{ij} &= \pi_i'P\pi_j - \alpha = y_i'Py_j + h'Py_i + h'Py_j + h'Ph - \alpha\\ &= y_i'Py_j + h'Py_i + h'Py_j.\end{aligned}$$

Hence, $\mathbb{E}[\Delta_{ij}] = 0$. We define the matrix $M = P - \alpha\mathbf{1}_K\mathbf{1}_K'$. For all $i \in [\![1, n]\!]$, $\pi_i'\mathbf{1}_K = h'\mathbf{1}_K = 1$, which implies that $y_i'\mathbf{1}_K = 0$. It follows that

$$\Delta_{ij} = y_i'My_j + h'My_i + h'My_j. \tag{J.2}$$

We plug (J.2) into (J.1) to decomposition $\Delta_{ij}\tilde{\Delta}_{ij}$ into 9 terms:

$$\begin{aligned}\Delta_{ij}\tilde{\Delta}_{ij} = (y_i'My_j)(\tilde{y}_i'M\tilde{y}_j) &+ \Big[(h'My_i)(h'M\tilde{y}_i) + (h'My_j)(h'M\tilde{y}_j)\Big]\\ &+ \Big[(y_i'My_j)(h'M\tilde{y}_i') + (y_i'My_j)(h'M\tilde{y}_j') + (h'My_i)(\tilde{y}_i'M\tilde{y}_j) + (h'My_j)(\tilde{y}_i'M\tilde{y}_j)\Big]\\ &+ \Big[(h'My_i)(h'M\tilde{y}_j) + (h'My_j)(h'M\tilde{y}_i)\Big].\end{aligned}$$

Summing over (i, j) such that $i \neq j$ gives a total of 9 partial sums, which we denote by S_1, S_{21}, S_{22}, S_{31}, S_{32}, S_{33}, S_{34}, S_{41} and S_{42}, respectively. For example,

$$\begin{aligned}S_1 &= \alpha^{-1}\sum_{i\neq j}(y_i'My_j)(\tilde{y}_i'M\tilde{y}_j),\\ S_{21} &= \alpha^{-1}(n-1)\sum_{i}(h'My_i)(h'M\tilde{y}_i),\\ S_{31} &= \alpha^{-1}\sum_{i\neq j}(y_i'My_j)(h'M\tilde{y}_i),\\ S_{41} &= \alpha^{-1}\sum_{i\neq j}(h'My_i)(h'M\tilde{y}_j).\end{aligned} \tag{J.3}$$

It follows that

$$S = S_1 + \sum_{m=1}^{2} S_{2m} + \sum_{m=1}^{4} S_{3m} + \sum_{m=1}^{2} S_{4m}.$$

Combining (J.1) and (J.4) with Jensen's inequality, we have

$$\begin{aligned}1 + D_{\chi^2}(P_0\|P_1) &\leq \mathbb{E}\left[\exp\left(\frac{S_1 + \sum_{m=2} S_{2m} + \sum_{m=1}^{4} S_{3m} + \sum_{m=1}^{2} S_{4m}}{2(1-\alpha)}\right)\right]\\ &\leq \frac{1}{9}\exp\left(\frac{9|S_1|}{2(1-\alpha)}\right) + \frac{1}{9}\sum_{m=1}^{2}\exp\left(\frac{9|S_{2m}|}{2(1-\alpha)}\right) +\end{aligned}$$

$$+\frac{1}{9}\sum_{m=1}^{4}\exp\left(\frac{9|S_{3m}|}{2(1-\alpha)}\right)+\frac{1}{9}\sum_{m=1}^{2}\exp\left(\frac{9|S_{4m}|}{2(1-\alpha)}\right).$$

Write $c_\alpha = 9/[2(1-\alpha)]$. To show the claim, it suffices to show that

$$\mathbb{E}\big[\exp(c_\alpha|X|)\big] = 1+o(1), \qquad \text{for each } X \in \{S_1, S_{21}, S_{22}, S_{31}, \ldots, S_{34}, S_{41}, S_{42}\}. \tag{J.4}$$

Below, we show (J.4) for each of X listed above.

First, consider $X = S_1$. Let $\delta_1, \delta_2, \ldots, \delta_K$ be the K eigenvalues of M, arranged in the descending order of magnitude, and let $b_1, b_2, \ldots, b_K$ be the associated eigenvectors. Then, $M = \sum_{k=1}^K \delta_k b_k b_k'$. It follows that

$$S_1 = \alpha^{-1}\sum_{k,l}\delta_k\delta_l\left(\sum_i (y_i'b_k)(\tilde{y}_i'b_l)\right)^2 - \alpha^{-1}\sum_{k,l}\delta_k\delta_l\sum_i (y_i'b_k)^2(\tilde{y}_i'b_l)^2.$$

Note that $\max_k|\delta_k| = \|M\|$, where $\|M\|$ is the operator norm of M. Therefore,

$$|S_1| \le \alpha^{-1}\|M\|^2\left[\sum_{k,l}\left(\sum_i (y_i'b_k)(\tilde{y}_i'b_l)\right)^2 + \sum_{k,l}\sum_i (y_i'b_k)^2(\tilde{y}_i'b_l)^2\right].$$

In addition, for any $i \in [\![1,n]\!]$ and $k \in [\![1,K]\!]$, by the Cauchy-Schwarz inequality, we have $(y_i'b_k)^2 \le \|y_i\|_2^2\|b_k\|_2^2 = \|y_i\|_2^2 \le \|y_i\|_1 \le 2$, given that $\|y_i\|_\infty \le 1$ and that $\|y_i\|_1 \le \|\pi_i\|_1 + \|h\|_1 \le 2$. It follows that

$$|S_1| \le 4n\alpha^{-1}K^2\|M\|^2 + R_1, \tag{J.5}$$

where

$$R_1 \equiv \alpha^{-1}K^2\|M\|^2\max_{k,l}\left(\sum_i (y_i'b_k)(\tilde{y}_i'b_l)\right)^2.$$

To bound R_1, we fix a tuple (k,l) and provide an upper bound for $Y_{kl} := \sum_i (y_i'b_k)(\tilde{y}_i'b_l)$. Note that Y_{kl} is a sum of independent, zero-mean random variables. In addition, $|(y_i'b_k)(\tilde{y}_i'b_l)| \le \|y_i\|_2\|\tilde{y}_i\|_2 \le 2$. We can apply Hoeffding's inequality, for any $t>0$:

$$\mathbb{P}(|Y_{kl}|>t) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n(2\|y_i\|_2\|\tilde{y}_i\|_2)^2}\right) = 2\exp\left(-\frac{t^2}{8n}\right).$$

Hence, denoting $Y_* := \max_{k,l}|Y_{kl}|$, we have

$$\mathbb{P}(Y_*>t) = \mathbb{P}\left(\bigcup_{k,l}\{|Y_{kl}|>t\}\right) \le \sum_{k,l}\mathbb{P}(|Y_{kl}|>t) \le 2K^2\exp\left(-\frac{t^2}{8n}\right).$$

It follows that, for any $t>0$,

$$\mathbb{P}(R_1>t) = \mathbb{P}\left(Y^* > \frac{\sqrt{t\alpha}}{K\|M\|}\right) \le 2K^2\exp\left(-\frac{\alpha t}{8nK^2\|M\|^2}\right). \tag{J.6}$$

We now use (J.5) and (J.6) to bound $\mathbb{E}[\exp(c_\alpha|S_1|)]$. For any non-negative variable X, it follows from integration by part that $\mathbb{E}[\exp(X)] = 1 + \int_0^\infty e^t \mathbb{P}(X > t)dt$. It follows that

$$\begin{aligned}\mathbb{E}\big[\exp(c_\alpha|S_1|)\big] &\leq e^{4c_\alpha n\alpha^{-1}K^2\|M\|^2} \cdot \mathbb{E}\big[\exp(c_\alpha R_1)\big] \\ &\leq e^{4c_\alpha n\alpha^{-1}K^2\|M\|^2}\left[1 + \int_0^\infty e^t \mathbb{P}\left(R_1 > c_\alpha^{-1}t\right)dt\right] \\ &\leq e^{4c_\alpha n\alpha^{-1}K^2\|M\|^2}\left[1 + \int_0^\infty 2e^{-\left(\frac{\alpha}{8c_\alpha nK^2\|M\|^2}-1\right)t}dt\right].\end{aligned}$$

In our assumption, $\beta_n \to 0$, which implies that

$$n\alpha^{-1}\|M\|^2 \to 0.$$

It follows that $e^{4c_\alpha n\alpha^{-1}K^2\|M\|^2} = \exp(o(1)) = 1 + o(1)$. Also, for n big enough, $\frac{\alpha}{8c_\alpha nK^2\|M\|^2} - 1 > 0$. Furthermore, we note that for any value $z > 0$, $\int_0^\infty e^{-zt}dt = z^{-1}$. Combining the above gives

$$\mathbb{E}\big[\exp(c_\alpha|S_1|)\big] \leq e^{4c_\alpha K^2 n\alpha^{-1}\|M\|^2}\left(1 + \frac{16c_\alpha K^2 n\alpha^{-1}\|M\|^2}{1 - 8c_\alpha K^2 n\alpha^{-1}\|M\|^2}\right) = 1 + o(1). \tag{J.7}$$

This proves (J.4) for $X = S_1$.

Second, consider $X = S_{21}$ (the analysis of S_{22} is similar and thus omitted). We define a unit-norm vector $u = \|Mh\|^{-1}(Mh)$. Then,

$$S_{21} = \alpha^{-1}(n-1)\|Mh\|^2\sum_i (y_i'u)(\tilde{y}_i'u).$$

The variables $\{(y_i'u)(\tilde{y}_i'u)\}_{1\leq i\leq n}$ are independent, with $|(y_i'u)(\tilde{y}_i'u)| \leq \|y_i\|\|\tilde{y}_i\|$. We have seen that $\|y_i\|^2 \leq 2$ and $\|\tilde{y}_i\|^2 \leq 2$. It follows that $|(y_i'u)(\tilde{y}_i'u)| \leq 2$. Applying Hoeffding's inequality, we obtain that, for any $t > 0$,

$$\begin{aligned}\mathbb{P}(|S_{21}| > t) &= \mathbb{P}\left(\left|\sum_i (y_iu)(\tilde{y}_iu)\right| > \frac{t\alpha}{(n-1)\|Mh\|^2}\right) \\ &\leq 2\exp\left(-\frac{t^2\alpha^2}{8n(n-1)^2\|Mh\|^4}\right) \leq 2\exp\left(-\frac{t^2\alpha^2}{8n^3\|Mh\|^4}\right).\end{aligned}$$

Our assumption $\beta_n \to 0$ implies that

$$n^3\alpha^{-2}\|Mh\|^4 \to 0.$$

Furthermore, for $z > 0$, we have $\int_0^\infty e^{-zt^2+t} \leq \sqrt{2\pi z^{-1}}e^{(4z)^{-1}}$. Combining these gives

$$\begin{aligned}\mathbb{E}\big[\exp(c_\alpha|S_{21}|)\big] &= 1 + \int_0^\infty e^t \mathbb{P}\left(|S_{21}| > c_\alpha^{-1}t\right)dt \\ &\leq 1 + \int_0^\infty 2e^{-\frac{\alpha^2}{8c_\alpha^2 n^3\|Mh\|^4}t^2+t}dt\end{aligned}$$

$$\leq 1 + 2\sqrt{2\pi}\sqrt{8c_\alpha^2 n^3 \alpha^{-2}\|Mh\|^4}\exp\left(-2c_\alpha^2 n^3\alpha^{-2}\|Mh\|^4\right)$$
$$= 1 + o(1). \tag{J.8}$$

This proves (J.4) for $X = S_{21}$.

Next, consider S_{31} (the analyses of S_{32}-S_{34} are similar and omitted). Recall that $M = \sum_{k=1}^K \delta_k b_k b_k'$ is the eigen-decomposition of M; additionally, we have defined $u = \|Mh\|^{-1}(Mh)$. It follows that

$$\begin{aligned}
S_{31} &= \alpha^{-1}\|Mh\|\sum_{i\neq j}(y_i' M y_j)(\tilde{y}_i' u) \\
&= \alpha^{-1}\|Mh\|\sum_{i\neq j}\left[\sum_k \delta_k (y_i' b_k)(y_j' b_k)\right](\tilde{y}_i' u) \\
&= \alpha^{-1}\|Mh\|\sum_k \delta_k\left[\sum_i (y_i' b_k)(\tilde{y}_i' u)\right]\left[\sum_j (y_j' b_k)\right] \\
&\quad - \alpha^{-1}\|Mh\|\sum_k \delta_k\left[\sum_i (y_i' b_k)^2(\tilde{y}_i' u)\right].
\end{aligned}$$

We have seen that $\|b_i\|^2 = 1$, $\|y_i\|^2 \leq 2$, $\|\tilde{y}_i\| \leq 2$, $\|u\| = 1$, and $|\delta_k| \leq \|M\|$. It follows that

$$|S_{31}| \leq R_{31} + 2\sqrt{2}\, n\alpha^{-1}K\|M\|\|Mh\|, \tag{J.9}$$

where

$$R_{31} := \alpha^{-1}\|M\|\|Mh\|K\max_k Z_k, \qquad \text{with} \quad Z_k = \left[\sum_i (y_i' b_k)(\tilde{y}_i' u)\right]\left[\sum_j (y_j' b_k)\right].$$

We can derive the tail probability bound for Z_k: Since $|y_i' b_k| \leq \|y_i\| \leq \sqrt{2}$ and $|\tilde{y}_i' u| \leq \|\tilde{y}_i\| \leq \sqrt{2}$, the Hoeffding's inequality yields that

$$\begin{aligned}
\mathbb{P}(|Z_k| > t) &\leq \mathbb{P}\left(\left|\sum_i (y_i' b_k)(\tilde{y}_i' u)\right| > \sqrt{t}\right) + \mathbb{P}\left(\left|\sum_j (y_j' b_k)\right| > \sqrt{t}\right) \\
&\leq 2\exp\left(-\frac{t}{8n}\right) + 2\exp\left(-\frac{t}{4n}\right) \leq 4\exp\left(-\frac{t}{8n}\right).
\end{aligned}$$

We thus have

$$\mathbb{P}(|R_{31}| > t) = \mathbb{P}\left(\max_k Z_k > \frac{t\alpha}{K\|M\|\|Mh\|}\right) \leq 4K\exp\left(-\frac{t\alpha}{8nK\|M\|\|Mh\|}\right). \tag{J.10}$$

We apply (J.9)-(J.10) to bound $\mathbb{E}[\exp(c_\alpha|S_{31}|)]$. Our assumption $\beta_n \to 0$ ensures that $n\alpha^{-1}\|M\|^2 \to 0$. Note that $\|Mh\| \leq \|M\|\|h\| \leq \|M\|\sqrt{\|h\|_1\|h\|_\infty} \leq \|M\|$. It follows that

$$n\alpha^{-1}\|M\|\|Mh\| \to 0.$$

We then mimic the proof of (J.7) to get

$$\begin{aligned}\mathbb{E}\big[\exp(c_\alpha|S_{31}|)\big] &\le e^{2\sqrt{2}c_\alpha n\alpha^{-1}K\|M\|\|Mh\|}\left[1+\int_0^\infty e'\mathbb{P}\left(|R_{31}|>c_\alpha^{-1}t\right)dt\right]\\ &\le e^{2\sqrt{2}c_\alpha n\alpha^{-1}K\|M\|\|Mh\|}\left[1+\int_0^\infty 4Ke^{-\left(\frac{\alpha}{4c_\alpha nK\|M\|\|Mh\|}-1\right)t}\right]\\ &\le e^{2\sqrt{2}c_\alpha n\alpha^{-1}K\|M\|\|Mh\|}\left(1+\frac{16c_\alpha K^2 n\alpha^{-1}\|M\|\|Mh\|}{1-4c_\alpha Kn\alpha^{-1}\|M\|\|Mh\|}\right)\\ &=1+o(1). \end{aligned}\tag{J.11}$$

This proves (J.4) for $X=S_{31}$.

Last, consider S_{41} (the analysis of S_{42} is similar and omitted). Since $u=\|Mh\|^{-1}Mh$, we have

$$\begin{aligned}S_{41}&=\alpha^{-1}\|Mh\|^2\sum_{i\neq j}(y_i'u)(\tilde{y}_j'u)\\ &=\alpha^{-1}\|Mh\|^2\left[\sum_i(y_i'u)\right]\left[\sum_j(\tilde{y}_j'u)\right]-\alpha^{-1}\|Mh\|^2\sum_i(y_i'u)(\tilde{y}_i{}'u).\end{aligned}$$

Note that $|(y_i'u)(\tilde{y}_i{}'u)|\le\|y_i\|\|\tilde{y}_i\|\le 2$. We immediately have

$$|S_{41}|\le R_{41}+2n\alpha^{-1}\|Mh\|^2,\tag{J.12}$$

where

$$R_{41}=\alpha^{-1}\|Mh\|^2\left[\sum_i(y_i'u)\right]\left[\sum_j(\tilde{y}_j'u)\right].$$

We apply Hoeffding's inequality to derive the tail probability bound: For all $t>0$,

$$\begin{aligned}\mathbb{P}(|R_{41}|>t)&=\mathbb{P}\left(\left|\sum_i(y_i'u)\right|>\frac{\sqrt{\alpha t}}{\|Mh\|}\right)+\mathbb{P}\left(\left|\sum_j(\tilde{y}_j'u)\right|>\frac{\sqrt{\alpha t}}{\|Mh\|}\right)\\ &\le 4\exp\left(-\frac{\alpha t}{8n\|Mh\|^2}\right).\end{aligned}\tag{J.13}$$

We have seen that $\|Mh\|\le\|M\|$. Therefore, the assumption of $\beta_n\to 0$ leads to

$$n\alpha^{-1}\|Mh\|^2\to 0.$$

Using (J.12) and (J.13), we have

$$\begin{aligned}\mathbb{E}\big[\exp(c_\alpha|S_{41}|)\big]&\le e^{2c_\alpha n\alpha^{-1}\|Mh\|^2}\left[1+\int_0^\infty e'\mathbb{P}\left(|R_{41}|>c_\alpha^{-1}t\right)dt\right]\\ &\le e^{2c_\alpha n\alpha^{-1}\|Mh\|^2}\left[1+\int_0^\infty 4e^{-\left(\frac{\alpha}{8c_\alpha n\|Mh\|^2}-1\right)t}\right].\end{aligned}$$

$$\leq e^{2c_\alpha n\alpha^{-1}\|Mh\|^2}\left(1+\frac{32c_\alpha n\alpha^{-1}\|Mh\|^2}{1-8c_\alpha n\alpha^{-1}\|Mh\|^2}\right)$$
$$=1+o(1).$$

This proves (J.4) for $X=S_{41}$. □

Appendix K: Proof of Theorem 3.6

Note: this proof requires Lemma K.1 and Lemma K.2, which are provided directly after the proof.

We start by studying the case $t_0=0$. We consider a sequence of null hypotheses indexed by n, where $\Omega_n=\alpha_n\mathbf{1}_K\mathbf{1}_K'\in\mathcal{M}_{0n}$ under $H_0^{(n)}$. For our sequence of alternatives, we consider $\Omega_n=\Pi_nP_n\Pi_n'$ under $H_1^{(n)}$, with

$$P_n=\alpha_n\left[\gamma_nI_K+(1-\gamma_n)\mathbf{1}_K\mathbf{1}_K'\right],\quad\text{and}\quad\pi_1,...,\pi_n\overset{\text{iid}}{\sim}F,$$

where for all $k\in\{1,...,K\}$,

$$\mathbb{P}_{\pi\sim F}(\pi=e_k)=\frac{1}{K}.$$

In the above definition, $\{e_k\}_{k=1}^K$ denotes the canonical basis of $\mathbb{R}^K$. It follows that

$$h:=\mathbb{E}_{\pi\sim F}[\pi]=\frac{1}{K}\mathbf{1}_K,\quad\text{and}\quad\Sigma:=\mathbb{E}_{\pi\sim F}[\pi\pi']=\frac{1}{K}I_K.$$

Under this random mixed membership model, it is straightforward to verify that

$$\alpha_0=\alpha_n\left(1-\frac{K-1}{K}\gamma_n\right),$$
$$\|P_nh-\alpha_0\mathbf{1}_K\|=0,$$
$$\|P_n-\alpha_0\mathbf{1}_K\mathbf{1}_K'\|=\alpha_n\gamma_n.$$

Hence

$$\beta_n=\max\{n^{3/2}\alpha_0^{-1}\|P_nh-\alpha_0\mathbf{1}_K\|^2,\quad n^2\alpha_0^{-2}\|P_n-\alpha_0\mathbf{1}_K\mathbf{1}_K'\|^4\}=n^2\alpha_0^{-2}\alpha_n^4\gamma_n^4.$$

By assumption, $\gamma_n\to0$, hence for n sufficiently large, $\alpha_n<2\alpha_0$, hence

$$\beta_n=O(n^2\alpha_n^2\gamma_n^4)=o(1),$$

under the assumption that $n^2\alpha_n^2\gamma_n^4=o(1)$. By Theorem 3.5, the χ^2-distance between the two distributions satisfies $D_{\chi^2}(f_0^{(n)}\|f_1^{(n)})=o(1)$. By connection between L_1-distance and χ^2-distance, it follows that

$$\|f_0^{(n)}-f_1^{(n)}\|_1=o(1).$$

We now slightly modify the alternative hypothesis. Let $\{\Pi_n^0\}_n$ be a sequence of non-random membership matrices such that $(P_n,\Pi_n^0)\in\mathcal{M}_{1n}(0)$. Such a sequence can be built e.g. by considering $\lfloor n/K\rfloor$

pure nodes in each community and all other nodes equally mixed across all communities. In the modified alternative hypothesis $\tilde{H}_1^{(n)}$,

$$\tilde{\Pi} = \begin{cases} \Pi_n, & \text{if } (\Pi_n, P_n) \in \mathcal{M}_{1n}(0), \\ \Pi_n^0, & \text{otherwise.} \end{cases}$$

Let $\tilde{f}_1^{(n)}$ be the probability measure associated with $\tilde{H}_1^{(n)}$. Under $\tilde{H}_1^{(n)}$, all realizations $\tilde{\Pi}_n P_n \tilde{\Pi}_n'$ are in the class $\mathcal{M}_{1n}(0)$, by definition. By the Neyman-Pearson lemma and elementary inequalities,

$$\begin{aligned} Risk_n^*(0) &\geq 1 - \inf_{f_0 \in \mathcal{M}_{0n}, f_1 \in \mathcal{M}_{1n}(0)} \{\|f_0 - f_1\|_1\} \\ &\geq 1 - \|f_0^{(n)} - \tilde{f}_1^{(n)}\|_1 \\ &\geq 1 - \|f_0^{(n)} - f_1^{(n)}\|_1 - \|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1 \\ &\geq 1 - o(1) - \|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1. \end{aligned}$$

It follows from Lemma K.1 that $\tilde{\Pi}_n = \Pi_n$ with probability $1 - o(1)$. As a result,

$$\|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1 = o(1),$$

from which we obtain that $\lim_{n\to\infty}\{Risk_n^*(0)\} = 1$.

Next, we study the case $0 < t_0$. Again, we consider a sequence of null hypotheses indexed by n, where $\Omega_n = \alpha_n \mathbf{1}_K \mathbf{1}_K' \in \mathcal{M}_{0n}$ under $H_0^{(n)}$. For our sequence of alternatives, we consider $\Omega_n = \Pi_n P_n \Pi_n'$ under $H_1^{(n)}$, with

$$P_n = \alpha_n \left[\gamma_n I_K + (1-\gamma_n)\mathbf{1}_K \mathbf{1}_K'\right], \quad \text{and} \quad \pi_1, ..., \pi_n \overset{\text{iid}}{\sim} F,$$

where

$$\mathbb{P}_{\pi\sim F}(\pi = e_1) = \frac{K+1}{2K}, \quad \text{and} \quad \mathbb{P}_{\pi\sim F}(\pi = e_1) = \frac{1}{2K} \quad \forall k \in \{2, ..., K\}.$$

It follows that

$$h := \mathbb{E}_{\pi\sim F}[\pi] = \frac{1}{2K}(Ke_1 + \mathbf{1}_K), \quad \text{and} \quad \Sigma := \mathbb{E}_{\pi\sim F}[\pi\pi'] = \frac{1}{2K}(Ke_1e_1' + I_K).$$

Under this random mixed membership model, it is straightforward to verify that

$$\begin{aligned} \alpha_0 &= \alpha_n\left(1 - \frac{3K-3}{4K}\gamma_n\right), \\ \|P_n h - \alpha_0 \mathbf{1}_K\| &= \alpha_n\gamma_n\sqrt{\frac{(K-1)(K+3)}{16K}}, \\ \|P_n - \alpha_0 \mathbf{1}_K \mathbf{1}_K'\| &= \max\left\{\alpha_n\gamma_n, \frac{K-1}{4}\alpha_n\gamma_n\right\}. \end{aligned}$$

Recall that

$$\beta_n = \max\{n^{3/2}\alpha_0^{-1}\|P_n h - \alpha_0 \mathbf{1}_K\|^2, \quad n^2\alpha_0^{-2}\|P_n - \alpha_0 \mathbf{1}_K \mathbf{1}_K'\|^4\}.$$

Hence

$$\beta_n = \max\left\{n^{3/2}\alpha_0^{-1}\alpha_n^2\gamma_n^2 \frac{(K-1)(K+3)}{16K}, \quad \max\left(1, \left(\frac{K-1}{4}\right)^4\right) n^2\alpha_0^{-2}\alpha_n^4\gamma_n^4\right\}$$

By assumption, $\gamma_n \to 0$, hence for n sufficiently large, $\alpha_n < 2\alpha_0$, hence

$$\beta_n = O\left(\max\left\{n^{3/2}\alpha_n\gamma_n^2, \quad n^2\alpha_n^2\gamma_n^4\right\}\right) = O\left(\max\left\{n^{3/2}\alpha_n\gamma_n^2, \quad \frac{(n^{3/2}\alpha_n\gamma_n^2)^2}{n}\right\}\right) = o(1),$$

under the assumption that $n^{3/2}\alpha_n\gamma_n^2 = o(1)$. By Theorem 3.5, the χ^2-distance between the two distributions satisfies $D_{\chi^2}(f_0^{(n)}\|f_1^{(n)}) = o(1)$. By connection between L_1-distance and χ^2-distance, it follows that

$$\|f_0^{(n)} - f_1^{(n)}\|_1 = o(1).$$

We now slightly modify the alternative hypothesis. Let $\{\Pi_n^0\}_n$ be a sequence of non-random membership matrices such that $(P_n, \Pi_n^0) \in \mathcal{M}_{1n}(t_0)$. Such a sequence can be built e.g. by considering $\lfloor n(K+1)/2K \rfloor$ pure nodes in community 1, $\lfloor n/2K \rfloor$ nodes in communities 2 to K and all other nodes with mixed membership vector $(2K)^{-1}(Ke_1 + \mathbf{1}_K)$. In the modified alternative hypothesis $\tilde{H}_1^{(n)}$,

$$\tilde{\Pi} = \begin{cases} \Pi_n, & \text{if } (\Pi_n, P_n) \in \mathcal{M}_{1n}(t_0), \\ \Pi_n^0, & \text{otherwise.} \end{cases}$$

Let $\tilde{f}_1^{(n)}$ be the probability measure associated with $\tilde{H}_1^{(n)}$. Under $\tilde{H}_1^{(n)}$, all realizations $\tilde{\Pi}_n P_n \tilde{\Pi}_n'$ are in the class $\mathcal{M}_{1n}(t_0)$, by definition. By the Neyman-Pearson lemma and elementary inequalities,

$$\begin{aligned} Risk_n^*(0) &\geq 1 - \inf_{f_0 \in \mathcal{M}_{0n}, f_1 \in \mathcal{M}_{1n}(t_0)} \{\|f_0 - f_1\|_1\} \\ &\geq 1 - \|f_0^{(n)} - \tilde{f}_1^{(n)}\|_1 \\ &\geq 1 - \|f_0^{(n)} - f_1^{(n)}\|_1 - \|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1 \\ &\geq 1 - o(1) - \|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1. \end{aligned}$$

It follows from Lemma K.2 that $\tilde{\Pi}_n = \Pi_n$ with probability $1 - o(1)$. As a result,

$$\|f_1^{(n)} - \tilde{f}_1^{(n)}\|_1 = o(1),$$

from which we obtain that $\lim_{n\to\infty}\{Risk_n^*(t_0)\} = 1$. □

Lemma K.1 (Case $t_0 = 0$). *Fix $K \geq 2$, a sequence $\{\alpha_n\}_n \in [0,1]^{\mathbb{N}}$, and a sequence $\{\gamma_n\}_n \in (\mathbb{R}_+)^{\mathbb{N}}$. Denote by $\{e_k\}_{k=1}^K$ the canonical basis of $\mathbb{R}^K$. Consider the sequence of alternative probability matrices $\Omega_n = \Pi_n P_n \Pi_n'$, with*

$$P_n = \alpha_n \left[\gamma_n I_K + (1-\gamma_n)\mathbf{1}_K \mathbf{1}_K'\right], \quad \text{and} \quad \pi_1, ..., \pi_n \overset{iid}{\sim} F,$$

where for all $k \in \{1, ..., K\}$, $\mathbb{P}_{\pi \sim F}(\pi = e_k) = \frac{1}{K}$. Suppose that $\alpha_n \to 0$, $n\alpha_n \to \infty$, and $\gamma_n \to 0$. Then, with probability $1 - o(1)$, $(P_n, \Pi_n) \in \mathcal{M}_{1n}(0)$.

Proof
From the proof of Theorem 3.6 for $t_0 = 0$, we know that

$$h := \mathbb{E}_{\pi \sim F}[\pi] = \frac{1}{K}\mathbf{1}_K, \quad \Sigma := \mathbb{E}_{\pi \sim F}[\pi\pi'] = \frac{1}{K} I_K' \quad \text{and} \quad \alpha_0 = \alpha_n \left(1 - \frac{K-1}{K}\gamma_n\right).$$

We introduce the following random quantities:

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^n \pi_i, \quad \tilde{G} = \frac{1}{n}\sum_{i=1}^n \pi_i \pi_i', \quad \text{and} \quad \tilde{\alpha}_0 = \tilde{h} P_n \tilde{h}'.$$

To show that $(P, \Pi) \in \mathcal{M}_{1n}(0)$, we will check that

1. $OSC(\tilde{h}) \leq C$ and $\|\tilde{G}^{-1}\| \leq C$,
2. $\tilde{\alpha}_0 \leq c$, $n\tilde{\alpha}_0 \geq c^{-1}$, and $\tilde{\alpha}_0 \geq \alpha_n/2$,
3. $2\tilde{\alpha}_0^{-1}\|P_n - \tilde{\alpha}_0 \mathbf{1}_K \mathbf{1}_K'\| \geq \gamma_n$.

First, recognize that $\tilde{h} = n^{-1}\sum_{i=1}^n \tilde{\pi}_i \xrightarrow{\text{as}} K^{-1}\mathbf{1}_K$ by the Strong Law of Large Numbers. As a consequence, for n sufficiently large, we have $OSC(\tilde{h}) < C$ with probability at least $1 - o(1)$. Next, let $y_i = \pi_i - \tilde{h}$. We have

$$\begin{aligned} n\tilde{G} &= \sum_{i=1}^n \pi_i \pi_i' = \sum_{i=1}^n (\tilde{h}\tilde{h}' + \tilde{h} y_i' + y_i \tilde{h}' + y_i y_i') \\ &= n\Sigma + \sum_{i=1}^n (y_i y_i' - \mathbb{E}[y_i y_i']) + \sum_{i=1}^n (\tilde{h} y_i') + \sum_{i=1}^n (y_i \tilde{h}') \\ &= n\Sigma + Z_0 + Z_1 + Z_2. \end{aligned}$$

Notice that Z_0 is a sum of n independent mean-zero random matrices, so we can apply the matrix Hoeffding inequality to bound its operator norm. Since $\|y_i y_i' - \mathbb{E}[y_i y_i']\| \leq C$, we obtain for $t > 0$,

$$\mathbb{P}(\|Z_0\| > t) \leq \exp\left(-\frac{Ct^2}{n}\right).$$

If we pick $t = C\sqrt{n\log(n)}$, then we have that $\|Z_0\| < C\sqrt{n\log(n)}$ with probability $1 - o(1)$. Similarly, it is straightforward to show that $\|Z_1 + Z_2\| \leq C\sqrt{n\log(n)}$ with probability $1 - o(1)$. Now, recall that $\lambda_{\min}(\Sigma) = K^{-1}$. As a result,

$$\lambda_{\min}(n\tilde{G}) = \lambda_{\min}(n\Sigma + Z_0 + Z_1 + Z_2) > \lambda_{\min}(n\Sigma) - \|Z_0 + Z_1 + Z_2\| > \frac{n}{K} - C\sqrt{n\log(n)}.$$

It follows that

$$\lambda_{\min}(\tilde{G}) > \frac{1}{K} - C\sqrt{\frac{\log(n)}{n}},$$

which shows that for n sufficiently large, $\|\tilde{G}^{-1}\| < C$ with probability $1 - o(1)$.

Next, we show that $\tilde{\alpha}_0 < c$ and $n\tilde{\alpha}_0 > c^{-1}$ with high probability. Denote $z := \tilde{h} - h$. We can rewrite

$$\tilde{\alpha}_0 = \tilde{h}'P\tilde{h} = z'Pz + 2h'Pz + \alpha_0.$$

Notice that both $\|z'Pz\| \leq C\|Pz\|$ and $\|h'Pz\| \leq C\|Pz\|$. We now provide a high-probability bound on the 2-norm of Pz, which can be written as a sum of mean-zero independent random variables

$$Pz = \frac{1}{n}\sum_{i=1}^{n}(P\pi_i - Ph),$$

where for all $i = 1, ..., n$ it holds that $\|P\pi_i - Ph\| \leq C\|P\| \leq C\alpha_n\gamma_n$. For $t > 0$, Hoeffding's inequality yields

$$\mathbb{P}\left(\|Pz\| > t\right) \leq C\exp\left(-\frac{Cnt^2}{\alpha_n^2\gamma_n^2}\right).$$

Pick $t = \alpha_n\gamma_n\sqrt{\log(n)/n}$. As a consequence, we obtain that $\|Pz\| < \alpha_n\gamma_n\sqrt{\log(n)/n}$ with probability $1 - o(1)$. Hence with probability $1 - o(1)$,

$$\tilde{\alpha}_0 = \alpha_0 + O\left(\alpha_n\gamma_n\sqrt{\frac{\log(n)}{n}}\right) = \alpha_n - \frac{K-1}{K}\alpha_n\gamma_n + o\left(\alpha_n\gamma_n\right) = \alpha_n + O(\alpha_n\gamma_n).$$

It follows that for n sufficiently large, $\tilde{\alpha}_0 < c$ and $n\tilde{\alpha}_0 > c^{-1}$ with probability $1 - o(1)$. We also obtain from this last equation that for n sufficiently large, $\tilde{\alpha}_0 \geq \alpha_n/2$ with probability $1 - o(1)$.

It remains to show that $2\tilde{\alpha}_0^{-1}\|P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\| \geq \gamma_n$. With probability $1 - o(1)$, the matrix $(P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K')$ has eigenvalues

$$\lambda_+ = K(\alpha_n - \tilde{\alpha}_0) - (K-1)\alpha_n\gamma_n = o(\alpha_n\gamma_n),$$
$$\lambda_- = \alpha_n\gamma_n.$$

hence for n sufficiently large, we must have $\|P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\| = \alpha_n\gamma_n \geq \tilde{\alpha}_0\gamma_n/2$. It follows that

$$2\tilde{\alpha}_0^{-1}\|P - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\| \geq \gamma_n,$$

which concludes the proof. □

Lemma K.2 (Case $0 < t_0$). *Fix $K \geq 2$, a sequence $\{\alpha_n\}_n \in [0,1]^{\mathbb{N}}$, and a sequence $\{\gamma_n\}_n \in (\mathbb{R}_+)^{\mathbb{N}}$. Denote by $\{e_k\}_{k=1}^K$ the canonical basis of $\mathbb{R}^K$. Consider the sequence of alternative probability matrices $\Omega_n = \Pi_n P_n \Pi_n'$, with*

$$P_n = \alpha_n\left[\gamma_n I_K + (1-\gamma_n)\mathbf{1}_K\mathbf{1}_K'\right], \quad \textit{and} \quad \pi_1, ..., \pi_n \overset{iid}{\sim} F,$$

where

$$\mathbb{P}_{\pi\sim F}(\pi = e_1) = \frac{K+1}{2K}, \quad \textit{and} \quad \mathbb{P}_{\pi\sim F}(\pi = e_k) = \frac{1}{2K} \quad \forall k \in \{2,...,K\}.$$

Suppose that $\alpha_n \to 0$*,* $n\alpha_n \to \infty$*,* $\gamma_n \to 0$*, and* $0 < t_0 < \sqrt{(K-1)(K+3)/(16K)}$*. Then, with probability* $1-o(1)$*,* $(P_n, \Pi_n) \in \mathcal{M}_{1n}(t_0)$*.*

Proof
From the proof of Theorem 3.6 for $t_0 > 0$, we know that

$$h := \mathbb{E}_{\pi\sim F}[\pi] = \frac{1}{2K}(Ke_1 + \mathbf{1}_K),$$
$$\Sigma := \mathbb{E}_{\pi\sim F}[\pi\pi'] = \frac{1}{2K}(Ke_1e_1' + I_K),$$
$$\text{and} \quad \alpha_0 = \alpha_n\left(1 - \frac{3K-3}{4K}\gamma_n\right).$$

We introduce the following random quantities:

$$\tilde{h} = \frac{1}{n}\sum_{i=1}^n \pi_i, \quad \tilde{G} = \frac{1}{n}\sum_{i=1}^n \pi_i\pi_i', \text{and} \quad \tilde{\alpha}_0 = \tilde{h}P_n\tilde{h}'.$$

To show that $(P,\Pi) \in \mathcal{M}_{1n}(t_0)$, we will check that

1. $OSC(\tilde{h}) \le C$ and $\|\tilde{G}^{-1}\| \le C$,
2. $\tilde{\alpha}_0 \le c$, $n\tilde{\alpha}_0 \ge c^{-1}$, and $\tilde{\alpha}_0 \ge \alpha_n/2$,
3. $2\tilde{\alpha}_0^{-1}\|P - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\| \ge \gamma_n$ and $\|P\tilde{h} - \tilde{\alpha}_0\mathbf{1}_K\| \ge t_0\|P - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\|$.

The first two points can be shown with probability at least $1-o(1)$ in the same way as in the proof of Lemma K.1. We will focus on the third point. For n sufficiently large, $\alpha_n\gamma_n$ is the largest eigenvalue of $(P - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K')$ in magnitude. Hence, we must have, for n sufficiently big

$$\|P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\| = \alpha_n\gamma_n \ge \tilde{\alpha}_0\gamma_n/2.$$

Now, introduce the (continuous) function with support $\mathbb{R}^K$:

$$g(x) := \left\|\begin{bmatrix} x_1(1-x_1) - \sum_{k\neq 1} x_k^2 \\ x_2(1-x_2) - \sum_{k\neq 2} x_k^2 \\ \vdots \\ x_K(1-x_K) - \sum_{k\neq K} x_k^2 \end{bmatrix}\right\|.$$

Notice that $\|P\tilde{h} - \tilde{\alpha}_0\mathbf{1}_K\| = \alpha_n\gamma_n g(\tilde{h})$ and $g(h) = \sqrt{(K-1)(K+3)/(16K)}$. As a consequence, for n sufficiently large,

$$\frac{\|P_n\tilde{h} - \tilde{\alpha}_0\mathbf{1}_K\|}{\|P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\|} = \frac{\alpha_n\gamma_n g(\tilde{h})}{\alpha_n\gamma_n} \xrightarrow{\text{as}} g(h) = \sqrt{\frac{(K-1)(K+3)}{16K}} > t_0.$$

It follows that for n sufficiently large, with probability at least $1-o(1)$,

$$\|P_n\tilde{h} - \tilde{\alpha}_0\mathbf{1}_K\| \ge t_0\|P_n - \tilde{\alpha}_0\mathbf{1}_K\mathbf{1}_K'\|,$$

which concludes the proof. □

Appendix L: Proof of Propositions 4.1-4.2

L.1. Proof of Proposition 4.1

We suppose that there exists an eligible tuple (Π_0, P_0, K_0) such that $\Omega = \Pi_0 P_0 \Pi_0'$. To show the first point of the proposition, define the set:

$$S = \left\{ k \in \mathbb{N}^* \quad \middle| \quad \exists (\Pi, P) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{k\times k} \text{ eligible such that } \Omega = \Pi P \Pi' \right\}.$$

Note that S is a discrete set lower bounded by 0 which is non-empty since $K_0 \in S$ by assumption. It follows that S has a lower bound, which we denote as k_Ω. It corresponds to the INC defined in Definition 4.2.

Now, we proceed to showing that when $K = k_\Omega$, the matrix P is identifiable up to permutation. Suppose that we have two pairs of eligible matrices $(\Pi, P), (\Pi^*, P^*) \in \mathbb{R}^{n\times k_\Omega} \times \mathbb{R}^{k_\Omega \times k_\Omega}$ such that $\Omega = \Pi P \Pi' = \Pi^* P^* (\Pi^*)'$. Because Π, Π^* are eligible, they contain the identity matrix as a submatrix. We assume without loss of generality that the first k_Ω rows of Π and Π^* correspond to k_Ω pure points, one per community. The submatrices

$$\tilde{\Pi} := \Pi_{|\{1,\ldots,k_\Omega\},\cdot} \quad \text{ and } \tilde{\Pi}^* := \Pi^*_{|\{1,\ldots,k_\Omega\},\cdot}$$

are permutations matrix. We have

$$\Omega_{|\{1,\ldots,k_\Omega\}\times\{1,\ldots,k_\Omega\}} = \tilde{\Pi} P \tilde{\Pi}' = \tilde{\Pi}^* P^* (\tilde{\Pi}^*)',$$

which implies that $P^* = DPD'$, where $D = (\tilde{\Pi}^*)'\tilde{\Pi}$ is a permutation matrix.

If, in addition, we have that $\text{rank}(P) = k_\Omega$, then P is invertible. It follows that

$$\tilde{\Pi} P = \tilde{\Pi}^* P^* (\tilde{\Pi}^*)' \tilde{\Pi} = \tilde{\Pi}^* P^* D = \tilde{\Pi}^* D P \quad \implies \quad \tilde{\Pi} = \tilde{\Pi}^* D.$$

In addition, since $\Omega = \Pi P \Pi' = \Pi^* P^* (\Pi^*)'$, Π and Π^* have full column rank, which means that there must exist an invertible matrix $B \in \mathbb{R}^{K\times K}$ such that $\Pi = \Pi^* B$. This implies that $\tilde{\Pi} = \tilde{\Pi}^* B$. As a result that $B = D$, so $\Pi = \Pi^* D$. This shows that if $\text{rank}(P) = k_\Omega$, then Π is also identifiable up to permutation.

Finally, it holds by definition of k_Ω that $K_0 \geq k_\Omega$. Since $\text{rank}(P_0) = \text{rank}(\Omega)$ and $k_\Omega = \dim(P) \geq \text{rank}(\Omega)$, we obtain that

$$K_0 \geq k_\Omega \geq \text{rank}(P_0).$$

Furthermore, if P_0 is non-singular, then $K_0 = \text{rank}(P_0)$, hence

$$K_0 = k_\Omega = \text{rank}(P_0).$$

L.2. Proof of Proposition 4.2

By Proposition 4.1, there exists a pair of eligible $\Pi \in \mathbb{R}^{n\times k_\Omega}$ and $P \in \mathbb{R}^{k_\Omega \times k\Omega}$ such that $\Omega = \Pi P \Pi'$, where k_Ω is the INC. Hence in the rest of the proof, we take $K = k_\Omega$.

Denote by $\Lambda \in \mathbb{R}^{r\times r}$ the matrix of eigenvalues of Ω. It follows that we can write $\Omega = \Xi\Lambda\Xi'$. Furthermore, note that the fact that $r = \text{rank}(\Omega)$ implies that we also have $\text{rank}(P) = r$. We can thus denote by $X \in \mathbb{R}^{k_\Omega\times r}$ the matrix of eigenvectors of P, and by $L \in \mathbb{R}^{r\times r}$ the corresponding matrix of non-zero eigenvalues, thus obtaining that $P = XLX'$. As a consequence,

$$\Omega = \Xi\Lambda\Xi' = \Pi X L(\Pi X)'.$$

Note that $\Lambda\Xi'$ and $L(\Pi X)'$ must have full row-rank r, so the column space of Ξ is equal to the column space of ΠX. There must exist a matrix $B \in \mathbb{R}^{r\times r}$ such that $\Xi = \Pi X B$. Hence there exists a matrix $V \in \mathbb{R}^{k_\Omega\times r}$ such that

$$\Xi = \Pi V. \tag{L.1}$$

Since Π is a membership matrix, it follows that the rows of Ξ are convex combinations of the k_Ω rows of V. Because Π is eligible, the identity matrix is a submatrix of Π. Without loss of generality, assume that $\Pi_{|\{1,\dots,k_\Omega\},\cdot} = I_{k_\Omega}$. It follows that $V = \Xi_{\{1,\dots,k_\Omega\},\cdot}$. This shows that $C(\Xi)$ is a polytope with at most k_Ω vertices and at least r vertices.

In the case that $k_\Omega = r$, the desired result follows immediately. In the case that $k_\Omega < r$, *hic jacet lepus*. Suppose by contradiction that V has only N distinct rows, where $r \le N < k_\Omega$. This means that we can write $\Xi = \Pi B\tilde{V}$, where $\tilde{V} \in \mathbb{R}^{N\times r}$ is the matrix containing the unique rows of V and $B \in \mathbb{R}^{k_\Omega\times N}$ is a row-replication matrix (which admits the identity matrix I_N as a submatrix). It follows that we can write:

$$\Omega = \Pi B\tilde{V}\Lambda\tilde{V}'B'\Pi'.$$

We denote $\tilde{\Pi} := \Pi B$ and $\tilde{P} = \tilde{V}\Lambda\tilde{V}'$, and proceed to showing that these matrices are eligible. First, it is straightforward to see that for any $i \in \{1,\dots,n\}$, the i-th row of $\tilde{\Pi}$ is positive and verifies $\tilde{\pi}_i'\mathbf{1}_N = \pi_i'B\mathbf{1}_N = \pi'\mathbf{1}_K = 1$. In addition, since both Π admits I_{k_Ω} as a submatrix and B admits I_N as a submatrix, it follows that $\tilde{\Pi}$ admits I_N as a submatrix. This shows that $\tilde{\Pi}$ is admissible.

Now, from Equation (L.1), we know that $\Omega = \Pi V\Lambda V'\Pi'$, so $P = V\Lambda V' = B\tilde{V}\Lambda\tilde{V}'B'$. By definition, B admits a left inverse, call it $Q \in \{0,1\}^{N\times k_\Omega}$, so that $QB = I_2$. Then $\tilde{P} = QPQ'$. Since both Q and P are nonnegative, it follows that $\tilde{P}$ is nonnegative, thus eligible.

We have shown that we can write $\Omega = \tilde{\Pi}\tilde{P}\tilde{\Pi}'$, where $(\tilde{\Pi},\tilde{P}) \in \mathbb{R}^{n\times N}\times\mathbb{R}^{N\times N}$ eligible and $N < k_\Omega$, QEA. □