Carnegie Mellon University
Dietrich College of Humanities and Social Sciences
Dissertation

Submitted in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy

Title: Network global testing and estimating number of network communities
Presented by: Shengming Luo
Accepted by: Department of Statistics & Data Science
Readers:

JIASHUN JIN, ADVISOR — DATE

ZHENG TRACY KE, ADVISOR — DATE

JING LEI — DATE

NYNKE NIEZINK — DATE

ALESSANDRO RINALDO — DATE

COSMA SHALIZI — DATE

LARRY WASSERMAN — DATE

Approved by the Committee on Graduate Degrees:

RICHARD SCHEINES, DEAN — DATE

Network global testing and estimating number of network communities using graphlets

Shengming Luo

April 30, 2021

A dissertation submitted in partial fulfillment
of the requirements for the Degree of Doctor of Philosophy

Department of Statistics & Data Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213

Thesis Committee:
Jiashun Jin, Chair
Zheng Tracy Ke, Chair
Jing Lei
Nynke Niezink
Alessandro Rinaldo
Cosma Shalizi
Larry Wasserman

To my beloved parents and family

Abstract

Given a symmetrical social network, the network global testing is where we use the adjacency matrix of the network to test whether it has only one community or multiple communities. It's also naturally connected to the problem of estimating the number of network communities, which is arguably one of the most important problem in network analysis area. Despite many interesting works in recent years, it remains unclear how to find test statistics and estimators that are (a) applicable to networks with severe degree heterogeneity and mixed-memberships with varying sparsity, and is (b) optimal. This thesis aims to design statistics to solve the above two problems, under a more realistic network model. To assess optimality, we use the phase transition framework, which includes the standard minimax argument, but is more informative.

In the first part of this thesis, we focus on the network global testing problem and propose the Signed Polygon as a class of new tests. Fixing $m \geq 3$, for each m-gon in the network, define a score using the centered adjacency matrix. The sum of such scores is then the m-th order Signed Polygon statistic. The Signed Quadrilateral (SgnQ) is special example of the Signed Polygon with $m = 4$. We show that SgnQ test satisfies (a) accommodate severe degree heterogeneity, (b) accommodate mixed-memberships, (c) have a tractable null distribution, and (d) adapt automatically to different levels of sparsity, and achieve the optimal phase diagram. and especially, it works well for very sparse and less sparse networks. Our proposed tests compare favorably with the existing tests and achieve the optimal phase diagram. Also, many existing tests do not allow for severe heterogeneity or mixed-memberships, and they behave unsatisfactorily in our settings.

In the second part of the thesis, we propose Stepwise Goodness-of-Fit (StGoF) as a new approach to estimating K, the number of network communities. For $m = 1, 2, \ldots$, StGoF alternately uses a community detection step (pretending m is the correct number of communities) and a goodness-of-fit step. We use SCORE Jin (2015) for community detection, and propose a new goodness-of-fit measure. Denote the goodness-of-fit statistic in step m by $\psi_n^{(m)}$. We show that as $n \to \infty$, $\psi_n^{(m)} \to N(0,1)$ when $m = K$ and $\psi_n^{(m)} \to \infty$ in probability when $m < K$. Therefore, with a proper threshold, StGoF terminates at $m = K$ as desired. We consider a broad setting that allows severe degree heterogeneity, a wide range of sparsity, and especially weak signals. In particular, we propose a measure for signal-to-noise ratio (SNR) and show that there is a phase transition: when SNR $\to 0$ as $n \to \infty$, consistent estimates for K do not exist, and when SNR $\to \infty$, StGoF is consistent, uniformly for a broad class of settings. In this sense, StGoF achieves the optimal phase transition. Stepwise

testing algorithms of similar kind (e.g., Wang et al. (2017); Ma et al. (2018)) are known to face analytical challenges. We overcome the challenges by using a different design in the stepwise algorithm and by deriving sharp results in the under-fitting case ($m < K$) and the null case ($m = K$). The key to our analysis is to show that SCORE has the Non-Splitting Property (NSP). The NSP is non-obvious, so additional to rigorous proofs, we also provide an intuitive explanation.

Acknowledgments

First and foremost, I would like to thank my advisors, Jiashun Jin and Zheng Tracy Ke, for their guidance, support and encouragement. Prof. Jin has great insight into maths and statistics and can always point out a direction whenever my research gets to a dead end. His enthusiasm and high standard towards academia have always been a role model to me. I am deeply grateful for him dedicating so much time teaching me, and pushing me towards academic perfection. Prof. Ke has a solid foundation in maths and probability. She has always been very supportive whenever I was stuck with technical issues during the projects. I am deeply thankful for her help and mentoring.

Second, I would like to thank Jing Lei, Nynke Niezink, Alessandro Rinaldo, Cosma Shalizi and Larry Wasserman for serving my thesis committee. Especially, I would like to thank Jing and Ale for their continuing help and illuminating discussions, which shed light on both my research and life. I would like to thank Larry for supporting my pursuing to academia. I would also like to thank Nynke for organizing the Network workshop each week and Cosma for the introductory and advanced courses for network analysis. Both resources are very helpful in forming my research ideas and generalizing research horizons.

Third, I would like to thank all the support and friendship from and outside of Carnegie Mellon University. Especially, I would like to thank all the Department of Statistics and Data Science members.

Finally, I would like to thank my parents Jinglei Luo, and Qiuping Yao for their continuing love, support, and understanding. I would like to thank Minshi Peng for her love and help throughout the whole Ph.D. career.

Contents

One

Introduction

Network data encodes connections between units of analysis, which introduces many interesting research questions with broad applications. This thesis focus on two important and related problems in network analysis area.

- *Network global testing problem* Given a symmetric social network, how to test whether it has only one community or multiple communities.
- *Estimating number of network communities* Given a symmetric social network, how many communities are there?

Real world networks have several characteristics that are ubiquitously found:

- *Severe degree heterogeneity.* The distribution of the node degrees usually has a power-law tail, implying severe degree heterogeneity.
- *Mixed-memberships.* Communities are tightly woven clusters of nodes where we have more edges within than between. Communities are rarely non-overlapping, and some nodes may belong to more than one community (and thus have mixed-memberships).
- *Sparsity.* Many networks are sparse. The sparsity levels may range significantly from one network to another.
- *Weak signal.* The community structure is masked by strong noise, and the signal-to-noise ratio (SNR) is usually relatively small.

These features pose great challenges to both the modeling and inference of network data. Most of existing works modeling the network with Stochastic Block Model (SBM), which is well known for oversimplifying the features observed in real world networks. Instead, this thesis would focus on more realistic and complicated network models. To understand the statistical limits of the above problems, we adopt the phase transition framework, which includes the classical minimax theory as a special case but is more informative.

1.1 Network global testing

Recently, the global testing problem has attracted much attention. A good understanding of the problem is useful for discovering non-obvious social groups and patterns Béjar et al.

(2016); Du and Yang (2011), measuring diversity of individual nodes Fu et al. (2015), determining stopping time in a recursive community detection scheme Li et al. (2018); Zhao et al. (2011). It may also help understand other related problems such as membership estimation Zhang et al. (2014), and estimating the number of the communities Saldana et al. (2017); Wang et al. (2017).

Many interesting approaches have been proposed. Mossel et al. (2015) and Banerjee and Ma (2017) (see also Banks et al. (2016)) considered a special case of the testing problem, where they assume a simple null of Erdos-Renyi random graph model and a special alternative which is an SBM with two equal-sized communities. They provided the asymptotic distribution of the log-likelihood ratio within the contiguous regime. Since the likelihood ratio test statistic is NP-hard to compute, Banerjee and Ma (2017) introduced an approximation by linear spectral statistics. Lei (2016) also considered the SBM model and studied the problem of testing whether $K = K_0$ or $K > K_0$, which is based on the Tracy-Widom law of extreme eigenvalues and requires delicate random matrix theory. Unfortunately, these work have been focused on the SBM (which allows neither severe degree heterogeneity nor mixed membership). Therefore, despite the elegant theory in these works, it remains unclear how to extend their ideas to our settings. The approach by Gao and Lafferty (2017) is probably the first that tackles the global testing problem in settings that allow severe degree heterogeneity, but still in a relatively idealized setting. Jin et al. (2018) considered the problem in much broader settings, with a very different theoretical framework. They suggested a general recipe for constructing test statistics that have $N(0, 1)$ as the asymptotic null distribution, and proposed a class of test statistics called the graphlet counting (GC), which includes the EZ test as a special case. They explained why both GC and EZ tests are reasonable ideas (for settings much broader than that of Gao and Lafferty (2017)) and showed that both tests have competitive power in many cases.

Compared to previous works, our contributions are as follows:

- Identify the Region of Impossibility and the Region of Possibility in the phase space.
- Propose the Signed Polygon as a class of new tests that are appropriate for networks with severe degree heterogeneity and mixed-memberships, with an easy-to-track asymptotic null distribution.
- Prove that the Signed Triangle and Signed Quadrilateral tests are optimally adaptive and perform well for all networks in the Region of Possibility, ranging from very sparse ones to the least sparse ones.

To show the success of the Signed Polygon test for the whole Region of Possibility is very subtle and extremely tedious. The main reason is that we hope to cover the whole spectrum of degree heterogeneity and sparsity levels. Crude bounds may work in one case but not another, and many seemingly negligible terms turn out to be non-negligible. The lower bound argument is also very subtle. Compared to work on SBM where there is only one unknown parameter under the null, our null has n unknown parameters. The difference provides a lot of freedom in constructing inseparable hypothesis pairs, and so the Region of Impossibility in our setting is much wider than that for SBM. Our construction of inseparable hypothesis pairs uses theorems on non-negative matrix scaling, a mathematical area pioneered by Sinkhorn (1974) and Marshall and Olkin (1968) among others (e.g., Brualdi (1974); Johnson and Reams (2009)).

1.2 Estimating number of network communities

In network analysis, how to estimate the number of communities K is a fundamental problem. In many recent approaches, K is assumed as known a priori (see for example Chen et al. (2018); Gao et al. (2018); Karrer and Newman (2011); Ma et al. (2020); Zhao et al. (2011); Xu et al. (2020) on community detection, Jin et al. (2017); Zhang et al. (2014) on mixed-membership estimation, and Liu et al. (2017) on dynamic community detection). Unfortunately, K is rarely known in applications, so the performance of these approaches hinges on how well we can estimate K.

In recent years, many interesting approaches for estimating K have been proposed. Le and Levina (2015) proposed to estimate K using the eigenvalues of the non-backtracking matrix or Bethe Hessian matrix, using ideas from mathematical graph theory. Unfortunately, the approach requires relatively strong conditions for consistency. Liu et al. (2019) proposed to estimate K by using the classical scree plot approach with careful theoretical justification, but the approach is known to be unsatisfactory in the presence of severe degree heterogeneity, for it is hard to derive a sharp bound for the spectral norm of the noise matrix W. Saldaña et al. (2017) used a BIC-type objective function and Daudin et al. (2008); Latouche et al. (2012) used an objective function of the Bayesian model selection flavor. However, these methods did not provide explicit theoretical guarantee on consistency (though a partial result was established in Li et al. (2020), which stated that under SBM, the proposed estimator $\widehat{K}$ is no greater than K with high probability). Wang et al. (2017) proposed to estimate K by solving a BIC type optimization problem, where the objective function is the sum of the log-likelihood and the model complexity. The major challenge here is that the likelihood is the sum of exponentially many terms and is hard to compute. In a remarkable paper, Ma et al. (2018) extended the idea of Wang et al. (2017) by proposing a new approach that is computationally more feasible.

Compared to previous works, our contributions are as follows.

- We propose StGoF as a new approach to estimating K. For $m = 1, 2, \ldots$, StGoF alternately uses two sub-steps, a community detection sub-step where we apply SCORE Jin (2015), assuming m is the correct number of communities, and a Goodness-of-Fit (GoF) sub-step.

- We derive $N(0, 1)$ as the explicit limiting null distribution for the GoF sub-step, and use the NSP of SCORE to derive tight bounds in the under-fitting case. These sharp results and the design of StGoF allow us to avoid the analysis in the over-fitting case and so to overcome the technical challenges faced by stepwise testing of this kind.

- We show that StGoF achieves the optimal phase transition under mild conditions and consistent in broad settings (e.g., weak signals, severe degree heterogeneity, and a wide range of sparsity).

1.3 Outline

The reminder of the thesis is organized as follows. In the second chapter, we propose Signed Polygons as a novel class of global testing statistic, with SgnQ as a special case. We prove SgnQ test is applicable to a wide range of networks, including those with severe degree heterogeneity and mixed memberships. Moreover, for a broad class of parameter settings where only minimum regularity conditions are required. For lower bound, we use a phase transition framework and show that SgnQ achieves the optimal phase transition diagram. For the third chapter, propose Stepwise Goodness-of-Fit (StGoF) as a new approach to estimating K, the number of network communities in a given network. We consider a broad setting where we allow severe degree heterogeneity, a wide range of sparsity, and especially weak signals. We also prove that StGoF achieves the optimal phase transition diagram.

The second chapter is based on the paper Jin et al. (2019), and is co-supervised by Professor Jiashun Jin and Professor Zheng Tracy Ke. The third chapter is based on Jin et al. (2020), and is co-supervised by Professor Jiashun Jin and Professor Zheng Tracy Ke, and Minzhe Wang and Shengming Luo have contributed equally.

My research on community detection Jin et al. (2021a), mixed membership estimation Jin et al. (2017) and network pairwise comparison Jin et al. (2021b) is not included in this thesis.

Two

Optimal Adaptivity of Signed-Polygon Statistics for Network Testing

2.1 Introduction

Given a symmetrical social network, we are interested in the *global testing problem* where we use the adjacency matrix of the network to test whether it has only one community or multiple communities. A good understanding of the problem is useful for discovering non-obvious social groups and patterns Béjar et al. (2016); Du and Yang (2011), measuring diversity of individual nodes Fu et al. (2015), determining stopping time in a recursive community detection scheme Li et al. (2018); Zhao et al. (2011). It may also help understand other related problems such as membership estimation Zhang et al. (2014), and estimating the number of the communities Saldana et al. (2017); Wang et al. (2017).

Phase transition is a well-known optimality framework Donoho and Jin (2004); Ingster et al. (2010); Ma and Wu (2015); Paul (2007). It is related to the minimax framework but can be more informative in many cases. Conceptually, for the global testing problem, in the two-dimensional phase space with the two axes calibrating the "sparsity" and "signal strength", respectively, there is a "Region of Possibility" and a "Region of Impossibility". In "Region of Possibility", any alternative is separable from the null. In "Region of Impossibility", any alternative is inseparable from the null. If a test is able to automatically adapt to different levels of sparsity and is able to separate any given alternative in the "Region of Possibility" from the null, then we call it "optimally adaptive".

We are interested in finding tests that satisfy the following requirements.

- (R1). Applicable to networks with severe degree heterogeneity.
- (R2). Applicable to networks with mixed-memberships.
- (R3). The asymptotic null distribution is easy to track, so the rejection regions are easy to set.
- (R4). Optimally adaptive: We desire a single test that is able to adapt to different levels of sparsity and is optimally adaptive.

2.1.1 The DCMM model

We adopt the *Degree Corrected Mixed Membership (DCMM)* model Zhang et al. (2014); Jin et al. (2017). Denote the adjacency matrix by A, where

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ and node } j \text{ have an edge,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.1.1}$$

Conventionally, self-edges are not allowed so all the diagonal entries of A are 0. In DCMM, we assume there are K perceivable communities $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$, and each node is associated with a mixed-membership weight vector $\pi_i = (\pi_i(1), \pi_i(2), \ldots, \pi_i(K))'$ where for $1 \leq k \leq K$ and $1 \leq i \leq n$,

$$\pi_i(k) = \text{the weight node } i \text{ puts in community } k. \tag{2.1.2}$$

Moreover, for a $K \times K$ symmetric nonnegative matrix P which models the community structure, and positive parameters $\theta_1, \theta_2, \ldots, \theta_n$ which model the degree heterogeneity, we assume the upper triangular entries of A are independent Bernoulli variables satisfying

$$\mathbb{P}(A_{ij} = 1) = \theta_i \theta_j \cdot \pi_i' P \pi_j \equiv \Omega_{ij}, \qquad 1 \leq i < j \leq n, \tag{2.1.3}$$

where Ω denotes the matrix $\Theta \Pi P \Pi' \Theta$, with Θ being the $n \times n$ diagonal matrix $\mathrm{d}iag(\theta_1, \ldots, \theta_n)$ and Π being the $n \times K$ matrix $[\pi_1, \pi_2, \ldots, \pi_n]'$. For identifiability (see Jin et al. (2017) for more discussion), we assume

$$\text{all diagonal entries of } P \text{ are } 1. \tag{2.1.4}$$

When $K = 1$, (3.2.4) implies $P = 1$, and so $\Omega_{ij} = \theta_i \theta_j$, $1 \leq i, j \leq n$.

Write for short $\mathrm{d}iag(\Omega) = \mathrm{d}iag(\Omega_{11}, \Omega_{22}, \ldots, \Omega_{nn})$, and let W be the matrix where for $1 \leq i, j \leq n$, $W_{ij} = A_{ij} - \Omega_{ij}$ if $i \neq j$ and $W_{ij} = 0$ otherwise. In matrix form, we have

$$A = \Omega - \mathrm{d}iag(\Omega) + W, \qquad \text{where } \Omega = \Theta \Pi P \Pi' \Theta. \tag{2.1.5}$$

DCMM includes three models as special cases, each of which is well-known and has been studied extensively recently.

- *Degree Corrected Block Model (DCBM)* Karrer and Newman (2011). If we do not allow mixed-memberships (i.e., each weight vector π_i is degenerate with one entry being nonzero), then DCMM reduces to the DCBM.
- *Mixed Membership Stochastic Block Model (MMSBM)* Airoldi et al. (2008). If $\theta_1 = \theta_2 = \ldots = \theta_n$ and we denote the common value by α_n, then Ω reduces to $\Omega = \alpha_n \Pi P \Pi'$. For identifiability in this special case, (3.2.4) is too strong, so we relax it to that the average of the diagonals of P is 1.
- *Stochastic Block Model (SBM)* Holland et al. (1983). MMSBM further reduces to the classical SBM if additionally we do not allow mixed-memberships.

Under DCMM, the global testing problem is the problem of testing

$$H_0^{(n)} : K = 1 \qquad \text{vs.} \qquad H_1^{(n)} : K \geq 2. \tag{2.1.6}$$

The seeming simplicity of the two hypotheses is deceiving, as both of them are highly composite, consisting of many different parameter configurations.

2.1.2 Phase transition: a preview of our main results

Let $\lambda_1, \lambda_2, \ldots, \lambda_K$ be the first K eigenvalues of Ω, arranged in the descending order in magnitude. We can view (a) $\sqrt{\lambda_1}$ both as the sparsity level and the noise level Jin (2015) (i.e., spectral norm of the noise matrix W), (b) $|\lambda_2|$ as the signal strength, so $|\lambda_2|/\sqrt{\lambda_1}$ is the Signal-to-Noise Ratio (SNR), and (c) $|\lambda_2|/\lambda_1$ as a measure for the similarity between different communities.

Now, in the two-dimensional *phase space* where the x-axis is $\sqrt{\lambda_1}$ which measures the sparsity level, and the y-axis is $|\lambda_2|/\lambda_1$ which measures the community similarity, we have two regions.

- *Region of Possibility (*$1 \ll \sqrt{\lambda_1} \ll \sqrt{n},\ |\lambda_2|/\sqrt{\lambda_1} \to \infty$*).* For any alternative hypothesis in this region, it is possible to distinguish it from any null hypothesis, by the Signed Polygon tests to be introduced.
- *Region of Impossibility (*$1 \ll \sqrt{\lambda_1} \ll \sqrt{n},\ |\lambda_2|/\sqrt{\lambda_1} \to 0$*).* In this region, any alternative hypothesis is inseparable from the null hypothesis, provided that some mild conditions hold.

See Figure 2.1 (left panel). The Signed Polygon test satisfies all requirements (R1)-(R4) aforementioned. Since the test is able to separate all alternatives (ranging from very sparse to less sparse) in the Region of Possibility from the null, it is *optimally adaptive.*

To further elucidate, consider the special DCMM in Example 1, where

$$\lambda_1 \sim (1 + (K-1)b_n)\|\theta\|^2, \qquad \lambda_k \sim (1-b_n)\|\theta\|^2, \ \ k = 2, 3, \ldots, K.$$

The sparsity level is $\sqrt{\lambda_1} \asymp \|\theta\|$, and the SNR is $|\lambda_2|/\sqrt{\lambda_1} \asymp \|\theta\|(1-b_n)$, where $(1-b_n)$ measures the community similarity. In this example, the Region of Possibility and Region of Impossibility are defined by

$$\{1 \ll \|\theta\| \ll \sqrt{n}, \|\theta\|(1-b_n) \to \infty\}, \text{ and } \{1 \ll \|\theta\| \ll \sqrt{n}, \|\theta\|(1-b_n) \to 0\},$$

respectively. See Figure 2.1 (right panel).

Remark 1. As the phase transition is hinged on $\lambda_2/\sqrt{\lambda_1}$, one may think that the statistic $\widehat{\lambda}_2/\sqrt{\widehat{\lambda}_1}$ is optimally adaptive, where $\widehat{\lambda}_k$ is the k-th eigenvalue of A, $1 \le k \le K$, arranged in the descending order in magnitude. This is however not true, for the consistency of $\widehat{\lambda}_2$ to λ_2 can not be guaranteed in our range of interest, unless with strong conditions on θ_{max} Jin (2015).

2.1.3 Literature review, the Signed Polygon and our contribution

Recently, the global testing problem has attracted much attention and many interesting approaches have been proposed. To name a few, Mossel et al. (2015) and Banerjee and Ma (2017) (see also Banks et al. (2016)) considered a special case of the testing problem, where they assume a simple null of Erdos-Renyi random graph model and a special alternative which is an SBM with two equal-sized communities. They provided the asymptotic distribution of

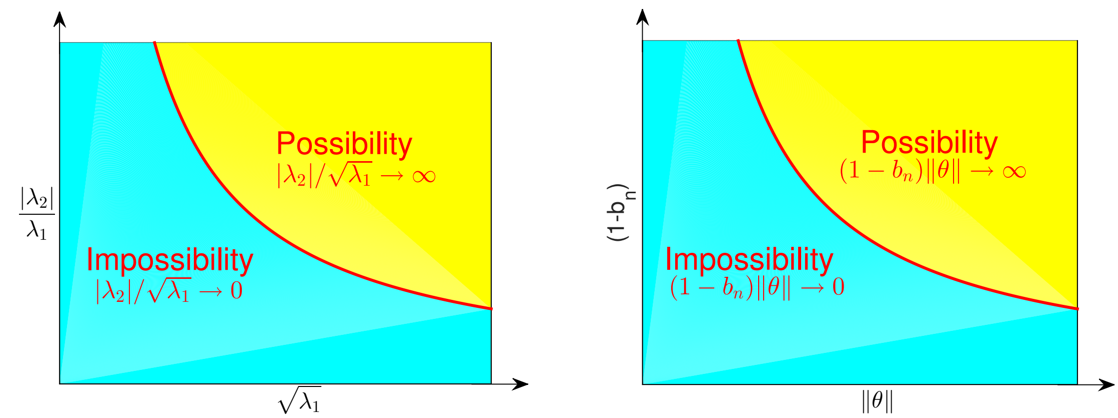

Figure 2.1: Left: Phase transition. In Region of Impossibility, any alternative hypothesis is indistinguishable from a null hypothesis, provided that some mild conditions hold. In Region of Possibility, the Signed Polygon test is able to separate any alternative hypothesis from a null hypothesis asymptotically. Right: Phase transition for the special DCMM model in Example 1, where $\sqrt{\lambda_1} \asymp \|\theta\|$, $|\lambda_2|/\lambda_1 \asymp (1-b_n)$, and $|\lambda_2|/\sqrt{\lambda_1} \asymp (1-b_n)\|\theta\|$.

the log-likelihood ratio within the contiguous regime. Since the likelihood ratio test statistic is NP-hard to compute, Banerjee and Ma (2017) introduced an approximation by linear spectral statistics. Lei (2016) also considered the SBM model and studied the problem of testing whether $K = K_0$ or $K > K_0$, where K_0 is the number of communities. His approach is based on the Tracy-Widom law of extreme eigenvalues and requires delicate random matrix theory. Unfortunately, these work have been focused on the SBM (which allows neither severe degree heterogeneity nor mixed membership). Therefore, despite the elegant theory in these works, it remains unclear how to extend their ideas to our settings.

The approach by Gao and Lafferty (2017) is probably the first that tackles the global testing problem in settings that allow severe degree heterogeneity. They showed that the EZ test has a null that is asymptotically $N(0,1)$, and has competitive powers in many interesting settings. However, they only considered a relatively idealized setting where the off-diagonal entries of P are all equal and where (θ_i, π_i)'s are *iid* generated (see details therein), and whether their ideas continue to work in our setting remains unclear.

Jin et al. (2018) considered the problem in much broader settings, with a very different theoretical framework. They suggested a general recipe for constructing test statistics that have $N(0,1)$ as the asymptotic null distribution, and proposed a class of test statistics called the *graphlet counting (GC)*, which includes the EZ test as a special case. They explained why both GC and EZ tests are reasonable ideas (for settings much broader than that of Gao and Lafferty (2017)) and showed that both tests have competitive power in many cases.

At that time, our hope was that the GC test is the desired test. We tried very hard to analyze the GC test, hoping that it satisfies (R1)-(R4). Unfortunately, after substantial time and efforts, we found that in the less sparse case, the variance of the GC test becomes unsatisfactorily large, and so the test loses power in many easy-to-test scenarios and is not optimally adaptive. Fortunately, right at the moment of despair, we came to realize that

- Especially for the less sparse case, the key to constructing a powerful test is not how to capture the signal, but to reduce the variance.
- The GC test is based on counts of *non-centered* cycles/paths. The variance can be much smaller if we count the *centered* cycles instead.

Centered and non-centered cycles are defined on the centered and non-centered adjacency matrix, respectively. See details below.

These insights motivate a class of new tests which we call *Signed Polygon*, including the Signed Triangle (SgnT) and the Signed Quadrilateral (SgnQ). The Signed Polygon statistics are related to the Signed Cycle statistics, first introduced by Bubeck et al. Bubeck et al. (2016) and later generalized by Banerjee Banerjee (2018).

The Signed Polygon and the Signed Cycle are cycle-counting approaches, both of which recognize the benefit of variance reduction by counting centered cycles instead of non-centered cycles, but there are some major differences. The study of the Signed Cycles has been focused on the SBM and similar models, where under the null, $\mathbb{P}(A_{ij} = 1) = \alpha$, $1 \leq i \neq j \leq n$, and α is the only unknown parameter. In this case, a natural approach to centering the adjacency matrix A is to first estimate α using the whole matrix A (say, $\widehat{\alpha}$), and then subtract all off-diagonal entries of A by $\widehat{\alpha}$. However, under the null of our setting, $\mathbb{P}(A_{ij} = 1) = \theta_i\theta_j$, $1 \leq i \neq j \leq n$, and there are n different unknown parameters $\theta_1, \theta_2, \ldots, \theta_n$. In this case, how to center the matrix A is not only unclear but also *worrisome*, especially when the network is very sparse, because we have to use limited data to estimate a large number of unknown parameters. Also, for any approaches we may have, the analysis is seen to be much harder than that of the previous case.

Note that the ways how two statistics are defined over the centered adjacency matrix are also different. See Section 2.1.4 and Bubeck et al. (2016).

In the Signed Polygon, we use a new approach to estimate $\theta_1, \theta_2, \ldots, \theta_n$ under the null, and use the estimates to center the matrix A. To our surprise, data limitation (though a challenge) does not ruin the idea, and even for very sparse networks, the estimation errors of $\theta_1, \theta_2, \ldots, \theta_n$ only have a negligible effect. The main contributions of the chapter are as follows.

- Identify the Region of Impossibility and the Region of Possibility in the phase space.
- Propose the Signed Polygon as a class of new tests that are appropriate for networks with severe degree heterogeneity and mixed-memberships.
- Prove that the Signed Triangle and Signed Quadrilateral tests satisfy all the requirements (R1)-(R4), and especially that they are optimally adaptive and perform well for all networks in the Region of Possibility, ranging from very sparse ones to the least sparse ones.

To show the success of the Signed Polygon test for the whole Region of Possibility is very subtle and extremely tedious. The main reason is that we hope to cover the *whole spectrum* of degree heterogeneity and sparsity levels. Crude bounds may work in one case but not another, and many seemingly negligible terms turn out to be non-negligible (see Sections 2.1.4 and 2.4). The lower bound argument is also very subtle. Compared to work on SBM where there is only one unknown parameter under the null, our null has n unknown parameters. The difference provides a lot of freedom in constructing inseparable hypothesis pairs, and so the Region of Impossibility in our setting is much wider than that for SBM. Our construction of inseparable hypothesis pairs uses theorems on non-negative matrix scaling, a mathematical area pioneered by Sinkhorn Sinkhorn (1974) and Olkin Marshall and Olkin (1968) among others (e.g., Brualdi (1974); Johnson and Reams (2009)).

2.1.4 The Signed Polygon statistic

Recall that A is the adjacency matrix of the network. Introduce a vector $\widehat{\eta}$ by ($\mathbf{1}_n$ denotes the vector of 1's)

$$\widehat{\eta} = (1/\sqrt{V})\, A\mathbf{1}_n, \qquad \text{where } V = \mathbf{1}_n' A \mathbf{1}_n. \tag{2.1.7}$$

Fixing $m \geq 3$, the order-m *Signed Polygon* statistic is defined by (notation: $(dist)$ is short for "distinct", which means any two of $i_1, \ldots, i_m$ are unequal)

$$U_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} (A_{i_1 i_2} - \widehat{\eta}_{i_1}\widehat{\eta}_{i_2})(A_{i_2 i_3} - \widehat{\eta}_{i_2}\widehat{\eta}_{i_3}) \ldots (A_{i_m i_1} - \widehat{\eta}_{i_m}\widehat{\eta}_{i_1}). \tag{2.1.8}$$

When $m = 4$, we call it the Signed-Quadrilateral (SgnQ) statistic:

$$Q_n = \sum_{i_1, i_2, i_3, i_4 (dist)} (A_{i_1 i_2} - \widehat{\eta}_{i_1}\widehat{\eta}_{i_2})(A_{i_2 i_3} - \widehat{\eta}_{i_2}\widehat{\eta}_{i_3})(A_{i_3 i_4} - \widehat{\eta}_{i_3}\widehat{\eta}_{i_4})(A_{i_4 i_1} - \widehat{\eta}_{i_4}\widehat{\eta}_{i_1}). \tag{2.1.9}$$

For analysis, we focus on Q_n, but the theoretical framework is extendable to general m.

The key to understanding and analyzing the Signed Polygon is the *Ideal Signed Polygon*. Introduce a *non-stochastic counterpart* of $\widehat{\eta}$ by

$$\eta^* = \Omega\mathbf{1}_n/\sqrt{v_0}, \qquad \text{where } v_0 = \mathbf{1}_n'\Omega\mathbf{1}_n. \tag{2.1.10}$$

Define the order-m *Ideal Signed Polygon* statistic by

$$\widetilde{U}_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} (A_{i_1 i_2} - \eta^*_{i_1}\eta^*_{i_2})(A_{i_2 i_3} - \eta^*_{i_2}\eta^*_{i_3}) \ldots (A_{i_m i_1} - \eta^*_{i_m}\eta^*_{i_1}). \tag{2.1.11}$$

We expect to see that

$$\widehat{\eta} \approx \mathbb{E}[\widehat{\eta}] \approx \eta^*.$$

We can view $\widetilde{U}_n^{(m)}$ as the oracle version of $U_n^{(m)}$, with η^* given. We can also view $U_n^{(m)}$ as the *plug-in* version of $\widetilde{U}_n^{(m)}$, where we replace η^* by $\widehat{\eta}$.

For implementation, it is desirable to rewrite T_n and Q_n in matrix forms, which allows us to avoid using a for loop and compute much faster (say, in MATLAB or R). For any two matrices $M, N \in \mathbb{R}^{n,n}$, let $\mathrm{tr}(M)$ be the trace of M, $diag(M) = diag(M_{11}, M_{22}, \ldots, M_{nn})$, and $M \circ N$ be the Hadamard product of M and N (i.e., $M \circ N \in \mathbb{R}^{n,n}$, $(M \circ N)_{ij} = M_{ij}N_{ij}$). Denote $\widetilde{A} = A - \widehat{\eta}\widehat{\eta}'$. The following theorem is proved in the supplementary material.

Theorem 2.1.1. *We have*

$$\begin{aligned} Q_n = {} & \mathrm{tr}(\widetilde{A}^4) - 4\,\mathrm{tr}(\widetilde{A}\circ\widetilde{A}^3) + 8\,\mathrm{tr}(\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}^2) - 6\,\mathrm{tr}(\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}) \\ & - 2\,\mathrm{tr}(\widetilde{A}^2\circ\widetilde{A}^2) + 2\cdot 1_n'[diag(\widetilde{A})(\widetilde{A}\circ\widetilde{A})diag(\widetilde{A})]1_n + 1_n'[\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}]1_n. \end{aligned}$$

The complexity of computing Q_n is $O(n^2\bar{d})$, where $\bar{d}$ is the average degree of the network.

Compared to the EZ and GC tests proposed in Gao and Lafferty (2017); Jin et al. (2018), the computational complexity of SgnQ is of the same order.

Remark 2 (*Connection to the Signed Cycle*). In the more idealized MMSBM or SBM model, we don't have degree heterogeneity, and $\Omega = \alpha_n \mathbf{1}_n\mathbf{1}_n'$ under the null, where α_n is the only unknown parameter. In this simple setting, it makes sense to estimate α_n by $\widehat{\alpha}_n = \bar{d}/(n-1)$, where $\bar{d}$ is the average degree. This gives rise to the *Signed Cycle* statistics Banerjee (2018); Bubeck et al. (2016):

$$C_n^{(m)} = \sum_{i_1,i_2,\ldots,i_m (dist)} (A_{i_1i_2} - \widehat{\alpha}_n)(A_{i_2i_3} - \widehat{\alpha}_n)\ldots(A_{i_mi_1} - \widehat{\alpha}_n).$$

Bubeck et al. (2016) first proposed $C_n^{(3)}$ for a global testing problem in a model similar to MMSBM. Although their test statistic is also called the Signed Triangle, it is different from our statistic, for their tests are only applicable to models without degree heterogeneity. The analysis of the Signed Polygon is also much more delicate than that of the Signed Cycle, as the error $(\widehat{\alpha}_n - \alpha_n)$ is much smaller than the errors in $(\widehat{\eta} - \eta^*)$.

It remains to understand (1) how the Signed Polygon manages to reduce variance, (2) what are the analytical challenges.

Consider the first question. We illustrate it with the Ideal Signed Polygon (2.1.11) and the null case. In this case, $\Omega = \theta\theta'$. It is seen $\eta^* = \theta$, $A_{ij} - \eta_i^*\eta_j^* = A_{ij} - \Omega_{ij} = W_{ij}$, for $i \neq j$ (see (2.1.5) for definition of W), and so

$$\widetilde{U}_n^{(m)} = \sum_{i_1,i_2,\ldots,i_m (dist)} W_{i_1i_2}W_{i_2i_3}\ldots W_{i_mi_1}.$$

In the sum, each term is an m-product of independent centered Bernoulli variables, and two terms $W_{i_1i_2}W_{i_2i_3}\ldots W_{i_mi_1}$ and $W_{i_1'i_2'}W_{i_2'i_3'}\ldots W_{i_m'i_1'}$ are correlated only when $\{i_1, i_2, \ldots, i_m\}$ and $\{i_1', i_2', \ldots, i_m'\}$ are the vertices of the same polygon. Such a construction is known to be efficient in variance reduction (e.g., Bubeck et al. (2016)).

In comparison, the main term of an order-m GC test statistic Jin et al. (2018) is

$$N_n^{(m)} = \sum_{i_1,i_2,\ldots,i_m (dist)} A_{i_1i_2}A_{i_2i_3}\ldots A_{i_mi_1}.$$

Since here the Bernoulli variables are not centered, we can split $N_n^{(m)}$ into two uncorrelated terms: $N_n^{(m)} = \widetilde{U}_n^{(m)} + (N_n^{(m)} - \widetilde{U}_n^{(m)})$. Compared to the Signed Polygon, the additional variance comes from the second term, which is undesirably large in the less sparse case (Ke, 2019).

Remark 3. The above argument also explains why the order-2 Signed Polygon does not work well. To see the point, note that when $m = 2$, $\widetilde{U}_n^{(m)} = \sum_{i_1\neq i_2} W_{i_1i_2}^2$ under the null, which has an unsatisfactory variance due to the square of the W-terms.

Consider the second question. We discuss with the SgnQ statistic. Recall that η^* is a non-stochastic proxy of $\widehat{\eta}$. For any $1 \le i, j \le n$ and $i \neq j$, we decompose $\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j = \delta_{ij} + r_{ij}$, where δ_{ij} is the main term, which is a linear function of $\widehat{\eta}_i$ and $\widehat{\eta}_j$, and r_{ij} is the remainder term. Introduce

$$\widetilde{\Omega} = \Omega - \eta^*(\eta^*)'. \tag{2.1.12}$$

We have $A_{ij} - \widehat{\eta}_i\widehat{\eta}_j = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}$. After inserting this into Q_n, each 4-product is now the product of 4 bracketed terms, where each bracketed term is the sum of 4 terms. Expanding the brackets and re-organizing, Q_n splits into $4 \times 4 \times 4 \times 4 = 256$ *post-expansion* sums, each of the form

$$\sum_{i_1, i_2, i_3, i_4 (dist)} a_{i_1 i_2} b_{i_2 i_3} c_{i_3 i_4} d_{i_4 i_1},$$

where a is a generic term which can be equal to either of the four terms $\widetilde{\Omega}$, W, δ, and r; same for b, c and d. While some of these terms may be equal to each other, the symmetry we can exploit is limited, due to (a) degree heterogeneity, (b) mixed-memberships, and (c) the underlying polygon structure. As a result, we still have more than 50 post-expansion sums to analyze.

The analysis of a post-expansion sum with the presence of one or more r-term is the most tedious of all, where we need to further decompose each r-term into three different terms. This requires analysis of more than 100 additional post-expansion sums.

At first glance, we may think most of the post-expansion sums are easy to control via a crude bound (e.g., the Cauchy-Schwarz inequality). Unfortunately, this is not the case, and many seemingly negligible terms turn out to be non-negligible. Here are some of the reasons.

- Due to the scarcity of data, the estimation error $(\widehat{\eta}_i - \eta_i)$ is not sufficiently small. Also, severe degree heterogeneity dictates that a crude bound may be enough for some $\widehat{\eta}_i$ but not for other $\widehat{\eta}_i$.
- We aim to cover all interesting sparsity levels: a crude bound may be enough for a specific range of sparsity levels, but not for others.
- We desire to have a *single* test that works for all levels of sparsity. Alternatively, we can find one test that works well for the more sparse case and another test that works well for the less sparse case, but this is less appealing from a practical viewpoint.

As a result, we have to analyze a large number of post-expansion sums, where the analysis is subtle, extremely tedious, and error-prone, involving delicate combinatorics, due to the underlying polygon structure. See Section 2.4.

2.1.5 Organization of the chapter

Section 3.2 focuses on the Region of Possibility and contains the upper bound argument. Section 2.3 focuses on the Region of Impossibility and contains the lower bound argument.

Section 2.4 presents the key proof ideas, with the proof of secondary lemmas deferred to the supplementary material. Section 3.5 presents the numerical study.

For any $q > 0$ and $\theta \in \mathbb{R}^n$, $\|\theta\|_q$ denotes the ℓ^q-norm of θ (when $q = 2$, we drop the subscript for simplicity). Also, θ_{min} and θ_{max} denote $\min\{\theta_1, \ldots, \theta_n\}$ and $\max\{\theta_1, \ldots, \theta_n\}$, respectively. For any $n > 1$, $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector of 1's. For two positive sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$, and we write $a_n \asymp b_n$ if for sufficiently large n, there are two constants $c_2 > c_1 > 0$ such that $c_1 \leq a_n/b_n \leq c_2$. We use $\sum_{i_1,i_2,\ldots,i_m(dist)}$ to denote the sum over all $(i_1, \ldots, i_m)$ such that $1 \leq i_k \leq n$ and $i_k \neq i_\ell$ for $1 \leq k \neq \ell \leq m$ (so the number of summands is $n(n-1)\cdots(n-m+1)$).

2.2 The Signed Polygon test and the upper bound

For reasons aforementioned, we focus our discussion on the SgnQ statistic Q_n, but the ideas are extendable to general Signed Polygon statistics. In Section 2.2.1, we establish the asymptotic normality of two statistics. In Section 2.2.2, we use two statistics to construct two tests, the SgnT test and the SgnQ test. In Section 2.2.3, we discuss the power of the two tests.

In a DCMM model, $\Omega = \Theta\Pi P\Pi'\Theta$, where $\Theta = diag(\theta_1, \ldots, \theta_n)$, and Π is the $n \times K$ membership matrix $[\pi_1, \pi_2, \ldots, \pi_n]'$. We assume as $n \to \infty$,

$$\|\theta\| \to \infty, \quad \theta_{max} \to 0, \quad \text{and} \quad (\|\theta\|^2/\|\theta\|_1)\sqrt{\log(\|\theta\|_1)} \to 0. \tag{2.2.13}$$

The first condition is necessary. In fact, if $\|\theta\| \to 0$, then the alternative is indistinguishable from the null, as suggested by lower bounds in Section 2.3. The second one is mild as the eligible range for θ_{max} is roughly $(n^{-1/2}, 1)$. The last one is weaker than that of $\theta_{max}\sqrt{\log(n)} \to 0$, and is very mild.

Moreover, introduce $G = \|\theta\|^{-2}\Pi'\Theta^2\Pi \in \mathbb{R}^{K\times K}$. This matrix is properly scaled and it can be shown that $\|G\| \leq 1$ (Appendix C, supplementary material). When the null is true, $K = P = G = 1$, and we don't need any additional condition. When the alternative is true, we assume

$$\frac{\max_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i\pi_i(k)\}}{\min_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i\pi_i(k)\}} \leq C, \qquad \|G^{-1}\| \leq C, \qquad \|P\| \leq C. \tag{2.2.14}$$

The conditions are mild. Take the first two for example. When there is no mixed membership, they only require the K classes to be relatively balanced.

2.2.1 Asymptotic normality of the null

The following two theorems are proved in Section 2.4.4.

Theorem 2.2.1 (Limiting null of the SgnQ statistic)**.** *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*, where the condition* (2.2.13) *is satisfied. Suppose the null hypothesis is true. As $n \to \infty$,*

$$\mathbb{E}[Q_n] = (2 + o(1))\|\theta\|^4, \qquad \textit{and} \qquad \mathrm{Var}(Q_n) \sim 8\|\theta\|^8,$$

and

$$\frac{Q_n - \mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \longrightarrow N(0,1), \qquad \textit{in law.}$$

2.2.2 The level-α SgnQ tests

By Theorems 2.2.1, the null variances of the two statistics depend on $\|\theta\|^2$. To use the two statistics as tests, we need to estimate $\|\theta\|^2$. For $\widehat{\eta}$ and η^* defined in (2.1.7) and (2.1.10), respectively, we have $\widehat{\eta} \approx \eta^*$ and $\eta^* = \theta$ under the null. A reasonable estimator for $\|\theta\|^2$ under the null is therefore $\|\widehat{\eta}\|^2$. We propose to estimate $\|\theta\|^2$ with $(\|\widehat{\eta}\|^2 - 1)$, which corrects the bias and is slightly more accurate than $\|\widehat{\eta}\|^2$. The following lemma is proved in the supplementary material.

Lemma 1 (Estimation of $\|\theta\|^2$). *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*, where the condition* (2.2.13) *holds when either hypothesis is true and condition (2.2.14) holds when the alternative is true. Then, under both hypotheses, as $n \to \infty$*

$$(\|\widehat{\eta}\|^2 - 1)/\|\eta^*\|^2 \to 1, \qquad \textit{in probability},$$

where

$$\|\eta^*\|^2 = (\mathbf{1}_n'\Omega^2\mathbf{1}_n)/(\mathbf{1}_n'\Omega\mathbf{1}_n) \begin{cases} = \|\theta\|^2, & \textit{under } H_0^{(n)}, \\ \asymp \|\theta\|^2, & \textit{under } H_1^{(n)}. \end{cases}$$

Combining Theorem 2.2.1 and Lemma 1, we have

$$\frac{Q_n - 2(\|\widehat{\eta}\|^2 - 1)^2}{\sqrt{8(\|\widehat{\eta}\|^2 - 1)^4}} \longrightarrow N(0,1), \qquad \text{in law.} \tag{2.2.15}$$

With the same α, we propose the following SgnQ test, which is a one-sided test where we reject the null hypothesis if and only if

$$Q_n \geq \big(2 + z_\alpha\sqrt{8}\big)(\|\widehat{\eta}\|^2 - 1)^2, \quad z_\alpha\text{: upper } \alpha \text{ quantile of } N(0,1). \tag{2.2.16}$$

As a result, for both tests we just defined, the levels satisfy

$$\mathbb{P}_{H_0^{(n)}}(\text{Reject the null}) \to \alpha, \qquad \text{as } n \to \infty.$$

2.2.3 Power analysis of the SgnQ tests

The matrices Ω and $\widetilde{\Omega}$ play a key role in power analysis. Recall that Ω is defined in (3.1.3) where $\text{rank}(\Omega) = K$, and $\widetilde{\Omega} = \Omega - \eta^*(\eta^*)'$ is defined in (2.1.12) with $\eta^* = \Omega\mathbf{1}_n/\sqrt{\mathbf{1}_n'\Omega\mathbf{1}_n}$ as in (2.1.10). Recall that $\lambda_1, \lambda_2, \ldots, \lambda_K$ are the K nonzero eigenvalues of Ω. Let $\xi_1, \xi_2, \ldots, \xi_K$ be the corresponding eigenvectors. The following theorems are proved in Section 2.4.4.

Theorem 2.2.2 (Limiting behavior of the SgnQ statistic (alternative)). *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*. Suppose the alternative hypothesis is true and the conditions* (2.2.13)-(2.2.14) *hold. As $n \to \infty$,*

$$\mathbb{E}[Q_n] = \text{tr}(\widetilde{\Omega}^4) + o((\lambda_2/\lambda_1)^4\|\theta\|^8) + o(\|\theta\|^4),$$

and

$$\text{Var}(Q_n) \leq C\big(\|\theta\|^8 + C(\lambda_2/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6\big).$$

By Theorem 2.2.2 and Lemma 1, under the alternative hypothesis,

$$\text{the mean and variance of } \tfrac{Q_n - 2(\|\widehat{\eta}\|^2-1)^2}{\sqrt{8(\|\widehat{\eta}\|^2-1)^4}} \text{ are } \tfrac{\text{tr}(\widetilde{\Omega}^4)}{\sqrt{8\|\eta^*\|^8}} \text{ and } \sigma_n^2,$$

respectively, where σ_n^2 denotes the asymptotic variance, which satisfies that

$$\sigma_n^2 \leq \begin{cases} C, & \text{if } |\lambda_2/\lambda_1| \ll \|\theta\|_3^{-1}, \\ C(\lambda_2/\lambda_1)^6 \cdot \|\theta\|_3^6, & \text{if } |\lambda_2/\lambda_1| \gg \|\theta\|_3^{-1}. \end{cases}$$

If we fix the degree heterogeneity vector θ and let (λ_2/λ_1) range, there is a *phase change* in the variance. We shall call:

- the case of $|\lambda_2/\lambda_1| \leq C\|\theta\|_3^{-1}$ as the *weak signal* case for SgnQ.
- the case of $|\lambda_2/\lambda_1| \gg \|\theta\|_3^{-1}$ as the *strong signal* case for SgnQ.

We now analyze $\mathrm{tr}(\widetilde{\Omega}^4)$. The following lemma is proved in the supplementary material.

Lemma 2 (Analysis of $\mathrm{tr}(\widetilde{\Omega}^4)$)**.** *Suppose the conditions of Theorem 2.2.2 hold. Under the alternative hypothesis,*

- *If* $|\lambda_2|/\lambda_1 \to 0$, *then* $\mathrm{tr}(\widetilde{\Omega}^4) = \mathrm{tr}(\Lambda^4) + (q'\Lambda q)^4 + 2(h'\Lambda^2 h)^2 + 4(h'\Lambda h)^2(h'\Lambda^2 h) + 4h'\Lambda^4 h + 4(h'\Lambda h)(h'\Lambda^3 h) + o(\lambda_2^4) \gtrsim \sum_{k=2}^4 \lambda_k^4$.
- *If* $|\lambda_2|/\lambda_1 \geq C$, *then* $\mathrm{tr}(\widetilde{\Omega}^4) \geq C\sum_{k=2}^K \lambda_k^4$.
- *In the special case where* $K = 2$, *the vector* h *is a scalar, and* $\mathrm{tr}(\widetilde{\Omega}^4) = [(h^2+1)^4 + o(1)] \cdot \lambda_2^4$.

As a result, it always holds that $\mathrm{tr}(\widetilde{\Omega}^4) \geq C\sum_{k=2}^K \lambda_k^4$. Then, in the *weak signal* case,

$$\frac{\mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \geq \frac{C\big(\sum_{k=2}^K \lambda_k^4\big)}{\|\theta\|^4} \geq C\Big(\lambda_1^{-2}\sum_{k=2}^K \lambda_k^4\Big),$$

In the *strong signal* case, since $(\lambda_2/\lambda_1)^3 \leq \lambda_1^{-3}(\sum_{k=2}^K \lambda_k^4)^{\frac{3}{4}}$,

$$\frac{\mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \geq \frac{C\big(\sum_{k=2}^K \lambda_k^4\big)}{\lambda_1^{-3}(\sum_{k=2}^K \lambda_k^4)^{\frac{3}{4}}\|\theta\|_3^3\|\theta\|^4} \geq \frac{C\|\theta\|^3}{\|\theta\|_3^3}\Big(\lambda_1^{-2}\sum_{k=2}^K \lambda_k^4\Big)^{\frac{1}{4}},$$

where $\|\theta\|^3/\|\theta\|_3^3 \to \infty$. So, in both cases, the power of the SgnQ test goes to 1 if $\lambda_1^{-2}\sum_{k=2}^K \lambda_k^4 \to \infty$. This is validated in Theorem 2.2.3, which is proved in Section 2.4.4.

Theorem 2.2.3 (Power of the SgnQ test)**.** *Under the conditions of Theorem 2.2.2, for any fixed* $\alpha \in (0,1)$, *consider the SgnQ test in* (2.2.16). *As* $n \to \infty$, *if*

$$\lambda_1^{-1/2}\Big(\sum_{k=2}^K \lambda_k^4\Big)^{1/4} \to \infty,$$

then the Type I error $\to \alpha$, *and the Type II error* $\to 0$.

In summary, Theorem 2.2.3 imply that as long as

$$|\lambda_2|/\sqrt{\lambda_1} \to \infty, \tag{2.2.17}$$

the level of SgnQ test tend to α as expected, and the power tend to 1. The SgnT test requires mild conditions to avoid "signal cancellation", but the SgnQ test has no such issue.

Our simulations further support that SgnQ may have better performance than SgnT. See Section 3.5.

Remark 3. Practically, we prefer to fix α, say, $\alpha = 5\%$. If we allow the level α to change with n, then when (2.2.17) holds, there is a sequence of α_n that tends to 0 slowly enough such that $|\lambda_2|/(z_{\alpha_n/2} \cdot \sqrt{\lambda_1}) \to \infty$. As a result, for either of the two tests, the Type I error $\to 0$ and the power $\to 1$, so the sum of Type I and Type II errors $\to 0$.

2.3 Optimal adaptivity, lower bound, and Region of Impossibility

We now focus on the Region of Impossibility, where $|\lambda_2|/\sqrt{\lambda_1} \to 0$. We first present a standard minimax lower bound, from which we can conclude that there is a sequence of hypothesis pairs (one alternative and one null) that are asymptotically indistinguishable. However, this does not answer the question whether *all alternatives* in the Region of Impossibility are indistinguishable from the null. To answer this question, we need much more sophisticated study; see Section 2.3.2.

2.3.1 Minimax lower bound

Given an integer $K \geq 1$, a constant $c_0 > 0$, and two positive sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\beta_n\}_{n=1}^{\infty}$, we define a class of parameters for DCMM (recall that $\Omega = \Theta\Pi P\Pi'\Theta$, $G = \|\theta\|^{-2}\Pi'\Theta^2\Pi$ and is properly scaled, and λ_k is the k-th largest eigenvalue of Ω in magnitude):

$$\mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$$
$$= \left\{ \begin{array}{ll} (\theta, \Pi, P): & \theta_{\max} \leq \beta_n, \|\theta\|^{-1} \leq \beta_n, \|\theta\|^2\|\theta\|_1^{-1}\sqrt{\log(\|\theta\|_1)} \leq \beta_n, \\ & \frac{\max_k\{\sum_{i=1}^n \theta_i\pi_i(k)\}}{\min_k\{\sum_{i=1}^n \theta_i\pi_i(k)\}} \leq c_0, \|G^{-1}\| \leq c_0, |\lambda_2|/\sqrt{\lambda_1} \leq \alpha_n \end{array} \right\}.$$

For the null case, $K = P = \pi_i = 1$, and the above defines a class of θ, which we write for short by

$$\mathcal{M}_n(1, c_0, \alpha_n, \beta_n) = \mathcal{M}_n^*(\beta_n).$$

The following theorem is proved in the supplementary material:

Theorem 2.3.1 (Minimax lower bound)**.** *Fix $K \geq 2$, a constant $c_0 > 0$, and any sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\beta_n\}_{n=1}^{\infty}$ such that $\alpha_n \to 0$ and $\beta_n \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\inf_{\psi}\Big\{ \sup_{\theta \in \mathcal{M}_n^*(\beta_n)} \mathbb{P}(\psi = 1) + \sup_{(\theta,\Pi,P)\in\mathcal{M}_n(K,c_0,\alpha_n,\beta_n)} \mathbb{P}(\psi = 0)\Big\} \to 1,$$

where the infimum is taken over all possible tests ψ.

The minimax theorem says that in the Region of Impossibility, *there exists a sequence of alternatives* that are inseparable from the null. This does not show what we desire, that is *any sequence in the Region of Impossibility* is inseparable from the null. This is discussed in the next section.

2.3.2 Region of Impossibility

Recall that under DCMM, $\Omega = \Theta\Pi P\Pi'\Theta$ and $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$. Since our model is a mixed-membership latent variable model, in order to characterize the *least favorable* *configuration*, it is conventional to use a *random mixed-membership (RMM) model* for the matrix Π, while (Θ, P) are still non-stochastic. In detail,

- Let $V = \{x \in \mathbb{R}^K, x_k \geq 0, \sum_{k=1}^K x_k = 1\}$.
- Let $V_0 = \{e_1, e_2, \ldots, e_K\}$, where e_k is the k-th Euclidean basis vector.

In DCMM-RMM, we fix a distribution F defined over V and assume

$$\pi_i \overset{iid}{\sim} F, \qquad \text{where } h \equiv \mathbb{E}[\pi_i].$$

If we further restrict that F is defined over V_0, then the network has no mixed-membership, and DCMM-RMM reduces to DCBM-RMM.

The desired result is to show that, for any given P and F, there is a sequence of hypothesis pairs (a null and an alternative)

$$H_0^{(n)}: \quad \Omega = \theta\theta', \qquad \text{and} \qquad H_1^{(n)}: \quad \Omega = \widetilde{\Theta}\Pi P \Pi' \widetilde{\Theta}, \tag{2.3.18}$$

where $\widetilde{\Theta} = diag(\widetilde{\theta}_1, \widetilde{\theta}_1, \ldots, \widetilde{\theta}_n)$ and $\widetilde{\theta}_i$ can be different from θ_i, such that the two hypotheses within each pair are asymptotically indistinguishable from each other, provided that under the alternative $|\lambda_2|/\sqrt{\lambda_1} \to 0$.

Here, since Ω depends on π_i, λ_k is random, and it is more convenient to translate the condition of $|\lambda_2|/\sqrt{\lambda_1} \to 0$ to the condition of

$$\|\theta\| \cdot |\mu_2(P)| \to 0, \tag{2.3.19}$$

where $\mu_k(P)$ is the k-th largest eigenvalue of P in magnitude. The equivalence of two conditions are justified in Appendix D.1 of the supplementary material. The regularity condition (2.2.14) can also be ensured with high probability, by assuming that all entries of $\mathbb{E}[\pi_i]$ are at the order of $O(1)$.

Under the DCBM, the desired result can be proved satisfactorily. The key is the following lemma, which is in the line of Sinkhorn's beautiful work on scalable matrices Sinkhorn (1974) (see also Brualdi (1974); Johnson and Reams (2009); Marshall and Olkin (1968)) and is proved in the supplement.

Lemma 3. *Fix a matrix $A \in \mathbb{R}^{K,K}$ with strictly positive diagonal entries and non-negative off-diagonal entries, and a strictly positive vector $h \in \mathbb{R}^K$, there exists a diagonal matrix $D = diag(d_1, d_2, \ldots, d_K)$ such that $DADh = 1_K$ and $d_k > 0$, $1 \leq k \leq K$.*

In detail, consider a DCBM-RMM setting where $\pi_i \overset{iid}{\sim} F$ and F is supported over V_0 (with possibly unequal probabilities on the K points). Recall $h = \mathbb{E}[\pi_i]$. By Lemma 3, there is a unique diagonal matrix D such that $DPDh = 1_K$. Let

$$\widetilde{\theta}_i = d_k \cdot \theta_i, \qquad \text{if } \pi_i = e_k, \qquad 1 \leq i \leq n, \ 1 \leq k \leq K. \tag{2.3.20}$$

The following theorem is proved in the supplementary material.

Theorem 2.3.2 (Region of Impossibility (DCBM))**.** *Fix $K > 1$ and a distribution F defined over V_0. Consider a sequence of DCBM model pairs indexed by n:*

$$H_0^{(n)}: \Omega = \theta\theta' \qquad \textit{and} \qquad H_1^{(n)}: \Omega = \widetilde{\Theta}\Pi P \Pi' \widetilde{\Theta},$$

where $\pi_i \overset{iid}{\sim} F$ *and* $\widetilde{\Theta} = diag(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$ *with* $\widetilde{\theta}_i$ *defined as in (2.3.20). If* $\theta_{max} \leq c_0$ *for a constant* $c_0 < 1$,

$$\min_{1\leq k\leq K}\{h_k\} \geq C, \qquad \text{and} \qquad \|\theta\| \cdot |\mu_2(P)| \to 0,$$

then for each pair of two hypotheses, the χ^2*-distance between the two joint distributions tends to* 0*, as* $n \to \infty$.

We now generalize the result to DCMM. Fix a distribution F defined over V. Given a set of (Θ, P, Π) with $\Theta = \mathrm{d}iag(\theta_1, \theta_2, \ldots, \theta_n)$ and $\pi_i \overset{iid}{\sim} F$, let $\widetilde{h}_D = \mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$ for any diagonal matrix $D \in \mathbb{R}^{K\times K}$ with positive diagonals. We assume that there exists D such that

$$DPD\widetilde{h}_D = 1_K, \qquad \min_{1\leq k\leq K}\{\widetilde{h}_{D,k}\} \geq C. \tag{2.3.21}$$

When such a D exists, we let

$$\widetilde{\theta}_i = \theta_i/\|D^{-1}\pi_i\|_1, \qquad 1 \leq i \leq n. \tag{2.3.22}$$

When the support of F is restricted to V_0, this reduces to the DCBM setting discussed above, in which (2.3.21) always holds, and $\widetilde{\theta}_i$ is the same as that in (2.3.20). When $K = 2$ (but the support of F is not restricted to V_0), condition (2.3.21) also holds for all matrices A in our setting. The proof is elementary so is omitted. The following theorem is proved in the supplementary material.

Theorem 2.3.3 (Region of Impossibility (DCMM))**.** *Fix* $K > 1$ *and a distribution* F *defined over* V*. Consider a sequence of DCMM model pairs indexed by* n*:*

$$H_0^{(n)} : \Omega = \theta\theta' \qquad \text{and} \qquad H_1^{(n)} : \Omega = \widetilde{\Theta}\Pi P\Pi'\widetilde{\Theta},$$

where $\pi_i \overset{iid}{\sim} F$ *and* $\widetilde{\Theta} = diag(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$ *with* $\widetilde{\theta}_i$ *defined as in (2.3.22). If (2.3.21) holds,* $\theta_{max} \leq c_0$ *for a constant* $c_0 < 1$*, and*

$$\|\theta\| \cdot |\mu_2(P)| \to 0,$$

then for each pair of two hypotheses, the χ^2*-distance between the two joint distributions tends to* 0*, as* $n \to \infty$.

In Theorems 2.3.2 and 2.3.3, we try to be as general as possible, where F and P are arbitrarily given, and we seek for a Θ-matrix in the alternative to make it most delicate to separate two hypotheses. We now consider a special case where P is arbitrarily given, but F is allowed to alter slightly. For any P and F, by Lemma 3, there is a unique positive diagonal matrix D such that

$$DPDh = 1_K, \qquad \text{where} \qquad h = \mathbb{E}[\pi_i]. \tag{2.3.23}$$

Let $\widetilde{\Pi} = [\widetilde{\pi}_1, \widetilde{\pi}_2, \ldots, \widetilde{\pi}_n]'$ and $\widetilde{\Theta} = \mathrm{d}iag(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$, with

$$\widetilde{\pi}_i = D\pi_i/\|D\pi_i\|_1, \qquad \widetilde{\theta}_i = \|D\pi_i\|_1 \cdot \theta_i. \tag{2.3.24}$$

The following theorem is proved in the supplementary material.

Theorem 2.3.4 (Region of Impossibility (DCMM with flexible Π))**.** *Fix* $K > 1$ *and a distribution* F *defined over* V*. Consider a sequence of DCMM model pairs indexed by* n*:*

$$H_0^{(n)} : \Omega = \theta\theta' \qquad \text{and} \qquad H_1^{(n)} : \Omega = \widetilde{\Theta}\widetilde{\Pi}P\widetilde{\Pi}'\widetilde{\Theta},$$

where $\widetilde{\Pi}$ *and* $\widetilde{\Theta}$ *are defined as in (2.3.23)-(2.3.24). If* $\theta_{max} \leq c_0$ *for a constant* $c_0 < 1$,

$$\min_{1 \leq k \leq K}\{h_k\} \geq C, \qquad \textit{and} \qquad \|\theta\| \cdot |\mu_2(P)| \to 0,$$

then for each pair of two hypotheses, the χ^2*-distance between the two joint distributions tends to* 0*, as* $n \to \infty$.

For completeness, one may wonder what happens if we require the null and the alternative have perfectly matching Θ matrix (up to an overall scaling). Such a scenario is natural when we focus on SBM or MMSBM, where degree heterogeneity is not allowed and so there is little freedom in choosing the Θ matrix. In this case, in order that the two hypotheses are indistinguishable, the expected node degrees under the alternative have to match those under the null. For each node $1 \leq i \leq n$, conditional on π_i and neglecting the effect of no self edges, the expected degree equals to

$$\|\theta\|_1 \cdot \theta_i \qquad \text{and} \qquad \|\theta\|_1 \cdot (\pi_i' P h) \cdot \theta_i,$$

under the null and under the alternative, respectively, where $\{\pi_j\}_{j \neq i} \overset{iid}{\sim} F$ and $h = \mathbb{E}[\pi_j]$. For the expected degrees to match under any realized π_i, it is necessary that

$$Ph = q_n 1_K, \qquad \text{for some scaling parameter } q_n > 0. \tag{2.3.25}$$

The following theorem is proved in the supplementary material.

Theorem 2.3.5 (Region of Impossibility (DCMM with matching Θ))**.** *Fix* $K > 1$ *and a distribution* F *defined over* V*. Consider a sequence of DCMM model pairs indexed by* n*:*

$$H_0^{(n)} : \Omega = q_n \cdot \theta\theta' \qquad \textit{and} \qquad H_1^{(n)} : \Omega = \Theta\Pi P \Pi'\Theta,$$

where $\Theta = diag(\theta_1, \theta_2, \ldots, \theta_n)$*,* $\pi_i \overset{iid}{\sim} F$*, and* (P, h, q_n) *satisfy (2.3.25). If* $\theta_{max} \leq c_0$ *for a constant* $c_0 < 1$*,*

$$\min_{1 \leq k \leq K}\{h_k\} \geq C, \qquad \textit{and} \qquad \|\theta\| \cdot |\mu_2(P)| \to 0,$$

then for each pair of two hypotheses, the χ^2*-distance between the two joint distributions tends to* 0*, as* $n \to \infty$.

Example 1 (contd). In Example 1, π_i is drawn from $e_1, e_2, \ldots, e_K$ with equal probabilities, and $P = (1 - b_n)I_K + b_n 1_K 1_K'$. Therefore, $h = \mathbb{E}[\pi_i] = (1/K)1_K$. In this case, all conditions of Theorem 2.3.5 hold, and especially, $q_n = (1/K) + (K-1)b_n/K$, and $\mu_2(P) = (1 - b_n)$.

Remark 6 (Least favorable configuration of LDA-DCMM). The Dirichlet model is often used for mixed-memberships Airoldi et al. (2008). Consider the model pairs

$$H_0^{(n)} : \Omega = q_n\theta\theta' \qquad \text{and} \qquad H_1^{(n)} : \Omega = \Theta\Pi P\Pi'\Theta, \quad \pi_i \overset{iid}{\sim} \text{Dir}(\alpha),$$

where $\text{Dir}(\alpha)$ is a Dirichlet distribution with parameters $\alpha = (\alpha_1, \ldots, \alpha_K)'$. By Theorem 2.3.5, as long as $P\alpha \propto 1_K$, the null and alternative hypotheses are asymptotically indistinguishable if $(1 - q_n)\|\theta\| \to 0$. One can easily construct P such that $P\alpha \propto 1_K$. For example, $P = (1 - q_n)MM' + q_n 1_K 1_K'$, where $M \in \mathbb{R}^{K \times (K-1)}$ is a matrix whose columns are from $Span^{\perp}(\alpha)$ and satisfy $\text{diag}(MM') = I_K$.

2.3.3 Optimal adaptivity

Recall that $\sqrt{\lambda_1}$, $|\lambda_2|/\lambda_1$, and $|\lambda_2|/\sqrt{\lambda_1}$ can be viewed as a measure for the sparsity, community similarity, and SNR, respectively. Combining Theorems 2.2.2, Theorems 2.3.2-2.3.5, and Remark 4 in Section 2.2.3, in the two-dimensional phase space where the x-axis is $\sqrt{\lambda_1}$ and the y-axis is the $|\lambda_2|/\lambda_1$, we have a partition to two regions, the Region of Possibility and the Region of Impossibility.

- **Region of Impossibility** ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} = o(1)$). In this region, any DCBM alternative is asymptotically inseparable from the null, and up to a mild condition, any DCMM alternative is also asymptotically inseparable from the null.

- **Region of Possibility** ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$). In this region, asymptotically, any alternative is completely separable form any null.

The SgnQ test is optimally adaptive: for any alternative in the Region of Possibility, the test is able to separate it from the null with a sum of Type I and Type II errors tending to 0.

To the best of our knowledge, the Signed Polygon is the only known test that is both applicable to general DCMM (where we allow severe degree heterogeneity and arbitrary mixed-memberships) and optimally adaptive. The EZ and GC tests are the only other tests we know that are applicable to general DCMM, but their variances are unsatisfactorily large for the less sparse case, so they are not optimally adaptive.

Remark 4. Most lower bound results in the literature Mossel et al. (2015); Banerjee (2018); Gao and Lafferty (2017) are in the standard minimax framework, where they focus on a particular sequence of alternative (e.g., the off-diagonals of P are equal). In our case, the standard minimax theorem only implies that in the Region of Impossibility, there is a sequence of alternative that are inseparable from the null. Our results (Theorems 2.3.2-2.3.5) are much stronger, implying that *any* alternative in the Region of Possibility is inseparable from the null.

Remark 5. Existing minimax lower bounds Mossel et al. (2015); Banks et al. (2016); Banerjee (2018) have been largely focused on the SBM. Though a least favorable scenario for SBM is also (one of the) least favorable scenario for DCMM, the former does not provide much insight on how the least favorable configurations and the separating boundary of the two regions (Possibility and Impossibility) depend on the degree heterogeneity and mixed-memberships. Moreover, our results suggest that $\|\theta\|$, not $\|\theta\|_1$, determines the separating boundary. In the SBM case, $\theta_1 = \ldots = \theta_n$ and $\|\theta\|_1 = \sqrt{n}\|\theta\|$, so it is hard to tell which of the two norms decides the boundary. In DCMM, there is no simple relationship between $\|\theta\|_1$ and $\|\theta\|$, and we can tell this clearly.

2.4 The behavior of the SgnQ test statistics

Recall that the SgnQ statistic Q_n are defined as

$$Q_n = \sum_{i_1,i_2,i_3,i_4 (dist)} (A_{i_1i_2} - \widehat{\eta}_{i_1}\widehat{\eta}_{i_2})(A_{i_2i_3} - \widehat{\eta}_{i_2}\widehat{\eta}_{i_3})(A_{i_3i_4} - \widehat{\eta}_{i_3}\widehat{\eta}_{i_4})(A_{i_4i_1} - \widehat{\eta}_{i_4}\widehat{\eta}_{i_1}),$$

where

$$\widehat{\eta} = A\mathbf{1}_n/\sqrt{V}, \qquad \text{where } V = \mathbf{1}_n' A\mathbf{1}_n.$$

In Section 2.1.4, we have introduced the following non-stochastic proxy of $\widehat{\eta}$:

$$\eta^* = \Omega\mathbf{1}_n/\sqrt{v_0}, \qquad \text{where } v_0 = \mathbf{1}_n\Omega\mathbf{1}_n.$$

We now introduce another non-stochastic proxy $\widetilde{\eta}$ by

$$\widetilde{\eta} = A\mathbf{1}_n/\sqrt{v}, \qquad \text{where } v = \mathbb{E}[\mathbf{1}_n' A\mathbf{1}_n] = \mathbf{1}_n(\Omega - \mathrm{d}iag(\Omega))\mathbf{1}_n. \tag{2.4.26}$$

Denoting the mean of $\widetilde{\eta}$ by η, it is seen that

$$\eta = ([\Omega - \mathrm{d}iag(\Omega)]\mathbf{1}_n)/\sqrt{\mathbf{1}_n'(\Omega - \mathrm{d}iag(\Omega))\mathbf{1}_n}. \tag{2.4.27}$$

Here, η and η^* are close to each other but η^* has a more explicit form. For example, under the null hypothesis, $\Omega = \theta\theta'$, and it is seen that $\eta^* = \theta$. Recall that

$$A = \Omega - \mathrm{d}iag(\Omega) + W, \qquad \text{and} \qquad \widetilde{\Omega} = \Omega - \eta^*(\eta^*)'.$$

Fix $1 \leq i, j \leq n$ and $i \neq j$. First, we write

$$A_{ij} - \widehat{\eta}_i\widehat{\eta}_j = (A_{ij} - \eta_i^*\eta_j^*) + (\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j) = \widetilde{\Omega}_{ij} + W_{ij} + (\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j).$$

Second, we write

$$\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j = \delta_{ij} + r_{ij},$$

where

$$\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i) \tag{2.4.28}$$

is the linear approximation term of $(\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j)$ and $r_{ij} \equiv (\eta_i^*\eta_j^* - \widehat{\eta}_i\widehat{\eta}_j) - \delta_{ij}$ is the remainder term. By definition and elementary algebra,

$$r_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) - (\eta_i - \widetilde{\eta}_i)(\eta_j - \widetilde{\eta}_j) + (1 - \frac{v}{V})\widetilde{\eta}_i\widetilde{\eta}_j. \tag{2.4.29}$$

It is seen that r_{ij} is of a smaller order than that of δ_{ij}. The remainder term can be shown to have a negligible effect over Q_n in terms of the variance. See Theorem 2.4.3.

Let X be the symmetric matrix where all diagonal entries are 0 and for $1 \leq i, j \leq n$ but $i \neq j$, $X_{ij} = A_{ij} - \widehat{\eta}_i\widehat{\eta}_j$, or equivalently,

$$X_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}. \tag{2.4.30}$$

If we omit the remainder term, then we have a proxy of X, denoted by X^*, where all diagonal entries of X^* are 0, and for $1 \leq i, j \leq n$ but $i \neq j$,

$$X_{ij}^* = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}. \tag{2.4.31}$$

If we further omit the δ term, then we have another proxy of X, denoted by $\widetilde{X}$, where all diagonal entries of $\widetilde{X}$ are 0, and for $1 \leq i, j \leq n$ but $i \neq j$,

$$\widetilde{X}_{ij} = \widetilde{\Omega}_{ij} + W_{ij}. \tag{2.4.32}$$

With the above notations, we can rewrite Q_n as follows:

$$Q_n = \sum_{i_1,i_2,i_3,i_4(dist)} X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}X_{i_4i_1}.$$

For the Ideal Signed Polygon in (2.1.11), we have the *Ideal SgnQ test statistic*

$$\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{X}_{i_1i_2}\widetilde{X}_{i_2i_3}\widetilde{X}_{i_3i_4}\widetilde{X}_{i_4i_1}. \tag{2.4.33}$$

The Ideal SgnQ test statistics can be viewed as proxies of the SgnQ test statistics, respectively, but such proxies are frequently not accurate enough. Therefore, we introduce another proxy for SgnQ, which we call the Proxy SgnQ test statistics, respectively. Recall that $X^*_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}$.

Definition 4. *The Proxy SgnQ test statistic is*

$$Q^*_n = \sum_{i_1,i_2,i_3,i_4(dist)} X^*_{i_1i_2}X^*_{i_2i_3}X^*_{i_3i_4}X^*_{i_4i_1}.$$

By these notations, we can partition SgnQ by

$$Q_n = \widetilde{Q}_n + (Q^*_n - \widetilde{Q}_n) + (Q_n - Q^*_n).$$

Below, first in Section 2.4.1, we analyze the Ideal SgnQ test statistics. Then in Section 2.4.2, we analyze the difference between the Ideal SgnQ and the Proxy SgnQ. Last, in Section 2.4.3, we analyze the difference between the Proxy SgnQ and the real SgnQ.

2.4.1 The behavior of the Ideal SgnQ test statistics

Recall the Ideal SgnQ test statistic is defined as

$$\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{X}_{i_1i_2}\widetilde{X}_{i_2i_3}\widetilde{X}_{i_3i_4}\widetilde{X}_{i_4i_1}, \tag{2.4.34}$$

where for any $i \neq j$, $\widetilde{X}_{ij} = \widetilde{\Omega}_{ij} + W_{ij}$. Under the null, since $\widetilde{\Omega}$ is a zero matrix, the statistic reduces to

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

Similarly, it can be shown that the statistic is asymptotically normal, with

$$\mathbb{E}[\widetilde{Q}_n] = 0, \qquad \text{and} \qquad \mathrm{Var}(\widetilde{Q}_n) \sim 8\|\theta\|^8.$$

Under the alternative, similarly, we obtain

$$2 \times 2 \times 2 \times 2 = 16$$

post-expansion sums, and divide them into 6 different types, according to $(N_{\widetilde{\Omega}}, N_W)$. See Table 3.2, where we recall $\alpha = |\lambda_2|/\lambda_1$.

From the table, among all 16 post-expansion sums, the total mean is

$$\sim \mathrm{tr}(\widetilde{\Omega}^4),$$

with Type V sum being the only contributor, and the total variance

$$\leq C\|\theta\|^8 + C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6,$$

Table 2.1: The 6 different types of the 16 post-expansion sums of $\widetilde{Q}_n$.

Type	#	$(N_{\widetilde{\Omega}}, N_W)$	Examples	Mean	Variance
I	1	(0, 4)	$\sum_{i,j,k,\ell(dist)} W_{ij}W_{jk}W_{k\ell}W_{\ell i}$	0	$\asymp \|\theta\|^8$
II	4	(1, 3)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}W_{k\ell}W_{\ell i}$	0	$\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
IIIa	4	(2, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i}$	0	$\leq C\alpha^4\|\theta\|^6\|\theta\|_3^6 = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
IIIb	2	(2, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}$	0	$\leq C\alpha^4\|\theta\|_3^{12} = o(\|\theta\|^8)$
IV	4	(3, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}$	0	$\leq \alpha^6\|\theta\|^8\|\theta\|_3^6$
V	1	(4, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}$	$\sim \mathrm{tr}(\widetilde{\Omega}^4)$	0

with Type I sum and Type IV sum being the major contributors. The following theorem is proved in the supplementary material.

Theorem 2.4.1 (Ideal SgnQ test statistic). *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*, where the condition* (2.2.14) *is satisfied under the alternative hypothesis. Suppose $\theta_{\max} \to 0$ and $\|\theta\| \to \infty$ as $n \to \infty$, and suppose $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ under the alternative hypothesis. Then, under the null hypothesis, as $n \to \infty$,*

$$\mathbb{E}[\widetilde{Q}_n] = 0, \qquad \mathrm{Var}(\widetilde{Q}_n) = 8\|\theta\|^8 \cdot [1 + o(1)],$$

and

$$\frac{\widetilde{Q}_n - \mathbb{E}[\widetilde{Q}_n]}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}} \longrightarrow N(0,1), \qquad \textit{in law.}$$

Furthermore, under the alternative hypothesis, as $n \to \infty$,

$$\mathbb{E}[\widetilde{Q}_n] = \mathrm{tr}(\widetilde{\Omega}^4) + o(\|\theta\|^4), \qquad \mathrm{Var}(\widetilde{T}_n) \leq C[\|\theta\|^8 + (|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6].$$

2.4.2 The behavior of $(Q_n^* - \widetilde{Q}_n)$

Consider $(Q_n^* - \widetilde{Q}_n)$, which is defined as

$$\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} X^*_{i_1i_2}X^*_{i_2i_3}X^*_{i_3i_4}X^*_{i_4i_1}. \tag{2.4.35}$$

Similarly, if we expand the bracket of all individual terms and re-organize, we have

$$3 \times 3 \times 3 \times 3 = 81$$

post-expansion sums. Out of the 81 post-expansion sums, $2 \times 2 \times 2 \times 2 = 16$ of them do not depend on δ, the sum of which equals to $\widetilde{Q}_n$. These leave us with 65 post-expansion sums, the total sum of which is $(Q_n^* - \widetilde{Q}_n)$. Similarly, according to $(N_{\widetilde{\Omega}}, N_W, N_\delta)$, we divide these 65 sums into 10 types. See Table 3.3, where we recall that $\alpha = |\lambda_2|/\lambda_1$.

Consider the null hypothesis first. Under the null, $\widetilde{\Omega}$ is a zero matrix, so the nonzero post-expansion sums only include Type Ia, Type IIa, Type IIIa, and Type IV. It is seen that

$$|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| \leq C\|\theta\|^4,$$

and that

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) = o(\|\theta\|^8).$$

Note that $\|\theta\|^8$ is the order of $\mathrm{Var}(\widetilde{Q}_n)$ under the null. The difference between the variance of Q_n^* and the variance of $\widetilde{Q}_n$ is negligible, but the difference between the mean of Q_n^* and the

Table 2.2: The 10 types of the post-expansion sums for $(Q_n^* - \widetilde{Q}_n)$. .

Type	#	$(N_\delta, N_{\widetilde{\Omega}}, N_W)$	Examples	Abs. Mean	Variance
Ia	4	(1, 0, 3)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} W_{k\ell} W_{\ell i}$	0	$\leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
Ib	8	(1, 1, 2)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} W_{\ell i}$	0	$\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	0	$\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
Ic	8	(1, 2, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	$\leq C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} \widetilde{\Omega}_{\ell i}$	0	$\leq \frac{C\alpha^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1} = o(\|\theta\|^8)$
Id	4	(1, 3, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	0	$\leq \frac{C\alpha^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1} = O(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
IIa	4	(2, 0, 2)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} W_{k\ell} W_{\ell i}$	$\leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
	2		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} \delta_{k\ell} W_{\ell i}$	$\leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIb	8	(2, 1, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	0	$\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} W_{\ell i}$	$\leq C\alpha\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIc	4	(2, 2, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	$\leq C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\alpha^4\|\theta\|^{14}}{\|\theta\|_1^2} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
	2		$\leq \sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	$\frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2} = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8)$
IIIa	4	(3, 0, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} W_{\ell i}$	$\leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIIb	4	(3, 1, 0)	$\leq \sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	$\leq \frac{C\alpha\|\theta\|^6}{\|\theta\|_1^3} = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IV	1	(4, 0, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} \delta_{\ell i}$	$\leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^{10}}{\|\theta\|_1^2} = o(\|\theta\|^8)$

mean of $\widetilde{Q}_n$ is non-negligible. With lengthy calculations (see the supplementary material), we can show that

$$\mathbb{E}[Q_n^* - \widetilde{Q}_n] \sim 2\|\theta\|^4.$$

Therefore, $(Q_n^* - 2\|\theta\|^4)$ and $\widetilde{Q}_n$ have a negligible difference under the null.

Consider the alternative hypothesis next. From Table 3.3,

$$|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| \leq C(|\lambda_2|/\lambda_1)^2\|\theta\|^6,$$

where the major contribution is from Type Ic and Type IIc post-expansion sums. Under our assumptions for the alternative, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ and $\lambda_1 \asymp \|\theta\|^4$. It is easy to see that $|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| = o(\lambda_2^4)$, where λ_2^4 is the order of $\mathrm{tr}(\widetilde{\Omega}^4)$ and $\mathbb{E}[\widetilde{Q}_n]$; see Lemma 2 and Theorem 2.4.1. Additionally, $\|\theta\|^4 = O(\lambda_1^2) = o(\lambda_2^4)$, which is also of a smaller order of $\mathbb{E}[\widetilde{Q}_n]$. We conclude that

$$\big|\mathbb{E}[Q_n^* - \widetilde{Q}_n - 2\|\theta\|^4]\big| = o(\mathbb{E}[\widetilde{Q}_n]).$$

From the table,

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq \frac{C(|\lambda_2|/\lambda_1)^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1} + o(\|\theta\|^8),$$

with the major contribution from Type Id. Here, the second term is smaller than $\mathrm{Var}(\widetilde{Q}_n)$, and the first term is upper bounded by (using the universal inequality of $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$)

$$C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6,$$

which has a comparable order as $\mathrm{Var}(\widetilde{Q}_n)$. It follows that

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n - 2\|\theta\|^4) = \mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C\mathrm{Var}(\widetilde{Q}_n).$$

Combining the above, we obtain that the SNR of $(Q_n^* - 2\|\theta\|^4)$ and $\widetilde{Q}_n$ are at the same order.

These results are summarized in the following theorem, which is proved in the supplementary material.

Theorem 2.4.2 (Proxy SgnQ test statistic). *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*, where the condition* (2.2.14) *is satisfied under the alternative hypothesis. Suppose $\theta_{\max} \to 0$ and $\|\theta\| \to \infty$ as $n \to \infty$, and suppose $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ under the alternative hypothesis. Then, under the null hypothesis, as $n \to \infty$,*

$$\mathbb{E}[(Q_n^* - 2\|\theta\|^4) - \widetilde{Q}_n] = o(\|\theta\|^4), \qquad \mathrm{Var}(Q_n^* - \widetilde{Q}_n) = o(\|\theta\|^8).$$

Furthermore, under the alternative hypothesis,

$$\mathbb{E}[(Q_n^* - 2\|\theta\|^4) - \widetilde{Q}_n] = o((|\lambda_2|/\lambda_1)^4\|\theta\|^8),$$
$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8).$$

2.4.3 The behavior of $(Q_n - Q_n^*)$

The SgnQ statistic we introduce in Section 2.1.4 is defined as

$$Q_n = \sum_{i_1,i_2,i_3,i_4(dist)} X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}X_{i_4i_1},$$

where $X_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}$ for any $i \neq j$. Similar to Sections 2.4.1-2.4.2, we first expand every bracket in the definitions and obtain $4 \times 4 \times 4 \times 4 = 256$ different post-expansion sums in Q_n. Out of the 256 post-expansion sums in Q_n, $3 \times 3 \times 3 \times 3 = 81$ of them do not involve any r term and are contained in Q_n^*; this leaves a total of

$$256 - 81 = 175$$

different post-expansion sums in $(Q_n - Q_n^*)$. In the appendix, we investigate the order of mean and variance of each of 175 post-expansion sums in $(Q_n - Q_n^*)$. The calculations are very tedious: although we expect these post-expansion sums to be of a smaller order than the post-expansion sums in Sections 2.4.1-2.4.2, it is impossible to prove this argument rigorously using only some crude bounds (such as Cauchy-Schwarz inequality). Instead, we still need to do calculations for each post-expansion sum.

Theorem 2.4.3 (Real SgnQ test statistic). *Consider the testing problem* (2.1.6) *under the DCMM model* (3.1.1)-(3.2.4)*, where the condition* (2.2.14) *is satisfied under the alternative hypothesis. Suppose $\theta_{\max} \to 0$ and $\|\theta\| \to \infty$ as $n \to \infty$, and suppose $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ under the alternative hypothesis. Then, under the null hypothesis, as $n \to \infty$,*

$$|\mathbb{E}[Q_n - Q_n^*]| = o(\|\theta\|^4), \qquad \textit{and} \qquad \mathrm{Var}(Q_n - Q_n^*) = o(|\theta\|^8).$$

Under the alternative hypothesis, as $n \to \infty$,

$$|\mathbb{E}[Q_n - Q_n^*]| = o((|\lambda_2|/\lambda_1)^4 \|\theta\|^8),$$
$$\mathrm{Var}(Q_n - Q_n^*) = o((|\lambda_2|/\lambda_1)^6 \|\theta\|^8 |\theta\|_3^6) + o(\|\theta\|^8).$$

Similarly, we can conclude that $(Q_n - Q_n^*)$ has a negligible effect to both the asymptotic normality under the null and the SNR under the alternative.

2.4.4 Proof of the main theorems

Consider Theorem 2.2.1. In this theorem, we assume the null is true. First, by Theorems 2.4.2 and 2.4.3 and elementary statistics,

$$\mathbb{E}[Q_n^* - \widetilde{Q}_n] \sim 2\|\theta\|^4, \qquad |\mathbb{E}[Q_n - Q_n^*]| = o(\|\theta\|^4), \tag{2.4.36}$$

and

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) = o(\|\theta\|^8), \qquad \mathrm{Var}(Q_n - Q_n^*) = o(\|\theta\|^8). \tag{2.4.37}$$

It follows that

$$\mathbb{E}[Q_n] - \mathbb{E}[\widetilde{Q}_n] = (2 + o(1))\|\theta\|^4, \qquad \mathrm{Var}(Q_n - \widetilde{Q}_n) = o(\|\theta\|^8). \tag{2.4.38}$$

By Theorem 2.4.1.

$$\mathbb{E}[\widetilde{Q}_n] = o(\|\theta\|^4), \quad \mathrm{Var}(\widetilde{Q}_n) \sim 8\|\theta\|^8, \quad \frac{\widetilde{Q}_n - \mathbb{E}[\widetilde{Q}_n]}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}} \to N(0,1). \tag{2.4.39}$$

Since for any random variables X and Y, $\mathrm{Var}(X+Y) \leq (1+a_n)\mathrm{Var}(X) + (1+\frac{1}{a_n})\mathrm{Var}(Y)$ for any number $a_n > 0$, combining the above and letting a_n tend to 0 appropriately slow,

$$\mathbb{E}[Q_n] \sim 2\|\theta\|^4, \qquad \mathrm{Var}(Q_n) \sim 8\|\theta\|^8. \tag{2.4.40}$$

Moreover, write

$$\frac{Q_n - \mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} = \sqrt{\frac{\mathrm{Var}(\widetilde{Q}_n)}{\mathrm{Var}(Q_n)}} \cdot \left[\frac{(Q_n - \widetilde{Q}_n)}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}} + \frac{\widetilde{Q}_n - \mathbb{E}[\widetilde{Q}_n]}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}} + \frac{\mathbb{E}[\widetilde{Q}_n] - \mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}}\right].$$

On the right hand side, by (2.4.38)-(2.4.40), as $n \to \infty$, the term outside the bracket $\to 1$, and for the three terms in the bracket, the first one has a mean and variance that tend to 0 so it tends to 0 in probability, the second one weakly converges to $N(0,1)$, and the last one $\to 0$. Combining these,

$$\frac{Q_n - \mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \to N(0,1), \qquad \text{in law.} \tag{2.4.41}$$

Combining (2.4.40) and (2.4.41) proves Theorem 2.2.1.

Next, we consider Theorem 2.2.2, where we assume the alternative is true. First, similarly, by Theorems 2.4.2 and 2.4.3,

$$\mathbb{E}[Q_n^* - \widetilde{Q}_n] = (2 + o(1))\|\theta\|^4 + o((|\lambda_2|/\lambda_1)^4 \|\theta\|^8),$$

and

$$\mathrm{Var}(Q_n - \widetilde{Q}_n) \leq C(\lambda_2/\lambda_1)^6 \|\theta\|^8 \|\theta\|_3^6 + o(\|\theta\|^8).$$

Second, by Theorems 2.4.1,

$$\mathbb{E}[\widetilde{Q}_n] = \mathrm{tr}(\widetilde{\Omega}^4) + o(\|\theta\|^4), \qquad \mathrm{Var}(\widetilde{Q}_n) \leq C[\|\theta\|^8 + (\lambda_2/\lambda_1)^6 \|\theta\|^8 \|\theta\|_3^6].$$

Combining these proves Theorem 2.2.2.

Last, we consider Theorems 2.2.3. First, by Theorem 2.2.1 and Lemma 1, under the null,

$$\frac{Q_n - 2(\|\widehat{\eta}\|^2 - 1)^2}{\sqrt{8(\|\widehat{\eta}\|^2 - 1)^4}} \to N(0,1),$$

so the Type I error is

$$\mathbb{P}_{H_0^{(n)}}\left(Q_n \geq (2 + z_\alpha\sqrt{8})(|\widehat{\eta}\|^2 - 1)^2\right) = P\left(\frac{Q_n - 2(\|\widehat{\eta}\|^2 - 1)^2}{\sqrt{8(\|\widehat{\eta}\|^2 - 1)^4}} \geq z_\alpha\right) = \alpha + o(1).$$

Second, fixing $0 < \varepsilon < 1$, let A_ε be the event $\{(\|\widehat{\eta}\|^2 - 1) \leq (1+\varepsilon)\|\eta^*\|^2\}$. By Lemma 1 and definitions, on one hand, over the event A_ε, $(\|\widehat{\eta}\|^2 - 1) \leq (1+\varepsilon)\|\eta^*\|^2 \leq C\|\theta\|^2$, and on the other hand, $\mathbb{P}(A_\varepsilon^c) = o(1)$. Therefore, the Type II error

$$\begin{aligned}
&\mathbb{P}_{H_1^{(n)}}\left(Q_n \leq (2 + z_\alpha\sqrt{8})(\|\widehat{\eta}\|^2 - 1)^2\right) \\
\leq &\mathbb{P}_{H_1^{(n)}}\left(Q_n \leq (2 + z_\alpha\sqrt{8})(\|\widehat{\eta}\|^2 - 1)^2, A_\varepsilon\right) + \mathbb{P}(A_\varepsilon^c) \\
\leq &\mathbb{P}_{H_1^{(n)}}\left(Q_n \leq C(2 + z_\alpha\sqrt{8})\|\theta\|^4\right) + o(1),
\end{aligned}$$

where by Chebyshev's inequality, the first term in the last line

$$\leq [\mathbb{E}(Q_n) - C(2 + z_\alpha\sqrt{8})\|\theta\|^4]^{-2} \cdot \mathrm{Var}(Q_n). \tag{2.4.42}$$

By Lemma D.2 of the supplementary material and our assumptions, $\lambda_1 \asymp \|\theta\|^2$, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$, and $\|\theta\| \to \infty$. Using Lemma 2 $\mathbb{E}[Q_n] \geq C\lambda_2^4 \gg \lambda_1^2$, and it follows that $\mathbb{E}(Q_n) \gg C(2 + z_\alpha\sqrt{8})\|\theta\|^4$, so for sufficiently large n,

$$\mathbb{E}(Q_n) - C(2 + z_\alpha\sqrt{8})\|\theta\|^4 \geq \frac{1}{2}\mathbb{E}[Q_n] \geq C\lambda_2^4.$$

At the same time, by Theorem 2.2.2,

$$\mathrm{Var}(Q_n) \leq C(\|\theta\|^8 + (\lambda_2/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6).$$

Combining these, the right hand side of (2.4.42) does not exceed

$$C\frac{\|\theta\|^8 + (\lambda_2/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6}{\lambda_2^8} = (I) + (II), \tag{2.4.43}$$

where $(I) = C\lambda_2^{-8}\|\theta\|^8$ and $(II) = C\lambda_2^{-8}(\lambda_2/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6$. Now, first, since $\lambda_1 \asymp \|\theta\|^2$ and $|\lambda_2|/\sqrt{\lambda_1} \to 0$, $(I) \leq C(\lambda_2/\sqrt{\lambda_1})^{-8} \to 0$. Second, since $\lambda_1 \asymp \|\theta\|^2$ and $\|\theta\|_3^6 \leq \|\theta\|^4$, $(II) = C\lambda_2^{-2}\lambda_1^{-6}\|\theta\|^8\|\theta\|_3^6 \leq C\lambda_2^{-2}$. As $|\lambda_2|/\sqrt{\lambda_1} \to \infty$, $\sqrt{\lambda_1} \asymp \|\theta\|$ with $\|\theta\| \to \infty$, $|\lambda_2| \to \infty$ and $(II) \to 0$. Inserting these into (2.4.43), the Type II error $\to 0$ and the claim follows. □

2.5 Simulations

We investigate the numerical performance of the SgnQ test (2.2.16). We include the SgnT test Jin et al. (2019), which is known for suffering from "signal cancellation" and is only optimal adaptive under additional mild conditions. We also include the EZ test Gao and Lafferty (2017) and the GC test Jin et al. (2018) for comparison. For reasons mentioned in Jin et al. (2018), we use a two-sided rejection region for EZ and a one-sided rejection region for GC.

Given (n, K), a scalar $\beta_n > 0$ that controls $\|\theta\|$, a symmetric nonnegative matrix

$P \in \mathbb{R}^{K\times K}$, a distribution $f(\theta)$ on $\mathbb{R}_+$, and a distribution $g(\pi)$ on the standard simplex of $\mathbb{R}^K$, we generate two network adjacency matrices A^{null} and A^{alt}, under the null and the alternative, respectively, as follows:

- Generate $\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n$ *iid* from $f(\theta)$. Let $\theta_i = \beta_n \cdot \widetilde{\theta}_i / \|\widetilde{\theta}\|$, $1 \le i \le n$.
- Generate $\pi_1, \pi_2, \ldots, \pi_n$ *iid* from $g(\pi)$.
- Let $\Omega^{alt} = \Theta\Pi P \Pi'\Theta'$, where $\Theta = \mathrm{diag}(\theta_1, \cdots, \theta_n)$ and $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$. Generate A^{alt} from Ω^{alt} according to Model (3.1.1).
- Let $\Omega^{null} = (a'Pa)\cdot\theta\theta'$, where $a = \mathbb{E}_g\pi \in \mathbb{R}^K$ is the mean vector of $g(\pi)$. Generate A^{null} from Ω^{null} according to Model (3.1.1).

The pair $(\Omega^{null}, \Omega^{alt})$ is constructed in a way such that the corresponding networks have approximately the same expected average degree. This is the most subtle case for distinguishing two hypotheses (see Section 2.3).

It is of interest to explore different sparsity levels and also to focus on the parameter settings where the SNR is neither too large or too small. Therefore, for most of the experiments, we let $\beta_n = \|\theta\|$ range but fix the SNR at a more or less the same level. See details below.

For each parameter setting, we generate 200 networks under the null hypothesis and 200 networks under the alternative hypothesis, run all the four tests with a targeting level $\alpha = 5\%$, and then record the sum of percent of type I errors and percent of type II errors.

We consider three experiments (and a total of 8 sub-experiments), exploring different sets of n, K, θ, Π, and P, etc.

Experiment 1: We study the role of degree heterogeneity. Fix $(n, K) = (2000, 2)$. Let P be a 2×2 matrix with unit diagonal entries and all off-diagonal entries equal to b_n. Let $g(\pi)$ be the uniform distribution on $\{(0,1), (1,0)\}$. We consider three sub-experiments, Exp 1a-1c, where respectively we take $f(\theta)$ to be the following:

- Uniform distribution $U(2,3)$.
- Two-point distribution $0.95\delta_1 + 0.05\delta_3$, where δ_a is a point mass at a.
- Pareto distribution $\mathrm{Pareto}(10, 0.375)$, where 10 is the shape parameter and 0.375 is the scale parameter.

The degree heterogeneity is moderate in the Exp 1a-1b, but more severe Exp 1c. In such a setting, SNR is at the order of $\|\theta\|(1-b_n)$. Therefore, for each sub-experiment, we let $\beta_n = \|\theta\|$ vary while fixing the SNR to be

$$\|\theta\|(1 - b_n) = 3.2.$$

The sum of Type I and Type II errors are displayed in Figure 2.2.

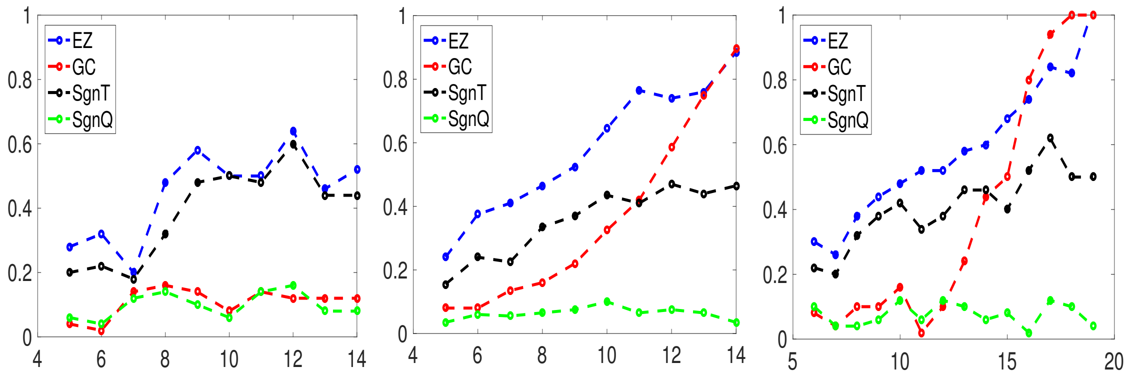

Figure 2.2: From left to right: Experiment 1a, 1b, and 1c. The y-axis are the sum of Type I and Type II errors (testing level is fixed at 5%). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on 200 repetitions.

First, both the SgnQ test and the GC test are based on the counts of 4-cycles, but the *GC* test counts *non-centered* cycles and the SgnQ test counts *centered* cycles. As we pointed out in Section 2.1, counting *centered* cycles may have much smaller variances than counting *non-centered* cycles, especially in the less sparse case, and thus improves the testing power. This is confirmed by numerical results here, where the SgnQ test is consistently better than the GC test, significantly so in the less sparse case. Similarly, both the SgnT test and the EZ test are based on the counts of 3-cycles, but the *EZ* test counts *non-centered* cycles and the SgnQ test counts *centered* cycles, and we expect the GC to significantly improve the EZ, especially in the less sparse case. This is also confirmed in the experiment.

Second, SgnQ and GC are order-4 test graphlet counting statistics, and SgnT and EZ are order-3 graphlet counting statistics. In comparison, SgnQ significantly outperforms SgnT, and GC significantly outperforms EZ (in the more sparse case; see discussion for the less sparse case below). A possible explanation is that higher-order graphlet counting statistics have larger SNR. Investigation on this is interesting, and we leave this to the future study. Note that SgnQ is the best among all four tests.

Last, GC outperforms EZ in the more sparse case, but underperforms in the less sparse case. The reason for the latter is as follows. The biases of both tests are negligible in the more sparse case, but are non-negligible in the less sparse case, with that of GC is much larger.

Experiment 2: We study the cases with larger K and more complicate matrix P. For a $b_n \in (0,1)$, let $\varepsilon_n = \frac{1}{6}\min(1-b_n, b_n)$, and let P be the matrix with 1 on the diagonal but the off-diagonal entries are iid drawn from Unif$(b_n - \varepsilon_n, b_n + \varepsilon_n)$; once the P matrix is drawn, it is fixed throughout different repetitions. We consider two sub-experiments, Exp 2a and 2b. In Exp 2a, we take $(n, K) = (1000, 5)$, $f(\theta)$ to be Pareto$(10, 0.375)$, and $g(\pi)$ to be the uniform distribution on $\{e_1, \cdots, e_K\}$ (the standard basis vectors of $\mathbb{R}^K$). We let β_n range but $\|\theta\|(1-b_n)$ is fixed at 4.5, so the SNR will not change drastically. In Exp 2b, we take $(n, K) = (3000, 10)$, $f(\theta)$ to be $0.95\delta_1 + 0.05\delta_3$, and $g(\pi) = 0.1\sum_{k=1}^{2}\delta_{e_k} + 0.15\sum_{k=3}^{6}\delta_{e_k} + 0.05\sum_{k=7}^{10}\delta_{e_k}$ (so to have unbalanced community sizes). Similarly, we let β_n range but fix $\|\theta\|(1-b_n) = 5.2$.

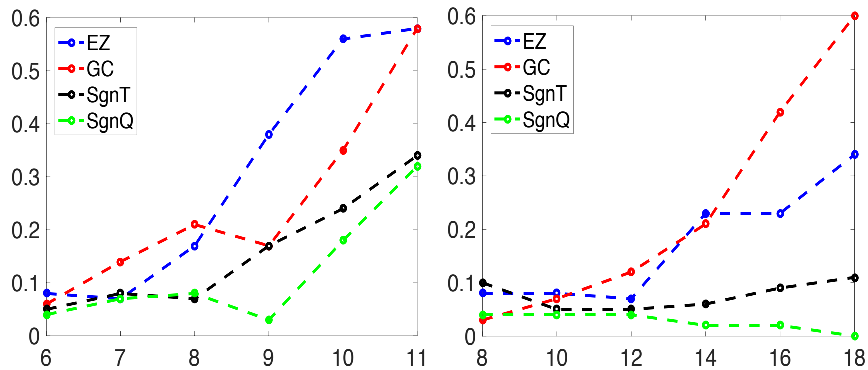

Figure 2.3: From left to right: Experiment 2a and 2b. The y-axis are the sum of Type I and Type II errors (testing level is fixed at 5%). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on 200 repetitions.

The sum of Type I and Type II errors are displayed in Figure 3.5.

In these examples, EZ and GC underperform SgnT and SgnQ, especially in the less sparse case, and the performances of the SgnT and SgnQ are more similar to each other, compared to those in Experiment 1. In these examples, we have larger K, more complicate P, and unbalanced community sizes, and the performance of the SgnT and SgnQ test statistics suggest that they are relatively robust.

Experiment 3: We investigate the role of mixed-membership. We have three sub-experiments, Exp 3a-3c. where the memberships are not-mixed, lightly mixed, and significantly mixed, respectively. For all sub-experiments, we take $(n, K) = (2000, 3)$ and $f(\theta)$ to be Unif$(2, 3)$. For Exp 3a, we let $g_1(\pi) = 0.4\delta_{e_1} + 0.3\delta_{e_2} + 0.3\delta_{e_3}$. In Exp 3b, we let $g_2(\pi) = 0.3\sum_{k=1}^{3}\delta_{e_k} + 0.1\cdot$ Dirichlet, and in Exp 3c, we let $g_3(\pi) = 0.25\sum_{k=1}^{3}\delta_{e_k} + 0.25\cdot$ Dirichlet, where Dirichlet represents the symmetric K-dimensional Dirichlet distribution. In Exp 3a-3b, we let β_n range while $(1-b_n)\|\theta\|$ is fixed at 4.2 so the SNR's are roughly the same. In Exp 3c, we also let β_n range but $(1-b_n)\|\theta\| = 4.5$ (the SNR's need to be slightly larger to counter the effect of mixed-membership, which makes the testing problem harder).

The sum of Type I and Type II errors are presented in Figure 3.6. First, the results confirm that mixed-memberships make the testing problem harder. For example, the value of $\|\theta\|(1-b_n)$ in Exp 3c is higher than that of Exp 3a-3b, but the testing errors are higher, due to that the memberships in Exp 3c are more mixed. Second, SgnQ consistently outperforms EZ and SgnT. Third, GC is comparable with SgnQ in the more sparse case, but performs unsatisfactorily in the less sparse case, for reasons explained before. Last, in these settings, SgnT is uniformly better than EZ, and more so when the memberships become more mixed.

2.A Matrix forms of Signed-Polygon statistics

We prove Theorem 2.1.1. Recall that $\widetilde{A} = A - \widehat{\eta}\widehat{\eta}$. By definition,

$$Q_n = \mathrm{tr}(\widetilde{A}^4) - \sum_{\substack{\text{at least two of} \\ i,j,k,\ell \text{ are equal}}} \widetilde{A}_{ij}\widetilde{A}_{jk}\widetilde{A}_{k\ell}\widetilde{A}_{\ell i}.$$

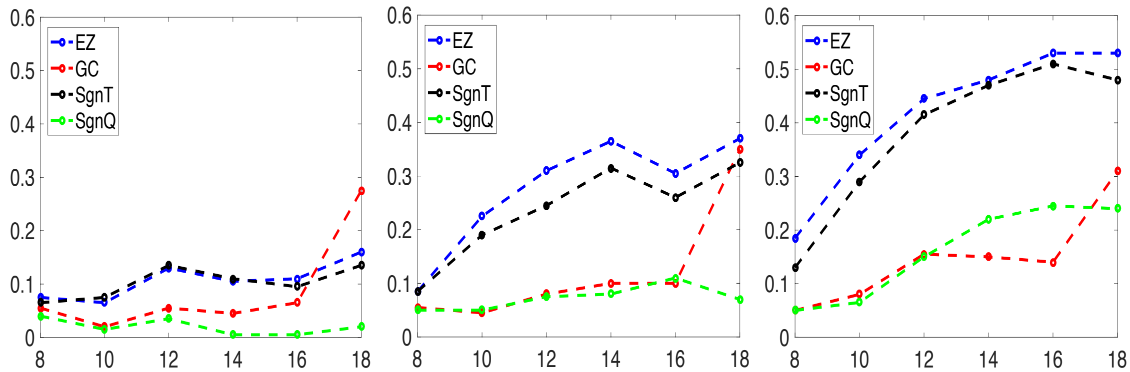

Figure 2.4: From left to right: Experiment 3a, 3b, and 3c. The y-axis are the sum of Type I and Type II errors (testing level is fixed at 5%). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on 200 repetitions.

When at least two of $\{i, j, k, \ell\}$ are equal, depending on how many indices are equal, we have four patterns: $\{i, i, i, i\}$, $\{i, i, i, j\}$, $\{i, i, j, j\}$, $\{i, i, j, k\}$, where (i, j, k) are distinct. For each pattern, depending on the appearing locations of the next distinct indices, there are a few variations. Take the pattern $\{i, i, j, k\}$ for example: (a) when a new distinct index appears at location 2 and at location 3, the variations are (i, j, k, i), (i, j, k, j), (i, j, k, k); (b) when a new distinct index appears at location 2 and at location 4, the variations are (i, j, i, k), (i, j, j, k); (c) when a new distinct index appears at location 3 and location 4, the variation is (i, i, j, k). Using similar arguments, we can find all variations of each pattern. Define

$$S_1 = \sum_{i,j,k(dist)} \widetilde{A}_{ii}\widetilde{A}_{ij}\widetilde{A}_{jk}\widetilde{A}_{ki}, \qquad S_2 = \sum_{i,j,k(dist)} \widetilde{A}_{ij}^2\widetilde{A}_{ik}^2,$$
$$S_3 = \sum_{i,j(dist)} \widetilde{A}_{ii}^2\widetilde{A}_{ij}^2, \qquad S_4 = \sum_{i,j(dist)} \widetilde{A}_{ij}^4,$$
$$S_5 = \sum_{i,j(dist)} \widetilde{A}_{ii}\widetilde{A}_{ij}^2\widetilde{A}_{jj}, \qquad S_6 = \sum_{i} \widetilde{A}_{ii}^4.$$

Therefore

$$Q_n = \mathrm{tr}(\widetilde{A}^4) - 4S_1 - 2S_2 - 4S_3 - S_4 - 2S_5 - S_6. \tag{2.A.44}$$

What remains is to derive the matrix form of S_1-S_6. By direct calculations,

$$\begin{aligned}
S_1 &= \sum_i \widetilde{A}_{ii}\Big[\sum_{j\neq i, k\neq i} \widetilde{A}_{ij}\widetilde{A}_{jk}\widetilde{A}_{ki} - \sum_{j\neq i}\widetilde{A}_{ij}\widetilde{A}_{jj}\widetilde{A}_{ji}\Big] \\
&= \sum_i \widetilde{A}_{ii}\Big[\Big(\sum_{j,k}\widetilde{A}_{ij}\widetilde{A}_{jk}\widetilde{A}_{ki} - 2\sum_j \widetilde{A}_{ij}^2\widetilde{A}_{ii} + \widetilde{A}_{ii}^3\Big) - \Big(\sum_j \widetilde{A}_{ij}^2\widetilde{A}_{jj} - \widetilde{A}_{ii}^3\Big)\Big] \\
&= \sum_{i,j,k}\widetilde{A}_{ii}\widetilde{A}_{ij}\widetilde{A}_{jk}\widetilde{A}_{ki} - 2\sum_{i,j}\widetilde{A}_{ii}^2\widetilde{A}_{ij}^2 - \sum_{i,j}\widetilde{A}_{ii}\widetilde{A}_{ij}^2\widetilde{A}_{jj} + 2\sum_i \widetilde{A}_{ii}^4 \\
&= \mathrm{tr}(\widetilde{A}\circ\widetilde{A}^3) - 2\,\mathrm{tr}(\widetilde{A}\circ\widetilde{A}\circ\widetilde{A}^2) - 1_n'[\mathrm{d}iag(\widetilde{A})(\widetilde{A}\circ\widetilde{A})\mathrm{d}iag(\widetilde{A})]1_n + 2S_6.
\end{aligned}$$

Moreover, we can derive that

$$\begin{aligned}
S_2 &= \sum_i \Big[\sum_{j\neq i, k\neq i} \widetilde{A}_{ij}^2 \widetilde{A}_{ik}^2 - \sum_{j\neq i} \widetilde{A}_{ij}^4 \Big] \\
&= \sum_i \Big[\Big(\sum_{j,k} \widetilde{A}_{ij}^2 \widetilde{A}_{ik}^2 - 2\sum_j \widetilde{A}_{ij}^2 \widetilde{A}_{ii}^2 + \widetilde{A}_{ii}^4 \Big) - \Big(\sum_j \widetilde{A}_{ij}^4 - \widetilde{A}_{ii}^4 \Big) \Big] \\
&= \sum_{i,j,k} \widetilde{A}_{ij}^2 \widetilde{A}_{ik}^2 - 2\sum_{i,j} \widetilde{A}_{ij}^2 \widetilde{A}_{ii}^2 - \sum_{i,j} \widetilde{A}_{ij}^4 + 2\sum_i \widetilde{A}_{ii}^4 \\
&= \mathrm{tr}(\widetilde{A}^2 \circ \widetilde{A}^2) - 2\,\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A}^2) - 1_n'[\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}]1_n + 2S_6.
\end{aligned}$$

It is also easy to see that

$$\begin{aligned}
S_3 &= \sum_{i,j} \widetilde{A}_{ii}^2 \widetilde{A}_{ij}^2 - \sum_i \widetilde{A}_{ii}^4 = \mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A}^2) - S_6, \\
S_4 &= \sum_{i,j} \widetilde{A}_{ij}^4 - \sum_i \widetilde{A}_{ii}^4 = 1_n'[\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}]1_n - S_6, \\
S_5 &= \sum_{i,j} \widetilde{A}_{ii} \widetilde{A}_{ij}^2 \widetilde{A}_{jj} - S_6 = 1_n'[\mathrm{d}iag(\widetilde{A})(\widetilde{A} \circ \widetilde{A})\mathrm{d}iag(\widetilde{A})]1_n - S_6, \\
S_6 &= \mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}).
\end{aligned}$$

Plugging the matrix forms of S_1-S_6 into (2.A.44), we obtain

$$\begin{aligned}
Q_n =& \,\mathrm{tr}(\widetilde{A}^4) - 4\,\mathrm{tr}(\widetilde{A} \circ \widetilde{A}^3) - 2\,\mathrm{tr}(\widetilde{A}^2 \circ \widetilde{A}^2) + 8\,\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A}^2) - 6\,\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}) \\
&+ 2 \cdot 1_n'[\mathrm{d}iag(\widetilde{A})(\widetilde{A} \circ \widetilde{A})\mathrm{d}iag(\widetilde{A})]1_n + 1_n'[\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}]1_n.
\end{aligned}$$

This gives the desired expression of Q_n.

Last, we discuss the complexity of computing Q_n. It involves the following operations:

- Compute the matrix $\widetilde{A} = A - \widehat{\eta}\widehat{\eta}'$.
- Compute the Hadamard product of finitely many matrices.
- Compute the trace of a matrix.
- Compute the matrix DMD for a matrix M and a diagonal matrix D.
- Compute $1_n' M 1_n$ for a matrix M.
- Compute the matrices $\widetilde{A}^k$, for $k = 2, 3, 4$.

Excluding the last operation, the complexity is $O(n^2)$. For the last operation, since we can compute $\widetilde{A}^k$ recursively from $\widetilde{A}^k = \widetilde{A}^{k-1}\widetilde{A}$, it suffices to consider the complexity of computing $B\widetilde{A}$, for an arbitrary $n \times n$ matrix B. Write

$$B\widetilde{A} = BA - B\widehat{\eta}(\widehat{\eta})'.$$

Consider computing BA. The (i,j)-th entry of BA is $\sum_{\ell: A_{\ell j}\neq 0} B_{i\ell} A_{\ell j}$, where the total number of nonzero $A_{\ell j}$ equals to d_j, the degree of node j. Hence, the complexity of computing the (i,j)-th entry of BA is $O(d_j)$. It follows that the complexity of computing BA is $O(\sum_{i,j=1}^n d_j) = O(n^2 \bar{d})$. Consider computing $B\widehat{\eta}(\widehat{\eta})'$. We first compute the vector

$v = B\widehat{\eta}$ and then compute $v(\widehat{\eta})'$, where the complexity of both steps is $O(n^2)$. Combining the above, the complexity of computing $B\widetilde{A}$ is $O(n^2\bar{d})$. We have seen that this is the dominating step in computing T_n and Q_n, so the complexity of the latter is also $O(n^2\bar{d})$.

2.B Estimation of $\|\theta\|$

We prove Lemma 1. First, we show that

$$\|\eta^*\|^2 \begin{cases} = \|\theta\|^2, & \text{under the null,} \\ \asymp \|\theta\|^2, & \text{under the alternative.} \end{cases}$$

Recall that $\eta^* = (1/\sqrt{1_n'\Omega 1_n})\Omega 1_n$. Hence,

$$\|\eta^*\|^2 = (1_n'\Omega^2 1_n)/(1_n'\Omega 1_n). \tag{2.B.45}$$

Under the null, $\Omega = \theta\theta'$, and the claim follows by direct calculations. Under the alternative, $\Omega = \sum_{k=1}^K \lambda_k \xi_k \xi_k'$, so

$$1_n'\Omega 1_n = \sum_{k=1}^K \lambda_k (1_n'\xi_k)^2, \qquad 1_n'\Omega^2 1_n = \sum_{k=1}^K \lambda_k^2 (1_n'\xi_k)^2.$$

By Lemma 6, $\lambda_1 \asymp \|\theta\|^2$. By Lemma 7, $1_n'\xi_1 \asymp \|\theta\|^{-1}\|\theta\|_1$ and $|1_n'\xi_k| = O(\|\theta\|^{-1}\|\theta\|_1)$. It follows that $1_n'\Omega^2 1_n \geq \lambda_1^2 (1_n'\xi_1)^2 \geq C\|\theta\|_1^2\|\theta\|^2$ and $1_n'\Omega^2 1_n \leq \lambda_1^2 \sum_{k=1}^K (1_n'\xi_k)^2 \leq C\|\theta\|_1^2\|\theta\|^2$. We conclude that

$$1_n'\Omega^2 1_n \asymp \|\theta\|_1^2\|\theta\|^2. \tag{2.B.46}$$

Moreover, $1_n'\Omega 1_n \leq |\lambda_1| \sum_{k=1}^K (1_n'\xi_k)^2 \leq C\|\theta\|_1^2$, and by Lemma 8, $1_n'\Omega 1_n \geq C\|\theta\|_1^2$. It follows that

$$1_n'\Omega 1_n \asymp \|\theta\|_1^2. \tag{2.B.47}$$

Plugging (2.B.46)-(2.B.47) into (2.B.45) gives the claim.

Next, we show $(\|\widehat{\eta}\|^2 - 1)/\|\eta^*\|^2 \to 1$ in probability. Since $\|\eta^*\| \asymp \|\theta\| \to \infty$ as $n \to \infty$, it suffices to show $\|\widehat{\eta}\|^2/\|\eta^*\|^2 \to 1$ in probability. By definition,

$$\|\widehat{\eta}\|^2 = \frac{1_n'A^2 1_n}{1_n'A 1_n}.$$

Compare this with (2.B.45), all we need to show is that in probability,

$$\frac{1_n'A 1_n}{1_n'\Omega 1_n} \to 1, \qquad \text{and} \qquad \frac{1_n'A^2 1_n}{1_n'\Omega^2 1_n} \to 1. \tag{2.B.48}$$

Since the proofs are similar, we only show the second one. By elementary probability, it is sufficient to show that as $n \to \infty$,

$$\frac{\mathbb{E}[1_n'A^2 1_n]}{1_n'\Omega^2 1_n} \to 1, \qquad \frac{\mathrm{Var}(1_n'A^2 1_n)}{(1_n'\Omega^2 1_n)^2} \to 0. \tag{2.B.49}$$

We now prove (2.B.49). Consider the first claim. Write

$$1_n'A^2 1_n = \sum_{i,j,k} A_{ij}A_{jk} = \sum_{i \neq j} A_{ij}^2 + \sum_{i,j,k(dist)} A_{ij}A_{jk}. \tag{2.B.50}$$

It follows that

$$\mathbb{E}[1_n'A^2 1_n] = \sum_{i \neq j} \Omega_{ij} + \sum_{i,j,k(dist)} \Omega_{ij}\Omega_{jk}.$$

Since $\Omega_{ij} \leq \theta_i\theta_j$ under both hypotheses, we have

$$\begin{aligned}\left|\mathbb{E}[1_n' A^2 1_n] - 1_n'\Omega 1_n - 1_n'\Omega^2 1_n\right| &\leq \Big|\sum_i \Omega_{ii} + \sum_{\substack{(i,j,k)\text{ are}\\ \text{not distinct}}} \Omega_{ij}\Omega_{jk}\Big| \\ &\leq \sum_i \theta_i^2 + C\sum_{i,j}\theta_i^2\theta_j^2 + C\sum_{i,k}\theta_i^3\theta_k \\ &\leq C\|\theta\|^2 + C\|\theta\|^4 + C\|\theta\|_3^3\|\theta\|_1 \\ &\leq C\|\theta\|_3^3\|\theta\|_1,\end{aligned}$$

where we have used the universal inequality $\|\theta\|^4 \leq \|\theta\|_3^3\|\theta\|_1$. Since $\|\theta\|_3^3 \leq \theta_{\max}^2\|\theta\|_1 = o(\|\theta\|_1)$, the right hand side is $o(\|\theta\|_1^2) = o(1_n'\Omega 1_n)$. So,

$$\mathbb{E}[1_n' A^2 1_n] = 1_n'\Omega^2 1_n + 1_n'\Omega 1_n + o(1_n'\Omega 1_n). \tag{2.B.51}$$

Combining this with (2.B.46)-(2.B.47) gives

$$\Big|\frac{\mathbb{E}[1_n' A^2 1_n]}{1_n'\Omega^2 1_n} - 1\Big| \lesssim \frac{1_n'\Omega 1_n}{1_n'\Omega^2 1_n} \asymp \frac{1}{\|\theta\|^2},$$

and the claim follows by $\|\theta\| \to \infty$.

Consider the second claim. By (2.B.50),

$$\mathrm{Var}(1_n' A^2 1_n) \leq 2\mathrm{Var}\Big(\sum_{i\neq j} A_{ij}^2\Big) + 2\mathrm{Var}\Big(\sum_{i,j,k(dist)} A_{ij}A_{jk}\Big). \tag{2.B.52}$$

We re-write $\sum_{i\neq j} A_{ij}^2 = \sum_{i\neq j} A_{ij} = 2\sum_{i<j} A_{ij}$. The variables $\{A_{ij}\}_{1\leq i<j\leq n}$ are mutually independent. It follows that

$$\mathrm{Var}\Big(\sum_{i\neq j} A_{ij}^2\Big) = 4\sum_{i<j}\mathrm{Var}(A_{ij}) \leq C\sum_{i,j}\Omega_{ij} \leq C\|\theta\|_1^2. \tag{2.B.53}$$

Moreover, since $A_{ij}A_{jk} = (\Omega_{ij} + W_{ij})(\Omega_{jk} + W_{jk})$, we have

$$\begin{aligned}\sum_{i,j,k(dist)} A_{ij}A_{jk} &= \sum_{i,j,k(dist)} \Omega_{ij}\Omega_{jk} + 2\sum_{i,j,k(dist)} \Omega_{ij}W_{jk} + \sum_{i,j,k(dist)} W_{ij}W_{jk} \\ &\equiv \sum_{i,j,k(dist)} \Omega_{ij}\Omega_{jk} + X_1 + X_2.\end{aligned}$$

By elementary probability,

$$\mathrm{Var}\Big(\sum_{i,j,k(dist)} A_{ij}A_{jk}\Big) \leq 2\mathrm{Var}(X_1) + 2\mathrm{Var}(X_2).$$

To compute the variance of X_1, we note that

$$X_1 = 4\sum_{j<k}\beta_{jk}W_{jk}, \qquad \beta_{jk} = \sum_{i\notin\{j,k\}} \Omega_{ij}.$$

The variables $\{W_{jk}\}_{1\leq j<k\neq n}$ are mutually independent, and $|\beta_{jk}| \leq C\sum_i \theta_i\theta_j \leq C\|\theta\|_1\theta_j$. It follows that

$$\mathrm{Var}(X_1) \leq C\sum_{j,k}(\|\theta\|_1\theta_j)^2(\theta_j\theta_k) \leq C\|\theta\|_1^3\|\theta\|_3^3.$$

To compute the variance of X_2, we note that

$$\mathrm{Var}(X_2) = \sum_{i,j,k(dist)}\ \sum_{i',j',k'(dist)} \mathbb{E}[W_{ij}W_{jk}W_{i'j'}W_{j'k'}].$$

The summand is nonzero only when the two variables $\{W_{i'j'}, W_{j'k'}\}$ are the same as the two variables $\{W_{ij}, W_{jk}\}$. This can only happen if $(i,j,k) = (i',j',k')$ or $(i,j,k) = (k',j',i')$, where in either case the summand equals to $\mathbb{E}[W_{ij}^2 W_{jk}^2]$. It follows that

$$\mathrm{Var}(X_2) = \sum_{i,j,k(dist)} 2\mathbb{E}[W_{ij}^2 W_{jk}^2] \leq C \sum_{i,j,k} \theta_i \theta_j^2 \theta_k \leq C\|\theta\|^2 \|\theta\|_1^2.$$

Combining the above gives

$$\mathrm{Var}\Big(\sum_{i,j,k(dist)} A_{ij} A_{jk} \Big) \leq C\|\theta\|_1^3 \|\theta\|_3^3 + C\|\theta\|^2 \|\theta\|_1^2 \leq C\|\theta\|_1^3 \|\theta\|_3^3, \tag{2.B.54}$$

where we have used the fact that $\|\theta\|_1 \|\theta\|_3^3 \geq \|\theta\|^4$ (Cauchy-Schwarz inequality) and $\|\theta\| \to \infty$. Plugging (2.B.53)-(2.B.54) into (2.B.52) gives

$$\mathrm{Var}(1_n' A^2 1_n) \leq C\|\theta\|_1^3 \|\theta\|_3^3. \tag{2.B.55}$$

Comparing this with (2.B.46) and using $\|\theta\|_3^3 \leq \theta_{\max}^2 \|\theta\|_1$, we obtain

$$\frac{\mathrm{Var}(1_n' A^2 1_n)}{(1_n' \Omega^2 1_n)^2} \leq \frac{C\|\theta\|_1^3 \|\theta\|_3^3}{\|\theta\|_1^4 \|\theta\|^4} \leq \frac{C\theta_{\max}^2}{\|\theta\|^4},$$

and the claim follows by $\|\theta\| \to \infty$.

2.C Spectral analysis for Ω and $\widetilde{\Omega}$

We state and prove some useful results about eigenvalues and eigenvectors of Ω and $\widetilde{\Omega}$. In Section 2.C.4, we prove Lemma 2.

For $1 \leq k \leq K$, let λ_k be the k-th largest (in absolute value) eigenvalue of Ω and let $\xi_k \in \mathbb{R}^n$ be the corresponding unit-norm eigenvector. We write

$$\Xi = [\xi_1, \xi_2, \ldots, \xi_K] = [u_1, u_2, \ldots, u_n]',$$

so that u_i is the i-th row of Ξ. Recall that G is the $K \times K$ matrix $\|\theta\|^{-2}(\Pi'\Theta^2\Pi)$.

2.C.1 Spectral analysis of Ω

The following lemma relates λ_k and ξ_k to the eigenvalues and eigenvectors of the $K \times K$ matrix $G^{\frac{1}{2}}PG^{\frac{1}{2}}$.

Lemma 5. *Consider the DCMM model. Let d_k be the k-th largest (in absolute value) eigenvalue of $G^{\frac{1}{2}}PG^{\frac{1}{2}}$ and let $\beta_k \in \mathbb{R}^K$ be the associated eigenvector, $1 \leq k \leq K$. Then under the null,*

$$\lambda_1 = \|\theta\|^2, \qquad \xi_1 = \pm\theta/\|\theta\|.$$

Under the alternative, for $1 \leq k \leq K$,

$$\lambda_k = d_k\|\theta\|^2, \qquad \xi_k = \|\theta\|^{-1}[\theta \circ (\Pi G^{-\frac{1}{2}}\beta_k)].$$

Under the alternative hypothesis, we further have the following lemma:

Lemma 6. *Under the DCMM model, as $n \to \infty$, suppose* (2.2.14) *holds. As $n \to \infty$, under the alternative hypothesis,*

$$\lambda_1 \asymp \|\theta\|^2, \qquad \|u_i\| \leq C\|\theta\|^{-1}\theta_i, \quad \text{for all } 1 \leq i \leq n.$$

The quantities $(1_n'\xi_k)$ play key roles in the analysis of the Signed Polygon tests. By Lemma 5,

$$\xi_1 = (\|\theta\|)^{-1}\Theta\Pi G^{-1/2}\beta_1,$$

where β_1 is the first eigenvector of $G^{1/2}PG^{1/2}$, corresponding to the largest eigenvalue of $G^{1/2}PG^{1/2}$. It is seen $G^{-1/2}\beta_1$ is the eigenvector of the matrix PG associated with the largest eigenvalue of GP, which is the same as the largest eigenvalue of $G^{1/2}PG^{1/2}$. Since PG is a non-negative matrix, by Perron's theorem, we can assume all entries of $G^{-1/2}\beta_1$ are non-negative. As a result, all entries of ξ_1 are non-negative, and

$$1_n'\xi_1 > 0.$$

The following lemma is proved in Section 2.C.3.

Lemma 7. *Under the DCMM model, as $n \to \infty$, suppose* (2.2.14) *holds. As $n \to \infty$,*

$$\max_{1 \leq k \leq K} |1_n'\xi_k| \leq C\|\theta\|^{-1}\|\theta\|_1, \qquad 1_n'\xi_1 \geq C\|\theta\|^{-1}\|\theta\|_1.$$

and so for any $2 \leq k \leq K$,

$$|1_n'\xi_k| \leq C|1_n'\xi_1|$$

We also have a lower bound for $1_n'\Omega 1_n$. The following lemma is proved in Section 2.C.3.

Lemma 8. *Under the DCMM model, as $n \to \infty$, suppose* (2.2.14) *holds. As $n \to \infty$, both under the null hypothesis and the alternative hypothesis,*

$$1_n'\Omega 1_n \geq C\|\theta\|_1^2.$$

2.C.2 Spectral analysis of $\widetilde{\Omega}$

Recall that

$$\widetilde{\Omega} = \Omega - (\eta^*)(\eta^*)', \qquad \text{where } \eta^* = (1/\sqrt{1_n'\Omega 1_n})\Omega 1_n,$$

and $\lambda_1, \ldots, \lambda_K$ are the K nonzero eigenvalues of Ω, arranged in the descending order in magnitude, and $\xi_1, \ldots, \xi_K$ are the corresponding unit-norm eigenvectors of Ω The following lemma is proved in Section 2.C.3.

Lemma 9. *Under the DCMM model, as $n \to \infty$, suppose* (2.2.14) *holds. Then,*

$$|\lambda_2| \leq \|\widetilde{\Omega}\| \leq C|\lambda_2|.$$

Moreover, for any fixed integer $m \geq 1$,

$$|(\widetilde{\Omega}^m)_{ij}| \leq C|\lambda_2|^m \cdot \|\theta\|^{-2}\theta_i\theta_j, \quad \textit{for all } 1 \leq i, j \leq n.$$

Recall that $d_1, \ldots, d_K$ are the nonzero eigenvalues of $G^{\frac{1}{2}}PG^{\frac{1}{2}}$. Introduce

$$D = \mathrm{d}iag(d_1, d_2, \ldots, d_K), \qquad \widetilde{D} = \mathrm{d}iag(d_2, d_3, \ldots, d_K),$$

and

$$h = \Big(\frac{1_n'\xi_2}{1_n'\xi_1}, \frac{1_n'\xi_3}{1_n'\xi_1}, \ldots, \frac{1_n'\xi_K}{1_n'\xi_1}\Big)', \qquad u_0 = \sum_{k=2}^{K} \frac{d_k(1_n'\xi_k)^2}{d_1(1_n'\xi_1)^2}.$$

By Lemma 7, $1_n'\xi_1 > 0$, so h and u_0 are both well-defined. Write $\Xi = [\xi_1, \xi_2, \ldots, \xi_K]$. The following lemma gives an alternative expression of $\widetilde{\Omega}$.

Lemma 10. *Under the DCMM model,*

$$\widetilde{\Omega} = \|\theta\|^2 \cdot \Xi M \Xi',$$

where M is a $K \times K$ matrix satisfying

$$M = \begin{bmatrix} (1+u_0)^{-1}h'\widetilde{D}h & -(1+u_0)^{-1}h'\widetilde{D} \\ -(1+u_0)^{-1}\widetilde{D}h & \widetilde{D} - (d_1(1+u_0))^{-1}\widetilde{D}hh'\widetilde{D} \end{bmatrix}.$$

If additionally $|\lambda_2|/\lambda_1 \to 0$, then for the matrix $\widetilde{M} \in \mathbb{R}^{K,K}$,

$$\widetilde{M} = \|\theta\|^2 \cdot \begin{bmatrix} h'\widetilde{D}h & -h'\widetilde{D} \\ -\widetilde{D}h & \widetilde{D} \end{bmatrix},$$

we have

$$|M_{ij} - \widetilde{M}_{ij}| \leq C\lambda_2^2/\lambda_1, \qquad \textit{for all } 1 \leq i, j \leq K.$$

We now study $\mathrm{tr}(\widetilde{\Omega}^4)$, which is related to the power of the SgnQ test. We discuss the two cases $|\lambda_2|/\lambda_1 \to 0$ and $|\lambda_2|/\lambda_1 \geq c_0$ separately. Consider the case of $|\lambda_2|/\lambda_1 = o(1)$. Since $\widetilde{\Omega} = \Xi M \Xi'$, where $\Xi'\Xi = I_K$, we have

$$\mathrm{tr}(\widetilde{\Omega}^4) = \mathrm{tr}(M^4).$$

The following lemma is proved in Section 2.C.3.

Lemma 11. *Consider the DCMM model, where (2.2.14) holds. As $n \to \infty$, if $|\lambda_2|/\lambda_1 \to 0$, then*

$$|\operatorname{tr}(\widetilde{\Omega}^4) - \operatorname{tr}(\widetilde{M}^4)| \le o(|\lambda_2|^3), \tag{2.C.56}$$

Moreover,

$$\begin{aligned}\operatorname{tr}(\widetilde{M}^4) &= \operatorname{tr}(\widetilde{D}^4) + (h'\widetilde{D}h)^4 + 4(h'\widetilde{D}^2h)^2 + 4(h'\widetilde{D}h)^2(h'\widetilde{D}^2h) + 4h'\widetilde{D}^4h + 4(h'\widetilde{D}h)(h'\widetilde{D}^3h)\\ &\ge \operatorname{tr}(\widetilde{D}^4) + (h'\widetilde{D}h)^4 + 2[(h'\widetilde{D}^2h)^2 + (h'\widetilde{D}h)^2(h'\widetilde{D}^2h) + h'\widetilde{D}^4h]\\ &\ge \operatorname{tr}(\widetilde{D}^4).\end{aligned}$$

- *In the special case where $\lambda_2, \lambda_3, \ldots, \lambda_K$ have the same signs,*

  $$|\operatorname{tr}(\widetilde{M}^3)| \ge |\sum_{k=2}^{K} \lambda_k^3| = \sum_{k=2}^{K} |\lambda_k|^3,$$

  and so

  $$|\operatorname{tr}(\widetilde{\Omega}^3)| \ge \sum_{k=2}^{K} |\lambda_k|^3 + o(|\lambda_2|^3).$$

- *In the special case where $K = 2$, the vector h is a scalar, and*

  $$\operatorname{tr}(\widetilde{M}^3) = (1+h^2)^3\lambda_2^3, \qquad \operatorname{tr}(\widetilde{M}^4) = (1+h^2)^4\lambda_2^4,$$

  and so

  $$\operatorname{tr}(\widetilde{\Omega}^3) = [(1+h^2)^3 + o(1)]\lambda_2^3, \qquad \operatorname{tr}(\widetilde{\Omega}^4) = [(1+h^2)^4 + o(1)]\lambda_2^4.$$

We now consider the case $|\lambda_2/\lambda_1| \ge c_0$. In this case, $\widetilde{M}$ is not a good proxy for M any more, so we can not derive a simple formula for $\operatorname{tr}(\widetilde{\Omega}^3)$ or $\operatorname{tr}(\widetilde{\Omega}^4)$ as above. However, for $\operatorname{tr}(\widetilde{\Omega}^4)$, since

$$\operatorname{tr}(\widetilde{\Omega}^4) \ge \|\widetilde{\Omega}\|^4,$$

by Lemma 9, we immediately have

$$\operatorname{tr}(\widetilde{\Omega}^4) \ge C\lambda_2^4 \ge C(\sum_{k=2}^{K} \lambda_k^4)/(K-1) \ge C\sum_{k=2}^{K} \lambda_k^4. \tag{2.C.57}$$

2.C.3 Proof of Lemmas 5-11

Proof of Lemma 5

The proof for the null case is straightforward, so we only prove the lemma for the alternative case. Consider the spectral decomposition

$$G^{1/2}PG^{1/2} = BDB'.$$

where

$$D = \mathrm{d}iag(d_1, \ldots, d_K) \qquad \text{and} \qquad B = [\beta_1, \ldots, \beta_K].$$

Combining this with $\Omega = \Theta\Pi P\Pi'\Theta$ gives

$$\begin{aligned}\Omega &= \Theta\Pi G^{-\frac{1}{2}}(G^{\frac{1}{2}}PG^{\frac{1}{2}})G^{-\frac{1}{2}}\Pi'\Theta \\ &= \Theta\Pi G^{-\frac{1}{2}}(BDB')G^{-\frac{1}{2}}\Pi'\Theta \\ &= (\|\theta\|^{-1}\Theta\Pi G^{-\frac{1}{2}}B)(\|\theta\|^2 D)(\|\theta\|^{-1}\Theta\Pi G^{-\frac{1}{2}}B)' \\ &= H(\|\theta\|^2 D)H',\end{aligned}$$

where

$$H = \|\theta\|^{-1}\Theta\Pi G^{-\frac{1}{2}}B.$$

Recalling that $G = (\|\theta\|^2)^{-1} \cdot \Pi'\Theta^2\Pi$, it is seen

$$H'H = \|\theta\|^{-2}B'G^{-\frac{1}{2}}(\Pi'\Theta^2\Pi)G^{-\frac{1}{2}}B = B'B = I_K, \tag{2.C.58}$$

Therefore,

$$\Omega = H(\|\theta\|^2 D)H'$$

is the spectral decomposition of Ω. Since $(\widetilde{D}_k, \xi_k)$ are the k-th eigenvalue of Ω and unit-norm eigenvector respectively, we have

$$\xi_k = \pm 1 \cdot \text{the } k\text{-th column of H} = \pm(\|\theta\|)^{-1}\Theta\Pi G^{-1/2}\beta_k.$$

This proves the claim. □

Proof of Lemma 6

Consider the first claim. By Lemma 5, $\lambda_1 = d_1\|\theta\|^2$, where d_1 is the maximum eigenvalue of $G^{\frac{1}{2}}PG^{\frac{1}{2}}$. It suffices to show that $d_1 \asymp 1$. Since all entries of P are upper bounded by constants, we have

$$\|P\| \leq C.$$

Additionally, since G is a nonnegative symmetric matrix,

$$\|G\| \leq \|G\|_{\max} = \max_{1\leq k\leq K}\sum_{\ell=1}^{K} G(k,\ell) = \|\theta\|^{-2}\max_{1\leq k\leq K}\sum_{\ell=1}^{K}\sum_{i=1}^{n}\pi_i(k)\pi_i(\ell)\theta_i^2 \leq 1. \tag{2.C.59}$$

It follows that

$$d_1 \leq \|G\|\|P\| \leq C. \tag{2.C.60}$$

At the same time, for any unit-norm non-negative vector $x \in \mathbb{R}^K$, since all entries of P are non-negative and all diagonal entries of P are 1,

$$x'Px \geq x'x = 1.$$

It follows that

$$d_1 = \|G^{\frac{1}{2}}PG^{\frac{1}{2}}\| \geq \frac{(G^{-\frac{1}{2}}x)'(G^{\frac{1}{2}}PG^{\frac{1}{2}})(G^{-\frac{1}{2}}x)}{\|(G^{-\frac{1}{2}}x)\|^2} = \frac{x'Px}{x'G^{-1}x} \geq \frac{1}{\|G^{-1}\|}.$$

Combining it with the assumption (2.2.14) gives

$$d_1 \geq C. \tag{2.C.61}$$

where we note C denotes a generic constant which may vary from occurrence to occurrence. Combining (2.C.60)-(2.C.61) gives the claim.

Consider the second claim. Let $B = [\beta_1, \beta_2, \ldots, \beta_K]$ and $D = diag(d_1, d_2, \ldots, d_K)$ as in the proof of Lemma 5, where we note B is orthonormal. By Lemma 5 and definitions,

$$u_i' = \|\theta\|^{-1}\theta_i\pi_i' G^{-\frac{1}{2}}B.$$

It follows that

$$\|u_i\| \leq \|\theta\|^{-1}\theta_i \cdot \|\pi_i\|\|G^{-\frac{1}{2}}\|\|B\| \leq (\|\theta\|)^{-1}\theta_i\|G^{-1/2}\|,$$

where we have used $\|B\| = 1$ and $\|\pi_i\| = [\sum_{k=1}^K \pi_i(k)^2]^{1/2} \leq 1$. Finally, by the assumption (2.2.14), $\|G^{-1}\| \leq C$ and so $\|G^{-1/2}\| \leq C$. Combining these gives the claim. $\square$

Proof of Lemma 7

It is sufficient to show the first two claims. Consider the first claim. By Lemma 6, for all $1 \leq k \leq K$ and $1 \leq i \leq n$,

$$|\xi_k(i)| \leq C\|\theta\|^{-1}\theta_i.$$

It follows that

$$|1_n'\xi_k| \leq C\sum_{i=1}^n \|\theta\|^{-1}\theta_i \leq C\|\theta\|^{-1}\|\theta\|_1, \qquad \text{for all } 1 \leq k \leq K, \tag{2.C.62}$$

and the claim follows.

Consider the second claim. By Lemma 5,

$$\xi_1 = \|\theta\|^{-1}\Theta\Pi(G^{-\frac{1}{2}}\beta_1), \tag{2.C.63}$$

where β_1 is the (unit-norm) eigenvector of $G^{\frac{1}{2}}PG^{\frac{1}{2}}$ associated with λ_1, which is the largest eigenvalue of $G^{1/2}PG^{1/2}$. By basic algebra, λ_1 is also the largest eigenvalue of the matrix PG, with $G^{-1/2}\beta_1$ being the corresponding eigenvector. Since PG is a nonnegative matrix, $G^{-\frac{1}{2}}\beta_1$ is a nonnegative vector (e.g., (Horn and Johnson, 1985, Theorem 8.3.1)). Denote for short by

$$h = G^{-1/2}\beta_1.$$

It follows from (2.C.63) that

$$1_n'\xi_1 = (\|\theta\|)^{-1} \cdot 1_n'\Theta\Pi h = \|\theta\|^{-1} \cdot \sum_{k=1}^K \Big(\sum_{i=1}^n \pi_i(k)\theta_i\Big)h_k. \tag{2.C.64}$$

We note that $\sum_{k=1}^K \big(\sum_{i=1}^n \pi_i(k)\theta_i\big) = \|\theta\|_1$. Combining it with the assumption (2.2.14) yields

$$\min_{1\leq k\leq K}\Big\{\sum_{i=1}^n \pi_i(k)\theta_i\Big\} \geq C\|\theta\|_1.$$

Inserting this into (2.C.64) gives

$$1_n'\xi_1 \geq C(\|\theta\|)^{-1}\|\theta\|_1 \cdot \|h\|_1. \tag{2.C.65}$$

We claim that $\|h\| \geq 1$. Otherwise, if $\|h\| < 1$, then every entry of h is no greater than 1 in magnitude, and so

$$\|h\|_1 \geq \|h\|^2 = \|G^{-1}\beta_1\|^2.$$

Since $\|G^{-1}\| = \|G\|^{-1} \geq 1$ (see (2.C.59)) and $\|\beta_1\| = 1$,

$$\|G^{-\frac{1}{2}}\beta_1\| \geq 1.$$

and so it follows $\|h\| \geq 1$. The contradiction show that $\|h\| \geq 1$. The claim follows by combining this with (2.C.65). $\square$

Proof of Lemma 8

For $1 \leq k \leq K$, let

$$c = (\|\theta\|_1)^{-1}\Pi'\Theta 1_n = (\|\theta\|_1)^{-1}(1_n'\Theta\Pi)'.$$

Since $\Omega = \Theta\Pi P\Pi'\Theta$ and all entries of P are non-negative,

$$1_n'\Omega 1_n = \|\theta\|_1^2(c'Pc) \geq \|\theta\|^2\Big(\sum_{k=1}^K c_k^2\Big). \tag{2.C.66}$$

Note that, first, $c_k \geq 0$, and second, $\|\theta\|_1 \sum_{k=1}^K c_k = 1_n'\Pi\Theta 1_n = 1_n'\Theta 1_n$, where the last term is $\|\theta\|_1$, and so

$$\sum_{k=1}^K c_k = 1.$$

Together with the Cauchy-Schwartz inequality, we have

$$\sum_{k=1}^K c_k^2 \geq (\sum_{k=1}^K c_k)^2/K = 1/K.$$

Combining this with (2.C.66) gives the claim. $\square$

Proof of Lemma 9

Consider the first claim. We first derive a lower bound for $\|\widetilde{\Omega}\|$. By Lemma 10,

$$\|\widetilde{\Omega}\| = \|\theta\|^2 \cdot \|M\|, \tag{2.C.67}$$

where with the same notations as in the proof of Lemma 10, $M = D - (1+u_0)^{-1}vv'$. Let M_0 be the top left 2×2 block of M. Let $D_0 = \mathrm{d}iag(d_1, d_2)$, and let v_0 be the sub-vector of v in (2.C.72) restricted to the first two coordinates. By (2.C.72),

$$M_0 = D_0 - (1+u_0)^{-1}v_0v_0' = D_0^{\frac{1}{2}}\Big(I_2 - (1+u_0)^{-1}D_0^{-1/2}v_0v_0'D_0^{-\frac{1}{2}}\Big)D_0^{\frac{1}{2}},$$

and so by $\|D_0^{-1/2}\| = |d_2|^{-1/2}$ we have

$$\|\Big(I_2 - (1+u_0)^{-1}D_0^{-1/2}v_0v_0'D_0^{-\frac{1}{2}}\Big)\| \leq \|D_0^{-1/2}M_0D_0^{-1/2}\| \leq |d_2|^{-1} \cdot \|M_0\|. \tag{2.C.68}$$

At the same time, since $(1+u_0)^{-1}D_0^{-1/2}v_0v_0'D_0^{-1/2}$ is a rank-1 matrix, there is an orthonormal matrix and a number b such that

$$Q(1+u_0)^{-1}D_0^{-1/2}v_0v_0'D_0^{-1/2}Q' = \mathrm{d}iag(b, 0).$$

It follows

$$\|\Big(I_2 - (1+u_0)^{-1}D_0^{-1/2}v_0v_0'D_0^{-\frac{1}{2}}\Big)\| = \|I_2 - \mathrm{d}iag(b,0)\| = \max\{|1-b|, 1\} \geq 1.$$

Inserting this into (2.C.68) gives

$$\|M_0\| \geq |d_2|,$$

Note that $\|M\| \geq \|M_0\|$. Combining this with (2.C.67) gives

$$\|\widetilde{\Omega}\| \geq |d_2|\|\theta\|^2.$$

Next, we derive an upper bound for $\|\widetilde{\Omega}\|$. By Lemma 7,

$$\max_{1\le k\le K} |1_n'\xi_k| \le C\|\theta\|^{-1}\|\theta\|_1, \qquad 1_n'\xi_1 \ge C\|\theta\|^{-1}\|\theta\|_1. \tag{2.C.69}$$

By (2.C.69), all the entries of M are upper bounded by $C|\lambda_2|$, which implies $\|M\| \le C|d_2|$. Plugging it into (2.C.67) gives

$$\|\widetilde{\Omega}\| \le \frac{C}{|1+u_0|}|d_2|\|\theta\|^2, \tag{2.C.70}$$

and all remains to show is

$$1 + u_0 \ge C > 0.$$

Now, recalling that $\Omega = \sum_{k=1}^K \lambda_k\xi_k\xi_k'$ and $\lambda_k = d_k\|\theta\|^2$, by definitions,

$$d_1(1_n'\xi_1)^2(1+u_0) = \sum_{k=1}^K d_k(1_n'\xi_k)^2 = \|\theta\|^{-2}1_n'\Omega 1_n.$$

By Lemma 8 which gives $1_n'\Omega 1_n \ge C\|\theta\|_1^2$. It follows that

$$1 + u_0 \ge \frac{\|\theta\|^{-2}1_n'\Omega 1_n}{d_1(1_n'\xi_1)^2} \ge C\frac{\|\theta\|^{-2}\cdot\|\theta\|_1^2}{\|\theta\|^{-2}\cdot\|\theta\|_1^2} \ge C,$$

where in the second inequality we have used (2.C.69) and $d_1 = (\|\theta\|)^{-2}\cdot\lambda_1 \le 1$ (see Lemma 6). Inserting this into (2.C.70) gives the claim.

Consider the second claim. By Lemma 10,

$$\widetilde{\Omega} = \Xi M \Xi',$$

where Ξ and M are the same there. Write

$$\Xi = [\xi_1, \xi_2, \ldots, \xi_K] = [u_1, u_2, \ldots, u_n]'.$$

Note that $\widetilde{\Omega}$ and M have the same spectral norm. It follows that

$$\widetilde{\Omega}^m = \Xi M^m \Xi',$$

and

$$|(\widetilde{\Omega}^m)_{ij}| = |u_i' M^m u_j| \le \|u_i\|\|M\|^m\|u_j\|.$$

By Lemma 6, $\|u_i\|\|u_j\| \le C\|\theta\|^{-2}\theta_i\theta_j$, and by the first part of the current lemma,

$$\|M\| = \|\widetilde{\Omega}\| \le C|d_2|\|\theta\|^2.$$

It follows that

$$|(\widetilde{\Omega}^m)_{ij}| \le C|d_2|^m\|\theta\|^{2m-2}\theta_i\theta_j.$$

This proves the claim. □

Proof of Lemma 10

Consider the first claim. By definitions,

$$\widetilde{\Omega} = \Omega - (\eta^*)(\eta^*)', \qquad \text{where } \eta^* = \frac{1}{\sqrt{1_n'\Omega 1_n}}\Omega 1_n. \tag{2.C.71}$$

Recalling $\widetilde{D}_k = d_k\|\theta\|^2$ and $\Xi = [\xi_1, \xi_2, \ldots, \xi_K]$, we have

$$\Omega = \sum_{k=1}^K \widetilde{D}_k\xi_k\xi_k' = \|\theta\|^2\cdot\Xi D\Xi'.$$

It follows that

$$1_n'\Omega 1_n = \|\theta\|^2 \sum_{k=1}^K d_k(1_n'\xi_k)^2,$$

and

$$\eta^* = \frac{\|\theta\|}{\sqrt{\sum_{s=1}^K d_s(1_n'\xi_s)^2}} \sum_{k=1}^K d_k(1_n'\xi_k)\xi_k = \frac{\|\theta\|}{\sqrt{(1+u_0)}}\Big[\sqrt{d_1}\,\xi_1 + \sum_{k=2}^K \frac{d_k(1_n'\xi_k)}{\sqrt{d_1}(1_n'\xi_1)}\xi_k\Big],$$

where the vector in the big bracket on the right is Ξv, if we let

$$v = (\sqrt{d_1}, \frac{d_2(1_n'\xi_2)}{\sqrt{d_1}(1_n'\xi_1)}, \ldots, \frac{d_K(1_n'\xi_K)}{\sqrt{d_1}(1_n'\xi_1)})'.$$

Combining these gives

$$\widetilde{\Omega} = \|\theta\|^2 \Xi D \Xi' - \frac{\|\theta\|^2}{1+u_0}\Xi v v'\Xi.$$

Plugging it into (2.C.71) gives

$$\widetilde{\Omega} = \|\theta\|^2 \Xi D \Xi' - \frac{\|\theta\|^2}{1+u_0}\Xi v v'\Xi = \|\theta\|^2 \Xi (D - (1+u_0)^{-1} v v')\Xi'. \tag{2.C.72}$$

By definitions,

$$D = \mathrm{d}iag(d_1, d_2, \ldots, d_K), \qquad \text{and} \qquad v = d_1^{-1/2}\cdot(d_1, h'\widetilde{D})'.$$

It follows

$$D - (1+u_0)^{-1}vv' = \begin{bmatrix} (1+u_0)^{-1}d_1u_0 & -(1+u_0)^{-1}h'\widetilde{D} \\ -(1+u_0)^{-1}\widetilde{D}h & \widetilde{D} - (d_1(1+u_0))^{-1}\widetilde{D}hh'\widetilde{D} \end{bmatrix},$$

where we note that

$$d_1 u_0 = \sum_{s=2}^K d_s \frac{(1_n'\xi_s)^2}{(1_n'\xi_1)^2} = h'\widetilde{D}h,$$

Combining these gives the claim.

Consider the second claim. By definitions,

$$M - \widetilde{M} = \|\theta\|^2 \cdot \begin{bmatrix} [(1+u_0)^{-1} - 1]d_1u_0 & (1-(1+u_0)^{-1})h'\widetilde{D} \\ (1-(1+u_0)^{-1})\widetilde{D}h & -(d_1(1+u_0))^{-1}\widetilde{D}hh'\widetilde{D} \end{bmatrix}.$$

Note that

$$|1 - (1+u_0)^{-1}| \leq C|u_0| \leq C|\widetilde{D}_2|/\widetilde{D}_1,$$

and that by Lemma 7,

$$|(1_n'\xi_k)| \leq C 1_n'\xi_1,$$

and so each entry of $\widetilde{D}h$ does not exceed $C|d_2|$. It follows that for all $2 \leq i, j \leq K$,

$$|M_{1i} - \widetilde{M}_{1i}| \leq C\|\theta\|^2(|\widetilde{D}_2|/\widetilde{D}_1)d_2^2 \leq C\widetilde{D}_2^2/\widetilde{D}_1,$$

and

$$|M_{ij} - \widetilde{M}_{ij}| \leq C\|\theta\|^2 d_1^{-1} d_2^2 \leq C\widetilde{D}_2^2/\widetilde{D}_1.$$

Finally,

$$d_1u_0^2 = d_1^{-1}(\sum_{s=2} d_2 \frac{(1_n'\xi_s)^2}{(1_n'\xi_1)^2})^2 \leq C d_2^2/d_1,$$

so

$$|M_{11} - \widetilde{M}_{11}| \leq C\|\theta\|^2 d_2^2/d_1 \leq C\widetilde{D}_2^2/\widetilde{D}_1.$$

Combining these gives the claim. □

Proof of Lemma 11

It is sufficient to show (2.C.56). In fact, once (2.C.56) is proved, other claims follow by direct calculations, except for the first inequality regarding $\mathrm{tr}(\widetilde{\Omega}^4)$, we have used

$$|(h'\widetilde{D}h)(h'\widetilde{D}^3h)| \leq |h'\widetilde{D}h|\sqrt{(h'\widetilde{D}^2h)(h'\widetilde{D}^4h)} \leq \frac{1}{2}\Big[(h'\widetilde{D}h)^2(h'\widetilde{D}^2h) + h'\widetilde{D}^4h\Big].$$

We now show (2.C.56). Since $\mathrm{tr}(\widetilde{\Omega}^m) = \mathrm{tr}(\widetilde{M}^m)$, for $m = 3, 4$, it is sufficient to show

$$|\,\mathrm{tr}(M^3) - \mathrm{tr}(\widetilde{M}^3)| \leq C\lambda_2^4/\lambda_1), \qquad |\,\mathrm{tr}(M^4) - \mathrm{tr}(\widetilde{M}^4)| \leq C|\lambda_2|^5/\lambda_1. \tag{2.C.73}$$

Since the proofs are similar, we only show the first one. By basic algebra,

$$\mathrm{tr}(M^3 - \widetilde{M}^3) = \mathrm{tr}((M - \widetilde{M})^3) + 3\,\mathrm{tr}(\widetilde{M}(M - \widetilde{M})^2) + 3\,\mathrm{tr}(\widetilde{M}^2(M - \widetilde{M})).$$

By Lemma 10, for all $1 \leq i, j \leq K$,

$$|M_{ij} - \widetilde{M}_{ij}| \leq C\lambda_2^2/\lambda_1.$$

Also, by Lemma 7, all entries of h are bounded, so for all $1 \leq i, j \leq K$,

$$|\widetilde{M}_{ij}| \leq |\lambda_2|.$$

It follows

$$|\,\mathrm{tr}((M - \widetilde{M})^3| \leq C(\lambda_2^2/\lambda_1)^3,$$

$$|\,\mathrm{tr}(\widetilde{M}(M - \widetilde{M})^2)| \leq C|\lambda_2|(\lambda^2/\lambda_1)^2 \leq C|\lambda_2|^5/\lambda_1^2,$$

and

$$|\,\mathrm{tr}(\widetilde{M}^2(M - \widetilde{M})| \leq C\lambda_2^2(\lambda^2/\lambda_1) \leq C\lambda_2^4/\lambda_1.$$

where we note that $\lambda_2/\lambda_1 = o(1)$. Combining these gives the claim.

2.C.4 Proof of Lemma 2

The second bullet point is a direct result of (2.C.57), and the other two bullet points follow directly from Lemma 11 of this appendix.

2.D Lower bounds, Region of Impossibility

We study the Region of Impossibility by considering a DCMM with random mixed memberships. First, in Section 2.D.1, we establish the equivalence between regularity conditions for a DCMM with non-random mixed memberships and those for a DCMM with random mixed memberships. Next, we prove Lemma 3, which is key to the construction of inseparable hypothesis pairs. Last, we prove Theorem 2.3.1-2.3.5 of the main article.

2.D.1 Equivalence of regularity conditions

Let $\mu_1, \mu_2, \ldots, \mu_K$ be the eigenvalues of P, arranged in the descending order in magnitude. Recall that $\lambda_1, \lambda_2, \ldots, \lambda_K$ are the eigenvalues of Ω. The following lemma is proved in Section 2.D.5.

Lemma 12 (Equivalent definition of Region of Impossibility)**.** *Consider the DCMM model* (3.1.1)-(3.2.4)*, where the alternative is true and the condition* (2.2.14) *holds. Suppose* $\theta_{\max} \to 0$ *and* $\|\theta\| \to \infty$ *as* $n \to \infty$*. Then, as* $n \to \infty$,

$$\mu_1 \asymp 1, \qquad \frac{|\mu_2|}{\mu_1} \asymp \frac{|\lambda_2|}{\lambda_1}, \qquad \max_{1 \leq i,j \leq K} |P_{ij} - 1| \leq C(|\lambda_2|/\lambda_1).$$

As a result, $|\lambda_2|/\sqrt{\lambda_1} \to 0$ *if and only if* $\|\theta\| \cdot |\mu_2(P)| \to 0$.

We now consider DCMM with random mixed memberships: Given (Θ, P) and a distribution F over V (the standard simplex in $\mathbb{R}^K$), let

$$\Omega = \Theta\Pi P \Pi'\Theta, \qquad \Pi = [\pi_1, \pi_2, \ldots, \pi_n]', \qquad \pi_i \overset{iid}{\sim} F. \tag{2.D.74}$$

We notice that the conclusion of Lemma 12 holds provided that the regularity condition (2.2.14) is satisfied. The next lemma shows that (2.2.14) holds with high probability. It is proved in Section 2.D.5.

Lemma 13 (Equivalence of regularity conditions)**.** *Consider the model* (2.D.74)*. Let* $h = \mathbb{E}[\pi_i]$ *and* $\Sigma = \mathrm{Cov}(\pi_i)$*. Suppose* $\|P\| \leq C$*,* $\min_{1\leq k\leq K}\{h_k\} \geq C$ *and* $\|\Sigma^{-1}\| \leq C$*. Suppose* $\theta_{\max} \to 0$*,* $\|\theta\| \to \infty$*, and* $(\|\theta\|^2/\|\theta\|_1)\sqrt{\log(\|\theta\|_1)} \to 0$*, as* $n \to \infty$*. Then, as* $n \to \infty$*, with probability* $1 - o(1)$*, the condition* (2.2.14) *is satisfied, i.e.,*

$$\frac{\max_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i \pi_i(k)\}}{\min_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i \pi_i(k)\}} \leq C_0, \qquad \|G^{-1}\| \leq C_0,$$

for a constant $C_0 > 0$ *and* $G = \|\theta\|^{-2}(\Pi'\Theta^2\Pi)$.

2.D.2 Proof of Lemma 3

Let $M = \mathrm{d}iag(\mu_1, \mu_2, \ldots, \mu_K)$. It is seen $\mu = M1_K$ and so the desired result is to find a D such that

$$DADM1_K = 1_K \Longleftrightarrow MDADM1_K = M1_K = \mu \Longleftrightarrow D(MAM)D1_K = \mu.$$

Since MAM has strictly positive entries, it is sufficient to show that for any matrix A (MAM in our case; a slight misuse notation here) with strictly positive entries, there is a unique diagonal matrix D with strictly positive diagonal entries such that

$$DAD1_k = \mu_K. \tag{2.D.75}$$

We now show the existence and uniqueness separately.

For existence, we follow the proof in Marshall and Olkin (1968). Consider $d'Ad$ for a vector $d \in \mathbb{R}^K$ with strictly positive entries. It is shown there that $d'Ad$ can be minimized using Lagrange multiplier:

$$\frac{1}{2}d'Ad - \lambda \sum_{k=1}^{K} \mu_k \log(d_k).$$

Differentiating with respect to d and set the derivative to 0 gives

$$Ad = \lambda \sum_{k=1}^{K} \mu_k / d_k, \tag{2.D.76}$$

where $\lambda = d'Ad/(\sum_{k=1}^{K} \mu_k) > 0$. Letting $D = \lambda^{-1/2} \mathrm{d}iag(d_1, d_2, \ldots, d_K)$. It is seen that (2.D.76) can be rewritten as

$$DAD1_K = \mu,$$

and the claim follows.

For uniqueness, we adapt the proof in Johnson and Reams (2009) to our case. Suppose there are two different eligible diagonal matrices D_1 and D_2 satisfying (2.D.75). Let $d_1 = D_1 1_K$ and $d_2 = D_2 1_K$, and let $M = \mathrm{d}iag(\mu_1, \mu_2, \ldots, \mu_K)$. It follows that

$$D_2 D_1 A d_1 = D_2 D_1 A D_1 1_K = D_2 \mu = M d_2,$$

and so

$$M^{-1} D_2 D_1 A d_1 = d_2.$$

Now, for a diagonal matrix S with strictly positive diagonal entries to be determined, we have

$$S^{-1} M^{-1} D_2 D_1 A S S^{-1} d_1 = S^{-1} d_2.$$

We pick S such that

$$S^{-1} M^{-1} D_2 D_1 = S,$$

and denote such an S by S_0. It follows

$$S_0 A S_0 (S_0^{-1} d_1) = S_0^{-1} d_2.$$

or equivalently, if we let $\widetilde{d}_1 = S_0^{-1} d_1$ and $\widetilde{d}_2 = S_0^{-1} d_2$,

$$S_0 A S_0 \widetilde{d}_1 = \widetilde{d}_2. \tag{2.D.77}$$

Similarly, by switching the places of D_1 and D_2, we have

$$S_0 A S_0 \widetilde{d}_2 = \widetilde{d}_1. \tag{2.D.78}$$

Combining (2.D.77) and (2.D.78) gives

$$S_0 A S_0 (\widetilde{d}_1 + \widetilde{d}_2) = (\widetilde{d}_1 + \widetilde{d}_2), \qquad \text{and} \qquad S_0 A S_0 (\widetilde{d}_1 - \widetilde{d}_2) = -(\widetilde{d}_1 - \widetilde{d}_2).$$

This implies that 1 and -1 are the two eigenvalues of $S_0 A S_0$, with $\widetilde{d}_1 + \widetilde{d}_2$ and $\widetilde{d}_1 - \widetilde{d}_2$ being the corresponding eigenvectors, respectively, where we note that especially, $\widetilde{d}_1 + \widetilde{d}_2$ has all strictly positive entries. By Perron's theorem Horn and Johnson (1985), since $S_0 A S_0$ have all strictly positive entries, the eigenvector corresponding to the largest eigenvalue (i.e., the Perron root) have all strictly positive entries. As for any symmetric matrix, we can only have one eigenvector that has all strictly positive entries, so 1 must be the Perron root of S_0AS_0. Using Perron's Theorem again, all eigenvalues of S_0AS_0 except the Perron root itself should be strictly smaller than 1 in magnitude. This contradicts with the fact that -1 is an eigenvalue of S_0AS_0. The contradiction proves the uniqueness. □

2.D.3 Proof of Theorem 2.3.1

This theorem follows easily from Theorems 2.3.2-2.3.5. Fix (Θ, P, F) such that $\theta \in \mathcal{M}_n^*(\beta_n/\log(n))$ and $\|\theta\| \cdot |\mu_2(P)| \leq \alpha_n/\log(n)$. Consider a sequence of hypotheses indexed by n, where $\Omega = \theta\theta'$ under $H_0^{(n)}$, and Ω follows the construction in any of Theorems 2.3.2-2.3.5 under $H_1^{(n)}$. Let $P_0^{(n)}$ and $P_1^{(n)}$ be the probability measures associated with two hypotheses, respectively. By those theorems, the χ^2-distance satisfy

$$\mathcal{D}(P_0^{(n)}, P_1^{(n)}) = o(1), \qquad \text{as } n \to \infty.$$

By connection between L^1-distance and χ^2-distance, it follows that

$$\|P_0^{(n)} - P_1^{(n)}\|_1 = o(1), \qquad \text{as } n \to \infty.$$

We now slightly modify the alternative hypothesis. Let Π_0 be a non-random membership matrix such that $(\theta, \Pi_0, P) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$. In the modified alternative hypothesis $\widetilde{H}_1^{(n)}$,

$$\Pi = \begin{cases} \widetilde{\Pi}, & \text{if } (\theta, \widetilde{\Pi}, P) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n), \\ \Pi_0, & \text{otherwise}, \end{cases} \qquad \text{where} \quad \widetilde{\pi}_i \overset{iid}{\sim} F.$$

Let $\widetilde{P}_1^{(n)}$ be the probability measure associated with $\widetilde{H}_1^{(n)}$. By Lemmas 12-13, $\Pi = \widetilde{\Pi}$, except for vanishing probability. It follows that

$$\|P_1^{(n)} - \widetilde{P}_1^{(n)}\|_1 = o(1), \qquad \text{as } n \to \infty.$$

Under $\widetilde{H}_1^{(n)}$, all realizations (θ, Π, P) are in the class $\mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$. By Neymann-Pearson lemma and elementary inequalities,

$$\begin{aligned}
&\inf_{\psi}\Big\{ \sup_{\theta \in \mathcal{M}_n^*(\beta_n)} \mathbb{P}(\psi = 1) + \sup_{(\theta,\Pi,P) \in \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)} \mathbb{P}(\psi = 0) \Big\} \\
&\geq 1 - \inf_{f_0 \in \mathcal{M}_n^*(\beta_n), f_1 \in \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)} \{\|f_0 - f_1\|_1\} \\
&\geq 1 - \|P_0^{(n)} - \widetilde{P}_1^{(n)}\|_1 \\
&\geq 1 - \|P_0^{(n)} - P_1^{(n)}\|_1 - \|P_1^{(n)} - \widetilde{P}_1^{(n)}\|_1 \\
&\geq 1 - o(1),
\end{aligned}$$

where in the second line we have mis-used the notation $f \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$ to denote the probability density for a DCMM with non-random mixed memberships whose parameters are in the class $\mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$.

2.D.4 Proof of Theorems 2.3.2-2.3.5

We note that Theorem 2.3.2, Theorem 2.3.4 and Theorem 2.3.5 can be deduced from Theorem 2.3.3. To see this, recall that Theorem 2.3.3 assumes there exists a positive diagonal matrix D such that

$$DPD\widetilde{h}_D = 1_K, \qquad \min_{1\le k\le K}\{\widetilde{h}_{D,k}\} \ge C, \tag{2.D.79}$$

where $\widetilde{h}_D = \mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$. We show that the condition (2.D.79) is implied by conditions of other theorems. Theorem 2.3.2 assumes $\pi_i \in \{e_1, e_2, \dots, e_K\}$. It follows that $D^{-1}\pi_i/\|D^{-1}\pi_i\|_1 = \pi_i$, and so $\widetilde{h}_D = h$. By Lemma 3, there exists D such that $DPDh = 1_K$, hence, (2.D.79) is satisfied. Theorem 2.3.4 constructs the alternative hypothesis using $\widetilde{\pi}_i = D\pi_i/\|D\pi_i\|_1$. Equivalently, $D^{-1}\widetilde{\pi}_i/\|D^{-1}\pi_i\|_1 = \pi_i$, and so $\widetilde{h}_D$ becomes h. Since $DPDh = 1_K$, condition (2.D.79) holds. Theorem 2.3.5 assumes $Ph = q_n 1_K$. Let $D = q_n^{-1/2} I_K$. Then, $\widetilde{h}_D = h$ and $DPDh = q_n^{-1}Ph = 1_K$. Again, (2.D.79) is satisfied.

We only need to prove Theorem 2.3.3. Let $P_0^{(n)}$ and $P_1^{(n)}$ be the probability measure associated with $H_0^{(n)}$ and $H_1^{(n)}$, respectively. Let $\mathcal{D}(P_0^{(n)}, P_1^{(n)})$ be the chi-square distance between two probability measures. By elementary probability,

$$\mathcal{D}(P_0^{(n)}, P_1^{(n)}) = \int \left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 dP_0^{(n)} - 1.$$

It suffices to show that, when $\|\theta\| \cdot \mu_2(P) \to 0$,

$$\int \left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 dP_0^{(n)} = 1 + o(1). \tag{2.D.80}$$

Let p_{ij} and $q_{ij}(\Pi)$ be the corresponding Ω_{ij} under the null and the alternative, respectively. It is seen that

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

Let $\widetilde{\Pi}$ be an independent copy of Π. Then,

$$\begin{aligned}\left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 &= \mathbb{E}_\Pi\Big[\prod_{i<j}\Big(\frac{q_{ij}(\Pi)}{p_{ij}}\Big)^{A_{ij}}\Big(\frac{1-q_{ij}(\Pi)}{1-p_{ij}}\Big)^{1-A_{ij}}\Big] \cdot \mathbb{E}_{\widetilde{\Pi}}\Big[\prod_{i<j}\Big(\frac{q_{ij}(\widetilde{\Pi})}{p_{ij}}\Big)^{A_{ij}}\Big(\frac{1-q_{ij}(\widetilde{\Pi})}{1-p_{ij}}\Big)^{1-A_{ij}}\Big] \\ &= \mathbb{E}_{\Pi,\widetilde{\Pi}}\underbrace{\Big[\prod_{i<j}\Big(\frac{q_{ij}(\Pi)q_{ij}(\widetilde{\Pi})}{p_{ij}^2}\Big)^{A_{ij}}\Big(\frac{[1-q_{ij}(\Pi)][1-q_{ij}(\widetilde{\Pi})]}{[1-p_{ij}]^2}\Big)^{1-A_{ij}}\Big]}_{S(A,\Pi,\widetilde{\Pi})}.\end{aligned}$$

It follows that

$$\begin{aligned}\int \left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 dP_0^{(n)} &= \mathbb{E}_A\left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 \\ &= \mathbb{E}_{A,\Pi,\widetilde{\Pi}}[S(A,\Pi,\widetilde{\Pi})] \\ &= \mathbb{E}_{\Pi,\widetilde{\Pi}}\big\{\mathbb{E}_A\big[S(A,\Pi,\widetilde{\Pi})|\Pi,\widetilde{\Pi}\big]\big\},\end{aligned}$$

where the distribution of $A|(\Pi,\widetilde{\Pi})$ is under the null hypothesis. Under the null hypothesis, A is independent of $(\Pi,\widetilde{\Pi})$, the upper triangular entries of A are independent of each other, and $A_{ij} \sim \text{Bernoulli}(p_{ij})$. It follows that

$$\begin{aligned}\mathbb{E}_A\big[S(A,\Pi,\widetilde{\Pi})|\Pi,\widetilde{\Pi}\big] &= \prod_{i<j}\mathbb{E}_A\bigg[\Big(\frac{q_{ij}(\Pi)q_{ij}(\widetilde{\Pi})}{p_{ij}^2}\Big)^{A_{ij}}\Big(\frac{[1-q_{ij}(\Pi)][1-q_{ij}(\widetilde{\Pi})]}{[1-p_{ij}]^2}\Big)^{1-A_{ij}}\bigg|\Pi,\widetilde{\Pi}\bigg] \\ &= \prod_{i<j}\bigg\{p_{ij}\frac{q_{ij}(\Pi)q_{ij}(\widetilde{\Pi})}{p_{ij}^2} + (1-p_{ij})\frac{[1-q_{ij}(\Pi)][1-q_{ij}(\widetilde{\Pi})]}{[1-p_{ij}]^2}\bigg\} \\ &= \prod_{i<j}\bigg\{\frac{q_{ij}(\Pi)q_{ij}(\widetilde{\Pi})}{p_{ij}} + \frac{[1-q_{ij}(\Pi)][1-q_{ij}(\widetilde{\Pi})]}{1-p_{ij}}\bigg\}.\end{aligned}$$

Let $\Delta_{ij} = q_{ij}(\Pi) - p_{ij}$ and $\widetilde{\Delta}_{ij} = q_{ij}(\widetilde{\Pi}) - p_{ij}$. By direct calculations,

$$\frac{q_{ij}(\Pi)q_{ij}(\widetilde{\Pi})}{p_{ij}} + \frac{[1-q_{ij}(\Pi)][1-q_{ij}(\widetilde{\Pi})]}{1-p_{ij}} = 1 + \frac{\Delta_{ij}\widetilde{\Delta}_{ij}}{p_{ij}(1-p_{ij})}.$$

Combining the above gives

$$\int\bigg[\frac{dP_1^{(n)}}{dP_0^{(n)}}\bigg]^2 dP_0^{(n)} = \mathbb{E}_{\Pi,\widetilde{\Pi}}\bigg[\prod_{i<j}\Big(1 + \frac{\Delta_{ij}\widetilde{\Delta}_{ij}}{p_{ij}(1-p_{ij})}\Big)\bigg]. \tag{2.D.81}$$

We then plug in the expressions of Δ_{ij} and $\widetilde{\Delta}_{ij}$ from the model. Let D be the matrix in (2.D.79). Introduce $M = DPD - \mathbf{1}_K\mathbf{1}_K'$. We re-write

$$DPD = \mathbf{1}_K\mathbf{1}_K' + M.$$

It is seen that $M\widetilde{h}_D = \mathbf{0}_K$. The following lemma is proved in Section 2.D.5.

Lemma 14. *Under the conditions of Theorem 2.3.3, $\|M\| \le C|\mu_2(P)|$.*

Write for short $\pi_i^D = \frac{1}{\|D^{-1}\pi_i\|_1}D^{-1}\pi_i$ and $y_i = \pi_i^D - \mathbb{E}[\pi_i^D] = \pi_i^D - \widetilde{h}_D$. Under the alternative hypothesis,

$$\begin{aligned} q_{ij}(\Pi) &= \theta_i\theta_j\|D^{-1}\pi_i\|_1\|D^{-1}\pi_j\|_1 \cdot \pi_i'P\pi_j \\ &= \theta_i\theta_j \cdot (\pi_i^D)'(DPD)(\pi_j^D) \\ &= \theta_i\theta_j \cdot (\pi_i^D)'(\mathbf{1}_K\mathbf{1}_K' + M)(\pi_j^D) \\ &= \theta_i\theta_j \cdot \big[1 + (\pi_i^D)'M(\pi_j^D)\big] \\ &= \theta_i\theta_j \cdot \big[1 + (\widetilde{h}_D + y_i)'M(\widetilde{h}_D + y_j)\big] \\ &= \theta_i\theta_j \cdot (1 + y_i'My_j).\end{aligned}$$

Here, the fourth line is due to $\mathbf{1}_K'\pi_i = 1$ and the last line is due to $M\widetilde{h}_D = \mathbf{0}_K$. Under the null hypothesis, $p_{ij} = \theta_i\theta_j$. As a result,

$$\Delta_{ij} = \theta_i\theta_j \cdot y_i'My_j, \qquad y_i \equiv \pi_i^D - \mathbb{E}[\pi_i^D].$$

Similarly, $\widetilde{\Delta}_{ij} = \theta_i\theta_j \cdot \widetilde{y}_i'M\widetilde{y}_j$, with $\widetilde{y}_i = \widetilde{\pi}_i^D - \mathbb{E}[\widetilde{\pi}_i^D]$. We plug them into (2.D.81) and use $p_{ij} = \theta_i\theta_j$. It gives

$$\int\bigg[\frac{dP_1^{(n)}}{dP_0^{(n)}}\bigg]^2 dP_0^{(n)} = \mathbb{E}\bigg[\prod_{i<j}\Big(1 + \frac{\theta_i\theta_j}{1-\theta_i\theta_j}(y_i'My_j)(\widetilde{y}_i'M\widetilde{y}_j)\Big)\bigg], \tag{2.D.82}$$

where $\{y_i, \widetilde{y}_i\}_{i=1}^n$ are *iid* random vectors with $\mathbb{E}[y_i] = \mathbf{0}_K$.

We bound the right hand side of (2.D.82). Since $1 + x \le e^x$ for all $x \in \mathbb{R}$,

$$\mathcal{D}(P_0^{(n)}, P_1^{(n)}) \le \mathbb{E}[\exp(S)], \qquad \text{where} \quad S \equiv \sum_{i<j} \frac{\theta_i\theta_j}{1-\theta_i\theta_j}(y_i' M y_j)(\widetilde{y}_i' M \widetilde{y}_j).$$

Let $M = \sum_{k=1}^K \delta_k b_k b_k'$ be the eigen-decomposition of M. Then,

$$(y_i' M y_j)(\widetilde{y}_i' M \widetilde{y}_j) = \sum_{1\le k,\ell\le K} \delta_k\delta_\ell (b_k' y_i)(b_k' y_j)(b_\ell' \widetilde{y}_i)(b_\ell' \widetilde{y}_j).$$

This allows us to decompose

$$S = \frac{1}{K^2}\sum_{1\le k,\ell\le K} S_{k\ell}, \qquad \text{where} \;\; S_{k\ell} = K^2\delta_k\delta_\ell \sum_{i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}(b_k' y_i)(b_k' y_j)(b_\ell' \widetilde{y}_i)(b_\ell' \widetilde{y}_j).$$

By Jensen's inequality, $\exp(\frac{1}{K^2}\sum_{k,\ell} S_{k\ell}) \le \frac{1}{K^2}\sum_{k,\ell}\exp(S_{k\ell})$. It follows that

$$\int \left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 dP_0^{(n)} \le \mathbb{E}[\exp(S)] \le \max_{1\le k,\ell\le K} \mathbb{E}[\exp(S_{k\ell})]. \tag{2.D.83}$$

We now fix (k,ℓ) and derive a bound for $\mathbb{E}[\exp(S_{k\ell})]$. For n large enough, $\theta_{\max} \le 1/2$ and $K^4\|M\|^2\|\theta\|^2 \le 1/9$. By Taylor expansion of $(1-\theta_i\theta_j)^{-1}$,

$$\begin{aligned} S_{k\ell} &= K^2\delta_k\delta_\ell \sum_{i<j}\sum_{m=1}^\infty \theta_i^m\theta_j^m (b_k' y_i)(b_k' y_j)(b_\ell' \widetilde{y}_i)(b_\ell' \widetilde{y}_j) \\ &\equiv \sum_{m=1}^\infty X_m, \qquad \text{where} \quad X_m \equiv K^2\delta_k\delta_\ell\sum_{i<j}\theta_i^m\theta_j^m (b_k' y_i)(b_k' y_j)(b_\ell' \widetilde{y}_i)(b_\ell' \widetilde{y}_j). \end{aligned}$$

Since $|X_m| \le C\|M\|^2\|\theta\|_m^{2m} \le C\|M\|\|\theta\|_1^2\theta_{\max}^{2(m-1)}$, where $\sum_{m=1}^\infty \theta_{\max}^{2(m-1)} < \infty$, the random variable $\sum_{m=1}^\infty X_m$ is always well-defined. For $m \ge 1$, let $a_m = \theta_{\max}^{2(m-1)}(1-\theta_{\max}^2)$. Then, $\sum_{m=1}^\infty a_m = 1$. By Jenson's inequality,

$$\exp\Big(\sum_{m=1}^\infty X_m\Big) = \exp\Big(\sum_{m=1}^\infty a_m\cdot a_m^{-1}|X_m|\Big) \le \sum_{m=1}^\infty a_m\cdot\exp(a_m^{-1}X_m).$$

Using Fatou's lemma, we have

$$\mathbb{E}[\exp(S_{k\ell})] \le \sum_{m=1}^\infty a_m\cdot\mathbb{E}\big[\exp(a_m^{-1}X_m)\big]. \tag{2.D.84}$$

By definition of X_m,

$$X_m = K^2\delta_k\delta_\ell\bigg\{\Big[\sum_i \theta_i^m (b_k' y_i)(b_\ell' \widetilde{y}_i)\Big]^2 - \sum_i \theta_i^{2m}(b_k' y_i)^2(b_\ell' \widetilde{y}_i)^2\bigg\}.$$

Note that $\max_i\{\|y_i\|, \|\widetilde{y}_i\|\} \le \sqrt{K}$ and $\max_k|\delta_k| = \|M\|$. Therefore,

$$|X_m| \le K^2\|M\|^2\Big[\sum_i \theta_i^m(b_k' y_i)(b_\ell' \widetilde{y}_i)\Big]^2 + K^4\|M\|^2\|\theta\|_{2m}^{2m}.$$

Write $Y = \sum_i \theta_i^m(b_k' y_i)(b_\ell' \widetilde{y}_i)$. We see that Y is sum of independent, mean-zero random variables. Since $|(b_k' y_i)(b_\ell' \widetilde{y}_i)| \le K$, by Hoeffding's inequality,

$$\mathbb{P}(|Y| > t) \le 2\exp\Big(-\frac{t^2}{4K^2\|\theta\|_{2m}^{2m}}\Big), \qquad \text{for any } t > 0.$$

Since $\|\theta\|_{2m}^{2m} \le \|\theta\|^2\theta_{\max}^{2(m-1)} \le 2a_m\|\theta\|^2$, we have $a_m^{-1}K^4\|M\|^2\|\theta\|_{2m}^{2m} \le 2K^4\|M\|^2\|\theta\|^2$. Note that $K^4\|M\|^2\|\theta\|^2 \leq 1/9$. By direct calculations,

$$\begin{aligned}
\mathbb{E}\big[\exp(a_m^{-1}|X_m|)\big] &\leq e^{a_m^{-1}K^4\|M\|^2\|\theta\|_{2m}^{2m}} \cdot \mathbb{E}\big[e^{a_m^{-1}K^2\|M\|^2Y^2}\big] \\
&\leq e^{2K^4\|M\|^2\|\theta\|^2} \cdot \mathbb{E}\big[e^{a_m^{-1}K^2\|M\|^2Y^2}\big] \\
&= e^{2K^4\|M\|^2\|\theta\|^2}\Big[1 + \int_0^\infty e^t \cdot \mathbb{P}\big(a_m^{-1}K^2\|M\|^2Y^2 > t\big)dt\Big] \\
&\leq e^{2K^4\|M\|^2\|\theta\|^2}\Big[1 + \int_0^\infty e^t \cdot e^{-\frac{t}{8K^4\|M\|^2\|\theta\|^2}}dt\Big] \\
&\leq e^{K^4\|M\|^2\|\theta\|^2} \cdot (1 + 72K^4\|M\|^2\|\theta\|^2).
\end{aligned}$$

We plug it into (2.D.84) and notice that $\sum_{m=1}^\infty a_m = 1$. It gives

$$\mathbb{E}[\exp(S_{k\ell})] \leq e^{K^4\|M\|^2\|\theta\|^2} \cdot (1 + 72K^4\|M\|^2\|\theta\|^2). \tag{2.D.85}$$

Combining (2.D.83) and (2.D.85) gives

$$\int \left[\frac{dP_1^{(n)}}{dP_0^{(n)}}\right]^2 dP_0^{(n)} \leq e^{K^4\|M\|^2\|\theta\|^2} \cdot (1 + 72K^4\|M\|^2\|\theta\|^2).$$

We recall that $\|\theta\| \cdot \|M\| \leq C\|\theta\| \cdot |\mu_2(P)| \to 0$. Hence, the right hand side is $1 + o(1)$. This proves (2.D.80).

2.D.5 Proof of Lemmas 12-14

Proof of Lemma 12

The first claim follows by our assumptions on P, so we omit the proof. Consider the second claim. Recall that $G = \|\theta\|^{-2}\Pi'\Theta^2\Pi$ and $d_1, d_2, \ldots, d_K$ are the eigenvalues of $G^{1/2}PG^{1/2}$, arranged in the descending order in magnitude. By Lemmas D.1 and D.2, $\lambda_k = \|\theta\|^2 d_k$, $1 \leq k \leq K$, and $d_1 \asymp 1$. Combining these, it suffices to show

$$|\mu_2| \asymp |d_2|.$$

We now prove for the cases where P is non-singular and singular, separately. Consider the first case. Since $1/d_k$ and $1/\mu_K$ are the largest eigenvalue of $G^{-1/2}P^{-1/2}G^{-1/2}$ and P^{-1} in magnitude, respectively, and $\|G\| \leq C$ and $\|G^{-1}\| \leq C$, it is seen that $|\mu_K| \asymp |d_K|$. To show the claim, it sufficient to show that for any $m \geq 2$, if $|\mu_k| \asymp |d_k|$ for $k = m+1, \ldots, K$, then $|\mu_m| \asymp |d_m|$.

We now fix $m \geq 2$, and assume $|\mu_k| \asymp |d_k|$ for $k = m+1, \ldots, K$. The goal is to show $|\mu_m| \asymp |d_m|$. By symmetry, it is sufficient to show that

$$|d_m| \leq C|\mu_m|. \tag{2.D.86}$$

Let $P = V\mathrm{d}iag(d_1, d_2, \ldots, d_K)V'$ be the SVD of P, where $V \in \mathbb{R}^{K,K}$ is orthonormal, and let V_m be the sub-matrix of V consisting the first m columns of V. Introduce

$$\widetilde{P}_m = V_m D_m V_m', \qquad \text{where } D_m = \mathrm{d}iag(d_1, d_2, \ldots, d_m).$$

Let $\mu_1^*, \mu_2^*, \ldots, \mu_m^*$ and $d_1^*, d_2^*, \ldots, d_m^*$ be the first m eigenvalues of $\widetilde{P}_m$ and $G^{1/2}P_mG^{1/2}$, respectively, arranged in the descending order in magnitude. Since $\|G\| \leq C$, we have

$$\|P - P_m\| \leq C|\mu_{m+1}|, \qquad \|G^{1/2}(P - P_m)G^{1/2}\| \leq C|\mu_{m+1}|.$$

By Theorem (Bai and Silverstein, 2010, Theorem A.46),

$$|\mu_m - \mu_m^*| \leq C\|P - P_m\| \leq |\lambda_{m+1}|, \tag{2.D.87}$$

and

$$|d_m - d_m^*| \leq \|G^{1/2}(P - P_m)G^{1/2}\| \leq C|\mu_{m+1}|. \tag{2.D.88}$$

and At the same time, note that the nonzero eigenvalues of $G^{1/2}P_mG^{1/2}$ are the same as the nonzero eigenvalues of $D_mV_m'GV_m$, and also the same as those of $(V_m'GV_m)^{1/2}D_m(V_m'GV_m)^{1/2}$. Since $\|G\| \leq C$ and $\|G^{-1}\| \leq C$, it is seen $\|V_m'GV_m\| \leq C$ and $\|V_m'GV_m)^{-1}\| \leq C$. Therefore, by similar arguments,

$$|\mu_m^*| \asymp |d_m^*|. \tag{2.D.89}$$

Combining (2.D.87), (2.D.88), and (2.D.89) gives

$$\begin{aligned}|\mu_m| &\leq |\mu_m^*| + |\mu_m - \mu_m^*| \leq C(|d_m^*| + |d_{m+1}|) \\ &\leq C[(|d_m| + |d_m - d_m^*|) + |d_{m+1}|] \leq C|d_m|.\end{aligned}$$

This proves (2.D.86) and the claim follows.

We now consider the case where P is singular, say, $rank(P) = r < K$, and the nonzero eigenvalues are $\mu_1, \mu_2, \ldots, \mu_r$. Let $P = UDU'$ be the SVD, where $U \in \mathbb{R}^{n,r}$ and $D = diag(\mu_1, \mu_2, \ldots, \mu_r)$. By similar argument, the nonzero eigenvalues of $G^{1/2}PG^{1/2}$ are the same as $(U'GU)^{1/2}D(U'GU)^{1/2}$, where $\|U'GU\| \leq C$ and $\|(U'GU)^{-1}\| \leq C$. The remaining part of the proof is similar so is omitted.

Consider the last claim. Let $\widetilde{P} = \eta\eta'$, where η is the first eigenvector of P, scaled to have a ℓ^2-norm of $\sqrt{\mu_1}$. Write

$$|P_{ij} - 1| = |P_{ij} - \eta_i\eta_j| + |\eta_i\eta_j - 1|. \tag{2.D.90}$$

Now, first, by definitions and elementary algebra, for $1 \leq i, j \leq K$,

$$|P_{ij} - \eta_i\eta_j| \leq |P_{ij} - \widetilde{P}_{ij}| \leq \|P - \widetilde{P}\| \leq \mu_2, \tag{2.D.91}$$

where by the second claim, $\mu_2 = o(1)$. Note that for $1 \leq i, j \leq K$, $P_{ii} = 1$ and $P_{ij} \geq 0$. It is seen that $|\eta_i| = 1 + o(1)$ and all η_i must have the positive sign. It follows $|\eta_i - 1| = (1 + \eta_i)^{-1}(1 - \eta_i^2) \leq \mu_2$, and so

$$|1 - \eta_i\eta_j| \leq |(1 - \eta_i)(1 - \eta_j)| + |1 - \eta_i| + |1 - \eta_j| \leq C\mu_2. \tag{2.D.92}$$

Combining (2.D.90)-(2.D.92) gives the claim. □

Proof of Lemma 13

Consider the first claim about $\sum_i \theta_i\pi_i(k)$. Write $X = \sum_{i=1}^n \theta_i(\pi_i(k) - h_k)$. It is seen that X is sum of independent mean-zero random variables, where $\theta_i|\pi_i(k) - h_k| \leq C\theta_{\max}$ and $\sum_{i=1}^n \mathrm{Var}(\theta_i(\pi_i(k) - h_k)) \leq C\|\theta\|^2$. By Bernstein's inequality, for any $t > 0$,

$$\mathbb{P}(|X| > t) \leq \exp\Big(-\frac{t^2}{C\|\theta\|^2 + C\theta_{\max}t}\Big).$$

It follows that, with probability $1 - \|\theta\|_1^{-1}$,

$$\Big|\sum_i \theta_i\pi_i(k) - h_k\|\theta\|_1\Big| = |X| \leq C\|\theta\|\sqrt{\log(\|\theta\|_1)} + C\theta_{\max}\log(\|\theta\|_1).$$

Since $\|\theta\| \to \infty$, $\theta_{\max} \to 0$, and $(\|\theta\|^2/\|\theta\|_1)\sqrt{\log(\|\theta\|_1)} \to 0$, the right hand side is $o(\|\theta\|_1)$. Combining it with the assumption of $\min_k\{h_k\} \geq C$, we have

$$\sum_i \theta_i \pi_i(k) \geq C\|\theta\|_1, \qquad \text{with probability } 1 - \|\theta\|^{-1} = 1 - o(1).$$

Additionally, since $\pi_i(k) \leq 1$, $\sum_i \theta_i \pi_i(k) \leq \|\theta\|_1$. Therefore, with probability $1 - o(1)$, each $\sum_i \theta_i \pi_i(k)$ is at the order of $\|\theta\|_1$. This proves the first claim.

Consider the second claim about G. Let $y_i = \pi_i - h$. We can write

$$\|\theta\|^2 G = \sum_{i=1}^n \theta_i^2 \pi_i \pi_i' = \|\theta\|^2(hh') + \underbrace{\sum_{i=1}^n \theta_i^2 y_i y_i'}_{\equiv Y} + \underbrace{\sum_{i=1}^n \theta_i^2 h y_i'}_{Z_1} + \underbrace{\sum_{i=1}^n \theta_i^2 y_i h'}_{\equiv Z_2}.$$

Note that $\mathbb{E}[y_i y_i'] = \Sigma$. Then, $Y - \|\theta\|^2\Sigma = \sum_i \theta_i^2(y_i y_i' - \Sigma)$ is sum of independent, mean-zero random matrices, where $\theta_i^2\|y_i y_i' - \Sigma\| \leq C\theta_i^2$. Using the matrix Hoeffding inequality Tropp (2012), $\mathbb{P}\big(\|Y - \|\theta\|^2\Sigma\| > t\big) \leq \exp\big(-\frac{t^2}{C\|\theta\|_4^4}\big)$, for any $t > 0$. With $t = \|\theta\|^{-1}$, we have

$$\big\|Y - \|\theta\|^2\Sigma\big\| \leq C\|\theta\|_4^2\sqrt{\log(\|\theta\|)}, \qquad \text{with probability } 1 - \|\theta\|^{-1}.$$

Similarly, we can apply matrix Hoeffding inequality to Z_1 and Z_2. It gives

$$\|Z_1 + Z_2\| \leq C\|\theta\|_4^2\sqrt{\log(\|\theta\|)}, \qquad \text{with probability } 1 - \|\theta\|^{-1}.$$

Since $\|\theta\|_4^2 \leq \theta_{\max}\|\theta\| \ll \|\theta\|^2$, it follows that, with probability $1 - o(1)$,

$$\big\|Y + Z_1 + Z_2 - \|\theta\|^2\Sigma\big\| = o(\|\theta\|^2).$$

At the same time, $\lambda_{\min}(\|\theta\|^2\Sigma) = \|\theta\|^2 \cdot \|\Sigma^{-1}\|^{-1} \geq C\|\theta\|^2$. We thus have, with probability $1 - o(1)$,

$$\lambda_{\min}(\|\theta\|^2 G) \geq \lambda_{\min}(Y + Z_1 + Z_2) \geq \lambda_{\min}(\|\theta\|^2\Sigma) - \big\|Y + Z_1 + Z_2 - \|\theta\|^2\Sigma\big\| \geq C\|\theta\|^2.$$

This guarantees $\|G^{-1}\| \leq C$. $\square$

Proof of Lemma 14

Let $Q = P - 1_K 1_K'$, and introduce $d \in \mathbb{R}^K$ such that $D = \text{diag}(d)$. By Lemma 12, $\|Q\| \leq C|\mu_2|$. With these notations,

$$DPD - 1_K 1_K' = dd' + DQD - 1_K 1_K'. \tag{2.D.93}$$

Using the same notations, the assumption $DPD\widetilde{h}_D = 1_K$ can be written as $D(1_K 1_K' + Q)D\widetilde{h}_D = 1_K$. It implies

$$1_K = (d'\widetilde{h}_D)d + DQD\widetilde{h}_D. \tag{2.D.94}$$

We multiply $\widetilde{h}_D'$ on both sides and notice that $1_K'\widetilde{h}_D = 1$. It gives

$$(d'\widetilde{h}_D)^2 = 1 - \widetilde{h}_D' DQD\widetilde{h}_D. \tag{2.D.95}$$

Combining (2.D.94)-(2.D.95) gives

$$\begin{aligned} dd' - 1_K 1_K' &= [1 - (d'\widetilde{h}_D)^2]dd' - (d'\widetilde{h}_D)(DQD\widetilde{h}_D d + d\widetilde{h}_D DQD) - DQD\widetilde{h}_D\widetilde{h}_D' DQD \\ &= (\widetilde{h}_D' DQD\widetilde{h}_D) \cdot dd' - (d'\widetilde{h}_D)(DQD\widetilde{h}_D d + d\widetilde{h}_D DQD) - DQD\widetilde{h}_D\widetilde{h}_D' DQD. \end{aligned}$$

Since $\|\widetilde{h}_D\| \leq C$ and $\|d\| \leq C$, we immediately have

$$\|dd' - 1_K 1_K'\| \leq C\|Q\| \leq C|\mu_2|.$$

Plugging it into (2.D.93) gives

$$\|DPD - 1_K 1_K'\| \leq C\|Q\| \leq C|\mu_2|.$$

□

2.E Properties of Signed Polygon statistics

We prove Table 3.3 and Theorem 2.4.3. Recall the following notations:

$$\widetilde{\Omega} = \Omega - (\eta^*)(\eta^*)', \qquad \text{where} \quad \eta^* = \frac{1}{\sqrt{v_0}}\Omega\mathbf{1}_n, \;\; v_0 = \mathbf{1}_n'\Omega\mathbf{1}_n;$$

$$\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i), \qquad \text{where} \quad \eta = \frac{1}{\sqrt{v}}(\mathbb{E}A)\mathbf{1}_n, \;\; \widetilde{\eta} = \frac{1}{\sqrt{v}}A\mathbf{1}_n, \;\; v = \mathbf{1}_n'(\mathbb{E}A)\mathbf{1}_n;$$

$$r_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) - (\eta_i - \widetilde{\eta}_i)(\eta_j - \widetilde{\eta}_j) + (1 - \frac{v}{V})\widetilde{\eta}_i\widetilde{\eta}_j, \qquad \text{where } V = \mathbf{1}_n'A\mathbf{1}_n.$$

Then, the Ideal SgnQ statistic equals to

$$\widetilde{Q}_n = \sum_{i,j,k,\ell(dist)} (\widetilde{\Omega}_{ij} + W_{ij})(\widetilde{\Omega}_{jk} + W_{jk})(\widetilde{\Omega}_{k\ell} + W_{k\ell})(\widetilde{\Omega}_{\ell i} + W_{\ell i}),$$

the Proxy SgnQ statistic equals to

$$Q_n^* = \sum_{i,j,k,\ell(dist)} (\widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij})(\widetilde{\Omega}_{jk} + W_{jk} + \delta_{jk})(\widetilde{\Omega}_{k\ell} + W_{k\ell} + \delta_{k\ell})(\widetilde{\Omega}_{\ell i} + W_{\ell i} + \delta_{\ell i}),$$

and the SgnQ statistic equals to

$$Q_n = \sum_{i,j,k,\ell(dist)} (\widetilde{\Omega}_{ij}+W_{ij}+\delta_{ij}+r_{ij})(\widetilde{\Omega}_{jk}+W_{jk}+\delta_{jk}+r_{jk})(\widetilde{\Omega}_{k\ell}+W_{k\ell}+\delta_{k\ell}+r_{k\ell})(\widetilde{\Omega}_{\ell i}+W_{\ell i}+\delta_{\ell i}+r_{\ell i}).$$

As explained in Section 2.4, each of $\widetilde{Q}_n, Q_n^*, Q_n$ is the sum of a finite number of post-expansion sums, each having the form

$$\sum_{i,j,k,\ell(dist)} a_{ij}b_{jk}c_{k\ell}d_{\ell i}, \tag{2.E.96}$$

where a_{ij} equals to one of $\{\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, r_{ij}\}$; same for b_{ij}, c_{ij} and d_{ij}. Let $N_{\widetilde{\Omega}}$ be the (common) number of $\widetilde{\Omega}$ terms in each product; similarly, we define N_W, N_δ, N_r. These numbers satisfy $N_{\widetilde{\Omega}} + N_W + N_\delta + N_r = 4$. For example, for the post-expansion sum $\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}W_{k\ell}W_{\ell i}$, $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_r) = (1, 3, 0, 0)$. In Section 2.E.1, we study $\widetilde{Q}_n$, and it involves these post-expansion sums such that

$$N_\delta = N_r = 0,$$

In Section 2.E.2, we study $(Q_n^* - \widetilde{Q}_n)$, which involves post-expansion sums such that

$$N_\delta > 0, \text{ and } N_r = 0,$$

In Section 2.E.3, we study $(Q_n - Q_n^*)$, which is related to the sums such that

$$N_r > 0.$$

2.E.1 Analysis of Table 3.2, proof of Theorem 2.4.1

Define

$$X_1 = \sum_{i,j,k,\ell(dist)} W_{ij}W_{jk}W_{k\ell}W_{\ell i}, \qquad X_2 = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}W_{k\ell}W_{\ell i},$$

$$X_3 = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i}, \qquad X_4 = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i},$$

$$X_5 = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}, \qquad X_6 = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}.$$

We first consider the null hypothesis. Since $\widetilde{\Omega}$ is a zero matrix, it is not hard to see that

$$\widetilde{Q}_n = X_1.$$

The following lemmas are proved in Section 2.E.4.

Lemma 15. *Suppose the conditions of Theorem 2.4.1 hold. Under the null hypothesis, as* $n \to \infty$, $\mathbb{E}[\widetilde{Q}_n] = 0$ *and* $\mathrm{Var}(\widetilde{Q}_n) = 8\|\theta\|^8 \cdot [1 + o(1)]$.

Lemma 16. *Suppose the conditions of Theorem 2.4.1 hold. Under the null hypothesis, as* $n \to \infty$,

$$\frac{\widetilde{Q}_n - E[\widetilde{Q}_n]}{\sqrt{\mathrm{Var}(\widetilde{Q}_n)}} \longrightarrow N(0,1), \qquad \textit{in law.}$$

We then consider the alternative hypothesis. By elementary algebra,

$$\widetilde{Q}_n = X_1 + 4X_2 + 4X_3 + 2X_4 + 4X_5 + X_6.$$

The following lemma characterizes the asymptotic mean and variance of X_1-X_6 under the alternative hypothesis. It gives rise to Columns 5-6 of Table 3.2.

Lemma 17 (Table 3.2)**.** *Suppose conditions of Theorem 2.4.1 hold. Write* $\alpha = |\lambda_2|/\lambda_1$. *Under the alternative hypothesis, as* $n \to \infty$,

- $\mathbb{E}[X_k] = 0$ *for* $1 \le k \le 5$, *and* $\mathbb{E}[X_6] = \mathrm{tr}(\widetilde{\Omega}^4) \cdot [1 + o(1)]$.
- $C^{-1}\|\theta\|^8 \le \mathrm{Var}(X_1) \le C\|\theta\|^8$.
- $\mathrm{Var}(X_2) \le C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $\mathrm{Var}(X_3) \le C\alpha^4\|\theta\|^6\|\theta\|_3^6 = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$.
- $\mathrm{Var}(X_4) \le C\alpha^4\|\theta\|_3^{12} = o(\|\theta\|^8)$.
- $\mathrm{Var}(X_5) \le C\alpha^6\|\theta\|^8\|\theta\|_3^6$.

As a result, $\mathbb{E}[\widetilde{Q}_n] \sim \mathrm{tr}(\widetilde{\Omega}^4)$ *and* $\mathrm{Var}(\widetilde{Q}_n) \le C(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6)$.

Theorem 2.4.1 follows directly from Lemmas 15-17.

2.E.2 Analysis of Table 3.3, proof of Theorem 2.4.2

We introduce U_a, U_b and U_c such that

$$Q_n^* - \widetilde{Q}_n = U_a + U_b + U_c,$$

where U_a, U_b and U_c contain post-expansion sums (2.E.96) with $N_\delta = 1$, $N_\delta = 2$, and $N_\delta \ge 3$, respectively.

First, we consider the post-expansion sums with $N_\delta = 1$. Define

$$U_a = 4Y_1 + 8Y_2 + 4Y_3 + 8Y_4 + 4Y_5 + 4Y_6, \tag{2.E.97}$$

where

$$Y_1 = \sum_{i,j,k,\ell(dist)} \delta_{ij} W_{jk} W_{k\ell} W_{\ell i}, \qquad Y_2 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} W_{\ell i},$$
$$Y_3 = \sum_{i,j,k,\ell(dist)} \delta_{ij} W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}, \qquad Y_4 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i},$$
$$Y_5 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} \widetilde{\Omega}_{\ell i}, \qquad Y_6 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}.$$

Under the null hypothesis, only Y_1 is nonzero, and

$$U_a = 4Y_1.$$

Lemma 18. *Suppose the conditions of Theorem 2.4.1 hold. Under the null hypothesis, as $n \to \infty$, $\mathbb{E}[U_a] = 0$ and $\mathrm{Var}(U_a) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$.*

Under the alternative hypothesis, the following lemma characterizes the asymptotic means and variances of Y_1-Y_6. It gives rise to Rows 1-6 of Table 3.3 and is proved in Section 2.E.4.

Lemma 19 (Table 3.3, Rows 1-6)**.** *Suppose the conditions of Theorem 2.4.1 hold. Let $\alpha = |\lambda_2|/\lambda_1$. Under the alternative hypothesis, as $n \to \infty$,*

- $\mathbb{E}[Y_k] = 0$ *for* $k \in \{1, 2, 3, 5, 6\}$*, and* $|\mathbb{E}[Y_4]| \leq C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$.
- $\mathrm{Var}(Y_1) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $\mathrm{Var}(Y_2) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $\mathrm{Var}(Y_3) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $\mathrm{Var}(Y_4) \leq \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$.
- $\mathrm{Var}(Y_5) \leq \frac{C\alpha^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $\mathrm{Var}(Y_6) \leq \frac{C\alpha^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1} = O(\alpha^6\|\theta\|^8\|\theta\|_3^6)$.

As a result, $\mathbb{E}[U_a] = o(\alpha^4\|\theta\|^8)$ *and* $\mathrm{Var}(U_a) \leq C\alpha^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8)$.

Next, we consider the post-expansion sums with $N_\delta = 2$. Define

$$U_b = 4Z_1 + 2Z_2 + 8Z_3 + 4Z_4 + 4Z_5 + 2Z_6, \tag{2.E.98}$$

where

$$Z_1 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \delta_{jk} W_{k\ell} W_{\ell i}, \qquad Z_2 = \sum_{i,j,k,\ell(dist)} \delta_{ij} W_{jk} \delta_{k\ell} W_{\ell i},$$
$$Z_3 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}, \qquad Z_4 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} W_{\ell i},$$
$$Z_5 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}, \qquad Z_6 = \sum_{i,j,k,\ell(dist)} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}.$$

Under the null hypothesis, only Z_1 and Z_2 are nonzero, and

$$U_b = 4Z_1 + 2Z_2.$$

Lemma 20. *Suppose the conditions of Theorem 2.4.1 hold. Under the null hypothesis, as $n \to \infty$,*

- $\mathbb{E}[Z_1] = \|\theta\|^4 \cdot [1 + o(1)]$, *and* $\mathrm{Var}(Z_1) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $\mathbb{E}[Z_2] = 2\|\theta\|^4 \cdot [1 + o(1)]$, *and* $\mathrm{Var}(Z_2) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.

As a result, $\mathbb{E}[U_b] \sim 8\|\theta\|^4$ *and* $\mathrm{Var}(U_b) = o(\|\theta\|^8)$.

Under the alternative hypothesis, the following lemma provides the asymptotic means and variances of Z_1-Z_6. It gives rise to Rows 7-12 of Table 3.3:

Lemma 21 (Table 3.3, Rows 7-12)**.** *Suppose conditions of Theorem 2.4.1 hold. Write $\alpha = |\lambda_2|/\lambda_1$. Under the alternative hypothesis, as $n \to \infty$,*

- $|\mathbb{E}[Z_1]| \leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(Z_1) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $|\mathbb{E}[Z_2]| \leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(Z_2) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $\mathbb{E}Z_3 = 0$, *and* $\mathrm{Var}(Z_3) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$.
- $|\mathbb{E}[Z_4]| \leq C\alpha\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(Z_4) \leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $|\mathbb{E}[Z_5]| \leq C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(Z_5) \leq \frac{C\alpha^4\|\theta\|^{14}}{\|\theta\|_1^2} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$.
- $|\mathbb{E}[Z_6]| \leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2} = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(Z_6) \leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8)$.

As a result, $\mathbb{E}[U_b] = o(\alpha^4\|\theta\|^8)$ *and* $\mathrm{Var}(U_b) = o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6)$.

Last, we consider the post-expansion sums with $N_\delta \geq 3$. Define

$$U_c = 4T_1 + 4T_2 + F, \tag{2.E.99}$$

where

$$T_1 = \sum_{i,j,k,\ell (dist)} \delta_{ij}\delta_{jk}\delta_{k\ell}W_{\ell i}, \qquad T_2 = \sum_{i,j,k,\ell (dist)} \delta_{ij}\delta_{jk}\delta_{k\ell}\widetilde{\Omega}_{\ell i},$$
$$F = \sum_{i,j,k,\ell (dist)} \delta_{ij}\delta_{jk}\delta_{k\ell}\delta_{\ell i}.$$

Under the null hypothesis, only T_1 and F are nonzero, and

$$U_b = 4T_1 + F.$$

Lemma 22. *Suppose the conditions of Theorem 2.4.1 hold. Under the null hypothesis, as $n \to \infty$,*

- $\mathbb{E}[T_1] = -2\|\theta\|^4 \cdot [1 + o(1)]$, *and* $\mathrm{Var}(T_1) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $|\mathbb{E}[F]| = 2\|\theta\|^4 \cdot [1 + o(1)]$, *and* $\mathrm{Var}(F) \leq \frac{C\|\theta\|^{10}}{\|\theta\|_1^2} = o(\|\theta\|^8)$.

As a result, $\mathbb{E}[U_c] \sim -6\|\theta\|^4$ *and* $\mathrm{Var}(U_c) = o(\|\theta\|^8)$.

Under the alternative hypothesis, the next lemma studies the asymptotic means and variances of T_1, T_2 and F. It gives rise to Rows 13-15 of Table 3.3:

Lemma 23 (Table 3.3, Rows 13-15)**.** *Suppose conditions of Theorem 2.4.1 hold. Write* $\alpha = |\lambda_2|/\lambda_1$. *Under the alternative hypothesis, as* $n \to \infty$,

- $|\mathbb{E}[T_1]| \leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(T_1) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $|\mathbb{E}[T_2]| \leq \frac{C\alpha\|\theta\|^6}{\|\theta\|_1^3} = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(T_2) \leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$.
- $|\mathbb{E}[F]| \leq C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$, *and* $\mathrm{Var}(F) \leq \frac{C\|\theta\|^{10}}{\|\theta\|_1^2} = o(\|\theta\|^8)$.

As a result, $\mathbb{E}|U_c| = o(\alpha^4\|\theta\|^8)$ *and* $\mathrm{Var}(U_c) = o(\|\theta\|^8)$.

We now prove Theorem 2.4.2. Since $Q_n^* - \widetilde{Q}_n = U_a + U_b + U_c$, we have

$$\mathbb{E}[Q_n^* - \widetilde{Q}_n] = \mathbb{E}[U_a] + \mathbb{E}[U_b] + \mathbb{E}[U_c],$$
$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq 3\mathrm{Var}(U_a) + 3\mathrm{Var}(U_b) + 3\mathrm{Var}(U_c).$$

Consider the null hypothesis. By Lemmas 18, 20, 22,

$$\mathbb{E}[Q_n^* - \widetilde{Q}_n] = 0 + 8\|\theta\|^4 - 6\|\theta\|^4 + o(\|\theta\|^4) \sim 2\|\theta\|^4,$$

and

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C\|\theta\|^2\|\theta\|_3^6 + \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} + \frac{C\|\theta\|^{10}}{\|\theta\|_1^2}.$$

Using the universal inequality $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, we further have

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8),$$

where $\|\theta\|_3^3 = o(\|\theta\|^2)$ and $\|\theta\| \to \infty$ in our range of interest. This proves claims for the null hypothesis. Consider the alternative hypothesis. By Lemmas 19, 21, 23,

$$\left|\mathbb{E}[Q_n^* - \widetilde{Q}_n]\right| \leq C\alpha^2\|\theta\|^6,$$

where the main contributors are Y_4 and Z_5. Since $\alpha\|\theta\| \to \infty$ in our range of interest, the above is $o(\alpha^4\|\theta\|^8)$. By Lemmas 19, 21, 23,

$$\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq \frac{C\alpha^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1},$$

where the main contributor is Y_6. Using the universal inequality of $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the above is $O(\alpha^6\|\theta\|^8\|\theta\|_3^6)$. This proves claims for the alternative hypothesis.

2.E.3 Analysis of $(Q_n - Q_n^*)$, proof of Theorem 2.4.3

By definition, $(Q_n - Q_n^*)$ expands to the sum of 175 post-expansion sums, where each has the form (2.E.96) and satisfies $N_r > 0$. Recall that

$$r_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) - (\eta_i - \widetilde{\eta}_i)(\eta_j - \widetilde{\eta}_j) + (1 - \frac{v}{V})\widetilde{\eta}_i\widetilde{\eta}_j.$$

Since $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$, we have $\widetilde{\eta}_i\widetilde{\eta}_j = \eta_i\eta_j - \delta_{ij} + (\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)$. Inserting it into the definition of r_{ij} gives

$$r_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) + (1 - \frac{v}{V})\eta_i\eta_j - (1 - \frac{v}{V})\delta_{ij} - \frac{v}{V}(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j). \tag{2.E.100}$$

Define

$$\widetilde{r}_{ij} = -\frac{v}{V}(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j), \qquad \epsilon_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) + (1 - \frac{v}{V})\eta_i\eta_j - (1 - \frac{v}{V})\delta_{ij}.$$

Then, we can write

$$r_{ij} = \widetilde{r}_{ij} + \epsilon_{ij}. \tag{2.E.101}$$

Using this notation, we re-write

$$Q_n = \sum_{i,j,k,\ell (dist)} M_{ij}M_{jk}M_{k\ell}M_{\ell i}, \qquad \text{where } M_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + \widetilde{r}_{ij} + \epsilon_{ij},$$

and

$$Q_n^* = \sum_{i,j,k,\ell (dist)} M_{ij}^*M_{jk}^*M_{k\ell}^*M_{\ell i}^*, \qquad \text{where } M_{ij}^* \equiv \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}.$$

We then introduce an intermediate variable:

$$\widetilde{Q}_n^* = \sum_{i,j,k,\ell (dist)} \widetilde{M}_{ij}^*\widetilde{M}_{jk}^*\widetilde{M}_{k\ell}^*\widetilde{M}_{\ell i}^*, \quad \text{where } \widetilde{M}_{ij}^* = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + \widetilde{r}_{ij}. \tag{2.E.102}$$

As a result, $(Q_n - Q_n^*)$ decomposes into

$$Q_n - Q_n^* = (\widetilde{Q}_n^* - Q_n^*) + (Q_n - \widetilde{Q}_n^*). \tag{2.E.103}$$

We note that Q_n can be expanded to the sum of $5^4 = 625$ post-expansion sums, each with the form

$$\sum_{i,j,k,\ell (dist)} a_{ij}b_{jk}c_{k\ell}d_{\ell i},$$

where each of $a_{ij}, b_{ij}, c_{ij}, d_{ij}$ takes values in $\{\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, \widetilde{r}_{ij}, \epsilon_{ij}\}$. Let $N_{\widetilde{\Omega}}$ be the (common) number of $\widetilde{\Omega}$ terms in each product and define $N_W, N_\delta, N_{\widetilde{r}}, N_\epsilon$ similarly. Among the 625 post-expansion sums,

- $3^4 = 81$ of them are contained in Q_n^*,
- $4^4 - 3^4 = 175$ of them are contained in $(\widetilde{Q}_n^* - Q_n^*)$,
- and $5^4 - 4^4 = 369$ of them are contained in $(Q_n - \widetilde{Q}_n^*)$.

We shall study $(\widetilde{Q}_n^* - Q_n^*)$ and $(Q_n - \widetilde{Q}_n^*)$, separately.

In our analysis, one challenge is to deal with the random variable V that appears in the denominator in the expression of r_{ij}. The following lemma is useful and proved in Section 2.E.4.

Lemma 24. *Suppose conditions of Theorem 2.4.3 hold. As $n \to \infty$, for any sequence x_n such that $\sqrt{\log(\|\theta\|_1)} \ll x_n \ll \|\theta\|_1$,*

$$\mathbb{E}\big[(\widetilde{Q}_n - Q_n)^2 \cdot I\{|V - v| > \|\theta\|_1 x_n\}\big] \to 0.$$

The next two lemmas are proved in Section 2.E.4.

Lemma 25. *Suppose conditions of Theorem 2.4.3 hold. Write $\alpha = |\lambda_2|/\lambda_1$. As $n \to \infty$,*

- *Under the null hypothesis, $|\mathbb{E}[\widetilde{Q}_n^* - Q_n^*]| = o(\|\theta\|^4)$ and $\mathrm{Var}(\widetilde{Q}_n^* - Q_n^*) = o(\|\theta\|^8)$.*
- *Under the alternative hypothesis, $|\mathbb{E}[\widetilde{Q}_n^* - Q_n^*]| = o(\alpha^4\|\theta\|^8)$ and $\mathrm{Var}(\widetilde{Q}_n^* - Q_n^*) = o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6)$.*

Lemma 26. *Suppose conditions of Theorem 2.4.3 hold. Write $\alpha = |\lambda_2|/\lambda_1$. As $n \to \infty$,*

- *Under the null hypothesis, $|\mathbb{E}[Q_n - \widetilde{Q}_n^*]| = o(\|\theta\|^4)$ and $\mathrm{Var}(Q_n - \widetilde{Q}_n^*) = o(\|\theta\|^8)$.*
- *Under the alternative hypothesis, $|\mathbb{E}[Q_n - \widetilde{Q}_n^*]| = o(\alpha^4\|\theta\|^8)$ and $\mathrm{Var}(\widetilde{Q}_n^* - Q_n^*) = O(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6)$.*

Theorem 2.4.3 follows directly from (2.E.103) and Lemmas 25-26.

2.E.4 Proof of Lemmas 15-26

Proof of Lemma 15

Under the null hypothesis,

$$\widetilde{Q}_n = X_1 = \sum_{i,j,k,\ell (dist)} W_{ij}W_{jk}W_{k\ell}W_{\ell i}.$$

For mutually distinct indices (i,j,k,ℓ), $(W_{ij}, W_{jk}, W_{k\ell}, W_{\ell i})$ are independent of each other, each with mean zero. So $\mathbb{E}[W_{ij}W_{jk}W_{k\ell}W_{\ell i}] = 0$. It follows that

$$\mathbb{E}[\widetilde{Q}_n] = 0.$$

We now calculate the variance of $\widetilde{Q}_n$. Under the null hypothesis, $\Omega_{ij} = \theta_i\theta_j$; hence, $\mathrm{Var}(W_{ij}) = \Omega_{ij}(1-\Omega_{ij}) = \theta_i\theta_j - \theta_i^2\theta_j^2 = \theta_i\theta_j[1 + O(\theta_{\max}^2)]$. It follows that

$$\begin{aligned}\mathrm{Var}(W_{ij}W_{jk}W_{k\ell}W_{\ell i}) &= \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 \cdot [1 + O(\theta_{\max}^2)]^4 \\ &= \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 \cdot [1 + O(\theta_{\max}^2)]. \end{aligned} \tag{2.E.104}$$

Note that each (i,j,k,ℓ) corresponds to a 4-cycle in a complete graph of n nodes. For (i,j,k,ℓ) and (i',j',k',ℓ'), we can write $W_{ij}W_{jk}W_{k\ell}W_{\ell i} \cdot W_{i'j'}W_{j'k'}W_{k'\ell'}W_{\ell' i'}$ in the form of $\prod_t (W_{i_tj_t})^{m_t}$, where $\{W_{i_tj_t}\}$ are mutually distinct with each other and m_t is the number of times that $W_{i_tj_t}$ appears in this product. If the two 4-cycles corresponding to (i,j,k,ℓ) and (i',j',k',ℓ') are not exactly overlapping, then at least two of m_t equals to 1. As a result, the mean of $\prod_t (W_{i_tj_t})^{m_t}$ is zero. In other words, we have argued that

$$\begin{aligned}&\mathrm{Cov}(W_{ij}W_{jk}W_{k\ell}W_{\ell i},\ W_{i'j'}W_{j'k'}W_{k'\ell'}W_{\ell' i'}) = 0 \text{ if the} \\ &\text{two cycles corresponding to } (i,j,k,\ell) \text{ and } (i',j',k',\ell') \\ &\text{are not exactly overlapping.}\end{aligned} \tag{2.E.105}$$

In the sum over all distinct (i,j,k,ℓ), each 4-cycle is repeatedly counted by 8 times

$$(i,j,k,\ell),(j,k,\ell,i),(k,\ell,i,j),(\ell,i,j,k),$$
$$(\ell,k,j,i),(k,j,i,\ell),(j,i,\ell,k),(i,\ell,k,j).$$

It follows that

$$\begin{aligned}
\mathrm{Var}(\widetilde{Q}_n) &= \mathrm{Var}\Bigg(8\sum_{\substack{\text{unique}\\ \text{4-cycles}}} W_{ij}W_{jk}W_{k\ell}W_{\ell i}\Bigg)\\
&= 64\cdot\mathrm{Var}\Bigg(\sum_{\substack{\text{unique}\\ \text{4-cycles}}} W_{ij}W_{jk}W_{k\ell}W_{\ell i}\Bigg)\\
&= 64\sum_{\substack{\text{unique}\\ \text{4-cycles}}} \mathrm{Var}\big(W_{ij}W_{jk}W_{k\ell}W_{\ell i}\big)\\
&= 8\sum_{i,j,k,\ell(dist)} \mathrm{Var}\big(W_{ij}W_{jk}W_{k\ell}W_{\ell i}\big)\\
&= [1+O(\theta_{\max}^2)]\cdot 8\sum_{i,j,k,\ell(dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2,
\end{aligned} \tag{2.E.106}$$

where the third line is from (2.E.105) and the last line is from (2.E.104). We then compute the right hand side of (2.E.106). Note that

$$\sum_{i,j,k,\ell(dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 = \|\theta\|^8 - \sum_{i,j,k,\ell(not\ dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2,$$

where

$$\sum_{i,j,k,\ell(not\ dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 \le \binom{4}{2}\sum_{i,j,k}\theta_i^2\theta_j^2\theta_k^4 \le C\|\theta\|^4\|\theta\|_4^4 = \|\theta\|^8\cdot O\Big(\frac{\|\theta\|_4^4}{\|\theta\|^4}\Big).$$

Combining the above gives

$$\sum_{i,j,k,\ell(dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 = \|\theta\|^8\cdot\Big[1+O\Big(\frac{\|\theta\|_4^4}{\|\theta\|^4}\Big)\Big]. \tag{2.E.107}$$

We combine (2.E.106)-(2.E.107) and note that $\theta_{\max}=o(1)$ and $\|\theta\|_4^4/\|\theta\|^4 \le (\|\theta\|^2\theta_{\max}^2)/\|\theta\|^4 = o(1)$. So,

$$\mathrm{Var}(\widetilde{Q}_n) = 8\|\theta\|^8\cdot[1+o(1)].$$

This completes the proof.

Proof of Lemma 16

Under the null hypothesis,

$$\widetilde{Q}_n = X_1 = \sum_{i,j,k,\ell(dist)} W_{ij}W_{jk}W_{k\ell}W_{\ell i}.$$

In the proof of Theorem 3.2 of Jin et al. (2018), it has been shown that $X_1/\sqrt{\mathrm{Var}(X_1)} \to N(0,1)$ in law (in the proof there, $X_1/\sqrt{\mathrm{Var}(X_1)}$ is denoted as $S_{n,n}$). Since $\mathbb{E}[X_1]=0$, we can directly quote their results to get the desired claim.

Proof of Lemma 17

We shall study the mean and variance of each of X_1-X_6 and then combine those results.

Consider X_1. We have analyzed this term under the null hypothesis. Under the alternative hypothesis, the difference is that we no longer have $\Omega_{ij} = \theta_i\theta_j$. Instead, we have an upper bound $\Omega_{ij} = \theta_i\theta_j(\pi_i' P\pi_j) \leq C\theta_i\theta_j$. Using similar proof as that for the null hypothesis, we can derive that

$$\mathbb{E}[X_1] = 0, \qquad \mathrm{Var}(X_1) \leq C\|\theta\|^8. \tag{2.E.108}$$

To get a lower bound for $\mathrm{Var}(X_1)$, we notice that $\mathrm{Var}(W_{ij}) = \Omega_{ij}(1-\Omega_{ij}) \geq \Omega_{ij}[1 - O(\theta_{\max}^2)] \geq \Omega_{ij}/2$; this inequality is true even when $\Omega_{ij} = 0$. It follows that

$$\mathrm{Var}(W_{ij}W_{jk}W_{k\ell}W_{\ell i}) \geq \frac{1}{16}\Omega_{ij}\Omega_{jk}\Omega_{k\ell}\Omega_{\ell i}.$$

Note that the second last line of (2.E.106) is still true. As a result,

$$\begin{aligned}
\mathrm{Var}(X_1) &= 8 \sum_{i,j,k,\ell (dist)} \mathrm{Var}\big(W_{ij}W_{jk}W_{k\ell}W_{\ell i}\big)\\
&\geq \frac{1}{2} \sum_{i,j,k,\ell (dist)} \Omega_{ij}\Omega_{jk}\Omega_{k\ell}\Omega_{\ell i}\\
&= \frac{1}{2}\,\mathrm{tr}(\Omega^4) - \frac{1}{2} \sum_{i,j,k,\ell (not\ dist)} \Omega_{ij}\Omega_{jk}\Omega_{k\ell}\Omega_{\ell i}\\
&\geq \frac{1}{2}\,\mathrm{tr}(\Omega^4) - C \sum_{i,j,k,\ell (not\ dist)} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\\
&\geq \frac{1}{2}\,\mathrm{tr}(\Omega^4) - o(\|\theta\|^8),
\end{aligned}$$

where the last inequality is due to (2.E.107). Recall that $\lambda_1, \ldots, \lambda_K$ denote the K nonzero eigenvalues of Ω. By Lemma 6, $\lambda_1 \geq C^{-1}\|\theta\|^2$. It follows that

$$\mathrm{tr}(\Omega^4) = \sum_{k=1}^K \lambda_k^4 \geq \lambda_1^4 \geq C^{-1}\|\theta\|^8.$$

Combining the above gives

$$\mathrm{Var}(X_1) \geq C^{-1}\|\theta\|^8. \tag{2.E.109}$$

So far, we have proved all claims about X_1.

Consider X_2. Recall that

$$X_2 = \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij} W_{jk}W_{k\ell}W_{\ell i}.$$

It is easy to see that $\mathbb{E}[X_2] = 0$. Below, we bound its variance. Each index choice (i,j,k,ℓ) defines a undirected path j-k-ℓ-i in the complete graph of n nodes. If the two paths j-k-ℓ-i and j'-k'-ℓ'-i' are not exactly overlapping, then $W_{jk}W_{k\ell}W_{\ell i} \cdot W_{j'k'}W_{k'\ell'}W_{\ell' i'}$ have mean zero. In the sum above, each unique path j-k-ℓ-i is counted twice as (i,j,k,ℓ) and (j,i,ℓ,k).

Mimicking the argument in (2.E.106), we immediately have

$$\begin{aligned}\mathrm{Var}(X_2) &= 2\sum_{i,j,k,\ell(dist)} \mathrm{Var}\big(\widetilde{\Omega}_{ij}W_{jk}W_{k\ell}W_{\ell i}\big)\\ &= 2\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}^2\cdot \mathrm{Var}\big(W_{jk}W_{k\ell}W_{\ell i}\big).\end{aligned}$$

By Lemma 9, $|\widetilde{\Omega}_{ij}| \leq |\lambda_2|\|\theta\|^{-2}\theta_i\theta_j$. In our notations, $\alpha = |\lambda_2|/\lambda_1$; additionally, by Lemma 6, $\lambda_1 \leq C\|\theta\|^2$. Combining them gives

$$|\widetilde{\Omega}_{ij}| \leq C\alpha\theta_i\theta_j. \tag{2.E.110}$$

Moreover, $\mathrm{Var}(W_{jk}W_{k\ell}W_{\ell i}) \leq \Omega_{jk}\Omega_{k\ell}\Omega_{\ell i} \leq C\theta_j\theta_k^2\theta_\ell^2\theta_i$. It follows that

$$\begin{aligned}\mathrm{Var}(X_2) &\leq C\sum_{i,j,k,\ell(dist)} (\alpha\theta_i\theta_j)^2\cdot\theta_j\theta_k^2\theta_\ell^2\theta_i\\ &\leq C\alpha^2\sum_{i,j,k,\ell}\theta_i^3\theta_j^3\theta_k^2\theta_\ell^2\\ &\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6.\end{aligned}$$

Since $\|\theta\|_3^3 \leq \theta_{\max}\sum_i\theta_i^2 = \theta_{\max}\|\theta\|^2$, the right hand side is $\leq C\alpha^2\|\theta\|^8\theta_{\max}^2$. Note that $\alpha \leq 1$ and $\theta_{\max}\to 0$. So, this term is $o(\|\theta\|^8)$. We have proved all claims about X_2.

Consider X_3. Recall that

$$X_3 = \sum_{i,j,k,\ell(dist)}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} = \sum_{i,k,\ell(dist)}\Big(\sum_{j\notin\{i,k,\ell\}}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\Big)W_{k\ell}W_{\ell i}.$$

It is easy to see that $\mathbb{E}[X_3]=0$. We then study its variance. We note that for $W_{k\ell}W_{\ell i}$ and $W_{k'\ell'}W_{\ell' i'}$ to be correlated, we must have that $(k',\ell',i') = (k,\ell,i)$ or $(k',\ell',i') = (i,\ell,k)$; in other words, the two underlying paths k-ℓ-i and k'-ℓ'-i' have to be equal. Mimicking the argument in (2.E.106), we have

$$\begin{aligned}\mathrm{Var}(X_3) &\leq C\sum_{i,k,\ell(dist)}\mathrm{Var}\Big[\Big(\sum_{j\notin\{i,k,\ell\}}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\Big)W_{k\ell}W_{\ell i}\Big]\\ &\leq C\sum_{i,k,\ell(dist)}\Big(\sum_{j\notin\{i,k,\ell\}}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\Big)^2\cdot\mathrm{Var}(W_{k\ell}W_{\ell i}).\end{aligned}$$

By (2.E.110),

$$\Big|\sum_{j\notin\{i,k,\ell\}}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\Big| \leq C\sum_j\alpha^2\theta_i\theta_j^2\theta_k \leq C\alpha^2\|\theta\|^2\cdot\theta_i\theta_k.$$

Combining the above gives

$$\begin{aligned}\mathrm{Var}(X_3) &\leq C\sum_{i,k,\ell}(\alpha^2\|\theta\|^2\theta_i\theta_k)^2\cdot\theta_k\theta_\ell^2\theta_i\\ &\leq C\alpha^4\|\theta\|^4\sum_{i,k,\ell}\theta_i^3\theta_k^3\theta_\ell^2\\ &\leq C\alpha^4\|\theta\|^6\|\theta\|_3^6.\end{aligned}$$

Since $\|\theta\|\to\infty$, the right hand side is $o(\alpha^4\|\theta\|^8\|\theta\|_3^6)$. We have proved all claims about X_3.

Consider X_4. Recall that

$$X_4 = \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij} W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i} = \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{k\ell} W_{jk} W_{\ell i}.$$

It is easy to see that $\mathbb{E}[X_4] = 0$. To calculate its variance, note that $W_{jk}W_{\ell i}$ and $W_{j'k'}W_{\ell' i'}$ are uncorrelated unless (i) $\{j', k'\} = \{j, k\}$ and $\{\ell', i'\} = \{\ell, i\}$ or (ii) $\{j', k'\} = \{\ell, i\}$ and $\{\ell', i'\} = \{j, k\}$. Mimicking the argument in (2.E.106), we immediately have

$$\begin{aligned}
\mathrm{Var}(X_4) &\le C \sum_{i,j,k,\ell (dist)} \mathrm{Var}\big(\widetilde{\Omega}_{ij} \widetilde{\Omega}_{k\ell} W_{jk} W_{\ell i}\big) \\
&\le C \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij}^2 \widetilde{\Omega}_{k\ell}^2 \cdot \mathrm{Var}(W_{jk} W_{\ell i}) \\
&\le C \sum_{i,j,k,\ell} (\alpha\theta_i\theta_j)^2 (\alpha\theta_k\theta_\ell)^2 \cdot \theta_j\theta_k\theta_\ell\theta_i \\
&\le C\alpha^4 \sum_{i,j,k,\ell} \theta_i^3\theta_j^3\theta_k^3\theta_\ell^3 \\
&\le C\alpha^4 \|\theta\|_3^{12}.
\end{aligned}$$

Since $\|\theta\|_3^3 \le \theta_{\max}\|\theta\|^2 = o(\|\theta\|^2)$, the right hand side is $o(\|\theta\|^8)$. This proves the claims of X_4.

Consider X_5. Recall that

$$X_5 = \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i} = 2 \sum_{i<\ell} \Big(\sum_{\substack{j,k \notin \{i,\ell\} \\ j \neq k}} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \Big) W_{\ell i}.$$

It is easily seen that $\mathbb{E}[X_5] = 0$. Furthermore, we have

$$\mathrm{Var}(X_5) = 2 \sum_{i<\ell} \Big(\sum_{\substack{j,k \notin \{i,\ell\} \\ j \neq k}} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \Big)^2 \cdot \mathrm{Var}(W_{\ell i}). \tag{2.E.111}$$

By (2.E.110),

$$\Big| \sum_{\substack{j,k \notin \{i,\ell\} \\ j \neq k}} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \Big| \le C \sum_{j,k} \alpha^3 \theta_i \theta_j^2 \theta_k^2 \theta_\ell \le C\alpha^3 \|\theta\|^4 \cdot \theta_i \theta_\ell$$

We plug it into (2.E.111) and use $\mathrm{Var}(W_{\ell i}) \le \Omega_{\ell i} \le C\theta_\ell\theta_i$. It yields that

$$\begin{aligned}
\mathrm{Var}(X_5) \quad &\le \quad C \sum_{\ell, i (dist)} (\alpha^3 \|\theta\|^4 \theta_i \theta_\ell)^2 \cdot \theta_\ell \theta_i \\
&\le \quad C\alpha^6 \|\theta\|^8 \sum_{\ell, i} \theta_i^3 \theta_\ell^3 \\
&\le \quad C\alpha^6 \|\theta\|^8 \|\theta\|_3^6.
\end{aligned} \tag{2.E.112}$$

This proves the claims of X_5.

Consider X_6. Recall that

$$X_6 = \sum_{i,j,k,\ell (dist)} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i} = \mathrm{tr}(\widetilde{\Omega}^4) - \sum_{i,j,k,\ell (not\ dist)} \widetilde{\Omega}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}.$$

This is a non-stochastic number, so its variance is zero and its mean is X_6 itself. By Lemma 9, $|\lambda_2| \le \|\widetilde{\Omega}\| \le C|\lambda_2|$. Since $\|\widetilde{\Omega}\|^4 \le \mathrm{tr}(\widetilde{\Omega}^4) \le K\|\widetilde{\Omega}\|^4$, we immediately have

$\mathrm{tr}(\widetilde{\Omega}^4) \asymp \|\widetilde{\Omega}\|^4 \asymp |\lambda_2|^4$. Additionally, $|\lambda_2| = \alpha\lambda_1$ in our notation, and $\lambda_1 \asymp \|\theta\|^2$ by Lemma 6. It follows that

$$\mathrm{tr}(\widetilde{\Omega}^4) \asymp |\lambda_2|^4 \asymp \alpha^4\|\theta\|^8.$$

At the same time, by (2.E.110), $|\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}| \leq C\alpha^4\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2$. We thus have

$$\begin{aligned}|X_6 - \mathrm{tr}(\widetilde{\Omega}^4)| &\leq C\alpha^4 \sum_{i,j,k,\ell(\textit{not dist})} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2 \\ &\leq C\alpha^4 \sum_{i,j,k} \theta_i^2\theta_j^2\theta_k^4 \\ &\leq C\alpha^4\|\theta\|^4\|\theta\|_4^4 = o(\alpha^4\|\theta\|^8),\end{aligned}$$

where the last equality is due to $\|\theta\|_4^4 \leq \theta_{\max}^2\|\theta\|^2 = o(\|\theta\|^4)$. Combining the above gives

$$X_6 = \mathrm{tr}(\widetilde{\Omega}^4)\cdot[1+o(1)].$$

This proves the claims of X_6.

Last, we combine the results for X_1-X_6 to study $\widetilde{Q}_n$. Note that

$$\widetilde{Q}_n = X_1 + 4X_2 + 4X_3 + 2X_4 + 4X_5 + X_6.$$

Only X_6 has a nonzero mean. So,

$$\mathbb{E}[\widetilde{Q}_n] = \mathbb{E}[X_6] = \mathrm{tr}(\widetilde{\Omega}^4)\cdot[1+o(1)].$$

At the same time, given random variables $Z_1, Z_2, \ldots, Z_m$, $\mathrm{Var}(\sum_{k=1}^m Z_k) = \sum_k \mathrm{Var}(Z_k) + \sum_{k\neq\ell}\mathrm{Cov}(Z_k, Z_\ell) \leq \sum_k \mathrm{Var}(Z_k) + \sum_{k\neq\ell}\sqrt{\mathrm{Var}(Z_k)\mathrm{Var}(Z_\ell)} \leq m^2\max_k\{\mathrm{Var}(Z_k)\}$. We thus have

$$\mathrm{Var}(\widetilde{Q}_n) \leq C\max_{1\leq k\leq 6}\mathrm{Var}(X_k) \leq C\big(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6\big).$$

The proof of this lemma is now complete.

Proof of Lemma 18

Recall that $U_a = 4Y_1 = 4\sum_{i,j,k,\ell(dist)}\delta_{ij}W_{jk}W_{k\ell}W_{\ell i}$. By definition, $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$. It follows that

$$U_a = 4\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j - \widetilde{\eta}_j)W_{jk}W_{k\ell}W_{\ell i} + 4\sum_{i,j,k,\ell(dist)}\eta_j(\eta_i - \widetilde{\eta}_i)W_{jk}W_{k\ell}W_{\ell i}.$$

In the second sum, if we relabel $(i,j,k,\ell) = (j', i', \ell', k')$, it becomes

$$4\sum_{i',j',k',\ell'(dist)}\eta_{i'}(\eta_{j'} - \widetilde{\eta}_{j'})W_{i'\ell'}W_{\ell'k'}W_{k'j'} = 4\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j - \widetilde{\eta}_j)W_{i\ell}W_{\ell k}W_{kj},$$

which is the same as the first term. It follows that

$$U_a = 8\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j - \widetilde{\eta}_j)W_{jk}W_{k\ell}W_{\ell i}.$$

By definition, $\eta_j = \frac{1}{\sqrt{v}}\sum_{s\neq j}\mathbb{E}A_{js}$ and $\widetilde{\eta}_j = \frac{1}{\sqrt{v}}\sum_{s\neq j}A_{js}$. Hence,

$$\widetilde{\eta}_j - \eta_j = \frac{1}{\sqrt{v}}\sum_{s\neq j}W_{js}. \tag{2.E.113}$$

We then re-write

$$\begin{aligned} U_a &= -8 \sum_{i,j,k,\ell (dist)} \eta_i \Big(\frac{1}{\sqrt{v}} \sum_{s\neq j} W_{js} \Big) W_{jk} W_{k\ell} W_{\ell i} \\ &= -\frac{8}{\sqrt{v}} \sum_{\substack{i,j,k,\ell (dist) \\ s\neq j}} \eta_i W_{js} W_{jk} W_{k\ell} W_{\ell i}. \end{aligned}$$

In the summand, (i,j,k,ℓ) are distinct, but s is only required to be distinct from j. We consider two different cases: (a) the case of $s=k$, where the summand becomes $W_{jk}^2 W_{k\ell} W_{\ell i}$, and (b) the case of $s\neq k$. Correspondingly, we write

$$\begin{aligned} U_a &= -\frac{8}{\sqrt{v}} \sum_{i,j,k,\ell (dist)} \eta_i W_{jk}^2 W_{k\ell} W_{\ell i} - \frac{8}{\sqrt{v}} \sum_{\substack{i,j,k,\ell (dist) \\ s\notin\{j,k\}}} \eta_i W_{js} W_{jk} W_{k\ell} W_{\ell i} \\ &\equiv U_{a1} + U_{a2}. \end{aligned} \tag{2.E.114}$$

It is easy to see that the summands in both sums have mean zero. Therefore,

$$\mathbb{E}[U_a] = 0.$$

Next, we bound the variance of U_a. Since $\mathrm{Var}(U_a) \le 2\mathrm{Var}(U_{a1}) + 2\mathrm{Var}(U_{a2})$, it suffices to bound the variances of U_{a1} and U_{a2}. Consider U_{a1}. Note that

$$\mathrm{Var}(U_{a1}) = \frac{64}{v} \sum_{\substack{i,j,k,\ell (dist) \\ i',j',k',\ell' (dist)}} \eta_i \eta_{i'} \cdot \mathbb{E}[W_{jk}^2 W_{k\ell} W_{\ell i} W_{j'k'}^2 W_{k'\ell'} W_{\ell' i'}]. \tag{2.E.115}$$

By definition, $v = 1_n'(\mathbb{E}A)1_n = 1_n'\Omega 1_n - \sum_i \Omega_{ii}$. Since $\Omega_{ii} \le \theta_i^2$, it implies $v = 1_n'\Omega 1_n - O(\|\theta\|^2) = 1_n'\Omega 1_n + o(\|\theta\|_1^2)$. Moreover, we note that $1_n'\Omega 1_n \le C\sum_{i,j}\theta_i\theta_j \le C\|\theta\|_1^2$, and by Lemma 8, $1_n'\Omega 1_n \ge C^{-1}\|\theta\|_1^2$. Combining these results gives

$$C^{-1}\|\theta\|_1^2 \le v \le C\|\theta\|_1^2. \tag{2.E.116}$$

Moreover, $\eta_i = \frac{1}{\sqrt{v}}\sum_{s\neq i}\Omega_{is} \le \frac{C}{\|\theta\|_1}\sum_s \theta_i\theta_s$. This gives

$$0 \le \eta_i \le C\theta_i, \qquad \text{for all } 1 \le i \le n. \tag{2.E.117}$$

We plug (2.E.116)-(2.E.117) into (2.E.115) and find out that

$$\mathrm{Var}(U_{a1}) \le \frac{C}{\|\theta\|_1^2} \sum_{\substack{i,j,k,\ell (dist) \\ i',j',k',\ell' (dist)}} \theta_i \theta_{i'} \cdot \mathbb{E}[W_{jk}^2 W_{k\ell} W_{\ell i} W_{j'k'}^2 W_{k'\ell'} W_{\ell' i'}].$$

In order for the summand to be nonzero, all W terms have to be perfectly paired. By elementary calculations,

$$\theta_i\theta_{i'}\mathbb{E}[W_{jk}^2 W_{k\ell} W_{\ell i} W_{j'k'}^2 W_{k'\ell'} W_{\ell' i'}] = \begin{cases} \theta_i^2\, \mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j'k}^2], & \text{if } (\ell',k',i')\text{=}(\ell,k,i); \\ \theta_i\theta_k\, \mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j'i}^2], & \text{if } (\ell',k',i')\text{=}(\ell,i,k); \\ \theta_i\theta_j\, \mathbb{E}[W_{jk}^3 W_{k\ell}^2 W_{\ell i}^3], & \text{if } (j',k')\text{=}(i,\ell),\ (i',\ell')\text{=}(j,k); \\ 0, & \text{otherwise.} \end{cases}$$

Here, (i,j,k,ℓ) are distinct. In the second case above, $(W_{jk}^2, W_{k\ell}^2, W_{\ell i}^2, W_{j'i}^2)$ are independent of each other, no matter $j=j'$ or $j\neq j'$ (we remark that $j'\neq\ell$, because $j'\notin\{i',k',\ell'\} = \{i,k,\ell\}$). It follows that $\mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j'i}^2] \le \Omega_{jk}\Omega_{k\ell}\Omega_{\ell i}\Omega_{j'i} \le C\theta_i^2\theta_j\theta_k^2\theta_\ell^2\theta_{j'}$. In the first case, when $j\neq j'$, $\mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j'k}^2] \le \Omega_{jk}\Omega_{k\ell}\Omega_{\ell i}\Omega_{j'k} \le C\theta_i\theta_j\theta_k^3\theta_\ell^2\theta_{j'}$; when $j=j'$, it holds that $\mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j'k}^2] = \mathbb{E}[W_{jk}^4 W_{k\ell}^2 W_{\ell i}^2] \le C\theta_i\theta_j\theta_k^2\theta_\ell^2$. In the third case, $(W_{jk}^3, W_{k\ell}^2, W_{\ell i}^3)$ are mutually independent, so $\mathbb{E}[W_{jk}^2 W_{k\ell}^2 W_{\ell i}^2] \le \Omega_{jk}\Omega_{k\ell}\Omega_{\ell i} \le C\theta_i\theta_j\theta_k^2\theta_\ell^2$. We then have

$$\theta_i\theta_{i'}\mathbb{E}[W_{jk}^2 W_{k\ell}W_{\ell i}W_{j'k'}^2 W_{k'\ell'}W_{\ell' i'}] \le \begin{cases} C\theta_i^3\theta_j\theta_k^2\theta_\ell^2, & \text{if } (\ell',k',i')=(\ell,k,i),\ j'=j; \\ C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_{j'}, & \text{if } (\ell',k',i')=(\ell,k,i),\ j'\neq j; \\ C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_{j'}, & \text{if } (\ell',k',i')=(\ell,i,k); \\ C\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2, & \text{if } (j',k')=(i,\ell),\ (i',\ell')=(j,k); \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned} \mathrm{Var}(U_{a1}) &\le \frac{C}{\|\theta\|_1^2}\Big(\sum_{i,j,k,\ell}\theta_i^3\theta_j\theta_k^2\theta_\ell^2 + \sum_{i,j,k,\ell,j'}\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_{j'} + \sum_{i,j,k,\ell}\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\Big) \\ &\le \frac{C}{\|\theta\|_1^2}\big(\|\theta\|^4\|\theta\|_3^3\|\theta\|_1 + \|\theta\|^2\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^8\big) \\ &\le C\|\theta\|^2\|\theta\|_3^6, \end{aligned} \tag{2.E.118}$$

where we obtain the last inequality as follows: By Cauchy-Schwarz inequality, $\|\theta\|^4 = (\sum_i \theta_i^{1/2}\cdot\theta^{3/2})^2 \le (\sum_i\theta_i)(\sum_i\theta_i^3) \le \|\theta\|_1\|\theta\|_3^3$; therefore, $\|\theta\|^8 \le \|\theta\|^4\|\theta\|_3^3\|\theta\|_1 \le \|\theta\|_3^6\|\theta\|_1^2$. We then consider U_{a2}. Define

$$\mathcal{P}_5^* = \left\{ \begin{array}{ll} \text{path } i\text{-}\ell\text{-}k\text{-}j\text{-}s \text{ in a complete}: & \text{nodes } i,j,k,\ell \text{ are distinct,} \\ \text{graph with } n \text{ nodes} & \text{and node } s \text{ is different from } j,\ k \end{array} \right\}.$$

Fix a path i-ℓ-k-j-s in $\mathcal{P}_5^*$. If $s\notin\{i,\ell\}$, then this path is counted twice in the definition of U_{a2}, as i-ℓ-k-j-s and s-j-k-ℓ-i, respectively. If $s\in\{i,\ell\}$, then it is counted only once in the definition of U_{a2}. Hence, we can re-write

$$U_{a2} = -\frac{8}{\sqrt{v}}\sum_{\substack{\text{path in } \mathcal{P}_5^* \\ s\notin\{i,\ell\}}}(\eta_i+\eta_s)W_{sj}W_{jk}W_{k\ell}W_{\ell i} - \frac{8}{\sqrt{v}}\sum_{\substack{\text{path in } \mathcal{P}_5^* \\ s\in\{i,\ell\}}}\eta_i W_{sj}W_{jk}W_{k\ell}W_{\ell i}.$$

For two distinct paths in $\mathcal{P}_5^*$, the corresponding summands are uncorrelated with each other. It follows that

$$\begin{aligned} \mathrm{Var}(U_{a2}) &= \frac{64}{v}\sum_{\substack{\text{path in } \mathcal{P}_5^* \\ s\notin\{i,\ell\}}}(\eta_i+\eta_s)^2\,\mathrm{Var}(W_{sj}W_{jk}W_{k\ell}W_{\ell i}) \\ &\quad + \frac{64}{v}\sum_{\substack{\text{path in } \mathcal{P}_5^* \\ s\in\{i,\ell\}}}\eta_i^2\,\mathrm{Var}(W_{sj}W_{jk}W_{k\ell}W_{\ell i}) \\ &\le \frac{C}{v}\sum_{i,j,k,\ell,s}(\eta_i^2+\eta_s^2)\cdot\theta_i\theta_j^2\theta_k^2\theta_\ell^2\theta_s \\ &\le \frac{C}{\|\theta\|_1^2}\sum_{i,j,k,\ell,s}(\theta_i^3\theta_j^2\theta_k^2\theta_\ell^2\theta_s + \theta_i\theta_j^2\theta_k^2\theta_\ell^2\theta_s^3) \\ &\le \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}. \end{aligned} \tag{2.E.119}$$

By Cauchy-Schwarz inequality, $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$, so the right hand side of (2.E.119) is

$\leq C\|\theta\|^2\|\theta\|_3^6$. Combining it with (2.E.118) gives

$$\mathrm{Var}(U_a) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8).$$

This proves the claim.

Proof of Lemma 19

It suffices to prove the claims for each of Y_1-Y_6. Consider Y_1. We have analyzed this term under the null hypothesis. Using similar proof, we can easily derive that

$$\mathbb{E}[Y_1] = 0, \qquad \mathrm{Var}(Y_1) \leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8).$$

Consider Y_2. Using the definition of Y_2 and the expression of $\widetilde{\eta}_i$ in (2.E.113), we have

$$\begin{aligned}
Y_2 &= \sum_{i,j,k,\ell(dist)} \delta_{ij}\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} \\
&= \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_j(\eta_i - \widetilde{\eta}_i)\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} \\
&= \frac{1}{\sqrt{v}} \sum_{i,j,k,\ell(dist)} \eta_i\Big(-\sum_{s\neq j} W_{js}\Big)\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} + \frac{1}{\sqrt{v}} \sum_{i,j,k,\ell(dist)} \eta_j\Big(-\sum_{s\neq i} W_{is}\Big)\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i} \\
&= -\frac{1}{\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq j}} \eta_i\widetilde{\Omega}_{jk}W_{js}W_{k\ell}W_{\ell i} - \frac{1}{\sqrt{v}} \sum_{\substack{i,k,\ell(dist)\\ s\neq i}} \Big(\sum_{j\notin\{i,k,\ell\}} \eta_j\widetilde{\Omega}_{jk}\Big)W_{is}W_{k\ell}W_{\ell i}.
\end{aligned}$$

In the second sum above, we further separate two cases, $s = \ell$ and $s \neq \ell$. It then gives rise to three terms:

$$\begin{aligned}
Y_2 \quad &= \quad -\frac{1}{\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq j}} \eta_i\widetilde{\Omega}_{jk}W_{js}W_{k\ell}W_{\ell i} \\
&\quad -\frac{1}{\sqrt{v}} \sum_{i,k,\ell(dist)} \Big(\sum_{j\notin\{i,k,\ell\}} \eta_j\widetilde{\Omega}_{jk}\Big)W_{i\ell}^2W_{k\ell} \\
&\quad -\frac{1}{\sqrt{v}} \sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}}} \Big(\sum_{j\notin\{i,k,\ell\}} \eta_j\widetilde{\Omega}_{jk}\Big)W_{is}W_{k\ell}W_{\ell i} \\
&\equiv \quad Y_{2a} + Y_{2b} + Y_{2c}.
\end{aligned} \tag{2.E.120}$$

Since (i, j, k, ℓ) are distinct, it is easy to see that all three terms have mean zero. We thus have

$$\mathbb{E}[Y_2] = 0.$$

Below, we calculate the variances. First, we bound the variance of Y_{2a}. Each (i, j, k, ℓ, s) is associated with a length-3 path i-k-ℓ and an edge j-s in the complete graph. For (i, j, k, ℓ, s) and (i', j', k', ℓ', s'), if the associated path and edge are the same, then we group them together. Given a length-3 path i-k-ℓ and an edge j-s (such that the edge is not in the path), they are counted four times in the definition of Y_{2a}, as (i) i-k-ℓ and j-s, (ii) i-k-ℓ and s-j, (iii) ℓ-k-i and j-s, (iv) ℓ-k-i and s-j, so we group these four summands together. After grouping the summands, we re-write

$$Y_{2a} = -\frac{1}{\sqrt{v}} \sum_{\substack{\text{length-3} \\ \text{path}}} \sum_{\substack{\text{edge not} \\ \text{in the path}}} \big(\eta_i\widetilde{\Omega}_{jk} + \eta_i\widetilde{\Omega}_{sk} + \eta_k\widetilde{\Omega}_{ji} + \eta_k\widetilde{\Omega}_{si}\big) W_{js}W_{k\ell}W_{\ell i}.$$

In this new expression of Y_{2a}, two summands are correlated only when the underlying path&edge pairs are exactly the same. Additionally, by (2.E.110) and (2.E.117),

$$\big|\eta_i\widetilde{\Omega}_{jk} + \eta_i\widetilde{\Omega}_{sk} + \eta_k\widetilde{\Omega}_{ji} + \eta_k\widetilde{\Omega}_{si}\big| \leq C\alpha(\theta_j+\theta_s)\theta_i\theta_k.$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Y_{2a}) &\leq \frac{C}{v} \sum_{i,j,k,\ell,s} \alpha^2(\theta_j+\theta_s)^2\theta_i^2\theta_k^2 \cdot \mathrm{Var}(W_{js}W_{k\ell}W_{\ell i}) \\
&\leq \frac{C}{\|\theta\|_1^2} \sum_{i,j,k,\ell,s} \alpha^2(\theta_j+\theta_s)^2\theta_i^2\theta_k^2 \cdot \theta_i\theta_j\theta_k\theta_\ell^2\theta_s \\
&\leq \frac{C\alpha^2}{\|\theta\|_1^2} \sum_{i,j,k,\ell,s} (\theta_i^3\theta_j^3\theta_k^3\theta_\ell^2\theta_s + \theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_s^3) \\
&\leq \frac{C\alpha^2\|\theta\|^2\|\theta\|_3^9}{\|\theta\|_1}. \qquad (2.E.121)
\end{aligned}$$

Second, we bound the variance of Y_{2b}. Write $\beta_{ik\ell} = \sum_{j\notin\{i,k,\ell\}} \eta_j\widetilde{\Omega}_{jk}$. By (2.E.110) and (2.E.117), $|\beta_{ik\ell}| \leq C\sum_j \theta_j \cdot \alpha\theta_j\theta_k \leq C\alpha\|\theta\|^2\theta_k$. Using this notation,

$$Y_{2b} = \frac{1}{v} \sum_{i,j,k,\ell(dist)} \beta_{ik\ell}W_{i\ell}^2W_{k\ell}, \qquad \text{where} \quad |\beta_{ik\ell}| \leq C\alpha\|\theta\|^2\theta_k.$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Y_{2b}) = \mathbb{E}[Y_{2b}^2] &\leq \frac{C}{v} \sum_{\substack{i,k,\ell(dist) \\ i',k',\ell'(dist)}} \beta_{ik\ell}\beta_{i'k'\ell'} \cdot \mathbb{E}[W_{i\ell}^2W_{k\ell}W_{i'\ell'}^2W_{k'\ell'}] \\
&\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^2} \sum_{\substack{i,k,\ell(dist) \\ i',k',\ell'(dist)}} \theta_k\theta_{k'} \cdot \mathbb{E}[W_{i\ell}^2W_{k\ell}W_{i'\ell'}^2W_{k'\ell'}].
\end{aligned}$$

The summand is nonzero only when the two variables $W_{k\ell}$ and $W_{k'\ell'}$ equal to each other or when each of them equals to some other squared variables. By elementary calculations,

$$\begin{aligned}
&\theta_k\theta_{k'} \cdot \mathbb{E}[W_{i\ell}^2W_{k\ell}W_{i'\ell'}^2W_{k'\ell'}] \\
&= \begin{cases}
\theta_k^2\,\mathbb{E}[W_{i\ell}^4W_{k\ell}^2] \leq C\theta_i\theta_k^3\theta_\ell^2, & \text{if } (k',\ell')=(k,\ell),\ i'=i; \\
\theta_k^2\,\mathbb{E}[W_{i\ell}^2W_{k\ell}^2W_{i'\ell}^2] \leq C\theta_i\theta_k^3\theta_\ell^3\theta_{i'}, & \text{if } (k',\ell')=(k,\ell),\ i'\neq i; \\
\theta_k\theta_\ell\,\mathbb{E}[W_{i\ell}^2W_{k\ell}^2W_{i'k}^2] \leq C\theta_i\theta_k^3\theta_\ell^3\theta_{i'}, & \text{if } (k',\ell')=(\ell,k); \\
\theta_k^2\,\mathbb{E}[W_{i\ell}^3W_{k\ell}^3] \leq C\theta_i\theta_k^3\theta_\ell^2, & \text{if } \ell'=\ell,\ (i',k')=(i,k); \\
\theta_k\theta_i\,\mathbb{E}[W_{i\ell}^3W_{k\ell}^3] \leq C\theta_i^2\theta_k^2\theta_\ell^2, & \text{if } \ell'=\ell,\ (i',k')=(k,i); \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

As a result,

$$\begin{aligned}\mathrm{Var}(Y_{2b}) &\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^2}\Big(\sum_{i,k,\ell}\theta_i\theta_k^3\theta_\ell^2 + \sum_{i,k,\ell,i'}\theta_i\theta_k^3\theta_\ell^3\theta_{i'} + \sum_{i,k,\ell}\theta_i^2\theta_k^2\theta_\ell^2\Big)\\ &\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^2}\big(\|\theta\|_3^3\|\theta\|^2\|\theta\|_1 + \|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^6\big)\\ &\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6, \end{aligned} \tag{2.E.122}$$

where to get the last inequality we have used $\|\theta\|^6 \ll \|\theta\|^8 \leq (\|\theta\|_1\|\theta\|_3^3)^2$ and $\|\theta\|_3^3\|\theta\|^2\|\theta\|_1 \ll \|\theta\|_3^3\|\theta\|^4\|\theta\|_1 \leq (\|\theta\|_1\|\theta\|_3^3)^2$. Last, we bound the variance of Y_{2c}. Let $\beta_{ik\ell} = \sum_{j\notin\{i,k,\ell\}}\eta_j\widetilde{\Omega}_{jk}$ be the same as above. We write

$$Y_{2c} = \frac{1}{\sqrt{v}}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}}}\beta_{ik\ell}W_{is}W_{k\ell}W_{\ell i}, \qquad \text{where} \quad |\beta_{ik\ell}| \leq C\alpha\|\theta\|^2\theta_k.$$

For $\mathbb{E}[W_{is}W_{k\ell}W_{\ell i}\cdot W_{i's'}W_{k'\ell'}W_{\ell' i'}]$ to be nonzero, it has to be the case that $(W_{is}, W_{k\ell}, W_{\ell i})$ and $(W_{i's'}, W_{k'\ell'}, W_{\ell'i'})$ are the same set of variables, up to an order permutation. For each fixed (i,k,ℓ,s), there are only a constant number of (i',k',ℓ',s') such that the above is satisfied. As we have argued many times before (e.g., see (2.E.106)), it is true that

$$\begin{aligned}\mathrm{Var}(Y_{2c}) &\leq \frac{C}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}}}\beta_{ik\ell}^2\cdot\mathrm{Var}(W_{is}W_{k\ell}W_{\ell i})\\ &\leq \frac{C}{\|\theta\|_1^2}\sum_{i,k,\ell,s}(\alpha\|\theta\|^2\theta_k)^2\cdot\theta_i^2\theta_k\theta_\ell^2\theta_s\\ &\leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1}. \end{aligned} \tag{2.E.123}$$

We now combine the variances of Y_{2a}-Y_{2c}. Since $\|\theta\|_3^3 \leq \theta_{\max}^2\|\theta\|_1 \ll \|\theta\|_1$, the right hand side is (2.E.121) is $o(\alpha^2\|\theta\|^2\|\theta\|_3^6) = o(\alpha^2\|\theta\|^4\|\theta\|_3^6)$. Since $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the right hand side is (2.E.123) is $\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6$. It follows that

$$\mathrm{Var}(Y_2) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8).$$

This proves the claims of Y_2.

Consider Y_3. By definition,

$$Y_3 = \sum_{i,j,k,\ell(dist)}\eta_i(\eta_j - \widetilde{\eta}_j)W_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)}\eta_j(\eta_i - \widetilde{\eta}_i)W_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}.$$

In the second sum, if we relabel $(i,j,k,\ell) = (j',i',\ell',k')$, it can be written as $\sum_{i',j',k',\ell'(dist)}\eta_{i'}(\eta_{j'} - \widetilde{\eta}_{j'})W_{i'\ell'}\widetilde{\Omega}_{\ell'k'}W_{k'j'}$. This shows that the second sum is indeed equal to the first sum. As a result,

$$
\begin{aligned}
Y_3 &= 2 \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j) W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i} \\
&= 2 \sum_{i,j,k,\ell(dist)} \eta_i \Big(-\frac{1}{\sqrt{v}} \sum_{s \neq j} W_{js} \Big) W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i} \\
&= -\frac{2}{\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist) \\ s \neq j}} \eta_i \widetilde{\Omega}_{k\ell} W_{js} W_{jk} W_{\ell i} \\
&= -\frac{2}{\sqrt{v}} \sum_{i,j,k,\ell(dist)} \eta_i \widetilde{\Omega}_{k\ell} W_{jk}^2 W_{\ell i} - \frac{2}{\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist) \\ s \notin \{j,k\}}} \eta_i \widetilde{\Omega}_{k\ell} W_{js} W_{jk} W_{\ell i} \\
&\equiv Y_{3a} + Y_{3b},
\end{aligned} \tag{2.E.124}
$$

where the second line is from (2.E.113) and the second last line is from dividing all summands into two cases of $s = k$ and $s \neq k$. Both terms have mean zero, so

$$
\mathbb{E}[Y_3] = 0.
$$

Below, first, we calculate the variance of Y_{3a}.

$$
\mathrm{Var}(Y_{3a}) = \frac{4}{v} \sum_{\substack{i,j,k,\ell(dist) \\ i',j',k',\ell'(dist)}} (\eta_i \widetilde{\Omega}_{k\ell} \eta_{i'} \widetilde{\Omega}_{k'\ell'}) \cdot \mathbb{E}[W_{jk}^2 W_{\ell i} W_{j'k'}^2 W_{\ell' i'}].
$$

The summand is nonzero only if either the two variables $W_{\ell i}$ and $W_{\ell' i'}$ are the same, or each of the two variables $W_{\ell i}$ and $W_{\ell' i'}$ equals to another squared W term. By (2.E.110), (2.E.117), and elementary calculations,

$$
\begin{aligned}
&(\eta_i \widetilde{\Omega}_{k\ell} \eta_{i'} \widetilde{\Omega}_{k'\ell'}) \cdot \mathbb{E}[W_{jk}^2 W_{\ell i} W_{j'k'}^2 W_{\ell' i'}] \\
&\leq C\alpha^2 \theta_i \theta_k \theta_\ell \theta_{i'} \theta_{k'} \theta_{\ell'} \cdot \mathbb{E}[W_{jk}^2 W_{\ell i} W_{j'k'}^2 W_{\ell' i'}] \\
&= \begin{cases}
C\alpha^2 \theta_i^2 \theta_\ell^2 \theta_k^2 \, \mathbb{E}[W_{jk}^4 W_{\ell i}^2] \leq C\alpha^2 \theta_i^3 \theta_j \theta_k^3 \theta_\ell^3, & \text{if } \{\ell', i'\} = \{\ell, i\},\ (j', k') = (j, k); \\
C\alpha^2 \theta_i^2 \theta_\ell^2 \theta_k \theta_j \, \mathbb{E}[W_{jk}^4 W_{\ell i}^2] \leq C\alpha^2 \theta_i^3 \theta_j^2 \theta_k^2 \theta_\ell^3, & \text{if } \{\ell', i'\} = \{\ell, i\},\ (j', k') = (k, j); \\
C\alpha^2 \theta_i^2 \theta_\ell^2 \theta_k \theta_{k'} \, \mathbb{E}[W_{jk}^2 W_{\ell i}^2 W_{j'k'}^2] \leq C\alpha^2 \theta_i^3 \theta_j \theta_k^2 \theta_\ell^3 \theta_{j'} \theta_{k'}^2, & \text{if } \{\ell', i'\} = \{\ell, i\},\ \{j', k'\} \neq \{j, k\}; \\
C\alpha^2 \theta_i^2 \theta_\ell \theta_j \theta_k^2 \, \mathbb{E}[W_{jk}^3 W_{\ell i}^3] \leq C\alpha^2 \theta_i^3 \theta_j^2 \theta_k^3 \theta_\ell^2, & \text{if } \{\ell', i'\} = \{j, k\},\ (j', k') = (\ell, i); \\
C\alpha^2 \theta_i \theta_\ell^2 \theta_j \theta_k^2 \, \mathbb{E}[W_{jk}^3 W_{\ell i}^3] \leq C\alpha^2 \theta_i^2 \theta_j^2 \theta_k^3 \theta_\ell^3, & \text{if } \{\ell', i'\} = \{j, k\},\ (j', k') = (i, \ell); \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}
$$

There are only three different cases in the bounds. It follows that

$$
\begin{aligned}
\mathrm{Var}(Y_{3a}) \quad &\leq \quad \frac{C\alpha^2}{\|\theta\|_1^2} \Big(\sum_{i,j,k,\ell} \theta_i^3 \theta_j \theta_k^3 \theta_\ell^3 + \sum_{i,j,k,\ell} \theta_i^3 \theta_j^2 \theta_k^2 \theta_\ell^3 + \sum_{i,j,k,\ell,j',k'} \theta_i^3 \theta_j \theta_k^2 \theta_\ell^3 \theta_{j'} \theta_{k'}^2 \Big) \\
&\leq \quad \frac{C\alpha^2}{\|\theta\|_1^2} \big(\|\theta\|_1 \|\theta\|_3^9 + \|\theta\|^4 \|\theta\|_3^6 + \|\theta\|^4 \|\theta\|_1^2 \|\theta\|_3^6 \big) \\
&\leq \quad C\alpha^2 \|\theta\|^4 \|\theta\|_3^6,
\end{aligned} \tag{2.E.125}
$$

where in the last line we have used $\|\theta\|_3^9 \leq \|\theta\|_3^6 (\theta_{\max} \|\theta\|^2) = o(\|\theta\|^2 \|\theta\|_3^6)$ and $\|\theta\|_1 \geq \theta_{\max}^{-1} \|\theta\|^2 \to \infty$. Next, we calculate the variance of Y_{3b}. We mimic the argument in (2.E.121) and group summands according to the underlying path s-j-k and edge ℓ-i in a complete graph. It yields

$$Y_{3b} = -\frac{2}{\sqrt{v}} \sum_{\substack{\text{length-3}\\ \text{path}}} \sum_{\substack{\text{edge not}\\ \text{in the path}}} \big(\eta_i\widetilde{\Omega}_{k\ell} + \eta_\ell\widetilde{\Omega}_{ki} + \eta_i\widetilde{\Omega}_{s\ell} + \eta_\ell\widetilde{\Omega}_{si}\big) W_{sj}W_{jk}W_{\ell i},$$

where

$$\big|\eta_i\widetilde{\Omega}_{k\ell} + \eta_\ell\widetilde{\Omega}_{ki} + \eta_i\widetilde{\Omega}_{s\ell} + \eta_\ell\widetilde{\Omega}_{si}\big| \leq C\alpha(\theta_k + \theta_s)\theta_i\theta_\ell.$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Y_{3b}) &\leq \frac{C}{v}\sum_{i,j,k,\ell,s} \alpha^2(\theta_k+\theta_s)^2\theta_i^2\theta_\ell^2 \cdot \mathrm{Var}(W_{sj}W_{jk}W_{\ell i})\\
&\leq \frac{C\alpha^2}{\|\theta\|_1^2}\sum_{i,j,k,\ell,s} (\theta_i^3\theta_j^2\theta_k^3\theta_\ell^3\theta_s + \theta_i^3\theta_j^2\theta_k\theta_\ell^3\theta_s^3)\\
&\leq \frac{C\alpha^2\|\theta\|^2\|\theta\|_3^9}{\|\theta\|_1}. \qquad (2.E.126)
\end{aligned}$$

Since $\|\theta\|_3^9 \leq \|\theta\|_3^6(\theta_{\max}\|\theta\|_1) = o(\|\theta\|_1\|\theta\|_3^6)$, so the right hand side of (2.E.126) is much smaller than the right hand side of (2.E.125). Together, we have

$$\mathrm{Var}(Y_3) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8).$$

This proves the claims of Y_3.

Consider Y_4. We plug in $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$ and the expression (2.E.113). It gives

$$\begin{aligned}
Y_4 &= \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_j(\eta_i - \widetilde{\eta}_i)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}\\
&= \sum_{i,j,k,\ell(dist)} \eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_j\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq i} W_{is}\Big)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}\\
&= -\frac{1}{\sqrt{v}}\sum_{\substack{i,j,\ell(dist)\\ s\neq j}}\Big(\sum_{k\notin\{i,j,\ell\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\Big)W_{js}W_{\ell i} - \frac{1}{\sqrt{v}}\sum_{\substack{i,\ell(dist)\\ s\neq i}}\Big(\sum_{j,k\notin\{i,\ell\}} \eta_j\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\Big)W_{is}W_{\ell i}\\
&\equiv Y_{4a} + Y_{4b}.
\end{aligned}$$

First, we analyze Y_{4a}. When (i,j,ℓ) are distinct, $W_{js}W_{\ell i}$ has a mean zero. Therefore,

$$\mathbb{E}[Y_{4a}] = 0.$$

To calculate the variance, we rewrite

$$Y_{4a} = -\frac{1}{\sqrt{v}}\sum_{\substack{i,j,\ell(dist)\\ s\neq j}} \beta_{ij\ell}W_{js}W_{\ell i}, \qquad \text{where} \quad \beta_{ij\ell} = \sum_{k\notin\{i,j,\ell\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}$$

By (2.E.110) and (2.E.117), $|\beta_{ij\ell}| \leq C\sum_k \alpha^2\theta_i\theta_j\theta_k^2\theta_\ell \leq C\alpha^2\|\theta\|^2\theta_i\theta_j\theta_\ell$. Also, for $W_{js}W_{\ell i}$ and $W_{j's'}W_{\ell'i'}$ to be correlated, there are only two cases: $(W_{js}, W_{\ell i}) = (W_{j's'}, W_{\ell'i'})$ or $(W_{js}, W_{\ell i}) = (W_{\ell'i'}, W_{j's'})$. Mimicking the argument in (2.E.121) or (2.E.126), we can easily obtain that

$$
\begin{aligned}
\mathrm{Var}(Y_{4a}) &\le \frac{C}{v}\sum_{\substack{i,j,\ell(dist)\\ s\neq j}} \beta_{ij\ell}^2\cdot \mathrm{Var}(W_{js}W_{\ell i})\\
&\le \frac{C}{\|\theta\|_1^2}\sum_{i,j,\ell,s}(\alpha^2\|\theta\|^2\theta_i\theta_j\theta_\ell)^2\cdot\theta_i\theta_j\theta_\ell\theta_s\\
&\le \frac{C\alpha^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1}. \qquad (2.E.127)
\end{aligned}
$$

Next, we analyze Y_{4b}. We re-write

$$
Y_{4b} = -\frac{1}{\sqrt{v}}\sum_{\substack{i,\ell(dist)\\ s\neq i}}\beta_{i\ell}W_{is}W_{\ell i}, \qquad \text{where} \quad \beta_{i\ell}=\sum_{j,k\notin\{i,\ell\}}\eta_j\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}.
$$

By separating the case of $s=\ell$ from the case of $s\neq\ell$, we have

$$
Y_{4b} = -\frac{1}{\sqrt{v}}\sum_{i,\ell(dist)}\beta_{i\ell}W_{\ell i}^2 - \frac{1}{\sqrt{v}}\sum_{\substack{i,\ell(dist)\\ s\notin\{i,\ell\}}}\beta_{i\ell}W_{is}W_{\ell i}\equiv \widetilde{Y}_{4b}+Y_{4b}^*.
$$

Only $\widetilde{Y}_{4b}$ has a nonzero mean. By (2.E.110) and (2.E.117),

$$
|\beta_{i\ell}|\le C\sum_{j,k}\alpha^2\theta_j^2\theta_k^2\theta_\ell \le C\alpha^2\|\theta\|^4\theta_\ell.
$$

It follows that

$$
|\mathbb{E}[Y_{4b}]| = |\mathbb{E}[\widetilde{Y}_{4b}]| \le \frac{C}{\|\theta\|_1}\sum_{i,\ell}(\alpha^2\|\theta\|^4\theta_\ell)\theta_i\theta_\ell \le C\alpha^2\|\theta\|^6. \qquad (2.E.128)
$$

We now bound the variances of $\widetilde{Y}_{4b}$ and Y_{4b}^*. By direct calculations,

$$
\begin{aligned}
\mathrm{Var}(\widetilde{Y}_{4b}) &= \frac{2}{v}\sum_{i,\ell(dist)}\beta_{i\ell}^2\cdot\mathrm{Var}(W_{i\ell}^2) \le \frac{C}{\|\theta\|_1^2}\sum_{i,\ell}(\alpha^2\|\theta\|^4\theta_\ell)^2\cdot\theta_i\theta_\ell \le \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1},\\
\mathrm{Var}(Y_{4b}^*) &\le \frac{C}{v}\sum_{\substack{i,\ell(dist)\\ s\notin\{i,\ell\}}}\beta_{i\ell}^2\cdot\mathrm{Var}(W_{is}W_{\ell i}) \le \frac{C}{\|\theta\|_1^2}\sum_{i,\ell,s}(\alpha^2\|\theta\|^4\theta_\ell)^2\cdot\theta_i^2\theta_\ell\theta_s \le \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1}.
\end{aligned}
$$

Together, we have

$$
\mathrm{Var}(Y_{4b}) \le 2\mathrm{Var}(\widetilde{Y}_{4b}) + 2\mathrm{Var}(Y_{4b}^*) \le \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1}. \qquad (2.E.129)
$$

We combine the results of Y_{4a} and Y_{4b}. Since $\|\theta\|_3^6\le(\theta_{\max}\|\theta\|^2)^2 = o(\|\theta\|^4)$, the right hand side of (2.E.128) dominates the right hand side of (2.E.127). It follows that

$$
|\mathbb{E}[Y_4]| \le C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8), \quad \mathrm{Var}(Y_4)\le \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6).
$$

Here, we explain the equalities. The first one is due to $\alpha^2\|\theta\|^2\to\infty$. To get the second equality, we compare $\mathrm{Var}(Y_4)$ with the order of $\alpha^6\|\theta\|^8\|\theta\|_3^6$. Note that $\frac{\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = \frac{\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}\|\theta\|^4 \le \frac{\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}\|\theta\|_1\|\theta\|_3^3 \le \|\theta\|^6\|\theta\|_3^6$. It follows that $\mathrm{Var}(Y_4)\le C\alpha^4\|\theta\|^6\|\theta\|_3^6 \ll C\alpha^6\|\theta\|^8\|\theta\|_3^6$, where the last inequality is due to $\alpha^2\|\theta\|^2\to\infty$. So far, we have proved all claims about Y_4.

Consider Y_5. Recall that

$$Y_5 = \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}W_{k\ell}\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}W_{k\ell}\widetilde{\Omega}_{\ell i}.$$

With relabeling of $(i,j,k,\ell) = (j',i',\ell',k')$, the second sum can be written as $\sum_{i',j',k',\ell'(dist)}(\eta_{j'} - \widetilde{\eta}_{j'})\eta_{i'}\widetilde{\Omega}_{i'\ell'}W_{\ell'k'}\widetilde{\Omega}_{k'j'}$. This suggests that it is actually equal to the first sum above. Hence,

$$\begin{aligned}
Y_5 &= 2\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}W_{k\ell}\widetilde{\Omega}_{\ell i} \\
&= \sum_{i,j,k,\ell(dist)} \eta_i\Big(-\frac{2}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\widetilde{\Omega}_{jk}W_{k\ell}\widetilde{\Omega}_{\ell i} \\
&= -\frac{2}{\sqrt{v}} \sum_{\substack{j,k,\ell(dist)\\ s\neq j}} \Big(\sum_{i\notin\{j,k,\ell\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{\ell i}\Big)W_{js}W_{k\ell} \\
&\equiv -\frac{2}{\sqrt{v}} \sum_{\substack{j,k,\ell(dist)\\ s\neq j}} \beta_{jk\ell}W_{js}W_{k\ell}, \qquad \text{where} \quad \beta_{jk\ell} \equiv \sum_{i\notin\{j,k,\ell\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{\ell i}.
\end{aligned}$$

It is easy to see that $\mathbb{E}[W_{js}W_{k\ell}] = 0$ when (j,k,ℓ) are distinct. Hence,

$$\mathbb{E}[Y_5] = 0.$$

By (2.E.110) and (2.E.117), $|\beta_{jk\ell}| \leq C\sum_i \theta_i \cdot \alpha^2\theta_j\theta_k\theta_\ell\theta_i \leq C\alpha^2\|\theta\|^2\theta_j\theta_k\theta_\ell$. Similar to the argument in (2.E.121) or (2.E.126), we can show that

$$\begin{aligned}
\mathrm{Var}(Y_5) &\leq \frac{C}{v}\sum_{\substack{j,k,\ell(dist)\\ s\neq j}} \beta_{jk\ell}^2 \cdot \mathrm{Var}(W_{js}W_{k\ell}) \\
&\leq \frac{C}{\|\theta\|_1^2}\sum_{j,k,\ell,s} (\alpha^2\|\theta\|^2\theta_j\theta_k\theta_\ell)^2\theta_j\theta_s\theta_k\theta_\ell \\
&\leq \frac{C\alpha^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1}.
\end{aligned}$$

Since $\|\theta\|_3^9 = (\|\theta\|_3^3)^2\|\theta\|_3^3 \leq (\theta_{\max}\|\theta\|^2)^2(\theta_{\max}^2\|\theta\|_1) = o(\|\theta\|^4\|\theta\|_1)$, the right hand side is $o(\|\theta\|^8)$. This proves the claims of Y_5.

Consider Y_6. By definition and elementary calculations,

$$\begin{aligned}
Y_6 &= \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_j(\eta_i - \widetilde{\eta}_i)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} \\
&= 2\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} \\
&= 2\sum_{i,j,k,\ell(dist)} \eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} \\
&= -\frac{2}{\sqrt{v}}\sum_{j,s(dist)} \Big(\sum_{i,k,\ell(dist)\notin\{j\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\Big)W_{js}.
\end{aligned}$$

Here, to get the second line above, we relabeled $(i,j,k,\ell) = (j',i',\ell',k')$ in the second sum and found out the two sums are equal; the third line is from (2.E.113). We immediately see that

$$\mathrm{E}[Y_6] = 0.$$

By (2.E.110) and (2.E.117),

$$\Big| \sum_{i,k,\ell(dist)\notin\{j\}} \eta_i \widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\Big| \leq \sum_{i,k,\ell} C\theta_i \cdot \alpha^3\theta_j\theta_k^2\theta_\ell^2\theta_i \leq C\alpha^3\|\theta\|^6\theta_j.$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Y_6) &= \frac{8}{v}\sum_{j,s(dist)}\Big(\sum_{i,k,\ell(dist)\notin\{j\}} \eta_i\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\Big)^2 \cdot \mathrm{Var}(W_{js})\\
&\leq \frac{C}{\|\theta\|_1^2}\sum_{j,s}(\alpha^3\|\theta\|^6\theta_j)^2\theta_j\theta_s\\
&\leq \frac{C\alpha^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1}.
\end{aligned}$$

Since $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the variance is bounded by $C\alpha^6\|\theta\|^8\|\theta\|_3^6$. This proves the claims of Y_6.

Proof of Lemma 20

It suffices to prove the claims for each of Z_1 and Z_2; then, the claims of U_b follow immediately.

We first analyze Z_1. Plugging $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$ into the definition of Z_1 gives

$$\begin{aligned}
Z_1 = &\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\eta_j(\eta_k - \widetilde{\eta}_k)W_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)^2\eta_k W_{k\ell}W_{\ell i}\\
&+ \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j^2(\eta_k - \widetilde{\eta}_k)W_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j(\eta_j - \widetilde{\eta}_j)\eta_k W_{k\ell}W_{\ell i}.
\end{aligned}$$

In the last term above, if we relabel $(i,j,k,\ell) = (k',j',i',\ell')$, it becomes $\sum_{i',j',k',\ell'(dist)}(\eta_{k'} - \widetilde{\eta}_{k'})\eta_{j'}(\eta_{j'} - \widetilde{\eta}_{j'})\eta_{i'}W_{i'\ell'}W_{\ell'k'}$. This shows that the last sum equals to the first sum. Therefore,

$$\begin{aligned}
Z_1 \quad = \quad &\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)^2\eta_k W_{k\ell}W_{\ell i}\\
&+2\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\eta_j(\eta_k - \widetilde{\eta}_k)W_{k\ell}W_{\ell i}\\
&+\sum_{i,j,k,\ell(dist)} (\widetilde{\eta}_i - \eta_i)\eta_j^2(\widetilde{\eta}_k - \eta_k)W_{k\ell}W_{\ell i}\\
\equiv \quad & Z_{1a} + Z_{1b} + Z_{1c}. \qquad (2.E.130)
\end{aligned}$$

Below, we compute the means and variances of Z_{1a}-Z_{1c}.

First, we study Z_{1a}. When (i,j,k,ℓ) are distinct, $W_{k\ell}W_{\ell i}$ has a mean zero and is independent of $(\widetilde{\eta}_j - \eta_j)^2$, so $\mathbb{E}[(\eta_j - \widetilde{\eta}_j)^2 W_{k\ell}W_{\ell i}] = 0$. It follows that

$$\mathbb{E}[Z_{1a}] = 0.$$

To bound the variance of Z_{1a}, we use (2.E.113) to re-write

$$Z_{1a} \quad = \quad \sum_{i,j,k,\ell(dist)} \eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq j} W_{jt}\Big)\eta_k W_{k\ell}W_{\ell i}$$

$$\begin{aligned}
&= \frac{1}{v} \sum_{\substack{i,j,k,\ell(dist)\\ s,t\notin\{j\}}} \eta_i\eta_k W_{js}W_{jt}W_{k\ell}W_{\ell i} \\
&= \frac{1}{v} \sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{j\}}} \eta_i\eta_k W_{js}^2 W_{k\ell}W_{\ell i} + \frac{1}{v} \sum_{\substack{i,j,k,\ell(dist)\\ s,t(dist)\notin\{j\}}} \eta_i\eta_k W_{js}W_{jt}W_{k\ell}W_{\ell i} \\
&\equiv \widetilde{Z}_{1a} + Z_{1a}^*.
\end{aligned}$$

We first bound the variance of $\widetilde{Z}_{1a}$. It is seen that

$$\mathrm{Var}(\widetilde{Z}_{1a}) = \frac{1}{v^2} \sum_{\substack{i,j,k,\ell(dist),s\notin\{j\}\\ i',j',k',\ell'(dist),s'\notin\{j'\}}} \eta_i\eta_k\eta_{i'}\eta_{k'} \cdot \mathbb{E}[W_{js}^2 W_{k\ell}W_{\ell i} \cdot W_{j's'}^2 W_{k'\ell'}W_{\ell' i'}].$$

The summand is nonzero only if $\ell' = \ell$ and $\{k', i'\} = \{k, i\}$. We also note that, if we switch i' and k', the summand remains unchanged. So, it suffices to consider the case of $\ell' = \ell$ and $(k', i') = (k, i)$. By (2.E.117) and elementary calculations,

$$\begin{aligned}
&\eta_i\eta_k\eta_{i'}\eta_{k'} \cdot \mathbb{E}[W_{js}^2 W_{k\ell}W_{\ell i} \cdot W_{j's'}^2 W_{k'\ell'}W_{\ell' i'}] \\
&= \begin{cases}
\eta_i^2\eta_k^2\,\mathbb{E}[W_{js}^4 W_{k\ell}^2 W_{\ell i}^2] \le C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_s, & \text{if } (\ell',k',i') = (\ell,k,i),\ \{j',s'\} = \{j,s\}; \\
\eta_i^2\eta_k^2\,\mathbb{E}[W_{js}^2 W_{k\ell}^2 W_{\ell i}^2 W_{j's'}^2] \le C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_s\theta_{j'}\theta_{s'}, & \text{if } (\ell',k',i') = (\ell,k,i),\ \{j',s'\} \neq \{j,s\}; \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}(\widetilde{Z}_{1a}) &\le \frac{C}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell,s} \theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_s + \sum_{i,j,k,\ell,s,j',s'} \theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_s\theta_{j'}\theta_{s'} \Big) \\
&\le \frac{C}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^2\|\theta\|_3^6\|\theta\|_1^4\big) \\
&\le C\|\theta\|^2\|\theta\|_3^6.
\end{aligned}$$

We then bound the variance of Z_{1a}^*. Note that

$$\begin{aligned}
&\eta_i\eta_k\eta_{i'}\eta_{k'} \cdot \mathbb{E}[W_{js}W_{jt}W_{k\ell}W_{\ell i} \cdot W_{j's'}W_{j't'}W_{k'\ell'}W_{\ell' i'}] \\
&= \begin{cases}
\eta_i^2\eta_k^2\,\mathbb{E}[W_{js}^2 W_{jt}^2 W_{k\ell}^2 W_{\ell i}^2] \le C\theta_i^3\theta_j^2\theta_k^3\theta_\ell^2\theta_s\theta_t, & \text{if } (j',\ell') = (j,\ell), \{s',t'\} = \{s,t\}, \{k',i'\} = \{k,i\}; \\
\eta_i\eta_k\eta_s\eta_t\,\mathbb{E}[W_{js}^2 W_{jt}^2 W_{k\ell}^2 W_{\ell i}^2] \le C\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s^2\theta_t^2, & \text{if } (j',\ell') = (\ell,j), \{s',t'\} = \{k,i\}, \{k',i'\} = \{s,t\}; \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Z_{1a}^*) &\le \frac{C}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell,s,t} \theta_i^3\theta_j^2\theta_k^3\theta_\ell^2\theta_s\theta_t + \sum_{i,j,k,\ell,s,t} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s^2\theta_t^2 \Big) \\
&\le \frac{C}{\|\theta\|_1^4}\big(\|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^{12}\big) \\
&\le \frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2},
\end{aligned}$$

where the last inequality is because of $\|\theta\|^{12} = \|\theta\|^4(\|\theta\|^4)^2 \le \|\theta\|^4(\|\theta\|_1\|\theta\|_3^3)^2 = \|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2$. Combining the above gives

$$\mathrm{Var}(Z_{1a}) \le 2\mathrm{Var}(\widetilde{Z}_{1a}) + 2\mathrm{Var}(Z_{1a}^*) \le C\|\theta\|^2\|\theta\|_3^6. \tag{2.E.131}$$

Second, we study Z_{1b}. Since $(\eta_j - \widetilde{\eta}_j)$, $(\eta_k - \widetilde{\eta}_k)W_{k\ell}$ and $W_{\ell i}$ are independent of each other, each summand in Z_{1b} has a zero mean. It follows that

$$\mathbb{E}[Z_{1b}] = 0.$$

We now compute its variance. By direct calculations,

$$\begin{aligned}
Z_{1b} &= 2 \sum_{i,j,k,\ell(dist)} \eta_i \Big(-\frac{1}{\sqrt{v}} \sum_{s\neq j} W_{js}\Big) \eta_j \Big(-\frac{1}{\sqrt{v}} \sum_{t\neq k} W_{kt}\Big) W_{k\ell} W_{\ell i} \\
&= \frac{2}{v} \sum_{\substack{i,j,k,\ell(dist) \\ s\neq j, t\neq k}} \eta_i \eta_j W_{js} W_{kt} W_{k\ell} W_{\ell i} \\
&= \frac{2}{v} \sum_{\substack{i,j,k,\ell(dist) \\ s\neq j}} \eta_i \eta_j W_{js} W_{k\ell}^2 W_{\ell i} + \frac{2}{v} \sum_{\substack{i,j,k,\ell(dist) \\ s\neq j, t\notin\{k,\ell\}}} \eta_i \eta_j W_{js} W_{kt} W_{k\ell} W_{\ell i} \\
&\equiv \widetilde{Z}_{1b} + Z_{1b}^*.
\end{aligned}$$

We first bound the variance of $\widetilde{Z}_{1b}$. Note that

$$\mathrm{Var}(\widetilde{Z}_{1b}) = \frac{4}{v} \sum_{\substack{i,j,k,\ell(dist), s\neq j \\ i',j',k',\ell'(dist), s'\neq j'}} \eta_i \eta_j \eta_{i'} \eta_{j'} \cdot \mathbb{E}[W_{js} W_{k\ell}^2 W_{\ell i} \cdot W_{j's'} W_{k'\ell'}^2 W_{\ell' i'}].$$

For this summand to be nonzero, there are only two cases. In the first case, $(W_{js}, W_{\ell i})$ are paired with $(W_{j's'}, W_{\ell' i'})$. It follows that

$$\eta_i \eta_j \eta_{i'} \eta_{j'} \cdot \mathbb{E}[W_{js} W_{k\ell}^2 W_{\ell i} W_{j's'} W_{k'\ell'}^2 W_{\ell' i'}] = \eta_i \eta_j \eta_{i'} \eta_{j'} \cdot \mathbb{E}[W_{js}^2 W_{k\ell}^2 W_{\ell i}^2 W_{k'\ell'}^2].$$

This happens only if (i) $\{j', s'\} = \{j, s\}$ and $\{\ell', i'\} = \{\ell, i\}$, or (ii) $\{j', s'\} = \{\ell, i\}$ and $\{\ell', i'\} = \{j, s\}$. By (2.E.117) and elementary calculations,

$$\begin{aligned}
&\eta_i \eta_j \eta_{i'} \eta_{j'} \cdot \mathbb{E}[W_{js} W_{k\ell}^2 W_{\ell i} \cdot W_{j's'} W_{k'\ell'}^2 W_{\ell' i'}] \\
&= \begin{cases}
\eta_i^2 \eta_j^2 \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'\ell}^2] \leq C\theta_i^3 \theta_j^3 \theta_k \theta_\ell^3 \theta_s \theta_{k'}, & \text{if } (j', s') = (j, s), (\ell', i') = (\ell, i); \\
\eta_i \eta_j^2 \eta_\ell \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'i}^2] \leq C\theta_i^3 \theta_j^3 \theta_k \theta_\ell^3 \theta_s \theta_{k'}, & \text{if } (j', s') = (j, s), (\ell', i') = (i, \ell); \\
\eta_i^2 \eta_j \eta_s \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'\ell}^2] \leq C\theta_i^3 \theta_j^2 \theta_k \theta_\ell^3 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (s, j), (\ell', i') = (\ell, i); \\
\eta_i \eta_j \eta_\ell \eta_s \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'i}^2] \leq C\theta_i^3 \theta_j^2 \theta_k \theta_\ell^3 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (s, j), (\ell', i') = (i, \ell); \\
\eta_i \eta_j \eta_\ell \eta_s \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'j}^2] \leq C\theta_i^2 \theta_j^3 \theta_k \theta_\ell^3 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (\ell, i), (\ell', i') = (j, s); \\
\eta_i \eta_j^2 \eta_\ell \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k's}^2] \leq C\theta_i^2 \theta_j^3 \theta_k \theta_\ell^3 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (\ell, i), (\ell', i') = (s, j); \\
\eta_i^2 \eta_j \eta_s \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k'j}^2] \leq C\theta_i^3 \theta_j^3 \theta_k \theta_\ell^2 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (i, \ell), (\ell', i') = (j, s); \\
\eta_i^2 \eta_j^2 \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{k\ell}^2 W_{k's}^2] \leq C\theta_i^3 \theta_j^3 \theta_k \theta_\ell^2 \theta_s^2 \theta_{k'}, & \text{if } (j', s') = (i, \ell), (\ell', i') = (s, j); \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

The upper bound on the right hand side only has two types $C\theta_i^3 \theta_j^3 \theta_k \theta_\ell^3 \theta_s \theta_{k'}$ and $C\theta_i^3 \theta_j^2 \theta_k \theta_\ell^3 \theta_s^2 \theta_{k'}$.

The contribution of this case to $\mathrm{Var}(\widetilde{Z}_{1b})$ is

$$
\begin{aligned}
&\le \frac{C}{v^2}\Big(\sum_{i,j,k,\ell,s,k'} \theta_i^3\theta_j^3\theta_k\theta_\ell^3\theta_s\theta_{k'} + \sum_{i,j,k,\ell,s,k'} \theta_i^3\theta_j^2\theta_k\theta_\ell^3\theta_s^2\theta_{k'}\Big)\\
&\le \frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^9\|\theta\|_1^3 + \|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2\big)\\
&\le \frac{C\|\theta\|_3^9}{\|\theta\|_1}.
\end{aligned}
$$

In the second case, $\{W_{js}, W_{k\ell}, W_{\ell i}\}$ and $\{W_{j's'}, W_{k'\ell'}, W_{\ell' i'}\}$ are two sets of same variables. Then,

$$
\eta_i\eta_j\eta_{i'}\eta_{j'}\cdot\mathbb{E}[W_{js}W_{k\ell}^2W_{\ell i}W_{j's'}W_{k'\ell'}^2W_{\ell' i'}] = \eta_i\eta_j\eta_{i'}\eta_{j'}\cdot\mathbb{E}[W_{js}^3W_{k\ell}^3W_{\ell i}^3].
$$

This can only happen if $\ell' = \ell$, $\{i', k'\} = \{i, k\}$, and $\{j', s'\} = \{j, s\}$. By (2.E.117) and elementary calculations,

$$
\begin{aligned}
&\eta_i\eta_j\eta_{i'}\eta_{j'}\cdot\mathbb{E}[W_{js}W_{k\ell}^2W_{\ell i}\cdot W_{j's'}W_{k'\ell'}^2W_{\ell' i'}]\\
&= \begin{cases}
\eta_i^2\eta_j^2\cdot\mathbb{E}[W_{js}^3W_{\ell i}^3W_{k\ell}^3] \le C\theta_i^3\theta_j^3\theta_k\theta_\ell^2\theta_s, & \text{if } \ell'=\ell,\ (i',k')=(i,k),\ (j',s')=(j,s);\\
\eta_i^2\eta_j\eta_s\cdot\mathbb{E}[W_{js}^3W_{\ell i}^3W_{k\ell}^3] \le C\theta_i^3\theta_j^2\theta_k\theta_\ell^2\theta_s^2, & \text{if } \ell'=\ell,\ (i',k')=(i,k),\ (j',s')=(s,j);\\
\eta_i\eta_k\eta_j^2\cdot\mathbb{E}[W_{js}^3W_{\ell i}^3W_{k\ell}^3] \le C\theta_i^2\theta_j^3\theta_k^2\theta_\ell^2\theta_s, & \text{if } \ell'=\ell,\ (i',k')=(k,i),\ (j',s')=(j,s);\\
\eta_i\eta_k\eta_j\eta_s\cdot\mathbb{E}[W_{js}^3W_{\ell i}^3W_{k\ell}^3] \le C\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s^2, & \text{if } \ell'=\ell,\ (i',k')=(i,k),\ (j',s')=(s,j);\\
0, & \text{otherwise.}
\end{cases}
\end{aligned}
$$

The upper bound on the right hand side has three types, and the contribution of this case to $\mathrm{Var}(\widetilde{Z}_{1b})$ is

$$
\begin{aligned}
&\le \frac{C}{v^2}\Big(\sum_{i,j,k,\ell,s} \theta_i^3\theta_j^3\theta_k\theta_\ell^2\theta_s + \sum_{i,j,k,\ell,s} \theta_i^3\theta_j^2\theta_k\theta_\ell^2\theta_s^2 + \sum_{i,j,k,\ell,s} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s^2\Big)\\
&\le \frac{C}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^6\|\theta\|_3^3\|\theta\|_1 + \|\theta\|^{10}\big)\\
&\le \frac{C\|\theta\|^2\|\theta\|_3^6}{\|\theta\|_1^2},
\end{aligned}
$$

where we use $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$ (Cauchy-Schwarz) in the last line. It is seen that the contribution of the first case is dominating, and so

$$
\mathrm{Var}(\widetilde{Z}_{1b}) \le \frac{C\|\theta\|_3^9}{\|\theta\|_1}.
$$

We then bound the variance of Z_{1b}^*. Note that

$$
\mathrm{Var}(Z_{1b}^*) = \frac{4}{v^2}\sum_{\substack{i,j,k,\ell(dist),s\neq j,t\notin\{k,\ell\}\\ i',j',k',\ell'(dist),s'\neq j',t'\notin\{k',\ell'\}}} \eta_i\eta_j\eta_{i'}\eta_{j'}\cdot\mathbb{E}[W_{js}W_{kt}W_{k\ell}W_{\ell i}\cdot W_{j's'}W_{k't'}W_{k'\ell'}W_{\ell' i'}].
$$

For the summand to be nonzero, all W terms have to be perfectly matched, so that the expectation in the summand becomes

$$
\mathbb{E}[W_{js}W_{kt}W_{k\ell}W_{\ell i}\cdot W_{j's'}W_{k't'}W_{k'\ell'}W_{\ell' i'}] = \mathbb{E}[W_{js}^2W_{kt}^2W_{k\ell}^2W_{\ell i}^2] \le C\theta_i\theta_j\theta_k^2\theta_\ell^2\theta_s\theta_t.
$$

For this perfect match to happen, we need $(t', k', \ell', i') = (t, k, \ell, i)$ or $(t', k', \ell', i') = (i, \ell, k, t)$, as well as $\{j', s'\} = \{j, s\}$. This implies that, i' can only take values in $\{i, t\}$ and j' can only take values in $\{j, s\}$. It follows that $\eta_i\eta_j\eta_{i'}\eta_{j'}$ belongs to one of the following cases:

$$\eta_i\eta_j(\eta_i\eta_j) \le C\theta_i^2\theta_j^2, \qquad \eta_i\eta_j(\eta_i\eta_s) = C\theta_i^2\theta_j\theta_s,$$
$$\eta_i\eta_j(\eta_t\eta_j) \le C\theta_i\theta_j^2\theta_t, \qquad \eta_i\eta_j(\eta_t\eta_s) \le C\theta_i\theta_j\theta_t\theta_s.$$

Combining the above gives

$$\begin{aligned}
\mathrm{Var}(Z_{1b}^*) &\le \frac{C}{v^2}\sum_{i,j,k,\ell,s,t}(\theta_i^2\theta_j^2 + \theta_i^2\theta_j\theta_s + \theta_i\theta_j^2\theta_t + \theta_i\theta_j\theta_t\theta_s)\cdot\theta_i\theta_j\theta_k^2\theta_\ell^2\theta_s\theta_t \\
&\le \frac{C}{\|\theta\|_1^4}\big(\|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2 + 2\|\theta\|^8\|\theta\|_3^3\|\theta\|_1 + \|\theta\|^{12}\big) \\
&\le \frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2}.
\end{aligned}$$

We combine the variances of $\widetilde{Z}_{1b}$ and Z_{1b}^*. Since $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$, the variance of $\widetilde{Z}_{1b}$ dominates. It follows that

$$\mathrm{Var}(Z_{1b}) \le 2\mathrm{Var}(\widetilde{Z}_{1b}) + 2\mathrm{Var}(Z_{1b}^*) \le \frac{C\|\theta\|_3^9}{\|\theta\|_1}. \tag{2.E.132}$$

Third, we study Z_{1c}. It is seen that

$$\begin{aligned}
Z_{1c} &= \sum_{i,j,k,\ell(dist)}\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq i}W_{is}\Big)\eta_j^2\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq k}W_{kt}\Big)W_{k\ell}W_{\ell i} \\
&= \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\neq i, t\neq k}}\Big(\sum_{j\notin\{i,k,\ell\}}\eta_j^2\Big)W_{is}W_{kt}W_{k\ell}W_{\ell i} \\
&\equiv \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\neq i, t\neq k}}\beta_{ik\ell}W_{is}W_{kt}W_{k\ell}W_{\ell i},
\end{aligned}$$

where

$$\beta_{ik\ell} \equiv \sum_{j\notin\{i,k,\ell\}}\eta_j^2 \le C\sum_j\theta_j^2 \le C\|\theta\|^2. \tag{2.E.133}$$

We divide all summands into four groups: (i) $s = t = \ell$; (ii) $s = \ell$, $t \neq \ell$; (iii) $s \neq \ell$, $t = \ell$; (iv) $s \neq \ell$, $t \neq \ell$. It yields that

$$\begin{aligned}
Z_{1c} = \frac{1}{v}\sum_{i,k,\ell(dist)}\beta_{ik\ell}W_{k\ell}^2W_{\ell i}^2 + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ t\neq\{k,\ell\}}}\beta_{ik\ell}W_{kt}W_{k\ell}W_{\ell i}^2 \\
+ \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}}}\beta_{ik\ell}W_{is}W_{k\ell}^2W_{\ell i} + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}, t\notin\{k,\ell\}}}\beta_{ik\ell}W_{is}W_{kt}W_{k\ell}W_{\ell i}.
\end{aligned}$$

In the third sum, if we relabel $(i,k,\ell,s) = (k',i',\ell',t')$, it then has the form of $\sum_{i',k',\ell'(dist), t'\notin\{k',\ell'\}}\beta_{k'i'\ell'}W_{k't'}W_{i'\ell'}^2W_{\ell'}$

This shows that this sum equals to the second sum. We thus have

$$\begin{aligned}
Z_{1c} &= \frac{1}{v}\sum_{i,k,\ell(dist)} \beta_{ik\ell} W_{k\ell}^2 W_{\ell i}^2 + \frac{2}{v}\sum_{\substack{i,k,\ell(dist)\\ t\neq\{k,\ell\}}} \beta_{ik\ell} W_{kt} W_{k\ell} W_{\ell i}^2 \\
&\quad + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}, t\notin\{k,\ell\}}} \beta_{ik\ell} W_{is} W_{kt} W_{k\ell} W_{\ell i} \\
&\equiv \widetilde{Z}_{1c} + Z_{1c}^* + Z_{1c}^\dagger.
\end{aligned}$$

Among all three terms, only $\widetilde{Z}_{1c}$ has a nonzero mean. It follows that

$$\begin{aligned}
\mathbb{E}[Z_{1c}] = \mathbb{E}[\widetilde{Z}_{1c}] &= \frac{1}{v}\sum_{i,k,\ell(dist)} \beta_{ik\ell}\Omega_{k\ell}(1-\Omega_{k\ell})\Omega_{\ell i}(1-\Omega_{\ell i}) \\
&= \frac{1}{v}\sum_{i,k,\ell(dist)} \beta_{ik\ell}\Omega_{k\ell}\Omega_{\ell i}[1+O(\theta_{\max}^2)].
\end{aligned}$$

Under the null hypothesis, $\Omega_{ij} = \theta_i\theta_j$. It follows that $\eta_j = \frac{\theta_j}{\sqrt{v}}\sum_{i:i\neq j}\theta_i = [1+o(1)]\frac{\theta_j\|\theta\|_1}{\sqrt{v}}$ and that $\beta_{ik\ell} = \sum_{j\notin\{i,k,\ell\}}\eta_j^2 = [1+o(1)]\frac{\|\theta\|_1^2}{v}\sum_{j\notin\{i,k,\ell\}}\theta_j^2 = [1+o(1)]\frac{\|\theta\|_1^2\|\theta\|^2}{v}$. Additionally, $v = \sum_{i\neq j}\theta_i\theta_j = \|\theta\|_1^2\cdot[1+o(1)]$. As a result,

$$\begin{aligned}
\mathbb{E}[Z_{1c}] &= \frac{1}{v}\sum_{i,k,\ell(dist)} [1+o(1)]\frac{\|\theta\|_1^2\|\theta\|^2}{v}\cdot\theta_k\theta_\ell^2\theta_i \\
&= [1+o(1)]\cdot\frac{\|\theta\|_1^2\|\theta\|^2}{v^2}\sum_{i,k,\ell(dist)}\theta_k\theta_\ell^2\theta_i \\
&= [1+o(1)]\cdot\frac{\|\theta\|_1^2\|\theta\|^2}{\|\theta\|_1^4}\big[\|\theta\|_1^2\|\theta\|^2 - O(\|\theta\|^4 + \|\theta\|_1\|\theta\|_3^3)\big] \\
&= [1+o(1)]\cdot\|\theta\|^4,
\end{aligned} \tag{2.E.134}$$

where in the last line we have used $\|\theta\|^2 = o(\|\theta\|_1)$, $\|\theta\|_3^3 = o(\|\theta\|_1)$ and $\|\theta\|_1\to\infty$. We then bound the variance of Z_{1c} by studying the variance of each of the three variables, $\widetilde{Z}_{1c}$, Z_{1c}^* and $Z_{1c}^\dagger$. Consider $\widetilde{Z}_{1c}$ first. For $W_{k\ell}^2W_{\ell i}^2$ and $W_{k'\ell'}^2W_{\ell' i'}^2$ to be correlated, it has to be the case of either $\{k',\ell'\} = \{k,\ell\}$ or $\{i',\ell'\} = \{i,\ell\}$. By symmetry between k and i in the expression, it suffices to consider $\{k',\ell'\} = \{k,\ell\}$. Direct calculations show that

$$\mathrm{Cov}(W_{k\ell}^2W_{\ell i}^2,\, W_{k'\ell'}^2W_{\ell' i'}^2) \leq \begin{cases}
\mathbb{E}[W_{k\ell}^4W_{\ell i}^4] \leq C\theta_k\theta_\ell^2\theta_i, & \text{if } (k',\ell') = (k,\ell),\ i' = i; \\
\mathbb{E}[W_{k\ell}^4W_{\ell i}^2W_{\ell i'}^2] \leq C\theta_k\theta_\ell^3\theta_i\theta_{i'}, & \text{if } (k',\ell') = (k,\ell),\ i' \neq i; \\
\mathbb{E}[W_{k\ell}^4W_{\ell i}^2W_{ki}^2] \leq C\theta_k^2\theta_\ell^2\theta_i^2, & \text{if } (k',\ell') = (\ell,k),\ i' = i; \\
\mathbb{E}[W_{k\ell}^4W_{\ell i}^2W_{ki'}^2] \leq C\theta_k^2\theta_\ell^2\theta_i\theta_{i'}, & \text{if } (k',\ell') = (\ell,k),\ i' \neq i; \\
0, & \text{otherwise.}
\end{cases}$$

Combining it with (2.E.133) and the fact of $v \geq C^{-1}\|\theta\|_1^2$, we have

$$\begin{aligned}
\mathrm{Var}(\widetilde{Z}_{1c}) &\leq \frac{C\|\theta\|^4}{\|\theta\|_1^4}\Big(\sum_{i,k,\ell}\theta_k\theta_\ell^2\theta_i + \sum_{i,k,\ell,i'}\theta_k\theta_\ell^3\theta_i\theta_{i'} + \sum_{i,k,\ell}\theta_k^2\theta_\ell^2\theta_i^2 + \sum_{i,k,\ell,i'}\theta_k^2\theta_\ell^2\theta_i\theta_{i'}\Big) \\
&\leq \frac{C\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_1^2 + \|\theta\|_3^3\|\theta\|_1^3 + \|\theta\|^6 + \|\theta\|^4\|\theta\|_1^2\big) \\
&\leq \frac{C\|\theta\|^4\|\theta\|_3^3}{\|\theta\|_1}.
\end{aligned}$$

Consider Z_{1c}^*. By direct calculations,

$$\begin{aligned}
&\mathbb{E}[W_{kt}W_{k\ell}W_{\ell i}^2W_{k't'}W_{k'\ell'}W_{\ell'i'}^2] \\
&= \begin{cases}
\mathbb{E}[W_{kt}^2W_{k\ell}^2W_{\ell i}^4] \leq C\theta_i\theta_k^2\theta_\ell^2\theta_t, & \text{if } (k',t',\ell') = (k,t,\ell),\ i = i'; \\
\mathbb{E}[W_{kt}^2W_{k\ell}^2W_{\ell i}^2W_{\ell i'}^2] \leq C\theta_i\theta_k^2\theta_\ell^3\theta_t\theta_{i'}, & \text{if } (k',t',\ell') = (k,t,\ell),\ i \neq i'; \\
\mathbb{E}[W_{kt}^2W_{k\ell}^2W_{\ell i}^2W_{ti'}^2] \leq C\theta_i\theta_k^2\theta_\ell^2\theta_t^2\theta_{i'}, & \text{if } (k',t',\ell') = (k,\ell,t); \\
\mathbb{E}[W_{kt}^3W_{k\ell}^2W_{\ell i}^3] \leq C\theta_i\theta_k^2\theta_\ell^2\theta_t, & \text{if } (k',t',\ell',i') = (\ell,i,k,t); \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

We combine it with (2.E.133) and find that

$$\begin{aligned}
\mathrm{Var}(Z_{1c}^*) &= \frac{4}{v^2}\sum_{\substack{i,k,\ell(dist),t\neq\{k,\ell\}\\ i',k',\ell'(dist),t'\neq\{k',\ell'\}}}\beta_{ik\ell}\beta_{i'k'\ell'}\cdot\mathbb{E}[W_{kt}W_{k\ell}W_{\ell i}^2W_{k't'}W_{k'\ell'}W_{\ell'i'}^2] \\
&\leq \frac{C\|\theta\|^4}{\|\theta\|_1^4}\Big(\sum_{i,k,\ell,t}\theta_i\theta_k^2\theta_\ell^2\theta_t + \sum_{i,k,\ell,t,i'}\theta_i\theta_k^2\theta_\ell^3\theta_t\theta_{i'} + \sum_{i,k,\ell,t,i'}\theta_i\theta_k^2\theta_\ell^2\theta_t^2\theta_{i'}\Big) \\
&\leq \frac{C\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|^4\|\theta\|_1^2 + \|\theta\|^2\|\theta\|_3^3\|\theta\|_1^3 + \|\theta\|^6\|\theta\|_1^2\big) \\
&\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}.
\end{aligned}$$

Consider $Z_{1c}^\dagger$. Re-write

$$Z_{1c}^\dagger = \frac{1}{v}\sum_{i,k,\ell(dist)}\beta_{ik\ell}W_{ik}^2W_{k\ell}W_{\ell i} + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\},t\notin\{k,\ell\}\\ (s,t)\neq(k,i)}}\beta_{ik\ell}W_{is}W_{kt}W_{k\ell}W_{\ell i}.$$

Regarding the first term, by direct calculations,

$$\begin{aligned}
&\mathbb{E}[W_{ik}^2W_{k\ell}W_{\ell i}\cdot W_{i'k'}^2W_{k'\ell'}W_{\ell'i'}] \\
&= \begin{cases}
\mathbb{E}[W_{ik}^4W_{k\ell}^2W_{\ell i}^2] \leq C\theta_i^2\theta_k^2\theta_\ell^2, & \text{if } \ell' = \ell,\ \{i',k'\} = \{i,k\}; \\
\mathbb{E}[W_{ik}^3W_{k\ell}^2W_{\ell i}^3] \leq C\theta_i^2\theta_k^2\theta_\ell^2, & \text{if } (\ell',k') = (k,\ell),\ i' = i; \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

Combining it with (2.E.133) gives

$$\mathrm{Var}\Big(\frac{1}{v}\sum_{i,k,\ell(dist)}\beta_{ik\ell}W_{ik}^2W_{k\ell}W_{\ell i}\Big) \leq \frac{C\|\theta\|^4}{\|\theta\|_1^4}\sum_{i,j,k,\ell}\theta_i^2\theta_k^2\theta_\ell^2 \leq \frac{C\|\theta\|^{10}}{\|\theta\|_1^4}.$$

Regarding the second term, for $W_{is}W_{kt}W_{k\ell}W_{\ell i}$ and $W_{i's'}W_{k't'}W_{k'\ell'}W_{\ell'i'}$ to be correlated, all

W terms have to be perfectly matched. For each fixed (i,k,ℓ,s,t), there are only a constant number of (i',k',ℓ',s',t') so that the above is satisfied. Mimicking the argument in (2.E.106), we have

$$\mathrm{Var}\Big(\frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\},t\notin\{k,\ell\}\\ (s,t)\neq(k,i)}}\beta_{ik\ell}W_{is}W_{kt}W_{k\ell}W_{\ell i}\Big)\leq\frac{C}{v^2}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\},t\notin\{k,\ell\}\\ (s,t)\neq(k,i)}}\beta_{ik\ell}^2\cdot\mathrm{Var}(W_{is}W_{kt}W_{k\ell}W_{\ell i})$$
$$\leq\frac{C}{\|\theta\|_1^4}\sum_{i,k,\ell,s,t}\|\theta\|^4\cdot\theta_i^2\theta_k^2\theta_\ell^2\theta_s\theta_t\leq\frac{C\|\theta\|^{10}}{\|\theta\|_1^2}.$$

It follows that

$$\mathrm{Var}(Z_{1c}^{\dagger})\leq\frac{C\|\theta\|^{10}}{\|\theta\|_1^2}.$$

Combining the above results and noticing that $\|\theta\|^4\leq\|\theta\|_1\|\theta\|_3^3$, we immediately have

$$\mathrm{Var}(Z_{1c})\leq 3\mathrm{Var}(\widetilde{Z}_{1c})+3\mathrm{Var}(Z_{1c}^{*})+3\mathrm{Var}(Z_{1c}^{\dagger})\leq\frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}.\tag{2.E.135}$$

We now combine (2.E.131), (2.E.132), (2.E.134), and (2.E.135). Since $Z_1=Z_{1a}+Z_{1b}+Z_{1c}$, it follows that

$$\mathbb{E}[Z_1]=\|\theta\|^4\cdot[1+o(1)],\qquad\mathrm{Var}(Z_1)\leq C\|\theta\|^2\|\theta\|_3^6=o(\|\theta\|^8).$$

This proves the claims of Z_1.

Next, we analyze Z_2. Since $\delta_{ij}=\eta_i(\eta_j-\widetilde{\eta}_j)+\eta_j(\eta_i-\widetilde{\eta}_i)$, by direct calculations,

$$Z_2=\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j-\widetilde{\eta}_j)W_{jk}\eta_k(\eta_\ell-\widetilde{\eta}_\ell)W_{\ell i}+\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j-\widetilde{\eta}_j)W_{jk}(\eta_k-\widetilde{\eta}_k)\eta_\ell W_{\ell i}$$
$$+\sum_{i,j,k,\ell(dist)}(\eta_i-\widetilde{\eta}_i)\eta_jW_{jk}\eta_k(\eta_\ell-\widetilde{\eta}_\ell)W_{\ell i}+\sum_{i,j,k,\ell(dist)}(\eta_i-\widetilde{\eta}_i)\eta_jW_{jk}(\eta_k-\widetilde{\eta}_k)\eta_\ell W_{\ell i}.$$

By relabeling the indices, we find out that the first and last sums are equal and that the second and third sums are equal. It follows that

$$\begin{aligned}Z_2 &= 2\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j-\widetilde{\eta}_j)W_{jk}\eta_k(\eta_\ell-\widetilde{\eta}_\ell)W_{\ell i}\\ &\quad+2\sum_{i,j,k,\ell(dist)}\eta_i(\eta_j-\widetilde{\eta}_j)W_{jk}(\eta_k-\widetilde{\eta}_k)\eta_\ell W_{\ell i}\\ &\equiv Z_{2a}+Z_{2b}.\end{aligned}\tag{2.E.136}$$

First, we study Z_{2a}. It is seen that

$$\begin{aligned}Z_{2a}&=2\sum_{i,j,k,\ell(dist)}\eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j}W_{js}\Big)W_{jk}\eta_k\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq\ell}W_{\ell t}\Big)W_{\ell i}\\ &=\frac{2}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\neq j,t\neq\ell}}\eta_i\eta_kW_{js}W_{jk}W_{\ell t}W_{\ell i}.\end{aligned}$$

We divide summands into four groups: (i) $s=k$ and $t=i$, (ii) $s=k$ and $t\neq i$, (iii) $s\neq k$ and $t=i$, (iv) $s\neq k$ and $t\neq i$. By symmetry between (j,k,s) and (ℓ,i,t), the sum of group

(ii) and group (iii) are equal. We end up with

$$
\begin{aligned}
Z_{2a} &= \frac{2}{v}\sum_{i,j,k,\ell(dist)} \eta_i\eta_k W_{jk}^2 W_{\ell i}^2 + \frac{4}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{j,k\}}} \eta_i\eta_k W_{js}W_{jk}W_{\ell i}^2 \\
&\quad + \frac{2}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{j,k\},t\notin\{\ell,i\}}} \eta_i\eta_k W_{js}W_{jk}W_{\ell t}W_{\ell i} \\
&\equiv \widetilde{Z}_{2a} + Z_{2a}^* + Z_{2a}^\dagger,
\end{aligned}
$$

Only $\widetilde{Z}_{2a}$ has a nonzero mean. It follows that

$$
\mathbb{E}[Z_{2a}] = \mathbb{E}[\widetilde{Z}_{2a}] = \frac{2}{v}\sum_{i,j,k,\ell(dist)} \eta_i\eta_k\Omega_{jk}(1-\Omega_{jk})\Omega_{\ell i}(1-\Omega_{\ell i}).
$$

Under the null hypothesis, $\Omega_{ij} = \theta_i\theta_j$. Hence, $\Omega_{jk}(1-\Omega_{jk})\Omega_{\ell i}(1-\Omega_{\ell i}) = \theta_j\theta_k\theta_\ell\theta_i\cdot[1+O(\theta_{\max}^2)]$. Additionally, in the proof of (2.E.134), we have seen that $v = [1+o(1)]\cdot\|\theta\|_1^2$ and $\eta_j = [1+o(1)]\cdot\theta_j$. Combining these results gives

$$
\begin{aligned}
\mathbb{E}[Z_{2a}] &= \frac{2[1+o(1)]}{\|\theta\|_1^2}\sum_{i,j,k,\ell(dist)} (\theta_i\theta_k)(\theta_j\theta_k\theta_\ell\theta_i) \\
&= \frac{2[1+o(1)]}{\|\theta\|_1^2}\Big[\sum_{i,j,k,\ell}\theta_i^2\theta_j\theta_k^2\theta_\ell - \sum_{\substack{i,j,k,\ell\\ \text{(not dist)}}}\theta_i^2\theta_j\theta_k^2\theta_\ell\Big] \\
&= \frac{2[1+o(1)]}{\|\theta\|_1^2}\Big[\|\theta\|^4\|\theta\|_1^2 - O\big(\|\theta\|_4^4\|\theta\|_1^2 + \|\theta\|_3^3\|\theta\|^2\|\theta\|_1 + \|\theta\|^6\big)\Big] \\
&= \frac{2[1+o(1)]}{\|\theta\|_1^2}\cdot\|\theta\|^4\|\theta\|_1^2[1+o(1)] \\
&= [1+o(1)]\cdot 2\|\theta\|^4. \qquad (2.E.137)
\end{aligned}
$$

We then bound the variance of Z_a. Consider $\widetilde{Z}_{2a}$ first. Note that $W_{jk}^2W_{\ell i}^2$ and $W_{j'k'}^2W_{\ell'i'}^2$ are correlated only if either $\{j',k'\} = \{j,k\}$ or $\{j',k'\} = \{\ell,i\}$. By symmetry, it suffices to consider $\{j',k'\} = \{j,k\}$. Direct calculations show that

$$
\begin{aligned}
&\mathrm{Cov}(\eta_i\eta_k W_{jk}^2W_{\ell i}^2,\ \eta_{i'}\eta_{k'}W_{j'k'}^2W_{\ell'i'}^2) \\
&\le \begin{cases}
\eta_k^2\eta_i^2\mathbb{E}[W_{jk}^4W_{\ell i}^4] \le C\theta_i^3\theta_j\theta_k^3\theta_\ell, & \text{if } (j',k')=(j,k),\ i=i',\ \ell=\ell'; \\
\eta_k^2\eta_i^2\mathbb{E}[W_{jk}^4W_{\ell i}^2W_{\ell' i}^2] \le C\theta_i^4\theta_j\theta_k^3\theta_\ell\theta_{\ell'}, & \text{if } (j',k')=(j,k),\ i=i',\ \ell\neq\ell'; \\
\eta_k^2\eta_i\eta_{i'}\mathbb{E}[W_{jk}^4W_{\ell i}^2W_{\ell' i'}^2] \le C\theta_i^2\theta_j\theta_k^3\theta_\ell\theta_{i'}^2\theta_{\ell'}, & \text{if } (j',k')=(j,k),\ i\neq i'; \\
\eta_j\eta_k\eta_i^2\mathbb{E}[W_{jk}^4W_{\ell i}^4] \le C\theta_i^3\theta_j^2\theta_k^2\theta_\ell, & \text{if } (j',k')=(k,j),\ i=i',\ \ell=\ell'; \\
\eta_j\eta_k\eta_i^2\mathbb{E}[W_{jk}^4W_{\ell i}^2W_{\ell' i}^2] \le C\theta_i^4\theta_j^2\theta_k^2\theta_\ell\theta_{\ell'}, & \text{if } (j',k')=(k,j),\ i=i',\ \ell\neq\ell'; \\
\eta_j\eta_k\eta_i\eta_{i'}\mathbb{E}[W_{jk}^4W_{\ell i}^2W_{\ell' i'}^2] \le C\theta_i^2\theta_j^2\theta_k^2\theta_\ell\theta_{i'}^2\theta_{\ell'}, & \text{if } (j',k')=(k,j),\ i\neq i'; \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}
$$

As a result,

$$\begin{aligned}
\mathrm{Var}(\widetilde{Z}_{2a}) &= \frac{4}{v^2} \sum_{\substack{i,j,k,\ell (dist)\\ i',j',k',\ell'(dist)}} \mathrm{Cov}(\eta_i\eta_k W_{jk}^2 W_{\ell i}^2,\ \eta_{i'}\eta_{k'} W_{j'k'}^2 W_{\ell' i'}^2)\\
&\leq \frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|_4^4\|\theta\|_3^3\|\theta\|_1^3 + \|\theta\|_3^3\|\theta\|^4\|\theta\|_1^3\\
&\quad + \|\theta\|_3^3\|\theta\|^4\|\theta\|_1 + \|\theta\|_4^4\|\theta\|^4\|\theta\|_1^2 + \|\theta\|^8\|\theta\|_1^2\big)\\
&\leq \frac{C\|\theta\|^4\|\theta\|_3^3}{\|\theta\|_1},
\end{aligned}$$

where the last line is obtained as follows: There are six terms in the brackets; since $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the last three terms are dominated by the first three terms; for the first three terms, since $\|\theta\|_3^3 \leq \theta_{\max}^2\|\theta\|_1 = o(\|\theta\|_1)$ and $\|\theta\|_4^4 \leq \theta_{\max}^2\|\theta\|^2 = o(\|\theta\|^2)$, the third term dominates. Consider Z_{2a}^* next. We note that for

$$\mathbb{E}[W_{js}W_{jk}W_{\ell i}^2 \cdot W_{j's'}W_{j'k'}W_{\ell' i'}^2]$$

to be nonzero, it has to be the case of either $(W_{j's'}, W_{j'k'}) = (W_{js}, W_{jk})$ or $(W_{j's'}, W_{j'k'}) = (W_{jk}, W_{js})$. This can only happen if $(j', s', k') = (j, s, k)$ or $(j', s', k') = (j, k, s)$. By elementary calculations,

$$\begin{aligned}
&\eta_i\eta_k\eta_{i'}\eta_{k'} \cdot \mathbb{E}[W_{js}W_{jk}W_{\ell i}^2 \cdot W_{j's'}W_{j'k'}W_{\ell' i'}^2]\\
&= \begin{cases}
\eta_i^2\eta_k^2\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^4] \leq C\theta_i^3\theta_j^2\theta_k^3\theta_\ell\theta_s, & \text{if } (j',s',k') = (j,s,k),\ i'=i,\ \ell'=\ell;\\
\eta_i^2\eta_k^2\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^2W_{\ell' i}^2] \leq C\theta_i^4\theta_j^2\theta_k^3\theta_\ell\theta_s\theta_{\ell'}, & \text{if } (j',s',k') = (j,s,k),\ i'=i,\ \ell'\neq\ell;\\
\eta_i\eta_{i'}\eta_k^2\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^2W_{\ell' i'}^2] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell\theta_s\theta_{i'}^2\theta_{\ell'}, & \text{if } (j',s',k') = (j,s,k),\ i\neq i';\\
\eta_i^2\eta_k\eta_s\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^4] \leq C\theta_i^3\theta_j^2\theta_k^2\theta_\ell\theta_s^2, & \text{if } (j',s',k') = (j,k,s),\ i'=i,\ \ell'=\ell;\\
\eta_i^2\eta_k\eta_s\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^2W_{\ell' i}^2] \leq C\theta_i^4\theta_j^2\theta_k^2\theta_\ell\theta_s^2\theta_{\ell'}, & \text{if } (j',s',k') = (j,k,s),\ i'=i,\ \ell'\neq\ell;\\
\eta_i\eta_{i'}\eta_k\eta_s\,\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell i}^2W_{\ell' i'}^2] \leq C\theta_i^2\theta_j^2\theta_k^2\theta_\ell\theta_s^2\theta_{i'}^2\theta_{\ell'}, & \text{if } (j',s',k') = (j,k,s),\ i\neq i';\\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}(Z_{2a}^*) &= \frac{16}{v^2} \sum_{\substack{i,j,k,\ell (dist)\\ i',j',k',\ell'(dist)}} \eta_i\eta_k\eta_{i'}\eta_{k'} \cdot \mathbb{E}[W_{js}W_{jk}W_{\ell i}^2 \cdot W_{j's'}W_{j'k'}W_{\ell' i'}^2]\\
&\leq \frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|^2\|\theta\|_1^2 + \|\theta\|_4^4\|\theta\|_3^3\|\theta\|^2\|\theta\|_1^3 + \|\theta\|_3^3\|\theta\|^6\|\theta\|_1^3\\
&\quad + \|\theta\|_3^3\|\theta\|^6\|\theta\|_1 + \|\theta\|_4^4\|\theta\|^6\|\theta\|_1^2 + \|\theta\|^{10}\|\theta\|_1^2\big)\\
&\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1},
\end{aligned}$$

where the last inequality is obtained similarly as in the calculation of $\mathrm{Var}(\widetilde{Z}_{2a})$. Last, consider $Z_{2a}^\dagger$. Write

$$Z_{2a}^\dagger = \frac{2}{v} \sum_{i,j,k,\ell (dist)} \eta_i\eta_k W_{j\ell}^2 W_{jk} W_{\ell i} + \frac{2}{v} \sum_{\substack{i,j,k,\ell (dist)\\ s\notin\{j,k\}, t\notin\{\ell,i\}\\ (s,t)\neq(\ell,j)}} \eta_i\eta_k W_{js}W_{jk}W_{\ell t}W_{\ell i} \tag{2.E.138}$$

Regarding the first term, we note that

$$\eta_i\eta_k\eta_{i'}\eta_{k'}\cdot\mathbb{E}[W_{j\ell}^2W_{jk}W_{\ell i}\cdot W_{j'\ell'}^2W_{j'k'}W_{\ell'i'}]$$
$$=\begin{cases}\eta_i^2\eta_k^2\,\mathbb{E}[W_{jk}^2W_{\ell i}^2W_{j\ell}^4]\leq C\theta_i^3\theta_j^2\theta_k^3\theta_\ell^2, & \text{if } (j',k')=(j,k),\ (i',\ell')=(i,\ell);\\ \eta_i\eta_k^2\eta_\ell\,\mathbb{E}[W_{jk}^2W_{\ell i}^2W_{j\ell}^2W_{ji}^2]\leq C\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3, & \text{if } (j',k')=(j,k),\ (i',\ell')=(\ell,i);\\ \eta_i^2\eta_k\eta_\ell\,\mathbb{E}[W_{jk}^2W_{\ell i}^2W_{j\ell}^2W_{k\ell}^2]\leq C\theta_i^3\theta_j^2\theta_k^3\theta_\ell^4, & \text{if } (j',k')=(k,j),\ (i',\ell')=(i,\ell);\\ \eta_i\eta_k\eta_\ell\eta_j\,\mathbb{E}[W_{jk}^2W_{\ell i}^2W_{j\ell}^2W_{ki}^2]\leq C\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3, & \text{if } (j',k')=(k,j),\ (i',\ell')=(\ell,i);\\ 0, & \text{otherwise.}\end{cases}$$

It follows that

$$\begin{aligned}&\mathrm{Var}\Big(\frac{2}{v}\sum_{i,j,k,\ell(dist)}\eta_i\eta_kW_{j\ell}^2W_{jk}W_{\ell i}\Big)\\ \leq&\frac{C}{\|\theta\|_1^4}\sum_{i,j,k,\ell}(\theta_i^3\theta_j^2\theta_k^3\theta_\ell^2+\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3+\theta_i^3\theta_j^2\theta_k^3\theta_\ell^4)\\ \leq&\frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|^4+\|\theta\|_3^{12}+\|\theta\|_4^4\|\theta\|_3^6\|\theta\|^2\big)\\ \leq&\frac{C\|\theta\|_3^6\|\theta\|^4}{\|\theta\|_1^4}.\end{aligned}$$

Regarding the second term in (2.E.138). We note that, for $\eta_i\eta_kW_{js}W_{jk}W_{\ell t}W_{\ell i}$ and $\eta_{i'}\eta_{k'}W_{j's'}W_{j'k'}W_{\ell't'}W_{\ell'i'}$ to be correlated, all the W terms have to be perfectly paired. It turns out that

$$\mathbb{E}[W_{js}W_{jk}W_{\ell t}W_{\ell i}\cdot W_{j's'}W_{j'k'}W_{\ell't'}W_{\ell'i'}]=\mathbb{E}[W_{js}^2W_{jk}^2W_{\ell t}^2W_{\ell i}^2].$$

To perfectly pair the W terms, there are two possible cases: (i) $(j',\ell')=(j,\ell)$, $\{s',k'\}=\{s,k\}$, $\{\ell',i'\}=\{\ell,i\}$. (ii) $(j',\ell')=(\ell,j)$, $\{s',k'\}=\{\ell,i\}$, $\{\ell',i'\}=\{s,k\}$. As a result, $\eta_i\eta_k\eta_{i'}\eta_{k'}$ only has the following possibilities:

$$\begin{aligned}&\eta_i\eta_k(\eta_i\eta_k)=\eta_i^2\eta_k^2,\quad \eta_i\eta_k(\eta_i\eta_s)=\eta_i^2\eta_k\eta_s,\quad \eta_i\eta_k(\eta_\ell\eta_k)=\eta_i\eta_k^2\eta_\ell,\quad \eta_i\eta_k(\eta_\ell\eta_s)=\eta_i\eta_k\eta_\ell\eta_s,\\ &\eta_i\eta_k(\eta_k\eta_i)=\eta_i^2\eta_k^2,\quad \eta_i\eta_k(\eta_k\eta_\ell)=\eta_i\eta_k^2\eta_\ell,\quad \eta_i\eta_k(\eta_s\eta_i)=\eta_i^2\eta_k\eta_s,\quad \eta_i\eta_k(\eta_s\eta_\ell)=\eta_i\eta_k\eta_\ell\eta_s.\end{aligned}$$

By symmetry, there are only three different types: $\eta_i^2\eta_k^2$, $\eta_i^2\eta_k\eta_s$, and $\eta_i\eta_k\eta_\ell\eta_s$. It follows that

$$\begin{aligned}&\mathrm{Var}\Big(\frac{2}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{j,k\},t\notin\{\ell,i\},(s,t)\neq(\ell,j)}}\eta_i\eta_kW_{js}W_{jk}W_{\ell t}W_{\ell i}\Big)\\ \leq&\frac{C}{\|\theta\|_1^4}\sum_{i,j,k,\ell,s,t}(\theta_i^2\theta_k^2+\theta_i^2\theta_k\theta_s+\theta_i\theta_k\theta_\ell\theta_s)\cdot\theta_j^2\theta_s\theta_k\theta_\ell^2\theta_t\theta_i\\ \leq&\frac{C}{\|\theta\|_1^4}\sum_{i,j,k,\ell,s,t}(\theta_i^3\theta_j^2\theta_k^3\theta_\ell^2\theta_s\theta_t+\theta_i^3\theta_j^2\theta_k^2\theta_\ell^2\theta_s^2\theta_t+\theta_i^2\theta_j^2\theta_k^2\theta_\ell^3\theta_s^2\theta_t)\\ \leq&\frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|^4\|\theta\|_1^2+\|\theta\|_3^3\|\theta\|^8\|\theta\|_1\big)\ \leq\frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2}.\end{aligned}$$

It follows that

$$\mathrm{Var}(Z_{2a}^\dagger)\leq\frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2}.$$

Comparing the variances of $\widetilde{Z}_{2a}$, Z_{2a}^* and $Z_{2a}^\dagger$, we find out that the variance of Z_{2a}^* dominates.

As a result,

$$\mathrm{Var}(Z_{2a}) \leq 3\mathrm{Var}(\widetilde{Z}_{2a}) + 3\mathrm{Var}(Z_{2a}^{*}) + 3\mathrm{Var}(Z_{2a}^{\dagger}) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}. \tag{2.E.139}$$

Second, we study Z_{2b}. It is seen that

$$\begin{aligned} Z_{2b} &= 2 \sum_{i,j,k,\ell(dist)} \eta_i \Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big) W_{jk} \Big(-\frac{1}{\sqrt{v}}\sum_{t\neq k} W_{kt}\Big)\eta_\ell W_{\ell i} \\ &= \frac{2}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\neq j, t\neq k}} \eta_i\eta_\ell W_{js}W_{jk}W_{kt}W_{\ell i}. \end{aligned}$$

We divide summands into four groups: (i) $s = k$ and $t = j$, (ii) $s = k$ and $t \neq j$, (iii) $s \neq k$ and $t = j$, (iv) $s \neq k$ and $t \neq j$. By index symmetry, the sums of group (ii) and group (iii) are equal. We end up with

$$\begin{aligned} Z_{2b} &= \frac{2}{v}\sum_{i,j,k,\ell(dist)} \eta_i\eta_\ell W_{jk}^3 W_{\ell i} + \frac{4}{v}\sum_{i,j,k,\ell(dist), t\notin\{k,j\}} \eta_i\eta_\ell W_{jk}^2 W_{kt} W_{\ell i} \\ &\quad + \frac{2}{v}\sum_{i,j,k,\ell(dist), s\neq\{j,k\}, t\neq\{j,k\}} \eta_i\eta_\ell W_{js}W_{jk}W_{kt}W_{\ell i} \\ &\equiv \widetilde{Z}_{2b} + Z_{2b}^{*} + Z_{2b}^{\dagger}. \end{aligned}$$

It is easy to see that all three terms have mean zero. Therefore,

$$\mathbb{E}[Z_{2b}] = 0. \tag{2.E.140}$$

We then bound the variances. Consider $\widetilde{Z}_{2b}$ first. By direct calculations,

$$\begin{aligned} &\eta_i\eta_\ell\eta_{i'}\eta_{\ell'} \cdot \mathbb{E}[W_{jk}^3 W_{\ell i} \cdot W_{j'k'}^3 W_{\ell' i'}] \\ &= \begin{cases} \eta_i^2\eta_\ell^2 \cdot \mathbb{E}[W_{jk}^6 W_{\ell i}^2] \leq C\theta_i^3\theta_j\theta_k\theta_\ell^3, & \text{if } \{j',k'\} = \{j,k\},\ \{\ell',i'\} = \{\ell,i\}; \\ \eta_i\eta_\ell\eta_j\eta_k \cdot \mathbb{E}[W_{jk}^4 W_{\ell i}^4] \leq C\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2, & \text{if } \{j',k'\} = \{\ell,i\},\ \{\ell',i'\} = \{j,k\}; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

It follows that

$$\begin{aligned} \mathrm{Var}(\widetilde{Z}_{2b}) &\leq \frac{C}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell} \theta_i^3\theta_j\theta_k\theta_\ell^3 + \sum_{i,j,k,\ell} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\Big) \\ &\leq \frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^8\big) \\ &\leq \frac{C\|\theta\|_3^6}{\|\theta\|_1^2}. \end{aligned}$$

Consider Z^*_{2b} next. By direct calculations,

$$\eta_i\eta_\ell\eta_{i'}\eta_{\ell'}\cdot\mathbb{E}[W_{jk}^2W_{kt}W_{\ell i}\cdot W_{j'k'}^2W_{k't'}W_{\ell'i'}]$$
$$=\begin{cases}\eta_i^2\eta_\ell^2\,\mathbb{E}[W_{jk}^4W_{kt}^2W_{\ell i}^2]\le C\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t, & \text{if } (k',t')=(k,t),\ \{\ell',i'\}=\{\ell,i\},\ j'=j;\\ \eta_i^2\eta_\ell^2\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^2W_{j'k}^2]\le C\theta_i^3\theta_j\theta_k^3\theta_\ell^3\theta_t\theta_{j'}, & \text{if } (k',t')=(k,t),\ \{\ell',i'\}=\{\ell,i\},\ j'\ne j;\\ \eta_i^2\eta_\ell^2\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^2W_{j't}^2]\le C\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t^2\theta_{j'}, & \text{if } (k',t')=(t,k),\ \{\ell',i'\}=\{\ell,i\};\\ \eta_i\eta_\ell\eta_k\eta_t\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^4]\le C\theta_i^2\theta_j\theta_k^3\theta_\ell^2\theta_t^2, & \text{if } (k',t')=(\ell,i),\ \{\ell',i'\}=\{k,t\},\ j'=i;\\ \eta_i\eta_\ell\eta_k\eta_t\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^2W_{j'\ell}^2]\le C\theta_i^2\theta_j\theta_k^3\theta_\ell^3\theta_t^2\theta_{j'}, & \text{if } (k',t')=(\ell,i),\ \{\ell',i'\}=\{k,t\},\ j'\ne i;\\ \eta_i\eta_\ell\eta_k\eta_t\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^4]\le C\theta_i^2\theta_j\theta_k^3\theta_\ell^2\theta_t^2, & \text{if } (k',t')=(i,\ell),\ \{\ell',i'\}=\{k,t\},\ j'=\ell;\\ \eta_i\eta_\ell\eta_k\eta_t\,\mathbb{E}[W_{jk}^2W_{kt}^2W_{\ell i}^2W_{j'i}^2]\le C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_t^2\theta_{j'}, & \text{if } (k',t')=(i,\ell),\ \{\ell',i'\}=\{k,t\},\ j'\ne \ell;\\ \eta_i^2\eta_\ell^2\,\mathbb{E}[W_{jk}^3W_{kt}^3W_{\ell i}^2]\le C\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t, & \text{if } (k',t',j')=(k,j,t),\ \{i',\ell'\}=\{i,\ell\};\\ 0, & \text{otherwise.}\end{cases}$$

There are only two four types on the right hand side. It follows that

$$\begin{aligned}\mathrm{Var}(Z^*_{2b})&\le\frac{C}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell,t,j'}\theta_i^3\theta_j\theta_k^3\theta_\ell^3\theta_t\theta_{j'}+\sum_{i,j,k,\ell,t,j'}\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t^2\theta_{j'}\\&\qquad+\sum_{i,j,k,\ell,t}\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t+\sum_{i,j,k,\ell,t}\theta_i^2\theta_j\theta_k^3\theta_\ell^2\theta_t^2\Big)\\&\le\frac{C}{\|\theta\|_1^4}\big(\|\theta\|_3^9\|\theta\|_1^3+\|\theta\|_3^6\|\theta\|^4\|\theta\|_1^2+\|\theta\|_3^6\|\theta\|^2\|\theta\|_1^2+\|\theta\|_3^3\|\theta\|^6\|\theta\|_1\big)\\&\le\frac{C\|\theta\|_3^9}{\|\theta\|_1}.\end{aligned}$$

Last, consider $Z^\dagger_{2b}$. By direct calculations,

$$\eta_i\eta_\ell\eta_{i'}\eta_{\ell'}\cdot\mathbb{E}[W_{js}W_{jk}W_{kt}W_{\ell i}\cdot W_{j's'}W_{j'k'}W_{k't'}W_{\ell'i'}]$$
$$=\begin{cases}\eta_i^2\eta_\ell^2\,\mathbb{E}[W_{js}^2W_{jk}^2W_{kt}^2W_{\ell i}^2]\le C\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_s\theta_t, & \text{if } (j',s')=(j,s),\ (k',t')=(k,t),\ \{\ell',i'\}=\{\ell,i\};\\ \eta_i^2\eta_\ell^2\,\mathbb{E}[W_{js}^2W_{jk}^2W_{kt}^2W_{\ell i}^2]\le C\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_s\theta_t, & \text{if } (j',s')=(k,t),\ (k',t')=(j,s),\ \{\ell',i'\}=\{\ell,i\};\\ 0, & \text{otherwise.}\end{cases}$$

It follows that

$$\mathrm{Var}(Z^\dagger_{2b})\le\frac{C}{\|\theta\|_1^4}\sum_{i,j,k,\ell,s,t}\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_s\theta_t\le\frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2}.$$

Since $\|\theta\|_1\|\theta\|_3^3\ge\|\theta\|^4\to\infty$, the variance of Z^*_{2b} dominates the variances of $\widetilde{Z}_{2b}$ and $Z^\dagger_{2b}$. We thus have

$$\mathrm{Var}(Z_{2b})\le 3\mathrm{Var}(\widetilde{Z}_{2b})+3\mathrm{Var}(Z^*_{2b})+3\mathrm{Var}(Z^\dagger_{2b})\le\frac{C\|\theta\|_3^9}{\|\theta\|_1}.\tag{2.E.141}$$

We now combine (2.E.137), (2.E.139), (2.E.140), and (2.E.141). Since $\|\theta\|_3^6\le\theta_{\max}^2\|\theta\|^4\ll\|\theta\|^6$, the right hand side of (2.E.141) is much smaller than the right hand side of (2.E.139). It yields that

$$\mathbb{E}[Z_2]=2\|\theta\|^4\cdot[1+o(1)],\qquad \mathrm{Var}(Z_2)\le\frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}=o(\|\theta\|^8).$$

This proves the claims of Z_2.

Proof of Lemma 21

It suffices to prove the claims for each of Z_1-Z_6. We have analyzed Z_1-Z_2 under the null hypothesis. The proof for the alternative hypothesis is similar and omitted. We obtain that

$$\big|\mathbb{E}[Z_1]\big| \le C\|\theta\|^4, \qquad \mathrm{Var}(Z_1) \le C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8),$$
$$\big|\mathbb{E}[Z_2]\big| \le C\|\theta\|^4, \qquad \mathrm{Var}(Z_2) \le \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8).$$

First, we analyze Z_3. Since $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$, we have

$$\begin{aligned} Z_3 &= \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\eta_j(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}W_{\ell i} + \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)^2\eta_k\widetilde{\Omega}_{k\ell}W_{\ell i} \\ &+ \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i - \widetilde{\eta}_i)\eta_j^2(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}W_{\ell i} + \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i - \widetilde{\eta}_i)\eta_j(\eta_j - \widetilde{\eta}_j)\eta_k\widetilde{\Omega}_{k\ell}W_{\ell i} \\ &\equiv Z_{3a} + Z_{3b} + Z_{3c} + Z_{3d}. \end{aligned} \tag{2.E.142}$$

First, we study Z_{3a}. By direct calculations,

$$\begin{aligned} Z_{3a} &= \sum_{i,j,k,\ell(dist)} \eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\eta_j\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq k} W_{kt}\Big)\widetilde{\Omega}_{k\ell}W_{\ell i} \\ &= \frac{1}{v}\sum_{\substack{i,j,k,\ell(dist) \\ s\neq j, t\neq k}} \beta_{ijk\ell}W_{js}W_{kt}W_{\ell i}, \qquad \text{where } \beta_{ijk\ell} = \eta_i\eta_j\widetilde{\Omega}_{k\ell}. \end{aligned}$$

Since (i,j,k,ℓ) are distinct, all summands have mean zero. Hence,

$$\mathbb{E}[Z_{3a}] = 0. \tag{2.E.143}$$

To bound its variance, re-write

$$\begin{aligned} Z_{3a} &= \frac{1}{v}\sum_{i,j,k,\ell(dist)} \beta_{ijk\ell}W_{jk}^2W_{\ell i} + \frac{1}{v}\sum_{\substack{i,j,k,\ell(dist) \\ s\neq j, t\neq k, (s,t)\neq(k,j)}} \beta_{ijk\ell}W_{js}W_{kt}W_{\ell i} \\ &\equiv \widetilde{Z}_{3a} + Z_{3a}^*. \end{aligned}$$

We note that $|\beta_{ijk\ell}| \le C\alpha\theta_i\theta_j\theta_k\theta_\ell$ by (2.E.110) and (2.E.117). Consider the variance of $\widetilde{Z}_{3a}$. By direct calculations,

$$\begin{aligned} &\beta_{ijk\ell}\beta_{i'j'k'\ell'} \cdot \mathrm{Cov}(W_{jk}^2W_{\ell i},\, W_{j'k'}^2W_{\ell' i'}) \\ &= \begin{cases} C\alpha^2\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{jk}^4W_{\ell i}^2] \le C\alpha^2\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3, & \text{if } \{\ell',i'\} = \{\ell,i\},\ \{j',k'\} = \{j,k\}; \\ C\alpha^2\theta_i^2\theta_j\theta_k\theta_\ell^2\theta_{j'}\theta_{k'}\,\mathbb{E}[W_{jk}^2W_{j'k'}^2W_{\ell i}^2] \le C\alpha^2\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_{j'}^2\theta_{k'}^2, & \text{if } \{\ell',i'\} = \{\ell,i\},\ \{j',k'\} \neq \{j,k\}; \\ C\alpha^2\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{jk}^3W_{\ell i}^3] \le C\alpha^2\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3, & \text{if } \{j',k'\} = \{\ell,i\},\ \{\ell',i'\} = \{j,k\}; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

It follows that

$$
\begin{aligned}
\mathrm{Var}(\widetilde{Z}_{3a}) &\le \frac{C\alpha^2}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell}\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3 + \sum_{i,j,k,\ell,j',k'}\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_{j'}^2\theta_{k'}^2\Big)\\
&\le \frac{C\alpha^2}{\|\theta\|_1^4}\big(\|\theta\|_3^{12} + \|\theta\|^8\|\theta\|_3^6\big)\\
&\le \frac{C\alpha^2\|\theta\|_3^{12}}{\|\theta\|_1^2}.
\end{aligned}
$$

Consider the variance of Z_{3a}^*. For $W_{js}W_{kt}W_{\ell i}$ and $W_{j's'}W_{k't'}W_{\ell' i'}$ to be correlated, all W terms have to be perfectly paired. By symmetry across indices, it reduces to three cases: (i) $(\ell',i')=(\ell,i)$, $(j',s')=(j,s)$, $(k',t')=(k,t)$; (ii) $(\ell',i')=(j,s)$, $(j',s')=(\ell,i)$, $(k',t')=(k,t)$; (iii) $(\ell',i')=(j,s)$, $(j',s')=(k,t)$, $(k',t')=(\ell,i)$. It follows that

$$
\begin{aligned}
&\beta_{ijk\ell}\beta_{i'j'k'\ell'}\cdot\mathbb{E}[W_{js}W_{kt}W_{\ell i}\cdot W_{j's'}W_{k't'}W_{\ell'i'}]\\
&\le C\alpha^2(\theta_i\theta_j\theta_k\theta_\ell)(\theta_{i'}\theta_{j'}\theta_{k'}\theta_{\ell'})\cdot\mathbb{E}[W_{js}^2W_{kt}^2W_{\ell i}^2]\\
&\le \begin{cases}
C\alpha^2\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\mathbb{E}[W_{js}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3\theta_s\theta_t, & \text{case (i)}\\
C\alpha^2(\theta_i\theta_j\theta_k\theta_\ell)(\theta_s\theta_\ell\theta_k\theta_j)\mathbb{E}[W_{js}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\theta_i^2\theta_j^3\theta_k^3\theta_\ell^3\theta_s^2\theta_t, & \text{case (ii)}\\
C\alpha^2(\theta_i\theta_j\theta_k\theta_\ell)(\theta_s\theta_k\theta_\ell\theta_j)\mathbb{E}[W_{js}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\theta_i^2\theta_j^3\theta_k^3\theta_\ell^3\theta_s^2\theta_t, & \text{case (iii)}\\
0, & \text{otherwise.}
\end{cases}
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\mathrm{Var}(Z_{3a}^*) &\le \frac{C}{\|\theta\|_1^4}\Big(\sum_{i,j,k,\ell,s,t}\alpha^2\theta_i^3\theta_j^3\theta_k^3\theta_\ell^3\theta_s\theta_t + \sum_{i,j,k,\ell,s,t}\alpha^2\theta_i^2\theta_j^3\theta_k^3\theta_\ell^3\theta_s^2\theta_t\Big)\\
&\le \frac{C\alpha^2}{\|\theta\|_1^4}\big(\|\theta\|_3^{12}\|\theta\|_1^2 + \|\theta\|^4\|\theta\|_3^9\|\theta\|_1\big)\\
&\le \frac{C\alpha^2\|\theta\|_3^{12}}{\|\theta\|_1^2}.
\end{aligned}
$$

Combining the variance of $\widetilde{Z}_{3a}$ and Z_{3a}^* gives

$$
\mathrm{Var}(Z_{3a}) \le \frac{C\alpha^2\|\theta\|_3^{12}}{\|\theta\|_1^2}. \tag{2.E.144}
$$

Second, we study Z_{3b}. It is seen that

$$
\begin{aligned}
Z_{3b} &= \sum_{i,j,k,\ell(dist)}\eta_i\Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j}W_{js}\Big)\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq j}W_{jt}\Big)\eta_k\widetilde{\Omega}_{k\ell}W_{\ell i}\\
&= \frac{1}{v}\sum_{\substack{i,j,\ell(dist)\\ s\neq j,t\neq j}}\Big(\sum_{k\notin\{i,j,\ell\}}\eta_i\eta_k\widetilde{\Omega}_{k\ell}\Big)W_{js}W_{jt}W_{\ell i}\\
&\equiv \frac{1}{v}\sum_{\substack{i,j,\ell(dist)\\ s\neq j,t\neq j}}\beta_{ij\ell}W_{js}W_{jt}W_{\ell i},
\end{aligned}
$$

where by (2.E.110) and (2.E.117),

$$
|\beta_{ij\ell}| \le \sum_{k\notin\{i,j,\ell\}}|\eta_i\eta_k\widetilde{\Omega}_{k\ell}| \le \sum_k C\alpha\theta_i\theta_k^2\theta_\ell \le C\alpha\|\theta\|^2\cdot\theta_i\theta_\ell. \tag{2.E.145}
$$

We further decompose Z_{3b} into

$$Z_{3b} = \frac{1}{v} \sum_{\substack{i,j,\ell(dist)\\ s\neq j}} \beta_{ij\ell} W_{js}^2 W_{\ell i} + \frac{1}{v} \sum_{\substack{i,j,\ell(dist)\\ s,t(dist)\notin\{j\}}} \beta_{ij\ell} W_{js} W_{jt} W_{\ell i} \equiv \widetilde{Z}_{3b} + Z_{3b}^*.$$

It is easy to see that both terms have mean zero. It follows that

$$\mathbb{E}[Z_{3b}] = 0. \tag{2.E.146}$$

To calculate the variance of $\widetilde{Z}_{3b}$, we note that

$$\begin{aligned}
&\beta_{ij\ell}\beta_{i'j'\ell'} \cdot \mathbb{E}[W_{js}^2 W_{\ell i} \cdot W_{j's'}^2 W_{\ell' i'}] \\
&\leq C\alpha^2\|\theta\|^4\theta_i\theta_{i'}\theta_\ell\theta_{\ell'} \cdot \mathbb{E}[W_{js}^2 W_{\ell i} \cdot W_{j's'}^2 W_{\ell' i'}] \\
&\leq \begin{cases}
C\alpha^2\|\theta\|^4\theta_i^2\theta_\ell^2 \cdot \mathbb{E}[W_{js}^4 W_{\ell i}^2] \leq C\alpha^2\|\theta\|^4\theta_i^3\theta_j\theta_\ell^3\theta_s & \text{if } \{\ell', i'\} = \{\ell, i\},\ \{j', s'\} = \{j, s\} \\
C\alpha^2\|\theta\|^4\theta_i^2\theta_\ell^2 \cdot \mathbb{E}[W_{js}^2 W_{\ell i}^2 W_{j's'}^2] \leq C\alpha^2\|\theta\|^4\theta_i^3\theta_j\theta_\ell^3\theta_s\theta_{j'}\theta_{s'}, & \text{if } \{\ell', i'\} = \{\ell, i\},\ \{j', s'\} \neq \{j, s\}; \\
C\alpha^2\|\theta\|^4\theta_i\theta_\ell\theta_j\theta_s \cdot \mathbb{E}[W_{js}^3 W_{\ell i}^3] \leq C\alpha^2\|\theta\|^4\theta_i^2\theta_j^2\theta_\ell^2\theta_s^2, & \text{if } \{\ell', i'\} = \{j, s\},\ \{j', s'\} = \{\ell, i\}; \\
0, & \text{otherwise.}
\end{cases}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Var}(\widetilde{Z}_{3b}) &\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\Big(\sum_{i,j,\ell,s}\theta_i^3\theta_j\theta_\ell^3\theta_s + \sum_{i,j,\ell,s,j',s'}\theta_i^3\theta_j\theta_\ell^3\theta_s\theta_{j'}\theta_{s'} + \sum_{i,j,\ell,s,j',s'}\theta_i^2\theta_j^2\theta_\ell^2\theta_s^2\Big) \\
&\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|_3^6\|\theta\|_1^4 + \|\theta\|^8\big) \\
&\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6.
\end{aligned}$$

To calculate the variance of Z_{3b}^*, we note that $\mathrm{E}[W_{js}W_{jt}W_{\ell i} \cdot W_{j's'}W_{j't'}W_{\ell' i'}]$ is nonzero only if $j' = j$, $\{s', t'\} = \{s, t\}$ and $\{\ell', i'\} = \{\ell, i\}$. Combining it with (2.E.148) gives

$$\begin{aligned}
\mathrm{Var}(Z_{3b}^*) &\leq \frac{C}{v^2} \sum_{\substack{i,j,\ell(dist)\\ s,t(dist)\notin\{j\}}} \beta_{ij\ell}^2 \cdot \mathbb{E}[W_{js}^2 W_{jt}^2 W_{\ell i}^2] \\
&\leq \frac{C}{\|\theta\|_1^4} \sum_{i,j,\ell,s,t} (\alpha\|\theta\|^2\theta_i\theta_\ell)^2 \cdot \theta_j^2\theta_s\theta_t\theta_\ell\theta_i \\
&\leq \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4} \sum_{i,j,\ell,s,t} \theta_i^3\theta_j^2\theta_\ell^3\theta_s\theta_t \\
&\leq \frac{C\alpha^2\|\theta\|^6\|\theta\|_3^6}{\|\theta\|_1^2}.
\end{aligned}$$

Since $\|\theta\|^6 \leq \|\theta\|^4\|\theta\|^2 \ll \|\theta\|^4\|\theta\|_1$, the variance of $\widetilde{Z}_{3b}$ dominates the variance of Z_{3b}^*. Combining the above gives

$$\mathrm{Var}(Z_{3b}) \leq 2\mathrm{Var}(\widetilde{Z}_{3b}) + 2\mathrm{Var}(Z_{3b}^*) \leq C\alpha^2\|\theta\|^4\|\theta\|_3^6. \tag{2.E.147}$$

Third, we study Z_{3c}. It is seen that

$$\begin{aligned} Z_{3c} &= \sum_{i,j,k,\ell(dist)} \Big(-\frac{1}{\sqrt{v}}\sum_{s\neq i} W_{is}\Big)\eta_j^2\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq k} W_{kt}\Big)\widetilde{\Omega}_{k\ell}W_{\ell i} \\ &= \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\neq i, t\neq k}} \Big(\sum_{j\notin\{i,k,\ell\}} \eta_j^2\widetilde{\Omega}_{k\ell}\Big) W_{is}W_{kt}W_{\ell i} \\ &\equiv \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\neq i, t\neq k}} \beta_{ik\ell}W_{is}W_{kt}W_{\ell i}, \end{aligned}$$

where by (2.E.110) and (2.E.117),

$$|\beta_{ik\ell}| \leq \sum_{j\notin\{i,k,\ell\}} |\eta_j^2\widetilde{\Omega}_{k\ell}| \leq \sum_j C\alpha\theta_j^2\theta_k\theta_\ell \leq C\alpha\|\theta\|^2\theta_k\theta_\ell. \tag{2.E.148}$$

There are two cases for the indices: $i=\ell$ and $i\neq\ell$. We further decompose Z_{3c} into

$$Z_{3c} = \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ t\neq k}} \beta_{ik\ell}W_{i\ell}^2W_{kt} + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\}, t\neq k}} \beta_{ik\ell}W_{is}W_{kt}W_{\ell i} \equiv \widetilde{Z}_{3c} + Z_{3c}^*.$$

It is easy to see that both terms have zero mean. Hence,

$$\mathbb{E}[Z_{3c}] = 0. \tag{2.E.149}$$

To calculate the variance of $\widetilde{Z}_{3c}$, we note that $W_{i\ell}^2W_{kt}$ and $W_{i'\ell'}^2W_{k't'}$ are correlated only when (i) $\{k',t'\}=\{k,t\}$ or (ii) $\{k',t'\}=\{i,\ell\}$ and $\{i',\ell'\}=\{k,t\}$. By direct calculations,

$$\begin{aligned} &\beta_{ik\ell}\beta_{i'k'\ell'}\cdot\mathbb{E}[W_{i\ell}^2W_{kt}\cdot W_{i'\ell'}^2W_{k't'}] \\ &\leq C\alpha^2\|\theta\|^4\theta_k\theta_{k'}\theta_\ell\theta_{\ell'}\cdot\mathbb{E}[W_{i\ell}^2W_{kt}\cdot W_{i'\ell'}^2W_{k't'}] \\ &\leq \begin{cases} C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{i\ell}^4W_{kt}^2] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^3\theta_\ell^3\theta_t, & \text{if } (k',t')=(k,t),\ (i',\ell')=(i,\ell); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell\theta_i\,\mathbb{E}[W_{i\ell}^4W_{kt}^2] \leq C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^2\theta_t, & \text{if } (k',t')=(k,t),\ (i',\ell')=(\ell,i); \\ C\alpha^2\|\theta\|^4\theta_k\theta_\ell^2\theta_t\,\mathbb{E}[W_{i\ell}^4W_{kt}^2] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^2\theta_\ell^3\theta_t^2, & \text{if } (k',t')=(t,k),\ (i',\ell')=(i,\ell); \\ C\alpha^2\|\theta\|^4\theta_k\theta_t\theta_\ell\theta_i\,\mathbb{E}[W_{i\ell}^4W_{kt}^2] \leq C\alpha^2\|\theta\|^4\theta_i^2\theta_k^2\theta_\ell^2\theta_t^2, & \text{if } (k',t')=(t,k),\ (i',\ell')=(\ell,i); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell\theta_{\ell'}\,\mathbb{E}[W_{i\ell}^2W_{kt}^2W_{i'\ell'}^2] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^3\theta_\ell^2\theta_t\theta_{i'}\theta_{\ell'}^2, & \text{if } (k',t')=(k,t),\ \{i',\ell'\}\neq\{i,\ell\}; \\ C\alpha^2\|\theta\|^4\theta_k\theta_t\theta_\ell\theta_{\ell'}\,\mathbb{E}[W_{i\ell}^2W_{kt}^2W_{i'\ell'}^2] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^2\theta_\ell^2\theta_t^2\theta_{i'}\theta_{\ell'}^2, & \text{if } (k',t')=(t,k),\ \{i',\ell'\}\neq\{i,\ell\}; \\ C\alpha^2\|\theta\|^4\theta_k\theta_i\theta_\ell\theta_t\,\mathbb{E}[W_{i\ell}^3W_{kt}^3] \leq C\alpha^2\|\theta\|^4\theta_i^2\theta_k^2\theta_\ell^2\theta_t^2, & \text{if } (k',t')=(i,\ell),\ (i',\ell')=(k,t); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_i\theta_\ell\,\mathbb{E}[W_{i\ell}^3W_{kt}^3] \leq C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^2\theta_t, & \text{if } (k',t')=(i,\ell),\ (i',\ell')=(t,k); \\ C\alpha^2\|\theta\|^4\theta_k\theta_\ell^2\theta_t\,\mathbb{E}[W_{i\ell}^3W_{kt}^3] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^2\theta_\ell^3\theta_t^2, & \text{if } (k',t')=(\ell,i),\ (i',\ell')=(k,t); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{i\ell}^3W_{kt}^3] \leq C\alpha^2\|\theta\|^4\theta_i\theta_k^3\theta_\ell^3\theta_t, & \text{if } (k',t')=(\ell,i),\ (i',\ell')=(t,k); \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

There are only five types on the right hand side. It follows that

$$\begin{aligned}\mathrm{Var}(\widetilde{Z}_{3c}) &\le \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\Big(\sum_{i,k,\ell,t}\theta_i\theta_k^3\theta_\ell^3\theta_t + \sum_{i,k,\ell,t}\theta_i^2\theta_k^3\theta_\ell^2\theta_t + \sum_{i,k,\ell,t}\theta_i^2\theta_k^2\theta_\ell^2\theta_t^2 \\ &\qquad + \sum_{i,k,\ell,t,i',\ell'}\theta_i\theta_k^3\theta_\ell^2\theta_t\theta_{i'}\theta_{\ell'}^2 + \sum_{i,k,\ell,t,i',\ell'}\theta_i\theta_k^2\theta_\ell^2\theta_t^2\theta_{i'}\theta_{\ell'}^2\Big) \\ &\le \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|_3^6\|\theta\|_1^2 + \|\theta\|^4\|\theta\|_3^3\|\theta\|_1 + \|\theta\|^8 + \|\theta\|^4\|\theta\|_3^3\|\theta\|_1^3 + \|\theta\|^8\|\theta\|_1^2\big) \\ &\le \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1},\end{aligned}$$

where the last inequality is obtained as follows: Among the five terms in the brackets, the first and third terms are dominated by the last term, and the second term is dominated by the fourth term; it remains to compare the fourth term and the last term, where the fourth term dominated because $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$. To calculate the variance of Z_{3c}^*, we write

$$Z_{3c}^* = \frac{1}{v}\sum_{i,k,\ell(dist)}\beta_{ik\ell}W_{ik}^2W_{\ell i} + \frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\},t\neq k,(s,t)\neq(k,i)}}\beta_{ik\ell}W_{is}W_{kt}W_{\ell i}.$$

Regarding the first term, we note that

$$\begin{aligned}&\beta_{ik\ell}\beta_{i'k'\ell'}\cdot\mathbb{E}[W_{ik}^2W_{\ell i}\cdot W_{i'k'}^2W_{\ell'i'}] \\ \le\,& C\alpha^2\|\theta\|^4\theta_k\theta_\ell\theta_{k'}\theta_{\ell'}\cdot\mathbb{E}[W_{ik}^2W_{\ell i}\cdot W_{i'k'}^2W_{\ell'i'}] \\ \le\,& \begin{cases} C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{ik}^4W_{\ell i}^2] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^3, & \text{if } (\ell',i')=(\ell,i),\ k'=k; \\ C\alpha^2\|\theta\|^4\theta_k\theta_\ell^2\theta_{k'}\,\mathbb{E}[W_{ik}^2W_{\ell i}^2W_{ik'}^2] \le C\alpha^2\|\theta\|^4\theta_i^3\theta_k^2\theta_\ell^3\theta_{k'}^2, & \text{if } (\ell',i')=(\ell,i),\ k'\neq k; \\ C\alpha^2\|\theta\|^4\theta_i\theta_k\theta_\ell\theta_{k'}\,\mathbb{E}[W_{ik}^2W_{\ell i}^2W_{\ell k'}^2] \le C\alpha^2\|\theta\|^4\theta_i^3\theta_k^2\theta_\ell^3\theta_{k'}^2, & \text{if } (\ell',i')=(i,\ell); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{ik}^3W_{\ell i}^3] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^3, & \text{if } (\ell',i')=(k,i),\ k'=\ell; \\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

It follows that

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{v}\sum_{i,k,\ell(dist)}\beta_{ik\ell}W_{ik}^2W_{\ell i}\Big) &\le \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\Big(\sum_{i,k,\ell}\theta_i^2\theta_k^3\theta_\ell^3 + \sum_{i,k,\ell,k'}\theta_i^3\theta_k^2\theta_\ell^3\theta_{k'}^2\Big) \\ &\le \frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_3^6 + \|\theta\|^4\|\theta\|_3^6\big) \le \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^4}.\end{aligned}$$

Regarding the second term, we note that

$$\begin{aligned}&\beta_{ik\ell}\beta_{i'k'\ell'}\cdot\mathbb{E}[W_{is}W_{kt}W_{\ell i}\cdot W_{i's'}W_{k't'}W_{\ell'i'}] \\ \le\,& C\alpha^2\|\theta\|^4\theta_k\theta_{k'}\theta_\ell\theta_{\ell'}\cdot\mathbb{E}[W_{is}W_{kt}W_{\ell i}\cdot W_{i's'}W_{k't'}W_{\ell'i'}] \\ \le\,& \begin{cases} C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell^2\,\mathbb{E}[W_{is}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^3\theta_s\theta_t, & \text{if } (i',s',\ell')=(i,s,\ell),\ (k',t')=(k,t); \\ C\alpha^2\|\theta\|^4\theta_k\theta_t\theta_\ell^2\,\mathbb{E}[W_{is}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^2\theta_\ell^3\theta_s\theta_t^2, & \text{if } (i',s',\ell')=(i,s,\ell),\ (k',t')=(t,k); \\ C\alpha^2\|\theta\|^4\theta_k^2\theta_\ell\theta_s\,\mathbb{E}[W_{is}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^3\theta_\ell^2\theta_s^2\theta_t, & \text{if } (i',s',\ell')=(i,\ell,s),\ (k',t')=(k,t); \\ C\alpha^2\|\theta\|^4\theta_k\theta_t\theta_\ell\theta_s\,\mathbb{E}[W_{is}^2W_{kt}^2W_{\ell i}^2] \le C\alpha^2\|\theta\|^4\theta_i^2\theta_k^2\theta_\ell^2\theta_s^2\theta_t^2, & \text{if } (i',s',\ell')=(i,\ell,s),\ (k',t')=(t,k); \\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

It follows that

$$\mathrm{Var}\Big(\frac{1}{v}\sum_{\substack{i,k,\ell(dist)\\ s\notin\{i,\ell\},t\neq k,\\ (s,t)\neq(k,i)}}\beta_{ik\ell}W_{is}W_{kt}W_{\ell i}\Big)\leq\frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\sum_{\substack{i,k,\ell,\\ s,t}}(\theta_i^2\theta_k^3\theta_\ell^3\theta_s\theta_t+\theta_i^2\theta_k^2\theta_\ell^3\theta_s\theta_t^2+\theta_i^2\theta_k^2\theta_\ell^2\theta_s^2\theta_t^2)$$

$$\leq\frac{C\alpha^2\|\theta\|^4}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^2+\|\theta\|^6\|\theta\|_3^3\|\theta\|_1+\|\theta\|^{10}\big)$$

$$\leq\frac{C\alpha^2\|\theta\|^6\|\theta\|_3^6}{\|\theta\|_1^2}.$$

We plug the above results into Z_{3c}^*. Since $\|\theta\|^2\leq\|\theta\|_1\theta_{\max}\ll\|\theta\|_1^2$, we have $\frac{C\alpha^2\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^4}\ll\frac{C\alpha^2\|\theta\|^6\|\theta\|_3^6}{\|\theta\|_1^2}$. It follows that

$$\mathrm{Var}(Z_{3c}^*)\leq\frac{C\alpha^2\|\theta\|^6\|\theta\|_3^6}{\|\theta\|_1^2}.$$

Since $\|\theta\|_3^6\ll\|\theta\|_3^3\|\theta\|_1$, the variance of Z_{3c}^* is dominated by the variance of $\widetilde{Z}_{3c}$. It follows that

$$\mathrm{Var}(Z_{3c})\leq 2\mathrm{Var}(\widetilde{Z}_{3c})+2\mathrm{Var}(Z_{3c}^*)\leq\frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1}. \tag{2.E.150}$$

Last, we study Z_{3d}. In the definition of Z_{3d}, if we switch the two indices (j,k), then it becomes

$$Z_{3d}=\sum_{\substack{i,j,k,\ell\\(dist)}}(\eta_i-\tilde{\eta}_i)\eta_k(\eta_k-\tilde{\eta}_k)\eta_j\widetilde{\Omega}_{j\ell}W_{\ell i}=\sum_{\substack{i,j,k,\ell\\(dist)}}(\eta_k\eta_j\widetilde{\Omega}_{j\ell})(\eta_i-\tilde{\eta}_i)(\eta_k-\tilde{\eta}_k).$$

At the same time, we recall that

$$Z_{3c}=\sum_{\substack{i,j,k,\ell\\(dist)}}(\eta_i-\tilde{\eta}_i)\eta_j^2(\eta_k-\tilde{\eta}_k)\widetilde{\Omega}_{k\ell}W_{\ell i}=\sum_{\substack{i,j,k,\ell\\(dist)}}(\eta_j^2\widetilde{\Omega}_{k\ell})(\eta_i-\tilde{\eta}_i)(\eta_k-\tilde{\eta}_k).$$

Here, Z_{3d} has a similar structure as Z_{3c}. Moreover, in deriving the bound for $\mathrm{Var}(Z_{3c})$, we have used $|\eta_j^2\widetilde{\Omega}_{k\ell}|\leq C\alpha\theta_j^2\theta_k\theta_\ell$. In the expression of Z_{3d} above, we also have $|\eta_k\eta_j\widetilde{\Omega}_{j\ell}|\leq C\alpha\theta_j^2\theta_k\theta_\ell$. Therefore, we can use (2.E.149) and (2.E.150) to directly get

$$\mathbb{E}[Z_{3d}]=0,\qquad \mathrm{Var}(Z_{3d})\leq\frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} \tag{2.E.151}$$

Now, we combine (2.E.143), (2.E.146), (2.E.149) and (2.E.150) to get

$$\mathbb{E}[Z_3]=0.$$

We also combine (2.E.144), (2.E.147), (2.E.150)-(2.E.151). Since $\|\theta\|^4\leq\|\theta\|_1\|\theta\|_3^3$, the right hand side of (2.E.150)-(2.E.151) is dominated by the right hand side of (2.E.147); since $\|\theta\|_3^6\ll\|\theta\|_1^2$, the right hand side of (2.E.144) is negligible to the right hand side of (2.E.147). It follows that

$$\mathrm{Var}(Z_3)\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6.$$

This proves the claims of Z_3.

Next, we analyze Z_4. Since $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$,

$$Z_4 = \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell W_{\ell i}$$
$$+ \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)W_{\ell i} + \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell W_{\ell i}.$$

If we relabel (i, j, k, ℓ) as (ℓ', k', j', i') in the last sum, it is equal to the first sum. Therefore,

$$\begin{aligned} Z_4 = 2 &\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)W_{\ell i} \\ &+ \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell W_{\ell i} \\ &+ \sum_{i,j,k,\ell(dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)W_{\ell i} \\ &\equiv Z_{4a} + Z_{4b} + Z_{4c}. \end{aligned} \tag{2.E.152}$$

First, we study Z_{4a} and Z_{4b}. We show that they have the same structure as Z_{3c} and Z_{3a}, respectively. In Z_{4a}, by relabeling (i, j, k, ℓ) as (ℓ, k, j, i), we have

$$Z_{4a} = 2 \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_\ell(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{kj}\eta_j(\eta_i - \widetilde{\eta}_i)W_{i\ell} = 2 \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_j\eta_\ell\widetilde{\Omega}_{kj})(\eta_i - \widetilde{\eta}_i)(\eta_k - \widetilde{\eta}_k)W_{\ell i}.$$

At the same time, we recall the definition of Z_{3c} in (2.E.142):

$$Z_{3c} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i - \widetilde{\eta}_i)\eta_j^2(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}W_{\ell i} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_j^2\widetilde{\Omega}_{k\ell})(\eta_i - \widetilde{\eta}_i)(\eta_k - \widetilde{\eta}_k)W_{\ell i}.$$

It is seen that Z_{4a} has a similar structure as Z_{3c} does. Also, by (2.E.110) and (2.E.117), in the expression of Z_{4a}, we have $|\eta_j\eta_\ell\widetilde{\Omega}_{kj}| \le C\alpha\theta_j^2\theta_k\theta_\ell$, while in the expression of Z_{3d}, we have $|\eta_j^2\widetilde{\Omega}_{k\ell}| \le C\alpha\theta_j^2\theta_k\theta_\ell$. As a result, if we use similar calculation as before, we will get the same conclusion for Z_{4a} and Z_{3d}. Hence, we use (2.E.149)-(2.E.150) to conclude that

$$\mathbb{E}[Z_{4a}] = 0, \qquad \mathrm{Var}(Z_{4a}) \le \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1}. \tag{2.E.153}$$

For Z_{4b}, we note that

$$Z_{4b} = \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell W_{\ell i} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i\eta_\ell\widetilde{\Omega}_{jk})(\eta_j - \widetilde{\eta}_j)(\eta_k - \widetilde{\eta}_k)W_{\ell i},$$

where $|\eta_i\eta_\ell\widetilde{\Omega}_{jk}| \le C\alpha\theta_i\theta_j\theta_k\theta_\ell$. At the same time, we recall the definition of Z_{3a} in (2.E.142):

$$Z_{3a} = \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\eta_j(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}W_{\ell i} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i\eta_j\widetilde{\Omega}_{k\ell})(\eta_j - \widetilde{\eta}_j)(\eta_k - \widetilde{\eta}_k)W_{\ell i},$$

where $|\eta_i\eta_j\widetilde{\Omega}_{k\ell}| \le C\alpha\theta_i\theta_j\theta_k\theta_\ell$. Therefore, we can quote the results for Z_{3a} in (2.E.143)-(2.E.144) to get

$$\mathbb{E}[Z_{4b}] = 0, \qquad \mathrm{Var}(Z_{4b}) \le \frac{C\alpha^2\|\theta\|_3^{12}}{\|\theta\|_1^2}. \tag{2.E.154}$$

Second, we study Z_{4c}. It is seen that

$$\begin{aligned} Z_{4c} &= \sum_{i,j,k,\ell(dist)} \Big(-\frac{1}{\sqrt{v}}\sum_{s\neq i} W_{is}\Big)\eta_j\widetilde{\Omega}_{jk}\eta_k\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq \ell} W_{\ell t}\Big)W_{\ell i} \\ &= \frac{1}{v}\sum_{\substack{i,\ell(dist)\\ s\neq i, t\neq \ell}} \Big(\sum_{j,k(dist)\notin\{i,\ell\}} \eta_j\eta_k\widetilde{\Omega}_{jk}\Big)W_{is}W_{\ell t}W_{\ell i} \\ &\equiv \frac{1}{v}\sum_{\substack{i,\ell(dist)\\ s\neq i, t\neq \ell}} \beta_{i\ell}W_{is}W_{\ell t}W_{\ell i}, \end{aligned}$$

where

$$|\beta_{i\ell}| \leq \sum_{j,k(dist)\notin\{i,\ell\}} |\eta_j\eta_k\widetilde{\Omega}_{jk}| \leq \sum_{j,k} C\alpha\theta_j^2\theta_k^2 \leq C\alpha\|\theta\|^4. \tag{2.E.155}$$

We divide the summands into four groups: (i) $s=\ell$, $t=i$; (ii) $s=\ell$, $t\neq i$; (iii) $s\neq \ell$, $t=i$; (iv) $s\neq\ell$, $t\neq i$. By symmetry, the sum of group (ii) and the sum of group (iii) are equal. It yields that

$$\begin{aligned} Z_{4c} &= \frac{1}{v}\sum_{i,\ell(dist)} \beta_{i\ell}W_{\ell i}^3 + \frac{2}{v}\sum_{\substack{i,\ell(dist)\\ s\notin\{i,\ell\}}} \beta_{i\ell}W_{is}W_{\ell i}^2 + \frac{1}{v}\sum_{\substack{i,\ell(dist)\\ s\notin\{i,\ell\}, t\notin\{\ell,i\}}} \beta_{i\ell}W_{is}W_{\ell t}W_{\ell i} \\ &\equiv \widetilde{Z}_{4c} + Z_{4c}^* + Z_{4c}^\dagger. \end{aligned}$$

Only $\widetilde{Z}_{4c}$ has a nonzero mean. By (2.E.116) and (2.E.155),

$$\big|\mathbb{E}[Z_{4c}]\big| = \big|\mathbb{E}[\widetilde{Z}_{4c}]\big| \leq \frac{C}{\|\theta\|_1^2}\sum_{i,\ell} \alpha\|\theta\|^4\theta_i\theta_\ell \leq C\alpha\|\theta\|^4. \tag{2.E.156}$$

We now compute the variances of these terms. It is seen that

$$\mathrm{Var}(\widetilde{Z}_{4c}) \leq \frac{C}{v^2}\sum_{i,\ell(dist)} \beta_{i\ell}^2\mathrm{Var}(W_{i\ell}^3) \leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^4}\sum_{i,\ell}\theta_i\theta_\ell \leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2}.$$

For Z_{4c}^*, by direct calculations,

$$\begin{aligned} &\beta_{i\ell}\beta_{i'\ell'}\cdot\mathbb{E}[W_{is}W_{\ell i}^2\cdot W_{i's'}W_{\ell' i'}^2] \\ &\leq C\alpha^2\|\theta\|^8\cdot\mathbb{E}[W_{is}W_{\ell i}^2\cdot W_{i's'}W_{\ell' i'}^2] \\ &\leq \begin{cases} C\alpha^2\|\theta\|^8\cdot\mathbb{E}[W_{is}^2W_{\ell i}^4] \leq C\alpha^2\|\theta\|^8\theta_i^2\theta_\ell\theta_s, & \text{if } i'=i,\ s'=s,\ \ell'=\ell; \\ C\alpha^2\|\theta\|^8\cdot\mathbb{E}[W_{is}^2W_{\ell i}^2W_{\ell' i}^2] \leq C\alpha^2\|\theta\|^8\theta_i^3\theta_\ell\theta_s\theta_{\ell'}, & \text{if } i'=i,\ s'=s,\ \ell'\neq\ell; \\ C\alpha^2\|\theta\|^8\cdot\mathbb{E}[W_{is}^3W_{\ell i}^3] \leq C\alpha^2\|\theta\|^8\theta_i^2\theta_\ell\theta_s, & \text{if } i'=i,\ s'=\ell,\ \ell'=s; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

It follows that

$$\begin{aligned} \mathrm{Var}(Z_{4c}^*) &\leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^4}\Big(\sum_{i,\ell,s}\theta_i^2\theta_\ell\theta_s + \sum_{i,\ell,s,\ell'}\theta_i^3\theta_\ell\theta_s\theta_{\ell'}\Big) \\ &\leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^4}\big(\|\theta\|^2\|\theta\|_1^2 + \|\theta\|_3^3\|\theta\|_1^3\big) \\ &\leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1}, \end{aligned}$$

where, to get the last line, we have used $\|\theta\|^2 \ll \|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$. Regarding the variance of $Z^{\dagger}_{4c}$, we note that $W_{is}W_{\ell t}W_{\ell i}$ and $W_{i's'}W_{\ell' t'}W_{\ell' i'}$ are correlated only when the two undirected paths s-i-ℓ-t and s'-i'-ℓ'-t' are the same. Mimicking the argument in (2.E.121) or (2.E.126), we can derive that

$$\begin{aligned}\mathrm{Var}(Z^{\dagger}_{4c}) &\leq \frac{C}{v^2}\sum_{\substack{i,\ell(dist)\\ s\notin\{i,\ell\},t\notin\{\ell,i\}}} \beta_{i\ell}^2\cdot\mathrm{Var}(W_{is}W_{\ell t}W_{\ell i})\\ &\leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^4}\sum_{i,\ell,s,t}\theta_i^2\theta_\ell^2\theta_s\theta_t\\ &\leq \frac{C\alpha^2\|\theta\|^{12}}{\|\theta\|_1^2}.\end{aligned}$$

Since $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the variance of $Z^{\dagger}_{4c}$ is dominated by the variance of Z^*_{4c}. Since $\|\theta\|\to\infty$, we have $\|\theta\|_3^3 \gg 1/\|\theta\|_1$; it follows that the variance of $\widetilde{Z}_{4c}$ is dominated by the variance of Z^*_{4c}. Combining the above gives

$$\mathrm{Var}(Z_{4c}) \leq 3\mathrm{Var}(\widetilde{Z}_{4c}) + 3\mathrm{Var}(Z^*_{4c}) + 3\mathrm{Var}(Z^{\dagger}_{4c}) \leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1}. \tag{2.E.157}$$

We combine (2.E.153), (2.E.154) and (2.E.156) to get

$$\big|\mathbb{E}[Z_4]\big| \leq C\alpha\|\theta\|^4 = o(\alpha^4\|\theta\|^8).$$

We then combine (2.E.153), (2.E.154) and (2.E.157). Since $\|\theta\|_3^6 \leq (\theta_{\max}^2\|\theta\|_1)(\theta_{\max}\|\theta\|^2) = o(\|\theta\|_1\|\theta\|^2)$, the variance of Z_{4b} is negligible compared to the variances of Z_{4a} and Z_{4c}. It follows that

$$\mathrm{Var}(Z_4) \leq \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8).$$

This proves the claims of Z_4.

Next, we analyze Z_5. By plugging in the definition of δ_{ij}, we have

$$\begin{aligned}Z_5 = &\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j-\widetilde{\eta}_j)\eta_j(\eta_k-\widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j-\widetilde{\eta}_j)^2\eta_k\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\\ &+ \sum_{i,j,k,\ell(dist)} (\eta_i-\widetilde{\eta}_i)\eta_j^2(\eta_k-\widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell(dist)} (\eta_i-\widetilde{\eta}_i)\eta_j(\eta_j-\widetilde{\eta}_j)\eta_k\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\\ =& \; 2\sum_{i,j,k,\ell(dist)} \eta_i(\eta_j-\widetilde{\eta}_j)\eta_j(\eta_k-\widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_i(\eta_j-\widetilde{\eta}_j)^2\eta_k\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\\ &+ \sum_{i,j,k,\ell(dist)} (\eta_i-\widetilde{\eta}_i)\eta_j^2(\eta_k-\widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\\ \equiv&\; Z_{5a}+Z_{5b}+Z_{5c}.\end{aligned} \tag{2.E.158}$$

First, we study Z_{5a}. By definition, $(\widetilde{\eta}_i-\eta_i)$ has the expression in (2.E.113). It follows that

$$\begin{aligned} Z_{5a} &= 2 \sum_{i,j,k,\ell(dist)} \eta_i \Big(-\frac{1}{\sqrt{v}} \sum_{s\neq j} W_{js}\Big) \eta_j \Big(-\frac{1}{\sqrt{v}} \sum_{t\neq k} W_{kt}\Big) \widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} \\ &= \frac{2}{v} \sum_{\substack{j,k(dist)\\ s\neq j, t\neq k}} \Big(\sum_{i,\ell(dist)\notin\{j,k\}} \eta_i\eta_j \widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\Big) W_{js}W_{kt} \\ &\equiv \frac{2}{v} \sum_{\substack{j,k(dist)\\ s\neq j, t\neq k}} \beta_{jk} W_{js} W_{kt}, \end{aligned}$$

where

$$|\beta_{jk}| \leq \sum_{i,\ell(dist)\notin\{j,k\}} |\eta_i\eta_j \widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}| \leq \sum_{i,\ell} (C\theta_i\theta_j)(C\alpha^2\theta_k\theta_\ell^2\theta_i) \leq C\alpha^2\|\theta\|^4\theta_j\theta_k. \tag{2.E.159}$$

In Z_{5a}, the summand has a nonzero mean only if $(s,t)=(k,j)$. We further decompose Z_{5a} into

$$Z_{5a} = \frac{2}{v}\sum_{j,k(dist)} \beta_{jk}W_{jk}^2 + \frac{2}{v} \sum_{\substack{j,k(dist)\\ s\neq j, t\neq k,\\ (s,t)\neq(k,j)}} \beta_{jk}W_{js}W_{kt} \equiv \widetilde{Z}_{5a} + Z_{5a}^*.$$

Only the first term has a nonzero mean. By (2.E.116) and (2.E.159), we have

$$\big|\mathbb{E}[Z_{5a}]\big| = \big|\mathbb{E}[\widetilde{Z}_{5a}]\big| \leq \frac{C}{\|\theta\|_1^2}\sum_{j,k}(\alpha^2\|\theta\|^4\theta_j\theta_k)(\theta_j\theta_k) \leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2}. \tag{2.E.160}$$

We then compute the variances. In each of $\widetilde{Z}_{5a}$ and Z_{5a}^*, two summands are uncorrelated unless they are exactly the same variables (e.g., when $(j',k')=(k,j)$ in $\widetilde{Z}_{5a}$). Mimicking the argument in (2.E.121) or (2.E.126), we can derive that

$$\begin{aligned} \mathrm{Var}(\widetilde{Z}_{5a}) &\leq \frac{C}{v^2}\sum_{j,k(dist)} \beta_{jk}^2\,\mathrm{Var}(W_{jk}^2) \leq \frac{C\alpha^4\|\theta\|^8}{\|\theta\|_1^4}\sum_{j,k}(\theta_j^2\theta_k^2)\theta_j\theta_k \leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^4}, \\ \mathrm{Var}(Z_{5a}^*) &\leq \frac{C}{v^2}\sum_{\substack{j,k(dist)\\ s\neq j, t\neq k,\\ (s,t)\neq(k,j)}} \beta_{jk}^2\,\mathrm{Var}(W_{js}W_{kt}) \leq \frac{C\alpha^4\|\theta\|^8}{\|\theta\|_1^4}\sum_{j,k}(\theta_j^2\theta_k^2)\theta_j\theta_s\theta_k\theta_t \leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2}. \end{aligned}$$

It immediately leads to

$$\mathrm{Var}(Z_{5a}) \leq 2\mathrm{Var}(\widetilde{Z}_{5a}) + 2\mathrm{Var}(Z_{5a}^*) \leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2}. \tag{2.E.161}$$

Second, we study Z_{5b}. It is seen that

$$\begin{aligned} Z_{5b} &= \sum_{i,j,k,\ell(dist)} \eta_i \Big(-\frac{1}{\sqrt{v}}\sum_{s\neq j} W_{js}\Big)\Big(-\frac{1}{\sqrt{v}}\sum_{t\neq j} W_{jt}\Big)\eta_k \widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} \\ &= \frac{1}{v}\sum_{j,s\neq j,t\neq j}\Big(\sum_{i,k,\ell(dist)\notin\{j\}} \eta_i\eta_k\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}\Big) W_{js}W_{jt} \\ &\equiv \frac{1}{v}\sum_{j,s\neq j,t\neq j} \beta_j W_{js}W_{jt}, \end{aligned}$$

where

$$|\beta_j| \leq \sum_{i,k,\ell (dist) \notin \{j\}} |\eta_i \eta_k \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}| \leq \sum_{i,k,\ell} (C\theta_i \theta_k)(C\alpha^2 \theta_i \theta_k \theta_\ell^2) \leq C\alpha^2 \|\theta\|^6. \tag{2.E.162}$$

In Z_{5b}, the summand has a nonzero mean only if $s = t$. We further decompose Z_{5b} into

$$Z_{5b} = \frac{1}{v} \sum_{j,s (dist)} \beta_j W_{js}^2 + \frac{1}{v} \sum_{\substack{j \\ s,t (dist) \notin \{j\}}} \beta_j W_{js} W_{jt} \equiv \widetilde{Z}_{5b} + Z_{5b}^*.$$

Only $\widetilde{Z}_{5b}$ has a nonzero mean. By (2.E.116) and (2.E.162),

$$\big|\mathbb{E}[Z_{5b}]\big| = \big|\mathbb{E}[\widetilde{Z}_{5b}]\big| \leq \frac{C}{\|\theta\|_1^2} \sum_{j,s} (\alpha^2 \|\theta\|^6) \theta_j \theta_s \leq C\alpha^2 \|\theta\|^6. \tag{2.E.163}$$

To compute the variance, we note that in each of $\widetilde{Z}_{5b}$ and Z_{5b}^*, two summands are uncorrelated unless they are exactly the same random variables (e.g., when $\{j', s'\} = \{s, j\}$ in $\widetilde{Z}_{5b}$, and when $j' = j$ and $\{s', t'\} = \{s, t\}$ in Z_{5b}^*). Mimicking the argument in (2.E.121) or (2.E.126), we can derive that

$$\mathrm{Var}(\widetilde{Z}_{5b}) \leq \frac{C}{v^2} \sum_{j,s (dist)} \beta_j^2 \,\mathrm{Var}(W_{js}^2) \leq \frac{C\alpha^4 \|\theta\|^{12}}{\|\theta\|_1^4} \sum_{j,s} \theta_j \theta_s \leq \frac{C\alpha^4 \|\theta\|^{12}}{\|\theta\|_1^2},$$

$$\mathrm{Var}(Z_{5b}^*) \leq \frac{C}{v^2} \sum_{\substack{j \\ s,t (dist) \notin \{j\}}} \beta_j^2 \,\mathrm{Var}(W_{js} W_{jt}) \leq \frac{C\alpha^4 \|\theta\|^{12}}{\|\theta\|_1^4} \sum_{j,s,t} \theta_j^2 \theta_s \theta_t \leq \frac{C\alpha^4 \|\theta\|^{14}}{\|\theta\|_1^2}.$$

Combining the above gives

$$\mathrm{Var}(Z_{5b}) \leq 2\mathrm{Var}(\widetilde{Z}_{5b}) + 2\mathrm{Var}(Z_{5b}^*) \leq \frac{C\alpha^4 \|\theta\|^{14}}{\|\theta\|_1^2}. \tag{2.E.164}$$

Third, we study Z_{5c}. If we relabel $(i, j, k, \ell) = (j, i, k, \ell)$, then Z_{5c} becomes

$$Z_{5c} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_j - \tilde{\eta}_j) \eta_i^2 (\eta_k - \tilde{\eta}_k) \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell j} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i^2 \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell j}) (\eta_j - \tilde{\eta}_j)(\eta_k - \tilde{\eta}_k),$$

where $|\eta_i^2 \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell j}| \leq C\alpha^2 \theta_i^2 \theta_j \theta_k \theta_\ell^2$. At the same time, we recall that

$$Z_{5a} = 2 \sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i (\eta_j - \tilde{\eta}_j) \eta_j (\eta_k - \tilde{\eta}_k) \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i \eta_j \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}) (\eta_j - \tilde{\eta}_j)(\eta_k - \tilde{\eta}_k),$$

where $|\eta_i \eta_j \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}| \leq C\alpha^2 \theta_i^2 \theta_j \theta_k \theta_\ell^2$. It is easy to see that Z_{5c} has a similar structure as Z_{5c}. As a result, from (2.E.160)-(2.E.161), we immediately have

$$\big|\mathbb{E}[Z_{5c}]\big| \leq \frac{C\alpha^2 \|\theta\|^8}{\|\theta\|_1^2}, \qquad \mathrm{Var}(Z_{5c}) \leq \frac{C\alpha^4 \|\theta\|^8 \|\theta\|_3^6}{\|\theta\|_1^2}. \tag{2.E.165}$$

We now combine the results for Z_{5a}-Z_{5c}. Since $\|\theta\|^2 \leq \theta_{\max} \|\theta\|_1 \ll \|\theta\|_1^2$, $\mathbb{E}[Z_{5a}]$ and $\mathbb{E}[Z_{5c}]$ are of a smaller order than the the right hand side of (2.E.163). Since $\|\theta\|_3^6 \leq \theta_{\max}^2 \|\theta\|^4 \ll \|\theta\|^6$, $\mathrm{Var}(Z_{5a})$ and $\mathrm{Var}(Z_{5c})$ are of a smaller order than the right hand side of (2.E.164). It follows that

$$\big|\mathbb{E}[Z_5]\big| \leq C\alpha^2 \|\theta\|^6 = o(\alpha^4 \|\theta\|^8), \qquad \mathrm{Var}(Z_5) \leq \frac{C\alpha^4 \|\theta\|^{14}}{\|\theta\|_1^2} = o(\alpha^6 \|\theta\|^8 \|\theta\|_3^6).$$

We briefly explain why $\mathrm{Var}(Z_5) = o(\alpha^6 \|\theta\|^8 \|\theta\|_3^6)$: since $\|\theta\|^4 \leq \|\theta\|_1 \|\theta\|_3^3$, we immediately have $\|\theta\|^{14} \leq \|\theta\|^6(\|\theta\|_1\|\theta\|_3^3)^2$; it follows that the bound for $\mathrm{Var}(Z_5)$ is $\leq C\alpha^4\|\theta\|^6\|\theta\|_3^6$; note that $\alpha\|\theta\| \to \infty$, we immediately have $\alpha^4\|\theta\|^6\|\theta\|_3^6 = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$. This proves the claims of Z_5.

Last, we analyze Z_6. Plugging in the definition of δ_{ij}, we have

$$
\begin{aligned}
Z_6 = &\sum_{i,j,k,\ell (dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell (dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell\widetilde{\Omega}_{\ell i} \\
&+ \sum_{i,j,k,\ell (dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)\widetilde{\Omega}_{\ell i} + \sum_{i,j,k,\ell (dist)} (\eta_i - \widetilde{\eta}_i)\eta_j\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell\widetilde{\Omega}_{\ell i} \\
= &\, 2\sum_{i,j,k,\ell (dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}\eta_k(\eta_\ell - \widetilde{\eta}_\ell)\widetilde{\Omega}_{\ell i} + 2\sum_{i,j,k,\ell (dist)} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell\widetilde{\Omega}_{\ell i} \\
\equiv &\, Z_{6a} + Z_{6b}.
\end{aligned}
$$

By relabeling (i,j,k,ℓ) as (i,j,ℓ,k), we can write

$$
Z_{6a} = 2\sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{j\ell}\eta_\ell(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{ki} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i\eta_\ell\widetilde{\Omega}_{j\ell}\widetilde{\Omega}_{ki})(\eta_j - \widetilde{\eta}_j)(\eta_k - \widetilde{\eta}_k),
$$

where $|\eta_i\eta_\ell\widetilde{\Omega}_{j\ell}\widetilde{\Omega}_{ki}| \leq C\alpha^2\theta_i^2\theta_j\theta_k\theta_\ell^2$. Also, we write

$$
Z_{6b} = 2\sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\widetilde{\Omega}_{jk}(\eta_k - \widetilde{\eta}_k)\eta_\ell\widetilde{\Omega}_{\ell i} = 2\sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i\eta_\ell\widetilde{\Omega}_{jk}\widetilde{\Omega}_{\ell i})(\eta_j - \widetilde{\eta}_j)(\eta_k - \widetilde{\eta}_k).
$$

where $|\eta_i\eta_\ell\widetilde{\Omega}_{jk}\widetilde{\Omega}_{\ell i}| \leq C\alpha^2\theta_i^2\theta_j\theta_k\theta_\ell^2$. At the same time, we recall that

$$
Z_{5a} = 2\sum_{\substack{i,j,k,\ell \\ (dist)}} \eta_i(\eta_j - \widetilde{\eta}_j)\eta_j(\eta_k - \widetilde{\eta}_k)\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i} = \sum_{\substack{i,j,k,\ell \\ (dist)}} (\eta_i\eta_j\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i})(\eta_j - \widetilde{\eta}_j)(\eta_k - \widetilde{\eta}_k),
$$

where $|\eta_i\eta_j\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}| \leq C\alpha^2\theta_i^2\theta_j\theta_k\theta_\ell^2$. It is clear that both Z_{6a} and Z_{6b} have a similar structure as Z_{5a}. From (2.E.160)-(2.E.161), we immediately have

$$
\big|\mathbb{E}[Z_6]\big| \leq \frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2} = o(\alpha^4\|\theta\|^8), \qquad \mathrm{Var}(Z_6) \leq \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8).
$$

This proves the claims of Z_6.

Proofs of Lemmas 22 and 23

Recall that $\lambda_1, \lambda_2, \ldots, \lambda_K$ are all the nonzero eigenvalues of Ω, arranged in the descending order in magnitude. Write for short $\alpha = |\lambda_2|/|\lambda_1|$. We shall repeatedly use the following results, which are proved in (2.E.110), (2.E.116), and (2.E.117):

$$
v \asymp \|\theta\|_1^2, \qquad 0 < \eta_i < C\theta_i, \qquad |\widetilde{\Omega}_{ij}| \leq C\alpha\theta_i\theta_j.
$$

Recall that $U_c = 4T_1 + F$, under the null hypothesis; $U_c = 4T_1 + 4T_2 + F$ under the alternative hypothesis. By definition,

$$T_1 = \sum_{i_1,i_2,i_3,i_4 (dist)} \delta_{i_1 i_2}\delta_{i_2 i_3}\delta_{i_3 i_4} W_{i_4 i_1},$$
$$T_2 = \sum_{i_1,i_2,i_3,i_4 (dist)} \delta_{i_1 i_2}\delta_{i_2 i_3}\delta_{i_3 i_4} \widetilde{\Omega}_{i_4 i_1},$$
$$F = \sum_{i_1,i_2,i_3,i_4 (dist)} \delta_{i_1 i_2}\delta_{i_2 i_3}\delta_{i_3 i_4} \delta_{i_4 i_1},$$

where $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$, for $1 \le i, j \le n$, $i \ne j$. By symmetry and elementary algebra, we further write

$$T_1 = 2T_{1a} + 2T_{1b} + 2T_{1c} + 2T_{1d}, \tag{2.E.166}$$

where

$$T_{1a} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot W_{i_4 i_1},$$
$$T_{1b} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}^2\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot W_{i_4 i_1},$$
$$T_{1c} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot W_{i_4 i_1},$$
$$T_{1d} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_3}^2\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot W_{i_4 i_1}.$$

Similarly, we write

$$T_2 = 2T_{2a} + 2T_{2b} + 2T_{2c} + 2T_{2d}, \tag{2.E.167}$$

where

$$T_{2a} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot \widetilde{\Omega}_{i_4 i_1},$$
$$T_{2b} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}^2\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot \widetilde{\Omega}_{i_4 i_1},$$
$$T_{2c} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot \widetilde{\Omega}_{i_4 i_1},$$
$$T_{2d} = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_3}^2\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot \widetilde{\Omega}_{i_4 i_1}.$$

Also, similarly, we have

$$F = 2F_a + 12F_b + 2F_c, \tag{2.E.168}$$

where

$$F_a = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big],$$
$$F_b = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}^2\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})^2(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big],$$
$$F_c = \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}^2\eta_{i_4}^2\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})^2\big].$$

To show the lemmas, it is sufficient to show the following 11 items (a)-(k), corresponding to $T_{1a}, T_{1b}, T_{1c}, T_{1d}, T_{2a}, T_{2b}, T_{2c}, T_{2d}, F_a, F_b, F_c$, respectively. Item (a) claims that both under the null and the alternative,

$$|\mathbb{E}[T_{1a}]| \leq C\|\theta\|^6/\|\theta\|_1^2, \qquad \mathrm{Var}(T_{1a}) \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2. \tag{2.E.169}$$

Item (b) claims that both under the null and the alternative,

$$|\mathbb{E}[T_{1b}]| \leq C\|\theta\|^6/\|\theta\|_1^2, , \qquad \mathrm{Var}(T_{1b}) \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.170}$$

Item (c) claims that both under the null and the alternative,

$$\mathbb{E}[T_{1c}] = 0, \qquad \mathrm{Var}(T_{1c}) \leq C\|\theta\|_3^9/\|\theta\|_1, \tag{2.E.171}$$

Item (d) claims that

$$\begin{aligned} &\mathbb{E}[T_{1d}] \asymp -\|\theta\|^4 \text{ under the null,} \\ &|\mathbb{E}[T_{1d}]| \leq C\|\theta\|^4 \text{ under the alternative,} \end{aligned} \tag{2.E.172}$$

and that both under the null and the alternative,

$$\mathrm{Var}(T_{1d}) \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.173}$$

Next, for item (e)-(h), we recall that under the null, $T_2 = 0$, and correspondingly $T_{2a} = T_{2b} = T_{2c} = T_{2d} = 0$, so we only need to consider the alternative. Recall that $\alpha = |\lambda_2/\lambda_1|$. Item (e) claims that under the alternative,

$$\mathbb{E}[T_{2a}] = 0, \qquad \mathrm{Var}(T_{2a}) \leq C\alpha^2 \cdot \|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^3. \tag{2.E.174}$$

Item (f) claims that under the alternative,

$$\mathbb{E}[T_{2b}] = 0, \qquad \mathrm{Var}(T_{2b}) \leq C\alpha^2 \cdot \|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.175}$$

Item (g) claims that under the alternative,

$$|\mathbb{E}[T_{2c}]| \leq C\alpha\|\theta\|^6/\|\theta\|_1^3, \qquad \mathrm{Var}(T_{2c}) \leq C\alpha^2 \cdot \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.176}$$

Item (h) claims that both under the null and the alternative,

$$|\mathbb{E}[T_{2d}]| \leq C\alpha\|\theta\|^6/\|\theta\|_1^3, \qquad \mathrm{Var}(T_{2d}) \leq C\alpha^2 \cdot \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.177}$$

Finally, for items (i)-(k). Item (i) claims that both under the null and the alternative,

$$|\mathbb{E}[F_a]| \leq C\|\theta\|^8/\|\theta\|_1^4, \qquad \mathrm{Var}(F_a) \leq C\|\theta\|_3^{12}/\|\theta\|_1^4. \tag{2.E.178}$$

Item (j) claims that both under the null and the alternative,

$$|\mathbb{E}[F_b]| \leq C\|\theta\|^6/\|\theta\|_1^2, \qquad \mathrm{Var}(F_b) \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2. \tag{2.E.179}$$

Item (k) claims that

$$\begin{aligned} &\mathbb{E}[F_c] \asymp \|\theta\|^4 \text{ under the null,} \\ &|\mathbb{E}[F_c]| \leq C\|\theta\|^4 \text{ under the alternative,} \end{aligned} \tag{2.E.180}$$

and that under both under the null and the alternative,

$$\mathrm{Var}(F_3) \leq C\|\theta\|^{10}/\|\theta\|_1^2. \tag{2.E.181}$$

We now show Lemmas 18 and 19 follow once (a)-(k) are proved. In detail, first, we note that $\|\theta\|^6/\|\theta\|_1^2 = o(\|\theta\|^4)$. Inserting (2.E.172) and the first equation in each of (2.E.169)-(2.E.171) into (2.E.166) gives that

$$\mathbb{E}[T_1] \asymp -2\|\theta\|^4 \text{ under the null,} \qquad |\mathbb{E}[T_1]| \leq C\|\theta\|^4 \text{ under the alternative,}$$

and inserting (2.E.173) and the second equation in each of (2.E.169)-(2.E.171) into (2.E.166) gives that both under the null and the alternative,

$$\mathrm{Var}(T_1) \leq C[\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2 + \|\theta\|^6\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|_3^9/\|\theta\|_1 + \|\theta\|^6\|\theta\|_3^3/\|\theta\|_1],$$

where since $\|\theta\|_3^3/\|\theta\|^2 = o(1)$ and $\|\theta\|^2/\|\theta\|_1 = o(1)$, the right hand side

$$\leq C[\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1^2 + \|\theta\|^6\|\theta\|_3^3/\|\theta\|_1] \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1.$$

Second, inserting the first equation in each of (2.E.174)-(2.E.177) into (2.E.167) gives that under the alternative (recall that $T_2 = 0$ under the null),

$$|\mathbb{E}[T_2]| \leq C\alpha\|\theta\|^6/\|\theta\|_1^3,$$

and inserting the second equation in each of (2.E.174)-(2.E.177) into (2.E.167) gives

$$\mathrm{Var}(T_2) \leq C\alpha^2[\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5] \leq C\alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1,$$

where we have used $\|\theta\|^2 = o(\|\theta\|_1^2)$. Third, note that $\|\theta\|^8/\|\theta\|_1^4 = o(\|\theta\|^4)$ and $\|\theta\|^6/\|\theta\|_1^2 = o(\|\theta\|^4)$. Inserting (2.E.180) and the first equation in each of (2.E.178)-(2.E.179) into (2.E.168) gives

$$\mathbb{E}[F] \sim 2\|\theta\|^4 \text{ under the null}, \qquad |\mathbb{E}[F]| \leq C\|\theta\|^4 \text{ under the alternative},$$

and inserting (2.E.181) and the second equation in each of (2.E.178)-(2.E.179) into (2.E.168) gives that both under the null and the alternative,

$$\mathrm{Var}(F) \leq C[\|\theta\|_3^{12}/\|\theta\|_1^4 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2 + \|\theta\|^{10}/\|\theta\|_1^2] \leq C\|\theta\|^{10}/\|\theta\|_1^2,$$

where we have used $\|\theta\|_3^3 \ll \theta\|^2 \ll \|\theta\|_1$ and $\|\theta\|_3^3/\|\theta\|^2 = o(1)$.

We now combine the above results for T_1, T_2 and F. First, since that $U_c = 4T_1 + F$ under the null, it follows that under the null,

$$\mathbb{E}[U_c] \sim -6\|\theta\|^4,$$

and

$$\mathrm{Var}(U_c) \leq C[\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^{10}/\|\theta\|_1^2] \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$ (a direct use of Cauchy-Schwartz inequality). Second, since $U_c = 4T_1 + 4T_2 + F$ under the alternative, it follows that under the alternative,

$$|\mathbb{E}[U_c]| \leq C\|\theta\|^4,$$

and

$$\mathrm{Var}(U_c) \leq C[\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1 + \alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^{10}/\|\theta\|_1^2] \leq C\|\theta\|^6\|\theta\|_3^3(\alpha^2\|\theta\|^2+1)/\|\theta\|_1,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$ and basic algebra. Combining the above gives all the claims in Lemmas 18 and 19.

It remains to show the 11 items (a)-(k). We consider them separately.

Consider Item (a). The goal is to show (2.E.169). Recall that

$$T_{1a} = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot W_{i_4 i_1},$$

and that

$$\widetilde{\eta} - \eta = v^{-1/2}W1_n. \tag{2.E.182}$$

Plugging (2.E.182) into T_{11} gives

$$\begin{aligned} T_{1a} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4} \Big(\sum_{j_1, j_1\neq i_1} W_{i_1j_1}\Big)\Big(\sum_{j_2,j_2\neq i_2} W_{i_2j_2}\Big)\Big(\sum_{j_3,j_3\neq i_3} W_{i_3j_3}\Big) W_{i_4i_1} \\ &= -\frac{1}{v^{3/2}} \sum_{\substack{i_1,i_2,i_3,i_4(dist) \\ j_1\neq i_1, j_2\neq i_2, j_3\neq i_3}} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1i_4}. \end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1i_4} = \begin{cases} W_{i_1i_4}^2 W_{i_2i_3}^2, & \text{if } j_1 = i_4, (j_2, j_3) = (i_3, i_2), \\ W_{i_1i_4}^2 W_{i_2j_2}W_{i_3j_3}, & \text{if } j_1 = i_4, (j_2, j_3) \neq (i_3, i_2), \\ W_{i_2i_3}^2 W_{i_1j_1}W_{i_1i_4}, & \text{if } j_1 \neq i_4, (j_2, j_3) = (i_3, i_2), \\ W_{i_1i_2}^2 W_{i_3j_3}W_{i_1i_4}, & \text{if } (j_1, j_2) = (i_2, i_1), \\ W_{i_1i_3}^2 W_{i_2j_2}W_{i_1i_4}, & \text{if } (j_1, j_3) = (i_3, i_1), \\ W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1i_4}, & \text{otherwise.} \end{cases} \tag{2.E.183}$$

This allows us to further split T_{11} into 6 different terms:

$$T_{1a} = X_a + X_{b1} + X_{b2} + X_{b3} + X_{b4} + X_c, \tag{2.E.184}$$

where

$$\begin{aligned} X_a &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_4}^2 W_{i_2i_3}^2, \\ X_{b1} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2,j_3 \\ (j_2,j_3)\neq\{i_3,i_2\}}} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_4}^2 W_{i_2j_2}W_{i_3j_3}, \\ X_{b2} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1(j_1\neq i_4)} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_1j_1}W_{i_1i_4}, \\ X_{b3} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_3(j_3\neq i_3)} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_2}^2 W_{i_3j_3}W_{i_1i_4}, \\ X_{b4} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2(j_2\neq i_2)} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_3}^2 W_{i_2j_2}W_{i_1i_4}, \\ X_c &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,j_2,j_3 \\ j_1\notin\{i_1,i_4\},(j_2,j_3)\neq(i_3,i_2) \\ (j_1,j_2)\neq(i_2,i_1),(j_1,j_3)\neq(i_3,i_1)}} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1i_4}. \end{aligned}$$

We now show (2.E.169). Consider the first claim of (2.E.169). It is seen that out of the 6 terms on the right hand side of (2.E.184), the mean of all terms are 0, except for the first term. Note that for any $1 \leq i, j \leq n$, $i \neq j$, $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1 - \Omega_{ij})$, where Ω_{ij} are upper bounded by $o(1)$ uniformly for all such i, j. It follows

$$\begin{aligned} \mathbb{E}[X_a] &= -v^{-3/2} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4} \mathbb{E}[W_{i_1i_4}^2]\mathbb{E}[W_{i_2i_3}^2] \\ &= -(1 + o(1)) \cdot v^{-3/2} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4} \Omega_{i_1i_4}\Omega_{i_2i_3}. \end{aligned}$$

Since for any $1 \leq i, j \leq n$, $i \neq j$, $0 < \eta_i \leq C\theta_i$, $\Omega_{ij} \leq C\theta_i\theta_j$ and $v \asymp \|\theta\|_1^2$,

$$|\mathbb{E}[X_a]| \leq C(\|\theta\|_1)^{-3} \sum_{i_1,i_2,i_3,i_4 (dist)} \theta_{i_1}\theta_{i_2}^2\theta_{i_3}^2\theta_{i_4}^2 \leq C\|\theta\|^6/\|\theta\|_1^2.$$

Inserting these into (2.E.184) gives

$$|\mathbb{E}[T_{1a}]| \leq C\|\theta\|^6/\|\theta\|_1^2, \tag{2.E.185}$$

and the first claim of (2.E.169) follows.

Consider the second claim of (2.E.169) next. By (2.E.184) and Cauchy-Schwartz inequality,

$$\begin{aligned} \mathrm{Var}(T_{1a}) &\leq C\mathrm{Var}(X_a) + \mathrm{Var}(X_{b1}) + \mathrm{Var}(X_{b2}) + \mathrm{Var}(X_{b3}) + \mathrm{Var}(X_{b4}) + \mathrm{Var}(X_c)) \\ &\leq C(\mathrm{Var}(X_a) + \mathbb{E}[X_{b1}^2] + \mathbb{E}[X_{b2}^2] + \mathbb{E}[X_{b3}^2] + \mathbb{E}[X_{b4}^2] + \mathbb{E}[X_c^2]). \end{aligned} \tag{2.E.186}$$

We now consider $\mathrm{Var}(X_a)$, $\mathbb{E}[X_{b1}^2] + \mathbb{E}[X_{b2}^2] + \mathbb{E}[X_{b3}^2] + \mathbb{E}[X_{b4}^2]$, and $\mathbb{E}[X_c^2]$, separately.

Consider $\mathrm{Var}(X_a)$. Write $\mathrm{Var}(X_a)$ as

$$v^{-3} \sum_{\substack{i_1,\cdots,i_4 (dist) \\ i_1',\cdots,i_4' (dist)}} \eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'}$$
$$\mathbb{E}\big[(W_{i_1i_4}^2 W_{i_2i_3}^2 - \mathbb{E}[W_{i_1i_4}^2 W_{i_2i_3}^2])(W_{i_1'i_4'}^2 W_{i_2'i_3'}^2 - \mathbb{E}[W_{i_1'i_4'}^2 W_{i_2'i_3'}^2])\big]. \tag{2.E.187}$$

In the sum, a term is nonzero only when one of the following cases happens.

- (A). $\{W_{i_1i_4}, W_{i_2i_3}, W_{i_1'i_4'}, W_{i_2'i_3'}\}$ has 2 distinct random variables.
- (B). $\{W_{i_1i_4}, W_{i_2,i_3}, W_{i_1'i_4'}, W_{i_2'i_3'}\}$ has 3 distinct random variables. This has 4 sub-cases: (B1) $W_{i_1i_4} = W_{i_1'i_4'}$, (B2) $W_{i_1i_4} = W_{i_2'i_3'}$, (B3) $W_{i_2i_3} = W_{i_1'i_4'}$, and (B4) $W_{i_2i_3} = W_{i_2'i_3'}$.

For Case (A), the two sets $\{i_1, i_2, i_3, i_4\}$ and $\{i_1', i_2', i_3', i_4'\}$ are identical. By basic statistics and independence between $W_{i_1i_4}$ and $W_{i_2i_3}$,

$$\begin{aligned} &\mathbb{E}[(W_{i_1i_4}^2 W_{i_2i_3}^2 - \mathbb{E}[W_{i_1i_4}^2 W_{i_2i_3}^2])(W_{i_1'i_4'}^2 W_{i_2'i_3'}^2 - \mathbb{E}[W_{i_1'i_4'}^2 W_{i_2'i_3'}^2])] \\ =&\mathbb{E}[(W_{i_1i_4}^2 W_{i_2i_3}^2 - \mathbb{E}[W_{i_1i_4}^2 W_{i_2i_3}^2])^2] \\ =&\mathbb{E}[W_{i_1i_4}^4]\mathbb{E}[W_{i_2i_3}^4] - (\mathbb{E}[W_{i_1i_4}^2])^2(\mathbb{E}[W_{i_2i_3}^2])^2 \\ \leq&\mathbb{E}[W_{i_1i_4}^4]\mathbb{E}[W_{i_2i_3}^4], \end{aligned} \tag{2.E.188}$$

where by basic statistics and that $\Omega_{ij} \leq C\theta_i\theta_j$ for all $1 \leq i, j \leq n$, $i \leq j$, the right hand side

$$\leq C\Omega_{i_1i_4}\Omega_{i_2i_3} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}.$$

Combining these with (2.E.187) and noting that $v \sim \|\theta\|_1^2$ and that $0 < \eta_i \leq C\theta_i$ for all $1 \leq i \leq n$, the contribution of this case to $\mathrm{Var}(X_a)$ is no more than

$$C(\|\theta\|_1)^{-6} \sum_{i_1,\cdots,i_4 (dist)} \sum_a \theta_{i_1}^{a_1+1}\theta_{i_2}^{a_2+2}\theta_{i_3}^{a_3+2}\theta_{i_4}^{a_4+2}, \tag{2.E.189}$$

where $a = (a_1, a_2, a_3, a_4)$ and each a_i is either 0 and 1, satisfying $a_1 + a_2 + a_3 + a_4 = 3$. Note that the right hand side of (2.E.189) is no greater than

$$C(\|\theta\|_1)^{-6} \max\{\|\theta\|_1\|\theta\|_3^9, \|\theta\|^4\|\theta\|_3^6\} \leq C\|\theta\|_3^9/\|\theta\|_1^5,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$.

Next, consider (B1). By independence between $W_{i_1i_4}$, $W_{i_2i_3}$, and $W_{i'_2i'_3}$ and basic algebra,

$$\begin{aligned}
&\mathbb{E}[(W^2_{i_1i_4}W^2_{i_2i_3} - \mathbb{E}[W^2_{i_1i_4}W^2_{i_2i_3}])(W^2_{i'_1i'_4}W^2_{i'_2i'_3} - \mathbb{E}[W^2_{i'_1i'_4}W^2_{i'_2i'_3}])]\\
=&\mathbb{E}[(W^2_{i_1i_4}W^2_{i_2i_3} - \mathbb{E}[W^2_{i_1i_4}W^2_{i_2i_3}])(W^2_{i_1i_4}W^2_{i'_2i'_3} - \mathbb{E}[W^2_{i_1i_4}W^2_{i'_2i'_3}])]\\
=&\mathbb{E}[W^4_{i_1i_4}]\mathbb{E}[W^2_{i_2i_3}]\mathbb{E}[W^2_{i'_2i'_3}] - (\mathbb{E}[W^2_{i_1i_4}])^2\mathbb{E}[W^2_{i_2i_3}]\mathbb{E}[W^2_{i'_2i'_3}]\\
=&\mathrm{Var}(W^2_{i_1i_4})\mathbb{E}[W^2_{i_2i_3}]\mathbb{E}[W^2_{i'_2i'_3}], \qquad (2.E.190)
\end{aligned}$$

where by similar arguments, the last term

$$\leq C\Omega_{i_1i_4}\Omega_{i_2i_3}\Omega_{i'_2i'_3} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}\theta_{i'_2}\theta_{i'_3}.$$

Combining this with (2.E.187) and using similar arguments, the contribution of this case to $\mathrm{Var}(X_a)$

$$\leq C(\|\theta\|_1)^{-6} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i'_2,i'_3(dist)}} C\theta_{i_1}^{b_1+1}\theta_{i_2}^2\theta_{i_3}^2\theta_{i_4}^{b_2+2}\theta_{i'_2}^2\theta_{i'_3}^2, \qquad (2.E.191)$$

where similarly b_1, b_2 are either 0 or 1 and $b_1 + b_2 = 1$. By similar argument, the right hand side

$$\leq C\|\theta\|_1\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^6 = C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5.$$

The discussion for (B2), (B3), and (B4) are similar so is omitted, and their contribution to $\mathrm{Var}(X_a)$ are respectively

$$\leq C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5, \qquad (2.E.192)$$

$$\leq C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5, \qquad (2.E.193)$$

and

$$\leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4. \qquad (2.E.194)$$

Finally, inserting (2.E.189), (2.E.191), (2.E.192), (2.E.193), and (2.E.194) into (2.E.187) gives

$$\mathrm{Var}(X_a) \leq C[\|\theta\|_3^9/\|\theta\|_1^5 + \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4] \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4, \quad (2.E.195)$$

where we have used $\|\theta\|_3^3 \ll \|\theta\|^2$ and $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$.

Consider $\mathbb{E}[X_{b1}^2] + \mathbb{E}[X_{b21}^2] + \mathbb{E}[X_{b3}^2] + \mathbb{E}[X_{b4}^2]$. We claim that both under the null and the alternative,

$$\mathbb{E}[X_{b1}^2] \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2, \qquad (2.E.196)$$
$$\mathbb{E}[X_{b2}^2] \leq C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^3, \qquad (2.E.197)$$
$$\mathbb{E}[X_{b3}^2] \leq C\|\theta\|^6\|\theta\|_3^6/\|\theta\|_1^4, \qquad (2.E.198)$$
$$\mathbb{E}[X_{b4}^2] \leq C\|\theta\|^6\|\theta\|_3^6/\|\theta\|_1^4, \qquad (2.E.199)$$

where the last two terms are seen to be negligible compared to the other two. Therefore,

$$\mathbb{E}[X_{b1}^2] + \mathbb{E}[X_{b2}^2] + \mathbb{E}[X_{b3}^2] + \mathbb{E}[X_{b4}^2] \leq C[\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2 + \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^3], \qquad (2.E.200)$$

where since $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$ (Cauchy-Schwartz inequality) the right hand side

$$\leq C[\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2.$$

We now prove (2.E.196)-(2.E.199). Since the study for $\mathbb{E}[X_{b1}^2], \mathbb{E}[X_{b2}^2], \mathbb{E}[X_{b3}^2]$ and $\mathbb{E}[X_{b4}^2]$ are similar, we only present the proof for $\mathbb{E}[X_{b1}^2]$. Write $\mathbb{E}[X_{b1}^2]$ as

$$v^{-3} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i_1',i_2',i_3',i_4'(dist)}} \sum_{\substack{j_2,j_3\\ (j_2,j_3)\neq(i_3,i_2)}} \sum_{\substack{j_2',j_3'\\ (j_2',j_3')\neq(i_3',i_2')}} \eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'} W_{i_1i_4}^2 W_{i_2j_2} W_{i_3j_3} W_{i_1'i_4'}^2 W_{i_2'j_2'} W_{i_3'j_3'}.$$

Consider the term

$$W_{i_1i_4}^2 W_{i_2j_2} W_{i_3j_3} W_{i_1'i_4'}^2 W_{i_2'j_2'} W_{i_3'j_3'}.$$

In order for the mean to be nonzero, we have two cases

- Case A. The two sets of random variables $\{W_{i_1i_4}, W_{i_2j_2}, W_{i_3j_3}\}$ and $\{W_{i_1'i_4'}, W_{i_2'j_2'}, W_{i_3'j_3'}\}$ are identical.
- Case B. The two sets $\{W_{i_2j_2}, W_{i_3j_3}\}$ and $\{W_{i_2'j_2'}, W_{i_3'j_3'}\}$ are identical.

Consider Case A. In this case, $\{i_2', i_3', i_4'\}$ are three distinct indices in $\{i_1, i_2, i_3, i_4, j_2, j_3\}$, and for some integers satisfying $0 \leq a_1, a_2, \ldots, a_6 \leq 1$, $a_1 + a_2 + \ldots + a_6 = 3$,

$$\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'} = \eta_{i_1}^{a_1}\eta_{i_2}^{1+a_2}\eta_{i_3}^{1+a_3}\eta_{i_4}^{1+a_4}\eta_{j_2}^{a_5}\eta_{j_3}^{a_6}$$

and for some integers satisfying $0 \leq b_1, b_2, b_3 \leq 1$, and $b_1 + b_2 + b_3 = 1$,

$$W_{i_1i_4}^2 W_{i_2j_2} W_{i_3j_3} W_{i_1'i_4'}^2 W_{i_2'j_2'} W_{i_3'j_3'} = W_{i_1i_4}^{b_1+3} W_{i_2j_2}^{b_2+2} W_{i_3j_3}^{b_3+2}.$$

Similarly, by $v \sim \|\theta\|_1^2$, $0 < \eta_i \leq C\theta_i$, and uniformly for all b_1, b_2, b_3 above,

$$0 < \mathbb{E}[W_{i_1i_4}^{b_1+3} W_{i_2j_2}^{b_2+2} W_{i_3j_3}^{b_3+2}] \leq C\Omega_{i_1i_4}\Omega_{i_2j_2}\Omega_{i_3j_3} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}\theta_{j_2}\theta_{j_3}.$$

Therefore under both the null and the alternative, the contribution of Case A to the variance is

$$\leq C(\|\theta\|_1)^{-6} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2,j_3\\ j_2\neq i_2, j_3\neq i_3, (j_2,j_3)\neq(i_3,i_2)}} [\sum_a \theta_{i_1}^{a_1+1}\theta_{i_2}^{a_2+2}\theta_{i_3}^{a_3+2}\theta_{i_4}^{a_4+2}\theta_{j_2}^{a_5+1}\theta_{j_3}^{a_6+1}], \tag{2.E.201}$$

where $a = (a_1, a_2, \ldots, a_6)$ and a_i satisfies the above constraints. Note that the right hand size

$$\leq C(\|\theta\|_1)^{-6} \cdot \max\{\|\theta\|_1^3\|\theta\|_3^9, \|\theta\|_1^2\|\theta\|^4\|\theta\|_3^6, \|\theta\|_1\|\theta\|^8\|\theta\|_3^3, \|\theta\|^{12}\} \leq C\|\theta\|_3^9/\|\theta\|_1^3.$$

Here in the last inequality we have used $\|\theta\|^2 \leq \sqrt{\|\theta\|_1\|\theta\|_3^3}$.

Consider Case B. In this case, $\{i_2, i_3, j_2, j_3\} = \{i_2', i_3', j_2', j_3'\}$, and for some integers $0 \leq c_1, c_2, c_3, c_4 \leq 1$, $c_1 + c_2 + c_3 + c_4 = 2$,

$$\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'} = \eta_{i_2}^{c_1+1}\eta_{i_3}^{c_2+1}\eta_{i_4}\eta_{j_2}^{c_3}\eta_{j_3}^{c_4}\eta_{i_4'},$$

and

$$W_{i_1i_4}^2 W_{i_2j_2} W_{i_3j_3} W_{i_1'i_4'}^2 W_{i_2'j_2'} W_{i_3'j_3'} = W_{i_1i_4}^2 W_{i_2j_2}^2 W_{i_3j_3}^2 W_{i_1'i_4'}^2,$$

where the four W terms on the right are independent of each other. Similarly, by $v \sim \|\theta\|_1^2$, $0 < \eta_i \leq C\theta_i$,

$$0 < \mathbb{E}[W_{i_1i_4}^2 W_{i_2j_2}^2 W_{i_3j_3}^2 W_{i_1'i_4'}^2] \leq C\Omega_{i_1i_4}\Omega_{i_2j_2}\Omega_{i_3j_3}\Omega_{i_1'i_4'} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}\theta_{j_2}\theta_{j_3}\theta_{i_1'}\theta_{i_4'},$$

we have that under both the null and the alternative, the contribution of Case B to the variance

$$\leq C(\|\theta\|_1)^{-6} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i_1',i_4'(dist)}} \sum_{\substack{j_2,j_3\\ (j_2,j_3)\neq(i_3,i_2)}} \theta_{i_1}\theta_{i_2}^{c_1+2}\theta_{i_3}^{c_2+2}\theta_{i_4}^2\theta_{j_2}^{c_3+1}\theta_{j_3}^{c_4+1}\theta_{i_1'}\theta_{i_4'}^2,$$

where the right hand size

$$\leq C(\|\theta\|_1)^{-6}\cdot\|\theta\|_1^2\|\theta\|^4\cdot\max\{\|\theta\|_1^2\|\theta\|_3^6, \|\theta\|_1\|\theta\|^4\|\theta\|_3^3, \|\theta\|^8\} \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2. \quad (2.E.202)$$

Here we have again used $\|\theta\|^2 \leq \sqrt{\|\theta\|_1\|\theta\|_3^3}$.

Finally, combining (2.E.201) and (2.E.202) gives

$$\mathbb{E}[X_{b1}^2] \leq C(\|\theta\|_3^9/\|\theta\|_1^3 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2) \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2,$$

which proves (2.E.196).

Consider $\mathbb{E}[X_c^2]$. Consider the terms in the sum,

$$\eta_{i_2}\eta_{i_3}\eta_{i_4}W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1i_4}, \qquad \text{and} \qquad \eta_{i_2'}\eta_{i_3'}\eta_{i_4'}W_{i_1'j_1'}W_{i_2'j_2'}W_{i_3'j_3'}W_{i_1'i_4'}.$$

Each term has a mean 0, and two terms are uncorrelated with each other if only if the two sets of random variables $\{W_{i_1j_1}, W_{i_2j_2}, W_{i_3j_3}, W_{i_1i_4}\}$ and $\{W_{i_1'j_1'}, W_{i_2'j_2'}, W_{i_3'j_3'}, W_{i_1'i_4'}\}$ are identical (however, it is possible that $W_{i_1j_1}$ does not equal to $W_{i_1j_1'}$ but equals to $W_{i_2'j_2'}$, say). Additionally, the indices $i_2', i_3', i_4' \in \{i_1, i_2, i_3, i_4, j_1, j_2, j_3\}$, and it follows

$$\mathbb{E}[X_c^2] \leq Cv^{-3} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,j_2,j_3\\ j_1\notin\{i_1,i_4\},(j_1,j_3)\neq(i_3,i_1)\\ (j_2,j_3)\neq(i_3,i_2),(j_2,j_1)\neq(i_2,i_1)}} [\sum_a \eta_{i_1}^{a_1}\eta_{i_2}^{a_2+1}\eta_{i_3}^{a_3+1}\eta_{i_4}^{a_4+1}\eta_{j_1}^{a_5}\eta_{j_2}^{a_6}\eta_{j_3}^{a_7}]\cdot\mathbb{E}[W_{i_1j_1}^2W_{i_2j_2}^2W_{i_3j_3}^2W_{i_1j_1}^2],$$

where $a = (a_1, a_2, \cdots, a_7)$ and the power $0 \leq a_1, a_2, \cdots, a_7 \leq 1$, and $a_1 + a_2 + \cdots + a_7 = 3$. Note that $W_{i_1j_1}, W_{i_2j_2}, W_{i_3j_3}$ and $W_{i_1i_4}$ are independent and $\mathbb{E}(W_{ij}^2) \leq \Omega_{ij} \leq C\theta_i\theta_j$, $1 \leq i, j \leq n$, $i \neq j$,

$$\mathbb{E}[W_{i_1j_1}^2W_{i_2j_2}^2W_{i_3j_3}^2W_{i_1i_4}^2] \leq \Omega_{i_1j_1}\Omega_{i_2j_2}\Omega_{i_3j_3}\Omega_{i_1i_4} \leq C\theta_{i_1}^2\theta_{i_2}\theta_{i_3}\theta_{i_4}\theta_{j_1}\theta_{j_2}\theta_{j_3}.$$

Also, recall that both under the null and the alternative, $v \asymp \|\theta\|_1^2$ and $0 < \eta_i \leq C\theta_i$, $1 \leq i \leq n$. Combining these gives

$$\mathbb{E}[X_c^2] \leq C(\|\theta\|_1)^{-6} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,j_2,j_3\\ j_1\notin\{i_1,i_4\},(j_1,j_3)\neq(i_3,i_1)\\ (j_2,j_3)\neq(i_3,i_2),(j_2,j_1)\neq(i_2,i_1)}} [\sum_a \eta_{i_1}^{a_1+2}\eta_{i_2}^{a_2+2}\eta_{i_3}^{a_3+2}\eta_{i_4}^{a_4+2}\eta_{j_1}^{a_5+1}\eta_{j_2}^{a_6+1}\eta_{j_3}^{a_7+1}],$$

where the last term

$$\leq C\sum_a \|\theta\|_{a_1+2}^{a_1+2}\cdot\|\theta\|_{a_2+2}^{a_2+2}\cdot\|\theta\|_{a_3+2}^{a_3+2}\cdot\|\theta\|_{a_4+2}^{a_4+2}\|\theta\|_{a_5+1}^{a_5+1}\|\theta\|_{a_6+1}^{a_6+1}\|\theta\|_{a_7+1}^{a_7+1}/\|\theta\|_1^6.$$

Since $a_1, a_2, \cdots, a_7$ have to take values from $\{0, 1\}$ and their sum is 3, the above term

$$\leq C\|\theta\|^2\|\theta\|_3^9/\|\theta\|_1^3 = o(\|\theta\|_3^3),$$

where we have used $\|\theta\|_3^3 \ll \|\theta\|_2^2 \ll \|\theta\|_1$. Combining these gives

$$\mathbb{E}[X_c^2] \leq C\|\theta\|^2\|\theta\|_3^9/\|\theta\|_1^3. \quad (2.E.203)$$

Finally, inserting (2.E.195), (2.E.200), and (2.E.203) into (2.E.184) gives that both under the null and the alternative,

$$\mathrm{Var}(T_{11}) \leq C[\|\theta\|^8/\|\theta\|_1^4 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2 + \|\theta\|^2\|\theta\|_3^9/\|\theta\|_1^3] \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$ and $\|\theta\|_3^3/\|\theta\|_1 = o(1)$. This gives (2.E.169) and completes the proof for Item (a).

Consider Item (b). The goal is to show (2.E.170). Recall that

$$T_{1b} = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot W_{i_4 i_1},$$

and that

$$\widetilde{\eta} - \eta = v^{-1/2}W1_n.$$

Plugging this into T_{1b} gives

$$\begin{aligned} T_{1b} &= -v^{-3/2} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2\Big(\sum_{j_1\neq i_1} W_{i_1 j_1}\Big)\Big(\sum_{j_2\neq i_2} W_{i_2 j_2}\Big)\Big(\sum_{j_4\neq i_4} W_{i_4 j_4}\Big)W_{i_1 i_4} \\ &= -\frac{1}{v^{3/2}} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ j_1\neq i_1, j_2\neq i_2, j_4\neq i_4}} \eta_{i_2}\eta_{i_3}^2 W_{i_1 j_1}W_{i_2 j_2}W_{i_4 j_4}W_{i_1 i_4}. \end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_1 j_1}W_{i_2 j_2}W_{i_4 j_4}W_{i_1 i_4} = \begin{cases} W_{i_1 i_4}^3 W_{i_2 j_2}, & \text{if } j_1 = i_4, j_4 = i_1, \\ W_{i_1 i_2}^2 W_{i_1 i_4}^2, & \text{if } j_1 = i_2, j_2 = i_1, j_4 = i_1, \\ W_{i_1 i_4}^2 W_{i_2 i_4}^2, & \text{if } j_1 = i_4, j_2 = i_4, j_4 = i_2, \\ W_{i_1 i_2}^2 W_{i_4 j_4}W_{i_1 i_4}, & \text{if } j_1 = i_2, j_2 = i_1, \\ W_{i_1 i_4}^2 W_{i_1 j_1}W_{i_2 j_2}, & \text{if } j_4 = i_1, \\ W_{i_1 i_4}^2 W_{i_2 j_2}W_{i_4 j_4}, & \text{if } j_1 = i_4, \{i_2, j_2\} \neq \{i_4, j_4\}, \\ W_{i_2 i_4}^2 W_{i_1 j_1}W_{i_1 i_4}, & \text{if } j_2 = i_4, j_4 = i_2, \\ W_{i_1 j_1}W_{i_2 j_2}W_{i_4 j_4}W_{i_1 i_4}, & \text{otherwise.} \end{cases} \tag{2.E.204}$$

This allows us to further split T_{1b} into 8 different terms:

$$T_{1b} = Y_{a1} + Y_{a2} + Y_{a3} + Y_{b1} + Y_{b2} + Y_{b3} + Y_{b4} + Y_c, \tag{2.E.205}$$

where

$$Y_{a1} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2(j_2\neq i_2)} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_4}^3 W_{i_2j_2},$$

$$Y_{a2} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_2}^2 W_{i_1i_4}^2,$$

$$Y_{a3} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_4}^2 W_{i_2i_4}^2,$$

$$Y_{b1} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_4(j_4\neq i_4)} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_2}^2 W_{i_4j_4} W_{i_1i_4},$$

$$Y_{b2} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1(j_1\neq i_1),j_2(j_2\neq i_2)\\ \{i_1,j_1\}\neq\{i_2,j_2\}}} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_4}^2 W_{i_1j_1} W_{i_2j_2},$$

$$Y_{b3} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2(j_2\neq i_2),j_4(j_4\neq i_4)\\ \{i_2,j_2\}\neq\{i_4,j_4\}}} \eta_{i_2}\eta_{i_3}^2 W_{i_1i_4}^2 W_{i_2j_2} W_{i_4j_4},$$

$$Y_{b4} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1(j_1\neq i_1)} \eta_{i_2}\eta_{i_3}^2 W_{i_2i_4}^2 W_{i_1j_1} W_{i_1i_4},$$

$$Y_c = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,j_2,j_4\\ j_1\notin\{i_2,i_4\},j_2\notin\{i_1,i_4\},j_4\notin\{i_1,i_2\}}} \eta_{i_2}\eta_{i_3}^2 W_{i_1j_1} W_{i_2j_2} W_{i_4j_4} W_{i_1i_4}.$$

We now show the two claims in (2.E.170) separately.

Consider the first claim of (2.E.170). It is seen that out of the 8 terms on the right hand side of (2.E.232), the mean of all terms are 0, except that of the Y_{a2} and Y_{a3}. Note that for any $1 \leq i,j \leq n$, $i \neq j$, $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1-\Omega_{ij})$, where Ω_{ij} are upper bounded by $o(1)$ uniformly for all such i,j. It follows

$$\begin{aligned}\mathbb{E}[Y_{a2}] &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2 \mathbb{E}[W_{i_1i_2}^2]\mathbb{E}[W_{i_1i_4}^2] \\ &= -(1+o(1)) \cdot v^{-3/2} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2 \Omega_{i_1i_2}\Omega_{i_1i_4}.\end{aligned}$$

Since for any $1 \leq i,j \leq n$, $i \neq j$, $0 < \eta_i \leq C\theta_i$, $\Omega_{ij} \leq C\theta_i\theta_j$ and $v \asymp \|\theta\|_1^2$,

$$|\mathbb{E}[Y_{a2}]| \leq C(\|\theta\|_1)^{-3} \sum_{i_1,i_2,i_3,i_4(dist)} \theta_{i_1}^2\theta_{i_2}^2\theta_{i_3}^2\theta_{i_4} \leq C\|\theta\|^6/\|\theta\|_1^2.$$

Therefore,

$$|\mathbb{E}[Y_{a2}]| \leq C\|\theta\|^6/\|\theta\|_1^2. \tag{2.E.206}$$

By symmetry, we similarly find

$$|\mathbb{E}[Y_{a3}]| \leq C\|\theta\|^6/\|\theta\|_1^2. \tag{2.E.207}$$

Combining (2.E.206) and (2.E.207) gives

$$\mathbb{E}[|T_{1b}|] \leq C\|\theta\|^6/\|\theta\|_1^2.$$

This completes the proof of the first claim of (2.E.170).

We now show the second claim of (2.E.170) . By Cauchy-Schwartz inequality,

$$\begin{aligned}\mathrm{Var}(T_{1b}) &\leq C(\mathrm{Var}(Y_{a1}) + \mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) + \sum_{s=1}^{4} \mathrm{Var}(Y_{bs}) + \mathrm{Var}(Y_c)) \\ &\leq C(\mathrm{Var}(Y_{a1}) + \mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) + \sum_{s=1}^{4} \mathbb{E}[Y_{bs}^2] + \mathbb{E}[Y_c^2]). \qquad (2.E.208)\end{aligned}$$

We now show $\mathrm{Var}(Y_{a1})$, $\mathrm{Var}(Y_{a2})$, $\mathrm{Var}(Y_{a3})$, $\sum_{s=1}^4 \mathbb{E}[Y_{bs}^2]$, and $\mathbb{E}[Y_c^2]$, separately.

Consider $\mathrm{Var}(Y_{a1})$. Recalling $\mathbb{E}[Y_{a1}] = 0$, we write $\mathrm{Var}(Y_{a1})$ as

$$v^{-3} \sum_{\substack{i_1,i_2,i_3,i_4 (dist) \\ i_1',i_2',i_3',i_4' (dist)}} \sum_{j_2 (j_2 \neq i_2)} \sum_{j_2' (j_2' \neq i_2')} \eta_{i_2}\eta_{i_3}^2 \eta_{i_2'}\eta_{i_3'}^2 \mathbb{E}\big[W_{i_1 i_4}^3 W_{i_2 j_2} W_{i_1' i_4'}^3 W_{i_2' j_2'}\big]. \qquad (2.E.209)$$

In the sum, a term is nonzero only when one of the following cases happens.

- (A). $\{W_{i_1 i_4}, W_{i_2 j_2}, W_{i_1' i_4'}, W_{i_2' j_2'}\}$ has 2 distinct random variables.
- (B). $\{W_{i_1 i_4}, W_{i_2 j_2}, W_{i_1' i_4'}, W_{i_2' j_2'}\}$ has 3 distinct random variables. While it may seem we have 4 possibilities in this case, but the only one that has a nonzero mean is when $W_{i_2 j_2} = W_{i_2' j_2'}$.

For Case (A), the two sets $\{i_1, i_2, i_4, j_2\}$ and $\{i_1', i_2', i_4', j_2'\}$ are identical, and so for two integers $0 \leq b_1, b_2 \leq 1$ and $b_1 + b_2 = 1$,

$$W_{i_1 i_4}^3 W_{i_2 j_2} W_{i_1' i_4'}^3 W_{i_2' j_2'} = W_{i_1 i_4}^{4+2b_1} W_{i_2 j_2}^{2+2b_2},$$

and so

$$\mathbb{E}[W_{i_1 i_4}^3 W_{i_2 j_2} W_{i_1' i_4'}^3 W_{i_2' j_2'}] = \mathbb{E}[W_{i_1 i_4}^{4+2b_1} W_{i_2 j_2}^{2+2b_2}] = \mathbb{E}[W_{i_1 i_4}^{4+2b_1}]\mathbb{E}[W_{i_2 j_2}^{2+2b_2}],$$

Note that for any integer $2 \leq b \leq 6$,

$$0 < \mathbb{E}[W_{ij}^b] \leq C\Omega_{ij},$$

where note that $\Omega_{ij} \leq C\theta_i\theta_j$ for all $1 \leq i, j \leq n$, $i \leq j$. Recall that $v \sim \|\theta\|_1^2$, and that $0 < \eta_i \leq C\theta_i$ for all $1 \leq i \leq n$. Combining these that, the contribution of Case (A) to $\mathrm{Var}(Y_{a1})$ is no more than

$$C(\|\theta\|_1)^{-6} \sum_{i_1,\cdots,i_4 (dist)} \sum_{i_3', j_2} \sum_{a} \theta_{i_1}^{a_1+1}\theta_{i_2}^{a_2+2}\theta_{i_3}^2\theta_{i_4}^{a_3+1}\theta_{i_3'}^2\theta_{j_2}^{a_4+1}, \qquad (2.E.210)$$

where $a = (a_1, a_2, a_3, a_4)$ and each a_i is either 0 and 1, satisfying $a_1 + a_2 + a_3 + a_4 = 1$. Note that the right hand side of (2.E.210) is no greater than

$$C(\|\theta\|_1)^{-6} \max\{\|\theta\|_1^3\|\theta\|^4\|\theta\|_3^3, \|\theta\|_1^2\|\theta\|^8\} \leq C\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1^3,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$.

Next, consider Case (B). In this case, $\{i_2, j_2\} = \{i_2', j_2'\}$ and

$$W_{i_1 i_4}^3 W_{i_2 j_2} W_{i_1' i_4'}^3 W_{i_2' j_2'} = W_{i_1 i_4}^3 W_{i_2 j_2}^2 W_{i_1' i_4'}^3,$$

and by similar argument,

$$0 < \mathbb{E}[W_{i_1 i_4}^3 W_{i_2 j_2}^2 W_{i_1' i_4'}^3] \leq C\Omega_{i_1 i_4}\Omega_{i_2 j_2}\Omega_{i_1' i_4'}. \qquad (2.E.211)$$

Recall that $\Omega_{ij} \leq C\theta_i\theta_j$ for all $1 \leq i,j \leq n$, $i \leq j$, that $v \sim \|\theta\|_1^2$, and that $0 < \eta_i \leq C\theta_i$ for all $1 \leq i \leq n$. Combining this with (2.E.209), the contribution of this case to $\mathrm{Var}(Y_{a1})$

$$\leq C(\|\theta\|_1)^{-6} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i_1',i_3',i_4'(dist)}} \sum_{j_2} C\theta_{i_1}\theta_{i_2}^{2+b_1}\theta_{i_3}^2\theta_{i_4}\theta_{i_1'}\theta_{i_3'}^2\theta_{i_4'}\theta_{j_2}^{1+b_2}, \tag{2.E.212}$$

where similarly b_1, b_2 are either 0 or 1 and $b_1 + b_2 = 1$. By similar argument, the right hand side

$$\leq C\|\theta\|_1^{-6} \cdot [\|\theta\|_1^5\|\theta\|^4\|\theta\|_3^3 + \|\theta\|_1^4\|\theta\|^8] \leq C\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1,$$

where we've used Cauchy-Schwartz inequality that $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$.

Now, inserting (2.E.210) and (2.E.212) into (2.E.209) gives

$$\mathrm{Var}(Y_{a1}) \leq C[\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1^3 + \|\theta\|^4\|\theta\|_3^3/\|\theta\|_1] \leq C\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.213}$$

where we have used $\|\theta\|_1 \to \infty$ and $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$. This shows

$$\mathrm{Var}(Y_{a1}) \leq C\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.214}$$

Next, we consider $\mathrm{Var}(Y_{a2})$ and $\mathrm{Var}(Y_{a3})$. The proofs are similar to that of $\mathrm{Var}(X_a)$ of Item (a), so we skip the detail, but claim that

$$\mathrm{Var}(Y_{a2}) \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4, \tag{2.E.215}$$

and

$$\mathrm{Var}(Y_{a3}) \leq C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4. \tag{2.E.216}$$

Combining (2.E.214), (2.E.215), and (2.E.216) gives

$$\mathrm{Var}(Y_{a1}) + \mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) \leq C[\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4] \leq C\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.217}$$

where we have used the universal inequality that $\|\theta\|_3^3 \leq \|\theta\|_1^3$.

Next, consider $\sum_{s=1}^4 \mathbb{E}[Y_{bs}^2]$. For each $1 \leq s \leq 4$, the study of $\mathbb{E}[Y_{bs}^2]$ is similar to that of $\mathbb{E}[X_{b1}^2]$ in Item (a), so we skip the details. We have that both under the null and the alternative,

$$\mathbb{E}[Y_{b1}^2] \leq C\|\theta\|^{12}/\|\theta\|_1^4, \tag{2.E.218}$$

$$\mathbb{E}[Y_{b2}^2] \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.219}$$

$$\mathbb{E}[Y_{b3}^2] \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.220}$$

$$\mathbb{E}[Y_{b4}^2] \leq C\|\theta\|^{12}/\|\theta\|_1^4. \tag{2.E.221}$$

Therefore,

$$\sum_{s=1}^4 \mathbb{E}[Y_{bs}^2] \leq C[\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^{12}/\|\theta\|_1^4] \leq C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.222}$$

Third, we consider $\mathbb{E}[Y_c^2]$. The proof is very similar to that of $\mathbb{E}[X_c^2]$ and we have that both under the null and the alternative,

$$\mathbb{E}[Y_c^2] \leq C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^3. \tag{2.E.223}$$

Finally, combining (2.E.217), (2.E.222), and (2.E.223) with (2.E.208) gives

$$\mathrm{Var}(T_{1b}) \le C[\|\theta\|^4\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^6\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^3] \le C\|\theta\|^6\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.224}$$

where we have used $\|\theta\| \to \infty$ and $\|\theta\|^2 \ll \|\theta\|_1$. This completes the proof of (2.E.170).

Consider Item (c). The goal is to show (2.E.171). Recall that

$$T_{1c} = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot W_{i_4 i_1},$$

and that

$$\widetilde{\eta} - \eta = v^{-1/2} W 1_n.$$

Plugging this into T_{1c} gives

$$\begin{aligned} T_{1c} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4}\Big(\sum_{j_2\neq i_2} W_{i_2j_2}\Big)\Big(\sum_{\ell_2\neq i_2} W_{i_2\ell_2}\Big)\Big(\sum_{j_3\neq i_3} W_{i_3j_3}\Big) W_{i_1i_4} \\ &= -\frac{1}{v^{3/2}} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ j_2\neq i_2,\ell_2\neq i_2,j_3\neq i_3}} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}W_{i_1i_4}. \end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}W_{i_1i_4} = \begin{cases} W_{i_2i_3}^3 W_{i_1i_4}, & \text{if } j_2 = \ell_2 = i_3, j_3 = i_2, \\ W_{i_2j_2}^2 W_{i_3j_3} W_{i_1i_4}, & \text{if } j_2 = \ell_2, (j_3, j_2) \neq (i_2, i_3), \\ W_{i_2i_3}^2 W_{i_2\ell_2} W_{i_1i_4}, & \text{if } j_2 = i_3, j_3 = i_2, \ell_2 \neq i_3, \\ W_{i_2i_3}^2 W_{i_2j_2} W_{i_1i_4}, & \text{if } \ell_2 = i_3, j_3 = i_2, j_2 \neq i_3, \\ W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}W_{i_1i_4}, & \text{otherwise.} \end{cases} \tag{2.E.225}$$

This allows us to further split T_{1c} into 5 different terms:

$$T_{1c} = Z_a + Z_{b1} + Z_{b2} + Z_{b3} + Z_c, \tag{2.E.226}$$

where

$$\begin{aligned} Z_a &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^3 W_{i_1i_4}, \\ Z_{b1} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2,(j_3,j_2)\neq(i_2,i_3)} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2j_2}^2 W_{i_3j_3} W_{i_1i_4}, \\ Z_{b2} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2=i_3,j_3=i_2\\ \ell_2\neq i_3}} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_2\ell_2} W_{i_1i_4}, \\ Z_{b3} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{\ell_2=i_3,j_3=i_2\\ j_2\neq i_3}} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_2j_2} W_{i_1i_4}, \\ Z_c &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2,\ell_2,j_3\\ j_2\neq\ell_2,j_2,\ell_2\neq i_3,j_3\neq i_2}} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}W_{i_1i_4}. \end{aligned}$$

We now show the two claims in (2.E.171) separately. The proof of the first claim is trivial, so we only show the second claim of (2.E.171).

Consider the second claim of (2.E.171). By Cauchy-Schwartz inequality,

$$\begin{aligned}\mathrm{Var}(T_{1c}) &\leq C(\mathrm{Var}(Z_a) + \mathrm{Var}(Z_{b1}) + \mathrm{Var}(Z_{b2}) + \mathrm{Var}(Z_{b3}) + \mathrm{Var}(Z_c)) \\ &\leq C(\mathbb{E}[Z_a^2] + \sum_{s=1}^{3} \mathbb{E}[Z_{bs}^2] + \mathbb{E}[Z_c^2]). \end{aligned} \tag{2.E.227}$$

Note that

- The proof of $\mathrm{Var}(Z_a)$ is similar to that of $\mathrm{Var}(Y_a)$ in Item (b).
- The proof of $\sum_{s=1}^{3} \mathbb{E}[Z_{bs}^2]$ is similar to that of $\sum_{s=1}^{4} \mathbb{E}[X_{bs}^2]$ in Item (a).
- The proof of $\mathbb{E}[Z_c^2]$ is similar to that of $\mathbb{E}[X_c^2]$ in Item (a).

For these reasons, we omit the proof details and only state the claims. We have that under both the null and the alternative,

$$\mathrm{Var}(Z_a) \leq C\|\theta\|^4 \|\theta\|_3^6 / \|\theta\|_1^4, \tag{2.E.228}$$

$$\sum_{s=1}^{3} \mathbb{E}[Z_{bs}^2] \leq C\|\theta\|_3^9 / \|\theta\|_1, \tag{2.E.229}$$

and

$$\mathbb{E}[Z_c^2] \leq C\|\theta\|^2 \|\theta\|_3^9 / \|\theta\|_1^3. \tag{2.E.230}$$

Inserting (2.E.228), (2.E.229), and (2.E.230) into (2.E.227) gives

$$\mathrm{Var}(T_{1c}) \leq C[\|\theta\|^4 \|\theta\|_3^6 / \|\theta\|_1^4 + \|\theta\|_3^9 / \|\theta\|_1 + \|\theta\|^2 \|\theta\|_3^9 / \|\theta\|_1^3] \leq C\|\theta\|_3^9 / \|\theta\|_1,$$

where we have used $\|\theta\|_3^3 \ll \|\theta\|^2 \ll \|\theta\|_1$, $\|\theta\|^4 \leq \|\theta\|_1 \|\theta\|_3^3$ and $\|\theta\|_1 \to \infty$. This proves (2.E.171).

Consider Item (d). The goal is to show (2.E.172) and (2.E.173). Recall that

$$T_{1d} = -\frac{1}{v^{3/2}} \sum_{i_1, i_2, i_3, i_4 (dist)} \eta_{i_1} \eta_{i_3}^2 \big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2 (\eta_{i_4} - \widetilde{\eta}_{i_4})\big] \cdot W_{i_4 i_1}.$$

and that

$$\widetilde{\eta} - \eta = v^{-1/2} W 1_n.$$

Plugging this into T_{1d} gives

$$\begin{aligned} T_{1d} &= -\frac{1}{v^{3/2}} \sum_{i_1, i_2, i_3, i_4 (dist)} \eta_{i_1} \eta_{i_3}^2 \Big(\sum_{j_2 \neq i_2} W_{i_2 j_2}\Big)\Big(\sum_{\ell_2 \neq i_2} W_{i_2 \ell_2}\Big)\Big(\sum_{j_4 \neq i_4} W_{i_4 j_4}\Big) W_{i_1 i_4} \\ &= -\frac{1}{v^{3/2}} \sum_{\substack{i_1, i_2, i_3, i_4 (dist) \\ j_2 \neq i_2, \ell_2 \neq i_2, j_4 \neq i_4}} \eta_{i_1} \eta_{i_3}^2 W_{i_2 j_2} W_{i_2 \ell_2} W_{i_4 j_4} W_{i_1 i_4}. \end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_2j_2}W_{i_2\ell_2}W_{i_4j_4}W_{i_1i_4} = \begin{cases} W_{i_2i_4}^3 W_{i_1i_4}, & \text{if } j_2=\ell_2=i_4, j_4=i_2, \\ W_{i_2j_2}^2 W_{i_1i_4}^2, & \text{if } j_2=\ell_2, j_4=i_1, \\ W_{i_2j_2}^2 W_{i_4j_4} W_{i_1i_4}, & \text{if } j_2=\ell_2, j_4\neq i_1, (j_2,j_4)\neq(i_4,i_2), \\ W_{i_2j_2}W_{i_2i_4}^2 W_{i_1i_4}, & \text{if } \ell_2=i_4, j_4=i_2, j_2\neq i_4, \\ W_{i_2\ell_2}W_{i_2i_4}^2 W_{i_1i_4}, & \text{if } j_2=i_4, j_4=i_2, \ell_2\neq i_4, \\ W_{i_2j_2}W_{i_2\ell_2}W_{i_1i_4}^2, & \text{if } j_4=i_1, j_2\neq\ell_2, \\ W_{i_2j_2}W_{i_2\ell_2}W_{i_4j_4}W_{i_1i_4}, & \text{otherwise.} \end{cases} \tag{2.E.231}$$

This allows us to further split T_{14} into 7 different terms:

$$T_{1d} = U_{a1}+U_{a2}+U_{b1}+U_{b2}+U_{b3}+U_{b4}+U_c, \tag{2.E.232}$$

where

$$U_{a1} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_3}^2 W_{i_2i_4}^3 W_{i_1i_4},$$

$$U_{a2} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2} \eta_{i_1}\eta_{i_3}^2 W_{i_2j_2}^2 W_{i_1i_4}^2,$$

$$U_{b1} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_2(j_2\neq i_2), j_4(j_4\neq i_4)\\ j_4\neq i_1, (j_2,j_4)\neq(i_4,i_2)}} \eta_{i_1}\eta_{i_3}^2 W_{i_2j_2}^2 W_{i_4j_4} W_{i_1i_4},$$

$$U_{b2} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2(j_2\neq i_4)} \eta_{i_1}\eta_{i_3}^2 W_{i_2j_2} W_{i_2i_4}^2 W_{i_1i_4},$$

$$U_{b3} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\ell_2(\ell_2\neq i_4)} \eta_{i_1}\eta_{i_3}^2 W_{i_2\ell_2} W_{i_2i_4}^2 W_{i_1i_4},$$

$$U_{b4} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2\neq\ell_2} \eta_{i_1}\eta_{i_3}^2 W_{i_2j_2} W_{i_2\ell_2} W_{i_1i_4}^2,$$

$$U_c = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2,\ell_2,j_4,W\,\mathrm{dist}} \eta_{i_1}\eta_{i_3}^2 W_{i_2j_2} W_{i_2\ell_2} W_{i_4j_4} W_{i_1i_4}.$$

We now show (2.E.172) and (2.E.173) separately.

Consider (2.E.172). It is seen that out of the 7 terms on the right hand side of (2.E.226), all terms are mean 0, except for the second term U_{a2}. Note that for any $1\le i,j\le n$, $i\neq j$, $\mathbb{E}[W_{ij}^2]=\Omega_{ij}(1-\Omega_{ij})$, where Ω_{ij} are upper bounded by $o(1)$ uniformly for all such i,j. It follows

$$\begin{aligned}\mathbb{E}[U_{a2}] &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2} \eta_{i_1}\eta_{i_3}^2 \mathbb{E}[W_{i_2j_2}^2]\mathbb{E}[W_{i_1i_4}^2] \\ &= -(1+o(1))\cdot v^{-3/2} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ j_2}} \eta_{i_1}\eta_{i_3}^2 \Omega_{i_2j_2}\Omega_{i_1i_4}.\end{aligned}$$

Under null, for any $1 \leq i, j \leq n$, $i \neq j$, $\eta_i = (1+o(1))\theta_i$, $\Omega_{ij} = (1+o(1))\theta_i\theta_j$ and $v \asymp \|\theta\|_1^2$,

$$\mathbb{E}[U_{a2}] = (\|\theta\|_1)^{-3} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2} \theta_{i_1}^2 \theta_{i_2} \theta_{i_3}^2 \theta_{i_4} \theta_{j_2} = -(1+o(1))\|\theta\|^4,$$

and under alternative, a similar arguments yields

$$|\mathbb{E}[U_{a1}]| \leq C\|\theta\|^4. \tag{2.E.233}$$

This proves (2.E.172).

We now consider (2.E.173). By Cauchy-Schwartz inequality,

$$\begin{aligned}\mathrm{Var}(T_{1d}) &\leq C(\mathrm{Var}(U_{a1}) + \mathrm{Var}(U_{a2}) + \sum_{s=1}^{4} \mathrm{Var}(U_{bs}) + \mathrm{Var}(U_c)) \\ &\leq C(\mathrm{Var}(U_{a1}) + \mathrm{Var}(U_{a2}) + \sum_{s=1}^{4} \mathbb{E}[U_{bs}^2] + \mathbb{E}[U_c^2]).\end{aligned} \tag{2.E.234}$$

Note that

- The proof of U_{a1} is similar to that of Y_{a1} in Item (b).
- The proof of U_{a2} is similar to that of X_{a1} in Item (a).
- The proof of U_{bs}, $1 \leq s \leq 4$, is similar to that of X_{b1} in Item (a).
- The proof of U_c is similar to that of X_c in Item (a).

For these reasons, we omit the proof details, and claim that

$$\mathrm{Var}(U_{a1}) \leq C\|\theta\|^4 \|\theta\|_3^6 / \|\theta\|_1^4, \tag{2.E.235}$$

$$\mathrm{Var}(U_{a2}) \leq C\|\theta\|^4 \|\theta\|_3^3 / \|\theta\|_1, \tag{2.E.236}$$

$$\sum_{s=1}^{4} \mathbb{E}[U_{bs}^2] \leq C\|\theta\|^6 \|\theta\|_3^3 / \|\theta\|_1, \tag{2.E.237}$$

and

$$\mathrm{Var}(U_c) \leq C\|\theta\|^8 \|\theta\|_3^3 / \|\theta\|_1^3, \tag{2.E.238}$$

Inserting (2.E.235), (2.E.236), (2.E.237), and (2.E.238) into (2.E.234) gives

$$\mathrm{Var}(T_{1d}) \leq C[\|\theta\|^4 \|\theta\|_3^6 / \|\theta\|_1^4 + \|\theta\|^4 \|\theta\|_3^3 / \|\theta\|_1 + \|\theta\|^6 \|\theta\|_3^3 / \|\theta\|_1 + \|\theta\|^8 \|\theta\|_3^3 / \|\theta\|_1^3] \tag{2.E.239}$$

$$\leq C\|\theta\|^6 \|\theta\|_3^3 / \|\theta\|_1, \tag{2.E.240}$$

where we have used $\|\theta\| \to \infty$ and $\|\theta\|_3^3 \leq \|\theta\|_1^3$. This proves (2.E.173).

We now consider Item (e) and Item (f). Since the proof is similar, we only prove Item (e). The goal is to show (2.E.174). Recall that

$$T_{2a} = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4} \big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})\big] \cdot \widetilde{\Omega}_{i_4 i_1}, \tag{2.E.241}$$

and

$$\widetilde{\eta} - \eta = v^{-1/2} W 1_n. \tag{2.E.242}$$

Plugging (2.E.242) into (2.E.241) gives

$$\begin{aligned} T_{2a} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}\eta_{i_4}\Big(\sum_{j_1\neq i_1} W_{i_1j_1}\Big)\Big(\sum_{j_2\neq i_2} W_{i_2j_2}\Big)\Big(\sum_{j_3\neq i_3} W_{i_3j_3}\Big)\widetilde{\Omega}_{i_4i_1} \\ &= -\frac{1}{v^{3/2}} \sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ j_1\neq i_1, j_2\neq i_2, j_3\neq i_3}} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}\widetilde{\Omega}_{i_1i_4}. \end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_1j_1}W_{i_2j_2}W_{i_3j_3} = \begin{cases} W_{i_1i_2}^2 W_{i_3j_3}, & \text{if } j_1 = i_2, j_2 = i_1, \\ W_{i_1i_3}^2 W_{i_2j_2}, & \text{if } j_1 = i_3, j_3 = i_1, \\ W_{i_2i_3}^2 W_{i_1j_1}, & \text{if } j_2 = i_3, j_3 = i_2, \\ W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}, & \text{otherwise.} \end{cases} \tag{2.E.243}$$

This allows us to further split T_{2a} into 4 different terms:

$$T_{2a} = X_{a1} + X_{a2} + X_{a3} + X_b, \tag{2.E.244}$$

where

$$\begin{aligned} X_{a1} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_3\neq i_3} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_2}^2 W_{i_3j_3}\widetilde{\Omega}_{i_1i_4}, \\ X_{a2} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_2\neq i_2} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_3}^2 W_{i_2j_2}\widetilde{\Omega}_{i_1i_4}, \\ X_{a3} &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1\neq i_1} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_1j_1}\widetilde{\Omega}_{i_1i_4}, \\ X_b &= -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,j_2,j_3\\ j_k\neq i_\ell, k,\ell=1,2,3}} \eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}\widetilde{\Omega}_{i_1i_4}. \end{aligned}$$

We now consider the two claims of (2.E.174) separately. Since the mean of $X_{a1}, X_{a2}, X_{a3}, X_b$ are all 0, the first claim of (2.E.174) follows trivially, so all remains to show is the second claim of (2.E.174).

We now consider the second claim of (2.E.174). By Cauchy-Schwartz inequality,

$$\begin{aligned} \mathrm{Var}(T_{2a}) &\leq C\mathrm{Var}(X_{a1}) + \mathrm{Var}(X_{a2}) + \mathrm{Var}(X_{a3}) + \mathrm{Var}(X_b)) \\ &\leq C(\mathbb{E}[X_{a1}^2] + \mathbb{E}[X_{a2}^2] + \mathbb{E}[X_{a3}^2] + \mathbb{E}[X_b^2]). \end{aligned} \tag{2.E.245}$$

We now consider $\mathbb{E}[X_{a1}^2] + \mathbb{E}[X_{a2}^2] + \mathbb{E}[X_{a3}^2]$, and $\mathbb{E}[X_b^2]$, separately.

Consider $\mathbb{E}[X_{a1}^2]+\mathbb{E}[X_{a2}^2]+\mathbb{E}[X_{a3}^2]$. We claim that both under the null and the alternative,

$$\mathbb{E}[X_{a1}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.246}$$

$$\mathbb{E}[X_{a2}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.247}$$

$$\mathbb{E}[X_{a3}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5. \tag{2.E.248}$$

Combining these gives that both under the null and the alternative,

$$\mathbb{E}[X_{a1}^2] + \mathbb{E}[X_{a2}^2] + \mathbb{E}[X_{a3}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5. \tag{2.E.249}$$

It remains to show (2.E.246)-(2.E.248). Since the proofs are similar, we only prove

(2.E.246). Write

$$\mathbb{E}[X_{a1}^2] = v^{-3} \sum_{\substack{i_1,i_2,i_3,i_4 (dist)\\ i_1',i_2',i_3',i_4' (dist)}} \sum_{\substack{j_3,j_3'\\ j_3\neq i_3, j_3'\neq i_3'}} \eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'}\mathbb{E}[W_{i_1i_2}^2 W_{i_3j_3} W_{i_1'i_2'}^2 W_{i_3'j_3'}]\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i_1'i_4'}.$$

Consider the term

$$W_{i_1i_2}^2 W_{i_3j_3} W_{i_1'i_2'}^2 W_{i_3'j_3'}.$$

In order for the mean is nonzero, we have three cases

- Case A. $W_{i_1i_2} = W_{i_3'j_3'}$ and $W_{i_3j_3} = W_{i_1'i_2'}$.
- Case B. $W_{i_3j_3} = W_{i_3'j_3'}$ and $W_{i_1i_2} = W_{i_1'i_2'}$.
- Case C. $W_{i_3j_3} = W_{i_3'j_3'}$ and $W_{i_1i_2} \neq W_{i_1'i_2'}$.

Consider Case A. In this case, $\{i_1', i_2', i_3'\}$ are three distinct indices in $\{i_1, i_2, i_3, j_3\}$. In this case,

$$W_{i_1i_2}^2 W_{i_3j_3} W_{i_1'i_2'}^2 W_{i_3'j_3'} = W_{i_1i_2}^3 W_{i_3j_3}^3,$$

where by similar arguments as before

$$0 < \mathbb{E}[W_{i_1i_2}^3 W_{i_3j_3}^3] \leq C\Omega_{i_1i_2}\Omega_{i_3j_3} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{j_3}.$$

At the same time, recall that that $0 < \eta_i \leq C\theta_i$ for any $1 \leq i \leq n$, and that $|\widetilde{\Omega}_{ij}| \leq C\alpha\theta_i\theta_j$ for any $1 \leq i, j \leq n$, $i \neq j$, where $\alpha = |\lambda_2/\lambda_1|$ with λ_k being the k-th largest (in magnitude) eigenvalue of Ω, $1 \leq k \leq K$. By basic algebra,

$$|\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'}\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i_1'i_4'}| \leq C\alpha^2\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}^2\theta_{i_1'}\theta_{i_2'}\theta_{i_3'}\theta_{i_4'}^2.$$

Note that in the current case, $\{i_1, i_2\} = \{i_3', j_3'\}$ and $\{i_3, j_3\} = \{i_1', i_2'\}$, so for some integers $0 \leq b_1, b_2 \leq 1$ and $b_1 + b_2 = 1$,

$$\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}^2\theta_{i_1'}\theta_{i_2'}\theta_{i_3'}\theta_{i_4'}^2 = \theta_{i_1}^{1+b_1}\theta_{i_2}^{1+b_2}\theta_{i_3}^2\theta_{j_3}\theta_{i_4}^2\theta_{i_4'}^2.$$

Recall that $v \asymp \|\theta\|_1^2$. Combining these, the contribution of Case (A) to $\mathbb{E}[X_{a1}^2]$ is no greater than

$$C\alpha^2(\|\theta\|_1)^{-6} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{i_4'} \sum_{j_3 (j_3\neq i_3)} \sum_{b_1,b_2 (b_1+b_2=1)} \theta_{i_1}^{2+b_1}\theta_{i_2}^{2+b_2}\theta_{i_3}^3\theta_{j_3}^2\theta_{i_4}^2\theta_{i_4'}^2,$$

where the right hand side $\leq C\alpha^2 \cdot \|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6$. This shows that the contribution of Case (A) to $\mathbb{E}[X_{a1}^2]$ is no greater than

$$C\alpha^2 \cdot \|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6. \tag{2.E.250}$$

Consider Case B. By similar arguments,

$$W_{i_1i_2}^2 W_{i_3j_3} W_{i_1'i_2'}^2 W_{i_3'j_3'} = W_{i_1i_2}^6 W_{i_3j_3}^2,$$

where

$$\mathbb{E}[W_{i_1i_2}^6 W_{i_3j_3}^2] \leq C\Omega_{i_1i_2}\Omega_{i_3j_3} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{j_3},$$

Also, by similar arguments,

$$|\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'}\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i_1'i_4'}| \leq C\alpha^2\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}^2\theta_{i_1'}\theta_{i_2'}\theta_{i_3'}\theta_{i_4'}^2,$$

where as $W_{i_1i_2} = W_{i'_1i'_2}$ and $W_{i_3j_3} = W_{i'_3j'_3}$, the right hand side

$$\leq C\alpha^2\theta_{i_1}^2\theta_{i_2}^2\theta_{i_3}^{1+c_1}\theta_{j_3}^{c_2}\theta_{i_4}^2\theta_{i'_4}^2,$$

where $0 < c_1, c_2 \leq$ are integers satisfying $c_1 + c_2 = 1$. Recall $v \sim \|\theta\|_1^2$. Combining these, the contribution of Case (B) to $\mathbb{E}[X_{a1}^2]$

$$\leq C\alpha^2(\|\theta\|_1)^{-6}\sum_{i_1,i_2,i_3,i_4(dist)}\sum_{i'_4}\sum_{j_3(j_3\neq i_3)}\sum_{b_1,b_2(b_1+b_2=1)}\theta_{i_1}^3\theta_{i_2}^3\theta_{i_3}^{2+c_1}\theta_{j_3}^{1+c_2}\theta_{i_4}^2\theta_{i'_4}^2,$$

where by $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$, the above term

$$\leq C\alpha^2[\|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^5, \|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6] \leq C\alpha^2\|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^5.$$

This shows that the contribution of Case (B) to $\mathbb{E}[X_{a1}^2]$ is no greater than

$$C\|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^5. \tag{2.E.251}$$

Consider Case (C). In this case,

$$W_{i_1i_2}^2W_{i_3j_3}W_{i'_1i'_2}^2W_{i'_3j'_3} = W_{i_1i_2}^2W_{i_3j_3}^2W_{i'_1i'_2}^2,$$

where by similar arguments,

$$\mathbb{E}[W_{i_1i_2}^2W_{i_3j_3}^2W_{i'_1i'_2}^2] \leq C\Omega_{i_1i_2}\Omega_{i_3j_3}\Omega_{i'_1i'_2} \leq C\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{j_3}\theta_{i'_1}\theta_{i'_2}.$$

Also, by similar arguments,

$$|\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i'_2}\eta_{i'_3}\eta_{i'_4}\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i'_1i'_4}| \leq C\alpha^2\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}^2\theta_{i'_1}\theta_{i'_2}\theta_{i'_3}\theta_{i'_4}^2,$$

where as $W_{i_3j_3} = W_{i'_3j'_3}$, the right hand side

$$\leq C\alpha^2\theta_{i_1}\theta_{i_2}\theta_{i_3}^{1+c_1}\theta_{j_3}^{c_2}\theta_{i_4}^2\theta_{i'_4}^2,$$

with the same c_1, c_2 as in the proof of Case B. Combining these and using $v \asymp \|\theta\|_1^2$, we have that under both the null and the alternative, the contribution of Case (C) to $\mathbb{E}[X_{a1}^2]$

$$\leq C\alpha^2(\|\theta\|_1)^{-6}\sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i'_1,i'_2,i'_4(dist)}}\sum_{j_3(j_3\neq i_3)}\theta_{i_1}^2\theta_{i_2}^2\theta_{i_3}^{2+c_1}\theta_{j_3}^{1+c_2}\theta_{i_4}^2\theta_{i'_1}^2\theta_{i'_2}^2\theta_{i'_4}^2,$$

where the right hand size

$$\leq C\alpha^2\cdot[\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5 + \|\theta\|^{12}\|\theta\|_3^6/\|\theta\|_1^6] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5. \tag{2.E.252}$$

Here we have again used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$.

Combining (2.E.250), (2.E.251), and (2.E.252) gives

$$\mathbb{E}[X_{a1}^2] \leq C\alpha^2(\|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6 + \|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^5 + \|\theta\|^8\|\theta\|_3^9/\|\theta\|^5] \leq C\alpha^2\|\theta\|^8\|\theta\|_3^9/\|\theta\|_1^5,$$

where we have used $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$ and $\|\theta\| \to \infty$. This proves (2.E.246).

We now consider $\mathbb{E}[X_b^2]$. Write

$$\mathbb{E}[X_b^2] = v^{-3}\sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ i'_1,i'_2,i'_3,i'_4(dist)}}\sum_{\substack{j_3,j'_3\\ j_3\neq i_3,j'_3\neq i'_3}}\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i'_2}\eta_{i'_3}\eta_{i'_4}$$
$$\mathbb{E}[W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i'_1j'_1}W_{i'_2j'_2}W_{i'_3j'_3}]\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i'_1i'_4}.$$

Consider

$$W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}, \qquad \text{and} \qquad W_{i'_1j'_1}W_{i'_2j'_2}W_{i'_3j'_3}.$$

Each term has a mean 0, and two terms are uncorrelated with each other if and only if the two sets of random variables $\{W_{i_1j_1}, W_{i_2j_2}, W_{i_3j_3}\}$ and $\{W_{i_1'j_1'}, W_{i_2'j_2'}, W_{i_3'j_3'}\}$ are identical (however, it is possible that $W_{i_1j_1}$ does not equal to $W_{i_1'j_1'}$ but equals to $W_{i_2'j_2'}$, say). When this happens, first, $\{i_1,i_2,i_3,j_1,j_2,j_3\} = \{i_1',i_2',i_3',j_1',j_2',j_3'\}$. Recall that $|\widetilde{\Omega}_{ij}| \le C\alpha\theta_i\theta_j$ for all $1 \le i,j \le n$, $i \ne j$, and that $0 < \eta_i \le C\theta_i$ for all $1 \le i \le n$. For integers $a_i \in \{0,1\}$, $1 \le i \le 4$, that satisfy $\sum_{i=1}^6 a_i = 3$, we have

$$\begin{aligned}|\eta_{i_2}\eta_{i_3}\eta_{i_4}\eta_{i_2'}\eta_{i_3'}\eta_{i_4'}\widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i_1'i_4'}| &\le C\eta_{i_1}^{a_1}\eta_{j_1}^{a_2}\eta_{i_2}^{1+a_3}\eta_{j_2}^{a_4}\eta_{i_3}^{1+a_5}\eta_{j_3}^{a_6}\eta_{i_4}\eta_{i_4'}|\widetilde{\Omega}_{i_1i_4}||\widetilde{\Omega}_{i_1'i_4'}| \\ &\le C\alpha^2\theta_{i_1}^{1+a_1}\eta_{j_1}^{a_2}\eta_{i_2}^{1+a_3}\eta_{j_2}^{a_4}\eta_{i_3}^{1+a_5}\eta_{j_3}^{a_6}\eta_{i_4}^2\eta_{i_4'}^2.\end{aligned}$$

Second,

$$\mathbb{E}[W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_1'j_1'}W_{i_2'j_2'}W_{i_3'j_3'}] = \mathbb{E}[W_{i_1j_1}^2W_{i_2j_2}^2W_{i_3j_3}^2],$$

where by similar arguments, the right hand side

$$\le C\Omega_{i_1j_1}\Omega_{i_2j_2}\Omega_{i_3j_3} \le C\theta_{i_1}\theta_{j_1}\theta_{i_2}\theta_{j_2}\theta_{i_3}\theta_{j_3}.$$

Recall that $v \sim \|\theta\|_1^2$. Combining these gives

$$\mathbb{E}[X_b^2] \le C\alpha^2\|\theta\|_1^{-6}\sum_{i_1,i_2,i_3,i_4(dist)}\sum_{i_4'}\sum_{\substack{j_1,j_2,j_3\\ j_1\ne i_1, j_2\ne i_2, j_3\ne i_3}}\sum_a \theta_{i_1}^{2+a_1}\eta_{j_1}^{1+a_2}\eta_{i_2}^{2+a_3}\eta_{j_2}^{1+a_4}\eta_{i_3}^{2+a_5}\eta_{j_3}^{1+a_6}\eta_{i_4}^2\eta_{i_4'}^2,$$

where $a = (a_1, a_2, \ldots, a_6)$ as above. By the way a_i are defined, the right hand side

$$\le C\alpha^2\|\theta\|^4\Big(\sum_a \|\theta\|_{a_1+2}^{a_1+2}\cdot\|\theta\|_{a_2+1}^{a_2+1}\cdot\|\theta\|_{a_3+2}^{a_3+2}\cdot\|\theta\|_{a_4+1}^{a_4+1}\|\theta\|_{a_5+2}^{a_5+2}\|\theta\|_{a_6+1}^{a_6+1}\Big)/\|\theta\|_1^6,$$

which by $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$, the term in the bracket does not exceed

$$C\max\{\|\theta\|^{12}, \|\theta\|_1\|\theta\|^8\|\theta\|_3^3, \|\theta\|_1^2\|\theta\|^4\|\theta\|_3^6, \|\theta\|_1^3\|\theta\|_3^9\} \le C\|\theta\|_1^3\|\theta\|_3^9.$$

Combining these gives

$$\mathbb{E}[X_b^2] \le C\alpha^2\|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^3. \tag{2.E.253}$$

Finally, inserting (2.E.249)-(2.E.253) into (2.E.245) gives

$$\mathrm{Var}(T_{2a}) \le C\alpha^2[\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5 + \|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^3] \le C\alpha^2\|\theta\|^4\|\theta\|_3^9/\|\theta\|_1^3,$$

and (2.E.174) follows.

Consider Item (g) and Item (h). The proof are similar, so we only show Item (g). The goal is to show (2.E.176). Recall that

$$T_{2c} = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_2} - \widetilde{\eta}_{i_2})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})\big]\cdot\widetilde{\Omega}_{i_4i_1}, \tag{2.E.254}$$

and

$$\widetilde{\eta} - \eta = v^{-1/2}W1_n.$$

Plugging this into T_{2c} gives (note symmetry in $\widetilde{\Omega}$)

$$\begin{aligned}T_{2c} &= -\frac{1}{v^{2/3}}\sum_{i_1,i_2,i_3,i_4(dist)}\eta_{i_1}\eta_{i_3}\eta_{i_4}\Big(\sum_{j_2\ne i_2}W_{i_2j_2}\Big)\Big(\sum_{\ell_2\ne i_2}W_{i_2\ell_2}\Big)\Big(\sum_{j_3\ne i_3}W_{i_3j_3}\Big)\widetilde{\Omega}_{i_4i_1} \\ &= -\frac{1}{v^{3/2}}\sum_{\substack{i_1,i_2,i_3,i_4(dist)\\ j_1\ne i_1, j_2\ne i_2, j_3\ne i_3}}\eta_{i_1}\eta_{i_3}\eta_{i_4}W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}\widetilde{\Omega}_{i_1i_4}.\end{aligned}$$

By basic combinatorics and careful observations, we have

$$W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3} = \begin{cases} W_{i_2i_3}^3, & \text{if } j_1 = \ell_2 = i_3, j_3 = i_2, \\ W_{i_2j_2}^2 W_{i_3j_3}, & \text{if } j_1 = \ell_2, (j_2, j_3) \neq (i_3, i_2), \\ W_{i_2i_3}^2 W_{i_2\ell_2}, & \text{if } j_2 = i_3, j_3 = i_2, \ell_2 \neq i_3, \\ W_{i_2i_3}^2 W_{i_2j_2}, & \text{if } \ell_2 = i_3, j_3 = i_2, j_2 \neq i_3, \\ W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3}, & \text{otherwise.} \end{cases} \tag{2.E.255}$$

This allows us to further split T_{2c} into 4 different terms:

$$T_{2c} = Y_a + Y_{b1} + Y_{b2} + Y_{b3} + Y_c, \tag{2.E.256}$$

$$Y_a = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_3 \neq i_3} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^3 \widetilde{\Omega}_{i_1i_4},$$

$$Y_{b1} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_3 \neq i_3} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2j_2}^2 W_{i_3j_3} \widetilde{\Omega}_{i_1i_4},$$

$$Y_{b2} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_2 \neq i_2} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_2\ell_2} \widetilde{\Omega}_{i_1i_4},$$

$$Y_{b3} = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_1 \neq i_1} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2i_3}^2 W_{i_2j_2} \widetilde{\Omega}_{i_1i_4},$$

$$Y_c = -\frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_2,\ell_2,j_3 \\ j_2 \neq i_2, \ell_2 \neq i_2, j_3 \neq i_3 \\ j_2 \neq i_3, \ell_2 \neq i_3, j_3 \neq i_2}} \eta_{i_1}\eta_{i_3}\eta_{i_4} W_{i_2j_2}W_{i_2\ell_2}W_{i_3j_3} \widetilde{\Omega}_{i_1i_4}.$$

We now show the two claims in (2.E.176) separately. Consider the first claim. It is seen that out of the 5 terms on the right hand side of (2.E.256), the mean of all terms are 0, except for the first one. Note that for any $1 \leq i, j \leq n, i \neq j$, $\mathbb{E}[W_{ij}^3] \leq C\Omega_{ij}$. Together with $\Omega_{ij} \leq C\theta_i\theta_j$, $\widetilde{\Omega}_{ij} \leq C\alpha\theta_i\theta_j$, $0 < \eta_i < C\theta_i$ and $v \sim \|\theta\|_1^2$, it follows

$$\begin{aligned} \mathbb{E}[|Y_a|] &\leq \frac{1}{v^{3/2}} \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_3}\eta_{i_4} \Omega_{i_2i_3} \widetilde{\Omega}_{i_1i_4} \\ &\leq C\alpha \cdot \frac{1}{\|\theta\|_1^3} \sum_{i_1,i_2,i_3,i_4 (dist)} \theta_{i_1}^2 \theta_{i_2} \theta_{i_3}^2 \eta_{i_4}^2, \end{aligned}$$

where the last term is no greater than $C\alpha \cdot \|\theta\|^6/\|\theta\|_1^3$, and the first claim of (2.E.176) follows.

Consider the second claim of (2.E.176). By Cauchy-Schwartz inequality,

$$\begin{aligned} \mathrm{Var}(T_{2c}) &\leq C(\mathrm{Var}(Y_a) + \mathrm{Var}(Y_{b1}) + \mathrm{Var}(Y_{b2}) + \mathrm{Var}(Y_{b3}) + \mathrm{Var}(Y_c)) \\ &\leq C(\mathrm{Var}(Y_a) + \mathbb{E}[Y_{b1}^2] + \mathbb{E}[Y_{b2}^2] + \mathbb{E}[Y_{b3}^2] + \mathbb{E}[Y_c^2]). \end{aligned} \tag{2.E.257}$$

We now study $\mathrm{Var}(Y_a)$. Write

$$\mathrm{Var}(Y_a) = v^{-3} \sum_{\substack{i_1,i_2,i_3,i_4 (dist) \\ i_1',i_2',i_3',i_4' (dist)}} \eta_{i_1}\eta_{i_3}\eta_{i_4}\eta_{i_1'}\eta_{i_3'}\eta_{i_4'} \mathbb{E}[(W_{i_2i_3}^3 - \mathbb{E}[W_{i_2i_3}^3])(W_{i_2'i_3'}^3 - \mathbb{E}[W_{i_2'i_3'}^3])] \cdot \widetilde{\Omega}_{i_1i_4}\widetilde{\Omega}_{i_1'i_4'}.$$

Fix a term $(W_{i_2i_3}^3 - \mathbb{E}[W_{i_2i_3}^3])(W_{i_2'i_3'}^3 - \mathbb{E}[W_{i_2'i_3'}^3])$. When the mean is nonzero, we must have

$\{i_2, i_3\} = \{i_2', i_3'\}$, and when this happens,

$$\mathbb{E}[(W_{i_2i_3}^3 - \mathbb{E}[W_{i_2i_3}^3])(W_{i_2'i_3'}^3 - \mathbb{E}[W_{i_2'i_3'}^3])] = \mathrm{Var}(W_{i_2i_3}^3).$$

For a random variable X, we have $\mathrm{Var}(X) \leq \mathbb{E}[X^2]$, and it follows that

$$\mathrm{Var}(W_{i_2i_3}^3) \leq \mathbb{E}[W_{i_2i_3}^6] \leq \mathbb{E}[W_{i_2i_3}^2],$$

where we have used the property that $0 \leq W_{i_2i_3}^2 \leq 1$. Notice that $\mathbb{E}[W_{i_2i_3}^2] \leq C\theta_{i_2}\theta_{i_3}$, and recall that $v \asymp \|\theta\|_1^2$, $\widetilde{\Omega}_{ij} \leq C\alpha\theta_i\theta_j$ and $0 < \eta_i \leq C\theta_i$ for all $1 \leq i \leq n$. Combining these gives

$$\mathrm{Var}(Y_a) \leq C\alpha^2(\|\theta\|_1^{-6}) \cdot \sum_{\substack{i_1,i_2,i_3,i_4(dist) \\ i_1',i_4'(dist)}} \theta_{i_1}^2\theta_{i_2}\theta_{i_3}^3\theta_{i_4}^2\theta_{i_1'}^2\theta_{i_4'}^2 \leq C\alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5. \tag{2.E.258}$$

Additionally, note that

- The proof of Y_{b1}, Y_{b2}, and Y_{b3} is similar to that of X_{a1} in Item (e).
- The proof of Y_c is similar to that of X_b in Item (e).

For these reasons, we skip the proof details, but only to state that, both under the null and the alternative,

$$\mathbb{E}[Y_{b1}^2] \leq C\alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1, \tag{2.E.259}$$

$$\mathbb{E}[Y_{b2}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.260}$$

$$\mathbb{E}[Y_{b3}^2] \leq C\alpha^2\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.261}$$

and therefore,

$$\mathbb{E}[Y_{b1}^2] + \mathbb{E}[Y_{b2}^2] + \mathbb{E}[Y_{b3}^2] \leq C\alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1. \tag{2.E.262}$$

At the same time, both under the null and the alternative,

$$\mathbb{E}[Y_c^2] \leq C\alpha^2 \cdot \|\theta\|^{10}\|\theta\|_3^3/\|\theta\|_1^3. \tag{2.E.263}$$

Inserting (2.E.262) and (2.E.263) into (2.E.257) gives

$$\mathbb{E}[T_{2c}^2] \leq C\alpha^2[\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5 + \|\theta\|^8\|\theta\|_3^3/\|\theta\|_1 + \|\theta\|^{10}\|\theta\|_3^3/\|\theta\|_1^3] \leq C\alpha^2\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1.$$

This proves (2.E.176).

Consider Item (i). The goal is to show (2.E.178). Recall that

$$F_a = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big], \tag{2.E.264}$$

and that for any $1 \leq i \leq n$,

$$\widetilde{\eta}_i - \eta_i = v^{-1/2}\sum_{j \neq i}^{n} W_{ij}.$$

Inserting it into (2.E.264) gives

$$F_a = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_3} - \widetilde{\eta}_{i_3})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big],$$

By basic combinatorics and basic algebra, we have

$$W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_4j_4} = \begin{cases} W_{i_1i_2}^2 W_{i_3i_4}^2, & \text{if } (i_1,j_1)=(j_2,i_2), (i_3,j_3)=(j_4,i_4), \\ W_{i_1i_3}^2 W_{i_2i_4}^2, & \text{if } (i_1,j_1)=(j_3,i_3), (i_2,j_2)=(j_4,i_4), \\ W_{i_1i_4}^2 W_{i_2i_3}^2, & \text{if } (i_1,i_4)=(j_4,i_1), (i_2,j_2)=(j_3,i_3), \\ W_{i_1i_2}^2 W_{i_3j_3}W_{i_4j_4}, & \text{if } (i_1,j_1)=(j_2,i_2), (j_4,j_3)\neq(i_3,i_4), \\ W_{i_1i_3}^2 W_{i_2j_2}W_{i_4j_4}, & \text{if } (i_1,j_1)=(j_3,i_3), (j_4,j_2)\neq(i_2,i_4), \\ W_{i_1i_4}^2 W_{i_2j_2}W_{i_3j_4}, & \text{if } (i_1,j_1)=(j_4,i_4), (j_3,j_2)\neq(i_2,i_3), \\ W_{i_2i_3}^2 W_{i_1j_1}W_{i_4j_4}, & \text{if } (i_2,j_2)=(j_3,i_3), (j_4,j_1)\neq(i_1,i_4), \\ W_{i_2i_4}^2 W_{i_1j_1}W_{i_3j_3}, & \text{if } (i_2,j_2)=(j_4,i_4), (j_3,j_1)\neq(i_1,i_3), \\ W_{i_3i_4}^2 W_{i_1j_1}W_{i_2j_2}, & \text{if } (i_3,j_3)=(j_4,i_4), (j_2,j_1)\neq(i_1,i_2). \\ W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_4j_4}, & \text{otherwise.} \end{cases}$$

By symmetry, it allows us to further split F_1 into 3 different terms:

$$F_1 = 3X_a + 6X_b + X_c, \tag{2.E.265}$$

where

$$X_a = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_2}^2 W_{i_3i_4}^2,$$

$$X_b = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_3,j_4 \\ (j_3,j_4)\neq(i_4,i_3)}} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1i_2}^2 W_{i_3j_3}W_{i_4j_4},$$

and

$$X_c = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_1,j_2,j_3,j_4 \\ j_k\neq i_\ell, k,\ell=1,2,3,4}} \eta_{i_1}\eta_{i_2}\eta_{i_3}\eta_{i_4} W_{i_1j_1}W_{i_2j_2}W_{i_3j_3}W_{i_4j_4}.$$

We now show the two claims in (2.E.178) separately. Consider the first claim of (2.E.178). Note that $\mathbb{E}[X_b] = \mathbb{E}[X_c] = 0$. Recall that both under the null and the alternative, for any $i \neq j$, $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1-\Omega_{ij}) \leq C\theta_i\theta_j$, and that $0 < \eta_i \leq C\theta_i$, and that $v \asymp \|\theta\|_1^2$, Therefore,

$$0 < \mathbb{E}[X_a] \leq v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4}\theta_{i_1}\theta_{i_2}\theta_{i_3}\theta_{i_4} \leq C\|\theta\|^8/\|\theta\|_1^4.$$

Inserting into (2.E.265) gives

$$\mathbb{E}[|F_1|] \leq C\|\theta\|^8/\|\theta\|_1^4,$$

and the first claim (2.E.178) follows.

Consider the second claim (2.E.178) next. By (2.E.265) and Cauchy-Schwarz inequality,

$$\mathrm{Var}(F_1) \leq C(\mathrm{Var}(X_a) + \mathrm{Var}(X_b) + \mathrm{Var}(X_c)) \leq C(\mathrm{Var}(X_a) + \mathbb{E}[X_b^2] + \mathbb{E}[X_c^2]). \tag{2.E.266}$$

We now consider $\mathrm{Var}(X_a)$, $\mathbb{E}[X_b^2]$, and $\mathbb{E}[X_c^2]$, separately. Note that

- The proof of $\mathrm{Var}(X_a)$ is similar to that of $\mathrm{Var}(X_a)$ in Item (a).
- The proof of $\mathbb{E}[X_b^2]$ is similar to that of $\sum_{s=1}^4 \mathbb{E}[X_{bs}^2]$ in Item (a).
- The proof of $\mathbb{E}[X_c^2]$ is similar to that of $\mathbb{E}[X_c^2]$ in Item (a).

For these reasons, we omit the proof details and only state the claims. We have that under both the null and the alternative,

$$\mathrm{Var}(X_a) \le C\|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^8. \tag{2.E.267}$$

$$\mathrm{Var}(X_b^2) + \mathrm{Var}(Y_{a3}) \le C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4, \tag{2.E.268}$$

$$\mathbb{E}[X_c^2] \le C\|\theta\|_3^{12}/\|\theta\|_1^4, \tag{2.E.269}$$

Finally, inserting (2.E.267), (2.E.268), and (2.E.269) into (2.E.265) gives that, both under the null and the alternative,

$$\mathrm{Var}(F_1) \le C[\|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^8 + \|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6 + \|\theta\|_3^{12}/\|\theta\|_1^4] \le C\|\theta\|^8\|\theta\|_3^6/\|\theta\|_1^6,$$

where we have used $\|\theta\| \to \infty$ and $\|\theta\|_3^3 \ll \|\theta\|^2 \ll \|\theta\|_1$. This gives (2.E.178) and completes the proof for Item (i).

Consider Item (j). The goal is to show (2.E.179). Recall that

$$F_b = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}\eta_{i_3}^2\eta_{i_4}\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})^2(\eta_{i_2} - \widetilde{\eta}_{i_2})(\eta_{i_4} - \widetilde{\eta}_{i_4})\big],$$

and that

$$\widetilde{\eta} - \eta = v^{-1/2}W1_n.$$

Plugging this into F_b, we have

$$F_b = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,\ell_1,j_2,j_4 \\ j_1\neq i_1,\ell_1\neq i_1,j_2\neq i_2,j_4\neq i_4}} \eta_{i_2}\eta_{i_3}^2\eta_{i_4}W_{i_1j_1}W_{i_1\ell_1}W_{i_2j_2}W_{i_4j_4}.$$

By basic combinatorics and basic algebra, we have

$$W_{i_1j_1}W_{i_1\ell_1}W_{i_2j_2}W_{i_4j_4} = \begin{cases} W_{i_1i_2}^3W_{i_4j_4}, & \text{if } j_1,\ell_1 = i_2, j_2 = i_1, \\ W_{i_1i_4}^3W_{i_2j_2}, & \text{if } j_1,\ell_1 = i_4, j_4 = i_1, \\ W_{i_1i_2}^2W_{i_1i_4}^2, & \text{if } (j_1,j_2) = (i_2,i_1), (\ell_1,j_4) = (i_4,i_1), \\ W_{i_1i_2}^2W_{i_1i_4}^2, & \text{if } (\ell_1,j_2) = (i_2,i_1), (j_1,j_4) = (i_4,i_1), \\ W_{i_1i_4}^2W_{i_1i_2}^2, & \text{if } (j_1,j_4) = (i_4,i_1), (\ell_1,j_2) = (i_2,i_1), \\ W_{i_1i_4}^2W_{i_1i_2}^2, & \text{if } (\ell_1,j_4) = (i_4,i_1), (j_1,j_2) = (i_2,i_1), \\ W_{i_1j_1}^2W_{i_2i_4}^2, & \text{if } j_1 = \ell_1, (j_2,j_4) = (i_4,i_2), \\ W_{i_1i_2}^2W_{i_1j_1}W_{i_4j_4}, & \text{if } \ell_1 = i_2, j_2 = i_1, j_1 \neq i_2, i_4, \\ W_{i_1i_2}^2W_{i_1\ell_1}W_{i_4j_4}, & \text{if } j_1 = i_2, j_2 = i_1, \ell_1 \neq i_2, i_4, \\ W_{i_1i_4}^2W_{i_1j_1}W_{i_2j_2}, & \text{if } \ell_1 = i_4, j_4 = i_1, \ell_1 \neq i_2, i_4, \\ W_{i_1i_4}^2W_{i_1\ell_1}W_{i_2j_2}, & \text{if } j_1 = i_4, j_4 = i_1, j_1 \neq i_2, i_4, \\ W_{i_2i_4}^2W_{i_1j_1}W_{i_1\ell_1}, & \text{if } j_1 \neq \ell_1, (j_2,j_4) = (i_4,i_2). \\ W_{i_1j_1}^2W_{i_2j_2}W_{i_4j_4}, & \text{if } j_1 = \ell_1, (j_1,j_2) \neq (i_2,i_1), (j_1,j_4) \neq (i_4,i_1), \\ W_{i_1j_1}W_{i_1\ell_1}W_{i_2j_2}W_{i_4j_4}, & \text{otherwise.} \end{cases}$$

By these and symmetry, we can further split F_b into 7 different terms,

We decompose

$$F_b = 2Y_{a1} + 4Y_{a2} + Y_{a3} + 4Y_{b1} + Y_{b2} + Y_{b3} + Y_c, \tag{2.E.270}$$

where

$$Y_{a1} = v^{-2}\sum_{i_1,i_2,i_3,i_4(dist)}\sum_{j_4,j_4\neq i_4}\eta_{i_2}\eta_{i_3}^2\eta_{i_4}W_{i_1i_2}^3W_{i_4j_4},$$

$$Y_{a2} = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_1 i_2}^2 W_{i_1 i_4}^2,$$

$$Y_{a3} = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_1, j_1 \neq i_1} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_1 j_1}^2 W_{i_2 i_4}^2,$$

$$Y_{b1} = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_1, j_4 \\ j_1 \neq i_1, j_4 \neq i_4}} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_1 i_2}^2 W_{i_1 j_1} W_{i_4 j_4},$$

$$Y_{b2} = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_1, \ell_1 \\ j_1, \ell_1 \neq i_1}} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_2 i_4}^2 W_{i_1 j_1} W_{i_1 \ell_1},$$

$$Y_{b3} = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_1, j_2, j_4 \\ j_1 \neq i_1, j_2 \neq i_2, j_4 \neq i_4}} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_1 j_1}^2 W_{i_2 j_2} W_{i_4 j_4},$$

$$Y_c = v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{\substack{j_1, \ell_1, j_2, j_4 \\ j_1, \ell_1 \notin \{i_1, i_2, i_4\} \\ j_2 \notin \{i_1, i_4\}, j_4 \notin \{i_1, i_2\}}} \eta_{i_2}\eta_{i_3}^2\eta_{i_4} W_{i_1 j_1} W_{i_1 \ell_1} W_{i_2 j_2} W_{i_4 j_4},$$

We now consider the two claims in (2.E.179) separately. Consider the first claim. It is seen that only the second and the third terms above have non-zero mean. Recall that both under the null and the alternative, for any $i \neq j$, $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1 - \Omega_{ij}) \leq C\theta_i\theta_j$, $0 < \eta_i \leq C\theta_i$, and that $v \asymp \|\theta\|_1^2$. It follows

$$0 < \mathbb{E}[Y_{a2}] \leq v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \theta_{i_2}\theta_{i_3}^2\theta_{i_4} \cdot \theta_{i_1}^2\theta_{i_2}\theta_{i_4} \leq C\|\theta\|^8/\|\theta\|_1^4. \tag{2.E.271}$$

and

$$0 < \mathbb{E}[Y_{a3}] \leq v^{-2} \sum_{i_1,i_2,i_3,i_4 (dist)} \sum_{j_1} \theta_{i_2}\theta_{i_3}^2\theta_{i_4} \cdot \theta_{i_1}\theta_{i_2}\theta_{j_1}\theta_{i_4} \leq C\|\theta\|^6/\|\theta\|_1^2. \tag{2.E.272}$$

Combining (2.E.271), (2.E.272) with (2.E.270) gives

$$\mathbb{E}[|F_2|] \leq C[\|\theta\|^8/\|\theta\|_1^4 + \|\theta\|^6/\|\theta\|_1^2] \leq C\|\theta\|^6/\|\theta\|_1^2,$$

where we've used the universal inequality that $\|\theta\|^2 \leq \|\theta\|_1$. It follows the first claim of (2.E.179).

We now show the second claim of (2.E.179). By Cauchy-Schwarz inequality,

$$\begin{aligned} \mathrm{Var}(F_b) &\leq C\big(\mathrm{Var}(Y_{a1}) + \mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) + \mathrm{Var}(Y_{b1}) + \mathrm{Var}(Y_{b2}) + \mathrm{Var}(Y_{b3}) + \mathrm{Var}(Y_c)\big) \\ &\leq C\big(\mathrm{Var}(Y_{a1}) + \mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) + \mathbb{E}[Y_{b1}^2] + \mathbb{E}[Y_{b2}^2] + \mathbb{E}[Y_{b3}^2] + \mathbb{E}[Y_c^2]\big). \end{aligned} \tag{2.E.273}$$

We now consider $\mathrm{Var}(Y_{a1})$, $\mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3})$, $\mathbb{E}[Y_{b1}^2] + \mathbb{E}[Y_{b2}^2] + \mathbb{E}[Y_{b3}^2]$, and $\mathbb{E}[Y_c^2]$, separately. Note that

- The proof of $\mathrm{Var}(Y_{a1})$ is similar to that of $\mathrm{Var}(Y_a)$ in Item (b).
- The proof of $\mathrm{Var}(Y_{a2})$ and $\mathrm{Var}(Y_{a3})$ are similar to that of $\mathrm{Var}(X_a)$ in Item (a).
- The proof of $\sum_{s=1}^3 \mathbb{E}[Y_{b_s}^2]$ is similar to that of $\sum_{s=1}^4 \mathbb{E}[X_{bs}^2]$ in Item (a).

- The proof of $\mathbb{E}[Y_c^2]$ is similar to that of $\mathbb{E}[X_c^2]$ in Item (a).

For these reasons, we omit the proof details and only state the claims. We have that under both the null and the alternative,

$$\mathrm{Var}(Y_{a1}) \le C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5. \tag{2.E.274}$$

$$\mathrm{Var}(Y_{a2}) + \mathrm{Var}(Y_{a3}) \le C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4, \tag{2.E.275}$$

$$\sum_{s=1}^{3}\mathbb{E}[Y_{b_s}^2] \le C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2, \tag{2.E.276}$$

$$\mathbb{E}[Y_c^2] \le C\|\theta\|^6\|\theta\|_3^6/\|\theta\|_1^4. \tag{2.E.277}$$

Finally, inserting (2.E.274), (2.E.275), (2.E.276), and (2.E.277) into (2.E.273) gives

$$\begin{aligned}\mathrm{Var}(F_2) &\le C[\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4 + \|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^2 + \|\theta\|^6\|\theta\|_3^6/\|\theta\|_1^4] \\ &\le C\|\theta\|^4\|\theta\|_3^6/\|\theta\|_1^4, \end{aligned} \tag{2.E.278}$$

where we have used $\|\theta\|_3^3 \ll \|\theta\|^2 \ll \|\theta\|_1$, $\|\theta\| \to \infty$ and $\|\theta\|^4 \le \|\theta\|_1\|\theta\|_3^3$. This completes the proof of (2.E.179).

Consider Item (k). The goal is to show (2.E.180) and (2.E.181). Recall that

$$F_c = \sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}^2\eta_{i_4}^2\big[(\eta_{i_1} - \widetilde{\eta}_{i_1})^2(\eta_{i_3} - \widetilde{\eta}_{i_3})^2\big],$$

and that $\widetilde{\eta} - \eta = v^{-1/2}W1_n$. Plugging this into F_3 gives

$$F_c = v^{-2}\sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1,\ell_1,j_2,j_4 \\ j_1\neq i_1,\ell_1\neq i_1,j_3\neq i_3,\ell_3\neq i_3}} \eta_{i_2}^2\eta_{i_4}^2 W_{i_1j_1}W_{i_1\ell_1}W_{i_3j_3}W_{i_3\ell_3}.$$

By basic combinatorics and basic algebra, we have

$$W_{i_1j_1}W_{i_1\ell_1}W_{i_3j_3}W_{i_3\ell_3} = \begin{cases} W_{i_1i_3}^4, & \text{if } j_1=\ell_1=i_1, j_3=\ell_3=i_1, \\ W_{i_1i_3}^3W_{i_1j_1}, & \text{if } j_3=\ell_3=i_1, \ell_1=i_3, \\ W_{i_1i_3}^3W_{i_1\ell_1}, & \text{if } j_3=\ell_3=i_1, j_1=i_3, \\ W_{i_1i_3}^3W_{i_3j_3}, & \text{if } j_1=\ell_1=i_3, \ell_3=i_1, \\ W_{i_1i_3}^3W_{i_3\ell_3}, & \text{if } j_1=\ell_1=i_3, j_3=i_1, \\ W_{i_1j_1}^2W_{i_3j_3}^2, & \text{if } j_1=\ell_1, j_3=\ell_3, \\ W_{i_1j_1}^2W_{i_3j_3}W_{i_3\ell_3}, & \text{if } j_1=\ell_1\neq i_3, j_3\neq \ell_3, \\ W_{i_3j_3}^2W_{i_1j_1}W_{i_1\ell_1}, & \text{if } j_3=\ell_3\neq i_1, j_1\neq \ell_1, \\ W_{i_1i_3}^2W_{i_1\ell_1}W_{i_3\ell_3}, & \text{if } j_1=i_3, j_3=i_1, \\ W_{i_1i_3}^2W_{i_1j_1}W_{i_3j_3}, & \text{if } \ell_1=i_3, \ell_3=i_1, \\ W_{i_1i_3}^2W_{i_1j_1}W_{i_3\ell_3}, & \text{if } \ell_1=i_3, j_3=i_1, \\ W_{i_1i_3}^2W_{i_1\ell_1}W_{i_3j_3}, & \text{if } j_1=i_3, \ell_3=i_1, \\ W_{i_1j_1}W_{i_1\ell_1}W_{i_3j_3}W_{i_3\ell_3}, & \text{otherwise.} \end{cases}$$

By these and symmetry, we can further split F_3 into 6 different terms:

$$F_c = Z_a + 4Z_{b1} + Z_{b2} + 2Z_{c1} + 4Z_{c2} + Z_d, \tag{2.E.279}$$

where

$$Z_a = v^{-2}\sum_{i_1,i_2,i_3,i_4(dist)} \eta_{i_2}^2\eta_{i_4}^2W_{i_1i_3}^4,$$

$$Z_{b1} = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_4, j_4 \neq i_4} \eta_{i_2}^2 \eta_{i_4}^2 W_{i_1 i_3}^3 W_{i_3 j_3},$$

$$Z_{b2} = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1, j_1 \neq i_1, j_3, j_3 \neq i_3} \eta_{i_2}^2 \eta_{i_4}^2 W_{i_1 j_1}^2 W_{i_3 j_3}^2,$$

$$Z_{c1} = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1, j_3, \ell_3 \\ j_1 \notin \{i_1, i_3\}, j_3, \ell_3}} \eta_{i_2}^2 \eta_{i_4}^2 W_{i_1 j_1}^2 W_{i_3 j_3} W_{i_3 \ell_3},$$

$$Z_{c2} = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{\ell_1, \ell_3 \\ \ell_1 \neq i_1, \ell_3 \neq i_3}} \eta_{i_2}^2 \eta_{i_4}^2 W_{i_1 i_3}^2 W_{i_1 \ell_1} W_{i_3 \ell_3},$$

$$Z_d = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{\substack{j_1, \ell_1, j_3, \ell_3 \\ j_1 \neq \ell_1, j_3 \neq \ell_3 \\ j_1, \ell_1 \neq i_3, j_3, \ell_3 \neq i_1}} \eta_{i_2}^2 \eta_{i_4}^2 W_{i_1 j_1} W_{i_1 \ell_1} W_{i_3 j_3} W_{i_3 \ell_3}.$$

We now show (2.E.180) and (2.E.181) separately. Consider (2.E.180) first. It is among all the 6 Z-terms, only Z_a and Z_{b2} have non-zero means. We now consider $\mathbb{E}[Z_a]$ and $\mathbb{E}[Z_{b2}]$ separately.

First, consider $\mathbb{E}[Z_a]$. By similar arguments, both under the null and the alternative,

$$\mathbb{E}[W_{i_1 i_3}^4] \leq C\Omega_{i_1 i_3} \leq C\theta_{i_1}\theta_{i_3}.$$

Recalling that $0 < \eta_i \leq C\theta_i$ and $v \asymp \|\theta\|^2$, it is seen that

$$\mathbb{E}[Z_a] \leq C(\|\theta\|_1)^{-4} \sum_{i_1,i_2,i_3,i_4(dist)} \theta_{i_2}^2 \theta_{i_4}^2 \theta_{i_1} \theta_{i_3} \leq C\|\theta\|^4/\|\theta\|_1^2. \tag{2.E.280}$$

Next, consider $\mathbb{E}[Z_{b2}]$. First, recall that under the null, $\Omega = \theta\theta'$, $v = 1_n'(\Omega - \mathrm{d}iag(\Omega))1_n$, and $\eta = v^{-1/2}(\Omega - \mathrm{d}iag(\Omega)1_n$. It is seen $v \sim \|\theta\|_1^2$, $\eta_i = (1 + o(1)\theta_i$, $1 \leq i \leq n$, where $o(1) \to 0$ uniformly for all $1 \leq i \leq n$, and for any $i \neq j$, $\mathbb{E}[W_{ij}^2] = (1 + o(1))\theta_i\theta_j$, where $o(1) \to 0$ uniformly for all $1 \leq i, j \leq n$. It follows

$$\mathbb{E}[Z_{b2}] = v^{-2} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1, j_1 \neq i_1, j_3, j_3 \neq i_3} \eta_{i_2}^2 \eta_{i_4}^2 \mathbb{E}[W_{i_1 j_1}^2 W_{i_3 j_3}^2], \tag{2.E.281}$$

which

$$\sim (\|\theta\|_1)^{-4} \sum_{i_1,i_2,i_3,i_4(dist)} \sum_{j_1, j_1 \neq i_1, j_3, j_3 \neq i_3} \theta_{i_1} \theta_{i_2}^2 \theta_{i_3} \theta_{i_4}^2 \theta_{j_1} \theta_{j_3} \sim \|\theta\|^4.$$

Second, under the alternative, by similar argument, we have that $v \asymp \|\theta\|_1^2$, $0 < \eta_i < C\theta_i$ for all $1 \leq i \leq n$, and $\mathbb{E}[W_{ij}^2] \leq C\theta_i\theta_j$ for all $1 \leq i, j \leq n$, $i \neq j$. Similar to that under the null, we have

$$0 < |\mathbb{E}[Z_{b2}]| \leq C\|\theta\|^4. \tag{2.E.282}$$

Inserting (2.E.280), (2.E.281), and (2.E.282) into (2.E.279) and recalling that the mean of all other Z terms are 0,

$$\mathbb{E}[F_3] \sim \|\theta\|^4, \qquad \text{under the null,}$$

and

$$\mathbb{E}[F_3] \leq C\|\theta\|^4, \qquad \text{under the alternative,}$$

where we have used $\|\theta\|_1 \to \infty$. This proves (2.E.180).

We now consider (2.E.181). By Cauchy-Schwarz inequality,

$$\begin{aligned}\mathrm{Var}(F_c) &\le C\big(\mathrm{Var}(Z_a) + \mathrm{Var}(Z_{b1}) + \mathrm{Var}(Z_{b2}) + \mathrm{Var}(Z_{c1}) + \mathrm{Var}(Z_{c2}) + \mathrm{Var}(Z_d)\big) \\ &\le C\big(\mathrm{Var}(Z_a) + \mathbb{E}[Z_{b1}^2] + \mathrm{Var}(Z_{b2}) + \mathbb{E}[Z_{c1}^2] + \mathbb{E}[Z_{c2}^2] + \mathbb{E}[Z_d^2]\big). \end{aligned} \tag{2.E.283}$$

Consider $\mathrm{Var}(Z_a)$. Write

$$\mathrm{Var}(Z_a) = v^{-4} \sum_{\substack{i_1,i_2,i_3,i_4 (dist)\\ i_1',i_2',i_3',i_4'(dist)}} \eta_{i_2}^2\eta_{i_4}^2\eta_{i_2'}^2\eta_{i_4'}^2 \mathbb{E}[(W_{i_1i_3}^4 - \mathbb{E}[W_{i_1i_3}^4])(W_{i_1'i_3'}^4 - \mathbb{E}[W_{i_1'i_3'}^4])].$$

Fix a term $(W_{i_1i_3}^4 - \mathbb{E}[W_{i_1i_3}^4])(W_{i_1'i_3'}^4 - \mathbb{E}[W_{i_1'i_3'}^4])$. When the mean is nonzero, we must have $\{i_1, i_3\} = \{i_1', i_3'\}$, and when this happens,

$$\mathbb{E}[(W_{i_1i_3}^4 - \mathbb{E}[W_{i_1i_3}^4])(W_{i_1'i_3'}^4 - \mathbb{E}[W_{i_1'i_3'}^4])] = \mathrm{Var}(W_{i_1i_3}^4).$$

For a random variable X, we have $\mathrm{Var}(X) \le \mathbb{E}[X^2]$, and it follows that

$$\mathrm{Var}(W_{i_1i_3}^4) \le \mathbb{E}[W_{i_1i_3}^8] \le \mathbb{E}[W_{i_1i_3}^2],$$

where we have used the property that $0 \le W_{i_1i_3}^2 \le 1$; note that $\mathbb{E}[W_{i_1i_3}^2] \le C\theta_{i_1}\theta_{i_3}$. Recall that $v \asymp \|\theta\|_1^2$ and $0 < \eta_i \le C\theta_i$ for all $1 \le i \le n$. Combining these gives

$$\mathrm{Var}(Z_a) \le C(\|\theta\|_1^{-8}) \cdot \sum_{\substack{i_1,i_2,i_3,i_4 (dist)\\ i_2',i_4'(dist)}} \theta_{i_2}^2\theta_{i_4}^2\theta_{i_2'}^2\theta_{i_4'}^2\theta_{i_1}\theta_{i_3} \le C\|\theta\|^8/\|\theta\|_1^6. \tag{2.E.284}$$

We now consider all other terms on the right hand side of (2.E.283). Note that

- The proof of $\mathbb{E}[Z_{b1}^2]$ is similar to that of Y_{a1} in Item (b).
- The proof of $\mathrm{Var}(Z_{b2})$ is similar to that of X_a in Item (a).
- The proof of $\mathbb{E}[Z_{c1}^2]$ and $\mathbb{E}[Z_{c2}^2]$ are similar to that of X_b in Item (a).
- The proof of $\mathbb{E}[Z_d^2]$ is similar to that of X_c in Item (a).

For these reasons, we skip the proof details. We have that, under both the null and the alternative,

$$\mathbb{E}[Z_{b1}^2] \le C\|\theta\|^8\|\theta\|_3^3/\|\theta\|_1^5, \tag{2.E.285}$$

$$\mathrm{Var}(Z_{b2}) \le C\|\theta\|^8/\|\theta\|_1^2, \tag{2.E.286}$$

$$\mathbb{E}[Z_{c1}^2] + \mathbb{E}[Z_{c2}^2] \le C\|\theta\|^{10}/\|\theta\|_1^2, \tag{2.E.287}$$

and

$$\mathbb{E}[Z_d^2] \le C\|\theta\|^{12}/\|\theta\|_1^4. \tag{2.E.288}$$

Inserting (2.E.284), (2.E.285), (2.E.286), (2.E.287) and (2.E.288) into (2.E.283) gives

$$\begin{aligned}\mathrm{Var}(F_c) &\le C[\|\theta\|^8/\|\theta\|_1^6 + \|\theta\|^8/\|\theta\|_1^2 + \|\theta\|^{10}/\|\theta\|_1^2 + \|\theta\|^{12}/\|\theta\|_1^4] \\ &\le C\|\theta\|^{10}/\|\theta\|_1^2,\end{aligned}$$

which completes the proof of (2.E.181).

Proof of Lemma 24

Define an event D as

$$D = \big\{|V - v| \leq \|\theta\|_1 \cdot x_n\big\}, \qquad \text{for} \quad \sqrt{\log(\|\theta\|_1)} \ll x_n \ll \|\theta\|_1.$$

We aim to show that

$$\mathbb{E}[(Q_n - Q_n^*)^2 \cdot I_{D^c}] = o(\|\theta\|^8). \tag{2.E.289}$$

First, we bound the tail probability of $|V - v|$. Write

$$V - v = 2\sum_{i<j}(A_{ij} - \Omega_{ij}).$$

The variables $\{A_{ij} - \Omega_{ij}\}_{1\leq i<j\leq n}$ are mutually independent with mean zero. They satisfy $|A_{ij} - \Omega_{ij}| \leq 1$ and $\sum_{i<j} \mathrm{Var}(A_{ij} - \Omega_{ij}) \leq \sum_{i<j} \Omega_{ij} \leq 1_n'\Omega 1_n/2 \leq \|\theta\|_1^2/2$. Applying the Bernstein's inequality, for any $t > 0$,

$$\mathbb{P}\Big(\Big|2\sum_{i<j}(A_{ij} - \Omega_{ij})\Big| > t\Big) \leq 2\exp\Big(-\frac{t^2/2}{2\|\theta\|_1^2 + t/3}\Big).$$

We immediately have that, for some positive constants $C_1, C_2 > 0$,

$$\mathbb{P}(|V - v| > t) \leq \begin{cases} 2\exp\big(-\frac{C_1}{\|\theta\|_1^2}t^2\big), & \text{when } x_n\|\theta\|_1 \leq t \leq \|\theta\|_1^2, \\ 2\exp\big(-C_2 t\big), & \text{when } t > \|\theta\|_1^2. \end{cases} \tag{2.E.290}$$

Especially, letting $t = x_n\|\theta\|_1$, we have

$$\mathbb{P}(D^c) \leq 2\exp(-C_1 x_n^2). \tag{2.E.291}$$

Next, we derive an upper bound of $(Q_n - Q_n^*)^2$ in terms of V. Recall that V is the total number of edges and that $Q_n = \sum_{i,j,k,\ell(dist)} M_{ij}M_{jk}M_{k\ell}M_{\ell i}$, where $M_{ij} = A_{ij} - \widehat{\eta}_i\widehat{\eta}_j$. If one node of i, j, k, ℓ has a zero degree (say, node i), then $A_{ij} = 0$ and $\widehat{\eta}_i = 0$, and it follows that $M_{ij} = 0$ and $M_{ij}M_{jk}M_{k\ell}M_{\ell i} = 0$. Hence, only when (i, j, k, ℓ) all have nonzero degrees, this quadruple has a contribution to Q_n. Since V is the total number of edges, there are at most V nodes that have a nonzero degree. It follows that

$$|Q_n| \leq CV^4.$$

Moreover, $Q_n^* = \sum_{i,j,k,\ell(dist)} M_{ij}^*M_{jk}^*M_{k\ell}^*M_{\ell i}^*$, where $M_{ij}^* = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}$. Re-write $M_{ij}^* = A_{ij} - \eta_i^*\eta_j^* + \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_j - \widetilde{\eta}_j)$. First, since $\eta_i^* \leq C\theta_i$ and $\eta_i \leq C\theta_i$ (see (2.E.117)), $|M_{ij}^*| \leq A_{ij} + C\theta_i\theta_j + C\theta_i|\eta_j - \widetilde{\eta}_j| + C\theta_j|\eta_i - \widetilde{\eta}_i|$. Second, note that $\widetilde{\eta}_i$ equals to $v^{-1/2}$ times degree of node i, where $v \asymp \|\theta\|_1^2$ according to (2.E.116). It follows that $|\eta_i - \widetilde{\eta}_i| \leq C(\theta_i + \|\theta\|_1^{-1}V)$. Therefore,

$$|M_{ij}^*| \leq A_{ij} + C\theta_i\theta_j + C\|\theta\|_1^{-1}V(\theta_i + \theta_j).$$

We plug it into the definition of Q_n^* and note that there are at most V pairs of (i, j) such that $A_{ij} \neq 0$. By elementary calculation,

$$|Q_n^*| \leq C(V^4 + \|\theta\|_1^4).$$

Combining the above gives

$$(Q_n - Q_n^*)^2 \leq 2Q_n^2 + 2(Q_n^*)^2 \leq C(V^8 + \|\theta\|_1^8). \tag{2.E.292}$$

Last, we show (2.E.289). By (2.E.292) and that $V^8 \leq Cv^8 + C|V-v|^8$, we have

$$\begin{aligned}\mathbb{E}[(Q_n - Q_n^*)^2 \cdot I_{D^c}] &\leq C\mathbb{E}[|V-v|^8 \cdot I_{D^c}] + C(v^8 + \|\theta\|_1^8)\cdot \mathbb{P}(D^c) \\ &\leq C\mathbb{E}[|V-v|^8 \cdot I_{D^c}] + C\|\theta\|_1^{16} \cdot \mathbb{P}(D^c), \qquad (2.E.293)\end{aligned}$$

where the second line is from $v \asymp \|\theta\|_1^2$. Note that $x_n \gg \sqrt{\log(\|\theta\|_1)}$. For n sufficiently large, $x_n^2 \geq 17C_1^{-1}\log(\|\theta\|_1)$. Combining it with (2.E.291), we have

$$\|\theta\|_1^{16} \cdot \mathbb{P}(D^c) \leq \|\theta\|_1^{16} \cdot 2e^{-C_1 x_n^2} \leq \|\theta\|_1^{16} \cdot 2e^{-17\|\theta\|_1} = o(1). \qquad (2.E.294)$$

We then bound $\mathbb{E}[|V-v|^8 \cdot I_{D^c}]$. Let $f(t)$ and $F(t)$ be the probability density and CDF of $|V-v|$, and write $\bar{F}(t) = 1 - F(t)$. Using integration by part, for any continuously differentiable function $g(t)$ and $x > 0$, $\int_x^\infty g(t)f(t)dt = g(x)\bar{F}(x) + \int_x^\infty g'(t)\bar{F}(t)dt$. We apply the formula to $g(t) = t^8$ and $x = x_n\|\theta\|_1$. It yields

$$\begin{aligned}\mathbb{E}[|V-v|^8 \cdot I_{D^c}] &= (x_n\|\theta\|_1)^8 \cdot \mathbb{P}(D^c) + \int_{x_n\|\theta\|_1}^\infty 8t^7 \cdot \mathbb{P}(|V-v| > t)dt \\ &\equiv I + II.\end{aligned}$$

Consider I. By (2.E.294) and $x_n \ll \|\theta\|_1$,

$$I \ll \|\theta\|_1^{16} \cdot \mathbb{P}(D^c) = o(1).$$

Consider II. By (2.E.290), (2.E.294), and elementary probability,

$$\begin{aligned}II &\leq 8(\|\theta\|_1^2)^7 \cdot \mathbb{P}\big(x_n\|\theta\|_1 < |V-v| \leq \|\theta\|_1^2\big) + \int_{\|\theta\|_1^2} 8t^7 \cdot \mathbb{P}(|V-v| > t)dt \\ &\leq C\|\theta\|_1^{14} \cdot \mathbb{P}(D^c) + \int_{\|\theta\|_1^2} 8t^7 \cdot 2e^{-C_2 t}dt \\ &= o(1),\end{aligned}$$

where in the last line we have used (2.E.294) and the fact that $\int_x^\infty t^7 e^{-C_2 t}dt \to 0$ as $x \to \infty$. Combining the bounds for I and II gives

$$\mathbb{E}[|V-v|^8 \cdot I_{D^c}] = o(1). \qquad (2.E.295)$$

Then, (2.E.289) follows by plugging (2.E.294)-(2.E.295) into (2.E.293).

Proof of Lemma 25

There are 175 post-expansion sums in $(\widetilde{Q}_n^* - Q_n^*)$. They divide into 34 different types, denoted by R_1-R_{34} as shown in Table 2.3. It suffices to prove that, for each $1 \leq k \leq 34$, under the null hypothesis,

$$\big|\mathbb{E}[R_k]\big| = o(\|\theta\|^4), \qquad \mathrm{Var}(R_k) = o(\|\theta\|^8), \qquad (2.E.296)$$

and under the alternative hypothesis,

$$\big|\mathbb{E}[R_k]\big| = o(\alpha^4\|\theta\|^8), \qquad \mathrm{Var}(R_k) = O(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6). \qquad (2.E.297)$$

We need some preparation. First, recall that $\widetilde{r}_{ij} = -\frac{v}{V}(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)$. It follows that each post-expansion sum has the form

$$\left(\frac{v}{V}\right)^{N_{\widetilde{r}}} \sum_{i,j,k,\ell(dist)} a_{ij}b_{jk}c_{k\ell}d_{\ell i}, \qquad (2.E.298)$$

Table 2.3: The 34 types of the 175 post-expansion sums for $(\widetilde{Q}_n^* - Q_n^*)$.

Notation	#	$N_{\widetilde{r}}$	$(N_\delta, N_{\widetilde{\Omega}}, N_W)$	Examples	N_W^*
R_1	4	1	(0, 0, 3)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} W_{jk} W_{k\ell} W_{\ell i}$	5
R_2	8	1	(0, 1, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} W_{k\ell} W_{\ell i}$	4
R_3	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	4
R_4	8	1	(0, 2, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	3
R_5	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} W_{k\ell} \widetilde{\Omega}_{\ell i}$	3
R_6	4	1	(0, 3, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	2
R_7	8	1	(1, 0, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} W_{k\ell} W_{\ell i}$	5
R_8	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} W_{jk} \delta_{k\ell} W_{\ell i}$	5
R_9	8	1	(1, 1, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	4
R_{10}	8			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} W_{k\ell} \delta_{\ell i}$	4
R_{11}	8			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} W_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	4
R_{12}	8	1	(1, 2, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	3
R_{13}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	3
R_{14}	8	1	(2, 0, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \delta_{k\ell} W_{\ell i}$	5
R_{15}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} W_{k\ell} \delta_{\ell i}$	5
R_{16}	8	1	(2, 1, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	4
R_{17}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} \delta_{\ell i}$	4
R_{18}	4	1	(3, 0, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \delta_{k\ell} \delta_{\ell i}$	5
R_{19}	4	2	(0, 0, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} W_{k\ell} W_{\ell i}$	6
R_{20}	2			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} W_{jk} \widetilde{r}_{k\ell} W_{\ell i}$	6
R_{21}	4	2	(0, 2, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	4
R_{22}	2			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} \widetilde{r}_{k\ell} \widetilde{\Omega}_{\ell i}$	4
R_{23}	4	2	(2, 0, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \delta_{k\ell} \delta_{\ell i}$	6
R_{24}	2			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{r}_{k\ell} \delta_{\ell i}$	6
R_{25}	8	2	(0, 1, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	5
R_{26}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{\Omega}_{jk} \widetilde{r}_{k\ell} W_{\ell i}$	5
R_{27}	8	2	(1, 1, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	5
R_{28}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{r}_{k\ell} \widetilde{\Omega}_{\ell i}$	5
R_{29}	8	2	(1, 0, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \delta_{k\ell} W_{\ell i}$	6
R_{30}	4			$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \delta_{jk} \widetilde{r}_{k\ell} W_{\ell i}$	6
R_{31}	4	3	(0, 0, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{r}_{k\ell} W_{\ell i}$	7
R_{32}	4	3	(0, 1, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{r}_{k\ell} \widetilde{\Omega}_{\ell i}$	6
R_{33}	4	3	(1, 0, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{r}_{k\ell} \delta_{\ell i}$	7
R_{34}	1	4	(0, 0, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{r}_{ij} \widetilde{r}_{jk} \widetilde{r}_{k\ell} \widetilde{r}_{\ell i}$	8

where a_{ij} takes values in $\{\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, -(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)\}$ and $b_{jk}, c_{k\ell}, d_{\ell i}$ are similar. The variable $\frac{v}{V}$ has a complicated correlation with each summand, so we want to get rid of it. Denote the variable in (2.E.298) by Y. Write $m = N_{\widetilde{r}}$ and

$$Y = \left(\frac{v}{V}\right)^m X, \qquad \text{where} \quad X = \sum_{i,j,k,\ell(dist)} a_{ij} b_{jk} c_{k\ell} d_{\ell i}. \tag{2.E.299}$$

We compare the mean and variance of X and Y. By assumption, $\sqrt{\log(\|\theta\|_1)} \ll \|\theta\|_1/\|\theta\|^2$. Then, there exists a sequence x_n such that

$$\sqrt{\log(\|\theta\|_1)} \ll x_n \ll \|\theta\|_1/\|\theta\|^2, \qquad \text{as } n \to \infty.$$

We introduce an event

$$D = \big\{|V - v| \leq \|\theta\|_1 x_n\big\}.$$

In Lemma 24, we have proved $\mathbb{E}[(Q_n - Q_n^*)^2 \cdot I_{D^c}] = o(1)$. By similar proof, we can show: as long as $|Y - X|$ is bounded by a polynomial of V and $\|\theta\|_1$,

$$\mathbb{E}[(Y - X)^2 \cdot I_{D^c}] = o(1). \tag{2.E.300}$$

Additionally, on the event D, since $v \asymp \|\theta\|_1^2 \gg \|\theta\|_1 x_n$, we have $|V - v| = o(v)$. It follows that $\frac{|V-v|}{V} \lesssim \frac{|V-v|}{v} \leq C\|\theta\|^{-1} x_n = o(1)$. For any fixed $m \geq 1$, $(1+x)^m \leq 1 + Cx$ for x being close to 0. Hence, $|1 - \frac{v^m}{V^m}| \leq C|1 - \frac{v}{V}| \leq C\|\theta\|_1^{-1} x_n = o(\|\theta\|^{-2})$. It implies

$$|Y - X| = o(\|\theta\|^{-2}) \cdot |X|, \qquad \text{on the event } D. \tag{2.E.301}$$

By (2.E.300)-(2.E.301) and elementary probability,

$$\begin{aligned}
|\mathbb{E}[Y - X]| &\leq |\mathbb{E}[(Y - X) \cdot I_D]| + |\mathbb{E}[(Y - X) \cdot I_{D^c}]| \\
&\leq o(\|\theta\|^{-2}) \cdot \mathbb{E}[|X| \cdot I_D] + \sqrt{\mathbb{E}[(Y - X)^2 \cdot I_{D^c}]} \\
&\leq o(\|\theta\|^{-2})\sqrt{\mathbb{E}[X^2]} + o(1),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(Y) &\leq 2\mathrm{Var}(X) + 2\mathrm{Var}(Y - X) \\
&\leq 2\mathrm{Var}(X) + 2\mathbb{E}[(Y - X)^2] \\
&= 2\mathrm{Var}(X) + 2\mathbb{E}[(Y - X)^2 \cdot I_D] + 2\mathbb{E}[(Y - X)^2 \cdot I_{D^c}] \\
&\leq 2\mathrm{Var}(X) + o(\|\theta\|^{-4}) \cdot \mathbb{E}[X^2] + o(1).
\end{aligned}$$

Under the null hypothesis, suppose we can prove that

$$\mathbb{E}[X^2] = o(\|\theta\|^8). \tag{2.E.302}$$

Since $\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \mathrm{Var}(X)$, it implies $|\mathbb{E}[X]| = o(\|\theta\|^4)$ and $\mathrm{Var}(X) = o(\|\theta\|^8)$. Therefore,

$$\begin{aligned}
&|\mathbb{E}[Y]| \leq |\mathbb{E}[X]| + |\mathbb{E}[Y - X]| = o(\|\theta\|^4), \\
&\mathrm{Var}(Y) \leq C\mathrm{Var}(X) + o(\|\theta\|^{-4}) \cdot \mathbb{E}[X^2] + o(1) = o(\|\theta\|^8).
\end{aligned}$$

Under the alternative hypothesis, suppose we can prove that

$$|\mathbb{E}[X]| = O(\alpha^2\|\theta\|^6), \qquad \mathrm{Var}(X) = o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6). \tag{2.E.303}$$

Since $\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \mathrm{Var}(X)$, we have $\mathbb{E}[X^2] = O(\alpha^4\|\theta\|^{12})$. Then,

$$\begin{aligned}
&|\mathbb{E}[Y]| \leq O(\alpha^2\|\theta\|^6) + o(\|\theta\|^{-2}) \cdot O(\alpha^2\|\theta\|^6) = o(\alpha^4\|\theta\|^8), \\
&\mathrm{Var}(Y) \leq o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6) + o(\|\theta\|^{-4}) \cdot O(\alpha^4\|\theta\|^{12}) = o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6).
\end{aligned}$$

In conclusion, to prove that Y satisfies the requirement in (2.E.296)-(2.E.297), it is sufficient to prove that X satisfies (2.E.302)-(2.E.303). We remark that (2.E.303) puts a more stringent requirement on the mean of the variable, compared to (2.E.297).

From now on, in the analysis of each R_k of the form (2.E.298), we shall always neglect the factor $(\frac{v}{V})^{N_{\tilde{r}}}$, and show that, after this factor is removed, the random variable satisfies

(2.E.302)-(2.E.303). This is equivalent to pretending

$$\widetilde{r}_{ij} = -(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)$$

and proving each R_k satisfies (2.E.302)-(2.E.303). Unless mentioned, we stick to this mis-use of notation $\widetilde{r}_{ij}$ in the proof below.

Second, we divide 34 terms into several groups using the *intrinsic order of* W defined below. Note that $\widetilde{r}_{ij} = -(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)$, $\delta_{ij} = \eta_i(\eta_j - \widetilde{\eta}_j) + \eta_j(\eta_i - \widetilde{\eta}_i)$, and $\widetilde{\eta}_i - \eta_i = \frac{1}{\sqrt{v}} \sum_{s \neq i} W_{is}$. We thus have

$$\widetilde{r}_{ij} = -\frac{1}{v}\Big(\sum_{s\neq i} W_{is}\Big)\Big(\sum_{t\neq j} W_{jt}\Big), \qquad \delta_{ij} = -\frac{1}{\sqrt{v}}\eta_i\Big(\sum_{t\neq j} W_{jt}\Big) - \frac{1}{\sqrt{v}}\eta_j\Big(\sum_{s\neq i} W_{is}\Big).$$

Each $\widetilde{r}_{ij}$ is a weighted sum of terms like $W_{is}W_{jt}$, and each δ_{ij} is a weighted sum of terms like W_{jt}. Intuitively, we view $\widetilde{r}$-term as an "order-2 W-term" and view δ-term as "order-1 W-term." It motivates the definition of *intrinsic order of* W as

$$N_W^* = N_W + N_\delta + 2N_{\widetilde{r}}. \tag{2.E.304}$$

We group 34 terms by the value of N_W^*; see the last column of Table 2.3.

.There are 14 such terms, including R_2-R_6, R_9-R_{13}, R_{16}-R_{17}, and R_{21}-R_{22}. They all equal to zero under the null hypothesis, so it is sufficient to show that they satisfy (2.E.303) under the alternative hypothesis. We prove by comparing each R_k to some previously analyzed terms. Take R_9 for example. Plugging in the definition of $\widetilde{r}_{ij}$ and δ_{ij} gives

$$\begin{aligned} R_9 &= \sum_{i,j,k,\ell(dist)} [(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)][(\widetilde{\eta}_j - \eta_j)\eta_k + \eta_j(\widetilde{\eta}_k - \eta_k)]\widetilde{\Omega}_{k\ell}W_{\ell i} \\ &= R_{9a} + R_{9b}, \end{aligned}$$

where

$$\begin{aligned} R_{9a} &= \sum_{i,j,k,\ell(dist)} \eta_k\widetilde{\Omega}_{k\ell} \cdot [(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)^2 W_{\ell i}], \\ R_{9b} &= \sum_{i,j,k,\ell(dist)} \eta_j\widetilde{\Omega}_{k\ell} \cdot [(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)(\widetilde{\eta}_k - \eta_k)W_{\ell i}]. \end{aligned} \tag{2.E.305}$$

At the same time, we recall that T_1 in Lemmas 22-23 is defined as

$$T_1 = \sum_{i,j,k,\ell(dist)} \delta_{ij}\delta_{jk}\delta_{k\ell}W_{\ell i} = \sum_{i,j,k,\ell(dist)} \delta_{\ell j}\delta_{jk}\delta_{ki}W_{i\ell}.$$

In the proof of the above two lemmas, we express T_1 as the weighted sum of T_{1a}-T_{1d}; see (2.E.166). Note that T_{1a} and T_{1d} in (2.E.166) can be re-written as

$$\begin{aligned} T_{1d} &= \sum_{i,j,k,\ell(dist)} [\eta_\ell(\widetilde{\eta}_j - \eta_j)][(\widetilde{\eta}_j - \eta_j)\eta_k][\eta_k(\widetilde{\eta}_i - \eta_i)]W_{i\ell} \\ &= \sum_{i,j,k,\ell(dist)} \eta_k^2\eta_\ell \cdot [(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)^2 W_{\ell i}], \\ T_{1a} &= \sum_{i,j,k,\ell(dist)} [\eta_\ell(\widetilde{\eta}_j - \eta_j)][\eta_j(\widetilde{\eta}_k - \eta_k)][\eta_k(\widetilde{\eta}_i - \eta_i)]W_{i\ell} \\ &= \sum_{i,j,k,\ell(dist)} \eta_j\eta_k\eta_\ell \cdot [(\widetilde{\eta}_i - \eta_i)(\widetilde{\eta}_j - \eta_j)(\widetilde{\eta}_k - \eta_k)W_{i\ell}]. \end{aligned} \tag{2.E.306}$$

Compare (2.E.305) and (2.E.306). It is seen that R_{9a} and T_{1d} have the same structure, where the non-stochastic coefficients in the summand satisfy $|\eta_k \widetilde{\Omega}_{k\ell}| \leq C\alpha\theta_k^2\theta_\ell$ and $|\eta_k^2\eta_\ell| \leq C\theta_k^2\theta_\ell$, respectively. This means we can bound $|\mathbb{E}(R_{9a})|$ and $\mathrm{Var}(R_{9a})$ in the same way as we bound $|\mathbb{E}[T_{1d}]|$ and $\mathrm{Var}(T_{1d})$, and the bounds have an extra factor of α and α^2, respectively. In detail, in the proof of Lemmas 22-23, we have shown

$$|\mathbb{E}[T_{1d}]| \leq C\|\theta\|^4, \qquad \mathrm{Var}(T_{1d}) \leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1}.$$

It follows immediately that

$$|\mathbb{E}[R_{9a}]| \leq C\alpha\|\theta\|^4 = o(\alpha^2\|\theta\|^6), \qquad \mathrm{Var}(T_{1d}) \leq \frac{C\alpha^2\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8).$$

Similarly, since we have proved

$$|\mathbb{E}[T_{1a}]| \leq \frac{C\|\theta\|^6}{\|\theta\|_1^2}, \qquad \mathrm{Var}(T_{1a}) \leq \frac{C\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2},$$

it follows immediately that

$$|\mathbb{E}[R_{9b}]| \leq \frac{C\alpha\|\theta\|^6}{\|\theta\|_1^2} = o(\alpha^2\|\theta\|^6), \qquad \mathrm{Var}(R_{9b}) \leq \frac{C\alpha^2\|\theta\|^4\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8).$$

This proves (2.E.303) for $X = R_{9a}$.

Analysis of post-expansion sums with $N_W^* \leq 4$ We use the same strategy to bound all other terms with $N_W^* \leq 4$. The details are in Table 2.4. In each row of the table, the left column displays a targeting variable X, and the right column displays a previously analyzed variable, which we call X^*, that has a similar structure as X. It is not hard to see that we can obtain upper bounds for $|\mathbb{E}[X]|$ and $\mathrm{Var}(X)$ from multiplying the upper bounds of $|\mathbb{E}[X^*]|$ and $\mathrm{Var}(X^*)$ by α^m and α^{2m}, respectively, where m is a nonnegative integer (e.g., $m = 1$ in the analysis of R_9). Using our previous results, each X^* in the right column satisfies

$$|\mathbb{E}[X^*]| = O(\alpha^2\|\theta\|^6), \qquad \mathrm{Var}(X^*) = o(\|\theta\|^8 + \alpha^6\|\theta\|^8\|\theta\|_3^6).$$

So, each X in the left column satisfies (2.E.303).

.There are 10 such terms, including R_1, R_7-R_8, R_{14}-R_{15}, R_{18}, and R_{25}-R_{28}. Using the the notation

$$G_i = \widetilde{\eta}_i - \eta_i,$$

Table 2.4: The 14 types of post-expansion sums with $N_W^* \leq 4$. The right column displays the post-expansion sums defined before which have similar forms as the post-expansion sums in the left column. Definitions of the terms in the right column can be found in (2.E.130), (2.E.136), (2.E.142), (2.E.152), (2.E.158), (2.E.166), (2.E.167), and (2.E.168). For some terms in the right column, we permute (i,j,k,ℓ) in the original definition for ease of comparison with the left column. (In all expressions, the subscript "$i,j,k,\ell(dist)$" is omitted.)

	Expression		Expression
R_2	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}W_{k\ell}W_{\ell i}$	Z_{1b}	$\sum(\tilde\eta_i-\eta_i)\eta_j(\tilde\eta_j-\eta_j)\eta_k W_{k\ell}W_{\ell i}$
R_3	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)W_{jk}\widetilde\Omega_{k\ell}W_{\ell i}$	Z_{2a}	$\sum\eta_\ell(\tilde\eta_j-\eta_j)W_{jk}\eta_k(\tilde\eta_i-\eta_i)W_{i\ell}$
R_4	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}\widetilde\Omega_{k\ell}W_{\ell i}$	Z_{3d}	$\sum(\tilde\eta_i-\eta_i)\eta_j(\tilde\eta_j-\eta_j)\eta_k\widetilde\Omega_{k\ell}W_{\ell i}$
R_5	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}W_{k\ell}\widetilde\Omega_{\ell i}$	Z_{4b}	$\sum\widetilde\Omega_{ij}(\tilde\eta_j-\eta_j)\eta_k W_{k\ell}\eta_\ell(\tilde\eta_i-\eta_i)$
R_6	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}\widetilde\Omega_{k\ell}\widetilde\Omega_{\ell i}$	Z_{5a}	$\sum\eta_i(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}\widetilde\Omega_{k\ell}\eta_\ell(\tilde\eta_i-\eta_i)$
R_9	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2\eta_k\widetilde\Omega_{k\ell}W_{\ell i}$	T_{1d}	$\sum\eta_\ell(\tilde\eta_j-\eta_j)^2\eta_k^2(\tilde\eta_i-\eta_i)W_{i\ell}$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\widetilde\Omega_{k\ell}W_{\ell i}$	T_{1a}	$\sum\eta_\ell(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\eta_k(\tilde\eta_i-\eta_i)W_{i\ell}$
R_{10}	$\sum(\tilde\eta_i-\eta_i)^2(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}W_{k\ell}\eta_\ell$	T_{1c}	$\sum(\tilde\eta_j-\eta_j)\eta_k W_{k\ell}\eta_\ell(\tilde\eta_i-\eta_i)^2\eta_j$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}W_{k\ell}(\tilde\eta_\ell-\eta_\ell)\eta_i$	T_{1a}	$\sum(\tilde\eta_j-\eta_j)\eta_k W_{k\ell}(\tilde\eta_\ell-\eta_\ell)\eta_i(\tilde\eta_i-\eta_i)\eta_j$
R_{11}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)W_{jk}\eta_k(\tilde\eta_\ell-\eta_\ell)\widetilde\Omega_{\ell i}$	T_{1a}	$\sum(\tilde\eta_i-\eta_i)\eta_k W_{kj}(\tilde\eta_j-\eta_j)\eta_\ell(\tilde\eta_\ell-\eta_\ell)\eta_i$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)W_{jk}(\tilde\eta_k-\eta_k)\eta_\ell\widetilde\Omega_{\ell i}$	T_{1b}	$\sum\eta_i(\tilde\eta_k-\eta_k)W_{kj}(\tilde\eta_j-\eta_j)\eta_\ell^2(\tilde\eta_i-\eta_i)$
R_{12}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2\eta_k\widetilde\Omega_{k\ell}\widetilde\Omega_{\ell i}$	T_{2c}	$\sum\eta_i(\tilde\eta_j-\eta_j)^2\eta_k\widetilde\Omega_{k\ell}\eta_\ell(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\widetilde\Omega_{k\ell}\widetilde\Omega_{\ell i}$	T_{2a}	$\sum\eta_i(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\widetilde\Omega_{k\ell}\eta_\ell(\tilde\eta_i-\eta_i)$
R_{13}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}(\tilde\eta_k-\eta_k)\eta_\ell\widetilde\Omega_{\ell i}$	T_{2b}	$\sum\eta_i(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}(\tilde\eta_k-\eta_k)\eta_\ell^2(\tilde\eta_i-\eta_i)$
R_{16}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2\eta_k(\tilde\eta_k-\eta_k)\eta_\ell\widetilde\Omega_{\ell i}$	F_b	$\sum\eta_i(\tilde\eta_j-\eta_j)^2\eta_k(\tilde\eta_k-\eta_k)\eta_\ell^2(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2\eta_k^2(\tilde\eta_\ell-\eta_\ell)\widetilde\Omega_{\ell i}$	F_b	$\sum\eta_i(\tilde\eta_j-\eta_j)^2\eta_k^2(\tilde\eta_\ell-\eta_\ell)\eta_\ell(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)^2\eta_\ell\widetilde\Omega_{\ell i}$	F_b	$\sum\eta_i(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)^2\eta_\ell^2(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\eta_k(\tilde\eta_\ell-\eta_\ell)\widetilde\Omega_{\ell i}$	F_a	$\sum\eta_i(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\eta_k(\tilde\eta_\ell-\eta_\ell)\eta_\ell(\tilde\eta_i-\eta_i)$
R_{17}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\widetilde\Omega_{k\ell}(\tilde\eta_\ell-\eta_\ell)\eta_i$	F_a	$\sum\eta_i(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\eta_k(\tilde\eta_\ell-\eta_\ell)\eta_\ell(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2\eta_k\widetilde\Omega_{k\ell}(\tilde\eta_\ell-\eta_\ell)\eta_i$	F_b	$\sum\eta_i(\tilde\eta_j-\eta_j)^2\eta_k^2(\tilde\eta_\ell-\eta_\ell)\eta_\ell(\tilde\eta_i-\eta_i)$
	$\sum(\tilde\eta_i-\eta_i)^2(\tilde\eta_j-\eta_j)^2\eta_k\widetilde\Omega_{k\ell}\eta_\ell$	F_c	$\sum\eta_\ell(\tilde\eta_i-\eta_i)^2\eta_k^2(\tilde\eta_j-\eta_j)^2\eta_\ell$
R_{21}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)^2(\tilde\eta_k-\eta_k)\widetilde\Omega_{k\ell}\widetilde\Omega_{\ell i}$	F_b	$\sum\eta_i(\tilde\eta_j-\eta_j)^2\eta_k(\tilde\eta_k-\eta_k)\eta_\ell^2(\tilde\eta_i-\eta_i)$
R_{22}	$\sum(\tilde\eta_i-\eta_i)(\tilde\eta_j-\eta_j)\widetilde\Omega_{jk}(\tilde\eta_k-\eta_k)(\tilde\eta_\ell-\eta_\ell)\widetilde\Omega_{\ell i}$	F_a	$\sum\eta_i(\tilde\eta_j-\eta_j)\eta_j(\tilde\eta_k-\eta_k)\eta_k(\tilde\eta_\ell-\eta_\ell)\eta_\ell(\tilde\eta_i-\eta_i)$

we get the following expressions (note: factors of $(\frac{v}{V})^m$ have been removed; see explanations in (2.E.302)-(2.E.303)):

$$R_1 = \sum_{i,j,k,\ell(dist)} G_iG_jW_{jk}W_{k\ell}W_{\ell i},$$

$$R_7 = \sum_{i,j,k,\ell(dist)} G_iG_j\eta_jG_kW_{k\ell}W_{\ell i} + \sum_{i,j,k,\ell(dist)} G_iG_j^2\eta_kW_{k\ell}W_{\ell i}$$
$$= \sum_{i,j,k,\ell(dist)} \eta_j(G_iG_jG_kW_{k\ell}W_{\ell i}) + \sum_{i,j,k,\ell(dist)} \eta_k(G_iG_j^2W_{k\ell}W_{\ell i}),$$

$$R_8 = 2\sum_{i,j,k,\ell(dist)} G_iG_jW_{jk}\eta_kG_\ell W_{\ell i} = 2\sum_{i,j,k,\ell(dist)} \eta_k(G_iG_jG_\ell W_{jk}W_{\ell i}),$$

$$R_{14} = \sum_{\substack{i,j,k,\ell\\(dist)}} G_iG_j^2\eta_k^2G_\ell W_{\ell i} + 2\sum_{\substack{i,j,k,\ell\\(dist)}} G_iG_j^2\eta_kG_k\eta_\ell W_{\ell i} + \sum_{\substack{i,j,k,\ell\\(dist)}} G_iG_j\eta_jG_k\eta_kG_\ell W_{\ell i}$$
$$= \sum_{\substack{i,j,k,\ell\\(dist)}} \eta_k^2(G_iG_j^2G_\ell W_{\ell i}) + 2\sum_{\substack{i,j,k,\ell\\(dist)}} \eta_k\eta_\ell(G_iG_j^2G_kW_{\ell i}) + \sum_{\substack{i,j,k,\ell\\(dist)}} \eta_j\eta_k(G_iG_jG_kG_\ell W_{\ell i}),$$

$$R_{15} = \sum_{\substack{i,j,k,\ell}} G_iG_j\eta_jG_kW_{k\ell}G_\ell\eta_i + 2\sum_{\substack{i,j,k,\ell}} G_iG_j^2\eta_kW_{k\ell}G_\ell\eta_i + \sum_{\substack{i,j,k,\ell}} G_iG_j^2\eta_kW_{k\ell}\eta_\ell G_i$$

Each expression above belongs to one of the following types:

$$J_1 = \sum_{i,j,k,\ell(dist)} G_iG_jW_{jk}W_{k\ell}W_{\ell i}, \qquad J_2 = \sum_{i,j,k,\ell(dist)} \eta_j(G_iG_jG_kW_{k\ell}W_{\ell i}),$$

$$J_3 = \sum_{i,j,k,\ell(dist)} \eta_k(G_iG_jG_\ell W_{jk}W_{\ell i}), \qquad J_4 = \sum_{i,j,k,\ell(dist)} \eta_k(G_iG_j^2W_{k\ell}W_{\ell i}),$$

$$J_5 = \sum_{i,j,k,\ell(dist)} \eta_j\eta_k(G_iG_jG_kG_\ell W_{\ell i}), \qquad J_5' = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{jk}(G_iG_jG_kG_\ell W_{\ell i}),$$

$$J_6 = \sum_{i,j,k,\ell(dist)} \eta_k\eta_\ell(G_iG_j^2G_kW_{\ell i}), \qquad J_6' = \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{k\ell}(G_iG_j^2G_kW_{\ell i}),$$

$$J_7 = \sum_{i,j,k,\ell(dist)} \eta_k^2(G_iG_j^2G_\ell W_{\ell i}), \qquad J_8 = \sum_{i,j,k,\ell(dist)} \eta_k\eta_\ell(G_i^2G_j^2W_{k\ell}),$$

$$J_9 = \sum_{i,j,k,\ell(dist)} \eta_k\widetilde{\Omega}_{\ell i}(G_iG_j^2G_kG_\ell), \qquad J_{10} = \sum_{i,j,k,\ell(dist)} \eta_\ell\widetilde{\Omega}_{\ell i}(G_iG_j^2G_k^2).$$

Since $|\eta_j\eta_k| \leq C\theta_j\theta_k$ and $|\widetilde{\Omega}_{jk}| \leq C\alpha\theta_j\theta_k$, the study of J_5 and J_5' are similar. Also, the study of J_6 and J_6' are similar. We now study J_1-J_{10}. Consider J_1. It is seen that

$$J_1 = \frac{1}{v}\sum_{i,j,k,\ell(dist)} \Big(\sum_{s\neq i} W_{is}\Big)\Big(\sum_{t\neq j} W_{jt}\Big)W_{jk}W_{k\ell}W_{\ell i} = \frac{1}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\neq i, t\neq j}} W_{is}W_{i\ell}W_{jt}W_{jk}W_{k\ell}.$$

Since s can be equal to ℓ and t can be equal to k, there are three different types:

$$J_{1a} = \frac{1}{v}\sum_{i,j,k,\ell(dist)} W_{i\ell}^2W_{jk}^2W_{k\ell}, \qquad J_{1b} = \frac{1}{v}\sum_{\substack{i,j,k,\ell(dist)\\ t\notin\{j,k\}}} W_{i\ell}^2W_{jt}W_{jk}W_{k\ell},$$

$$J_{1c} = \frac{1}{v}\sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{i,\ell\}, t\notin\{j,k\}}} W_{is}W_{i\ell}W_{jt}W_{jk}W_{k\ell}.$$

We now calculate $\mathbb{E}[J_{1a}^2]$-$\mathbb{E}[J_{1c}^2]$. Take J_{1a} for example. In order to get nonzero $\mathbb{E}[W_{i\ell}^2W_{jk}^2W_{k\ell}W_{i'\ell'}^2W_{j'k'}^2W_{k'\ell'}]$, we need either $W_{k\ell} = W_{k'\ell'}$ or each of the two variables $(W_{k\ell}, W_{k',\ell'})$ equals to another squared-W term. The leading term of $\mathbb{E}[J_{1a}^2]$ comes from the first case. In this case, we have $W_{k\ell} = W_{k'\ell'}$ but allow for $W_{i\ell} \neq W_{i'\ell'}$ and $W_{jk} \neq W_{j'k'}$. It has to be the case of either $(k',\ell') = (k,\ell)$ or $(k',\ell') = (\ell,k)$. Therefore, we have $\mathbb{E}[W_{i\ell}^2W_{jk}^2W_{k\ell}W_{i'\ell'}^2W_{j'k'}^2W_{k'\ell'}] = \mathbb{E}[W_{i\ell}^2W_{jk}^2W_{i'\ell'}^2W_{j'k'}^2W_{k\ell}^2]$. Using similar arguments, we have the following results, where details are omitted, as they are similar to the calculations in the proof of Lemmas 18-23.

$$\mathbb{E}[J_{1a}^2] \leq \frac{C}{v^2}\sum_{\substack{i,j,k,\ell\\ i',j'}} \mathbb{E}[W_{i\ell}^2W_{jk}^2W_{i'\ell}^2W_{j'k}^2W_{k\ell}^2] \leq \frac{C}{\|\theta\|_1^4}\sum_{\substack{i,j,k,\ell\\ i',j'}} \theta_i\theta_j\theta_k^3\theta_\ell^3\theta_{i'}\theta_{j'} \leq C\|\theta\|_3^6,$$

$$\mathbb{E}[J_{1b}^2] \leq \frac{C}{v^2}\sum_{\substack{i,j,k,\ell,t\\ i'}} \mathbb{E}[W_{i\ell}^2W_{i'\ell}^2W_{jt}^2W_{jk}^2W_{k\ell}^2] \leq \frac{C}{\|\theta\|_1^4}\sum_{\substack{i,j,k,\ell,t\\ i'}} \theta_i\theta_j^2\theta_k^2\theta_\ell^3\theta_t\theta_{i'} \leq \frac{C\|\theta\|^4\|\theta\|_3^3}{\|\theta\|_1},$$

$$\mathbb{E}[J_{1c}^2] \leq \frac{C}{v^2}\sum_{i,j,k,\ell,s,t} \mathbb{E}[W_{is}^2W_{i\ell}^2W_{jt}^2W_{jk}^2W_{k\ell}^2] \leq \frac{C}{\|\theta\|_1^4}\sum_{i,j,k,\ell,s,t} \theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s\theta_t \leq \frac{C\|\theta\|^8}{\|\theta\|_1^2}.$$

The right hand sides are all $o(\|\theta\|^8)$. It follows that

$$\mathbb{E}[J_1^2] = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Consider J_2-J_4. By definition,

$$J_2 = \frac{1}{v\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq k}} \eta_j W_{is}W_{jt}W_{kq}W_{k\ell}W_{\ell i}, \qquad J_3 = \frac{1}{v\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq \ell}} \eta_k W_{is}W_{jt}W_{\ell q}W_{jk}W_{\ell i},$$

$$J_4 = \frac{1}{v\sqrt{v}} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq j}} \eta_k W_{is}W_{jt}W_{jq}W_{k\ell}W_{\ell i}.$$

The analysis is summarized in Table 2.5. In the first column of this table, we study different types of summands. For example, in the expression of J_2, $W_{is}W_{kq}W_{k\ell}W_{\ell i}$ have four different cases: (a) $W_{k\ell}^2W_{\ell i}^2$, (b) $W_{k\ell}^2W_{\ell i}W_{is}$ or $W_{k\ell}W_{\ell i}^2W_{kq}$, (c) $W_{k\ell}W_{\ell i}W_{ik}^2$, and (d) $W_{k\ell}W_{\ell i}W_{is}W_{kq}$. In cases (b) and (d), W_{is} or W_{kq} may further equal to W_{jt}. Having explored all variants and considered index symmetry, we end up with 6 different cases, as listed in the first column of Table 2.5. In the second column, we study the mean of the squares of the sum of each type of summands. Take the first row for example. We aim to study

$$\mathrm{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ t\neq j}} \eta_j (W_{k\ell}^2W_{\ell i}^2)W_{jt}\Big)\Big].$$

The naive expansion gives the sum of $\eta_j\eta_{j'}\,\mathbb{E}[W_{k\ell}^2W_{\ell i}^2W_{jt}W_{k'\ell'}^2W_{\ell' i'}^2W_{j't'}]$ over $(i,j,k,\ell,t,i',j',k',\ell',t')$. However, for this term to be nonzero, all single-W terms have to be paired (either with another single-W term or with a squared-W term). The main contribution is from the case of $W_{jt} = W_{j't'}$. This is satisfied only when $(j',s') = (j,s)$ or $(j',s') = (s,j)$. By calculations which are omitted here, we can show that $(j',s') = (j,s)$ yields a larger bound. Therefore, it reduces to the sum of $\eta_j^2\,\mathbb{E}[(W_{jt}^2)W_{k\ell}^2W_{\ell i}^2W_{k'\ell'}^2W_{\ell' i'}^2]$ over $(i,j,k,\ell,t,i',k',\ell')$, which is displayed in the second column of the table. In the last column, we sum the quantity in the second column over indices; it gives rise to a bound for the mean of the square of sum. See the table for details. Recall that the definition of J_2-J_4 contains a factor of $\frac{1}{v\sqrt{v}}$ in front of the sum, where $v \asymp \|\theta\|_1^2$ by (2.E.116). Hence, to get a desired bound, we only need that each row in the third column of Table 2.5 is

$$o(\|\theta\|^8\|\theta\|_1^6).$$

This is true. We thus conclude that

$$\max\big\{\mathbb{E}[J_2^2],\ \mathbb{E}[J_3^2],\ \mathbb{E}[J_4^2]\big\} = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Analysis of post-expansion sums with $N_W^* = 5$ Consider J_5-J_8. It is seen that

$$J_5 = \frac{1}{v^2} \sum_{i,j,k,\ell(dist)} \eta_j\eta_k W_{is}W_{jt}W_{kq}W_{\ell m}W_{\ell i}, \quad J_6 = \frac{1}{v^2} \sum_{i,j,k,\ell(dist)} \eta_k\eta_\ell W_{is}W_{jt}W_{jq}W_{km}W_{\ell i},$$

$$J_7 = \frac{1}{v^2} \sum_{i,j,k,\ell(dist)} \eta_k^2 W_{is}W_{jt}W_{jq}W_{\ell m}W_{\ell i}, \qquad J_8 = \frac{1}{v^2} \sum_{i,j,k,\ell(dist)} \eta_k\eta_\ell W_{is}W_{it}W_{jq}W_{jm}W_{k\ell},$$

Table 2.5: Analysis of J_2-J_4. In the second column, the variables in brackets are paired W terms.

	Types of summand	Terms in mean-squared	Bound
J_2	$\eta_j(W_{k\ell}^2W_{\ell i}^2)W_{jt}$	$\eta_j^2\,\mathbb{E}[(W_{jt}^2)W_{k\ell}^2W_{\ell i}^2W_{k'\ell'}^2W_{\ell' i'}^2] \leq \theta_i\theta_j^3\theta_k\theta_\ell^2\theta_t\theta_{i'}\theta_{k'}\theta_{\ell'}^2$	$\|\theta\|^4\|\theta\|_3^3\|\theta\|_1^5$
	$\eta_j(W_{k\ell}W_{\ell i}W_{ik}^2)W_{jt}$	$\eta_j^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{jt}^2)W_{ik}^4] \leq C\theta_i^2\theta_j^3\theta_k^2\theta_\ell^2\theta_t$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1$
	$\eta_j(W_{k\ell}^2W_{\ell i}W_{is})W_{jt}$	$\eta_j^2\,\mathbb{E}[(W_{\ell i}^2W_{is}^2W_{jt}^2)W_{k\ell}^2W_{k'\ell}^2] \leq C\theta_i^2\theta_j^3\theta_k\theta_\ell^3\theta_s\theta_t\theta_{k'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^4$
	$\eta_j(W_{k\ell}^2W_{\ell i})W_{ij}^2$	$\eta_j\eta_{j'}\,\mathbb{E}[(W_{\ell i}^2)W_{k\ell}^2W_{ij}^2W_{k'\ell}^2W_{ij'}^2] \leq C\theta_i^3\theta_j^2\theta_k\theta_\ell^3\theta_{j'}^2\theta_{k'}$	$\|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2$
	$\eta_j(W_{k\ell}W_{\ell i}W_{kq}W_{is})W_{jt}$	$\eta_j^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{kq}^2W_{is}^2W_{jt}^2)] \leq C\theta_i^2\theta_j^3\theta_k^2\theta_\ell^2\theta_s\theta_t\theta_q$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1^3$
	$\eta_j(W_{k\ell}W_{\ell i})W_{kq}W_{ij}^2$	$\eta_j\eta_{j'}\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{kq}^2)W_{ij}^2W_{ij'}^2] \leq C\theta_i^3\theta_j^2\theta_k^2\theta_\ell^2\theta_q\theta_{j'}^2$	$\|\theta\|^8\|\theta\|_3^3\|\theta\|_1$
J_3	$\eta_kW_{\ell i}^3W_{jk}^2$	$\eta_k\eta_{k'}\,\mathbb{E}[W_{\ell i}^3W_{jk}^2W_{\ell' i'}^3W_{j'k'}^2] \leq C\theta_i\theta_j\theta_k^2\theta_\ell\theta_{i'}\theta_{j'}\theta_{k'}^2\theta_{\ell'}$	$\|\theta\|^4\|\theta\|_1^6$
	$\eta_kW_{\ell i}^3(W_{jk}W_{jt})$	$\eta_k^2\,\mathbb{E}[(W_{jk}^2W_{jt}^2)W_{\ell i}^3W_{\ell' i'}^3] \leq C\theta_i\theta_j^2\theta_k^3\theta_\ell\theta_t\theta_{i'}\theta_{\ell'}$	$\|\theta\|^2\|\theta\|_3^3\|\theta\|_1^5$
	$\eta_k(W_{\ell i}^2W_{is})W_{jk}^2$	$\eta_k\eta_{k'}\,\mathbb{E}[(W_{is}^2)W_{\ell i}^2W_{jk}^2W_{\ell' i}^2W_{j'k'}^2] \leq C\theta_i^3\theta_j\theta_k^2\theta_\ell\theta_s\theta_{j'}\theta_{k'}^2\theta_{\ell'}$	$\|\theta\|^4\|\theta\|_3^3\|\theta\|_1^5$
	$\eta_k(W_{\ell i}^2W_{is})W_{jk}W_{jt}$	$\eta_k^2\,\mathbb{E}[(W_{is}^2W_{jk}^2W_{jt}^2)W_{\ell i}^2W_{\ell' i}^2] \leq C\theta_i^3\theta_j^2\theta_k^3\theta_\ell\theta_s\theta_t\theta_{\ell'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^4$
	$\eta_kW_{\ell i}^2W_{ij}^2W_{jk}$	$\eta_k^2\,\mathbb{E}[(W_{jk}^2)W_{\ell i}^2W_{ij}^2W_{\ell' i'}^2W_{i'j}^2] \leq C\theta_i^2\theta_j^3\theta_k^3\theta_\ell\theta_{i'}^2\theta_{\ell'}$	$\|\theta\|^4\|\theta\|_3^6\|\theta\|_1^2$
	$\eta_k(W_{\ell i}W_{is}W_{\ell q})W_{jk}^2$	$\eta_k\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2W_{is}^2W_{\ell q}^2)W_{jk}^2W_{j'k'}^2] \leq C\theta_i^2\theta_j\theta_k^2\theta_\ell^2\theta_s\theta_q\theta_{j'}\theta_{k'}^2$	$\|\theta\|^8\|\theta\|_1^4$
	$\eta_k(W_{\ell i}W_{is}W_{\ell q})W_{jk}W_{jt}$	$\eta_k^2\,\mathbb{E}[(W_{\ell i}^2W_{is}^2W_{\ell q}^2W_{jk}^2W_{jt}^2)] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^2\theta_s\theta_t\theta_q$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1^3$
	$\eta_kW_{\ell i}W_{ij}^2W_{\ell q}W_{jk}$	$\eta_k^2\,\mathbb{E}[(W_{\ell i}^2W_{\ell q}^2W_{jk}^2)W_{ij}^4] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^2\theta_q$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1$
J_4	$\eta_k(W_{k\ell}W_{\ell i}^2)W_{jt}^2$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2)W_{\ell i}^2W_{jt}^2W_{\ell i'}^2W_{j't'}^2] \leq C\theta_i\theta_j\theta_k^3\theta_\ell^3\theta_t\theta_{i'}\theta_{j'}\theta_{t'}$	$\|\theta\|_3^6\|\theta\|_1^6$
	$\eta_k(W_{k\ell}W_{\ell i}^2)W_{jt}W_{jq}$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2W_{jt}^2W_{jq}^2)W_{\ell i}^2W_{\ell i'}^2] \leq C\theta_i\theta_j^2\theta_k^3\theta_\ell^3\theta_t\theta_q\theta_{i'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^4$
	$\eta_k(W_{k\ell}W_{\ell i}W_{is})W_{jt}^2$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{is}^2)W_{jt}^2W_{j't'}^2] \leq C\theta_i^2\theta_j\theta_k^3\theta_\ell^2\theta_s\theta_t\theta_{j'}\theta_{t'}$	$\|\theta\|^4\|\theta\|_3^3\|\theta\|_1^5$
	$\eta_kW_{k\ell}W_{\ell i}W_{ij}^3$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2)W_{ij}^3W_{ij'}^3] \leq C\theta_i^3\theta_j\theta_k^3\theta_\ell^2\theta_{j'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta\|_1^2$
	$\eta_k(W_{k\ell}W_{\ell i}W_{is})W_{jt}W_{jq}$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{is}^2W_{jt}^2W_{jq}^2] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^2\theta_s\theta_t\theta_q$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1^3$
	$\eta_kW_{k\ell}W_{\ell i}W_{ij}^2W_{jq}$	$\eta_k^2\,\mathbb{E}[(W_{k\ell}^2W_{\ell i}^2W_{jq}^2)W_{ij}^4] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^2\theta_q$	$\|\theta\|^6\|\theta\|_3^3\|\theta\|_1$

The analysis is summarized in Table 2.6. We note that J_7 can be written as

$$J_7 = \frac{1}{v^2}\sum_{i,j,\ell(dist)} \beta_{ij\ell}W_{is}W_{jt}W_{jq}W_{\ell m}W_{\ell i}, \qquad \text{where} \quad \beta_{ij\ell} \equiv \sum_{k\notin\{i,j,\ell\}} \eta_k^2.$$

Although the values of $\beta_{ij\ell}$ change with indices, they have a common upper bound of $C\|\theta\|^2$. We treat $\beta_{ij\ell}$ as $\|\theta\|^2$ in Table 2.6, as this doesn't change the bounds but simplifies notations. Recall that the definition of J_5-J_8 contains a factor of $\frac{1}{v^2}$ in front of the sum, where $v \asymp \|\theta\|_1^2$ by (2.E.116). Hence, to get a desired bound, we only need that each row in the third column of Table 2.5 is

$$o(\|\theta\|^8\|\theta\|_1^8).$$

This is true. We thus conclude that

$$\max\left\{\mathbb{E}[J_5^2],\ \mathbb{E}[J_6^2],\ \mathbb{E}[J_7^2],\ \mathbb{E}[J_8^2]\right\} = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Consider J_9-J_{10}. They can be analyzed in the same way as we did for J_1-J_8. To save space, we only give a simplified proof for the case of $\|\theta\| \gg \alpha[\log(n)]^{5/2}$. For $1 \ll \|\theta\| \leq C\alpha[\log(n)]^{5/2}$, the proof is similar to those in Tables 2.5-2.6, which is omitted. For a constant $C_0 > 0$ to be decided, we introduce an event

$$E = \cap_{i=1}^n E_i, \qquad \text{where} \quad E_i = \left\{\sqrt{v}|G_i| \leq C_0\sqrt{\theta_i\|\theta\|_1\log(n)}\right\}. \tag{2.E.307}$$

Recall that $\sqrt{v}G_i = \sqrt{v}(\widetilde{\eta}_i - \eta_i) = \sum_{j\neq i}(A_{ij} - \mathbb{E}A_{ij})$. The variables $\{A_{ij}\}_{j\neq i}$ are mutually independent, satisfying that $|A_{ij} - \mathbb{E}A_{ij}| \leq 1$ and $\sum_j \mathrm{Var}(A_{ij}) \leq \sum_j \theta_i\theta_j \leq \theta_i\|\theta\|_1$. By

Table 2.6: Analysis of J_5-J_8. In the second column, the variables in brackets are paired W terms.

	Types of summand	Terms in mean-squared	Bound
	$\eta_j\eta_k W_{\ell i}^3 W_{jk}^2$	$\eta_j\eta_k\eta_{j'}\eta_{k'}\,\mathbb{E}[W_{\ell i}^3 W_{jk}^2 W_{\ell' i'}^3 W_{j'k'}^2] \leq C\theta_i\theta_j^2\theta_k^2\theta_\ell\theta_{i'}\theta_{j'}^2\theta_{k'}^2\theta_{\ell'}$	$\|\theta\|^8\|\theta\|_1^4$
	$\eta_j\eta_k W_{\ell i}^3(W_{jt}W_{kq})$	$\eta_j^2\eta_k^2\,\mathbb{E}[(W_{jt}^2 W_{kq}^2)W_{\ell i}^3 W_{\ell' i'}^3] \leq C\theta_i\theta_j^3\theta_k^3\theta_\ell\theta_t\theta_q\theta_{i'}\theta_{\ell'}$	$\|\theta\|_3^6\|\theta\|_1^6$
	$\eta_j\eta_k(W_{\ell i}^2 W_{is})W_{jk}^2$	$\eta_j\eta_k\eta_{j'}\eta_{k'}\,\mathbb{E}[(W_{is}^2)W_{\ell i}^2 W_{jk}^2 W_{\ell' i}^2 W_{j'k'}^2] \leq C\theta_i^3\theta_j^2\theta_k^2\theta_\ell\theta_s\theta_{j'}^2\theta_{k'}^2\theta_{\ell'}$	$\|\theta\|^8\|\theta\|_3^3\|\theta$
J_5	$\eta_j\eta_k(W_{\ell i}^2 W_{is})(W_{jt}W_{kq})$	$\eta_j^2\eta_k^2\,\mathbb{E}[(W_{is}^2)W_{jt}^2 W_{kq}^2)W_{\ell i}^2 W_{\ell' i}^2] \leq C\theta_i^3\theta_j^3\theta_k^3\theta_\ell\theta_s\theta_t\theta_q\theta_{\ell'}$	$\|\theta\|_3^9\|\theta\|_1^5$
	$\eta_j\eta_k W_{\ell i}^2 W_{ij}^2 W_{kq}$	$\eta_j\eta_k^2\eta_{j'}\,\mathbb{E}[(W_{kq}^2)W_{\ell i}^2 W_{ij}^2 W_{\ell' i'}^2 W_{i'j'}^2] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell\theta_q\theta_{i'}^2\theta_{j'}^2\theta_{\ell'}$	$\|\theta\|^8\|\theta\|_3^3\|\theta$
	$\eta_j\eta_k(W_{\ell i}W_{is}W_{\ell m})W_{jk}^2$	$\eta_j\eta_k\eta_{j'}\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{\ell m}^2)W_{jk}^2 W_{j'k'}^2] \leq C\theta_i^2\theta_j^2\theta_k^2\theta_\ell^2\theta_s\theta_m\theta_{j'}^2\theta_{k'}^2$	$\|\theta\|^{12}\|\theta\|_1^2$
	$\eta_j\eta_k(W_{\ell i}W_{is}W_{\ell m})(W_{jt}W_{kq})$	$\eta_j^2\eta_k^2\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{\ell m}^2 W_{jt}^2 W_{kq}^2)] \leq C\theta_i^2\theta_j^3\theta_k^3\theta_\ell^2\theta_s\theta_t\theta_q\theta_m$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_j\eta_k W_{\ell i}W_{ij}^2 W_{\ell m}W_{kq}$	$\eta_j\eta_k^2\eta_{j'}\,\mathbb{E}[(W_{\ell i}^2 W_{\ell m}^2 W_{kq}^2)W_{ij}^2 W_{ij'}^2] \leq C\theta_i^3\theta_j^2\theta_k^3\theta_\ell^2\theta_q\theta_m\theta_{j'}^2$	$\|\theta\|^6\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{\ell i}^2 W_{jt}^2 W_{km}$	$\eta_k^2\eta_\ell\eta_{\ell'}\,\mathbb{E}[(W_{km}^2)W_{\ell i}^2 W_{jt}^2 W_{\ell' i'}^2 W_{j't'}^2] \leq C\theta_i\theta_j\theta_k^3\theta_\ell^2\theta_t\theta_m\theta_{i'}\theta_{j'}\theta_{\ell'}^2\theta_{t'}$	$\|\theta\|^4\|\theta\|_3^3\|\theta$
	$\eta_k\eta_\ell W_{\ell i}^2 W_{jk}^3$	$\eta_k\eta_\ell\eta_{k'}\eta_{\ell'}\,\mathbb{E}[W_{\ell i}^2 W_{jk}^3 W_{\ell' i'}^2 W_{j'k'}^3] \leq C\theta_i\theta_j\theta_k^2\theta_\ell^2\theta_{i'}\theta_{j'}\theta_{k'}^2\theta_{\ell'}^2$	$\|\theta\|^8\|\theta\|_1^4$
	$\eta_k\eta_\ell W_{\ell i}^2(W_{jt}W_{jq})W_{km}$	$\eta_k^2\eta_\ell\eta_{\ell'}\,\mathbb{E}[(W_{jt}^2 W_{jq}^2 W_{km}^2)W_{\ell i}^2 W_{\ell' i'}^2] \leq C\theta_i\theta_j^2\theta_k^3\theta_\ell^2\theta_t\theta_q\theta_m\theta_{i'}\theta_{\ell'}^2$	$\|\theta\|^6\|\theta\|_3^3\|\theta$
	$\eta_k\eta_\ell W_{\ell i}^2 W_{jk}^2 W_{jq}$	$\eta_k\eta_\ell\eta_{k'}\eta_{\ell'}\,\mathbb{E}[(W_{jq}^2)W_{\ell i}^2 W_{jk}^2 W_{\ell' i'}^2 W_{jk'}^2] \leq C\theta_i\theta_j^3\theta_k^2\theta_\ell^2\theta_q\theta_{i'}\theta_{k'}^2\theta_{\ell'}^2$	$\|\theta\|^8\|\theta\|_3^3\|\theta$
	$\eta_k\eta_\ell(W_{\ell i}W_{is})W_{jt}^2 W_{km}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{km}^2)W_{jt}^2 W_{j't'}^2] \leq C\theta_i^2\theta_j\theta_k^3\theta_\ell^3\theta_s\theta_t\theta_m\theta_{j'}\theta_{t'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta$
J_6	$\eta_k\eta_\ell W_{\ell i}W_{ij}^3 W_{km}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{\ell i}^2 W_{km}^2)W_{ij}^3 W_{ij'}^3] \leq C\theta_i^3\theta_j\theta_k^3\theta_\ell^3\theta_m\theta_{j'}$	$\|\theta\|_3^9\|\theta\|_1^3$
	$\eta_k\eta_\ell W_{\ell i}W_{is}W_{jk}^3$	$\eta_k\eta_\ell^2\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2)W_{jk}^3 W_{j'k'}^3] \leq C\theta_i^2\theta_j\theta_k^2\theta_\ell^3\theta_s\theta_{j'}\theta_{k'}^2$	$\|\theta\|^6\|\theta\|_3^3\|\theta$
	$\eta_k\eta_\ell W_{\ell i}W_{ik}^2 W_{jt}^2$	$\eta_k\eta_\ell^2\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2)W_{ik}^2 W_{jt}^2 W_{ik'}^2 W_{j't'}^2] \leq C\theta_i^3\theta_j\theta_k^2\theta_\ell^3\theta_t\theta_{j'}\theta_{k'}^2\theta_{t'}$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell(W_{\ell i}W_{is})(W_{jt}W_{jq})W_{km}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{jt}^2 W_{jq}^2 W_{km}^2)] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^3\theta_s\theta_t\theta_q\theta_m$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{\ell i}W_{ij}^2 W_{jq}W_{km}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{\ell i}^2 W_{jq}^2 W_{km}^2)W_{ij}^4] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^3\theta_q\theta_m$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{\ell i}W_{is}W_{jk}^2 W_{jq}$	$\eta_k\eta_\ell^2\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{jq}^2)W_{jk}^2 W_{jk'}^2] \leq C\theta_i^2\theta_j^3\theta_k^2\theta_\ell^3\theta_s\theta_q\theta_{k'}^2$	$\|\theta\|^6\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{\ell i}W_{ik}^2 W_{jt}W_{jq}$	$\eta_k\eta_\ell^2\eta_{k'}\,\mathbb{E}[(W_{\ell i}^2 W_{jt}^2 W_{jq}^2)W_{ik}^2 W_{ik'}^2] \leq C\theta_i^3\theta_j^2\theta_k^2\theta_\ell^3\theta_t\theta_q\theta_{k'}^2$	$\|\theta\|^6\|\theta\|_3^6\|\theta$
	$\|\theta\|^2 W_{\ell i}^3 W_{jt}^2$	$\|\theta\|^4\,\mathbb{E}[W_{\ell i}^3 W_{jt}^2 W_{\ell' i'}^3 W_{j't'}^2] \leq C\|\theta\|^4\theta_i\theta_j\theta_\ell\theta_t\theta_{i'}\theta_{j'}\theta_{\ell'}\theta_{t'}$	$\|\theta\|^4\|\theta\|_1^8$
	$\|\theta\|^2 W_{\ell i}^3(W_{jt}W_{jq})$	$\|\theta\|^4\,\mathbb{E}[(W_{jt}^2 W_{jq}^2)W_{\ell i}^3 W_{\ell' i'}^3] \leq C\|\theta\|^4\theta_i\theta_j^2\theta_\ell\theta_t\theta_q\theta_{i'}\theta_{\ell'}$	$\|\theta\|^6\|\theta\|_1^6$
	$\|\theta\|^2(W_{\ell i}^2 W_{is})W_{jt}^2$	$\|\theta\|^4\,\mathbb{E}[(W_{is}^2)W_{\ell i}^2 W_{jt}^2 W_{\ell' i}^2 W_{j't'}^2] \leq C\|\theta\|^4\theta_i^3\theta_j\theta_\ell\theta_s\theta_t\theta_{j'}\theta_{\ell'}\theta_{t'}$	$\|\theta\|^4\|\theta\|_3^3\|\theta$
	$\|\theta\|^2 W_{\ell i}^2 W_{ij}^3$	$\|\theta\|^4\,\mathbb{E}[W_{\ell i}^2 W_{ij}^3 W_{\ell' i'}^2 W_{i'j'}^3] \leq C\|\theta\|^4\theta_i^2\theta_j\theta_\ell\theta_{i'}^2\theta_{j'}\theta_{\ell'}$	$\|\theta\|^8\|\theta\|_1^4$
	$\|\theta\|^2(W_{\ell i}^2 W_{is})(W_{jt}W_{jq})$	$\|\theta\|^4\,\mathbb{E}[(W_{is}^2 W_{jt}^2 W_{jq}^2)W_{\ell i}^2 W_{\ell' i}^2] \leq C\|\theta\|^4\theta_i^3\theta_j^2\theta_\ell\theta_s\theta_t\theta_q\theta_{\ell'}$	$\|\theta\|^6\|\theta\|_3^3\|\theta$
J_7	$\|\theta\|^2 W_{\ell i}^2 W_{ij}^2 W_{jq}$	$\|\theta\|^4\,\mathbb{E}[(W_{jq}^2)W_{\ell i}^2 W_{ij}^2 W_{\ell' i'}^2 W_{i'j}^2] \leq C\|\theta\|^4\theta_i^2\theta_j^3\theta_\ell\theta_q\theta_{i'}^2\theta_{\ell'}$	$\|\theta\|^8\|\theta\|_3^3\|\theta$
	$\|\theta\|^2(W_{\ell i}W_{is}W_{\ell m})W_{jt}^2$	$\|\theta\|^4\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{\ell m}^2)W_{jt}^2 W_{j't'}^2] \leq C\|\theta\|^4\theta_i^2\theta_j\theta_\ell^2\theta_s\theta_t\theta_m\theta_{j'}\theta_{t'}$	$\|\theta\|^8\|\theta\|_1^6$
	$\|\theta\|^2 W_{\ell i}W_{ij}^3 W_{\ell m}$	$\|\theta\|^4\,\mathbb{E}[(W_{\ell i}^2 W_{\ell m}^2)W_{ij}^3 W_{ij'}^3] \leq C\|\theta\|^4\theta_i^3\theta_j\theta_\ell^2\theta_m\theta_{j'}$	$\|\theta\|^6\|\theta\|_3^3\|\theta$
	$\|\theta\|^2(W_{\ell i}W_{is}W_{\ell m})(W_{jt}W_{jq})$	$\|\theta\|^4\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2 W_{\ell m}^2 W_{jt}^2 W_{jq}^2)] \leq C\|\theta\|^4\theta_i^2\theta_j^2\theta_\ell^2\theta_s\theta_t\theta_q\theta_m$	$\|\theta\|^{10}\|\theta\|_1^4$
	$\|\theta\|^2 W_{\ell i}W_{ij}^2 W_{\ell m}W_{jq}$	$\|\theta\|^4\,\mathbb{E}[(W_{\ell i}^2 W_{\ell m}^2 W_{jq}^2)W_{ij}^4] \leq C\|\theta\|^4\theta_i^2\theta_j^2\theta_\ell^2\theta_q\theta_m$	$\|\theta\|^{10}\|\theta\|_1^2$
	$\|\theta\|^2 W_{\ell i}W_{ij}^2 W_{\ell j}^2$	$\|\theta\|^4\,\mathbb{E}[(W_{\ell i}^2)W_{ij}^2 W_{\ell j}^2 W_{ij'}^2 W_{\ell j'}^2] \leq C\|\theta\|^4\theta_i^3\theta_j^2\theta_\ell^3\theta_{j'}^2$	$\|\theta\|^8\|\theta\|_3^6$
	$\eta_k\eta_\ell W_{ij}^4 W_{k\ell}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{k\ell}^2)W_{ij}^4 W_{i'j'}^4] \leq C\theta_i\theta_j\theta_k^3\theta_\ell^3\theta_{i'}\theta_{j'}$	$\|\theta\|_3^6\|\theta\|_1^4$
J_8	$\eta_k\eta_\ell(W_{ij}^3 W_{is})W_{k\ell}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{is}^2 W_{k\ell}^2)W_{ij}^3 W_{ij'}^3] \leq C\theta_i^3\theta_j\theta_k^3\theta_\ell^3\theta_s\theta_{j'}$	$\|\theta\|_3^9\|\theta\|_1^3$
	$\eta_k\eta_\ell(W_{ij}^2 W_{is}W_{jq})W_{k\ell}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{is}^2 W_{jq}^2 W_{k\ell}^2)W_{ij}^4] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^3\theta_s\theta_q$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell(W_{is}W_{it}W_{jq}W_{jm})W_{k\ell}$	$\eta_k^2\eta_\ell^2\,\mathbb{E}[(W_{is}^2 W_{it}^2 W_{jq}^2 W_{jm}^2 W_{k\ell}^2)] \leq C\theta_i^2\theta_j^2\theta_k^3\theta_\ell^3\theta_s\theta_t\theta_q\theta_m$	$\|\theta\|^4\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{is}^2 W_{jq}W_{jm}W_{k\ell}$	$\eta_k^2\eta_\ell^2\mathbb{E}[(W_{jq}^2 W_{jm}^2 W_{k\ell}^2)W_{is}^2 W_{i's'}^2] \leq C\theta_i\theta_j^2\theta_k^3\theta_\ell^3\theta_s\theta_q\theta_m\theta_{i'}\theta_{s'}$	$\|\theta\|^2\|\theta\|_3^6\|\theta$
	$\eta_k\eta_\ell W_{is}^2 W_{jq}^2 W_{k\ell}$	$\eta_k^2\eta_\ell^2\mathbb{E}[(W_{k\ell}^2)W_{is}^2 W_{jq}^2 W_{i's'}^2 W_{j'q'}^2] \leq C\theta_i\theta_j\theta_k^3\theta_\ell^3\theta_s\theta_q\theta_{i'}\theta_{j'}\theta_{s'}\theta_{q'}$	$\|\theta\|_3^6\|\theta\|_1^8$

Bernstein's inequality, for large n, the probability of E_i^c is $O(n^{-C_0/4.1})$. Applying the probability union bound, we find that the probability of E^c is $O(n^{-C_0/2.01})$. Recall that $V = \sum_{i,j:i\neq j} A_{ij}$. On the event E^c, if $V = 0$ (i.e., the network has no edges), then

$\widetilde{Q}_n^* = Q_n^* = 0$; otherwise, $V \geq 1$ and $|\widetilde{Q}_n^* - Q_n^*| \leq n^4$. Combining these results gives
$$\mathbb{E}\big[|\widetilde{Q}_n^* - Q_n^*|^2 \cdot I_{E^c}\big] \leq n^4 \cdot O(n^{-C_0/2.01}).$$
With an properly large C_0, the right hand side is $o(\|\theta\|^8)$. Hence, it suffices to focus on the event E. On the event E,
$$\begin{aligned}
|J_9| &\leq \sum_{i,j,k,\ell} |\eta_k \widetilde{\Omega}_{\ell i}||G_i G_j^2 G_k G_\ell| \\
&\leq C \sum_{i,j,k,\ell} (\alpha\theta_i\theta_k\theta_\ell) \frac{\sqrt{\theta_i\theta_j^2\theta_k\theta_\ell\|\theta\|_1^5[\log(n)]^5}}{\sqrt{v^5}} \\
&\leq \frac{C\alpha[\log(n)]^{5/2}}{\sqrt{\|\theta\|_1^5}} \Big(\sum_i \theta_i^{3/2}\Big)\Big(\sum_j \theta_j\Big)\Big(\sum_k \theta_k^{3/2}\Big)\Big(\sum_\ell \theta_\ell^{3/2}\Big) \\
&\leq \frac{C\alpha[\log(n)]^{5/2}}{\sqrt{\|\theta\|_1^3}} \Big(\sum_i \theta_i^{3/2}\Big)^3 \\
&\leq \frac{C\alpha[\log(n)]^{5/2}}{\sqrt{\|\theta\|_1^3}} \Big(\sum_i \theta_i^{2}\Big)^{3/2}\Big(\sum_i \theta_i\Big)^{3/2} \\
&\leq C\alpha[\log(n)]^{5/2}\|\theta\|^3,
\end{aligned}$$
where the second last line is from the Cauchy-Schwarz inequality. Since $\|\theta\| \gg \alpha[\log(n)]^{5/2}$, the right hand side is $o(\|\theta\|^4)$, which implies that $|J_9|^2 = o(\|\theta\|^8)$. Similarly, on the event E,
$$\begin{aligned}
|J_{10}| &\leq \sum_{i,j,k,\ell} |\eta_\ell \widetilde{\Omega}_{\ell i}||G_i G_j^2 G_k^2| \\
&\leq C \sum_{i,j,k,\ell} (\alpha\theta_i\theta_\ell^2) \frac{\sqrt{\theta_i\theta_j^2\theta_k^2\|\theta\|_1^5[\log(n)]^5}}{\sqrt{v^5}} \\
&\leq \frac{C\alpha[\log(n)]^{5/2}}{\sqrt{\|\theta\|_1^5}} \Big(\sum_i \theta_i^{3/2}\Big)\Big(\sum_j \theta_j\Big)\Big(\sum_k \theta_k\Big)\Big(\sum_\ell \theta_\ell^{2}\Big) \\
&\leq \frac{C\alpha[\log(n)]^{5/2}}{\sqrt{\|\theta\|_1^5}} \big(\|\theta\|\sqrt{\|\theta\|_1}\big)\|\theta\|_1^2\|\theta\|^2 \\
&\leq C\alpha[\log(n)]^{5/2}\|\theta\|^3;
\end{aligned}$$
again, the right hand side is $o(\|\theta\|^4)$. Combining the above gives
$$\max\big\{\mathbb{E}[J_9^2],\, \mathbb{E}[J_{10}^2]\big\} = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$
So far, we have proved: for each R_k with $N_W^* = 5$, it satisfies $\mathbb{E}[R_k^2] = o(\|\theta\|^8)$. This is sufficient to guarantee (2.E.302)-(2.E.303) for $X = R_k$.

.There are 7 such terms, including R_{19}-R_{20}, R_{23}-R_{24}, R_{29}-R_{30}, and R_{32}. We plug in the definition of $\widetilde{r}_{ij}$ and δ_{ij} and neglect all factors of $\frac{v}{V}$ (see the explanation in (2.E.302)-

(2.E.303)). It gives ($G_i = \widetilde{\eta}_i - \eta_i$):

$$
\begin{aligned}
R_{19} &= \sum_{i,j,k,\ell(dist)} G_i G_j^2 G_k W_{k\ell} W_{\ell i},\\
R_{20} &= \sum_{i,j,k,\ell(dist)} G_i G_j W_{jk} G_k G_\ell W_{\ell i},\\
R_{23} &= \sum_{i,j,k,\ell(dist)} G_i G_j^2 G_k (\eta_k G_\ell^2 \eta_i + 2G_k \eta_\ell G_\ell \eta_i + G_k \eta_\ell^2 G_i)\\
&= \sum_{i,j,k,\ell(dist)} \eta_i\eta_k G_i G_j^2 G_k G_\ell^2 + 2\sum_{i,j,k,\ell(dist)} \eta_i\eta_\ell G_i G_j^2 G_k^2 G_\ell + \sum_{i,j,k,\ell(dist)} \eta_\ell^2 G_i^2 G_j^2 G_k^2\\
&= 3\sum_{i,j,k,\ell(dist)} \eta_i\eta_k G_i G_j^2 G_k G_\ell^2 + \sum_{i,j,k,\ell(dist)} \eta_\ell^2 G_i^2 G_j^2 G_k^2,\\
R_{24} &= \sum_{i,j,k,\ell(dist)} G_i G_j (\eta_j G_k + G_j \eta_k) G_k G_\ell (\eta_\ell G_i + G_\ell \eta_i)\\
&= 4\sum_{i,j,k,\ell(dist)} \eta_j \eta_\ell G_i^2 G_j G_k^2 G_\ell,\\
R_{29} &= \sum_{i,j,k,\ell(dist)} G_i G_j^2 G_k (\eta_k G_\ell + G_k \eta_\ell) W_{\ell i}\\
&= \sum_{i,j,k,\ell(dist)} \eta_k G_i G_j^2 G_k G_\ell W_{\ell i} + \sum_{i,j,k,\ell(dist)} \eta_\ell G_i G_j^2 G_k^2 W_{\ell i},\\
R_{30} &= 2\sum_{i,j,k,\ell(dist)} G_i G_j (\eta_j G_k) G_k G_\ell W_{\ell i} = 2\sum_{i,j,k,\ell(dist)} \eta_j G_i G_j G_k^2 G_\ell W_{\ell i},\\
R_{32} &= \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{\ell i} G_i G_j^2 G_k^2 G_\ell.
\end{aligned}
$$

Each expression above belongs to one of the following types:

$$
\begin{aligned}
K_1 &= \sum_{i,j,k,\ell(dist)} G_i G_j^2 G_k W_{k\ell} W_{\ell i}, & K_2 &= \sum_{i,j,k,\ell(dist)} G_i G_j G_k G_\ell W_{jk} W_{\ell i},\\
K_3 &= \sum_{i,j,k,\ell(dist)} \eta_k G_i G_j^2 G_k G_\ell W_{\ell i}, & K_4 &= \sum_{i,j,k,\ell(dist)} \eta_\ell G_i G_j^2 G_k^2 W_{\ell i},\\
K_5 &= \sum_{i,j,k,\ell(dist)} \eta_i \eta_k G_i G_j^2 G_k G_\ell^2, & K_5' &= \sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ik} G_i G_j^2 G_k G_\ell^2,\\
K_6 &= \sum_{i,j,k,\ell(dist)} \eta_\ell^2 G_i^2 G_j^2 G_k^2.
\end{aligned}
$$

Since $|\eta_i\eta_k| \le C\theta_i\theta_k$ and $|\widetilde{\Omega}_{ik}| \le C\alpha\theta_i\theta_k$, the study of K_5 and K_5' are similar; we thus omit the analysis of K_5'. We now study K_1-K_6.

Analysis of post-expansion sums with $N_W^* = 6$ Consider K_1. Re-write

$$
K_1 = \frac{1}{v^2} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq i, t\neq j, q\neq j, m\neq k}} W_{is} W_{jt} W_{jq} W_{km} W_{k\ell} W_{\ell i}.
$$

Note that $W_{km}W_{k\ell}W_{\ell i}W_{is}$ has four different cases: (a) $W_{k\ell}^2 W_{\ell i}^2$, (b) $W_{k\ell}^2 W_{\ell i} W_{is}$, (c) $W_{k\ell} W_{\ell i} W_{ik}^2$, and (d) $W_{k\ell}W_{\ell i}W_{km}W_{is}$. At the same time, $W_{jt}W_{jq}$ has two cases: (i) W_{jk}^2 and (ii) $W_{jt}W_{jq}$.

This gives at least $4 \times 2 = 8$ cases. Each case may have sub-cases, e.g., for $(W_{k\ell}^2 W_{\ell i} W_{is}) W_{jt}^2$, if $(s,t) = (j,i)$, it becomes $W_{k\ell}^2 W_{\ell i} W_{ij}^3$. By direct calculations, all possible cases of the summand are as follows:

$$\begin{aligned}
&(W_{k\ell}^2 W_{\ell i}^2) W_{jt}^2, \quad (W_{k\ell}^2 W_{\ell i}^2)(W_{jt} W_{jq}), \quad (W_{k\ell}^2 W_{\ell i} W_{is}) W_{jt}^2, \\
&W_{k\ell}^2 W_{\ell i} W_{ij}^3, \quad (W_{k\ell}^2 W_{\ell i} W_{is})(W_{jt} W_{jq}), \quad W_{k\ell}^2 W_{\ell i} W_{ij}^2 W_{jq}, \\
&(W_{k\ell} W_{\ell i} W_{ik}^2) W_{jt}^2, \quad (W_{k\ell} W_{\ell i} W_{ik}^2)(W_{jt} W_{jq}), \\
&(W_{k\ell} W_{\ell i} W_{km} W_{is}) W_{jt}^2, \quad W_{k\ell} W_{\ell i} W_{km} W_{ij}^3, \\
&(W_{k\ell} W_{\ell i} W_{km} W_{is})(W_{jt} W_{jq}), \quad W_{k\ell} W_{\ell i} W_{km} W_{ij}^2 W_{jq}, \\
&W_{k\ell} W_{\ell i} W_{kj}^2 W_{ij}^2.
\end{aligned} \tag{2.E.308}$$

Take the second type for example. We aim to bound $\mathbb{E}[(\sum_{i,j,k,\ell,t,q} W_{k\ell}^2 W_{\ell i}^2 W_{jt} W_{jq})^2]$, which is equal to $\sum_{\substack{i,j,k,\ell,t,q \\ i',j',k',\ell',t',q'}} \mathbb{E}[W_{k\ell}^2 W_{\ell i}^2 W_{jt} W_{jq} W_{k'\ell'}^2 W_{\ell' i'}^2 W_{j't'} W_{j'q'}]$. For the expectation to be nonzero, each single W term has to be paired with another term. The main contribution comes from the case that $W_{j't'} W_{j'q'} = W_{jt} W_{jq}$. It implies $(j', t', q') = (j, t, q)$ or $(j', t', q') = (j, q, t)$. Then, the expression above becomes

$$\begin{aligned}
\sum_{\substack{i,j,k,\ell,t,q \\ i',k',\ell'}} \mathbb{E}[(W_{jt}^2 W_{jq}^2) W_{k\ell}^2 W_{\ell i}^2 W_{k'\ell'}^2 W_{\ell' i'}^2] &\le C \sum_{\substack{i,j,k,\ell,t,q \\ i',k',\ell'}} \theta_i \theta_j^2 \theta_k \theta_\ell^2 \theta_t \theta_q \theta_{i'} \theta_{k'} \theta_{\ell'}^2 \\
&\le C \|\theta\|^6 \|\theta\|_1^6.
\end{aligned}$$

There are a total of 9 indices in this sum, which are $(i, j, k, \ell, t, q, i', k', \ell')$. Similarly, for each type of summand, when we bound the expectation of the square of its sum, we count how many indices appear in the ultimate sum. This number equals to twice of the total number of indices appearing in the summand, minus the total number of indices appearing in single W terms. For the above example, all indices appearing in the summand are (i, j, k, ℓ, t, q), while indices appearing in single W terms are (j, t, q); so, the aforementioned number is $2 \times 6 - 3 = 9$. If this number if m_0, then the expectation of the square of sum of this type is bounded by $C\|\theta\|_1^{m_0}$. We note that K_1 has a factor $\frac{1}{v^2}$ in front of the sum, which brings in a factor of $\frac{C}{\|\theta\|_1^8}$ in the bound. Therefore, for any type of summand with $m_0 \le 8$, the expectation of the square of its sum is $O(1)$, which is $o(\|\theta\|^8)$. As a result, among the types in (2.E.308), we only need to consider those with $m_0 \ge 9$. We are left with

$$(W_{k\ell}^2 W_{\ell i}^2) W_{jt}^2, \qquad (W_{k\ell}^2 W_{\ell i}^2)(W_{jt} W_{jq}), \qquad (W_{k\ell}^2 W_{\ell i} W_{is}) W_{jt}^2.$$

We have proved that the expectation of the square of sum of the second type of summands is bounded by $C\|\theta\|^2\|\theta\|_1^6 = o(\|\theta\|^8\|\theta\|_1^8)$. For the other two types, by direct calculations,

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ t\neq j}} W_{k\ell}^2 W_{\ell i}^2 W_{jt}^2\Big)^2\Big] &\leq \sum_{\substack{i,j,k,\ell,t\\ i',j',k',\ell',t'}} \mathbb{E}[W_{k\ell}^2 W_{\ell i}^2 W_{jt}^2 W_{k'\ell'}^2 W_{\ell' i'}^2 W_{j't'}^2] \\
&\leq \sum_{\substack{i,j,k,\ell,t\\ i',j',k',\ell',t'}} \theta_i\theta_j\theta_k\theta_\ell^2\theta_t\theta_{i'}\theta_{j'}\theta_{k'}\theta_{\ell'}^2\theta_{t'} \\
&\leq C\|\theta\|^4\|\theta\|_1^8 = o(\|\theta\|^8\|\theta\|_1^8), \\
\mathbb{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ s\notin\{i,\ell\},t\neq j,\\ (s,t)\neq(j,i)}} W_{k\ell}^2 W_{\ell i} W_{is} W_{jt}^2\Big)^2\Big] &\leq \sum_{\substack{i,j,k,\ell,s,t\\ j',k',t'}} \mathbb{E}[(W_{\ell i}^2 W_{is}^2) W_{k\ell}^2 W_{jt}^2 W_{k'\ell}^2 W_{j't'}^2] \\
&\leq C\sum_{\substack{i,j,k,\ell,s,t\\ j',k',t'}} \theta_i^2\theta_j\theta_k\theta_\ell^3\theta_s\theta_t\theta_{j'}\theta_{k'}\theta_{t'} \\
&\leq C\|\theta\|^2\|\theta\|_3^3\|\theta\|_1^7 = o(\|\theta\|^8\|\theta\|_1^8).
\end{aligned}$$

Combining the above gives

$$\mathbb{E}[K_1^2] = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Consider K_2. Re-write

$$K_2 = \frac{1}{v^2}\sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq k,m\neq \ell}} W_{is}W_{jt}W_{kq}W_{\ell m}W_{jk}W_{\ell i}.$$

Note that $W_{qk}W_{kj}W_{jt}$ has three cases: (a) W_{kj}^3, (b) $W_{kj}^2W_{jt}$ (or $W_{qk}W_{kj}^2$), and (c) $W_{qk}W_{kj}W_{jt}$. Simiarly, $W_{m\ell}W_{\ell i}W_{is}$ has three cases: (a) $W_{\ell i}^3$, (b) $W_{\ell i}^2W_{is}$ (or $W_{m\ell}W_{\ell i}^2$), and (c) $W_{m\ell}W_{\ell i}W_{is}$. By index symmetry, this gives $3+2+1=6$ different cases. Some case may have sub-cases, due to that (s,t) may equal to (j,i), say. By direct calculations, all possible cases of the summand are as follows:

$$\begin{aligned}
&W_{kj}^3W_{\ell i}^3, \quad W_{kj}^3(W_{\ell i}^2W_{is}), \quad W_{kj}^3(W_{m\ell}W_{\ell i}W_{is}), \quad (W_{kj}^2W_{jt})(W_{\ell i}^2W_{is}),\\
&W_{kj}^2W_{ji}^2W_{\ell i}^2, \quad (W_{kj}^2W_{jt})(W_{m\ell}W_{\ell i}W_{is}), \quad W_{kj}^2W_{ji}^2W_{m\ell}W_{\ell i},\\
&(W_{qk}W_{kj}W_{jt})(W_{m\ell}W_{\ell i}W_{is}), \quad W_{qk}W_{kj}W_{ji}^2W_{m\ell}W_{\ell i}, \quad W_{kj}W_{ji}^2W_{k\ell}^2W_{\ell i}.
\end{aligned}$$

As in the analysis of (2.E.308), we count the effective number of indices, m_0, which equals to twice of the total number of indices appearing in the summand minus the total number of indices appearing in all single-W terms. For the above types of summand, m_0 equals to $8, 8, 8, 8, 8, 8, 7, 8, 6, 4$, respectively. None is larger than 8. We conclude that the expectation of the square of sum of each type of summand is bounded by $C\|\theta\|_1^8$. We immediately have

$$\mathbb{E}[K_2^2] = \frac{1}{v^4}\cdot C\|\theta\|_1^8 = O(1) = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Consider K_3. Re-write

$$K_3 = \frac{1}{v^2\sqrt{v}}\sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq j,m\neq k,p\neq \ell}} \eta_k W_{is}W_{jt}W_{jq}W_{km}W_{\ell p}W_{\ell i}$$

Note that $W_{jt}W_{jq}W_{km}$ has four cases: (a) W_{jk}^3, (b) $W_{jk}^2 W_{jt}$ (or $W_{jk}^2 W_{jq}$), (c) $W_{jt}^2 W_{km}$, and (d) $W_{jt}W_{jq}W_{km}$. At the same time, $W_{is}W_{\ell p}W_{\ell i}$ has three cases: (a) $W_{\ell i}^3$, (b) $W_{\ell i}^2 W_{is}$ (or $W_{\ell i}^2 W_{\ell p}$), and (c) $W_{\ell i}W_{is}W_{\ell p}$. This gives $4 \times 3 = 12$ different cases. Each case may have sub-cases. For example, in the case of $\eta_k(W_{jk}^2 W_{jt})(W_{\ell i}^2 W_{is})$, if $(s,t) = (j,i)$, it becomes $\eta_k W_{jk}^2 W_{ji}^2 W_{\ell i}^2$. By direct calculations, we obtain all possible cases of summands as follows:

$$
\begin{aligned}
&\eta_k W_{jk}^3 W_{\ell i}^3, \quad \eta_k W_{jk}^3(W_{\ell i}^2 W_{is}), \quad \eta_k W_{jk}^3(W_{\ell i}W_{is}W_{\ell p}), \quad \eta_k(W_{jk}^2 W_{jt})W_{\ell i}^3,\\
&\eta_k(W_{jk}^2 W_{jt})(W_{\ell i}^2 W_{is}), \quad \eta_k W_{jk}^2 W_{ji}^2 W_{\ell i}^2, \quad \eta_k(W_{jk}^2 W_{jt})(W_{\ell i}W_{is}W_{\ell p}),\\
&\eta_k W_{jk}^2 W_{ji}^2 W_{\ell i}W_{\ell p}, \quad \eta_k(W_{jt}^2 W_{km})W_{\ell i}^3, \quad \eta_k(W_{jt}^2 W_{km})(W_{\ell i}^2 W_{is}), \quad \eta_k W_{jt}^2 W_{ki}^2 W_{\ell i}^2,\\
&\eta_k(W_{jt}^2 W_{km})(W_{\ell i}W_{is}W_{\ell p}), \quad \eta_k W_{jt}^2 W_{ki}^2 W_{\ell i}W_{\ell p}, \quad \eta_k(W_{jt}W_{jq}W_{km})W_{\ell i}^3,\\
&\eta_k(W_{jt}W_{jq}W_{km})(W_{\ell i}^2 W_{is}), \quad \eta_k W_{jt}W_{ji}^2 W_{km}W_{\ell i}^2, \quad \eta_k W_{jt}W_{jq}W_{ki}^2 W_{\ell i}^2,\\
&\eta_k(W_{jt}W_{jq}W_{km})(W_{\ell i}W_{is}W_{\ell p}), \quad \eta_k W_{jt}W_{ji}^2 W_{km}W_{\ell i}W_{\ell p}, \quad \eta_k W_{jt}W_{jq}W_{ki}^2 W_{\ell i}W_{\ell p}.
\end{aligned}
$$

Same as before, let m_0 be the effective number of indices for each type of summand, which equals to twice of number of distinct indices appearing in the summand minus the number of distinct indices appearing in single-W terms (see (2.E.308) and text therein). By direct calculations, $m_0 \leq 10$ for all types above. It follows that, for each type of summand, the expectation of the square of their sums is bounded by

$$
\frac{1}{(v\sqrt{v})^2} \cdot C\|\theta\|_1^{m_0} \leq C\|\theta\|_1^{m_0-10} = O(1) = o(\|\theta\|^8).
$$

We immediately have

$$
\mathbb{E}[K_3^2] = o(\|\theta\|^8), \qquad \text{under both hypotheses.}
$$

Consider K_4. Re-write

$$
K_4 = \frac{1}{v^2\sqrt{v}} \sum_{\substack{i,j,k,\ell (dist)\\ s,t,q,m,p}} \eta_\ell W_{is}W_{jt}W_{jq}W_{km}W_{kp}W_{\ell i}.
$$

Note that $W_{is}W_{\ell i}$ has two cases: (a) $W_{\ell i}^2$ and (b) $W_{\ell i}W_{is}$. Moreover, there are a total of six cases for $W_{jt}W_{jq}W_{km}W_{kp}$: (a) W_{jk}^4, (b) $W_{jk}^3 W_{jt}$, (c) $W_{jk}^2 W_{jt}W_{km}$, (d) $W_{jt}^2 W_{km}^2$, (e) $W_{jt}W_{jq}W_{km}^2$, and (f) $W_{jt}W_{jq}W_{km}W_{kp}$. It gives $2 \times 6 = 12$ different cases. Each case may have some sub-cases. It turns out all different types of summand are as follows:

$$
\begin{aligned}
&\eta_\ell W_{\ell i}^2 W_{jk}^4, \quad \eta_\ell W_{\ell i}^2(W_{jk}^3 W_{jt}), \quad \eta_\ell W_{\ell i}^2(W_{jk}^2 W_{jt}W_{km}), \quad \eta_\ell W_{\ell i}^2(W_{jt}^2 W_{km}^2),\\
&\eta_\ell W_{\ell i}^2(W_{jt}W_{jq}W_{km}^2), \quad \eta_\ell W_{\ell i}^2(W_{jt}W_{jq}W_{km}W_{kp}), \quad \eta_\ell(W_{\ell i}W_{is})W_{jk}^4,\\
&\eta_\ell(W_{\ell i}W_{is})(W_{jk}^3 W_{jt}), \quad \eta_\ell W_{\ell i}W_{jk}^3 W_{ji}^2, \quad \eta_\ell(W_{\ell i}W_{is})(W_{jk}^2 W_{jt}W_{km}),\\
&\eta_\ell W_{\ell i}W_{jk}^2 W_{ji}^2 W_{km}, \quad \eta_\ell(W_{\ell i}W_{is})(W_{jt}^2 W_{km}^2), \quad \eta_\ell W_{\ell i}W_{ij}^3 W_{km}^2,\\
&\eta_\ell(W_{\ell i}W_{is})(W_{jt}W_{jq}W_{km}^2), \quad \eta_\ell W_{\ell i}W_{ij}^2 W_{jq}W_{km}^2, \quad \eta_\ell W_{\ell i}W_{jt}W_{jq}W_{ki}^3,\\
&\eta_\ell(W_{\ell i}W_{is})(W_{jt}W_{jq}W_{km}W_{kp}), \quad \eta_\ell W_{\ell i}W_{ij}^2 W_{jq}W_{km}W_{kp}.
\end{aligned}
$$

Same as before, for each type, let m_0 be the effective number of indices. It suffices to focus on cases where $m_0 \geq 11$. We are left with

$$
\eta_\ell W_{\ell i}^2(W_{jt}^2 W_{km}^2), \qquad \eta_\ell W_{\ell i}^2(W_{jt}W_{jq}W_{km}^2), \qquad \eta_\ell(W_{\ell i}W_{is})(W_{jt}^2 W_{km}^2).
$$

By direct calculations,

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ t\neq j, m\neq k}} \eta_\ell W_{\ell i}^2 W_{jt}^2 W_{km}^2\Big)\Big] &\leq \sum_{\substack{i,j,k,\ell,t,m\\ i',j',k',\ell',t',m'}} \eta_\ell\eta_{\ell'}\,\mathbb{E}[W_{\ell i}^2 W_{jt}^2 W_{km}^2 W_{\ell' i'}^2 W_{j' t'}^2 W_{k' m'}^2]\\
&\leq C \sum_{\substack{i,j,k,\ell,t,m\\ i',j',k',\ell',t',m'}} \theta_i\theta_j\theta_k\theta_\ell^2\theta_t\theta_m\theta_{i'}\theta_{j'}\theta_{k'}\theta_{\ell'}^2\theta_{t'}\theta_{m'}\\
&\leq C\|\theta\|^4\|\theta\|_1^{10} = o(\|\theta\|^8\|\theta\|_1^{10}),\\
\mathbb{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ t\neq j, q\neq j, m\neq k\\ t\neq q}} \eta_\ell W_{\ell i}^2 W_{jt} W_{jq} W_{km}^2\Big)\Big] &\leq \sum_{\substack{i,j,k,\ell,t,q,m\\ i',k',\ell',m'}} \eta_\ell\eta_{\ell'}\,\mathbb{E}[(W_{jt}^2 W_{jq}^2) W_{\ell i}^2 W_{km}^2 W_{\ell' i'}^2 W_{k' m'}^2]\\
&\leq C \sum_{\substack{i,j,k,\ell,t,q,m\\ i',k',\ell',m'}} \theta_i\theta_j^2\theta_k\theta_\ell^2\theta_t\theta_q\theta_m\theta_{i'}\theta_{k'}\theta_{\ell'}^2\theta_{m'}\\
&\leq C\|\theta\|^6\|\theta\|_1^{8} = o(\|\theta\|^8\|\theta\|_1^{10}),\\
\mathbb{E}\Big[\Big(\sum_{\substack{i,j,k,\ell(dist)\\ s\neq i, t\neq j, m\neq k\\ (s,t)\neq(j,i), (s,m)\neq(k,i)}} \eta_\ell W_{\ell i} W_{is} W_{jt}^2 W_{km}^2\Big)\Big] &\leq C \sum_{\substack{i,j,k,\ell,s,t,m\\ j',k',t',m'}} \eta_\ell^2\,\mathbb{E}[(W_{\ell i}^2 W_{is}^2) W_{jt}^2 W_{km}^2 W_{j' t'}^2 W_{k' m'}^2]\\
&\leq C \sum_{\substack{i,j,k,\ell,s,t,m\\ j',k',t',m'}} \theta_i^2\theta_j\theta_k\theta_\ell^3\theta_s\theta_t\theta_m\theta_{j'}\theta_{k'}\theta_{t'}\theta_{m'}\\
&\leq C\|\theta\|^2\|\theta\|_3^3\|\theta\|_1^{9} = o(\|\theta\|^8\|\theta\|_1^{10}).
\end{aligned}$$

It follows that

$$\mathbb{E}[K_4^2] \leq \frac{1}{(v^2\sqrt{v})^2}\cdot o(\|\theta\|^8\|\theta\|_1^{10}) = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Consider K_5-K_6. To save space, we only present the proof for the case of $\|\theta\| \gg [\log(n)]^{3/2}$. When $1 \ll \|\theta\| \leq C[\log(n)]^{3/2}$, we can bound $\mathbb{E}[K_5^2]$ and $\mathbb{E}[K_6^2]$ in the same way as in the study of J_1-J_8, so the proof is omitted. Let E be the event defined in (2.E.307). We have argued that it suffices to focus on the event E. On this event, $|G_i| \leq C\sqrt{\theta_i\|\theta\|_1\log(n)/v}$. It follows that

$$\begin{aligned}
|K_5| &\leq C\sum_{i,j,k,\ell}(\theta_i\theta_k)\frac{\sqrt{\theta_i\theta_j^2\theta_k\theta_\ell^2}\|\theta\|_1^3[\log(n)]^3}{v^3}\\
&\leq \frac{C[\log(n)]^3}{\|\theta\|_1^3}\Big(\sum_i\theta_i^{3/2}\Big)\Big(\sum_j\theta_j\Big)\Big(\sum_k\theta_k^{3/2}\Big)\Big(\sum_\ell\theta_\ell\Big)\\
&\leq \frac{C[\log(n)]^3}{\|\theta\|_1^3}\big(\|\theta\|\sqrt{\|\theta\|_1}\big)^2\|\theta\|_1^2\\
&\leq C[\log(n)]^3\|\theta\|^2,
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality $(\sum_i \theta_i^{3/2}) \leq \|\theta\|\sqrt{\|\theta\|_1}$. Similarly,

$$\begin{aligned}|K_6| &\leq C \sum_{i,j,k,\ell} \theta_\ell^2 \cdot \frac{\theta_i\theta_j\theta_k\|\theta\|_1^3[\log(n)]^3}{v^3} \\ &\leq \frac{C[\log(n)]^3}{\|\theta\|_1^3} \sum_{i,j,k,\ell} \theta_i\theta_j\theta_k\theta_\ell^2 \\ &\leq C[\log(n)]^3\|\theta\|^2.\end{aligned}$$

When $\|\theta\| \gg [\log(n)]^{3/2}$, both right hand sides are $o(\|\theta\|^4)$. We immediately have

$$\max\{\mathbb{E}[K_5^2],\ \mathbb{E}[K_6^2]\} = o(\|\theta\|^8).$$

We have proved: Each R_k with $N_W^* = 6$ satisfies $\mathbb{E}[R_k^2] = o(\|\theta\|^8)$. This is sufficient to guarantee (2.E.302)-(2.E.303) for $X = R_k$.

.There are 3 such terms, R_{31}, R_{33} and R_{34}. Consider R_{31}. By definition,

$$R_{31} = \sum_{i,j,k,\ell(dist)} G_iG_j^2G_k^2G_\ell W_{\ell i} = \frac{1}{v^3} \sum_{\substack{i,j,k,\ell(dist)\\ s\neq i,t\neq j,q\neq j,\\ m\neq k,p\neq k,y\neq \ell}} W_{is}W_{jt}W_{jq}W_{km}W_{kp}W_{\ell y}W_{\ell i}.$$

We note that $W_{\ell i}W_{is}W_{\ell y}$ has three cases: (a) $W_{\ell i}^3$, (b) $W_{\ell i}^2W_{is}$, and (c) $W_{\ell i}W_{is}W_{\ell y}$. Moreover, $W_{jt}W_{jq}W_{km}W_{kp}$ has six cases: (a) W_{jk}^4, (b) $W_{jk}^3W_{jt}$, (c) $W_{jk}^2W_{jt}W_{km}$, (d) $W_{jt}^2W_{km}^2$, (e) $W_{jt}W_{jq}W_{km}^2$, and (f) $W_{jt}W_{jq}W_{km}W_{kp}$. This gives $3 \times 6 = 18$ different cases. Since each case may have sub-cases, we end up with the following different types:

$$\begin{aligned}
&W_{\ell i}^3W_{jk}^4, \quad W_{\ell i}^3(W_{jk}^3W_{jt}), \quad W_{\ell i}^3(W_{jk}^2W_{jt}W_{km}), \quad W_{\ell i}^3(W_{jt}^2W_{km}^2),\\
&W_{\ell i}^3(W_{jt}W_{jq}W_{km}^2), \quad W_{\ell i}^3(W_{jt}W_{jq}W_{km}W_{kp}), \quad (W_{\ell i}^2W_{is})W_{jk}^4,\\
&(W_{\ell i}^2W_{is})(W_{jk}^3W_{jt}), \quad W_{\ell i}^2W_{jk}^3W_{ji}^2, \qquad (W_{\ell i}^2W_{is})(W_{jk}^2W_{jt}W_{km}),\\
&W_{\ell i}^2W_{jk}^2W_{ji}^2W_{km}, \quad (W_{\ell i}^2W_{is})(W_{jt}^2W_{km}^2), \quad W_{\ell i}^2W_{ij}^3W_{km}^2,\\
&(W_{\ell i}^2W_{is})(W_{jt}W_{jq}W_{km}^2), \quad W_{\ell i}^2W_{ij}^2W_{jq}W_{km}^2, \quad W_{\ell i}^2W_{jt}W_{jq}W_{ki}^3,\\
&(W_{\ell i}^2W_{is})(W_{jt}W_{jq}W_{km}W_{kp}), \quad W_{\ell i}^2W_{ij}^2W_{jq}W_{km}W_{kp},\\
&(W_{\ell i}W_{is}W_{\ell y})W_{jk}^4, \quad (W_{\ell i}W_{is}W_{\ell y})(W_{jk}^3W_{jt}), \quad W_{\ell i}W_{\ell y}W_{jk}^3W_{ji}^2,\\
&(W_{\ell i}W_{is}W_{\ell y})(W_{jk}^2W_{jt}W_{km}), \quad W_{\ell i}W_{\ell y}W_{jk}^2W_{ji}^2W_{km}, \quad W_{\ell i}W_{jk}^2W_{ji}^2W_{k\ell}^2,\\
&(W_{\ell i}W_{is}W_{\ell y})(W_{jt}^2W_{km}^2), \quad W_{\ell i}W_{\ell y}W_{ji}^3W_{km}^2, \quad W_{\ell i}W_{ji}^3W_{k\ell}^3,\\
&(W_{\ell i}W_{is}W_{\ell y})(W_{jt}W_{jq}W_{km}^2), \quad W_{\ell i}W_{\ell y}W_{ji}^2W_{jq}W_{km}^2, \quad W_{\ell i}W_{\ell y}W_{jt}W_{jq}W_{ki}^3,\\
&W_{\ell i}W_{ji}^2W_{jq}W_{ki}^3, \quad (W_{\ell i}W_{is}W_{\ell y})(W_{jt}W_{jq}W_{km}W_{kp}),\\
&W_{\ell i}W_{\ell y}W_{ji}^2W_{jq}W_{km}W_{kp}, \quad W_{\ell i}W_{ji}^2W_{jq}W_{k\ell}^2W_{kp}.
\end{aligned}$$

For each type, we count m_0, the effective number of indices. It equals to twice of the number of distinct indices in the summand, minus the number of distinct indices appearing in all single-W terms. It turns out that $m_0 \leq 12$ for all types above. By similar arguments as in (2.E.308), we conclude that

$$\mathbb{E}[R_{31}^2] \leq \frac{1}{v^6} \cdot C\|\theta\|_1^{m_0} \leq C\|\theta\|_1^{m_0-12} = O(1) = o(\|\theta\|^8), \qquad \text{under both hypotheses.}$$

Analysis of terms with $N_W^* \geq 7$ Consider R_{33}-R_{34}. We only give the proof when $\|\theta\|^6 \gg [\log(n)]^7$, as it is much simpler. In the case of $1 \ll \|\theta\|^6 \leq C[\log(n)]^7$, we can follow similar steps above to obtain desired bounds, where details are omitted. On the event E (see (2.E.307) for definition),

$$
\begin{aligned}
|R_{33}| &\leq \sum_{i,j,k,\ell} |\eta_\ell||G_i^2 G_j^2 G_k^2 G_\ell| \\
&\leq C \sum_{i,j,k,\ell} \theta_\ell \frac{\sqrt{\theta_i^2\theta_j^2\theta_k^2\theta_\ell \|\theta\|_1^7 [\log(n)]^7}}{(\sqrt{v})^7} \\
&\leq \frac{C[\log(n)]^{7/2}}{\sqrt{\|\theta\|_1^7}} \Big(\sum_i \theta_i\Big)\Big(\sum_j \theta_j\Big)\Big(\sum_k \theta_k\Big)\Big(\sum_\ell \theta_\ell^{3/2}\Big) \\
&\leq \frac{C[\log(n)]^{7/2}}{\sqrt{\|\theta\|_1^7}} \cdot \|\theta\|_1^3 (\|\theta\|\sqrt{\|\theta\|_1}) \\
&\leq C[\log(n)]^{7/2}\|\theta\|,
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality $\sum_\ell \theta_\ell^{3/2} \leq \|\theta\|\sqrt{\|\theta\|_1}$ in the second last line. When $\|\theta\|^6 \gg [\log(n)]^7$, the right hand side is $o(\|\theta\|^4)$. Similarly,

$$
\begin{aligned}
|R_{34}| &\leq \sum_{i,j,k,\ell} |G_i^2 G_j^2 G_k^2 G_\ell^2| \\
&\leq C \sum_{i,j,k,\ell} \frac{\theta_i\theta_j\theta_k\theta_\ell \|\theta\|_1^4 [\log(n)]^4}{v^4} \\
&\leq C[\log(n)]^4.
\end{aligned}
$$

When $\|\theta\|^6 \gg [\log(n)]^7$, the right hand side is $o(\|\theta\|^4)$. As we have argued in (2.E.307), the event E^c has a negligible effect. It follows that

$$
\max\big\{\mathbb{E}[R_{31}^2],\ \mathbb{E}[R_{33}^2],\ \mathbb{E}[R_{34}^2]\big\} = o(\|\theta\|^8), \qquad \text{under both hypotheses.}
$$

This is sufficient to guarantee (2.E.302)-(2.E.303) for R_k.

We have analyzed all 34 terms in Table 2.3. The proof is now complete.

Proof of Lemma 26

Consider an arbitrary post-expansion sum of the form

$$
\sum_{i,j,k,\ell (dist)} a_{ij} b_{jk} c_{k\ell} d_{\ell i}, \qquad \text{where} \quad a, b, c, d \in \{\widetilde{\Omega}, W, \delta, \widetilde{r}, \epsilon\}. \tag{2.E.309}
$$

Let $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}}, N_\epsilon)$ be the number of each type in the product, where these numbers have to satisfy $N_{\widetilde{\Omega}} + N_W + N_\delta + N_{\widetilde{r}} + N_\epsilon = 4$. As discussed in Section 2.E.3, $(Q_n - Q_n^*)$ equals to the sum of all post-expansion sums such that $N_\epsilon > 0$. Recall that

$$
\epsilon_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) + (1 - \frac{v}{V})\eta_i\eta_j - (1 - \frac{v}{V})\delta_{ij}.
$$

Define

$$
\epsilon_{ij}^{(1)} = \eta_i^*\eta_j^* - \eta_i\eta_j, \quad \epsilon_{ij}^{(2)} = (1 - \frac{v}{V})\eta_i\eta_j, \quad \epsilon_{ij}^{(3)} = -(1 - \frac{v}{V})\delta_{ij}.
$$

Then, $\epsilon_{ij} = \epsilon_{ij}^{(1)} + \epsilon_{ij}^{(2)} + \epsilon_{ij}^{(3)}$. It follows that each post-expansion sum of the form (2.E.309) can be further expanded as the sum of terms like

$$\sum_{i,j,k,\ell(dist)} a_{ij}b_{jk}c_{k\ell}d_{\ell i}, \quad \text{where} \;\; a,b,c,d \in \{\widetilde{\Omega}, W, \delta, \widetilde{r}, \epsilon^{(1)}, \epsilon^{(2)}, \epsilon^{(3)}\}. \tag{2.E.310}$$

Let $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}})$ have the same meaning as before, and let $N_\epsilon^{(m)}$ be the number of $\epsilon^{(m)}$ term in the product, for $m \in \{1,2,3\}$. These numbers have to satisfy $N_{\widetilde{\Omega}} + N_W + N_\delta + N_{\widetilde{r}} + N_\epsilon^{(1)} + N_\epsilon^{(2)} + N_\epsilon^{(3)} = 4$. Now, $(Q_n - Q_n^*)$ equals to the sum of all post-expansion sums of the form (2.E.310) with

$$N_\epsilon^{(1)} + N_\epsilon^{(2)} + N_\epsilon^{(3)} \geq 1. \tag{2.E.311}$$

Fix such a post-expansion sum and denote it by Y. We shall bound $|\mathbb{E}[Y]|$ and $\mathrm{Var}(Y)$.

We need some preparation. First, we derive a bound for $|\epsilon_{ij}^{(1)}|$. By definition, $\eta_i = (1/\sqrt{v})\sum_{j\neq i}\Omega_{ij}$ and $\eta_i^* = (1/\sqrt{v_0})\sum_j \Omega_{ij}$. It follows that

$$\eta_i^* = \frac{\sqrt{v}}{\sqrt{v_0}}\eta_i + \frac{1}{\sqrt{v_0}}\Omega_{ii}.$$

We then have

$$\eta_i^*\eta_j^* = \frac{v}{v_0}\eta_i\eta_j + \frac{\sqrt{v}}{v_0}(\eta_i\Omega_{jj} + \eta_j\Omega_{ii}) + \frac{1}{v_0}\Omega_{ii}\Omega_{jj}.$$

Note that $v = \sum_{i\neq j}\Omega_{ij}$ and $v_0 = \sum_{ij}\Omega_{ij} \asymp \|\theta\|_1^2$. It follows that $v_0 - v = \sum_i \Omega_{ii} \leq \sum_i \theta_i^2 \leq \|\theta\|^2$. Therefore,

$$\begin{aligned}
|\eta_i^*\eta_j^* - \eta_i\eta_j| &\leq \Big|1 - \frac{v}{v_0}\Big|\eta_i\eta_j + \frac{\sqrt{v}}{v_0}(\eta_i\Omega_{jj} + \eta_j\Omega_{ii}) + \frac{1}{v_0}\Omega_{ii}\Omega_{jj} \\
&\leq \frac{C\|\theta\|^2}{\|\theta\|_1^2}\cdot\theta_i\theta_j + \frac{C}{\|\theta\|_1}(\theta_i\theta_j^2 + \theta_j\theta_i^2) + \frac{C}{\|\theta\|_1^2}\cdot\theta_i^2\theta_j^2 \\
&\leq C\theta_i\theta_j\cdot\Big(\frac{\|\theta\|^2}{\|\theta\|_1^2} + \frac{\theta_i+\theta_j}{\|\theta\|_1} + \frac{\theta_i\theta_j}{\|\theta\|_1^2}\Big).
\end{aligned}$$

Since $\|\theta\|^2 \leq \theta_{\max}\|\theta\|_1$, the term in the brackets is bounded by $C\theta_{\max}/\|\theta\|_1$. We thus have

$$|\epsilon_{ij}^{(1)}| \leq \frac{C\theta_{\max}}{\|\theta\|_1}\cdot\theta_i\theta_j, \qquad \text{for all } 1 \leq i \neq j \leq n. \tag{2.E.312}$$

Second, in Lemmas 15-25, we have studied all post-expansion sums of the form

$$Z \equiv \sum_{i,j,k,\ell(dist)} a_{ij}b_{jk}c_{k\ell}d_{\ell i}, \qquad \text{where} \quad a,b,c,d \in \{\widetilde{\Omega}, W, \delta, \widetilde{r}\},$$

where $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}})$ are the numbers of each type in the product. We hope to take advantage of these results. Using the proved bounds for $|\mathbb{E}[Z]|$ and $\mathrm{Var}(Z)$, we can get

$$\mathbb{E}[Z^2] \leq C(\alpha^2)^{N_{\widetilde{\Omega}}}\cdot f(\theta; N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}}), \tag{2.E.313}$$

where $\alpha = |\lambda_2|/\lambda_1$ and $f(\theta; m_1, m_2, m_3, m_4)$ is a function of θ whose form is determined by

(m_1, m_2, m_3, m_4). For example,

$$\begin{cases} f(\theta; 0,4,0,0) = \|\theta\|^8, & \text{by claims of } X_1 \text{ in Lemmas 15\&17;} \\ f(\theta; 4,0,0,0) = \|\theta\|^{16}, & \text{by claims of } X_6 \text{ in Lemma 17;} \\ f(\theta; 3,1,0,0) = \|\theta\|^8\|\theta\|_3^6, & \text{by claims of } X_5 \text{ in Lemma 17;} \\ f(\theta; 1,2,1,0) = \|\theta\|^4\|\theta\|_3^6, & \text{by claims of } Y_2,\, Y_3 \text{ in Lemma 19;} \\ f(\theta; 1,1,1,1) = \|\theta\|^8, & \text{by claims of } R_9\text{-}R_{11} \text{ in the proof of Lemma 25.} \end{cases}$$

If there are more than one post-expansion sum that corresponds to the same $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}})$, we use the largest bound to define $f(\theta; N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}})$. Thanks to previous lemmas, we have known the function $f(\theta; m_1, m_2, m_3, m_4)$ for all possible (m_1, m_2, m_3, m_4).

We now show the claim. Recall that Y is the post-expansion sum in (2.E.310). The key is to prove the following argument: For any sequence x_n such that $\sqrt{\log(\|\theta\|_1)} \ll x_n \ll \|\theta\|_1$,

$$\mathbb{E}[Y^2] \le C(\alpha^2)^{N_{\widetilde{\Omega}}} \times \Big(\frac{\theta_{\max}^2}{\|\theta\|_1^2}\Big)^{N_\epsilon^{(1)}} \times \Big(\frac{x_n^2}{\|\theta\|_1^2}\Big)^{N_\epsilon^{(2)}+N_\epsilon^{(3)}} \times f(\theta; m_1, m_2, m_3, m_4)\Big|_{\substack{m_1 = N_{\widetilde{\Omega}} + N_\epsilon^{(1)} + N_\epsilon^{(2)},\ m_2 = N_W, \\ m_3 = N_\delta + N_\epsilon^{(3)},\ m_4 = N_{\widetilde{r}},}} \tag{2.E.314}$$

where $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_{\widetilde{r}}, N_\epsilon^{(1)}, N_\epsilon^{(2)}, N_\epsilon^{(3)})$ are the same as in (2.E.310)-(2.E.311), and $f(\theta; m_1, m_2, m_3, m_4)$ is the known function in (2.E.313).

We prove (2.E.314). Let D be the event

$$D = \{|V - v| \le \|\theta\|_1 x_n\}.$$

In Lemma 24, we have proved $\mathbb{E}[(Q_n - Q_n^*)^2 \cdot I_{D^c}] = o(1)$. By similar proof, we can show: when $|Y|$ is bounded by a polynomial of V and $\|\theta\|_1$ (which is always the case here),

$$\mathbb{E}[Y^2 \cdot I_{D^c}] = o(1).$$

It follows that

$$\mathbb{E}[Y^2] \le \mathbb{E}[Y^2 \cdot I_D] + o(1). \tag{2.E.315}$$

We then bound $\mathbb{E}[Y^2 \cdot I_D]$. In the definition of Y, each $\epsilon^{(2)}$ term introduces a factor of $(1 - \frac{v}{V})$, and each $\epsilon^{(3)}$ term introduces a factor of $-(1 - \frac{v}{V})$. We bring all these factors to the front and re-write the post-expansion sum as

$$Y = (-1)^{N_\epsilon^{(3)}}\Big(1 - \frac{v}{V}\Big)^{N_\epsilon^{(2)}+N_\epsilon^{(3)}} X, \qquad X \equiv \sum_{i,j,k,\ell (dist)} a_{ij} b_{jk} c_{k\ell} d_{\ell i}.$$

After the factor $(1 - \frac{v}{V})$ is removed, $\epsilon^{(2)}$ becomes $\eta_i\eta_j$; similarly, $\epsilon^{(3)}$ becomes δ_{ij}. Therefore, in the expression of X,

$$\begin{cases} a_{ij}, b_{ij}, c_{ij}, d_{ij} \in \{\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, \widetilde{r}_{ij}, \epsilon_{ij}^{(1)}, \eta_i\eta_j\}, \\ \text{number of } \eta_i\eta_j \text{ in the product is } N_\epsilon^{(2)}, \\ \text{number of } \delta_{ij} \text{ in the product is } N_\delta + N_\epsilon^{(3)}, \\ \text{number of any other term in the product is same as before.} \end{cases} \tag{2.E.316}$$

On the event D, $|1-\frac{v}{V}| \le \frac{x_n\|\theta\|_1}{C\|\theta\|_1^2} = O(\frac{x_n}{\|\theta\|_1})$. Hence,

$$|Y| \le C\Big(\frac{x_n}{\|\theta\|_1}\Big)^{N_\epsilon^{(2)}+N_\epsilon^{(3)}}|X|, \qquad \text{on the event } D.$$

It follows that

$$\mathbb{E}[Y^2\cdot I_D] \le C\Big(\frac{x_n^2}{\|\theta\|_1^2}\Big)^{N_\epsilon^{(2)}+N_\epsilon^{(3)}}\cdot \mathbb{E}[X^2]. \tag{2.E.317}$$

To bound $\mathbb{E}[X^2]$, we compare X and Z. In obtaining (2.E.313), the only property of $\widetilde{\Omega}$ we have used is

$$|\widetilde{\Omega}_{ij}| \le \alpha\cdot C\theta_i\theta_j.$$

In comparison, in the expression of X, we have (by (2.E.312) and (2.E.117))

$$|\widetilde{\Omega}_{ij}| \le \alpha\cdot C\theta_i\theta_j, \qquad |\epsilon_{ij}^{(1)}| \le \frac{\theta_{\max}}{\|\theta\|_1}\cdot C\theta_i\theta_j, \qquad |\eta_i\eta_j| \le C\theta_i\theta_j. \tag{2.E.318}$$

If we consider $(\alpha^{N_{\widetilde{\Omega}}}\cdot(\frac{\theta_{\max}}{\|\theta\|_1})^{N_\epsilon^{(1)}}\cdot 1^{N_\epsilon^{(2)}})^{-1}X$ and $(\alpha^{N_{\widetilde{\Omega}}})^{-1}Z$, we can derive the same upper bound for the second moment of both variables, except that the effective N_δ in X should be $N_\delta+N_\epsilon^{(3)}$ and the effective $N_{\widetilde{\Omega}}$ in X should be $N_{\widetilde{\Omega}}+N_\epsilon^{(1)}+N_\epsilon^{(2)}$. It follows that

$$\begin{aligned}\mathbb{E}[X^2] \le C(\alpha^2)^{N_{\widetilde{\Omega}}} &\times \Big(\frac{\theta_{\max}^2}{\|\theta\|_1^2}\Big)^{N_\epsilon^{(1)}} \\ &\times f(\theta;m_1,m_2,m_3,m_4)\Big|_{\substack{m_1=N_{\widetilde{\Omega}}+N_\epsilon^{(1)}+N_\epsilon^{(2)},\ m_2=N_W,\\ m_3=N_\delta+N_\epsilon^{(3)},\ m_4=N_{\widetilde{r}}.}}\end{aligned} \tag{2.E.319}$$

We plug (2.E.319) into (2.E.317), and then plug it into (2.E.315). It gives (2.E.314).

Next, we use (2.E.314) to prove the claims of this lemma. Under our assumption, we can choose a sequence x_n such that $\sqrt{\log(\|\theta\|_1)} \ll x_n \ll \|\theta\|_1/\|\theta\|^2$. Also, note that $\|\theta\|_1 \ge \theta_{\max}^{-1}\|\theta\|^2 \gg \|\theta\|^2$. Then,

$$\frac{\theta_{\max}}{\|\theta\|_1} = o(\|\theta\|^{-2}), \qquad \frac{x_n}{\|\theta\|_1} = o(\|\theta\|^{-2}). \tag{2.E.320}$$

As a result, since $N_\epsilon^{(1)}+N_\epsilon^{(2)}+N_\epsilon^{(3)} \ge 1$, (2.E.314) implies

$$\mathbb{E}[Y^2] = o(\|\theta\|^{-4})\cdot f(\theta;m_1,m_2,m_3,m_4), \tag{2.E.321}$$

for $m_1 = N_{\widetilde{\Omega}}+N_\epsilon^{(1)}+N_\epsilon^{(2)}$, $m_2 = N_W$, $m_3 = N_\delta+N_\epsilon^{(3)}$ and $m_4 = N_{\widetilde{r}}$. We then extract $f(\theta;m_1,m_2,m_3,m_4)$ from previous lemmas. Recall the following facts:

- Under the null hypothesis, for any previously analyzed post-expansion sum Z, $|\mathbb{E}[Z]| \le C\|\theta\|^4$ and $\mathrm{Var}(Z) \le C\|\theta\|^8$.

- Under the alternative hypothesis, except $\sum_{i,j,k,\ell(dist)}\widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}$, for all previously analyzed post-expansion sum Z , $|\mathbb{E}[Z]| \le C\alpha^2\|\theta\|^6$ and $\mathrm{Var}(Z) \le C\|\theta\|^8 + C\alpha^6\|\theta\|^8\|\theta\|_3^6$.

Therefore, under both hypotheses, except for $(m_1,m_2,m_3,m_4) = (4,0,0,0)$,

$$f(\theta;m_1,m_2,m_3,m_4) \le C(\|\theta\|^8 + \|\theta\|^{12} + \|\theta\|^8\|\theta\|_3^6) \le C\|\theta\|^{12}. \tag{2.E.322}$$

Consider two cases for Y. The first case is $N_{\widetilde{\Omega}} + N_\epsilon^{(1)} + N_\epsilon^{(2)} \neq 4$. Combining (2.E.321)-(2.E.322) gives

$$\mathbb{E}[Y^2] = o(\|\theta\|^{-4}) \cdot C\|\theta\|^{12} = o(\|\theta\|^8).$$

The claims follow immediately. The second case is $N_{\widetilde{\Omega}} + N_\epsilon^{(1)} + N_\epsilon^{(2)} = 4$. In this case,

$$f(\theta; m_1, m_2, m_3, m_4) = f(\theta; 4, 0, 0, 0) = \|\theta\|^{16}.$$

If $N_\epsilon^{(1)} + N_\epsilon^{(2)} \geq 2$, then by (2.E.314) and (2.E.320),

$$\mathbb{E}[Y^2] = o(\|\theta\|^{-8}) \cdot C\|\theta\|^{16} = o(\|\theta\|^8).$$

The claims follow. It remains to consider $N_\epsilon^{(1)} + N_\epsilon^{(2)} = 1$ (and so $N_{\widetilde{\Omega}} = 3$). Write for short $S = 1 - \frac{v}{V}$. By (2.E.316),

$$Y = S^{N_\epsilon^{(2)}} \cdot X, \qquad \text{where} \quad X = \sum_{i,j,k,\ell (dist)} a_{ij} b_{jk} c_{k\ell} d_{\ell i},$$

and $a_{ij}, b_{ij}, c_{ij}, d_{ij}$ can only take values from $\{\widetilde{\Omega}_{ij}, \epsilon_{ij}^{(1)}, \eta_i \eta_j\}$. So, X is a non-stochastic number. Using (2.E.318), we can easily show

$$|X| \leq C\alpha^{N_{\widetilde{\Omega}}} \Big(\frac{\theta_{\max}}{\|\theta\|_1}\Big)^{N_\epsilon^{(1)}} \|\theta\|^8.$$

When $(N_\epsilon^{(1)}, N_\epsilon^{(2)}) = (1, 0)$, we have $Y = X$. By (2.E.320), $\frac{\theta_{\max}}{\|\theta\|_1} = o(\|\theta\|^{-2})$. It follows that

$$\mathrm{Var}(Y) = 0, \qquad |\mathbb{E}[Y]| = |X| \leq C\alpha^3 \cdot o(\|\theta\|^{-2}) \cdot \|\theta\|^8 = o(\alpha^4 \|\theta\|^8).$$

This gives the desired claims. When $(N_\epsilon^{(1)}, N_\epsilon^{(2)}) = (0, 1)$, we have $Y = S \cdot X$. So,

$$|Y| = |X| \cdot |S| \leq C\alpha^3 \|\theta\|^8 \cdot |S|.$$

Note that $S = 1 - \frac{v}{V}$, where $v = \mathbb{E}[V]$. Using the tail bound (2.E.290), we can prove $\mathbb{E}[S^2] \leq C\|\theta\|_1^{-2}$. Therefore,

$$\mathbb{E}[Y^2] \leq \frac{C\alpha^6 \|\theta\|^{16}}{\|\theta\|_1^2} \leq C\alpha^6 \|\theta\|^8 \|\theta\|_3^6,$$

where the last inequality is due to $\|\theta\|^4 \leq \|\theta\|_1 \|\theta\|_3^3$ (Cauchy-Schwarz). The claims follow immediately. $\square$

Three

Estimating the number of communities by Stepwise Goodness-of-fit

3.1 Introduction

In network analysis, how to estimate the number of communities K is a fundamental problem. In many recent approaches, K is assumed as known a priori. See for example Chen et al. (2018); Gao et al. (2018); Karrer and Newman (2011); Ma et al. (2020); Zhao et al. (2011); Xu et al. (2020) on community detection, Jin et al. (2017); Zhang et al. (2014) on mixed-membership estimation, and Liu et al. (2017) on dynamic community detection. Unfortunately, K is rarely known in applications, so the performance of these approaches hinges on how well we can estimate K.

The primary interest of this chapter is how to estimate K. Given a symmetric and connected social network with n nodes and K communities, let A be the adjacency matrix:

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ and node } j \text{ have an edge,} \\ 0, & \text{otherwise,} \end{cases} \qquad 1 \leq i \neq j \leq n. \tag{3.1.1}$$

As a convention, self-edges are not allowed so all the diagonal entries of A are 0. Denote the K perceivable communities by $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_K$. We model the network by the widely-used degree-corrected block model (DCBM) Karrer and Newman (2011). For each $1 \leq i \leq n$, we encode the community label of node i by a vector $\pi_i \in \mathbb{R}^K$ where

$$i \in \mathcal{N}_k \qquad \Longleftrightarrow \qquad \pi_i(k) = 1 \text{ and } \pi_i(m) = 0 \text{ for } m \neq k. \tag{3.1.2}$$

Moreover, for a $K \times K$ symmetric nonnegative matrix P which models the community structure and positive parameters $\theta_1, \theta_2, \ldots, \theta_n$ which model the degree heterogeneity, we assume the upper triangular entries of A are independent Bernoulli variables satisfying

$$\mathbb{P}(A_{ij} = 1) = \theta_i \theta_j \cdot \pi_i' P \pi_j \equiv \Omega_{ij}, \qquad 1 \leq i < j \leq n, \tag{3.1.3}$$

where Ω denotes the matrix $\Theta \Pi P \Pi' \Theta$, with Θ being the $n \times n$ diagonal matrix $\mathrm{d}iag(\theta_1, \ldots, \theta_n)$ and Π being the $n \times K$ matrix $[\pi_1, \pi_2, \ldots, \pi_n]'$. For identifiability, we assume

$$\text{all diagonal entries of } P \text{ are } 1. \tag{3.1.4}$$

Write for short $\mathrm{d}iag(\Omega) = \mathrm{d}iag(\Omega_{11}, \Omega_{22}, \ldots, \Omega_{nn})$, and let W be the matrix where for $1 \leq i, j \leq n$, $W_{ij} = A_{ij} - \Omega_{ij}$ if $i \neq j$ and $W_{ij} = 0$ otherwise. In matrix form, we have

$$A = \Omega - \mathrm{d}iag(\Omega) + W, \qquad \text{where we recall } \ \Omega = \Theta \Pi P \Pi' \Theta. \tag{3.1.5}$$

In the special case of $\theta_1 = \theta_2 = \ldots = \theta_n$, DCBM reduces to the stochastic block model (SBM) Holland et al. (1983). In this paper, we focus on DCBM, but the idea is extendable to the degree-corrected mixed-membership (DCMM) model Zhang et al. (2014); Jin et al. (2017), where mixed membership is allowed; see Remark 3 below.

Real world networks have a few interesting features that we frequently observe.

- *Severe degree heterogeneity.* The distribution of the node degrees has a power-law tail, implying severe degree heterogeneity. Therefore, the sparsity level for individual nodes (measured by the number of edges) may vary significantly from one to another.
- *Network sparsity.* The overall network sparsity may range significantly from one network to another.
- *Weak signal.* The community structure is masked by strong noise, and the signal-to-noise ratio (SNR) is usually relatively small.

For analysis, we let n be the driving asymptotic parameter, and allow (Θ, Π, P) to depend on n, so that DCBM is broad enough to cover all interesting range of these metrics. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_n)'$, $\theta_{max} = \max\{\theta_1, \ldots, \theta_n\}$, and $\theta_{min} = \min\{\theta_1, \theta_2, \ldots, \theta_n\}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_K$ be the K nonzero eigenvalues of Ω, arranged in the descending order of magnitudes. The following were suggested by existing literature (e.g., Jin et al. (2019); Jin (2015)). First, a reasonable metic for network sparsity is $\|\theta\|$ and a reasonable metric for the degree heterogeneity is $\theta_{max}/\theta_{min}$. Second, the range of interest for $\|\theta\|$ is

$$C\sqrt{\log(n)} \leq \|\theta\| \leq C\sqrt{n}, \tag{3.1.6}$$

where $C > 0$ is a generic constant. Third, the signal strength and noise level are captured by $|\lambda_K|$ and $\|W\|$, respectively. When $\theta_{max} \leq C\theta_{min}$ and some mild conditions hold (e.g., $\|P\| \leq C$),

$$\lambda_1 \asymp \|\theta\|^2, \qquad \text{and} \qquad \|W\| = \text{a multi-}\log(n) \text{ term} \cdot \sqrt{\lambda_1} \text{ with high probability}, \tag{3.1.7}$$

(examples for multi-log(n)-terms are $\sqrt{\log(n)}, \log\log(n)$, etc.), so a reasonable metric for the signal to noise ratio (SNR) is $|\lambda_K|/\sqrt{\lambda_1}$. When $\theta_{max}/\theta_{min} \to \infty$, we need an adjusted SNR; see Section 3.2. We consider two extreme cases.

- *Strong signal case.* $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_K|$ are at the same magnitude, and so SNR $\asymp \sqrt{\lambda_1}$.
- *Weak signal case.* $|\lambda_K|/\sqrt{\lambda_1}$ is much smaller than $\sqrt{\lambda_1}$ and grows to ∞ slowly as $n \to \infty$ (in our range of interest, λ_1 may grow to ∞ rapidly as $n \to \infty$, so for example, we may have SNR $= \log\log(n)$ and $\lambda_1 = \sqrt{n}$).

Section 3.2.4 suggests that when SNR $= o(1)$, consistent estimate for K does not exist, so the weak signal case is a very challenging case. Motivated by the above observations, it is desirable to find a consistent estimate for K that satisfies the following requirements.

- (R1). Allow severe degree heterogeneity (i.e., $\theta_{max}/\theta_{min}$ may tend to ∞).

- (R2). Optimally adaptive to network sparsity, where $\|\theta\|$ may be as small as $O(\sqrt{\log(n)})$ or be as large as $O(\sqrt{n})$.

- (R3). *Attain the information lower bound.* Consistent for both the strong signal case where SNR is large and the weak signal case where SNR may be as small as $\log\log(n)$ (say).

Example 1. Recently, a frequently considered DCBM is to assume $P = P_0$ and $\theta_i \asymp \sqrt{\alpha_n}$ for all $1 \leq i \leq n$, where $\alpha_n > 0$ is a scaling parameter and P_0 is a fixed matrix. It is seen that $\lambda_1, \ldots, \lambda_K$ are at the same order, so the model only considers the strong signal case.

Example 2 *(A special DCBM).* Let $e_1, \ldots, e_K$ be the standard basis vectors of $\mathbb{R}^K$. Fixing a positive vector $\theta \in \mathbb{R}^n$ and a scalar $b_n \in (0, 1)$, we consider a DCBM where K is fixed, each community has n/K nodes, and $P = (1-b_n)I_K + b_n 1_K 1_K'$. In this model, $(1-b_n)$ measures the "dis-similarity" of different communities and is small in the more challenging case when different communities are similar. By basic algebra, $\lambda_1 \asymp \|\theta\|^2$, $\lambda_2 = \ldots = \lambda_K \asymp \|\theta\|^2(1-b_n)$, and SNR $\asymp \|\theta\|(1-b_n)$. In the very sparse case, $\|\theta\| = O(\sqrt{\log(n)})$. In the dense case, $\|\theta\| = O(\sqrt{n})$. When $b_n \leq c_0$ for a constant $c_0 < 1$, $|\lambda_K| \geq C|\lambda_1|$ and SNR $\asymp \|\theta\|$; we are in the strong signal case if $\|\theta\| \geq n^a$ for a constant $a > 0$. When $b_n = 1 + o(1)$ and $\|\theta\|(1-b_n) = \log\log(n)$ (say), SNR $\asymp \log\log(n)$ and we are in the weak signal case.

3.1.1 Literature review and our contributions

In recent years, many interesting approaches for estimating K have been proposed, which can be roughly divided into the spectral approaches, the cross validation approaches, the penalization approaches, and the likelihood ratio approaches.

Among the spectral approaches, Le and Levina (2015) proposed to estimate K using the eigenvalues of the non-backtracking matrix or Bethe Hessian matrix. The approach uses ideas from mathematical graph theory, and is quite interesting for it is different from most statistical approaches. Unfortunately, the approach requires relatively strong conditions for consistency. For example, their Theorem 4.1 only considers the idealized SBM model in the very sparse case, where $\theta_1 = \theta_2 = \ldots = \theta_n = 1/\sqrt{n}$ and $P = P_0$ for a fixed matrix P_0. Liu et al. (2019) proposed to estimate K by using the classical scree plot approach with careful theoretical justification, but the approach is known to be unsatisfactory in the presence of severe degree heterogeneity, for it is hard to derive a sharp bound for the spectral norm of the noise matrix W (e.g., Jin (2015)). Therefore, their approach requires the condition of $\theta_{max} \leq C\theta_{min}$. The paper also imposed the condition of $\|\theta\| = O(\sqrt{n})$ so it did not address the settings of sparse networks (see (3.1.6) for the interesting range of $\|\theta\|$). Among the cross-validation approaches, we have Chen and Lei (2018); Li et al. (2020), and among the penalization approaches, we have Saldaña et al. (2017); Daudin et al. (2008); Latouche et al. (2012), where K is estimated by the integer that optimizes some objective functions. For example, Saldaña et al. (2017) used a BIC-type objective function and Daudin et al. (2008); Latouche et al. (2012) used an objective function of the Bayesian model selection flavor. However, these methods did not provide explicit theoretical guarantee on consistency (though a partial result was established in (Li et al., 2020), which stated that under SBM, the proposed estimator $\widehat{K}$ is no greater than K with high probability).

For likelihood ratio approaches, Wang et al. (2017) proposed to estimate K by solving a BIC type optimization problem, where the objective function is the sum of the log-likelihood and the model complexity. The major challenge here is that the likelihood is the sum of exponentially many terms and is hard to compute. In a remarkable paper, Ma et al. (2018) extended the idea of Wang et al. (2017) by proposing a new approach that is computationally more feasible.

On a high level, we can recast their methods as a *stepwise testing or sequential testing* algorithm. Consider a stepwise testing scheme where for $m = 1, 2, \ldots$, we construct a test statistic $\ell_n^{(m)}$ (e.g. log-likelihood) assuming m is the correct number of communities. We estimate K as the smallest m such that the pairwise log-likelihood ratio $(\ell_n^{(m+1)} - \ell_n^{(m)})$ falls below a *threshold*. As mentioned in Wang et al. (2017); Ma et al. (2018), such an approach faces challenges. Call the cases $m < K$, $m = K$, and $m > K$ the *under-fitting*, *null*, and *over-fitting* cases, respectively.

- We have to analyze $\ell_n^{(m)}$ for both the under-fitting case and the over-fitting case, but we do not have efficient technical tools to address either case.
- It is hard to derive sharp results on the limiting distribution of $\ell_n^{(m+1)} - \ell_n^{(m)}$ in the null case, and so it is unclear how to pin down the threshold.

Ma et al. (2018) (see also Wang et al. (2017)) made interesting progress but unfortunately the problems are not resolved satisfactorily. For example, they require hard-to-check strong conditions on both the under-fitting and over-fitting cases. Also, in the over-fitting case, it is unclear whether their results are sharp, and in the under-fitting case, it is unclear how to standardize $\ell_n^{(m+1)} - \ell_n^{(m)}$ as the variance term is unknown; as a result, how to pin down the threshold remains unclear. Most importantly, both papers focus on the setting in Example 1 (see above), where severe degree heterogeneity is not allowed and they only consider the strong signal case.

In this chapter, we propose *Stepwise Goodness-of-Fit (StGoF)* as a new approach to estimating K. Our approach follows a different vein, so it is different not only by the particular procedures we use, but also in the design of the stepwise testing. In detail, for $m = 1, 2, \ldots$, StGoF alternately uses two sub-steps, a community detection sub-step where we apply SCORE Jin (2015) assuming m is the correct number of communities, and a Goodness-of-Fit (GoF) sub-step. We propose a new GoF approach and let $\psi_n^{(m)}$ be the GoF test statistic in step m. Assuming SNR $\to \infty$, we show that

$$\psi_n^{(m)} \begin{cases} \to N(0,1), & \text{when } m = K \text{ (null case)}, \\ \to \infty \text{ in probability}, & \text{when } 1 \leq m < K \text{ (under-fitting case)}. \end{cases} \tag{3.1.8}$$

This gives rise to a consistent estimate for K. Note that we have derived $N(0,1)$ as the explicit limiting null distribution which is crucial in our study. To prove (3.1.8), the key is to show that in the under-fitting case, SCORE has the so-called *Non-Splitting Property (NSP)*, meaning that all nodes in each (true) community are always clustered together. In the over-fitting case, $m > K$. The NSP does not hold and so the analytical challenge remains, but the design of StGoF and the sharp results in (3.1.8) help avoid the analysis in this case.

For the stepwise testing algorithms in Wang et al. (2017); Ma et al. (2018), analysis in the over-fitting case can not be avoided, as we need to analyze $\ell_n^{(m+1)} - \ell_n^{(m)}$ for $m = 1, 2, \ldots, K$; see details therein.

To assess the optimality, we use the phase transition, a well-known optimality framework. It is related to the minimax framework but can be frequently more informative Donoho and Jin (2004); Ingster et al. (2010); Ma and Wu (2015); Paul (2007). We show that when SNR $\to \infty$, (3.1.8) gives rise to an estimator that is consistent in a broad setting. We also obtain an information lower bound by showing that when SNR $\to 0$, consistent estimates for K do not exist. This suggests that our consistency result is sharp in terms of the rate of SNR, so we say that StGoF achieves the optimal phase transition; see Section 3.2.4. As far as we know, such a phase transition result on estimating K is new.

In order to achieve the optimal phase transition, a procedure needs to work well in the weak signal case. Since most existing methods have been focused on the strong signal case, it is unclear whether they achieve the optimal phase transition. Our contributions are as follows.

- We propose StGoF as a new approach to estimating K, where we use both a different design for stepwise testing and a new GoF test.
- We derive $N(0, 1)$ as the explicit limiting null distribution, and use the NSP of SCORE to derive tight bounds in the under-fitting case. These sharp results and the design of StGoF allow us to avoid the analysis in the over-fitting case and so to overcome the technical challenges faced by stepwise testing of this kind. Such an analytical strategy is extendable to other settings (e.g., the study of directed or bipartite graphs).
- We show that StGoF achieves the optimal phase transition when $\theta_{\max} \leq C\theta_{min}$ and consistent in broad settings (e.g., weak signals, severe degree heterogeneity, and a wide range of sparsity). In particular, StGoF satisfies all requirements (R1)-(R3) as desired.

Compared to Jin (2015), both papers study SCORE, but the goal of Jin (2015) is community detection where K is assumed as known, and the analysis were focused on the null case ($m = K$). Here, the goal is to estimate K: SCORE is only used as part of our stepwise algorithm, and the analysis of SCORE is focused on the under-fitting case ($m < K$), where the property of SCORE is largely unknown, and our results on the NSP of SCORE are new.

The proof of NSP is non-trivial when $m < K$. It depends on the row-wise distances of the matrix Ξ consisting of the first m columns of $[\xi_1, \ldots, \xi_K]\Gamma$, where ξ_k is the k-th eigenvector of Ω and Γ is an orthogonal matrix dictated by the Davis Kahan $\sin(\theta)$ theorem Davis and Kahan (1970). Γ is hard to track without a strong eigen-gap assumption, and when it ranges, the row-wise distances of Ξ are the same when $m = K$ but may vary significantly when

$m < K$. This is why the study on SCORE is much harder in the under-fitting case than in the null case. See Section 3.3.

3.1.2 Content

Sections 3.2-3.3 contain main theoretical results. In Section 3.2, we first propose a new GoF test for DCBM. we then show that StGoF is consistent for K uniformly in a broad class of settings. We also present the information lower bound and show that StGoF achieves the optimal phase transition. In Section 3.3, we show that SCORE has the Non-Splitting Property (NSP) for $1 \leq m \leq K$. We also shed light on why SCORE has the NSP and what the technical challenges are. In Section 3.4, we prove the main results. Section 3.5 presents numerical results with real and simulated data. The supplementary material contains the proofs for secondary theorems and lemmas.

In this chapter, $C > 0$ denotes a generic constant which may vary from case to case. For any numbers $\theta_1, \ldots, \theta_n$, $\theta_{max} = \max\{\theta_1, \ldots, \theta_n\}$, and $\theta_{min} = \min\{\theta_1, \ldots, \theta_n\}$. For any vectors $\theta = (\theta_1, \ldots, \theta_n)'$, both $diag(\theta)$ and $diag(\theta_1, \ldots, \theta_n)$ denote the $n \times n$ diagonal matrix with θ_i being the i-th diagonal entry, $1 \leq i \leq n$. For any vector $a \in \mathbb{R}^n$, $\|a\|_q$ denotes the Euclidean ℓ^q-norm (we write $\|a\|$ for short when $q = 2$). For any matrix $P \in \mathbb{R}^{n,n}$, $\|P\|$ denotes the matrix spectral norm, and $\|P\|_{\max}$ denotes the entry-wise maximum norm. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we say $a_n \sim b_n$ if $\lim_{n\to\infty}\{a_n/b_n\} = 1$ and $a_n \asymp b_n$ if there are constants $c_2 > c_1 > 0$ such that $c_1 a_n \leq b_n \leq c_2 a_n$ for sufficiently large n.

3.2 Optimal phase transition

This section contains the first part of our main results, where we discuss the consistency and optimality of StGoF. Section 3.3 contains the second part of our main results, where we discuss the NSP of SCORE Jin (2015).

Consider a DCBM with K communities as in (3.1.5). We assume

$$\|P\| \leq C, \qquad \|\theta\| \to \infty, \qquad \text{and} \qquad \theta_{\max}\sqrt{\log(n)} \to 0. \tag{3.2.1}$$

The first one is a mild regularity condition on the $K \times K$ community structure matrix P. The other two are mild conditions on sparsity. See (3.1.6) for the interesting range of $\|\theta\|$. We exclude the case where $\theta_i = O(1)$ for all $1 \leq i \leq n$ for convenience, but our results continue to hold in this case provided that we make some small changes in our proofs. Moreover, for $1 \leq k \leq K$, let $\mathcal{N}_k$ be the set of nodes belonging to community k, let n_k be the cardinality of $\mathcal{N}_k$, and let $\theta^{(k)}$ be the n-dimensional vector where $\theta_i^{(k)} = \theta_i$ if $i \in \mathcal{N}_k$ and $\theta_i^{(k)} = 0$ otherwise. We assume the K communities are balanced in the sense that

$$\min_{\{1\leq k\leq K\}}\{n_k/n,\ \|\theta^{(k)}\|_1/\|\theta\|_1,\ \|\theta^{(k)}\|/\|\theta\|\} \geq C. \tag{3.2.2}$$

In the presence of severe degree heterogeneity, the valid SNR for SCORE is

$$s_n = a_0(\theta)(|\lambda_K|/\sqrt{\lambda_1}), \qquad \text{where } a_0(\theta) = (\theta_{min}/\theta_{max}) \cdot (\|\theta\|/\sqrt{\theta_{max}\|\theta\|_1}) \leq 1.$$

In the special case of $\theta_{max} \leq C\theta_{min}$, it is true that $a_0(\theta) \asymp 1$ and $s_n \asymp |\lambda_K|/\sqrt{\lambda_1}$. In this case, s_n is the SNR introduced (3.1.7). We assume

$$s_n \geq C_0\sqrt{\log(n)}, \qquad \text{for a sufficiently large constant } C_0 > 0. \tag{3.2.3}$$

In the special case of $\theta_{max} \leq C\theta_{min}$, (3.2.3) is equivalent to $|\lambda_K|/\sqrt{\lambda_1} \geq C\sqrt{\log(n)}$, which is mild. See Remark 6 for more discussion. Define a $K \times K$ diagonal matrix H by $H_{kk} = \|\theta^{(k)}\|/\|\theta\|$, $1 \leq k \leq K$. For the matrix HPH and $1 \leq k \leq K$, let (largest means largest in magnitude)

μ_k be the k-th largest eigenvalue and η_k be the corresponding eigenvector.

By Perron's theorem Horn and Johnson (1985), if P is irreducible, then the multiplicity of μ_1 is 1, and all entries of η_1 are all strictly positive. Note also the size of the matrix P is small. It is therefore only a mild condition to assume that for a constant $0 < c_0 < 1$,

$$\min_{2 \leq k \leq K} |\mu_1 - \mu_k| \geq c_0|\mu_1|, \qquad \text{and} \qquad \frac{\max_{1 \leq k \leq K}\{\eta_1(k)\}}{\min_{1 \leq k \leq K}\{\eta_1(k)\}} \leq C. \tag{3.2.4}$$

In fact, (3.2.4) holds if all entries of P are lower bounded by a positive constant or $P \to P_0$ for a fixed irreducible matrix P_0. We also note that the most challenging case for network analysis is when the matrix P is close to the matrix of 1's (where it is hard to distinguish one community from another), and (3.2.4) always holds in such a case. In this paper, we implicitly assume K is fixed. This is mostly for simplicity, as there is really no technical hurdle for the case of diverging K. See Remark 5 for more discussion.

3.2.1 The StGoF algorithm and a DCBM Goodness-of-Fit test

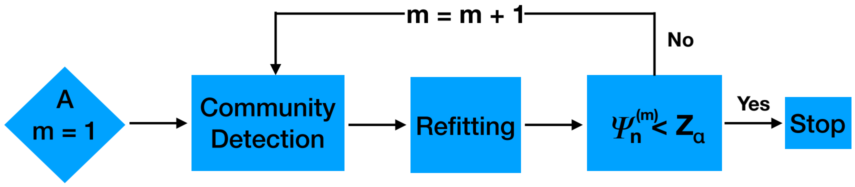

Figure 3.1: The flow chart of StGoF.

The Stepwise Goodness-Of-Fit algorithm (StGoF) is a stepwise algorithm where for $m = 1, 2, \ldots$, we alternately use a community detection step and a Goodness-of-Fit (GoF) step. In principle, we can view StGoF as a general framework, and for both steps, we may use different algorithms. However, for most existing community detection algorithms (e.g., Chen et al. (2018); Gao et al. (2018); Zhang et al. (2014)), it is unclear whether they have the desired theoretical properties (especially the NSP), so we may face analytical challenges. For this reason, we choose to use SCORE Jin (2015), which we prove to have the NSP. For GoF, existing algorithms (e.g., Hu et al. (2020); Lei (2016); see Remark 2 for more discussion) do not apply to the current setting, so we propose a new GoF measure called the Refitted Quadrilateral (RQ).

In detail, fixing a tolerance parameter $0 < \alpha < 1$ and letting z_α be the α upper-quantile of $N(0,1)$, StGoF runs as follows. Input the adjacency matrix A and initialize $m = 1$.

- *(a). Community detection.* If $m = 1$, let $\widehat{\Pi}^{(m)}$ be the n-dimensional vector of 1's. If $m > 1$, apply SCORE to A assuming m is the correct number of communities and obtain an $n \times m$ matrix $\widehat{\Pi}^{(m)}$ for the estimated community labels.
- *(b). Goodness-of-Fit.* Assuming $\widehat{\Pi}^{(m)}$ is the matrix of true community labels, we obtain an estimate $\widehat{\Omega}^{(m)}$ for Ω by refitting the DCBM, following (3.2.6)-(3.2.7) below.

Obtain the Refitted Quadrilateral test score $\psi_n^{(m)}$ as in (3.2.9)-(3.2.12).

- *(c). Termination.* If $\psi_n^{(m)} \geq z_\alpha$, repeat (a)-(b) with $m = m + 1$. Otherwise, output m as the estimate for K. Denote the final estimate by $\widehat{K}_\alpha^*$.

We recommend $\alpha = 1\%$ or 5%. See Figure 3.1 for the flow chart of the algorithm.

We now fill in the details for steps (a)-(b). Consider (a) first. The case of $m = 1$ is trivial so we only consider the case of $m > 1$. Let $\widehat{\lambda}_k$ be the k-th largest (in magnitude) eigenvalue of A, and let $\widehat{\xi}_k$ be the corresponding eigenvector. For each $m > 1$, we apply SCORE as follows.

Input: A and m. Output: the estimated $n \times m$ matrix of community labels $\widehat{\Pi}^{(m)}$.

- Obtain the first m eigenvectors $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_m$ of A. Define the $n \times (m-1)$ matrix of entry-wise ratios $\widehat{R}^{(m)}$ by $\widehat{R}^{(m)}(i,k) = \widehat{\xi}_{k+1}(i)/\widehat{\xi}_1(i)$, $1 \leq i \leq n, 1 \leq k \leq m-1$. [1]
- Cluster the rows of $\widehat{R}^{(m)}$ by the classical k-means assuming we have m clusters. Output $\widehat{\Pi}^{(m)} = [\widehat{\pi}_1^{(m)}, \ldots, \widehat{\pi}_n^{(m)}]'$ ($\widehat{\pi}_i^{(m)}(k) = 1$ if node i is clustered to cluster k and 0 otherwise).

Existing study of SCORE has been focused on the null case of $m = K$. Our interest here is on the under-fitting case ($1 < m < K$), where the property of SCORE is largely unknown.

Consider (b). The idea is to *pretend* that the SCORE estimate $\widehat{\Pi}^{(m)}$ is accurate. We then use it to estimate Ω by re-fitting, and check how well the estimated Ω fits with the adjacency matrix A. In detail, let d_i be the degree of node i, $1 \leq i \leq n$, and let $\widehat{\mathcal{N}}_k^{(m)}$ be the set of nodes that SCORE assigns to group k, $1 \leq k \leq m$. We decompose $\mathbf{1}_n$ as follows

$$\mathbf{1}_n = \sum_{k=1}^{m} \widehat{\mathbf{1}}_k^{(m)}, \qquad \text{where } \widehat{\mathbf{1}}_k^{(m)}(j) = 1 \text{ if } j \in \widehat{\mathcal{N}}_k^{(m)} \text{ and 0 otherwise.} \tag{3.2.5}$$

For most quantities that have superscript (m), we may only include the superscript when introducing these quantities for the first time, and omit it later for notational simplicity when there is no confusion. Introduce a vector $\widehat{\theta}^{(m)} = (\widehat{\theta}_1^{(m)}, \widehat{\theta}_2^{(m)}, \ldots, \widehat{\theta}_n^{(m)})' \in \mathbb{R}^n$ and a matrix $\widehat{P}^{(m)} \in \mathbb{R}^{m,m}$ where for all $1 \leq i \leq n$ and $1 \leq k, \ell \leq m$,

$$\widehat{\theta}_i^{(m)} = [d_i/(\widehat{\mathbf{1}}_k' A \mathbf{1}_n)] \cdot \sqrt{\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_k}, \qquad \widehat{P}_{k\ell}^{(m)} = (\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_\ell)/\sqrt{(\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_k)(\widehat{\mathbf{1}}_\ell' A \widehat{\mathbf{1}}_\ell)}. \tag{3.2.6}$$

Let $\widehat{\Theta}^{(m)} = \mathrm{d}iag(\widehat{\theta})$. We refit Ω by

$$\widehat{\Omega}^{(m)} = \widehat{\Theta}^{(m)} \widehat{\Pi}^{(m)} \widehat{P}^{(m)} (\widehat{\Pi}^{(m)})' \widehat{\Theta}^{(m)}. \tag{3.2.7}$$

Recall that $\Omega = \Theta \Pi P \Pi' \Theta$ and P has unit diagonal entries. In the ideal case where $m = K$, $\widehat{\Pi}^{(m)} = \Pi$, and $A = \Omega$, we can verify that $(\widehat{\Theta}^{(m)}, \widehat{P}^{(m)}, \widehat{\Omega}^{(m)}) = (\Theta, P, \Omega)$. This suggests that the refitting in (3.2.7) is reasonable. The Refitted Quadrilateral (RQ) test statistic is then

$$Q_n^{(m)} = \sum_{i_1,i_2,i_3,i_4 (dist)} (A_{i_1 i_2} - \widehat{\Omega}_{i_1 i_2}^{(m)})(A_{i_2 i_3} - \widehat{\Omega}_{i_2 i_3}^{(m)})(A_{i_3 i_4} - \widehat{\Omega}_{i_3 i_4}^{(m)})(A_{i_4 i_1} - \widehat{\Omega}_{i_4 i_1}^{(m)}), \tag{3.2.8}$$

[1] As the network is connected, $\widehat{\xi}_1$ is uniquely defined with all positive entries, by Perron's theorem Jin (2015).

("dist" means the indices are distinct). Without the refitted matrix $\widehat{\Omega}^{(m)}$, $Q_n^{(m)}$ reduces to

$$C_n = \sum_{i_1,i_2,i_3,i_4(dist)} A_{i_1i_2}A_{i_2i_3}A_{i_3i_4}A_{i_4i_1}, \tag{3.2.9}$$

which is the total number of quadrilaterals in the networks Jin et al. (2018). This is why we call $Q_n^{(m)}$ the refitted quadrilaterals.

We now discuss the mean and variance of $Q_n^{(m)}$ in the null case of $m = K$. In this case, first, it turns out that the variance can be well-approximated by $8C_n$. Second, while that $\mathbb{E}[Q_n^{(K)}] = 0$ in the ideal case of $\widehat{\Omega}^{(K)} = \Omega$, in the real case, $\widehat{\Omega}^{(K)} \neq \Omega$ and $\mathbb{E}[Q_n^{(K)}]$ is comparable to the standard deviation of $Q_n^{(K)}$. Therefore, the mean is not negligible in the null case, and we need bias correction.

Motivated by these, for any $m \geq 1$, we introduce two vectors $\widehat{g}^{(m)}, \widehat{h}^{(m)} \in \mathbb{R}^m$ where

$$\widehat{g}_k^{(m)} = (\widehat{\mathbf{1}}_k'\widehat{\theta})/\|\widehat{\theta}\|_1, \qquad \widehat{h}_k^{(m)} = (\widehat{\mathbf{1}}_k'\widehat{\Theta}^2\widehat{\mathbf{1}}_k)^{1/2}/\|\widehat{\theta}\|, \qquad 1 \leq k \leq m. \tag{3.2.10}$$

Write for short $\widehat{V}^{(m)} = \mathrm{d}iag(\widehat{P}\widehat{g})$ and $\widehat{H}^{(m)} = \mathrm{d}iag(\widehat{h})$. We estimate the mean of $Q_n^{(m)}$ by

$$B_n^{(m)} = 2\|\widehat{\theta}\|^4 \cdot [\widehat{g}'\widehat{V}^{-1}(\widehat{P}\widehat{H}^2\widehat{P} \circ \widehat{P}\widehat{H}^2\widehat{P})\widehat{V}^{-1}\widehat{g}], \tag{3.2.11}$$

where for matrixes A and B, $A \circ B$ is their Hadamard product Horn and Johnson (1985). We show that in the null case, $B_n^{(m)}$ is a good estimate for $\mathbb{E}[Q_n^{(m)}]$, and in the under-fitting case, it is much smaller than the leading term of $Q_n^{(m)}$ and so is negligible. Finally, the StGoF statistic is defined by

$$\psi_n^{(m)} = [Q_n^{(m)} - B_n^{(m)}]/\sqrt{8C_n}. \tag{3.2.12}$$

The computational cost of the StGoF algorithm is determined by (i) the number of iterations, (ii) the cost of SCORE, and (iii) the cost of computing $\psi_n^{(m)}$ in (3.2.12). For (i), we show in Section 3.2 that, under mild conditions, StGoF terminates in exactly K steps with high probability. For (ii), the costs are from implementing PCA and k-means (Jin, 2015). PCA is manageable even for very large networks, and the complexity is $O(n^2\bar{d})$ for each m if we use the power method, where $\bar{d}$ is the average degree. In practice, the k-means is usually implemented with the Lloyd's algorithm which is fast (e.g., only a few seconds when n is a few thousands). In theory, the computational cost of k-means for our setting is polynomial-time, since the dimension of each row of $\widehat{R}^{(m)}$ is $(m-1)$. For (iii), the following lemma shows the complexity is polynomial time. Lemma 27 is proved in the supplementary material.

Lemma 27. *For each $m = 1, 2, \ldots, K$, the complexity for computing $\psi_n^{(m)}$ is $O(n^2\bar{d})$, where $\bar{d}$ is average degree of the network.*

Remark 1. The RQ test has some connections to the SgnQ test in Jin et al. (2019), but is for different problem and is more sophisticated. The RQ test is for goodness-of-fit. It depends on the matrix $\widehat{\Omega}^{(m)}$, refitted for each m using the community detection results by SCORE. The SgnQ test is for global testing, where the goal is to test $K = 1$ vs. $K > 1$. The SgnQ test is not stepwise, and does not depend on any results of community detection. In particular, to analyze RQ, we need new technical tools, where the NSP of SCORE plays a key role.

Remark 2. Existing GoF algorithms include Hu et al. (2020); Lei (2016), but they only address the much narrower settings (e.g., dense networks with stochastic block model and strong signals). As mentioned in Hu et al. (2020), it remains unclear how to generalize these approaches to the DCBM setting here. In principle, a GoF approach only focuses on the null case, and can not be used for estimating K without sharp results in the under-fitting case, or the over-fitting case, or both.

Remark 3. We are primarily interested in DCBM, but the idea can be extended to the broader DCMM Zhang et al. (2014); Jin et al. (2017), where mixed-memberships are allowed. To this end, we need to replace SCORE by Mixed-SCORE Jin et al. (2017) (an adapted version of SCORE for networks with mixed memberships), and modify the refitting step accordingly. The analysis of the resultant procedure is much more challenging so we leave it to the future.

3.2.2 The null case and a confidence lower bound for K

In the null case, $m = K$. In this case, if we apply SCORE to the rows of $\widehat{R}^{(m)}$ assuming m clusters, then we have perfect community recovery. As a result, StGoF provides a confidence lower bound for K.

Theorem 3.2.1. *Fix $0 < \alpha < 1$. Suppose we apply StGoF to a DCBM model where (3.2.1)-(3.2.4) hold. As $n \to \infty$, up to a permutation of the columns of $\widehat{\Pi}^{(K)}$, $\mathbb{P}(\widehat{\Pi}^{(K)} \neq \Pi) \leq Cn^{-3}$, $\psi_n^{(K)} \to N(0, 1)$ in law, and $\mathbb{P}(\widehat{K}_\alpha^* \leq K) \geq (1 - \alpha) + o(1)$.*

Theorem 3.2.1 is proved in Section 3.4. Theorem 3.2.1 allows for severe degree heterogeneity. If the degree heterogeneity is moderate, $s_n \asymp |\lambda_K|/\sqrt{\lambda_1}$, and we have the following corollary.

Corollary 3.2.1. *Fix $0 < \alpha < 1$. Suppose we apply StGoF to a DCBM model where (3.2.1)-(3.2.2) and (3.2.4) hold. Suppose $\theta_{max} \leq C\theta_{min}$ and $|\lambda_K|/\sqrt{\lambda_1} \geq C_0\sqrt{\log(n)}$ for a sufficiently large constant $C_0 > 0$. As $n \to \infty$, up to a permutation of the columns of $\widehat{\Pi}^{(K)}$, $\mathbb{P}(\widehat{\Pi}^{(K)} \neq \Pi) \leq Cn^{-3}$, $\psi_n^{(K)} \to N(0, 1)$ in law, and $\mathbb{P}(\widehat{K}_\alpha^* \leq K) \geq (1 - \alpha) + o(1)$.*

Theorem 3.2.1 and Corollary 3.2.1 show that $\widehat{K}_\alpha^*$ provides a level-$(1 - \alpha)$ confidence lower bound for K. If α depends on n and tends to 0 slowly enough, these results continue to hold. In this case, $\mathbb{P}(\widehat{K}_\alpha^* \leq K) \geq 1 + o(1)$. In cases (e.g., when the SNR is slightly smaller than those above) where perfect community recovery is impossible but the fraction of of misclassified nodes is small, the asymptotic normality continues to hold. Same comments apply to Theorem 3.2.3 and Corollary 3.2.2.

3.2.3 The under-fitting case and consistency of StGoF

In the under-fitting case, $m < K$. We focus on the case of $1 < m < K$ as the case of $m = 1$ is trivial. Suppose we apply SCORE to the rows of $\widehat{R}^{(m)}$ assuming m is the correct number of communities and let $\widehat{\Pi}^{(m)}$ be the matrix of estimated community labels as before. When $1 < m < K$, we underestimate the number of clusters, so perfect community recovery is impossible. However, SCORE satisfies the *Non-Splitting Property (NSP)*. Recall that Π is the matrix of true community labels.

Definition 28. *Fix $K > 1$ and $m \leq K$. We say that a realization of the $n \times m$ matrix of estimated labels $\widehat{\Pi}^{(m)}$ satisfies the NSP if for any pair of nodes in the same (true) community, the estimated community labels are the same. When this happens, we write $\Pi \preceq \widehat{\Pi}^{(m)}$, meaning the partition (into clusters) on the left is finer than that on the right.*

Theorem 3.2.2. *Consider a DCBM where (3.2.1)-(3.2.4) hold. With probability at least $1 - O(n^{-3})$, for each $1 < m \leq K$, $\Pi \preceq \widehat{\Pi}^{(m)}$ up to a permutation in the columns.*

Theorem 3.2.2 says that SCORE has the NSP and is proved in Section 3.3. The theorem is the key to our study of the upper bound below. In Section 3.3, we explain the main technical challenges we face in proving the theorem, and present the key theorem and lemmas required for the proof. Why SCORE has the NSP is non-obvious, so to shed light on this, we present an intuitive explanation in Section 3.3. The following theorem is proved in Section 3.4.

Theorem 3.2.3. *Fix $0 < \alpha < 1$. Suppose we apply StGoF to a DCBM model where (3.2.1)-(3.2.4) hold. As $n \to \infty$, $\min_{1 \leq m < K}\{\psi_n^{(m)}\} \to \infty$ in probability and $\mathbb{P}(\widehat{K}_\alpha^* \neq K) \leq \alpha + o(1)$.*

Theorem 3.2.3 allows for severe degree heterogeneity. When the degree heterogeneity is moderate, SNR $\asymp |\lambda_K|/\sqrt{\lambda_1}$ and we have the following corollary.

Corollary 3.2.2. *Fix $0 < \alpha < 1$. Suppose we apply StGoF to a DCBM model where (3.2.1)-(3.2.2) and (3.2.4) hold. Suppose $\theta_{max} \leq C\theta_{min}$ and $|\lambda_K|/\sqrt{\lambda_1} \geq C_0\sqrt{\log(n)}$ for a sufficiently large constant $C_0 > 0$. As $n \to \infty$, $\min_{1 \leq m < K}\{\psi_n^{(m)}\} \to \infty$ in probability and $\mathbb{P}(\widehat{K}_\alpha^* \neq K) \leq \alpha + o(1)$.*

Note that in Theorem 3.2.3 and Corollary 3.2.2, if we let α depend on n and tend to 0 slowly enough, then we have $\mathbb{P}(\widehat{K}_\alpha^* = K) \to 1$.

Remark 4. While the NSP of SCORE largely facilitates the analysis, it does not mean that StGoF ceases to work well once NSP does not hold; it is just harder to analyze in such cases. Numerical experiments confirm that StGoF continues to behave well even when NSP does not hold exactly. How to analyze StGoF in such cases is an interesting problem for the future.

Remark 5. In this chapter, we assume K is fixed. For diverging K, the main idea continues to be valid, but we need to revise several things (e.g., definition of consistency and SNR, some regularity conditions, phase transition) to reflect the role of K. The proof for the case of diverging K can be much more tedious, but aside from that, we do not see a major technical hurdle. Especially, the NSP of SCORE continues to hold for a diverging K. Then, with some mild conditions, we can show that $\widehat{\Pi}^{(m)}$ has very few realizations, so the analysis of StGoF is readily extendable. That we assume K as fixed is not only for simplicity but also for practical relevance. For example, real networks may have hierarchical tree structure, and in each layer, the number of leaves (i.e., clusters) is small (e.g., Ji and Jin (2016); Ji et al. (2020); Lei et al. (2020); Li et al. (2018)). Therefore, we have small K in each layer when we perform hierarchical network analysis. Also, the goal of real applications is to have interpretable results. For example, for community detection, results with a large K is hard to interpret, so we may prefer a DCBM with a small K to an SBM with a large K. In this sense, a small K is practically more relevant.

Remark 6. Conditions (3.2.3) is the main condition that ensures (a) SCORE yields exact community recovery when $m = K$, and (b) SCORE has the NSP when $1 \leq m < K$. The condition is much weaker than those in existing works (e.g., Wang et al. (2017), Ma et al. (2018)), and can not be significantly improved in the case of $\theta_{max} \leq C\theta_{min}$ (see phase transition results in Section 3.2.4). The more difficult case where $\theta_{max}/\theta_{min}$ tends to ∞ rapidly has never been studied before, at least for estimating K, and it is unclear whether we can find an alternative algorithm that satisfies (a)-(b) under a significantly weaker condition than (3.2.3). On the other hand, we can view StGoF as a general framework for estimating K, where SCORE may be improved or replaced by some other procedures satisfying (a)-(b) in the future as researchers continue to make advancements in this area, so whether (3.2.3) can be further improved does not affect our main contributions (see Section 1.1 for our contributions).

3.2.4 Information lower bound and phase transition

In Theorem 3.2.3 and Corollary 3.2.2, we require the SNR, $|\lambda_K|/\sqrt{\lambda_1}$, to tend to ∞ at a speed of at least $\sqrt{\log(n)}$. Such a condition cannot be significantly relaxed. For example, if SNR $\to 0$, then it is impossible to have a consistent estimate for K. The exact meaning of this is described below.

We say two DCBM models are asymptotically indistinguishable if for any test that tries to decide which model is true, the sum of Type I and Type II errors is no smaller than $1 + o(1)$, as $n \to \infty$. Given a DCBM with K communities, our idea is to construct a DCBM with $(K + m)$ communities for any $m \geq 1$, and show that two DCBM are asymptotically indistinguishable, provided that the SNR of the latter is $o(1)$.

In detail, fixing $K_0 \geq 1$, consider a DCBM with K_0 communities that satisfies (3.1.1)-(3.1.4). Let $(\Theta, \widetilde{\Pi}, \widetilde{P})$ be the parameters of this DCBM, and let $\widetilde{\Omega} = \Theta\widetilde{\Pi}\widetilde{P}\widetilde{\Pi}'\Theta$. When $K_0 > 1$, let $(\beta', 1)'$ be the last column of $\widetilde{P}$, and let S be the sub-matrix of $\widetilde{P}$ excluding the last row and the last column. Given $m \geq 1$ and $b_n \in (0, 1)$, we construct a DCBM model with $(K_0 + m)$ communities as follows. We define a $(K_0 + m) \times (K_0 + m)$ matrix P:

$$P = \begin{bmatrix} S & \beta \mathbf{1}'_{m+1} \\ \mathbf{1}_{m+1}\beta' & \frac{m+1}{1+mb_n} M \end{bmatrix}, \qquad \text{where} \quad M = (1 - b_n)I_{m+1} + b_n \mathbf{1}_{m+1}\mathbf{1}'_{m+1}. \tag{3.2.13}$$

When $K_0 = 1$, we simply let $P = \frac{m+1}{1+mb_n} M$. Let $\widetilde{\ell}_i \in \{1, \ldots, K_0\}$ be the community label of node i defined by $\widetilde{\Pi}$. We generate labels $\ell_i \in \{1, \ldots, K_0 + m\}$ by

$$\ell_i = \begin{cases} \widetilde{\ell}_i, & \text{if } \widetilde{\ell}_i \in \{1, \ldots, K_0 - 1\}, \\ \text{uniformly drawn from } \{K_0, K_0 + 1, \ldots K_0 + m\}, & \text{if } \widetilde{\ell}_i = K_0. \end{cases} \tag{3.2.14}$$

Let Π be the corresponding matrix of community labels. This gives rise to a DCBM model with $(K_0 + m)$ communities, where $\Omega = \Theta\Pi P\Pi'\Theta$. Note that P does not have unit diagonals, but we can re-parametrize so that it has unit diagonals. In detail, introduce a $(K_0+m)\times(K_0+m)$ diagonal matrix D where $D_{kk} = \sqrt{P_{kk}}$, $1 \le k \le K_0 + m$. Now, if we let $P^* = D^{-1}PD^{-1}$, $\theta_i^* = \theta_i\|D\pi_i\|_1$, and $\Theta^* = diag(\theta_1^*, \ldots, \theta_n^*)$, then P^* has unit-diagonals and $\Omega = \Theta^*\Pi P^*\Pi'\Theta^*$.

Here some rows of Π are random (so we may call the corresponding model the random-label DCBM), but this is conventional in the study of lower bounds. Let λ_k be the kth largest eigenvalue (in magnitude) of Ω. Since Ω is random, λ_k's are also random (but we can bound $|\lambda_K|/\sqrt{\lambda_1}$ conveniently). The following theorem is proved in the supplementary material.

Theorem 3.2.4. *Fix $K_0 \ge 1$ and consider a DCBM model with n nodes and K_0 communities, whose parameters $(\theta, \widetilde{\Pi}, \widetilde{P})$ satisfy (3.2.1)-(3.2.2). Let $(\beta', 1)'$ be the last column of $\widetilde{P}$, and let S be the sub-matrix of $\widetilde{P}$ excluding the last row and last column. We assume $|\beta' S^{-1}\beta - 1| \ge C$.*

- *Fix $m \ge 1$. Given any $b_n \in (0,1)$, we can construct a random-label DCBM model with $K = K_0 + m$ communities as in (3.2.13)-(3.2.14). Then, as $n \to \infty$, $|\lambda_K|/\sqrt{\lambda_1} \le C\|\theta\|(1-b_n)$ with probability $1 - o(n^{-1})$. Moreover, if $(1-b_n)/|\lambda_{\min}(S)| = o(1)$, where $\lambda_{\min}(S)$ is the minimum eigenvalue (in magnitude) of S, then $|\lambda_K|/\sqrt{\lambda_1} \ge C^{-1}\|\theta\|(1-b_n)$ with probability $1 - o(n^{-1})$. Here $C > 1$ is a constant that does not depend on b_n.*

- *Fix $m_1, m_2 \ge 1$ with $m_1 \ne m_2$. As $n \to \infty$, if $\|\theta\|(1-b_n) \to 0$, then the two random-label DCBM models associated with m_1 and m_2 are asymptotically indistinguishable.*

By Theorem 3.2.4, starting from a (fixed-label) DCBM with K_0 communities, we can construct a collection of random-label DCBM, with $K_0+1, K_0+2, \ldots, K_0+m$ communities, respectively, where (a) for the model with (K_0+m) communities, $|\lambda_{K_0+m}|/\sqrt{\lambda_1} \asymp \|\theta\|(1-b_n)$, with an overwhelming probability, and (b) each pair of models are asymptotically indistinguishable if $\|\theta\|(1-b_n) = o(1)$. Therefore, for a broad class of DCBM with unknown K where SNR $= o(1)$ for some models, a consistent estimate for K does not exist.

Fixing $m_0 > 1$ and a sequence of numbers $a_n > 0$, let $\mathcal{M}_n(m_0, a_n)$ be the collection of DCBM for an n-node network with K communities, where $1 \le K \le m_0$, $|\lambda_K|/\sqrt{\lambda_1} \ge a_n$, and (3.2.1)-(3.2.2) hold. In Section 3.2.3, we show that if $a_n \ge C_0\sqrt{\log(n)}$ for a sufficiently large constant C_0, then for each DCBM in $\mathcal{M}_n(m_0, a_n)$, StGoF provides a consistent estimate for K. The following theorem says that, if we allow $a_n \to 0$, then $\mathcal{M}_n(m_0, a_n)$ is too broad, and a consistent estimate for K does not exist.

Theorem 3.2.5. *Fix $m_0 > 1$ and let $\mathcal{M}_n(m_0, a_n)$ be the class of DCBM as above. As $n \to \infty$, if $a_n \to 0$, then $\inf_{\widehat{K}}\{\sup_{\mathcal{M}_n(m_0,a_n)} \mathbb{P}(\widehat{K} \ne K)\} \ge (1/6 + o(1))$, where the probability is evaluated at any given model in $\mathcal{M}_n(m_0, a_n)$ and the supremum is over all such models.*

Combining Theorems 3.2.1, 3.2.5, and Corollary 3.2.2, we have a phase transition result.

- *Impossibility.* If $a_n \to 0$, then $\mathcal{M}_n(m_0, a_n)$ defines a class of DCBM that is too broad where some pairs of models in the class are asymptotically indistinguishable. Therefore, no estimator can consistently estimate the number of communities for each model in the class. In this case, we can say "a consistent estimate for K does not exist" for short.

- *Possibility.* If $a_n \geq C_0\sqrt{\log(n)}$ for a sufficiently large C_0, then for every DCBM in $\mathcal{M}_n(m_0, a_n)$, StGoF provides a consistent estimate for the number of communities if the model only has moderate degree heterogeneity (i.e., $\theta_{max} \leq C\theta_{min}$). StGoF continues to be consistent in the presence of severe degree heterogeneity if the adjusted SNR satisfies that $s_n \geq C_0\sqrt{\log(n)}$ with a sufficiently large C_0.

The case of $C \leq a_n < C_0\sqrt{\log(n)}$ is more delicate. Sharp results are possible if we consider more specific models (e.g., for a scaling parameter $\alpha_n > 0$, (θ_i/α_n) are iid from a fixed distribution F, and the off-diagonals of P are the same). We leave this to the future.

3.3 The non-splitting property (NSP) of SCORE

This section contains the second part of our main theoretical results. We first present the main technical tools for proving Theorem 3.2.2 (i.e., the NSP of SCORE), and then prove Theorem 3.2.2. Why NSP holds is non-obvious, so in Section 3.3.3, we also shed light by providing an intuitive explanation and several examples. The NSP may hold in many other unsupervised learning settings, and the gained insight in Section 3.3.3 may serve as a good starting point for studying NSP in these settings.

Here, the primary focus of our study on SCORE is on the under-fitting case of $m < K$, while existing study on SCORE (e.g., Jin (2015)) has been focused on the null case of $m = K$. In the last two paragraphs of Section 3.1.1, we have briefly explained why the study in the under-fitting case is much harder. This section will further explain this with details.

Recall that in the SCORE step, for each $1 < m \leq K$, we apply the k-means to the rows of an $n \times (m-1)$ matrix $\widehat{R}^{(m)}$, where $\widehat{R}^{(m)}(i,k) = \widehat{\xi}_{k+1}(i)/\widehat{\xi}_1(i)$, $1 \leq i \leq n$, $1 \leq k \leq m-1$, and $\widehat{\xi}_k$ is the k-th eigenvector (eigenvectors are arranged in the descending order in magnitudes of corresponding eigenvalues) of the adjacency matrix A. Viewing each row of $\widehat{R}^{(m)}$ as a point in $\mathbb{R}^{m-1}$, we will show that there is a polytope in $\mathbb{R}^{m-1}$ with vertices $v_1, v_2, \ldots, v_K$ such that with large probability, row i of $\widehat{R}^{(m)}$ falls close to v_k if node i belongs to the true community k, for all $1 \leq i \leq n$. Therefore, the n rows form K clusters (but K and true cluster labels are unknown), each being a true community. To show that SCORE satisfies the NSP, the goal is to show that the k-means algorithm will not split any of these K clusters. See Figure 3.2 where we illustrate the NSP with an example with $(K, m) = (4, 3)$.

Definition 29 (Bottom up pruning and minimum pairwise distances)**.** *Fixing $K > 1$ and $1 < m \leq K$, consider a $K \times (m-1)$ matrix $U = [u_1, u_2, \ldots, u_K]'$. First, let $d_K(U)$ be the minimum pairwise distance of all K rows. Second, let u_k and u_ℓ ($k < \ell$) be the pair that satisfies $\|u_k - u_\ell\| = d_K(U)$ (if this holds for multiple pairs, pick the first pair in the lexicographical order). Remove row ℓ from the matrix U and let $d_{K-1}(U)$ be the*

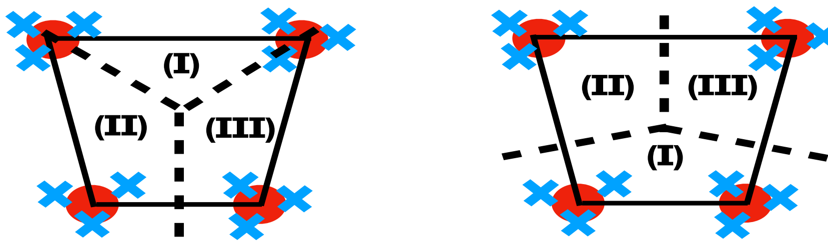

Figure 3.2: Illustration for what NSP means ($(K, m) = (4, 3)$). The rows of $\widehat{R}^{(m)}$ (blue crosses) form K clusters (red: cluster centers) each of which is a true community (K and true cluster labels are unknown). SCORE aims to cluster all rows of $\widehat{R}^{(m)}$ into m clusters. Left: Voronoi diagram of k-means when the NSP does not hold (which will not happen according to our proof). Right: Voronoi diagram when the NSP holds.

minimum pairwise distance for the remaining $(K-1)$ rows. Repeat this step and define $d_{K-2}(U), d_{K-3}(U), \ldots, d_2(U)$ recursively. Note that $d_K(U) \le d_{K-1}(U) \le \ldots \le d_2(U)$.

For example, if $(K, m) = (4, 3)$, and the rows of U are $(1, 0), (1, 0), (0, 1)$ and $(1, 1)$, then $d_4(U) = 0$, $d_3(U) = 1$, and $d_2(U) = \sqrt{2}$. The following theorem is the key to prove the NSP of SCORE, and is proved in the supplementary material.

Theorem 3.3.1. *Fix $1 < m \le K$ and let n be sufficiently large. Suppose $x_1, x_2, \ldots, x_n \in \mathbb{R}^{m-1}$ take K distinct values $u_1, u_2, \ldots, u_K$. Letting $U = [u_1, u_2, \ldots, u_K]'$ and $F_k = \{1 \le i \le n : x_i = u_k\}$, for $1 \le k \le K$, suppose $\min_{1 \le k \le K} |F_k| \ge \alpha_0 n$ and $\max_{1 \le k \le K} \|u_k\| \le C_0 \cdot d_m(U)$, for constants $0 < \alpha_0 < 1$, $C_0 > 0$. Suppose we apply k-means to a set of n points $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n$ assuming m clusters. Let $\widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_m$ be the resultant clusters (which are not necessarily unique). There is a number $c = c(\alpha_0, C_0, m) > 0$ such that if $\max_{1 \le i \le n} \|\widehat{x}_i - x_i\| \le c \cdot d_m(U)$, then $\#\{1 \le j \le m : \widehat{S}_j \cap F_k \neq \emptyset\} = 1$, for each $1 \le k \le K$.*

When we apply Theorem 3.3.1 to prove Theorem 3.2.2, all conditions required in Theorem 3.3.1 can be deduced from those in Theorem 3.2.2, so we do not need any additional conditions. See Lemma 33 and Section 3.3.2. Theorem 3.3.1 is a general result on k-means and may be useful in many other unsupervised settings. The proof is non-trivial for the following reasons.

- The objective function of the k-means is complicate, and the k-means solution is not necessarily unique. See Example 3.
- Theorem 3.3.1 only requires that there are at least m true cluster centers the minimum pairwise distance of which is large. If we assume a stronger condition, say, the minimum pairwise distance of all K cluster centers is large (i.e., $\max_{1 \le k \le K} \|u_k\| \le C_0 \cdot d_K(U)$), the proof is much easier, but unfortunately, such a condition does not always hold in our settings. See Example 4 below.

Example 3. Suppose $(K, m) = (4, 3)$ and F_1, F_2, F_3, F_4 have equal sizes. We view $u_1, u_2, \ldots, u_K$ as the vertices of a quadrilateral in $\mathbb{R}^2$. Suppose we apply the k-means to

$x_1, x_2, \ldots, x_n$ and let C_1, C_2, C_3 be the resultant clusters. Suppose that among the 6 different pairs of vertices, (u_1, u_2) is the pair with the smallest distance. In this case, the three clusters are $\mathcal{C}_1 = F_1 \cup F_2$, $C_2 = F_3$, and $C_3 = F_4$, and the cluster centers are $(u_1 + u_2)/2$, u_3, and u_4. If the quadrilateral is a square or rectangle, then among the 6 pairs of indices, more than one pairs have the smallest pairwise distance, so the k-means solutions are not unique.

Now, to prove Theorem 3.2.2, the idea is to apply Theorem 3.3.1 with $\widehat{x}_i$ being row i of $\widehat{R}^{(m)}$. To do this, we study the geometrical structure underlying $\widehat{R}^{(m)}$ in the under-fitting case, where the ideal polytope and tight row-wise large deviation bounds for $\widehat{R}^{(m)}$ play a key role.

3.3.1 Geometric structure, ideal polytopes, and row-wise bounds

Fix $1 \leq k \leq K$. Let λ_k be the k-th largest (in magnitude) eigenvalue of the $n \times n$ matrix Ω and let ξ_k be the corresponding unit-ℓ^2-norm eigenvector. By Davis-Kahan $\sin(\theta)$-theorem Davis and Kahan (1970), the two matrices $[\xi_1, \ldots, \xi_K]$ and $[\widehat{\xi}_1, \ldots, \widehat{\xi}_K]$ only match well with each other by a rotation matrix Γ: $[\widehat{\xi}_1, \ldots, \widehat{\xi}_K] \approx [\xi_1, \ldots, \xi_K]\Gamma$. Let Ξ be the matrix consisting of the first m columns of $[\xi_1, \ldots, \xi_K]\Gamma$. The geometrical structure underlying Ξ is the key to our study.

In the null case of $m = K$, the geometric structure was studied in Jin (2015); Jin et al. (2017). For the under-fitting case of $1 < m < K$, the study is much harder. The reason is that, Γ is hard to track without a strong condition on the eigen-gap of Ω, and as Γ ranges, the row-wise distances of Ξ remain the same when $m = K$, but may vary significantly when $m < K$. To deal with this, we need relatively tedious notations and harder proofs, compared to those in Jin (2015); Jin et al. (2017).

Recall that μ_k is the k-th largest (in magnitude) eigenvalue of the $K \times K$ matrix HPH, and η_k is the corresponding unit-ℓ^2-norm eigenvector. We now relate (μ_k, η_k) to (λ_k, ξ_k) above. The following lemma is proved in the supplementary material.

Lemma 30. *Consider a DCBM where (3.2.4) holds and let $\lambda_k, \mu_k, \eta_k, \xi_k$ be as above. We have the following claims. First, $\lambda_k = \|\theta\|^2 \mu_k$ for $1 \leq k \leq K$. Second, the multiplicity of μ_1 is 1 and all entries of η_1 have the same sign, and the same holds for λ_1 and ξ_1. Last, if η_k is an eigenvector of HPH corresponding to μ_k, then $\|\theta\|^{-1}\Theta\Pi H^{-1}\eta_k$ is an eigenvector of Ω corresponding to λ_k, and conversely, if ξ_k is an eigenvector of Ω corresponding to λ_k, then $\|\theta\|^{-1}H^{-1}\Pi'\Theta\xi_k$ is an eigenvector of HPH corresponding to μ_k.*

From now on, let η_1 be the unique unit-ℓ^2-norm eigenvector of HPH corresponding to λ_1 that have all positive entries. Note that $\eta_2, \ldots, \eta_K$ may not be unique. Fix a particular candidate for $\eta_2, \ldots, \eta_K$, say, $\eta_2^*, \ldots, \eta_K^*$. Let

$$[\xi_1, \xi_2^*, \ldots, \xi_K^*] = \|\theta\|^{-1}\Theta\Pi H^{-1}[\eta_1, \eta_2^*, \ldots, \eta_K^*]. \tag{3.3.15}$$

Definition 31. *Given any $(K-1) \times (K-1)$ orthogonal matrix Γ and $2 \leq k \leq K$, let $\eta_k(\Gamma)$ be the $(k-1)$-th column of $[\eta_2^*, \eta_3^*, \ldots, \eta_K^*]\Gamma$, with $\eta_k(i, \Gamma)$ being the i-th entry, $1 \leq i \leq K$,* *and let* $\xi_k(\Gamma)$ *be the* $(k-1)$*-th column of* $[\xi_2^*, \xi_3^*, \ldots, \xi_K^*]\Gamma$*, with* $\xi_k(j, \Gamma)$ *being the* j*-th entry,* $1 \leq j \leq n$.

Note that (η_1, ξ_1) are uniquely defined (up to a factor of ± 1), but $\{(\eta_k, \xi_k)\}_{2\leq k\leq K}$ are not necessarily unique. However, by Lemma 30 and basic linear algebra, there is a collection of $(K-1)\times(K-1)$ orthogonal matrices, denoted by $\mathcal{A}$, such that when Γ ranges in $\mathcal{A}$, $\{\eta_2(\Gamma), \ldots, \eta_K(\Gamma)\}$ give all possible candidates of $\{\eta_2, \ldots, \eta_K\}$, and $\{\xi_2(\Gamma), \ldots, \xi_K(\Gamma)\}$ give all possible candidates of $\{\xi_2, \ldots, \xi_K\}$. In the special case where $\mu_2, \ldots, \mu_K$ are distinct, $\mathcal{A}$ is the set of all $(K-1)\times(K-1)$ diagonal orthogonal matrices, and in the special case where $\mu_2 = \ldots = \mu_K$, $\mathcal{A}$ is the set of all $(K-1)\times(K-1)$ orthogonal matrices.

Fix $1 < m \leq K$ and a $(K-1)\times(K-1)$ orthogonal matrix Γ (which is not necessarily in $\mathcal{A}$). We define a $K\times(m-1)$ matrix $V^{(m)}(\Gamma)$ and an $n\times(m-1)$ matrix $R^{(m)}(\Gamma)$ by

$$V^{(m)}(k, \ell; \Gamma) = \eta_{\ell+1}(k; \Gamma)/\eta_1(k), \qquad 1 \leq k \leq K,\ 1 \leq \ell \leq m-1, \tag{3.3.16}$$

and

$$R^{(m)}(i, \ell; \Gamma) = \xi_{\ell+1}(i; \Gamma)/\xi_1(i), \qquad 1 \leq i \leq n,\ 1 \leq \ell \leq m-1. \tag{3.3.17}$$

We note that $V^{(m)}(\Gamma)$ is the sub-matrix of $V^{(K)}(\Gamma)$ consisting the first $(m-1)$ columns; same comments for $R^{(m)}(\Gamma)$. Write $V^{(m)}(\Gamma) = [v_1^{(m)}(\Gamma), \ldots, v_K^{(m)}(\Gamma)]'$ and $R^{(m)}(\Gamma) = [r_1^{(m)}(\Gamma), \ldots, r_n^{(m)}(\Gamma)]'$, so that $(v_k^{(m)}(\Gamma))'$ is the k-th row of $V^{(m)}(\Gamma)$ and $(r_i^{(m)}(\Gamma))'$ is the i-th row of $R^{(m)}(\Gamma)$, $1 \leq k \leq K, 1 \leq i \leq n$. For notational simplicity, we may drop "Γ" when there is no confusion. Recall that for $1 \leq k \leq K$, $\mathcal{N}_k$ denotes the k-th true community. The following lemma is proved in the appendix.

Lemma 32 (The ideal polytope)**.** *Consider a DCBM model where (3.2.4) holds. For any* $1 < m \leq K$ *and fixed* $(K-1)\times(K-1)$ *orthogonal matrix* Γ, $r_i^{(m)}(\Gamma) = v_k^{(m)}(\Gamma)$*, for any* $i \in \mathcal{N}_k$ *and* $1 \leq k \leq K$.

Therefore, the n rows of $R^{(m)}(\Gamma)$ have at most K distinct values, $(v_1^{(m)}(\Gamma))'$, $(v_2^{(m)}(\Gamma))'$, ..., $(v_K^{(m)}(\Gamma))'$. For an "easy" setting, $d_K(V^{(m)}(\Gamma)) \geq C$, so the minimum pairwise distance of these K rows are large. In a more "difficult" case, we may have $d_K(V^{(m)}(\Gamma)) = 0$. However, we can always find m rows of $V^{(m)}(\Gamma)$ so that the minimum pairwise distance of which is no smaller than a constant C. This is the following lemma, which is proved in the supplementary material.

Lemma 33. *Consider a DCBM model where* (3.2.2) *and* (3.2.4) *hold. Fix* $1 \leq m \leq K$ *and an* $(K-1)\times(K-1)$ *orthogonal matrix* Γ*, we have* $d_m(V^{(m)}(\Gamma)) \geq \sqrt{2}$ *when* $m = K$*, and* $d_m(V^{(m)}(\Gamma)) \geq C$ *when* $1 < m < K$*, where the constant* $C > 0$ *does not depend on* Γ.

We should not expect that $d_K(V^{(m)}(\Gamma)) \geq C$ holds for all rotation Γ. We can only show a weaker claim of $d_m(V^{(m)}(\Gamma)) \geq C$ as in Lemma 33. Below, we use a special example to illustrate how Γ affect $d_K(V^{(m)}(\Gamma))$.

Example 4. Consider a special case of Example 2 where $P = (1-b_n)I_K + b_n\mathbf{1}_K\mathbf{1}_K'$, $0 < b_n < 1$, and $\|\theta^{(k)}\| = \|\theta\|/\sqrt{K}$, $1 \leq k \leq K$ (as a result, $HPH = (1/K)P$). Note that the eigenvectors of HPH, denoted by $\eta_1, \eta_2, \ldots, \eta_K$, do not depend on b_n. We take the case of $(K, m) = (3, 2)$ for example. In this case, $\eta_1 = (1/\sqrt{3})[1, 1, 1]'$, and a candidate for $\{\eta_2, \eta_3\}$ is $\eta_2^* = (1/\sqrt{2})[1, -1, 0]'$, and $\eta_3^* = (1/\sqrt{6})[1, 1, -2]'$, and all possible candidates for $\{\eta_2, \eta_3\}$ are given by

$$[\eta_2^*, \eta_3^*]\Gamma, \qquad \Gamma = \Gamma(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \qquad 0 \leq \theta < 2\pi.$$

Now, $d_3(V^{(2)}(\Gamma))$ changes continuously in θ and take values in $[0, \sqrt{3}/\sqrt{2}]$, and hits 0 when $\theta \in \{\pi/6,\ \pi/2, 5\pi/6, 7\pi/6, 3\pi/2, 11\pi/6\}$. However, $d_2(V^{(2)}(\Gamma)) \geq \sqrt{3}/\sqrt{2}$ for all θ.

Similarly, we write $\widehat{R}^{(m)} = [\widehat{r}_1^{(m)}, \widehat{r}_2^{(m)}, \ldots, \widehat{r}_n^{(m)}]'$, so that $(\widehat{r}_i^{(m)})'$ is the i-th row of $\widehat{R}^{(m)}$. The following lemma provides a tight row-wise large-deviation bound for $\widehat{R}^{(m)}$ and is proved in the supplementary material.

Lemma 34. *Consider a DCBM model where (3.2.1)-(3.2.4) hold. With probability* $1 - O(n^{-3})$, *there exists a* $(K-1) \times (K-1)$ *orthogonal matrix* Γ *(which may depend on* n *and* $\widehat{R}^{(K)}$*) such that as* $n \to \infty$, $\|\widehat{r}_i^{(m)} - r_i^{(m)}(\Gamma)\| \leq \|\widehat{r}_i^{(K)} - r_i^{(K)}(\Gamma)\| \leq C s_n^{-1}\sqrt{\log(n)}$, *for all* $1 < m < K$ *and* $1 \leq i \leq n$ *.*

For illustration, we assume $d_K(V^{(m)}) \geq C$ for all $1 < m \leq K$ (we have dropped "Γ" to simplify notations) so the minimum pairwise distance of the K rows of $V^{(m)}$ is no smaller than C. In this case, Lemmas 32-34 say that the n rows of $R^{(m)}$ have K distinct values, $(v_1^{(m)})', (v_2^{(m)})', \ldots, (v_K^{(m)})'$, and partitioning the rows with respect to different values gives exactly K true communities. Note that we can view $v_1^{(m)}, v_2^{(m)}, \ldots, v_K^{(m)}$ as the vertices of a polytope in $\mathbb{R}^{m-1}$. See Figure 3.3 for an illustration of $K = 4$. In this case, $v_1^{(m)}, v_2^{(m)}, \ldots, v_K^{(m)}$ are the vertices of a tetrahedron when $m = 4$, the vertices of a quadrilateral when $m = 3$, and K scalars when $K = 2$. By Lemma 34 and the condition (3.2.3), for all $1 \leq i \leq n$, $\|\widehat{r}_i^{(m)} - r_i^{(m)}\|$ is much smaller than $d_K(V^{(m)})$. Therefore, the n rows of $\widehat{R}^{(m)}$ also form K clusters, each being a true community. If we apply k-means assuming K clusters, then we can fully recover the true communities. Unfortunately, K is unknown. In the under-fitting case, $m < K$ and we under-estimate the number of clusters. However, Theorem 3.3.1 guarantees that, although we are not able to recover all true communities, the NSP holds.

3.3.2 Proof of Theorem 3.2.2

By Lemma 34, there is an event E, where $\mathbb{P}(E^c) = O(n^{-3})$, and on this event there exists a $(K-1) \times (K-1)$ orthogonal matrix Γ (which may depend on n and $\widehat{R}^{(K)}$) such that

$$\max_{1 \leq i \leq n} \|\widehat{r}_i^{(m)} - r_i^{(m)}(\Gamma)\| \leq C s_n^{-1}\sqrt{\log(n)}, \qquad \text{for all } 1 < m \leq K.$$

Fix $1 < m \leq K$. By Lemma 32, $r_i^{(m)}(\Gamma) = v_k^{(m)}(\Gamma)$ for each $i \in \mathcal{N}_k$ and $1 \leq k \leq K$. Suppose $v_1^{(m)}(\Gamma), \ldots, v_K^{(m)}(\Gamma)$ have L distinct values, where L may depend on m and Γ and $L \geq m$ by Lemma 33. Note that whenever two vectors (say) $v_1^{(m)}(\Gamma)$ and $v_2^{(m)}(\Gamma)$ are identical, we can always treat $\mathcal{N}_1$ and $\mathcal{N}_2$ as the same cluster before we apply Theorem 3.3.1. Therefore,

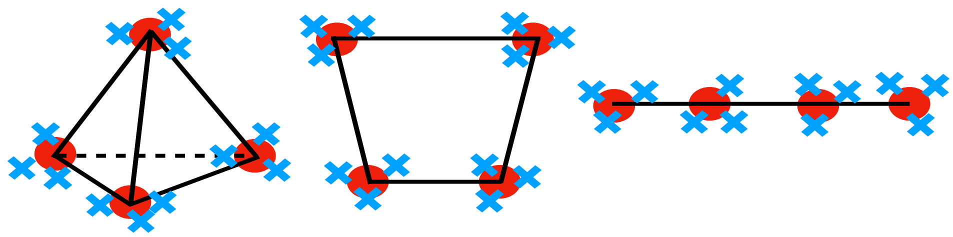

Figure 3.3: An example ($K = 4$). From left to right: $m = 4, 3, 2$. Red dots: the 4 distinct rows of $R^{(m)}$, $v_1^{(m)}, v_2^{(m)}, v_3^{(m)}, v_4^{(m)}$. Blue crosses: rows of $\widehat{R}^{(m)}$. The red dots are the vertices of a tetrahedron when $m = 4$, vertices of a quadrilateral when $m = 3$, and scalars when $m = 2$. For each m, the n rows of $\widehat{R}^{(m)}$ are seen to have K clusters, each of which is a true community.

without loss of generality, we assume $L = K$, so $v_1^{(m)}(\Gamma), \ldots, v_K^{(m)}(\Gamma)$ are distinct. It suffices to show that, on the event E, none of $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_K$ is split by the k-means.

We now apply Theorem 3.3.1 with $\widehat{x}_i = \widehat{r}_i^{(m)}$, $x_i = r_i^{(m)}(\Gamma)$, $F_k = \mathcal{N}_k$, and $U = V^{(m)}(\Gamma)$. Note that by Lemma 33, $d_m(U) \geq C$. Also, in the proof of Lemma 33, we have shown that $\max_{1\leq k\leq K} \|v_k^{(m)}(\Gamma)\| \leq C$. It follows that the ℓ^2-norm of each row of U is bounded by $C \cdot d_m(U)$. Additionally, on the event E, $\max_{1\leq i\leq n} \|\widehat{x}_i - x_i\| \leq C s_n^{-1}\sqrt{\log(n)}$. As long as $s_n \geq C_0\sqrt{\log(n)}$ for a sufficiently large constant C_0, we have $\max_{1\leq i\leq n} \|\widehat{x}_i - x_i\| \leq c \cdot d_m(U)$ for a sufficiently small constant c. The claim now follows by applying Theorem 3.3.1. $\square$

3.3.3 Why NSP holds: intuitive explanations and examples

Why NSP holds is non-obvious, so we provide an intuitive explanation and some examples. The NSP may hold for many other unsupervised learning settings, and this section may be especially helpful if we wish to extend our ideas to other settings. Since the NSP in general settings is already proved above and the purpose here is to provide some insight, we consider settings where

$$d_K(V^{(m)}(\Gamma)) \geq C. \tag{3.3.18}$$

This condition is stronger than the condition $d_m(V^{(m)}(\Gamma) \geq C$ needed in Theorem 3.3.1 (e.g., see Example 4). Also, for notational simplicity, we drop "Γ" below.

We start by introducing the *minimum gap* as a measure for the stability of the clustering results by k-means. Fixing $1 < m \leq K$, consider n points $u_1, u_2, \ldots, u_n \in \mathbb{R}^{m-1}$ and let $U = [u_1, u_2, \ldots, u_n]'$. Suppose we cluster $u_1, u_2, \ldots, u_n$ into m clusters using the k-means.

Definition 35. *Let $c_1, c_2, \ldots, c_m$ be any possible cluster centers from k-means (the set is not necessarily unique). Let $d_1(u_i; c_1, \ldots, c_m)$ and $d_2(u_i; c_1, \ldots, c_m)$ be the distances between u_i and its closest cluster center and the distance between u_i and its second closest cluster center, respectively. The minimum gap for the clustering results is defined by*

$$g_m(U) = \min_{\{\text{all possible } c_1, c_2, \ldots, c_m\}} \min_{1\leq i\leq n} \big\{d_2(u_i; c_1, \ldots, c_m) - d_1(u_i; c_1, \ldots, c_m)\big\}.$$

We now explain why NSP holds for the under-fitting case. We start by considering the oracle case where we apply k-means to the n rows of the non-stochastic matrix $R^{(m)}(\Gamma)$.

Theorem 3.3.2. *Consider a DCBM model where* (3.2.2) *holds. Fix* $1 < m < K$ *and any* $(K-1) \times (K-1)$ *orthogonal matrix* Γ*. Let* $V^{(m)}(\Gamma)$ *and* $R^{(m)}(\Gamma)$ *be as in (3.3.16) and (3.3.17), respectively. If* $d_K(V^{(m)}(\Gamma)) > 0$ *and we apply the k-means to rows of* $R^{(m)}(\Gamma)$*, then NSP holds and* $g_m(R^{(m)}(\Gamma)) \geq C d_K(V^{(m)}(\Gamma))$*, where* C *only depends on the constant in* (3.2.2).

Theorem 3.3.2 is proved in the supplementary material. In the oracle case, since $r_i^{(m)} = r_j^{(m)}$ when i and j are in the same community, the NSP must hold once we have $g_m(R^{(m)}) > 0$ (otherwise we can easily find a contradiction). At the same time, it is less obvious why $g_m(R^{(m)}) \geq C d_K(V^{(m)})$ holds. Below, we use two examples for further illustration. In these examples, we assume $K = 4$, and let $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \mathcal{N}_4$ be the true communities. We assume these communities have equal sizes. We consider the cases of $m = 2$ and $m = 3$, separately.

Example 5a. When $m = 3$, the four points $v_1^{(m)}, \ldots, v_4^{(m)}$ are the vertices of a quadrilateral in $\mathbb{R}^2$. Following Example 3, it is seen $g_m(R^{(m)}) \geq (1/2)\|v_1^{(m)} - v_2^{(m)}\| \equiv (1/2) d_K(V^{(m)})$.

Example 5b. When $m = 2$, $v_1^{(m)}, \ldots, v_4^{(m)}$ are scalars. Without loss of generality, we assume $v_1^{(m)} < v_2^{(m)} < v_3^{(m)} < v_4^{(m)}$. In Section 3.B.7, we show that $g_m(R^{(m)}) \geq [(3-\sqrt{3})/2] \cdot d_K(V^{(m)})$.

In the real case, we take an intuitive approach to explain why NSP holds for the k-means (see Theorem 3.3.1 for a rigorous proof). Recall that $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_K$ are the true communities. Suppose we apply the k-means to the rows of $\widehat{R}^{(m)}$ and obtain m clusters with centers $\widehat{c}_1, \widehat{c}_2, \ldots, \widehat{c}_m$. Suppose we also apply the k-means to the rows of $R^{(m)}$ and obtain m clusters $c_1, c_2, \ldots, c_m$. Under some regularity conditions, we expect to see that

$$\max_{1 \leq k \leq m} \|\widehat{c}_k - c_k\| \leq C \max_{1 \leq i \leq n} \|\widehat{r}_i - r_i\|, \qquad \text{up to a permutation of } c_1, c_2, \ldots, c_m. \tag{3.3.19}$$

By Lemma 34, the right hand side is $\leq C s_n^{-1}\sqrt{\log(n)}$ with large probability. In the k-means on rows of $R^{(m)}$, it follows from Theorem 3.3.2 that every row i for $i \in \mathcal{N}_k$ is clustered into a cluster with center c_j, for some $1 \leq j \leq m$. By Definition 35,

$$\|r_i - c_j\| + g_m(R^{(m)}) \leq \|r_i - c_\ell\|, \qquad \text{for any } \ell \neq j.$$

Combining it with (3.3.19), except for a small probability, for all $i \in \mathcal{N}_k$ and $\ell \neq j$,

$$\|\widehat{r}_i - \widehat{c}_j\| \leq \|r_i - c_j\| + \|\widehat{r}_i - r_i\| + \|\widehat{c}_j - c_j\| \leq \|r_i - c_j\| + C s_n^{-1}\sqrt{\log(n)},$$
$$\|\widehat{r}_i - \widehat{c}_\ell\| \geq \|r_i - c_\ell\| - \|\widehat{r}_i - r_i\| - \|\widehat{c}_\ell - c_\ell\| \geq \|r_i - c_\ell\| - C s_n^{-1}\sqrt{\log(n)}.$$

It follows that

$$\|\widehat{r}_i - \widehat{c}_j\| \leq \|\widehat{r}_i - \widehat{c}_\ell\| + \big[2C s_n^{-1}\sqrt{\log(n)} - g_m(R^{(m)})\big].$$

Therefore, as long as $2C s_n^{-1}\sqrt{\log(n)} < g_m(R^{(m)})$, $\widehat{c}_j$ is the closest cluster center to $\widehat{r}_i$, for every $i \in \mathcal{N}_k$. This shows that except for a small probability, the whole set $\mathcal{N}_k$ is assigned to the cluster with center $\widehat{c}_j$, i.e., NSP holds.

While the above explanation is intuitive and easy to understand, quite strong conditions are needed when we try to solidify each step. For example, while (3.3.19) sounds correct intuitively, it may not hold in some cases when the k-means solutions are not unique. Condition (3.3.18) may not hold in some cases either, due to the rotation aforementioned. To show NSP in the general settings as in our paper, we need Theorem 3.3.1 and Lemmas 32-34. On the other hand, the intuitive explanation here is easy-to-understand, and may provide a starting point for proving NSP in other unsupervised learning settings.

Remark 7. A simpler version of Theorem 3.3.1 was proved in Ma et al. (2018), under stronger conditions of (a) when we apply the k-means to $\{x_1, x_2, \ldots, x_n\}$, the k-means solution is unique, and (b) $d_K(U) \geq C$ (with the same notations as in Theorem 3.3.1). Unfortunately, Ma et al. (2018) only proved their claim for the special case of $(K, m) = (3, 2)$ (for general (K, m), the proof is non-trivial due to complex combinatorics). Also, conditions (a)-(b) are hard to check especially as we need them to hold for $U = V^{(m)}(\Gamma)$ with all Γ and all m; see Examples 3-4. For example, as illustrated in Example 4, when Γ ranges continuously, (b) tends to fail for some m. To make sure (b) holds, Ma et al. (2018) assumes a relatively strong condition (b1): $P \to P_0$ for a fixed matrix P_0 with distinct eigenvalues. This is a strong signal case where $\lambda_1, \lambda_2, \ldots, \lambda_K$ (eigenvalues of Ω) are at the same magnitude, and the eigen-gaps are also at the same magnitude; see Example 1. In this case, the Γ in David-Kahan $\sin(\theta)$ theorem is uniquely determined, so (b) holds. However, our primary interest is in the more challenging weak signal case, where typically $|\lambda_2|/\lambda_1 \to 0$. In this case, (b1) won't hold, because the only P_0 that can be the limit of P is the $K \times K$ matrix of all ones, where the K eigenvalues are not distinct.

3.4 The behavior of the RQ test statistic

In this section, we prove Theorems 3.2.1 and 3.2.3. Corollaries 3.2.1-3.2.2 follow directly from Theorems 3.2.1 and 3.2.3, respectively, so the proofs are omitted. All other theorems and lemmas are proved in the supplementary material.

3.4.1 Proof of Theorem 3.2.1 (the null case of $m = K$)

First, it is seen that the first item is a direct result of Theorem 3.2.2. Second, by definitions,

$$\mathbb{P}(\widehat{K}^*_\alpha \leq K) \geq \mathbb{P}(\psi_n^{(K)} \leq z_\alpha),$$

and so the last item follows once the second item is proved. Therefore, we only need to show the second item. Recall that when $m = K$,

$$\psi_n^{(K)} = [Q_n^{(K)} - B_n^{(K)}]/\sqrt{8C_n},$$

where $Q_n^{(K)}$, $B_n^{(K)}$, and C_n are defined in (3.2.9), (3.2.8) and (3.2.11), respectively, which we reiterate below:

$$Q_n^{(K)} = \sum_{i_1,i_2,i_3,i_4 (dist)} (A_{i_1 i_2} - \widehat{\Omega}^{(K)}_{i_1 i_2})(A_{i_2 i_3} - \widehat{\Omega}^{(K)}_{i_2 i_3})(A_{i_3 i_4} - \widehat{\Omega}^{(K)}_{i_3 i_4})(A_{i_4 i_1} - \widehat{\Omega}^{(K)}_{i_4 i_1}),$$

$$C_n = \sum_{i_1,i_2,i_3,i_4 (dist)} A_{i_1 i_2} A_{i_2 i_3} A_{i_3 i_4} A_{i_4 i_1}, \quad B_n^{(K)} = 2\|\widehat{\theta}\|^4 \cdot [\widehat{g}' \widehat{V}^{-1} (\widehat{P}\widehat{H}^2\widehat{P} \circ \widehat{P}\widehat{H}^2\widehat{P}) \widehat{V}^{-1} \widehat{g}].$$

In the first equation here, $\widehat{\Omega}^{(K)}$ depends on the estimated community label matrix $\widehat{\Pi}^{(K)}$. To facilitate the analysis, it's desirable to replace $\widehat{\Pi}^{(K)}$ by the true membership matrix Π. By the first claim of the current theorem, this replacement only has a negligible effect.

Formally, we introduce $\widehat{\Omega}^{(K,0)}$ to be the proxy of $\widehat{\Omega}^{(K)}$ with $\widehat{\Pi}^{(K)}$ in its definition replaced by Π. Moreover, define $Q_n^{(K,0)}$ to be the proxy of $Q_n^{(K)}$ with $\widehat{\Omega}^{(K)}$ replaced by $\widehat{\Omega}^{(K,0)}$ in its definition, and define the corresponding counterpart of $\psi_n^{(K)}$ as

$$\psi_n^{(K,0)} = [Q_n^{(K,0)} - B_n^{(K)}]/\sqrt{8C_n}.$$

Then, for any fixed number $t \in \mathbb{R}$ we have

$$\left|\mathbb{P}(\psi_n^{(K)} \leq t) - \mathbb{P}(\psi_n^{(K,0)} \leq t)\right| \leq \mathbb{P}(\widehat{\Pi}^{(K)} \neq \Pi) \to 0, \qquad \text{as } n \to \infty,$$

where the last step follows from the first claim in the current theorem. Hence by elementary probability, to prove $\psi_n^{(K)} \to N(0,1)$ in law, if suffices to show $\psi_n^{(K,0)} \to N(0,1)$ in law.

Recall that if we neglect the difference in the main diagonal entries, then $A - \Omega = W$. By definition, we expect that $\widehat{\Omega}^{(K,0)} \approx \Omega$, and so $(A - \widehat{\Omega}^{(K,0)}) \approx W$. This motivates us to define

$$\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}. \tag{3.4.20}$$

At the same time, for short, let b_n and c_n be the oracle counterparts of $B_n^{(K)}$ and C_n

$$c_n = \sum_{i_1,i_2,i_3,i_4(dist)} \Omega_{i_1i_2}\Omega_{i_2i_3}\Omega_{i_3i_4}\Omega_{i_4i_1}, \qquad b_n = 2\|\theta\|^4 \cdot [g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g]. \tag{3.4.21}$$

Here, two vectors $g, h \in \mathbb{R}^K$ are defined as $g_k = (\mathbf{1}_k'\theta)/\|\theta\|_1$ and $h_k = (\mathbf{1}_k'\Theta\mathbf{1}_k)^{1/2}/\|\theta\|$, where $\mathbf{1}_k$ is for short of $\mathbf{1}_k^{(K)}$, which is defined as

$$\mathbf{1}_k^{(K)}(i) = 1 \text{ if } i \in \mathcal{N}_k \text{ and } 0 \text{ otherwise.}$$

Moreover, $V = \mathrm{d}iag(Pg)$, and $H = \mathrm{d}iag(h)$. The following lemmas are proved in the supplementary material.

Lemma 36. *Under the conditions of Theorem 3.2.1, we have* $\mathbb{E}[C_n] = c_n \asymp \|\theta\|^8$ *and* $\mathrm{Var}(C_n) \leq C\|\theta\|^8 \cdot [1 + \|\theta\|_3^6]$, *and so* $C_n/c_n \to 1$ *in probability for* c_n *defined in* (3.4.21).

Lemma 37. *Under the conditions of Theorem 3.2.1,* $\widetilde{Q}_n/\sqrt{8c_n} \to N(0,1)$ *in law.*

Lemma 38. *Under the conditions of Theorem 3.2.1,* $\mathbb{E}(Q_n^{(K,0)} - \widetilde{Q}_n - b_n)^2 = o(\|\theta\|^8)$.

Lemma 39. *Under the conditions of Theorem 3.2.1, we have* $b_n \asymp \|\theta\|^4$ *and* $B_n^{(K)}/b_n \to 1$ *in probability for* b_n *defined in* (3.4.21).

Among these lemmas, the proof of Lemma 38 is the most complicated one, as it requires computing the bias in $Q_n^{(m)}$ caused by the refitting step; see Section 3.C.8 in the supplementary material for details.

We now prove Theorem 3.2.1. Rewrite $\psi_n^{(K,0)}$ as

$$\sqrt{\frac{c_n}{C_n}}\left[\frac{\widetilde{Q}_n}{\sqrt{8c_n}} + \frac{(Q_n^{(K,0)} - \widetilde{Q}_n - b_n)}{\sqrt{8c_n}} + \frac{(b_n - B_n^{(K)})}{\sqrt{8c_n}}\right] = \sqrt{\frac{c_n}{C_n}} \cdot [(I) + (II) + (III)], \tag{3.4.22}$$

where $(I) = \widetilde{Q}_n/\sqrt{8c_n}$, $(II) = (Q_n^{(K,0)} - \widetilde{Q}_n - b_n)/\sqrt{8c_n}$, and $(III) = (b_n - B_n^{(K)})/\sqrt{8c_n}$. Now, first by Lemmas 36-37,

$$c_n/C_n \to 1 \text{ in probability,} \qquad \text{and} \qquad (I) \to N(0,1) \text{ in law.} \tag{3.4.23}$$

Second, by Lemma 37,

$$\mathbb{E}[(II)^2] \leq (8c_n)^{-1} \cdot \mathbb{E}[(Q_n^{(K,0)} - \widetilde{Q}_n^{(K,0)} - b_n)^2] \leq c_n^{-1} \cdot o(\|\theta\|^8), \tag{3.4.24}$$

where the right hand side is $o(1)$ for $c_n \asymp \|\theta\|^8$ by Lemma 36. Last, by Lemma 36- 39, we have $b_n \asymp \sqrt{c_n} \asymp \|\theta\|^4$ and $B_n^{(K)}/b_n \xrightarrow{p} 1$, and so

$$(III) = \left(\frac{b_n}{\sqrt{8c_n}}\right) \cdot \left(\frac{B_n^{(K)}}{b_n} - 1\right) \xrightarrow{p} 0. \tag{3.4.25}$$

Inserting (3.4.23)-(3.4.25) into (3.4.22) gives the claim and concludes the proof of Theorem 3.2.1.

3.4.2 Proof of Theorem 3.2.3 (the under-fitting case of $m < K$)

In the proof of Theorem 3.2.1, we start from replacing $\widehat{\Pi}^{(K)}$ with the true community label matrix Π. However, when $m < K$, $\widehat{\Pi}^{(m)}$ does not concentrate on one particular label matrix. Below, we introduce a collection of label matrices, $\mathcal{G}_m$, consisting of all possible realizations of $\widehat{\Pi}^{(m)}$ when NSP holds. We then study the GoF statistic on the event that $\widehat{\Pi}^{(m)} = \Pi_0$, for a fixed $\Pi_0 \in \mathcal{G}_m$.

Recall that Π is the true community label matrix. Fix $1 \leq m < K$. Let $\mathcal{G}_m$ be the class of $n \times m$ matrices Π_0, where each Π_0 is formed as follows: let $\{1, 2, \ldots, K\} = S_1 \cup S_2 \ldots \cup S_m$ be a partition, column ℓ of Π_0 is the sum of all columns of Π in S_ℓ, $1 \leq \ell \leq m$. Let L_0 be the $K \times m$ matrix of 0 and 1 where

$$L_0(k, \ell) = 1 \text{ if and only if } k \in S_\ell, \qquad 1 \leq k \leq K,\ 1 \leq \ell \leq m. \tag{3.4.26}$$

Therefore, for each $\Pi_0 \in \mathcal{G}_m$, we can find an L_0 such that $\Pi_0 = \Pi L_0$. Note that each Π_0 is the community label matrix where each community implied by it (i.e., "pseudo community") is formed by merging one or more (true) communities of the original network.

Fix a Π_0 and let $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, \cdots, \mathcal{N}_m^{(m,0)}$ be the m "pseudo communities" associated with Π_0. Recall that $\widehat{\theta}^{(m)}$, $\widehat{\Theta}^{(m)}$ and $\widehat{P}^{(m)}$ are refitted quantities obtained by using the adjacency matrix A and $\widehat{\Pi}^{(m)}$; see (3.2.5)-(3.2.6). To misuse the notations a little bit, let $\widehat{\theta}^{(m,0)}$, $\widehat{\Theta}^{(m,0)}$ and $\widehat{P}^{(m,0)}$ be the proxy of $\widehat{\theta}^{(m)}$, $\widehat{\Theta}^{(m)}$ and $\widehat{P}^{(m)}$ respectively, constructed similarly by (3.2.5)-(3.2.6), but with $\widehat{\Pi}^{(m)}$ replaced by Π_0. Introduce

$$\widehat{\Omega}^{(m,0)} = \widehat{\Theta}^{(m,0)} \Pi_0 \widehat{P}^{(m,0)} \Pi_0' \widehat{\Theta}^{(m,0)}, \tag{3.4.27}$$

$$Q_n^{(m,0)} = \sum_{i_1, i_2, i_3, i_4 (dist)} (A_{i_1 i_2} - \widehat{\Omega}_{i_1 i_2}^{(m,0)})(A_{i_2 i_3} - \widehat{\Omega}_{i_2 i_3}^{(m,0)})(A_{i_3 i_4} - \widehat{\Omega}_{i_3 i_4}^{(m,0)})(A_{i_4 i_1} - \widehat{\Omega}_{i_4 i_1}^{(m,0)}),$$

and

$$\psi_n^{(m,0)} = [Q_n^{(m,0)} - B_n^{(m)}]/\sqrt{8C_n}. \tag{3.4.28}$$

These are the proxies of $\Omega^{(m)}$, $Q_n^{(m)}$, and $\psi_n^{(m)}$, respectively, where $\widehat{\Pi}^{(m)}$ is now frozen at a non-stochastic matrix Π_0.

In the under-fitting case, $m < K$, and we do not expect $\widehat{\Omega}^{(m,0)}$ to be close to Ω. We define a non-stochastic counterpart of $\widehat{\Omega}^{(m,0)}$ as follows. Let $\theta^{(m,0)}$, $\Theta^{(m,0)}$ and $P^{(m,0)}$ be constructed similarly by (3.2.5)-(3.2.6), except that $(A, \widehat{\Pi}^{(m)})$ and the vector $d = (d_1, d_2, \ldots, d_n)'$ are replaced with (Ω, Π_0) and $\Omega\mathbf{1}_n$, respectively. Let

$$\Omega^{(m,0)} = \Theta^{(m,0)}\Pi_0 P^{(m,0)}\Pi_0'\Theta^{(m,0)}. \tag{3.4.29}$$

The following lemma gives an equivalent expression of $\Omega^{(m,0)}$ and is proved in the supplementary material.

Lemma 40. *Fix $K > 1$ and $1 \leq m \leq K$. Let $\Pi_0 = \Pi L_0 \in \mathcal{G}_m$ and $\Omega^{(m,0)}$ be as above. Write $D = \Pi'\Theta\Pi \in \mathbb{R}^{K,K}$ and $D_0 = \Pi_0'\Theta\Pi \in \mathbb{R}^{m,K}$. Let P_0 be the $K \times K$ matrix given by*

$$P_0 = diag(PD\mathbf{1}_K) \cdot L_0 \cdot diag(D_0PD\mathbf{1}_K)^{-1}(D_0PD_0')diag(D_0PD\mathbf{1}_K)^{-1} \cdot L_0' \cdot diag(PD\mathbf{1}_K),$$

where the rank of P_0 is m. Then, $\Omega^{(m,0)} = \Theta\Pi P_0\Pi'\Theta$.

This lemmas says that $\Omega^{(m,0)}$ has a similar expression as Ω, with P replaced by a rank-m matrix P_0. When $m = K$, $\mathcal{G}_m$ has only one element Π; then $(P_0, \Omega^{(m,0)})$ reduces to (P, Ω).

We expect $\widehat{\Omega}^{(m,0)}$ to concentrate at $\Omega^{(m,0)}$. This motivates the following proxy of $Q_n^{(m,0)}$.

$$\widetilde{Q}_n^{(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} (A_{i_1i_2} - \Omega^{(m,0)}_{i_1i_2})(A_{i_2i_3} - \Omega^{(m,0)}_{i_2i_3})(A_{i_3i_4} - \Omega^{(m,0)}_{i_3i_4})(A_{i_4i_1} - \Omega^{(m,0)}_{i_4i_1}). \tag{3.4.30}$$

Introduce

$$\widetilde{\Omega}^{(m,0)} = \Omega - \Omega^{(m,0)}. \tag{3.4.31}$$

Recall that $A = (\Omega - \mathrm{d}iag(\Omega)) + W$, we rewrite $\widetilde{Q}_n^{(m,0)}$ as

$$\widetilde{Q}_n^{(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} (W_{i_1i_2} + \widetilde{\Omega}^{(m,0)}_{i_1i_2})(W_{i_2i_3} + \widetilde{\Omega}^{(m,0)}_{i_2i_3})(W_{i_3i_4} + \widetilde{\Omega}^{(m,0)}_{i_3i_4})(W_{i_4i_1} + \widetilde{\Omega}^{(m,0)}_{i_4i_1}). \tag{3.4.32}$$

Note that when $m = K$ and $\Pi_0 = \Pi$, the statistic $\widetilde{Q}_n^{(m,0)}$ reduces to $\widetilde{Q}_n$ defined in (3.4.20).

The matrix $\widetilde{\Omega}^{(m,0)}$ captures the signal strength in $\widetilde{Q}_n^{(m,0)}$. From now on, for notation simplicity, we write $\widetilde{\Omega}^{(m,0)} = \widetilde{\Omega}$ in the rest of the proof. Let $\widetilde{\lambda}_k$ be the k-th largest (in magnitude) eigenvalue of $\widetilde{\Omega}$ and recall that λ_k is the k-th largest (in magnitude) eigenvalue of Ω. In light of (3.4.31), we write $\Omega = \Omega^{(m,0)} + \widetilde{\Omega}$ and apply Weyl's theorem for singular values (see equation (7.3.13) of Horn and Johnson (1985)). Note that $\Omega^{(m,0)}$ has a rank m and Ω has a rank K. By Weyl's theorem, for all $1 \leq k \leq K - m$, $|\lambda_{m+k}| \leq |\lambda_{m+1}(\Omega^{(m,0)})| + |\widetilde{\lambda}_k| = |\widetilde{\lambda}_k|$. It follows that

$$\mathrm{tr}(\widetilde{\Omega}^4) \geq \sum_{k=1}^{K-m} |\widetilde{\lambda}_k|^4 \geq \sum_{k=m+1}^{K} |\lambda_k|^4.$$

As we will see in Lemma 42 below, $\mathrm{tr}(\widetilde{\Omega}^4)$ is the dominating term of $\mathbb{E}[\widetilde{Q}_n^{(m,0)}]$. Define

$$\tau^{(m,0)} = |\widetilde{\lambda}_1|/\lambda_1. \tag{3.4.33}$$

For notation simplicity, we write $\tau^{(m,0)} = \tau$, but keep in mind both $\widetilde{\Omega}$ and τ actually depend on m and $\Pi_0 \in \mathcal{G}_m$. The following lemmas are proved in the supplementary material.

Lemma 41. *Under the conditions of Theorem 3.2.3, for each $1 \leq m \leq K$, let $\widetilde{\Omega}$ and τ be defined as in* (3.4.29) *and* (3.4.33). *The following statements are true:*

- *There exists a constant $C > 0$ such that $|\widetilde{\Omega}_{ij}| \leq C\tau\theta_i\theta_j$, for all $1 \leq i, j \leq n$.*
- *$c_n \asymp \|\theta\|^8$, $\lambda_1 \asymp \|\theta\|^2$, and $\tau = O(1)$.*
- *$\mathrm{tr}(\widetilde{\Omega}^4) \geq C\tau^4\|\theta\|^8$, and $\tau\|\theta\| \to \infty$.*

Lemma 42. *Under the condition of Theorem 3.2.3, for $1 \leq m < K$,*

$$\mathbb{E}[\widetilde{Q}_n^{(m,0)}] = \mathrm{tr}(\widetilde{\Omega}^4) + o(\|\theta\|^4), \quad \mathrm{Var}(\widetilde{Q}_n^{(m,0)}) \leq C(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6).$$

Lemma 43. *Under the condition of Theorem 3.2.3, for $1 \leq m < K$,*

$$\mathbb{E}[Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}] = o(\tau^4\|\theta\|^8), \quad \mathrm{Var}(Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}) \leq o(\|\theta\|^8) + C\tau^6\|\theta\|^8\|\theta\|_3^6.$$

Lemma 44. *Under the conditions of Theorem 3.2.3, for $1 \leq m < K$, there exists a constant $C > 0$, such that $\mathbb{P}(B_n^{(m)} \leq C\|\theta\|^4) \geq 1 + o(1)$.*

We now prove Theorem 3.2.3. Note that by Theorem 3.2.1, the second item of Theorem 3.2.3 follows once the first item is proved. Therefore we only consider the first item, where it is sufficient to show that for all $1 < m < K$,

$$\psi_n^{(m)} \to \infty, \qquad \text{in probability.}$$

By the NSP of the solutions produced by SCORE, which is shown in Theorem 3.2.2, there exists an event A_n with $\mathbb{P}(A_n^c) \leq Cn^{-3}$ as $n \to \infty$, such that on event A_n we have $\widehat{\Pi}^{(m)} \in \mathcal{G}_m$. This further indicates that on event A_n we have

$$\psi_n^{(m)} \geq \min_{\Pi_0 \in \mathcal{G}_m} \psi_n^{(m,0)}, \tag{3.4.34}$$

where $\psi_n^{(m,0)}$ is defined in (3.4.28). The LHS is hard to analyze, but the RHS is relatively easy to analyze. Then further notice that the cardinality of $\mathcal{G}_m$ is $|\mathcal{G}_m| = m^K$, which is of constant order as long as K is constant. Therefore to prove $\psi_n^{(m)} \to \infty$ in probability, it suffices to show that for any fixed $\Pi_0 \in \mathcal{G}_m$,

$$\psi_n^{(m,0)} \to \infty, \qquad \text{in probability.} \tag{3.4.35}$$

We now show (3.4.35). Rewrite $\psi_n^{(m,0)}$ as

$$\sqrt{\frac{c_n}{C_n}} \cdot \left[\frac{Q_n^{(m,0)}}{\sqrt{8c_n}} - \frac{B_n^{(m)}}{\sqrt{8c_n}}\right] = \sqrt{\frac{c_n}{C_n}} \cdot [(I) - (II)], \tag{3.4.36}$$

where $(I) = Q_n^{(m,0)}/\sqrt{8c_n}$, and $(II) = B_n^{(m)}/\sqrt{8c_n}$. First, by Lemma 36 (since C_n and c_n do not depend on m, this lemma applies to both the null case and the under-fitting case),

$$c_n/C_n \to 1 \qquad \text{in probability.} \tag{3.4.37}$$

Second, by Lemma 41, $c_n \asymp \|\theta\|^8$. Combining it with Lemma 44 gives that there is a constant $C > 0$ such that

$$\mathbb{P}((II) \leq C) \geq 1 + o(1). \tag{3.4.38}$$

Last, by Lemma 41-43,

$$\mathbb{E}[(I)] \geq C\tau^4\|\theta\|^4 \cdot [1 + o(1)] \to \infty, \quad \mathrm{Var}((I)) \leq C(1 + \tau^6\|\theta\|_3^6).$$

Therefore, by Chebyshev's inequality, for any constant $M > 0$,

$$\mathbb{P}((I) < M) \leq (\mathbb{E}[(I)] - M)^{-2}\mathrm{Var}((I)) \leq C\left[\frac{1 + \tau^6\|\theta\|_3^6}{(\tau^4\|\theta\|^4[1 + o(1)] - M)^2}\right], \tag{3.4.39}$$

where on the denominator, $\tau\|\theta\| \to \infty$ by Lemma 41. Note that under our conditions, $\|\theta\|_3^3 = o(\|\theta\|^2)$ and $\|\theta\| \to \infty$. Combining these, the RHS of (3.4.39) tends to 0 as $n \to \infty$. Inserting (3.4.37)-(3.4.39) into (3.4.36) proves the claim, and concludes the proof of Theorem 3.2.3.

3.5 Real data analysis and simulation study

In theory, a good approximation for the null distribution of $\psi_n^{(m)}$ is $N(0,1)$ and Theorem 3.2.1, where we show $\psi_n^{(m)} \to N(0,1)$ in the null case). Such a result requires some model assumptions, which may be violated in real applications (e.g., outliers, artifacts). When this happens, a good approximation for the null distribution of $\psi_n^{(m)}$ is no longer $N(0,1)$ (i.e., theoretical null), but $N(u,\sigma^2)$ (i.e., empirical null) for some $(u,\sigma) \neq (0,1)$. Such a phenomenon has been repeatedly noted in the literature. For example, Efron Efron (2004) argued that due to artifacts or model misspecification, the *empirical null* frequently works better for real data than the *theoretical null*. The problem is then how to estimate the parameters (u,σ^2) of the empirical null.

We propose a bootstrap approach to estimating (u,σ^2). Recall that $\widehat{\lambda}_k$ is the k-th largest eigenvalue of A and $\widehat{\xi}_k$ is the corresponding eigenvector. Fixing $N > 1$ and $m > 1$, letting $\widehat{M}^{(m)} = \sum_{k=1}^m \widehat{\lambda}_k \widehat{\xi}_k \widehat{\xi}_k'$ and let $\widehat{S}^{(m)} = A - \widehat{M}^{(m)}$. For $b = 1,2,\ldots,N$, we simultaneously permute the rows and columns of $\widehat{S}^{(m)}$ and denote the resultant matrix by $\widehat{S}^{(m,b)}$. Truncating all entries of $(\widehat{M}^{(m)} + \widehat{S}^{(m,b)})$ at 1 at the top and 0 at the bottom, and denote the resultant matrix by $\widehat{\Omega}^{(b)}$. Generate an adjacency matrix $A^{(b)}$ such that for all $1 \leq i < j \leq n$, $A_{ij}^{(b)}$ are independent Bernoulli samples with parameters $\widehat{\Omega}_{ij}^{(b)}$ (we may need to repeat this step until the network is connected). Apply StGoF to $A^{(b)}$ and denote the resultant statistic by $Q_n^{(b)}$. We estimate u and σ by the empirical mean and standard deviation of $\{Q_n^{(b)}\}_{b=1}^N$, respectively. Denote the estimates by $\widehat{u}^{(m)}$ and $\widehat{\sigma}^{(m)}$, respectively. The bootstrap StGoF statistic is then

$$\psi_n^{(m,*)} = [Q_n^{(m)} - \widehat{u}^{(m)}]/\widehat{\sigma}^{(m)}, \qquad m = 1,2,\ldots, \tag{3.5.40}$$

where $Q_n^{(m)}$ is the same as in (3.2.12). Similarly, we estimate K as the smallest integer m such that $\psi_n^{(m,*)} \leq z_\alpha$, for the same z_α in StGoF. We recommend $N = 25$, as it usually gives stable estimates for $\widehat{u}^{(m)}$ and $\widehat{\sigma}^{(m)}$.

The original StGoF works well for real data where the DCBM is reasonable, but for data sets where DCBM is significantly violated, bootstrap StGoF may help. For the 6 data sets considered in Section 3.5, two methods perform similarly for all but one data set. This particular data set is suspected to have many outliers, and bootstrap StGoF performs significantly better. For theoretical analysis, we focus on the original StGoF statistics $\psi_n^{(m)}$ as in (3.2.12).

3.5.1 Real data analysis

For real data analysis, we consider 6 different data sets as in Table 3.1, which can be downloaded from `http://www-personal.umich.edu/~mejn/netdata/`. We now discuss the true K. For the dolphin network, it was argued in Liu et al. (2016) that both $K = 2$ or

$K = 4$ are reasonable. For UKfaculty network, we symmetrize the network by ignoring the directions of the edges. There are 4 school affiliations for the faculty members so we take $K = 4$. For the football network, we take $K = 11$. The network was manually labelled as 12 groups, but the 12th group only consist of the 5 "independent" teams that do not belong to any conference and do not form a conference themselves. For polbooks network, Le and Levina Le and Levina (2015) suggest that $K = 3$, but it was argued by Jin et al. (2017) that a more appropriate model for the network is a degree corrected mixed-membership (DCMM) model with two communities, so $K = 2$ is also appropriate.

We compare StGoF and bootstrap StGoF (StGoF*) with the BIC approach by Wang and Bickel Wang et al. (2017), BH approach by Le and Levina Le and Levina (2015), ECV approach by Li *et al.* Li et al. (2020), and NCV approach by Chen and Lei Chen and Lei (2018). For all these methods, we use the R package "randnet" to implement them. Note that among these approaches, ECV and NCV are cross validation (CV) approaches and the results vary from one repetition to the other. Therefore, we run each method for 25 times and report the mean and SD. The StGoF* uses bootstrapping mean and standard deviation and is also random, but the SDs are 0 for five data sets. Most methods require a feasible range of K as a priori. We take $\{1, 2, ..., 15\}$ as the range in this section.

Table 3.1: Comparison of estimated K. Take ECV for Dolphins for example: for 25 independent repetitions, the estimated K have a mean of 3.08 and a SD of 0.91, ranging from 2 to 5 (SD of StGoF* are 0 for the first 5 data sets).

Name	n	K	BIC	BH	ECV	NCV	StGoF	StGoF*
Dolphins	62	2, 4	2	2	3.08(0.91) [2,5]	2.20(2.71) [1,15]	2	3
Football	115	11	10	10	11.28(0.61) [11,13]	12.36(1.15) [11, 15]	10	10
Karate	34	2	2	2	2.60(1.00) [1,6]	2.56(0.58) [2,4]	2	2
UKfaculty	81	4	4	3	5.56(1.61) [3,11]	2.40(0.28) [2,3]	4	4
Polblogs	1222	2	6	8	4.88(1.13) [4, 8]	2(0.00) [2, 2]	2*	2
Polbooks	105	2, 3	3	4	7.56(2.66) [2,15]	2.08(0.71) [2, 5]	5	2.4(0.25) [2, 3]

The polblogs network is suspected to have outliers, so most of the methods do not work well. For this particular network, the mean of StGoF is much larger than expected, so we choose to estimate K by the m that minimizes $\psi_n^{(m)}$ for $1 \leq m \leq 15$ (for this reason, we put a $*$ next to 2 in the table). Note that StGoF* correctly estimates K as 2. The polbooks network is suspected have a signifiant faction of mixing nodes (e.g., Jin et al. (2017)), which explains why StGoF overestimates K. Fortunately, for both data sets, StGoF* estimates K correctly, suggesting the bootstrapping means and standard deviations help standardize $Q_n^{(m)}$.

3.5.2 Simulations

We now study StGoF with simulated data. We compare StGoF with BIC, ECV, NCV via a small scale simulations (for StGoF, $\alpha = 0.05$). We do not include StGoF* since there is no model specification. We do not include BH for comparison either: the method is designed for very sparse stochastic block model and the performance is unsatisfactory for most of our settings.

Given (n, K), a scalar $\beta_n > 0$ that controls the sparsity, a symmetric non-negative matrix $P \in \mathbb{R}^{K\times K}$, a distribution $f(\theta)$ on $(0, \infty)$, and a distribution $g(\pi)$ on the standard simplex of $\mathbb{R}^K$, we generate the adjacency matrix $A \in \mathbb{R}^{n,n}$ as follows:

1. Generate $\widetilde{\theta}_1, \widetilde{\theta}_2, ..., \widetilde{\theta}_n$ *iid* from $f(\theta)$. Let $\theta_i = \beta_n \cdot \widetilde{\theta}_i / \|\widetilde{\theta}\|$ and $\Theta = \mathrm{d}iag(\theta_1, ..., \theta_n)$.
2. Generate $\pi_1, \pi_2, ..., \pi_n$ iid from $g(\pi)$, and let $\Pi = [\pi_1, \pi_2, ..., \pi_n]'$.
3. Let $\Omega = \Theta\Pi P\Pi'\Theta$. For each experiment below, once Ω is generated, we keep it fixed, and use it to generate A according to the DCBM, for 100 times independently.

For all algorithms, we measure the performance by the fraction of times the algorithm correctly estimates the true number of communities K (i.e., accuracy). Note that $\|\theta\| = \beta_n$, and SNR $\asymp \|\theta\|(1 - b_n)$. For the experiments, we let β_n range so to cover many different sparsity levels, but let $\|\theta\|(1 - b_n)$ be at a more or less the same level, so the problem of estimating K is not too difficult or too easy; see details below. We consider three experiments, and each experiment has some sub-experiments.

Experiment 1. In this experiment, we study how degree heterogeneity affect the results and comparisons. Fixing $(n, K) = (600, 4)$, we let P be the 4×4 matrix with unit diagonals and off-diagonals $P(k, \ell) = 1 - [(1 - b_n)(|k - \ell| + 1)]/K$, where $1 \leq k, \ell \leq 4$ and $k \neq \ell$. Such matrix is called a Toeplitz matrix. Let $g(\pi)$ be the uniform distribution over e_1, e_2, e_3, e_4 (the standard basis vectors of $\mathbb{R}^4$).

We consider three sub-experiments, Exp 1a-1c. In these sub-experiments, we keep $(1 - b_n)\|\theta\|$ fixed at 9.5 so the SNR's are roughly at the same level. We let β_n range from 10 to 14 so to cover both the more sparse and the more dense cases. Moreover, for the three sub-experiments, we take $f(\theta)$ to be $U(2, 3)$ (uniform distribution), Pareto(8, .375) (8 is the shape parameter and .375 is the scale parameter), and two point mixture $0.95\delta_1 + 0.05\delta_3$ (δ_a is a point mass at a), respectively. Note that from Exp 1a to Exp 1c, the degree heterogeneity is increasingly more severe on average.

The estimation accuracy is presented in Figure 3.4, where StGoF is seen to consistently outperform other approaches. Also, from Exp 1a to Exp1c, the estimation accuracy for all algorithms get consistently lower, suggesting that when the degree heterogeneity gets more severe, the problem of estimating K gets more challenging.

Experiment 2. In this experiment, we study how the relative sizes of different communities affect the results and comparisons. For $b_n > 0$ to be determined, we set $(n, K) = (1200, 3)$, $f(\theta)$ as Pareto(10, 0.375), and let P be the 3×3 matrix satisfying $P(k, \ell) = 1-|k-\ell|(1-b_n)/2$, $1 \leq k, \ell \leq 3$. We let β_n range in $\{12, 13, ..., 17\}$ and keep $(1 - b_n)\|\theta\|$ fixed at 10 so the SNR's are roughly at the same level. We take $g(\pi)$ as the distribution with weights a, b, and $(1 - a - b)$ on vectors e_1, e_2, e_3 (the standard basis vectors of $\mathbb{R}^3$), respectively. Consider three sub-experiments, Exp 2a-2c, where we take $(a, b) = (.30, .35), (.25, .375)$, and $(.20, .40)$,

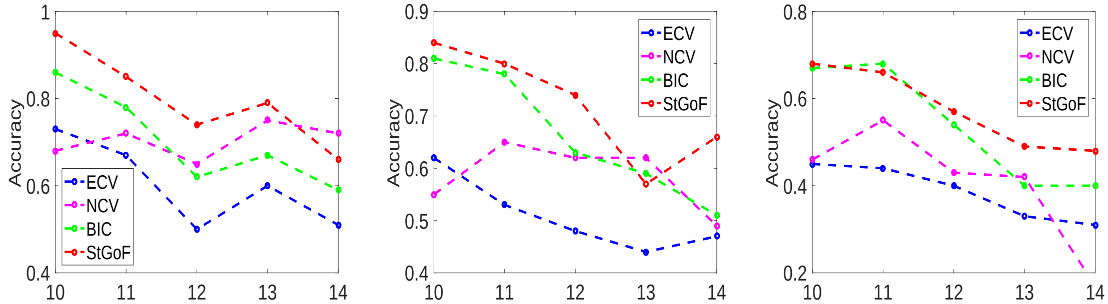

Figure 3.4: Left to right: Experiment 1a, 1b, and 1c, where the degree heterogeneity are increasingly more severe (x-axis: sparsity. y-axis: accuracy). Results are based on 100 repetitions.

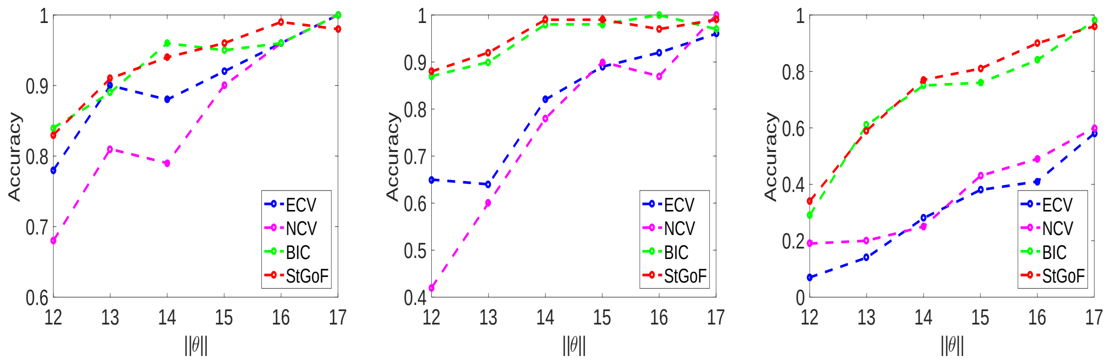

Figure 3.5: Left to right: Experiment 2a, 2b, and 2c (x-axis: $\|\theta\|$ (sparsity level); y-axis: estimation accuracy. The results are based on 100 repetitions.

respectively, so the three communities in the network are slightly unbalanced, moderately unbalanced, and slightly unbalanced, respectively.

Figure 3.5 presents the estimation accuracy. First, StGoF consistently outperforms NCV, ECV and BIC. Second, when the three communities get increasingly unbalanced, all methods become less accurate, suggesting that estimating K gets increasingly harder. Last, the performance of ECV and NCV are relatively close to that of StGOF when the communities are relatively balanced (e.g., Exp 2a), but are comparably more unsatisfactorily when the models are more unbalanced (e.g., Exp 2b-2c).

Experiment 3. We study how robust these algorithms are in cases of model misspecification. Fix $(n, K) = (600, 4)$. We let $f(\theta)$ be the uniform distribution $U(2, 3)$, and let P be the 4×4 matrix with unit diagonals and where for $1 \leq k, \ell \leq 4$ and $k \neq \ell$, $P(k, \ell) = 1 - (1 - b_n)(|k - \ell| + 1)/K$. We consider two sub-experiments, Exp 3a-3b. For sparsity, we let β_n range from 11 to 16 in Exp 3a and range from 11 to 18 in Exp 3b. For different β_n, we choose b_n so that $(1 - b_n)\|\theta\|$ is fixed at 10.5. Moreover, in Exp 3a, we allow mixed-memberships. We take $g(\pi)$ to be the mixing distribution which puts probability .2 on e_1, e_2, e_3, e_4 (standard basis vectors of $\mathbb{R}^4$), respectively, and let π be the symmetric K-dimensional Dirichlet distribution for the remaining probability of .2. Once we have θ_i, π_i,

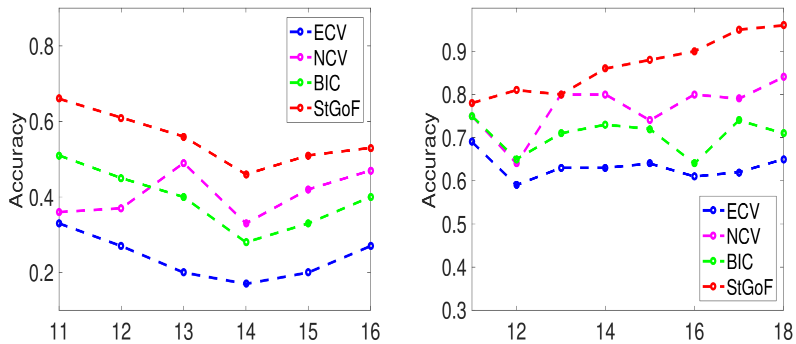

Figure 3.6: Experiment 3a (left) and 3b (right) (x-axis: $\|\theta\|$ (sparsity level). y-axis: estimation accuracy). The results are based on 100 repetitions.

and P, we let $\Omega_{ij} = \theta_i \theta_j \pi_i' P \pi_j$, $1 \leq i, j \leq n$, similar to that in DCBM. In Exp 3b, we allow outliers. First, we let $g(\pi)$ be the mixing distribution that puts masses .25 on e_1, e_2, e_3, e_4, and obtain Ω as in DCBM. We then randomly select 10% of the nodes and re-define Ω_{ij} as ρ_n if either i or j is selected, where $\rho_n = n^{-2} \sum_{1 \leq i,j \leq n} \Omega_{ij}$.

Figure 3.6 presents the estimation accuracy. The two cross-validation methods (ECV and NCV) are not model based algorithms and are expected to be less affected by model misspecification, so we can use their results as a benchmark to evaluate the performances of StGoF and the likelihood-based approach BIC. Figure 3.6 shows that StGoF continues to perform well in all settings, suggesting that it is not sensitive to model misspecification. The performance of BIC, if compared to those in Experiments 1-2, is less satisfactory, suggesting that the method is more sensitive to the model misspecification.

3.A Proof of results in Sections 3.1-3.2

3.A.1 Proof of Lemma 27

For the goodness-of-fit test, it contains calculation of (a) $\widehat{\Omega}^{(m)}$ as the refitted Ω, (b) $Q_n^{(m)}$ as the main term, (c) $B_n^{(m)}$ as the bias correction term and (d) C_n as the variance estimator.

For (a), it requires calculation of d_i for $1 \leq i \leq n$, and $\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_\ell$ and $\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_n$ for $1 \leq k, \ell \leq m$ with $m \leq K$. Since d_i needs $O(d_i)$ operations, it takes $O(n\bar{d})$ for calculating d_i, $1 \leq i \leq n$. Similarly, it takes $O(\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_\ell)$ to calculate $\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_\ell$ and $O(\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_n)$ to calculate $\widehat{\mathbf{1}}_k' A \widehat{\mathbf{1}}_n$, $1 \leq k, \ell \leq m$. The total complexity is then $O(n\bar{d})$. By (3.2.8),

$$\widehat{\Omega}^{(m)}(i,j) = \widehat{\theta}^{(m)}(i)\widehat{\theta}^{(m)}(j)(\widehat{\pi}_i^{(m)})' \widehat{P}^{(m)} \widehat{\pi}_j^{(m)},$$

whose calculation takes $O(m^2)$ operations. Hence, calculation of $\widehat{\Omega}^{(m)}$ needs $O(m^2 n^2)$ operations. Combining together, we conclude that step (a) costs $O(m^2 n^2)$.

For (b), $Q_n^{(m)}$ can be calculated using the same form in Theorem 1.1 of Jin et al. (2019). As is shown there, this step requires $O(n^2 \bar{d})$ operations.

For (c), given $\widehat{\Omega}^{(m)}$ and $\widehat{P}^{(m)}$, the calculation of $\widehat{g}^{(m)}$, $\widehat{V}^{(m)}$ and $\widehat{H}^{(m)}$ only takes $O(n)$. By (3.2.11), calculation of $B_n^{(m)}$ only involves calculate $\|\widehat{\theta}\|$ and $\widehat{g}' \widehat{V}^{-1} (\widehat{P}\widehat{H}^2\widehat{P} \circ \widehat{P}\widehat{H}^2\widehat{P}) \widehat{V}^{-1} \widehat{g}$. The first part needs $O(n)$ operations. The second part only involves vectors in $\mathbb{R}^m$ and matrices in $\mathbb{R}^{m,m}$. Moreover since $m \leq K$ and K is fixed, it takes at most $o(n)$ operations. Combining above, step (c) costs $O(n)$.

For (d), the calculation follows from Proposition A.1 of Jin et al. (2018). It should be noted C_n is denoted as $\widehat{C}_4$ there, and it requires calculation of (i) trace of a matrix, (ii) A^4 for matrix A and (iii) quadratic form of matrix A and A^2. For (i), it only takes $O(n)$. For (iii), it takes at most $O(n^2)$. For (ii), we can compute A^k recursively from $A^k = A^{k-1}A$. it suffices to consider the complexity of computing BA, for an arbitrary $n \times n$ matrix B. The (i,j)-th entry of BA is $\sum_{\ell: A_{\ell j} \neq 0} B_{i\ell} A_{\ell j}$, where the total number of nonzero $A_{\ell j}$ equals to d_j, the degree of node j. Hence, the complexity of computing the (i,j)-th entry of BA is $O(d_j)$. It follows that the complexity of computing BA is $O(n^2 \bar{d})$.

Combining above, the goodness-of-fit test needs $O(n^2 \bar{d})$ operations. □

3.A.2 Proof of Theorem 3.2.4

First, we show the claims on $|\lambda_K|/\sqrt{\lambda_1}$. Define a diagonal matrix H by $H_{kk} = \|\theta\|^{-1} \sqrt{\sum_{i: \ell_i = k} \theta_i^2}$, for $1 \leq k \leq K$. Note that H is also stochastic. By Lemma 30, the eigenvalues of Ω are equal to the eigenvalues of $\|\theta\|^2 HPH$, i.e.,

$$\lambda_k = \|\theta\|^2 \cdot \lambda_k(HPH), \qquad 1 \leq k \leq K.$$

It follows that

$$|\lambda_K|/\sqrt{\lambda_1} = \|\theta\| \cdot |\lambda_K(HPH)|/\sqrt{\lambda_1(HPH)}. \tag{3.A.41}$$

Below, we first study the matrix H and then show the claims.

Consider the matrix H. Let $\widetilde{N}_1, \widetilde{N}_2, \ldots, \widetilde{N}_{K_0}$ be the (non-stochastic) communities of the DCBM with K_0 communities. For each $1 \leq k \leq K_0$, let $\theta^{(k)} \in \mathbb{R}^n$ be such that $\theta_i^{(k)} = \theta_i \cdot 1\{i \in \widetilde{N}_k\}$. By definition,

$$H_{kk}^2 = \|\theta\|^{-2} \begin{cases} \|\theta^{(k)}\|^2, & \text{for } 1 \leq k \leq K_0 - 1, \\ \sum_{i \in \widetilde{N}_{K_0}} \theta_i^2 \cdot 1\{\ell_i = k\}, & \text{for } K_0 \leq k \leq K_0 + m. \end{cases}$$

Since (3.2.2) is satisfied, $\|\theta\|^2 \geq \|\theta^{(k)}\|^2 \geq C\|\theta\|^2$, for $1 \leq k \leq K_0$. It implies that

$$C^{-1} \leq H_{kk} \leq C, \qquad \text{for } 1 \leq k \leq K_0 - 1. \tag{3.A.42}$$

Fix $k \geq K_0$. The n indicators $1\{\ell_i = k\}$ are *iid* Bernoulli variables with a success probability of $\frac{1}{m+1}$. Therefore, $\mathbb{E}H_{kk}^2 = \frac{1}{m+1}\|\theta\|^{-2}\|\theta^{(K_0)}\|^2$. Furthermore, by Hoeffding's inequality,

$$\mathbb{P}\Big(\big|\|\theta\|^2(H_{kk}^2 - \mathbb{E}H_{kk}^2)\big| > t\Big) \leq 2\exp\Big(-\frac{t^2}{2\sum_{i \in \widetilde{N}_{K_0}} \theta_i^4}\Big).$$

By (3.2.1), $\theta_{\max}\sqrt{\log(n)} \to 0$. Hence, $\sum_{i \in \widetilde{N}_{K_0}} \theta_i^4 \leq \theta_{\max}^2\|\theta^{(K_0)}\|^2 \ll \|\theta\|^2/\log(n)$. Taking $t = \|\theta\|$ in the above equation yields $\big|H_{kk}^2 - \mathbb{E}H_{kk}^2\big| \leq \|\theta\|^{-1}$ with probability $1 - o(n^{-1})$. We have seen that $\mathbb{E}H_{kk}^2 = \frac{1}{m+1}\|\theta\|^{-2}\|\theta^{(K_0)}\|^2$, which is bounded above and below by constants. Additionally, $\|\theta\|^{-1} = o(1)$. Combining these results gives

$$C^{-1} \leq H_{kk} \leq C, \qquad \text{with probability } 1 - o(n^{-1}), \text{ for any } k \geq K_0. \tag{3.A.43}$$

It follows from (3.A.42) and (3.A.43) that

$$\|H\| \leq C, \qquad \|H^{-1}\| \leq C, \qquad \text{with probability } 1 - o(n^{-1}). \tag{3.A.44}$$

Consider the the upper bound for $|\lambda_K|/\sqrt{\lambda_1}$. It suffices to get an upper bound for $|\lambda_K(HPH)|$ and a lower bound for $\lambda_1(HPH)$. Note that $|\lambda_K(HPH)|$ is the smallest singular value of HPH, which can be different from the absolute value of the smallest eigenvalue. Therefore, we cannot use Cauchy's interlacing theorem (Horn and Johnson, 1985) to relate $|\lambda_K(HPH)|$ to the smallest eigenvalue of M. We need a slightly longer proof. Write

$$P = \begin{bmatrix} S & \beta\mathbf{1}_{m+1}' \\ \mathbf{1}_{m+1}\beta' & \mathbf{1}_{m+1}\mathbf{1}_{m+1}' \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{(K_0-1)\times(K_0-1)} & \mathbf{0}_{(K_0-1)\times 1} \\ \mathbf{0}_{1\times(K_0-1)} & \frac{m+1}{1+mb_n}M - \mathbf{1}_{m+1}\mathbf{1}_{m+1}' \end{bmatrix} \equiv P^* + \Delta.$$

The matrix P^* can be re-expressed as (e_{K_0} is the K_0th standard basis of $\mathbb{R}^{K_0}$)

$$P^* = \begin{bmatrix} I_{K_0} \\ \mathbf{1}_m e_{K_0}' \end{bmatrix} \begin{bmatrix} S & \beta \\ \beta' & 1 \end{bmatrix} \begin{bmatrix} I_{K_0} & e_{K_0}\mathbf{1}_m' \end{bmatrix}.$$

Therefore, the rank of P^* is only K_0. Then, HP^*H is also a rank-K_0 matrix. Consequently, for $K = K_0 + m$,

$$\lambda_K(HP^*H) = 0.$$

By Weyl's inequality (Horn and Johnson, 1985), $|\lambda_K(HPH) - \lambda_K(HP^*H)| \leq \|H\Delta H\|$. Combining these results gives

$$|\lambda_K(HPH)| \leq \|H\Delta H\|. \tag{3.A.45}$$

Note that $\|\Delta\| = \|\frac{m+1}{1+mb_n}M - \mathbf{1}_{m+1}\mathbf{1}_{m+1}'\|$. M is a matrix whose diagonals are 1 and off-diagonals are equal to b_n. As a result, Δ is a matrix whose diagonals are equal to $\frac{m(1-b_n)}{1+mb_n}$

and off-diagonals are equal to $\frac{-(1-b_n)}{1+mb_n}$. It follows immediately that

$$\|\Delta\| \le C(1-b_n).$$

We plug it into (3.A.45) and apply (3.A.44). It yields that

$$|\lambda_K(HPH)| \le C(1-b_n). \tag{3.A.46}$$

Furthermore, $\lambda_1(P) \ge P_{11} = 1$ and $\lambda_1(P) \le \|H^{-1}\|^2\lambda_1(HPH)$. Combining it with (3.A.44) gives

$$\lambda_1(HPH) \ge C^{-1}. \tag{3.A.47}$$

Note that (3.A.46)-(3.A.47) hold with probability $1-o(n^{-1})$, because their derivation utilizes (3.A.44). We plug (3.A.46)-(3.A.47) into (3.A.41) to get $|\lambda_K|/\sqrt{\lambda_1} \le C\|\theta\|(1-b_n)$, with probability $1-o(n^{-1})$. This proves the upper bound of $|\lambda_K|/\sqrt{\lambda_1}$.

Consider the the lower bound for $|\lambda_K|/\sqrt{\lambda_1}$. Using (3.A.44), we have

$$|\lambda_K(HPH)|^{-1} = \|(HPH)^{-1}\| \le \|H^{-1}\|^2 \cdot \|P^{-1}\| \le C\|P^{-1}\|. \tag{3.A.48}$$

We then bound $\|P^{-1}\|$. Write

$$P = A + B, \qquad \text{where} \quad A = \begin{bmatrix} S & \\ & \frac{m+1}{1+mb_n}M \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \mathbf{0} & \beta\mathbf{1}'_{m+1} \\ \mathbf{1}_{m+1}\beta' & \mathbf{0} \end{bmatrix}.$$

The matrix B is a rank-2 matrix, which can be re-expressed as

$$B = XD^{-1}X', \qquad \text{where} \quad X = \begin{bmatrix} \beta & \beta \\ \mathbf{1}_{m+1} & -\mathbf{1}_{m+1} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & \\ & -2 \end{bmatrix}.$$

We use the matrix inversion formula to get

$$\begin{aligned} \|P^{-1}\| &= \|(A + XD^{-1}X')^{-1}\| \\ &= \|A^{-1} - A^{-1}X(D + X'A^{-1}X)^{-1}X'A^{-1}\| \\ &\le \|A^{-1}\| \cdot \big(1 + \|X(D + X'A^{-1}X)^{-1}X'A^{-1}\|\big) \\ &= \|A^{-1}\| \cdot \big(1 + \|(D + X'A^{-1}X)^{-1}(X'A^{-1}X)\|\big). \end{aligned} \tag{3.A.49}$$

By direct calculations, writing $M_0 = \frac{1+mb_n}{m+1}M$ and $\mathbf{1} = \mathbf{1}_{m+1}$ for short, we have

$$X'A^{-1}X = \begin{bmatrix} \beta'S^{-1}\beta + \mathbf{1}'M_0^{-1}\mathbf{1} & \beta'S^{-1}\beta - \mathbf{1}'M_0^{-1}\mathbf{1} \\ \beta'S^{-1}\beta - \mathbf{1}'M_0^{-1}\mathbf{1} & \beta'S^{-1}\beta + \mathbf{1}'M_0^{-1}\mathbf{1} \end{bmatrix}.$$

Note that $M\mathbf{1} = (1+mb_n)\mathbf{1}$. It implies that $M^{-1}\mathbf{1} = \frac{1}{1+mb_n}\mathbf{1}$. As a result,

$$\mathbf{1}'M_0^{-1}\mathbf{1} = \frac{1+mb_n}{m+1}\mathbf{1}'M_0^{-1}\mathbf{1} = \frac{1+mb_n}{m+1}\mathbf{1}'\Big(\frac{1}{1+mb_n}\mathbf{1}\Big) = 1.$$

Plugging it into the expression of $X'A^{-1}X$ gives

$$X'A^{-1}X = \begin{bmatrix} \beta'S^{-1}\beta + 1 & \beta'S^{-1}\beta - 1 \\ \beta'S^{-1}\beta - 1 & \beta'S^{-1}\beta + 1 \end{bmatrix}.$$

It follows from direct calculations that

$$(D + X'A^{-1}X)^{-1}(X'A^{-1}X) = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & \frac{3\beta'S^{-1}\beta+1}{\beta'S^{-1}\beta-1} \end{bmatrix}. \tag{3.A.50}$$

Under the condition $|\beta'S^{-1}\beta - 1| \ge C$, the absolute value of $\frac{3\beta'S^{-1}\beta+1}{\beta'S^{-1}\beta-1}$ is bounded by a constant. Therefore, the spectral norm of the matrix in (3.A.50) is bounded by a constant.

We plug it into (3.A.49) to get

$$\|P^{-1}\| \le C\|A^{-1}\| \le C\max\big\{|\lambda_{\min}(S)|^{-1},\ |\lambda_{\min}(M)|^{-1}\big\}.$$

The minimum eigenvalue of M is $(1-b_n)$. Hence, under the condition of $|\lambda_{\min}(S)| \gg 1-b_n$, we immediately have $\|P^{-1}\| \le C(1-b_n)^{-1}$. We plug it into (3.A.48) to get

$$|\lambda_K(HPH)| \ge C^{-1}(1-b_n). \tag{3.A.51}$$

Additionally, $\|\widetilde{P}\| \le C$ by (3.2.1). It follows from the connection between P and $\widetilde{P}$ in (3.2.13) that $\|P\| \le C$. Combining it with (3.A.44) gives $\|HPH\| \le C$, i.e.,

$$\lambda_1(HPH) \le C. \tag{3.A.52}$$

Here (3.A.51) and (3.A.52) are satisfied with probability $1-o(1)$, because their derivation uses (3.A.44). We plug (3.A.51)-(3.A.52) into (3.A.41). It yields that $|\lambda_K|/\sqrt{\lambda_1} \ge C^{-1}\|\theta\|(1-b_n)$, with probability $1-o(1)$. This proves the lower bound of $|\lambda_K|/\sqrt{\lambda_1}$.

Next, we show that, if $\|\theta\|(1-b_n) \to 0$, the two random-label DCBM models associated with m_1 and m_2 are asymptotically indistinguishable. It is sufficient to show that each random-label DCBM is asymptotically indistinguishable from the (fixed-label) DCBM with K_0 communities.

Fix $m \ge 1$. Let $f_0(A)$ and $f_1(A)$ be the respective likelihood of the (fixed-label) DCBM and the random-label DCBM. Write $\widetilde{\Omega} = \Theta\widetilde{\Pi}\widetilde{P}\widetilde{\Pi}'\Theta$ and $\Omega = \Theta\Pi P\Pi'\Theta$. It is seen that

$$f_0(A) = \prod_{1\le i<j\le n} \widetilde{\Omega}_{ij}^{A_{ij}}(1-\widetilde{\Omega}_{ij})^{1-A_{ij}}, \qquad f_1(A) = \int \prod_{1\le i<j\le n} \Omega_{ij}^{A_{ij}}(1-\Omega_{ij})^{1-A_{ij}} d\mathbb{P}(\Pi).$$

Recall that $\widetilde{\mathcal{N}}_1, \widetilde{\mathcal{N}}_2, \ldots, \widetilde{\mathcal{N}}_{K_0}$ are the (non-stochastic) communities in the first DCBM. We observe that $\widetilde{\Omega}_{ij} \neq \Omega_{ij}$ only when both i and j are in $\widetilde{\mathcal{N}}_{K_0}$. Therefore, the likelihood ratio is

$$L(A) \equiv \frac{f_1(A)}{f_0(A)} = \int \prod_{\{i,j\}\subset \widetilde{N}_{K_0}, i<j} \Big(\frac{\Omega_{ij}}{\widetilde{\Omega}_{ij}}\Big)^{A_{ij}} \Big(\frac{1-\Omega_{ij}}{1-\widetilde{\Omega}_{ij}}\Big)^{1-A_{ij}} d\mathbb{P}(\Pi). \tag{3.A.53}$$

When i, j are both in $\widetilde{N}_{K_0}$, it is seen that

$$\widetilde{\Omega}_{ij} = \theta_i\theta_j, \qquad \Omega_{ij} = \theta_i\theta_j \cdot \pi_i'\Big(\frac{(m+1)}{1+mb_n}M\Big)\pi_j,$$

where $\pi_i = e_k$ if and only if $\ell_i = K_0 - 1 + k$, $1 \le k \le m+1$, and $e_1, e_2, \ldots, e_{m+1}$ are the standard bases of $\mathbb{R}^{m+1}$. Here we have mis-used the notation π_i; previously, we use π_i' to denote the i-th row of Π, but currently, the i-th row of Π is $(\mathbf{0}_{K_0-1}', \pi_i')'$. Define

$$z_i = \pi_i - \frac{1}{m+1}\mathbf{1}_{m+1}, \qquad \text{for all } i \in \widetilde{\mathcal{N}}_{K_0}.$$

The random vectors $\{z_i\}_{i\in\widetilde{\mathcal{N}}_{K_0}}$ are independently and identically distributed, satisfying $\mathbb{E}z_i = \mathbf{0}$ and $\|z_i\| \le 1$. In the paragraph below (3.A.45), we have seen that

$$\frac{m+1}{1+mb_n}M = \mathbf{1}_{m+1}\mathbf{1}_{m+1}' + \frac{1-b_n}{1+mb_n}\begin{bmatrix} m & -1 & \cdots & -1 \\ -1 & m & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & -1 & m \end{bmatrix} \equiv \mathbf{1}_{m+1}\mathbf{1}_{m+1}' + G.$$

The matrix G satisfies that $G\mathbf{1}_{m+1} = \mathbf{0}$ and $\|G\| \le C(1-b_n)$. It follows that

$$\begin{aligned}
\Omega_{ij} &= \theta_i\theta_j \cdot \pi_i' \left(\mathbf{1}_{m+1}\mathbf{1}_{m+1}' + G\right)\pi_j \\
&= \theta_i\theta_j + \theta_i\theta_j(\pi_i' G\pi_j) \\
&= \theta_i\theta_j + \theta_i\theta_j\Big(\frac{1}{m+1}\mathbf{1}_{m+1} + z_i\Big)' G\Big(\frac{1}{m+1}\mathbf{1}_{m+1} + z_j\Big) \\
&= \theta_i\theta_j(1 + z_i' G z_j).
\end{aligned} \tag{3.A.54}$$

We plug it into (3.A.53) to get

$$L(A) \equiv \frac{f_2(A)}{f_1(A)} = \mathbb{E}_z\left\{\prod_{\substack{i,j\in\widetilde{\mathcal{N}}_{K-1}\\ i<j}} (1+z_i'Gz_j)^{A_{ij}}\left[\frac{1-\theta_i\theta_j(1+z_i'Gz_j)}{1-\theta_i\theta_j}\right]^{1-A_{ij}}\right\}. \tag{3.A.55}$$

The χ^2-distance between two models is $\mathbb{E}_{A\sim f_0}[(L(A)-1)^2]$. To show that the two models are asymptotically indistinguishable, it suffices to show that the χ^2-distance is $o(1)$ Tsybakov (2008). Using the property that $\mathbb{E}_{A\sim f_0}[(L(A)-1)^2] = \mathbb{E}_{A\sim f_0}[L^2(A)] - 1$, we only need to show

$$\mathbb{E}_{A\sim f_0}[L^2(A)] \le 1 + o(1). \tag{3.A.56}$$

We now show (3.A.56). Write $L(A) = \mathbb{E}_z[g(A,z)]$, where $g(A,z)$ is the term inside the expectation in (3.A.55). Let $\{\widetilde{z}_i\}_{i\in\widetilde{\mathcal{N}}_{K_0}}$ be an independent copy of $\{z_i\}_{i\in\widetilde{\mathcal{N}}_{K_0}}$. Then,

$$\mathbb{E}_{A\sim f_0}[L^2(A)] = \mathbb{E}_{A\sim f_0}\Big\{\mathbb{E}_z[g(A,z)]\cdot\mathbb{E}_{\widetilde{z}}[g(A,\widetilde{z})]\Big\} = \mathbb{E}_{z,\widetilde{z}}\Big\{\mathbb{E}_{A\sim f_0}[g(A,z)g(A,\widetilde{z})]\Big\}. \tag{3.A.57}$$

Using the expression of $g(A,z)$ in (3.A.55), we have

$$g(A,z)g(A,\widetilde{z}) = \prod_{\substack{i,j\in\widetilde{\mathcal{N}}_{K_0}\\ i<j}} \big[(1+z_i'Gz_j)(1+\widetilde{z}_i'G\widetilde{z}_j)\big]^{A_{ij}}\left\{\frac{[1-\theta_i\theta_j(1+z_i'Gz_j)][1-\theta_i\theta_j(1+\widetilde{z}_i'G\widetilde{z}_j)]}{(1+\theta_i\theta_j)^2}\right\}^{1-A_{ij}}.$$

Here A_{ij}'s are independent Bernoulli variables, where $\mathbb{P}(A_{ij}=1) = \theta_i\theta_j$. If we take expectation with respect to A_{ij} in each term of the product, it gives

$$\begin{aligned}
&(1+z_i'Gz_j)(1+\widetilde{z}_i'G\widetilde{z}_j)\cdot\mathbb{P}(A_{ij}=1) + \frac{[1-\theta_i\theta_j(1+z_i'Gz_j)][1-\theta_i\theta_j(1+\widetilde{z}_i'G\widetilde{z}_j)]}{(1-\theta_i\theta_j)^2}\cdot\mathbb{P}(A_{ij}=0) \\
=\ &\theta_i\theta_j(1+z_i'Gz_j)(1+\widetilde{z}_i'G\widetilde{z}_j) + \frac{[1-\theta_i\theta_j(1+z_i'Gz_j)][1-\theta_i\theta_j(1+\widetilde{z}_i'G\widetilde{z}_j)]}{1-\theta_i\theta_j} \\
=\ &(1+z_i'Gz_j)(1+\widetilde{z}_i'G\widetilde{z}_j)\Big(\theta_i\theta_j + \frac{\theta_i^2\theta_j^2}{1-\theta_i\theta_j}\Big) + \frac{1-\theta_i\theta_j(1+z_i'Gz_j)-\theta_i\theta_j(1+\widetilde{z}_i'G\widetilde{z}_j)}{1-\theta_i\theta_j} \\
=\ &(1+z_i'Gz_j)(1+\widetilde{z}_i'G\widetilde{z}_j)\frac{\theta_i\theta_j}{1-\theta_i\theta_j} + 1 - \frac{\theta_i\theta_j(1+z_i'Gz_j+\widetilde{z}_i'G\widetilde{z}_j)}{1-\theta_i\theta_j} \\
=\ &1 + \frac{\theta_i\theta_j}{1-\theta_i\theta_j}(z_i'Gz_j)(\widetilde{z}_i'G\widetilde{z}_j).
\end{aligned}$$

As a result,

$$\mathbb{E}_{A\sim f_0}[g(A,z)g(A,\widetilde{z})] = \prod_{\{i,j\}\subset\widetilde{\mathcal{N}}_{K_0},i<j}\left[1+\frac{\theta_i\theta_j}{1-\theta_i\theta_j}(z_i'Gz_j)(\widetilde{z}_i'G\widetilde{z}_j)\right]$$
$$\leq \exp\left(\sum_{\{i,j\}\subset\widetilde{\mathcal{N}}_{K_0},i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}(z_i'Gz_j)(\widetilde{z}_i'G\widetilde{z}_j)\right),$$

where the second line is from the inequality that $1+x\leq e^x$ for all $x\in\mathbb{R}$. We plug it into (3.A.57). Then, to show (3.A.56), it suffices to show that

$$\mathbb{E}_{z,\widetilde{z}}[\exp(Y)]\leq 1+o(1), \qquad \text{where} \quad Y\equiv\sum_{\{i,j\}\subset\widetilde{\mathcal{N}}_{K_0},i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}(z_i'Gz_j)(\widetilde{z}_i'G\widetilde{z}_j). \quad (3.A.58)$$

We now show (3.A.58). We drop the subscript $\{i,j\}\subset\widetilde{\mathcal{N}}_{K_0}$ in most places to make notations simpler. The matrix G can be re-written as

$$G=\frac{1-b_n}{1+mb_n}\Big[(m+1)I_{m+1}-\mathbf{1}_{m+1}\mathbf{1}_{m+1}'\Big].$$

Additionally, $z_i'\mathbf{1}_{m+1}\equiv 0$. It follows that $z_i'Gz_j=\frac{(m+1)(1-b_n)}{1+mb_n}(z_i'z_j)$. As a result,

$$Y=\frac{(m+1)^2(1-b_n)^2}{(1+mb_n)^2}\sum_{i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}(z_i'z_j)(\widetilde{z}_i'\widetilde{z}_j')$$
$$=\frac{1}{(m+1)^2}\sum_{1\leq k,\ell\leq m+1}\underbrace{\frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2}\sum_{i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}z_i(k)z_j(k)\widetilde{z}_i(\ell)\widetilde{z}_j(\ell)}_{\equiv Y_{k\ell}}.$$

By Jensen's inequality, $\exp(Y)=\exp\big(\frac{1}{(m+1)^2}\sum_{k,\ell}Y_{k\ell}\big)\leq\frac{1}{(m+1)^2}\sum_{k,\ell}\exp(Y_{k\ell})$. It follows that

$$\mathbb{E}_{z,\widetilde{z}}[\exp(Y)]\leq\frac{1}{(m+1)^2}\sum_{1\leq k,\ell\leq m+1}\mathbb{E}_{z,\widetilde{z}}\big[\exp(Y_{k\ell})\big]\leq\max_{1\leq k,\ell\leq m+1}\mathbb{E}_{z,\widetilde{z}}\big[\exp(Y_{k\ell})\big].$$

Therefore, to show (3.A.58), it suffices to show that, for each $1\leq k,\ell\leq m+1$,

$$\mathbb{E}_{z,\widetilde{z}}\big[\exp(Y_{k\ell})\big]\leq 1+o(1). \quad (3.A.59)$$

Fix (k,ℓ). We now show (3.A.59). Define $\sigma_i=z_i(k)\widetilde{z}(\ell)$, for all $i\in\widetilde{N}_{K_0}$. Then,

$$Y_{k\ell}=\frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2}\sum_{i<j}\frac{\theta_i\theta_j}{1-\theta_i\theta_j}\sigma_i\sigma_j$$
$$=\frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2}\sum_{i<j}\sum_{s=1}^{\infty}\theta_i^s\theta_j^s\sigma_i\sigma_j$$
$$=\sum_{s=1}^{\infty}\underbrace{(1-\theta_{\max}^2)\theta_{\max}^{2s-2}}_{\equiv w_s}\underbrace{\frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2(1-\theta_{\max}^2)\theta_{\max}^{2s-2}}\sum_{i<j}\theta_i^s\theta_j^s\sigma_i\sigma_j}_{\equiv X_s}.$$

In the second line above, we used the Taylor expansion $\frac{\theta_i\theta_j}{1-\theta_i\theta_j}=\sum_{s=1}^{\infty}\theta_i^s\theta_j^s$. It is valid because $|\theta_i\theta_j|\leq\theta_{\max}^2=o(1)$. In the third line, we have switched the order of summation.

It is valid because the double sum is finite if we take the absolute value of each summand. The numbers $\{w_s\}_{s=1}^{\infty}$ satisfy that $\sum_{s=1}^{\infty} w_s = 1$. By Jenson's inequality,

$$\exp(Y_{k\ell}) = \exp\Big(\sum_{s=1}^{\infty} w_s \cdot X_s\Big) \leq \sum_{s=1}^{\infty} w_s \cdot \exp(X_s).$$

By Fatou's lemma,

$$\mathbb{E}_{\sigma}[\exp(Y_{k\ell})] \leq \sum_{s=1}^{\infty} w_s \cdot \mathbb{E}_{\sigma}[\exp(X_s)] \leq \max_{s\geq 1}\ \mathbb{E}_{\sigma}[\exp(X_s)] \tag{3.A.60}$$

It remains to study X_s. Note that

$$\begin{aligned}
X_s &= \frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2(1-\theta_{\max}^2)\theta_{\max}^{2s-2}} \sum_{i<j} \theta_i^s\theta_j^s\sigma_i\sigma_j \\
&= \frac{(m+1)^4(1-b_n)^2}{(1+mb_n)^2(1-\theta_{\max}^2)\theta_{\max}^{2s-2}} \left[\frac{1}{2}\sum_{i,j} \theta_i^s\theta_j^s\sigma_i\sigma_j - \sum_i \theta_i^{2s}\sigma_i^2\right] \\
&\leq \frac{(m+1)^4(1-b_n)^2}{2(1+mb_n)^2(1-\theta_{\max}^2)\theta_{\max}^{2s-2}} \left(\sum_i \theta_i^s\sigma_i\right)^2.
\end{aligned}$$

Note that the summation is over $i \in \widetilde{N}_{K_0}$. Let $\theta^* \in \mathbb{R}^n$ be defined by $\theta_i^* = \theta_i \cdot 1\{i \in \widetilde{N}_{K_0}\}$. Since $1-\theta_{\max}^2 \geq 1/2$ and $\|\theta^*\|_{2s}^{2s} \leq \theta_{\max}^{2s-2}\|\theta^*\|^2 \leq \theta_{\max}^{2s-2}\|\theta\|^2$, we have

$$X_s \leq \frac{a_0(1-b_n)^2\|\theta\|^2}{\|\theta^*\|_{2s}^{2s}} \left(\sum_i \theta_i^s\sigma_i\right)^2, \tag{3.A.61}$$

for a constant $a_0 > 0$. We apply Hoeffding's inequality to get that, for all $t > 0$,

$$\mathbb{P}\Big(\Big|\sum_i \theta_i^s\sigma_i\Big| > t\Big) \leq 2\exp\Big(-\frac{t^2}{2\sum_i \theta_i^{2s}}\Big) = 2\exp\Big(-\frac{t^2}{2\|\theta^*\|_{2s}^{2s}}\Big). \tag{3.A.62}$$

For any nonnegative variable X, using the formula of integration by part, we can derive that $\mathbb{E}[\exp(aX)] = 1 + a\int_0^{\infty}\exp(at)\mathbb{P}(X > t)dt$. As a result,

$$\begin{aligned}
\mathbb{E}_{\sigma}\big[\exp(X_s)\big] &\leq \mathbb{E}_{\sigma}\left\{\exp\Big[\frac{a_0(1-b_n)^2\|\theta\|^2}{\|\theta^*\|_{2s}^{2s}}\Big(\sum_i \theta_i^s\sigma_i\Big)^2\Big]\right\} \\
&= 1 + \frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}} \int_0^{\infty} \exp\left(\frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}}t\right) \cdot \mathbb{P}\Big\{\Big(\sum_i \theta_i^m\sigma_i\Big)^2 > t\Big\}dt \\
&\leq 1 + \frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}} \int_0^{\infty} \exp\left(\frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}}t\right) \cdot \exp\left(-\frac{t}{2\|\theta^*\|_{2s}^{2s}}\right) dt \\
&= 1 + \frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}} \int_0^{\infty} \exp\left(-\frac{1-2a_0\|\theta\|^2(1-b_n)^2}{2\|\theta^*\|_{2s}^{2s}}t\right) dt \\
&= 1 + \frac{a_0\|\theta\|^2(1-b_n)^2}{\|\theta^*\|_{2s}^{2s}} \cdot \frac{2\|\theta^*\|_{2s}^{2s}}{1-2a_0\|\theta\|^2(1-b_n)^2} \\
&= 1 + \frac{2a_0\|\theta\|^2(1-b_n)^2}{1-2a_0\|\theta\|^2(1-b_n)^2}.
\end{aligned}$$

The right hand side does not depend on s, so the same bound holds for $\max_{s\geq 1}\{\mathbb{E}_{\sigma}[\exp(X_s)]\}$.

When $\|\theta\|^2(1-b)^2 \to 0$, this upper bound is $1+o(1)$. Plugging it into (3.A.60) gives (3.A.59). Then, the second claim follows. $\square$

3.A.3 Proof of Theorem 3.2.5

We show a slightly stronger argument. Given $1 \leq K_1 < K_2 \leq m_0$, let $\mathcal{M}_n(K_1, K_2, a_n)$ be the sub-collection of $\mathcal{M}_n(m_0, a_n)$ corresponding to $K_1 \leq K \leq K_2$. Note that

$$\inf_{\widehat{K}}\Big\{\sup_{\mathcal{M}_n(m_0,a_n)} \mathbb{P}(\widehat{K} \neq K)\Big\} \geq \inf_{\widehat{K}}\Big\{\sup_{\mathcal{M}_n(K_1,K_2,a_n)} \mathbb{P}(\widehat{K} \neq K)\Big\}.$$

It suffices to lower bound the right hand side.

Fix an arbitrary DCBM model with (K_1-1) communities. For each $1 \leq m \leq K_2-K_1+1$, we use (3.2.13)-(3.2.14) to construct a random-label DCBM with (K_1-1+m) communities, where $b_n = 1 - c\,\|\theta\|^{-1}a_n$, for a constant c to be decided. Let $\mathbb{P}_k$ denote the probability measure associated with the k-community random-label DCBM, for $K_1 \leq k \leq K_2$. By Theorem 3.2.4, we can choose an appropriately small constant c such that $|\lambda_K|/\sqrt{\lambda_1} \geq a_n$ with probability $1-o(n^{-1})$, under each $\mathbb{P}_k$. Additionally, using a proof similar to that of (3.A.43), we can show that (3.2.1)-(3.2.2) are satisfied with probability $1-o(n^{-1})$. Therefore, under each $\mathbb{P}_k$, the realization of (Θ, Π, P) belongs to $\mathcal{M}_n(K_1, K_2, a_n)$ with probability $1-o(n^{-1})$. Then, for any $\widehat{K}$,

$$\sup_{\mathcal{M}_n(K_1,K_2,a_n)} \mathbb{P}(\widehat{K} \neq K) \geq \max_{K_1 \leq k \leq K_2} \mathbb{P}_k(\widehat{K} \neq K) + o(n^{-1}). \tag{3.A.63}$$

To bound the right hand side of (3.A.63), consider a multi-hypothesis testing problem: Given an adjacency matrix A, choose one out of the models $\{\mathbb{P}_k\}_{K_1 \leq k \leq K_2}$. For any test ψ, define

$$\bar{p}(\psi) = \frac{1}{K_2 - K_1 + 1} \sum_{k=K_1}^{K_2} \mathbb{P}_k(\psi \neq k).$$

We apply (Tsybakov, 2008, Proposition 2.4). It yields that

$$\frac{1}{K_2 - K_1} \sum_{k=K_1+1}^{K_2} \chi^2(\mathbb{P}_k, \mathbb{P}_{K_1}) \leq \alpha^* \quad \Longrightarrow \quad \inf_{\psi} \bar{p}(\psi) \geq \sup_{0<\tau<1} \left\{ \frac{\tau(K_2-K_1)}{1+\tau(K_2-K_1)}[1-\tau(\alpha^*+1)] \right\}.$$

We have shown in Theorem 3.2.4 that $\alpha^* = o(1)$. By letting $\tau = 1/2$ in the above, we immediately find that

$$\inf_{\psi} \bar{p}(\psi) \gtrsim \frac{K_2 - K_1}{2 + (K_2 - K_1)}\Big(1 - \frac{1+o(1)}{2}\Big) \geq 1/6 + o(1). \tag{3.A.64}$$

Now, given any estimator $\widehat{K}$, it defines a test $\psi_{\widehat{K}}$, where $\psi_{\widehat{K}} = \widehat{K}$ if $K_1 \leq \widehat{K} \leq K_2$ and $\psi_{\widehat{K}} = K_1$ otherwise. It is easy to see that

$$\bar{p}(\psi_{\widehat{K}}) \leq \max_{K_1 \leq k \leq K_2} \mathbb{P}_k(\widehat{K} \neq k). \tag{3.A.65}$$

Combining (3.A.64)-(3.A.65) gives that $\max_{K_1 \leq k \leq K_2} \mathbb{P}_k(\widehat{K} \neq k) \geq 1/6 + o(1)$. We plug it into (3.A.63) to get the claim. $\square$

3.B Proof of results in Section 3.3

3.B.1 Proof of Lemma 30

By definition of H, we have $\Pi^2\Theta\Pi = \|\theta\|^2 \cdot H^2$. As a result, the matrix $U = \|\theta\|^{-1}\Theta\Pi H^{-1}$ satisfies that $U'U = I_K$. We now write

$$\Omega = \Theta\Pi P\Pi'\Theta = \|\theta\|^2 \cdot U \cdot (HPH) \cdot U', \qquad \text{where} \quad U'U = I_K.$$

Since U contains orthonormal columns, the nonzero eigenvalues of Ω are the nonzero eigenvalues of $\|\theta\|^2(HPH)$. This proves that $\lambda_k = \|\theta\|^2\mu_k$. Furthermore, there is a one-to-one correspondence between the eigenvectors of Ω and the eigenvectors of HPH through

$$[\xi_1, \xi_2, \ldots, \xi_k] = U[\eta_1, \eta_2, \ldots, \eta_K].$$

It follows that $\xi_k = U\eta_k = \|\theta\|^{-1}\Theta\Pi H^{-1}\eta_k$. This proves the claim about ξ_k. We can multiply both sides of the equation $\xi_k = U\eta_k$ by $\|\theta\|^{-1}H^{-1}\Pi'\Theta$ from the left. It yields that

$$\begin{aligned}\|\theta\|^{-1}H^{-1}\Pi'\Theta\xi_k &= (\|\theta\|^{-1}H^{-1}\Pi'\Theta)(\|\theta\|^{-1}\Theta\Pi H^{-1}\eta_k)\\ &= \|\theta\|^{-2}H^{-1}(\Pi'\Theta^2\Pi)H^{-1}\eta_k \;=\; \eta_k.\end{aligned}$$

This proves the claim about η_k. Last, the condition (3.2.4) ensures that the multiplicity of μ_1 is 1 and that μ_1 is a strictly positive vector. It follows that λ_1 has a multiplicity of 1. Note that $\xi_k = U\eta_k$ implies

$$\xi_1(i) = \|\theta\|^{-1}\theta_i \sum_{k=1}^{K} H_{kk}^{-1}\pi_i(k)\eta_1(k) \geq \|\theta\|^{-1}\theta_i \min_{1\leq k\leq K}\{H_{kk}^{-1}\eta_1(k)\}.$$

Since η_1 is a positive vector and H is a positive diagonal matrix, we conclude that all entries of ξ_1 are positive. □

3.B.2 Proof of Lemma 32

We fix an arbitrary $(K-1)\times(K-1)$ orthogonal matrix Γ and drop "Γ" in the notations η_k, ξ_k, r_i, v_k. By Definition 31,

$$[\eta_1, \eta_2, \ldots, \eta_K] = [\eta_1, \eta_2^*, \ldots, \eta_K^*]\begin{bmatrix}1 & \\ & \Gamma\end{bmatrix}, \qquad [\xi_1, \xi_2, \ldots, \eta_K] = [\xi_1, \xi_2^*, \ldots, \xi_K^*]\begin{bmatrix}1 & \\ & \Gamma\end{bmatrix}.$$

Here, $\eta_1, \eta_2^*, \ldots, \eta_K^*$ is a particular candidate of the eigenvectors of HPH and $\xi_1, \xi_2^*, \ldots, \xi_K^*$ is linked to $\eta_1, \eta_2^*, \ldots, \eta_K^*$ through

$$[\xi_1, \xi_2^*, \ldots, \xi_K^*] = \|\theta\|^{-1}\Theta\Pi H^{-1}[\eta_1, \eta_2^*, \ldots, \eta_K^*].$$

It follows immediately that

$$[\xi_1, \xi_2, \ldots, \xi_K] = \|\theta\|^{-1}\Theta\Pi H^{-1}[\eta_1, \eta_2, \ldots, \eta_K]. \tag{3.B.66}$$

As a result, for any true community $\mathcal{N}_k$,

$$\xi_\ell(i) = [\theta_i/(\|\theta\|H_{kk})]\cdot\eta_\ell(k), \qquad \text{for all } i \in \mathcal{N}_k.$$

We plug it into the definition of $R^{(m)}$ to get that for each $i \in \mathcal{N}_k$ and $1 \leq \ell \leq m-1$,

$$R^{(m)}(i,\ell) = \frac{\xi_{\ell+1}(i)}{\xi_1(i)} = \frac{[\theta_i/(\|\theta\|H_{kk})]\cdot\eta_{\ell+1}(k)}{[\theta_i/(\|\theta\|H_{kk})]\cdot\eta_1(k)} = \frac{\eta_{\ell+1}(k)}{\eta_1(k)} = V^{(m)}(k,\ell).$$

It follows that $r_i^{(m)} = v_k^{(m)}$ for each $i \in \mathcal{N}_k$. □

3.B.3 Proof of Lemma 33

The matrix $V^{(K)}(\Gamma)$ was studied in Jin (2015); Jin et al. (2017). Since the pairwise distances for rows in $V^{(K)}(\Gamma)$ are invariant of Γ, the quantity $d_K(V^{(K)}(\Gamma))$ does not change with Γ either. Using Lemma B.3 of Jin (2015), we immediately know that $d_K(V^{(K)}(\Gamma)) \geq \sqrt{2}$.

Below, we fix $1 < m < K$ and a $(K-1)\times(K-1)$ orthogonal matrix Γ, and study $d_m(V^{(m)}(\Gamma))$. For notation simplicity, we drop "Γ" when there is no confusion.

We apply a bottom up pruning procedure (same as in Definition 29) to $V^{(m)}$. First, we find two rows $v_k^{(m)}$ and $v_\ell^{(m)}$ that attain the minimum pairwise distance (if there is a tie, pick the first pair in the lexicographical order) and change the ℓth row to $v_k^{(m)}$ (suppose $k < \ell$). Denote the resulting matrix by $V^{(m,K-1)}$. Next, we consider the rows of $V^{(m,K-1)}$ and similarly find two rows attaining the minimum pairwise distance and replace one row by the other. Denote the resulting matrix by $V^{(m,K-2)}$. We repeat these steps to get a sequence of matrices:

$$V^{(m,K)},\ V^{(m,K-1)},\ V^{(m,K-2)},\ \ldots, V^{(m,2)},\ V^{(m,1)},$$

where $V^{(m,K)} = V^{(m)}$ and for each $1 \leq k \leq K$, $V^{(m,k)}$ has at most k distinct rows. Comparing it with the definition of $d_k(V^{(m)})$ (see Definition 29), we find that $V^{(m,k-1)}$ differs from $V^{(m,k)}$ in only 1 row, and the difference on this row is a vector whose Euclidean norm is exactly $d_k(V^{(m)})$. As a result,

$$\|V^{(m,k)} - V^{(m,k-1)}\| = d_k(V^{(m)}), \qquad 2 \leq k \leq K. \tag{3.B.67}$$

By triangle inequality and the fact that $d_k(V^{(m)}) \leq d_{k-1}(V^{(m)})$, we have

$$\|V^{(m,K)} - V^{(m,m-1)}\| \leq \sum_{k=m}^{K} d_k(V^{(m)}) \leq (K-m+1)\cdot d_m(V^{(m)}).$$

To show the claim, it suffices to show that

$$\|V^{(m,K)} - V^{(m,m-1)}\| \geq C. \tag{3.B.68}$$

We now show (3.B.68). Introduce two matrices

$$V_*^{(m,K)} = [\mathbf{1}_K, V^{(m,K)}], \qquad V_*^{(m,m-1)} = [\mathbf{1}_K, V^{(m,m-1)}],$$

where $\mathbf{1}_K$ is the K-dimensional vector of 1s. Adding the vector $\mathbf{1}_K$ as the first column changes neither the number of distinct rows nor pairwise distances among rows. Additionally,

$$\|V^{(m,K)} - V^{(m,m-1)}\| = \|V_*^{(m,K)} - V_*^{(m,m-1)}\|. \tag{3.B.69}$$

Let $\sigma_m(U)$ denote the m-th singular value of a matrix U. Since $V_*^{(m,m-1)}$ has at most $(m-1)$ distinct rows, its rank is at most $(m-1)$. As a result,

$$\sigma_m(V_*^{(m,m-1)}) = 0. \tag{3.B.70}$$

We then study $\sigma_m(V_*^{(m,K)})$. Note that

$$V_*^{(m,K)} = [\mathbf{1}_K, V^{(m)}] = \begin{bmatrix} 1 & v_1^{(m)} \\ \vdots & \vdots \\ 1 & v_K^{(m)} \end{bmatrix} = [\mathrm{d}iag(\eta_1)]^{-1}\cdot\big[\eta_1, \eta_2(\Gamma), \ldots, \eta_m(\Gamma)\big], \tag{3.B.71}$$

where $\eta_1, \eta_2(\Gamma), \ldots, \eta_K(\Gamma)$ is one choice of eigenvectors of HPH indexed by Γ (see Definition 31) and $\mathrm{d}iag(\eta_1)$ is the diagonal matrix whose diagonal entries are from η_1. Write for short $Q = [\eta_1, \eta_2(\Gamma), \ldots, \eta_m(\Gamma)]$. We have

$$(V_*^{(m,K)})' V_*^{(m,K)} = Q'[\mathrm{d}iag(\eta_1)]^{-2} Q.$$

By the last item of (3.2.4) and that $\|\eta_1\| = 1$, we conclude that $\eta_1(k) \asymp 1/\sqrt{K}$ for all $1 \le k \le K$. In particular, there exists a constant $c > 0$ such that $\big([\mathrm{d}iag(\eta_1)]^{-2} - cI_K\big)$ is a positive semi-definite matrix. It follows that $\big(Q'[\mathrm{d}iag(\eta_1)]^{-2}Q - cQ'Q\big)$ is a positive semi-definite matrix. Therefore,

$$\lambda_m\big((V_*^{(m,K)})' V_*^{(m,K)}\big) \ge \lambda_m(cQ'Q) = c \cdot \lambda_m(Q'Q), \tag{3.B.72}$$

where $\lambda_m(\cdot)$ denotes the m-th largest eigenvalue of a symmetric matrix. By (3.3.15), for some pre-specified choice of eigenvectors, $\eta_1, \eta_2^*, \ldots, \eta_K^*$, of HPH,

$$Q \text{ is the first } m \text{ columns of the matrix } [\eta_1, \eta_2^*, \ldots, \eta_K^*] \cdot \mathrm{d}iag(1, \Gamma).$$

Note that $[\eta_1, \eta_2^*, \ldots, \eta_K^*]$ and $\mathrm{d}iag(1, \Gamma)$ are both $K \times K$ orthogonal matrices. Then, their product is also an orthogonal matrix, and the columns in Q are orthonormal. It follows that

$$Q'Q = I_m.$$

This shows that the right hand side of (3.B.72) is equal to c. The left hand side of (3.B.72) is equal to $\sigma_m^2(V_*^{(m,K)})$. It follows that

$$\sigma_m(V_*^{(m,K)}) \ge C. \tag{3.B.73}$$

We now combine (3.B.70) and (3.B.73), and apply Weyl's inequality for singular values (Horn and Johnson, 1985, Corollary 7.3.5). It gives

$$C \le \sigma_m(V_*^{(m,K)}) - \sigma_m(V_*^{(m,m-1)}) \le \|V_*^{(m,K)} - V_*^{(m,m-1)}\|.$$

Combining it with (3.B.69) gives (3.B.68). The claim follows immediately. □

Remark. The proof of Theorem 3.2.2 uses $\max_{1\le k\le K} \|v_k^{(m)}(\Gamma)\| \le C$, and we prove this claim here. Note that $v_k^{(m)}(\Gamma)$ is a sub-vector of the kth row of $V_*^{(m,K)}$. In light of (3.B.71), the row-wise ℓ_2-norms of $V_*^{(m,K)}$ are uniformly bounded by $C\|\mathrm{d}iag^{-1}(\eta_1)\|$. We have argued that $\eta_1(k) \asymp 1/\sqrt{K} \le C$ for all $1 \le k \le K$. As a result, $\max_{1\le k\le K} \|v_k^{(m)}(\Gamma)\| \le C$.

3.B.4 Proof of Lemma 34

Since $\|\widehat{r}^{(m)} - r_i^{(m)}(\Gamma)\| \le \|\widehat{r}^{(K)} - r_i^{(K)}(\Gamma)\|$, we only need to show the claim for $m = K$. Write $r_i^{(K)}(\Gamma) = r_i(\Gamma)$ for short. In the special case of $\Gamma = I_{K-1}$ (i.e., $\eta_k(\Gamma) = \eta_k^*$ for $2 \le k \le K$, by Definition 31), we further write $r_i = r_i(I_{K-1})$ for short. It is easy to see that

$$r_i(\Gamma) = \Gamma' \cdot r_i, \qquad \text{for any orthogonal matrix } \Gamma \in \mathbb{R}^{(K-1)\times(K-1)}.$$

It suffices to show that with probability $1 - O(n^{-3})$ there exists a $(K-1)\times(K-1)$ orthogonal matrix Γ, which may depend on n and $\widehat{R}^{(K)}$, such that

$$\max_{1\le i\le n} \|\widehat{r}_i - \Gamma' \cdot r_i\| \le C s_n^{-1} \sqrt{\log(n)}.$$

Such a bound was given by Theorem 4.1 of Jin et al. (2021a) (also, see Lemma 2.1 of Jin et al. (2017) for a special case where $\lambda_2, \ldots, \lambda_K$ are at the same order). □

3.B.5 Proof of Theorem 3.3.2

The key of proof is the following lemma, which characterizes the change of the k-means objective under perturbation of cluster assignment. Consider the problem of clustering points $y_1, y_2, \ldots, y_n$ into two disjoint clusters A and B. The k-means objective is the residual sum of squares by setting the two cluster centers as the within-cluster means. Now, we move a subset C from cluster A to cluster B. The new clusters are $\widetilde{A} = A \setminus C$ and $\widetilde{B} = B \cup C$, and the cluster centers are updated accordingly. There is an explicit formula for the change of the k-means objective:

Lemma 45. *For any $y_1, y_2, \ldots, y_n \in \mathbb{R}^d$ and subset $M \subset \{1, 2, \ldots, n\}$, define $\bar{y}_M = \frac{1}{|M|}\sum_{i\in M} y_i$. Let $\{1, 2, \ldots, n\} = A \cup B$ be a partition, and let C be a strict subset of A. Write $\widetilde{A} = A\backslash C$ and $\widetilde{B} = B \cup C$. Define*

$$RSS = \sum_{i\in A} \|y_i - \bar{y}_A\|^2 + \sum_{i\in B} \|y_i - \bar{y}_B\|^2, \qquad \widetilde{RSS} = \sum_{i\in\widetilde{A}} \|y_i - \bar{y}_{\widetilde{A}}\|^2 + \sum_{i\in\widetilde{B}} \|y_j - \bar{y}_{\widetilde{B}}\|^2.$$

Then,

$$\widetilde{RSS} - RSS = \frac{|B||C|}{|B| + |C|}\|\bar{y}_C - \bar{y}_B\|^2 - \frac{|A||C|}{|A| - |C|}\|\bar{y}_C - \bar{y}_A\|^2.$$

This lemma is proved by elementary calculation, which is relegated to Section 3.D.1. It shows that the change of k-means objective depends on the distances from $\bar{y}_C$ to two previous cluster centers.

We now apply Lemma 45 to prove the claim. For notation simplicity, we drop "Γ" and omit the superscript m, i.e., we write $r_i^{(m)}(\Gamma) = r_i$ and $v_k^{(m)}(\Gamma) = v_k$. By Lemma 32 and the condition (3.2.2),

- The n points $r_1, r_2, \ldots, r_n$ take K distinct values, $v_1, \ldots, v_K$.
- The minimum pairwise distance of $v_1, v_2, \ldots, v_K$ is defined as $d_K(V) > 0$.
- For each v_k, there are at least $a_0 n$ points, corresponding to nodes in community $\mathcal{N}_k$, that are equal to v_k, where $a_0 > 0$ is a constant determined by condition (3.2.2).

First, we show that any optimal solution of the k-means clustering on $\{r_1, r_2, \ldots, r_n\}$ satisfies NSP. We prove by contradiction. If this is not true, there must exist a community $\mathcal{N}_k$ and two clusters, say, S_1 and S_2, such that $\mathcal{N}_k \cap S_1 \neq \emptyset$ and $\mathcal{N}_k \cap S_2 \neq \emptyset$. Note that we have either $S_1 \setminus \mathcal{N}_k \neq \emptyset$ or $S_2 \setminus \mathcal{N}_k \neq \emptyset$ (if both S_1 and S_2 are contained in $\mathcal{N}_k$, then we can combine these two clusters and construct another cluster assignment with a smaller residual sum of squares, which conflicts with the optimality of the solution). Without loss of generality, we assume $S_1 \setminus \mathcal{N}_k \neq \emptyset$. We now move an arbitrary $r_i \in \mathcal{N}_k \cap S_1$ to S_2 and update the cluster centers (i.e., within-cluster means) accordingly. Let RSS and $\widetilde{RSS}$ be the respective k-means objective before and after the change. We apply Lemma 45 to get that

$$\widetilde{RSS} - RSS = \frac{|S_2|}{|S_2| + 1}\|r_i - c_2\|^2 - \frac{|S_1|}{|S_1| - 1}\|r_i - c_1\|^2. \tag{3.B.74}$$

Since i is clustered to S_1 in the optimal solution, it must be true that $\|r_i - c_1\| \leq \|r_i - c_2\|$, which further implies that $\|v_k - c_1\| \leq \|v_k - c_2\|$. At the same time, if we take any $j \in \mathcal{N}_k \cap S_2$, we can similarly derive that $\|v_k - c_2\| \leq \|v_k - c_1\|$. Combining the above gives $\|v_k - c_1\| = \|v_k - c_2\|$. It follows that

$$\|r_i - c_1\| = \|r_i - c_2\|.$$

We immediately see that

$$\widetilde{RSS} - RSS = \Big(\frac{|S_2|}{|S_2|+1} - \frac{|S_1|}{|S_1|-1}\Big)\|r_i - c_1\|^2 = -\frac{|S_1|+|S_2|}{(|S_2|+1)(|S_1|-1)}\|r_i - c_1\|^2.$$

The optimality of k-means solutions ensures that $\widetilde{RSS} - RSS \geq 0$. Therefore, the above equality is possible only if $\|r_i - c_1\| = 0$. However, $\|r_i - c_1\| = 0$ implies $c_1 = c_2$, which is impossible.

Second, define $g(r_i; c_1, c_2, \ldots, c_m) \equiv d_2(r_i; c_1, \ldots, c_m) - d_1(r_i; c_1, \ldots, c_m)$, which is the gap between the distances from r_i to the closest and second closest cluster centers. We aim to show that $g(r_i; c_1, c_2, \ldots, c_m)$ has a uniform lower bound for all $1 \leq i \leq n$. Fix i. Without loss of generality, we assume c_1 and c_2 are the cluster centers closest and second closest to r_i. Then, i is clustered to S_1. Suppose $i \in \mathcal{N}_k$. The NSP we proved above implies that

$$\mathcal{N}_k \subset S_1.$$

Again, by NSP, there are only two possible cases: (a) $S_1 = \mathcal{N}_k$, and (b) S_1 is the union of $\mathcal{N}_k$ and some other true communities.

In case (a), we immediately have $c_1 = v_k$. It follows that

$$\|r_i - c_1\| = \|v_k - c_1\| = 0.$$

Furthermore, for any $j \in S_2$, r_j equals to some v_ℓ that is distinct from v_k. Therefore,

$$\|r_i - c_2\| = \|v_k - c_2\| \geq \min_{j \in S_2} \|v_k - r_j\| \geq \min_{\ell \neq k} \|v_k - v_\ell\| = d_K(V).$$

As a result,

$$g(r_i; c_1, c_2, \ldots, c_m) = \|r_i - c_2\| - \|r_i - c_1\| = \|r_i - c_2\| \geq d_K(V).$$

The proves the claim in case (a).

In case (b), we consider moving $\mathcal{N}_k$ from S_1 to S_2, and let RSS and $\widetilde{RSS}$ denote the respective k-means objective before and after the change. Applying Lemma 45, we obtain

$$\widetilde{RSS} - RSS = \frac{|S_2||\mathcal{N}_k|}{|S_2| + |\mathcal{N}_k|}\|v_k - c_2\|^2 - \frac{|S_1||\mathcal{N}_k|}{|S_1| - |\mathcal{N}_k|}\|v_k - c_1\|^2. \tag{3.B.75}$$

Let $\Delta = \|v_k - c_2\|^2 - \|v_k - c_1\|^2$. By direct calculations,

$$\widetilde{RSS} - RSS = \frac{|S_2||\mathcal{N}_k|}{|S_2| + |\mathcal{N}_k|}\Delta - \frac{|\mathcal{N}_k|^2(|S_1| + |S_2|)}{(|S_2| + |\mathcal{N}_k|)(|S_1| - |\mathcal{N}_k|)}\|v_k - c_1\|^2.$$

The optimality of k-means solutions implies that $\widetilde{RSS} \geq RSS$. It follows that

$$\Delta \geq \frac{|\mathcal{N}_k|(|S_1| + |S_2|)}{|S_2|(|S_1| - |\mathcal{N}_k|)}\|v_k - c_1\|^2.$$

Note that $|\mathcal{N}_k| \geq a_0 n$, $|S_1| - |\mathcal{N}_k| \leq n$, and $\frac{|S_1|+|S_2|}{|S_2|} \geq 1$. It is seen that $\frac{|\mathcal{N}_k|(|S_1|+|S_2|)}{|S_2|(|S_1|-|\mathcal{N}_k|)} \geq a_0$. As a result,

$$\|v_k - c_2\|^2 - \|v_k - c_1\|^2 = \Delta \geq a_0\|v_k - c_1\|^2.$$

It implies that $\|v_k - c_2\|^2 \geq (1+a_0)\|v_k - c_1\|^2$, i.e.,

$$\|v_k - c_2\| - \|v_k - c_1\| \geq \big(\sqrt{1+a_0} - 1\big)\,\|v_k - c_1\|. \tag{3.B.76}$$

We then derive a lower bound on $\|v_k - c_1\|$. Here, c_1 is the mean of r_i's in S_1. For any $j \in S_1 \backslash \mathcal{N}_k$, r_j equals to some v_ℓ that is distinct from v_k. As a result, $\|v_k - r_j\| \geq \min_{\ell \neq k} \|v_k - v_\ell\| \geq d_K(V)$, for all $j \in S_1 \backslash \mathcal{N}_k$. It follows that

$$\begin{aligned}
\|v_k - c_1\| &= \Big\| v_k - \Big(\frac{|\mathcal{N}_k|}{|S_1|} v_k + \frac{1}{|S_1|} \sum_{j \in S_1 \backslash \mathcal{N}_k} r_j \Big) \Big\| \\
&= \Big\| \frac{1}{|S_1|} \sum_{j \in S_1 \backslash \mathcal{N}_k} (r_j - v_k) \Big\| \\
&= \frac{|S_1 \backslash \mathcal{N}_k|}{|S_1|} \, \Big\| \Big(\frac{1}{|S_1 \backslash \mathcal{N}_k|} \sum_{j \in S_1 \backslash \mathcal{N}_k} r_j \Big) - v_k \Big\| \\
&\geq \frac{|S_1 \backslash \mathcal{N}_k|}{|S_1|} \min_{j \in S_1 \backslash \mathcal{N}_k} \|r_j - v_k\| \\
&\geq a_0 \cdot d_K(V),
\end{aligned} \tag{3.B.77}$$

where in the last inequality we have used $|S_1| \leq n$ and $|S_1 \backslash \mathcal{N}_k| \geq a_0 n$ (because S_1 is the union of $\mathcal{N}_k$ and at least one other community). Combing (3.B.76) and (3.B.77) gives

$$g(r_i; c_1, c_2, \ldots, c_m) \geq a_0 \big(\sqrt{1+a_0} - 1\big)\, d_K(V).$$

This proves the claim in case (b). $\square$

3.B.6 Proof of Theorem 3.3.1

Write for short $d_m = d_m(U)$ and $\delta = \max_{1 \leq i \leq n} \|\widehat{x}_i - x_i\|$. Given any partition $\{1, 2, \ldots, n\} = \cup_{k=1}^m B_k$ and vectors $b_1, b_2, \ldots, b_m \in \mathbb{R}^d$, define

$$R(B_1, \ldots, B_m; b_1, \ldots, b_m) = n^{-1} \sum_{k=1}^m \sum_{i \in B_k} \|x_i - b_k\|^2. \tag{3.B.78}$$

Fixing $B_1, \ldots, B_m$, the value of $R(B_1, \ldots, B_m; b_1, \ldots, b_m)$ is minimized when b_k is the average of x_i's within each B_k. When $b_1, \ldots, b_m$ take these special values, we skip them in the notation. Namely, define

$$R(B_1, \ldots, B_m) = R(B_1, \ldots, B_m; \underline{x}_1, \ldots, \underline{x}_m), \quad \text{where } \underline{x}_k = |B_k|^{-1} \sum_{i \in B_k} x_i, \tag{3.B.79}$$

We define $\widehat{R}(B_1, \ldots, B_m; b_1, \ldots, b_m)$ and $\widehat{R}(B_1, \ldots, B_m)$ similarly but replace x_i by $\widehat{x}_i$. We shall prove the claim by contradiction. Suppose there is $1 \leq k \leq K$ such that F_k intersects with more than one $\widehat{S}_j$. By pigeonhole principle, there exists j_1, such that $|F_k \cap \widehat{S}_{j_1}| \geq m^{-1}|F_k|$. Let $\widehat{S}_{j_2}$ be another cluster that intersects with F_k. We have

$$|F_k \cap \widehat{S}_{j_1}| \geq m^{-1}\alpha_0 n, \qquad F_k \cap \widehat{S}_{j_2} \neq \emptyset,$$

Below, we aim to show: There exists $C_1 = C_1(\alpha_0, C_0, m)$ such that

$$\min_{\widetilde{S}_1, \ldots, \widetilde{S}_m} R(\widetilde{S}_1, \ldots, \widetilde{S}_m) \geq R(\widehat{S}_1, \ldots, \widehat{S}_m) - C_1 \delta \cdot d_m, \tag{3.B.80}$$

where the minimum on the left hand side is taken over possible partitions of $\{1, 2, \ldots, n\}$ into m clusters. We also aim to show that there exists $C_2 = C_2(\alpha_0, C_0, m)$ such that we can construct a clustering structure $\widetilde{S}_1, \widetilde{S}_2, \ldots, \widetilde{S}_m$ satisfying that

$$R(\widetilde{S}_1, \ldots, \widetilde{S}_m) \leq R(\widehat{S}_1, \ldots, \widehat{S}_m) - C_2 \cdot d_m^2. \tag{3.B.81}$$

Combining (3.B.80)-(3.B.81) gives

$$R(\widehat{S}_1, \ldots, \widehat{S}_m) - C_1\delta \cdot d_m \leq R(\widehat{S}_1, \ldots, \widehat{S}_m) - C_2 \cdot d_m^2$$

This is impossible if $C_1\delta \cdot d_m < C_2 \cdot d_m^2$. Hence, we can take

$$c(\alpha_0, C_0, m) < C_2/C_1.$$

There is a contradiction between (3.B.80) and (3.B.81) whenever $\delta \leq c \cdot d_m$. The claim follows.

It remains to prove (3.B.80) and (3.B.81). Consider (3.B.80). For an arbitrary cluster structure $B_1, B_2, \ldots, B_m$, let $\widehat{R}(B_1, \ldots, B_m)$, $R(B_1, \ldots, B_m)$, $\underline{\widehat{x}}_k$ and $\underline{x}_k$ be defined as in (3.B.79). By direct calculations,

$$(\widehat{x}_i - \underline{\widehat{x}}_k) - (x_i - \underline{x}_k) = \frac{|B_k| - 1}{|B_k|}(\widehat{x}_i - x_i) - \frac{1}{|B_k|} \sum_{j \in B_k : j \neq i} (\widehat{x}_j - x_j).$$

Since $\|\widehat{x}_j - x_j\| \leq \delta$ for all $1 \leq j \leq n$, the above equality implies that $\|(\widehat{x}_i - \underline{\widehat{x}}_k) - (x_i - \underline{x}_k)\| \leq \delta$. As a result, $\|\widehat{x}_i - \underline{\widehat{x}}_k\|^2 \leq \|x_i - \underline{x}_k\|^2 + 2\delta\|x_i - \underline{x}_k\| + \delta^2$. It follows that

$$\begin{aligned}
\widehat{R}(B_1, \ldots, B_m) &\leq R(B_1, \ldots, B_m) + 2\delta n^{-1} \sum_{k=1}^{m} \sum_{i \in B_k} \|x_i - \underline{x}_k\| + \delta^2 \\
&\leq R(B_1, \ldots, B_m) + 2\delta\sqrt{R(B_1, \ldots, B_m)} + \delta^2 \\
&\leq \left(\sqrt{R(B_1, \ldots, B_m)} + \delta\right)^2,
\end{aligned}$$

where the second line is from the Cauchy-Schwarz inequality. It follows that $\sqrt{\widehat{R}(B_1, \ldots, B_m)} \leq \sqrt{R(B_1, \ldots, B_m)} + \delta$. We can switch $\widehat{R}(B_1, \ldots, B_m)$ and $R(B_1, \ldots, B_m)$ to get a similar inequality. Combining them gives

$$\sqrt{R(B_1, \ldots, B_m)} - \delta \leq \sqrt{\widehat{R}(B_1, \ldots, B_m)} \leq \sqrt{R(B_1, \ldots, B_m)} + \delta. \tag{3.B.82}$$

This inequality holds for an arbitrary partition $(B_1, B_2, \ldots, B_m)$. We now apply it to $(\widehat{S}_1, \ldots, \widehat{S}_m)$, which are the clusters obtained from applying k-means on $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n$. We also consider applying k-means on $x_1, x_2, \ldots, x_n$ and let $S_1, S_2, \ldots, S_m$ denote the resultant clusters. By optimality of the k-means solutions,

$$\widehat{R}(\widehat{S}_1, \ldots, \widehat{S}_m) \leq \widehat{R}(S_1, \ldots, S_m).$$

Combining it with (3.B.82) gives

$$\begin{aligned}
\sqrt{R(\widehat{S}_1, \ldots, \widehat{S}_m)} &\leq \sqrt{\widehat{R}(\widehat{S}_1, \ldots, \widehat{S}_m)} + \delta \\
&\leq \sqrt{\widehat{R}(S_1, \ldots, S_m)} + \delta \\
&\leq \sqrt{R(S_1, \ldots, S_m)} + 2\delta.
\end{aligned} \tag{3.B.83}$$

Since $\max_{1 \leq i \leq n} \|x_i\| \leq C_0 \cdot d_m$, we can easily see that $R(S_1, \ldots, S_m) \leq C_0^2 \cdot d_m^2$. It follows that, as long as $\delta \leq d_m/4$,

$$\begin{aligned}R(\widehat{S}_1, \ldots, \widehat{S}_m) &\leq R(S_1, \ldots, S_m) + 4\delta\sqrt{R(S_1, \ldots, S_m)} + 4\delta^2 \\ &\leq R(S_1, \ldots, S_m) + 4C_0\delta \cdot d_m + \delta \cdot d_m \\ &\leq R(S_1, \ldots, S_m) + (4C_0 + 1)\delta \cdot d_m.\end{aligned}$$

As a result,

$$\min_{\widetilde{S}_1, \ldots, \widetilde{S}_m} R(\widetilde{S}_1, \ldots, \widetilde{S}_m) = R(S_1, \ldots, S_m) \geq R(\widehat{S}_1, \ldots, \widehat{S}_m) - (4C_0 + 1)\delta \cdot d_m.$$

This proves (3.B.80) for $C_1 = 4(C_0 + 1)$.

Consider (3.B.81). Define

$$w_j = |\widehat{S}_j|^{-1} \sum_{i \in \widehat{S}_j} x_i, \qquad \text{for each } 1 \leq j \leq m. \tag{3.B.84}$$

Using the notations in (3.B.78)-(3.B.79), we write $R(\widehat{S}_1, \ldots, \widehat{S}_m) = R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m)$. We aim to construct $\{(\widetilde{S}_j, \widetilde{w}_j)\}_{1 \leq j \leq m}$ such that

$$R(\widetilde{S}_1, \ldots, \widetilde{S}_m, \widetilde{w}_1, \ldots, \widetilde{w}_m) \leq R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m) - C_2 \cdot d_m^2. \tag{3.B.85}$$

Since $R(\widetilde{S}_1, \ldots, \widetilde{S}_m) = \min_{b_1, \ldots, b_m} R(\widetilde{S}_1, \ldots, \widetilde{S}_m, b_1, \ldots, b_m)$, we immediately have

$$R(\widetilde{S}_1, \ldots, \widetilde{S}_m) \leq R(\widetilde{S}_1, \ldots, \widetilde{S}_m, \widetilde{w}_1, \ldots, \widetilde{w}_m) \leq R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m) - C_2 \cdot d_m^2.$$

This proves (3.B.81).

What remains is to construct $\{(\widetilde{S}_j, \widetilde{w}_j)\}_{j=1}^m$ so that (3.B.85) is satisfied. Let $\widehat{w}_j = |\widehat{S}_j|^{-1} \sum_{i \in \widehat{S}_j} \widehat{x}_i$, for $1 \leq j \leq m$. Then, $\{(\widehat{S}_j, \widehat{w}_j)\}_{1 \leq j \leq m}$ are the clusters and cluster centers obtained by applying the k-means algorithm on $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n$. The k-means solution guarantees to assign each point to the closest center. Take $i \in F_k \cap \widehat{S}_{j_1}$ and $i' \in F_k \cap \widehat{S}_{j_2}$. It follows that

$$\|\widehat{x}_i - \widehat{w}_{j_1}\| \leq \|\widehat{x}_i - \widehat{w}_{j_2}\|, \qquad \|\widehat{x}_{i'} - \widehat{w}_{j_2}\| \leq \|\widehat{x}_{i'} - \widehat{w}_{j_1}\|.$$

Since $x_i = x_{i'} = u_k$ and $\max\{\|\widehat{x}_i - x_i\|, \|\widehat{x}_{i'} - x_{i'}\|, \|\widehat{w}_{j_1} - w_{j_1}\|, \|\widehat{w}_{j_2} - w_{j_2}\|\} \leq \delta$, we have

$$\|u_k - w_{j_1}\| \leq \|\widehat{x}_i - \widehat{w}_{j_1}\| + 2\delta \leq \|\widehat{x}_i - \widehat{w}_{j_2}\| + 2\delta \leq \|u_k - w_{j_2}\| + 4\delta.$$

Similarly, we can derive that $\|u_k - w_{j_2}\| \leq \|u_k - w_{j_1}\| + 4\delta$. Combining them gives

$$\big|\|u_k - w_{j_1}\| - \|u_k - w_{j_2}\|\big| \leq 4\delta. \tag{3.B.86}$$

This inequality tells us that $\|u_k - w_{j_1}\|$ and $\|u_k - w_{j_2}\|$ are sufficiently close. Introduce

$$C_3 = \frac{m^{-1}\alpha_0}{36 \times 12C_0}.$$

Below, we consider two cases: $\|u_k - w_{j_1}\| < C_3 \cdot d_m$ and $\|u_k - w_{j_1}\| \geq C_3 \cdot d_m$.

In the first case, $\|u_k - w_{j_1}\| < C_3 \cdot d_m$. The definition of d_m guarantees that there are m points from $\{u_1, u_2, \ldots, u_K\}$ such that their minimum pairwise distance is d_m. Without loss of generality, we assume these m points are $u_1, u_2, \ldots, u_m$. If $k \in \{1, 2, \ldots, m\}$, then the distance from u_k to any of the other $(m-1)$ points is at least d_m. If $k \notin \{1, 2, \ldots, m\}$, then u_k cannot be simultaneously within a distance of $< d_m/2$ to two or more points of $u_1, u_2, \ldots, u_m$. In other words, there exists at least $(m-1)$ points from $u_1, u_2, \ldots, u_m$ whose distance to u_k is at least $\geq d_m/2$. Combining the above situations, we conclude that there exist $(m-1)$ points from $\{u_1, u_2, \ldots, u_K\}$, which we assume to be $u_1, u_2, \ldots, u_{m-1}$ without loss of generality, such that

$$\min_{1\le \ell\ne s\le m-1}\|u_\ell-u_s\|\ge d_m,\qquad \min_{1\le \ell\le m-1}\|u_\ell-u_k\|\ge d_m/2. \tag{3.B.87}$$

We then consider two sub-cases. In the first sub-case, there exists $\ell\in\{1,2,\dots,m-1\}$ such that $|F_\ell\cap(\widehat{S}_{j_1}\cup\widehat{S}_{j_2})|\ge m^{-1}\alpha_0 n$. Then, at least one of $\widehat{S}_{j_1}$ and $\widehat{S}_{j_2}$ contains more than $(m^{-1}\alpha_0/2)n$ nodes from F_ℓ. We only study the situation of $|F_\ell\cap\widehat{S}_{j_2}|\ge(m^{-1}\alpha_0/2)n$. The proof for the situation of $|F_\ell\cap\widehat{S}_{j_1}|\ge(m^{-1}\alpha_0/2)n$ is similar and omitted. We modify the clusters and cluster centers $\{(\widehat{S}_j,w_j)\}_{1\le j\le m}$ as follows:

(i) Combine $\widehat{S}_{j_2}\backslash F_\ell$ and $\widehat{S}_{j_1}$ into one cluster and set the cluster center to be w_{j_1}.

(ii) Create a new cluster as $\widehat{S}_{j_2}\cap F_\ell$ and set the cluster center to be u_ℓ.

The other clusters and cluster centers remain unchanged. Namely, we let

$$\widetilde{S}_j=\begin{cases}\widehat{S}_{j_1}\cup(\widehat{S}_{j_2}\backslash F_\ell), & \text{if } j=j_1,\\ \widehat{S}_{j_2}\cap F_\ell, & \text{if } j=j_2,\\ \widehat{S}_j, & \text{if } j\notin\{j_1,j_2\},\end{cases}\qquad \widetilde{w}_j=\begin{cases}u_\ell, & \text{if } j=j_2,\\ w_j, & \text{otherwise.}\end{cases}$$

Recall that $n\cdot R(B_1,\dots,B_m,b_1,\dots,b_m)=\sum_{j=1}^m\sum_{i\in B_j}\|x_i-b_j\|^2$. By direct calculations,

$$\begin{aligned}\Delta &\equiv n\cdot R(\widehat{S}_1,\dots,\widehat{S}_m,w_1,\dots,w_m)-n\cdot R(\widetilde{S}_1,\dots,\widetilde{S}_m,\widetilde{w}_1,\dots,\widetilde{w}_m)\\ &=\sum_{i\in(\widehat{S}_{j_2}\cap F_\ell)}\big(\|x_i-w_{j_2}\|^2-\|x_i-u_\ell\|^2\big)-\sum_{i\in(\widehat{S}_{j_2}\backslash F_\ell)}\big(\|x_i-w_{j_1}\|^2-\|x_i-w_{j_2}\|^2\big)\\ &\equiv\Delta_2-\Delta_1.\end{aligned}$$

Here Δ_1 is the increase of the residual sum of squares (RSS) caused by the operation (i) and Δ_2 is the decrease of RSS caused by the operation (ii).

$$\begin{aligned}\Delta_1&=\sum_{i\in(\widehat{S}_{j_2}\backslash F_\ell)}(\|x_i-w_{j_1}\|-\|x_i-w_{j_2}\|)(\|x_i-w_{j_1}\|+\|x_i-w_{j_2}\|)\\ &\le\sum_{i\in(\widehat{S}_{j_2}\backslash F_\ell)}\|w_{j_1}-w_{j_2}\|\cdot(\|x_i-w_{j_1}\|+\|x_i-w_{j_2}\|)\\ &\le|\widehat{S}_{j_2}\backslash F_\ell|\cdot\|w_{j_1}-w_{j_2}\|\cdot 4C_0\cdot d_m,\end{aligned}$$

where the third line is from the triangle inequality and the last line is because $\max_{1\le j\le m}\|w_j\|\le\max_{1\le i\le n}\|x_i\|\le C_0\cdot d_m$. Note that $\|w_{j_1}-w_{j_2}\|\le\|u_k-w_{j_1}\|+\|u_k-w_{j_2}\|$. We have assumed $\|u_k-w_{j_1}\|<C_3\cdot d_m$ in this case. Combing it with (3.B.86), as long as $\delta<(C_3/4)\cdot d_m$,

$$\|w_{j_1}-w_{j_2}\|\le 2\|u_k-w_{j_1}\|+4\delta\le 3C_3\cdot d_m.$$

It follows that

$$\Delta_1\le 12C_0C_3\cdot nd_m^2. \tag{3.B.88}$$

Since $x_i=u_\ell$ for $i\in F_\ell$, we immediately have

$$\Delta_2=|\widehat{S}_{j_2}\cap F_\ell|\cdot\|u_\ell-w_{j_2}\|^2.$$

We have assumed $\|u_k - w_{j_1}\| \le C_3 \cdot d_m$ in this case. Combining it with (3.B.86) and (3.B.87) gives

$$\begin{aligned}\|u_\ell - w_{j_2}\| &\ge \|u_\ell - u_k\| - \|u_k - w_{j_2}\| \\ &\ge \|u_\ell - u_k\| - \big(\|u_k - w_{j_1}\| + 4\delta\big) \\ &\ge d_m/2 - (C_3 \cdot d_m + 4\delta).\end{aligned}$$

Recall that $C_3 = \frac{m^{-1}\alpha_0}{36\times 12C_0} < 1/12$. Then, as long as $\delta < (1/48)d_m$, we have $\|u_\ell - w_{j_2}\| \ge d_m/3$. It follows that

$$\Delta_2 \ge (m^{-1}\alpha_0/2)n \cdot (d_m/3)^2 \ge \frac{m^{-1}\alpha_0}{18} \cdot nd_m^2. \tag{3.B.89}$$

As a result,

$$\Delta = \Delta_2 - \Delta_1 \ge \Big(\frac{m^{-1}\alpha_0}{18} - 12C_0C_3\Big) \cdot nd_m^2.$$

We plug in the expression of C_3, the right hand side is $(m^{-1}\alpha_0/36) \cdot nd_m^2$. It follows that

$$R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m) - R(\widetilde{S}_1, \ldots, \widetilde{S}_m, \widetilde{w}_1, \ldots, \widetilde{w}_m) \ge \frac{m^{-1}\alpha_0}{36} \cdot d_m^2. \tag{3.B.90}$$

This gives (3.B.85) in the first sub-case.

In the second sub-case, $|F_\ell \cap (\widehat{S}_{j_1} \cup \widehat{S}_{j_2})| < m^{-1}\alpha_0 n$ for all $1 \le \ell \le m-1$. For each F_ℓ, by pigeonhole principle, there exists at least one $j \in \{1, 2, \ldots, m\}$ such that $|F_\ell \cap \widehat{S}_j| \ge m^{-1}|F_\ell| \ge m^{-1}\alpha_0 n$. Denote such a j by j_ℓ^*; if there are multiple indices satisfying the requirement, we pick one of them. This gives

$$j_1^*, j_2^*, \ldots, j_{m-1}^* \quad \in \quad \{1, 2, \ldots, m\}\backslash\{j_1, j_2\}.$$

These $(m-1)$ indices take at most $(m-2)$ distinct values. By pigeonhole principle, there exist $1 \le \ell_1 \ne \ell_2 \le m-1$ such that $j_{\ell_1}^* = j_{\ell_2}^* = j^*$, for some $j^* \notin \{j_1, j_2\}$. Recalling (3.B.84), we let w_{j^*} denote the average of x_i's in $\widehat{S}_{j^*}$. Since $\|u_{\ell_1} - u_{\ell_2}\| \ge d_m$, the point w_{j^*} cannot be simultaneously within a distance of $d_m/2$ to both u_{ℓ_1} and u_{ℓ_2}. Without loss of generality, suppose

$$\|u_{\ell_1} - w_{j^*}\| \ge d_m/2.$$

We modify the clusters and cluster centers $\{(\widehat{S}_j, w_j)\}_{1\le j\le m}$ as follows:

(i) Combine $\widehat{S}_{j_1}$ and $\widehat{S}_{j_2}$ into one cluster and set the cluster center to be w_{j_1}.

(ii) Split $\widehat{S}_{j^*}$ into two clusters, where one is $(\widehat{S}_{j^*} \cap F_{\ell_1})$, and the other is $(\widehat{S}_{j^*}\backslash F_{\ell_1})$; the two cluster centers are set as u_{ℓ_1} and w_{j^*}, respectively.

The other clusters and cluster centers remain unchanged. Namely, we let

$$\widetilde{S}_j = \begin{cases} \widehat{S}_{j_1} \cup \widehat{S}_{j_2}, & \text{if } j = j_1, \\ \widehat{S}_{j^*} \cap F_{\ell_1}, & \text{if } j = j_2, \\ \widehat{S}_{j^*}\backslash F_{\ell_1}, & \text{if } j = j^*, \\ \widehat{S}_j, & \text{if } j \notin \{j_1, j_2, j^*\}, \end{cases} \qquad \widetilde{w}_j = \begin{cases} u_{\ell_1}, & \text{if } j = j_2, \\ w_j, & \text{otherwise.} \end{cases}$$

By direct calculations,

$$\begin{aligned}\Delta &\equiv n\cdot R(\widehat{S}_1,\ldots,\widehat{S}_m,w_1,\ldots,w_m) - n\cdot R(\widetilde{S}_1,\ldots,\widetilde{S}_m,\widetilde{w}_1,\ldots,\widetilde{w}_m)\\ &= \sum_{i\in(\widehat{S}_{j^*}\cap F_{\ell_1})}\left(\|x_i-w_{j^*}\|^2-\|x_i-u_{\ell_1}\|^2\right) - \sum_{i\in\widehat{S}_{j_2}}\left(\|x_i-w_{j_1}\|^2-\|x_i-w_{j_2}\|^2\right)\\ &\equiv \Delta_2-\Delta_1,\end{aligned}$$

where Δ_1 is the increase of RSS caused by (i) and Δ_2 is the decrease of RSS caused by (ii). We can bound Δ_1 in a similar way as in the previous sub-case, and the details are omitted. It gives

$$\Delta_1 \leq 12C_0C_3\cdot nd_m^2.$$

Since $x_i = u_{\ell_1}$ for all $i\in F_{\ell_1}$, we immediately have

$$\Delta_2 = |\widehat{S}_{j^*}\cap F_{\ell_1}|\cdot\|u_{\ell_1}-w_{j^*}\|^2 \geq (m^{-1}\alpha_0 n)\cdot(d_m/2)^2 \geq \frac{m^{-1}\alpha_0}{4}\cdot nd_m^2.$$

As a result, $\Delta \geq (\frac{m^{-1}\alpha_0}{4} - 12C_0C_3)m^{-1}\alpha_0\cdot nd_m^2$. If we plug in the expression of C_3, it becomes $\geq (\frac{2}{9}m^{-1}\alpha_0)\cdot nd_m^2$. This gives

$$R(\widehat{S}_1,\ldots,\widehat{S}_m,w_1,\ldots,w_m) - R(\widetilde{S}_1,\ldots,\widetilde{S}_m,\widetilde{w}_1,\ldots,\widetilde{w}_m) \geq \frac{2m^{-1}\alpha_0}{9}\cdot d_m^2. \tag{3.B.91}$$

This gives (3.B.85) in the second sub-case.

In the second case, $\|u_k - w_{j_1}\| \geq C_3\cdot d_m$. We recall that $|F_k\cap\widehat{S}_{j_1}| \geq m^{-1}\alpha_0 n$. Let E be a subset of $F_k\cap\widehat{S}_{j_1}$ such that $|E| = \lceil |F_k\cap\widehat{S}_{j_1}|/2\rceil$. Note that $|\widehat{S}_{j_1}\backslash E|\leq n$. We have

$$\widehat{S}_{j_1}\backslash E\neq\emptyset, \qquad \text{and} \qquad \frac{|E|}{|\widehat{S}_{j_1}\backslash E|}\geq m^{-1}\alpha_0/2.$$

We now modify the clusters and cluster centers $\{(\widehat{S}_j,w_j)\}_{1\leq j\leq m}$ as follows:

- Move the subset E from $\widehat{S}_{j_1}$ to $\widehat{S}_{j_2}$, and update each cluster center to be the within cluster average of x_i's.

The other clusters and cluster centers are unchanged. Namely, we let

$$\widetilde{S}_j = \begin{cases}\widehat{S}_{j_1}\backslash E, & \text{if } j=j_1,\\ \widehat{S}_{j_2}\cup E, & \text{if } j=j_2,\\ \widehat{S}_j, & \text{if } j\notin\{j_1,j_2\},\end{cases} \qquad \widetilde{w}_j = \begin{cases}\frac{1}{|\widetilde{S}_j|}\sum_{i\in\widetilde{S}_j}x_i, & \text{if } j\in\{j_1,j_2\},\\ w_j, & \text{otherwise.}\end{cases}$$

We apply Lemma 45 to $A=\widehat{S}_{j_1}$, $B=\widehat{S}_{j_2}$, and $C=E$, and note that $x_i=u_k$ for all $i\in E$.

It follows that

$$
\begin{aligned}
\Delta &\equiv n \cdot R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m) - n \cdot R(\widetilde{S}_1, \ldots, \widetilde{S}_m, \widetilde{w}_1, \ldots, \widetilde{w}_m) \\
&= -\left(\frac{|\widehat{S}_{j_2}| \cdot |E|}{|\widehat{S}_{j_2}| + |E|} \|u_k - w_{j_2}\|^2 - \frac{|\widehat{S}_{j_1}| \cdot |E|}{|\widehat{S}_{j_1}| - |E|} \|u_k - w_{j_1}\|^2 \right) \\
&= \frac{|E|^2 \cdot (|\widehat{S}_{j_1}| + |\widehat{S}_{j_2}|)}{(|\widehat{S}_{j_2}| + |E|)(|\widehat{S}_{j_1}| - |E|)} \|u_k - w_{j_1}\|^2 + \frac{|\widehat{S}_{j_2}| \cdot |E|}{|\widehat{S}_{j_2}| + |E|} \big(\|u_k - w_{j_1}\|^2 - \|u_k - w_{j_2}\|^2\big) \\
&\geq \frac{|E|^2}{|\widehat{S}_{j_1}| - |E|} \|u_k - w_{j_1}\|^2 + \frac{|\widehat{S}_{j_2}| \cdot |E|}{|\widehat{S}_{j_2}| + |E|} \big(\|u_k - w_{j_1}\|^2 - \|u_k - w_{j_2}\|^2\big). \qquad (3.B.92)
\end{aligned}
$$

By (3.B.86), $\|u_k - w_{j_2}\| \leq \|u_k - w_{j_1}\| + 4\delta$. It follows that, as long as $\delta < (C_3/16) \cdot d_m$,

$$
\begin{aligned}
\|u_k - w_{j_1}\|^2 - \|u_k - w_{j_2}\|^2 &\geq -8\delta \cdot \|u_k - w_{j_1}\| - 16\delta^2 \\
&\geq -9\delta \cdot \|u_k - w_{j_1}\|,
\end{aligned}
$$

where the last line is because $16\delta^2 \leq C_3\delta \cdot d_m \leq \delta \cdot \|u_k - w_{j_1}\|$. We plug it into (3.B.92) to get

$$
\begin{aligned}
\Delta &\geq \frac{|E|^2}{|\widehat{S}_{j_1} \backslash E|} \|u_k - w_{j_1}\|^2 - \frac{|\widehat{S}_{j_2}| \cdot |E|}{|\widehat{S}_{j_2}| + |E|} \cdot 9\delta \cdot \|u_k - w_{j_1}\| \\
&\geq |E| \cdot (m^{-1}\alpha_0/2) \cdot \|u_k - w_{j_1}\|^2 - |E| \cdot 9\delta \cdot \|u_k - w_{j_1}\| \\
&\geq |E| \cdot \|u_k - w_{j_1}\| \cdot \Big(\frac{C_3 m^{-1}\alpha_0}{2} d_m - 9\delta \Big),
\end{aligned}
$$

where the second line is because $|E| \geq (m^{-1}\alpha_0/2) \cdot |\widehat{S}_{j_1} \backslash E|$ and the last line is because we have assumed $\|u_k - w_{j_1}\| \geq C_3 \cdot d_m$ in the current case. As long as $\delta < \frac{C_3 m^{-1}\alpha_0}{27} \cdot d_m$, the number in brackets is $\geq \frac{C_3 m^{-1}\alpha_0}{6} d_m$. We also plug in $|E| = \lceil m^{-1}\alpha_0/2 \rceil n$ and $\|u_k - w_{j_1}\| \geq C_3 \cdot d_m$ to get

$$
\Delta \geq \frac{m^{-1}\alpha_0}{2} n \cdot C_3 d_m \cdot \frac{C_3 m^{-1}\alpha_0}{6} d_m \geq \frac{C_3^2 m^{-2}\alpha_0^2}{12} \cdot n d_m^2.
$$

It follows that

$$
R(\widehat{S}_1, \ldots, \widehat{S}_m, w_1, \ldots, w_m) - R(\widetilde{S}_1, \ldots, \widetilde{S}_m, \widetilde{w}_1, \ldots, \widetilde{w}_m) \geq \frac{C_3^2 m^{-2}\alpha_0^2}{12} \cdot d_m^2. \qquad (3.B.93)
$$

This gives (3.B.85) in the second case. We combine (3.B.90), (3.B.91) and (3.B.93), and take the minimum of the right hand sides of three inequalities. Since $m^{-1}\alpha_0 < 1$ and $C_3^2 < 1/3$, we choose

$$
C_2 = (1/12) C_3^2 m^{-2} \alpha_0^2.
$$

Then, (3.B.85) is satisfied for all cases. This completes the proof of (3.B.81).

We remark that the scalar $c = c(\alpha_0, C_0, m)$ is not exactly C_2/C_1. In the derivation of (3.B.80) and (3.B.81), we have imposed other restrictions on δ, which can be expressed as $\delta \leq C_4 \cdot d_m$, where C_4 is determined by (C_0, α, m) and (C_1, C_2, C_3). Since (C_1, C_2, C_3) only depend on (α_0, C_0, m), C_4 is a function of (α_0, C_0, m) only. We take $c = \min\{C_2/C_1,\ C_4\}$. $\square$

3.B.7 Proof of the claim in Example 4b of Section 3.3

In Example 4b of Section 3.3, we have the following claim.

Lemma 46. *Let $R^{(m)}$ and $V^{(m)}$ be as in (3.3.17) and (3.3.16), respectively. If $(K, m) = (4, 2)$ and all 4 communities have equal sizes, then $g_m(R^{(m)}) \geq [(3 - \sqrt{3})/2] d_K(V^{(m)})$.*

We now show the claim. For short, let $x_k = v_k^{(m)}$ for all $1 \leq k \leq 4$ and let $d_* = g_m(R^{(m)})$. Without loss of generality, we assume $x_1 = 0$, $x_2 = 1$, $x_3 = x$, and $x_4 = y$, where $y > x > 1$. Let $z = y - x$. It is seen $d_K(V^{(m)}) = \min\{1, x - 1, z\}$. To show the claim, it is sufficient to show

$$d_* \geq \frac{3 - \sqrt{3}}{2} \min\{1, x - 1, z\}. \tag{3.B.94}$$

By definitions,

$$d_* = \min_{\{\text{all possible } c_1, c_2\}} \min_{1 \leq i \leq 4} \{d_i(c_1, c_2)\}, \tag{3.B.95}$$

where for $1 \leq i \leq 4$, $d_i(c_1, c_2) \geq 0$ is the difference between the distance from x_i to the center of the cluster to which x_i does not belong and the distance from x_i to the center of the cluster to which x_i belongs. For simplicity, we write $d_i = d_i(c_1, c_2)$ when there is no confusion.

For the four points x_1, x_2, x_3, x_4, we have three possible candidates (a)-(c) for the clustering results (which of them is the actual clustering result depends on the values of (x, y)):

- (a). The left most point forms one cluster, the other three form the other cluster.
- (b). The left two points form one cluster, the other two points form the other cluster.
- (c). The left three points form one cluster, the right most point forms the other cluster.

Recall that for any n points $x_1, x_2, \ldots, x_n$, the RSS for the k-means solution with K clusters is

$$RSS = \sum_{k=1}^{K} \sum_{\{i \in \text{cluster } k\}} (x_i - c_k)^2,$$

where $c_1, c_2, \ldots, c_K$ are the cluster centers. For (a), the two cluster centers are $c_1 = 0$ and $c_2 = (1 + x + y)/3$. In this case, the RSS is $S_1 = x^2 + y^2 + 1 - (1/3)(x + y + 1)^2$. For (b), the two cluster centers are $c_1 = 1/2$ and $c_2 = (x + y)/2$, and the RSS is $S_2 = (1/2) + (1/2)(x - y)^2$. For (c), the two cluster centers are $c_1 = (1 + x)/3$ and $x_2 = y$, and the RSS is $S_3 = x^2 + 1 - (1/3)(x + 1)^2$. It is seen that the actual clustering result is as in (a) if and only if $S_1 \leq S_2$ and $S_1 \leq S_3$; similar for (b) and (c).

Recall that $z = y - x$. Consider the two-dimensional space with x and z being the two axes. As in Figure 3.7, we partition the region $\{(x, z) : x > 1, z > 0\}$ into three sub-regions as follows.

- Region (I). $\{(x, z) : 2x + z < 2 + \sqrt{3}, z < 1\}$.
- Region (II). $\{(x, z) : z < (2x - 1)/\sqrt{3}, 2x + z > 2 + \sqrt{3}\}$.

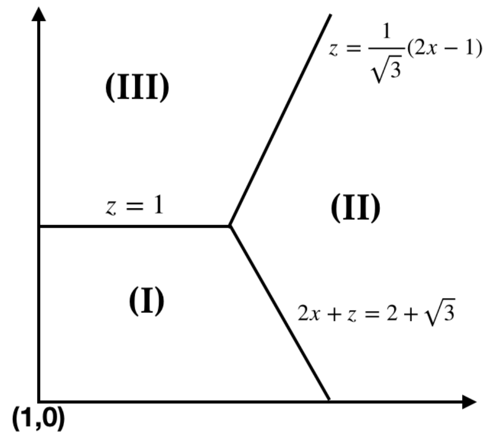

Figure 3.7: In the two dimensional space with x and z being the two axes, the whole region $\{(x, z) : x > 1, z > 0\}$ partitions into three sub-regions (I), (II), and (III), respectively.

- Region (III). $\{(x, z) : z > 1, z > (2x - 1)/\sqrt{3}\}$.

Note that any point (x, z) in our range of interest either belongs to one of the three regions, or falls on one of the boundaries of these regions. We now show the claim by consider the three regions in Figure 3.7 separately. The discussions for the case where (x, z) fall on the boundaries of these regions are similar so are omitted.

Consider Region (I). In this region, by elementary algebra, we have $S_1 < S_2$ and $S_1 < S_3$. Therefore, case (a) is the final clustering result, where the two clusters are $\{x_1\}$ and $\{x_2, x_3, x_4\}$, respectively, with cluster centers $c_1 = 0$ and $c_2 = (x+y+1)/3$. By definitions, for (x, z) in Region (I), $d_1 = |c_2 - 0| - |c_1 - 0| = (1+x+y)/3$, $d_2 = |c_1 - 1| - |c_2 - 1| = (5-x-y)/3$, $d_3 = |c_1 - x| - |c_2 - x| = (x + y + 1)/3$ if $2x > y + 1$ and $d_3 = (5x - y - 1)/3$ otherwise, and $d_4 = |c_1 - y| - |c_2 - y| = (x + y + 1)/3$. By elementary algebra, it is seen that d_2 is the smallest among $\{d_1, d_2, d_3, d_4\}$. Combining this with (3.B.95) gives that for (x, z) in Region (I), $d_* = (5 - x - y)/3 = (5 - 2x - z)/3$. Note that for (x, z) in Region (I), $2x + z < 2 + \sqrt{3}$. It follows $2(x - 1) + z < \sqrt{3}$ and so $\min\{1, x - 1, z\} \leq \sqrt{3}/3$. Combining these,

$$\frac{d_*}{\min\{1, x-1, z\}} \geq \frac{1}{3}\frac{5 - (2 + \sqrt{3})}{\sqrt{3}/3} \geq (\sqrt{3} - 1). \tag{3.B.96}$$

Consider Region (II). In this region, by elementary algebra, $S_2 \leq S_1$ and $S_2 < S_3$. Therefore, case (b) is the actual clustering result, so the two cluster centers are $c_1 = 1/2$ and $c_2 = (x + y)/2$, respectively. By definitions, $d_1 = |c_2 - 0| - |c_1 - 0| = (x + y - 1)/2$, $d_2 = |c_2 - 1| - |c_1 - 1| = (x + y - 3)/2$, $d_3 = |c_1 - x| - |c_2 - x| = (3x - y - 1)/2$, and $d_4 = |c_1 - y| - |c_2 - y| = (x + y - 1)/2$. By elementary algebra, among the four numbers $\{d_1, d_2, d_3, d_4\}$, d_2 is the smallest when $z < 1$ and d_3 is smallest when $z > 1$. Combining this with (3.B.95) gives that for (x, z) in Region (II),

$$d_* = \begin{cases} (x + y - 3)/2 = (2x + z - 3)/2, & \text{if } z < 1, \\ (3x - y - 1)/2 = (2x - z - 1)/2, & \text{if } z \geq 1. \end{cases}$$

Consider the case of $z < 1$ first. In this case, $\min\{1, x-1, z\} = \min\{x-1, z\} > 0$, and $2x + z - 3 > (2 - 2/\sqrt{3})(x-1) + (1 - 1/\sqrt{3})z$ since $2x + z > (2 + \sqrt{3})$ in Region (II). Therefore,

$$\frac{d_*}{\min\{1, x-1, z\}} = \frac{2x + z - 3}{2\min\{x-1, z\}} \geq \frac{(2 - 2/\sqrt{3})(x-1) + (1 - 1/\sqrt{3})z}{2\min\{x-1, z\}},$$

where the right hand side is no smaller than

$$[(2 - 2/\sqrt{3}) + (1 - 1/\sqrt{3})]/2 = (3 - \sqrt{3})/2.$$

Consider the case $z \geq 1$. In this case, $\min\{1, x-1, z\} = \min\{x-1, 1\} > 0$, and $(2x - z - 1) \geq (2 - 2/\sqrt{3})(x-1) + (1 - 1/\sqrt{3})$ since $z \leq (2x-1)/\sqrt{3}$. Therefore,

$$\frac{d_*}{\min\{1, x-1, z\}} = \frac{2x - z - 1}{2\min\{x-1, 1\}} \geq \frac{(2 - 2/\sqrt{3})(x-1) + (1 - 1/\sqrt{3})}{2\min\{x-1, 1\}},$$

where the right hand side is no smaller than

$$[(2 - 2/\sqrt{3}) + (1 - 1/\sqrt{3})]/2 = (3 - \sqrt{3})/2.$$

Combining the above, we have that in Region (II),

$$d_* \geq \frac{(3 - \sqrt{3})}{2} \min\{1, x-1, z\}. \tag{3.B.97}$$

Consider Region (III). By elementary algebra, it is seen $S_3 < S_1$ and $S_3 < S_2$ in this case. Therefore, case (c) is the actual clustering result, so the two cluster centers are $c_1 = (1+x)/3$ and $c_2 = y$, respectively. By definitions, $d_1 = |c_2 - 0| - |c_1 - 0| = (3y - x - 1)/3$, $d_2 = |c_2 - 1| - |c_1 - 1| = (-1 - x + 3y)/3$ if $x > 2$ and $d_2 = (x + 3y - 5)/3$ elsewise, $d_3 = |c_1 - x| - |c_2 - x| = (1 - 5x + 3y)/3$, and $d_4 = |c_1 - y| - |c_2 - y| = (-1 - x + 3y)/3$. By elementary algebra, d_3 is the smallest in $\{d_1, d_2, d_3, d_4\}$. Combining these with (3.B.95) gives that for (x, z) in Region (III),

$$d_* = (1 - 5x + 3y)/3 = (1 - 2x + 3z)/3, \qquad \min\{1, x-1, z\} = \min\{1, x-1\}.$$

When $x > 2$, $\min\{1, x-1\} = 1$, and the minimum of d_* in Region (III) is $(\sqrt{3} - 1)$ attained at $(x, z) = (2, \sqrt{3})$. When $x < 2$, $\min\{1, x-1\} = x - 1$. Therefore, $d_*/\min\{1, x-1\} = (z - 1/3)/(x-1) - 2/3$, where the minimum in Region (III) is $2/\sqrt{3}$, attained at $(x, z) = ((\sqrt{3}+1)/2, 1)$. Combining these, we have that for (x, z) in Region (III),

$$d_* \geq (\sqrt{3} - 1)\min\{1, x-1, z\}. \tag{3.B.98}$$

Combining (3.B.96)-(3.B.98) gives the claim. □

3.C Proof of results in Section 3.4

3.C.1 Proof of Lemma 36

Consider the first two claims. It is easy to see that $\mathbb{E}[C_n] = c_n$. In the proof of Theorem 3.1 of Jin et al. (2018), it has been shown that

$$c_n = \mathrm{tr}(\Omega^4) + O(\|\theta\|_4^4\|\theta\|^4) = \mathrm{tr}(\Omega^4) + o(\|\theta\|^8).$$

Moreover, $\lambda_1^4 \leq \mathrm{tr}(\Omega^4) \leq K\lambda_1^4$. In the proof of Theorem 3.2.4, we have seen that $\lambda_1 = \|\theta\|^2 \cdot \lambda_1(HPH')$. Using the condition (3.2.2) and the fact that P has unit diagonals, we have $\lambda_1(HPH') \geq C\lambda_1(P) \geq C$. Similarly, since we have assumed $\|P\| \leq C$ in (3.2.1),

$\lambda_1(HPH') \leq C\lambda_1(P) \leq C$. Here, C is a generic constant. We have proved that

$$\mathbb{E}[C_n] = c_n \asymp \|\theta\|^8.$$

To compute the variance of C_n, write

$$C_n = \widetilde{Q}_n + \Delta, \qquad \text{where} \quad \widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}.$$

The variance of Δ is computed in the proof of Lemma B.2 of Jin et al. (2018). Using the upper bound of the variance of $\left(\sum_{CC(I_n)} \Delta^{(k)}_{i_1i_2i_3i_4}\right)$ for $k = 1, 2, 3$ there, we have

$$\mathrm{Var}(\Delta) \leq C\|\theta\|_3^6\|\theta\|^8.$$

Furthermore, we show in the proof of Lemma 37 that $\mathrm{Var}(\widetilde{Q}_n) = 8c_n \cdot [1 + o(1)]$. It follows that $\mathrm{Var}(\widetilde{Q}_n) \asymp c_n \asymp \|\theta\|^8$. Combining these results gives

$$\mathrm{Var}(C_n) \leq C\|\theta\|^8 \cdot [1 + \|\theta\|_3^6].$$

Consider the last claim. For any $\varepsilon > 0$, using Chebyshev's inequality, we have

$$\mathbb{P}(|C_n/c_n - 1| \geq \varepsilon) \leq (c_n\varepsilon)^{-2}\mathrm{Var}(C_n) \leq \frac{C(1 + \|\theta\|_3^6)}{\varepsilon^2\, \|\theta\|^8}.$$

Here we have used the first two claims. Since $\|\theta\|_3^3 \leq \theta_{\max}\|\theta\|^2 = o(\|\theta\|^8)$, the rightmost term is $o(1)$ as $n \to \infty$. This proves that $C_n/c_n \to 1$ in probability. □

3.C.2 Proof of Lemma 37

In the proof of Theorem 3.2 of Jin et al. (2018), it was shown that $\widetilde{Q}_n/\sqrt{\mathrm{Var}(\widetilde{Q}_n)} \to N(0, 1)$ in law (in the proof there, $\widetilde{Q}_n/\sqrt{\mathrm{Var}(\widetilde{Q}_n)}$ is denoted as $S_{n,n}$). It remains to prove $\mathrm{Var}(\widetilde{Q}_n) = 8c_n \cdot [1 + o(1)]$.

Note that for each ordered quadruple (i, j, k, ℓ) with four distinct indices, there are 8 summands in the definition of $\widetilde{Q}_n$ whose values are exactly the same; these summands correspond to $(i_1, i_2, i_3, i_4) \in \{(i, j, k, \ell), (j, k, \ell, i), (k, \ell, i, j), (\ell, i, j, k), (k, j, i, \ell), (j, i, \ell, k), (i, \ell, k, j), (\ell, k, j, i)\}$. We treat these 8 summands as in an equivalent class. Denote by CC_4 the collection of all such equivalent classes. Then, for any doubly indexed sequence $\{x_{ij}\}_{1\leq i\neq j\leq n}$ such that $x_{ij} = x_{ji}$, it is true that $\sum_{i_1,i_2,i_3,i_4(dist)} x_{i_1i_2}x_{i_2i_3}x_{i_3i_4}x_{i_4i_1} = 8\sum_{CC_4} x_{i_1i_2}x_{i_2i_3}x_{i_3i_4}x_{i_4i_1}$. In particular,

$$\widetilde{Q}_n = 8\sum_{CC_4} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}.$$

The summands are independent of each other, and the variance of $W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$ is equal to $\Omega^*_{i_1i_2}\Omega^*_{i_2i_3}\Omega^*_{i_3i_4}\Omega^*_{i_4i_1}$, where $\Omega^*_{ij} = \Omega_{ij}(1 - \Omega_{ij})$. As a result,

$$\mathrm{Var}(\widetilde{Q}_n) = 64\sum_{CC_4} \Omega^*_{i_1i_2}\Omega^*_{i_2i_3}\Omega^*_{i_3i_4}\Omega^*_{i_4i_1} = 8\sum_{i_1,i_2,i_3,i_4(dist)} \Omega^*_{i_1i_2}\Omega^*_{i_2i_3}\Omega^*_{i_3i_4}\Omega^*_{i_4i_1}.$$

Recall that $c_n = \sum_{i_1,i_2,i_3,i_4(dist)} \Omega_{i_1i_2}\Omega_{i_2i_3}\Omega_{i_3i_4}\Omega_{i_4i_1}$. Then,

$$\begin{aligned}
|\mathrm{Var}(\widetilde{Q}_n) - 8c_n| &\leq 8 \sum_{i_1,i_2,i_3,i_4(dist)} |\Omega_{i_1i_2}\Omega_{i_2i_3}\Omega_{i_3i_4}\Omega_{i_4i_1} - \Omega^*_{i_1i_2}\Omega^*_{i_2i_3}\Omega^*_{i_3i_4}\Omega^*_{i_4i_1}| \\
&\leq 8 \sum_{i_1,i_2,i_3,i_4(dist)} \Omega_{i_1i_2}\Omega_{i_2i_3}\Omega_{i_3i_4}\Omega_{i_4i_1} \cdot C\|\Omega\|_{\max} \\
&= 8c_n \cdot O(\theta^2_{\max}).
\end{aligned}$$

Since $\theta_{\max} = o(1)$ by the condition (3.2.1), we immediately have $\mathrm{Var}(\widetilde{Q}_n) = 8c_n\cdot[1+o(1)]$. $\square$

3.C.3 Proof of Lemma 38

The proof is combined with the proof of Lemma 43; see below.

3.C.4 Proof of Lemma 39

Consider the first claim. Since $b_n = 2\|\theta\|^4 \cdot [g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g]$ (see (3.4.21)), it suffices to show that

$$g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g \asymp 1.$$

The vectors $g, h \in \mathbb{R}^K$ are defined by $g_k = (\mathbf{1}_k'\theta)/\|\theta\|_1$ and $h_k = (\mathbf{1}_k'\Theta^2\mathbf{1}_k)^{1/2}/\|\theta\|$, where $\mathbf{1}_k$ is for short of $\mathbf{1}_k^{(K)}$. By condition (3.2.2), $c_1 \leq g_k \leq 1$ and $c_1 \leq h_k^2 \leq 1$ for $1 \leq k \leq K$, and $\|P\| \leq c_2$, for some constants $c_1, c_2 \in (0,1)$.

For the upper bound, by $h_k^2 \leq 1$ and $\|P\| \leq c_2$, we have $\|(PH^2P) \circ (PH^2P)\| \leq C$. Since P has unit diagonals and $g_k \geq c_1$, the diagonal elements of $V = \mathrm{d}iag(Pg)$ is no less than c_1. Hence

$$g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g \leq \|g'V^{-1}\|^2 \cdot \|PH^2P \circ PH^2P\| \leq C. \tag{3.C.99}$$

For the lower bound, since P has unit diagonals and $h_k^2 \geq c_1$, we can lower bound diagonal elements of $PH^2P \circ PH^2P$ by c_1^2. Since $g \in \mathbb{R}^K$ is a non-negative vector with entries summing to 1, the diagonal elements of $V = \mathrm{d}iag(Pg)$ is no more than $\max_{k,l} P_{k,\ell} \leq \|P\| \leq c_2$. Therefore each entry of vector gV^{-1} is at least c_1/c_2. Since $PH^2P \circ PH^2P \in \mathbb{R}^{(K,K)}$ is non-negative matrix and $g'V^{-1} \in \mathbb{R}^{(K}$ is non-negative vector, we can lower bound

$$g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g \geq c_1^2\|g'V^{-1}\|^2 \geq C, \tag{3.C.100}$$

Combining (3.C.99)-(3.C.100), we completes the proof of the first claim.

Consider the second claim. Introduce the following event

$$A_n = \big\{\widehat{\Pi}^{(K)} = \Pi, \text{ up to a permutation in the columns of } \widehat{\Pi}^{(K)}\big\}. \tag{3.C.101}$$

By Theorem 3.2.1, when $m = K$, SCORE exactly recovers Π with probability $1 - o(n^{-3})$, i.e.,

$$\mathbb{P}(A_n^c) \leq Cn^{-3} = o(1).$$

This means if we replace every $\widehat{\Pi}^{(K)}$ in the definition of $B_n^{(K)}$ with Π, and denote the resulting quantity as $B_n^{(K,0)}$, the above inequality immediately implies that $B_n^{(K)}/B_n^{(K,0)} \xrightarrow{p} 1$. So we only need to prove $B_n^{(K,0)}/b_n \xrightarrow{p} 1$. Since we will never use the original definition of $B_n^{(K)}$ in the rest of the proof, without causing any confusion we will suspend the original definitions of $B_n^{(K)}$ and the quantities used to define $B_n^{(K)}$, including $(\widehat{\theta}, \widehat{g}, \widehat{V}, \widehat{P}, \widehat{H})$, and use them to actually denote the correspondents with every $\widehat{\Pi}^{(K)}$ replaced by Π.

Recall the formulas for $B_n^{(K)}$ and b_n in (3.2.11) and (3.4.21), we have

$$\frac{B_n^{(K)}}{b_n} = \frac{\|\widehat{\theta}\|^4}{\|\theta\|^4} \cdot \frac{\widehat{g}'\widehat{V}^{-1}(\widehat{P}\widehat{H}^2\widehat{P} \circ \widehat{P}\widehat{H}^2\widehat{P})\widehat{V}^{-1}\widehat{g}}{g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g}. \tag{3.C.102}$$

To show that $B_n^{(K)}/b_n \to 1$, we need the follow lemma, which is proved in Section 3.D.2.

Lemma 47. *Suppose the conditions of Theorem 3.2.1 hold. Let $\mathbf{1}_n \in \mathbb{R}^n$ be the vector of 1's, and let $\mathbf{1}_k \in \mathbb{R}^n$ be the vector such that $\mathbf{1}_k(i) = 1\{i \in \mathcal{N}_k\}$, for $1 \le i \le n$ and $1 \le k \le K$. As $n \to \infty$, for all $1 \le k \le K$,*

$$\frac{\mathbf{1}_n' A \mathbf{1}_n}{\mathbf{1}_n' \Omega \mathbf{1}_n} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' A \mathbf{1}_n}{\mathbf{1}_k' \Omega \mathbf{1}_n} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' A \mathbf{1}_k}{\mathbf{1}_k' \Omega \mathbf{1}_k} \xrightarrow{p} 1.$$

Moreover, let d_i be the degree of node i and let $d_i^* = (\Omega \mathbf{1}_n)_i$, for $1 \le i \le n$. Write $D = diag(d) \in \mathbb{R}^{n,n}$ and $D^* = diag(d^*) \in \mathbb{R}^{n,n}$. As $n \to \infty$, for all $1 \le k \le K$,

$$\frac{\|\widehat{\theta}\|_1}{\|\theta\|_1} \xrightarrow{p} 1, \qquad \frac{\|\widehat{\theta}\|}{\|\theta\|} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' D^2 \mathbf{1}_k}{\mathbf{1}_k' (D^*)^2 \mathbf{1}_k} \xrightarrow{p} 1.$$

First, by Lemma 47, $\|\widehat{\theta}\|/\|\theta\| \xrightarrow{p} 1$. It follows from the continuous mapping theorem that

$$\|\widehat{\theta}\|^4/\|\theta\|^4 \xrightarrow{p} 1. \tag{3.C.103}$$

Second, recall that $g_k = (\mathbf{1}_k'\theta)/\|\theta\|_1$ and $\widehat{g}_k = (\mathbf{1}_k'\widehat{\theta})/\|\widehat{\theta}\|_1$, where by (3.2.6), we have the equality $\mathbf{1}_k'\widehat{\theta} = (\mathbf{1}_k'd) \cdot \sqrt{\mathbf{1}_k' A \mathbf{1}_k}/(\mathbf{1}_k' A \mathbf{1}_n)$. Here, keep in mind that we have replaced $\widehat{\Pi}^{(K)}$ with Π, which implies that $\widehat{\mathbf{1}}_k = \mathbf{1}_k$. The vector d is such that $d = A\mathbf{1}_n$. It follows that $\mathbf{1}_k'\widehat{\theta} = \sqrt{\mathbf{1}_k' A \mathbf{1}_k}$. Furthermore, $\mathbf{1}_k'\Omega\mathbf{1}_k = (\mathbf{1}_k'\theta)^2$, because P has unit diagonals. Combining the above gives

$$\frac{\widehat{g}_k}{g_k} = \frac{\mathbf{1}_k'\widehat{\theta}}{\mathbf{1}_k'\theta} \cdot \frac{\|\theta\|_1}{\|\widehat{\theta}\|_1} = \frac{\sqrt{\mathbf{1}_k' A \mathbf{1}_k}}{\sqrt{\mathbf{1}_k' \Omega \mathbf{1}_k}} \cdot \frac{\|\theta\|_1}{\|\widehat{\theta}\|_1} \xrightarrow{p} 1, \qquad 1 \le k \le K. \tag{3.C.104}$$

Third, note that by definition and basic algebra, both P and $\widehat{P}$ have unit diagonals. We compare their off-diagonals. By (3.2.6), $\widehat{P}_{k\ell} = \mathbf{1}_k' A \mathbf{1}_\ell / \sqrt{(\mathbf{1}_k' A \mathbf{1}_k)(\mathbf{1}_\ell' A \mathbf{1}_\ell)}$. At the same time, it can be easily verified that $P_{k\ell} = \mathbf{1}_k' \Omega \mathbf{1}_\ell / \sqrt{(\mathbf{1}_k' \Omega \mathbf{1}_k)(\mathbf{1}_\ell' \Omega \mathbf{1}_\ell)}$. Introduce

$$X = \frac{\sqrt{(\mathbf{1}_k' \Omega \mathbf{1}_k)(\mathbf{1}_\ell' \Omega \mathbf{1}_\ell)}}{\sqrt{(\mathbf{1}_k' A \mathbf{1}_k)(\mathbf{1}_\ell' A \mathbf{1}_\ell)}}.$$

By Lemma 47, $X \xrightarrow{p} 1$. We re-write

$$\widehat{P}_{k\ell} - P_{k\ell} = \frac{\mathbf{1}_k' A \mathbf{1}_\ell - \mathbf{1}_k' \Omega \mathbf{1}_\ell}{\sqrt{(\mathbf{1}_k' A \mathbf{1}_k)(\mathbf{1}_\ell' A \mathbf{1}_\ell)}} + P_{k\ell}(X-1) = \frac{\mathbf{1}_k' W \mathbf{1}_\ell}{(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)} X + P_{k\ell}(X-1),$$

where in the last inequality we have used the fact that $\mathbf{1}_k'\Omega\mathbf{1}_k = (\mathbf{1}_k'\theta)^2$ for all $1 \le k \le K$. Note that $\mathbb{E}[\mathbf{1}_k' W \mathbf{1}_\ell] = 0$. Moreover, $\mathrm{Var}(W_{ij}) \le \|P\|_{\max}\theta_i\theta_j \le C\theta_i\theta_j$. It follows that $\mathrm{Var}(\mathbf{1}_k' W \mathbf{1}_\ell) \le C(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)$. Therefore,

$$\mathbb{E}\left[\frac{\mathbf{1}_k' W \mathbf{1}_\ell}{(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)}\right]^2 \le \frac{C}{(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)} = O(\|\theta\|_1^{-2}) = o(1).$$

Hence, $\frac{\mathbf{1}_k' W \mathbf{1}_\ell}{(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)} \xrightarrow{p} 0$. Combining the above results, we have

$$\widehat{P}_{k\ell} - P_{k\ell} \xrightarrow{p} 0, \qquad 1 \le k, \ell \le K. \tag{3.C.105}$$

Fourth, since $V = \mathrm{d}iag(Pg)$ and $\widehat{V} = \mathrm{d}iag(\widehat{P}\widehat{g})$, it follows from (3.C.104) and (3.C.105) that

$$\widehat{V}_{kk}/V_{kk} \xrightarrow{p} 1, \qquad 1 \le k \le K. \tag{3.C.106}$$

Last, note that $H^2, \widehat{H}^2 \in \mathbb{R}^{K,K}$ are diagonal matrices, with k-th diagonal elements being h_k^2 and $\widehat{h}_k^2$, respectively. By (3.2.10), $\widehat{h}_k^2 = (\mathbf{1}_k'\widehat{\Theta}^2\mathbf{1}_k)/\|\widehat{\theta}\|^2$. In addition, by (3.2.6), for any $i \in \mathcal{N}_k$, we have $\widehat{\theta}_i^2 = d_i^2(\mathbf{1}_k' A \mathbf{1}_k)/(\mathbf{1}_k' A \mathbf{1}_n)^2$. We thus re-write

$$\widehat{H}_{kk} \equiv \widehat{h}_k^2 = \frac{(\mathbf{1}_k' D^2 \mathbf{1}_k)\cdot(\mathbf{1}_k' A \mathbf{1}_k)}{(\mathbf{1}_k' A \mathbf{1}_n)^2 \cdot \|\widehat{\theta}\|^2}.$$

Additionally, $h_k = (\mathbf{1}_k'\Theta^2\mathbf{1}_k)/\|\theta\|^2$, as defined in the paragraph below (3.4.21). By direct calculations, $(\mathbf{1}_k'\Omega\mathbf{1}_n)/\sqrt{\mathbf{1}_k'\Omega\mathbf{1}_k} = \big[(\mathbf{1}_k'\theta)\sum_\ell P_{k\ell}(\mathbf{1}_\ell'\theta)\big]/(\mathbf{1}_k'\theta) = \sum_\ell P_{k\ell}(\mathbf{1}_\ell'\theta)$. Also, for any $i \in \mathcal{N}_k$, we have $d_i^* = (\Omega\mathbf{1}_n)_i = \theta_i[\sum_\ell P_{k\ell}(\mathbf{1}_\ell'\theta)]$. It implies that $\mathbf{1}_k'(D^*)^2\mathbf{1}_k = (\mathbf{1}_k'\Theta^2\mathbf{1}_k)[\sum_\ell P_{k\ell}(\mathbf{1}_\ell'\theta)]^2$. We can use these expressions to verify that

$$H_{kk} \equiv h_k^2 = \frac{[\mathbf{1}_k'(D^*)^2\mathbf{1}_k]\cdot(\mathbf{1}_k'\Omega\mathbf{1}_k)}{(\mathbf{1}_k'\Omega\mathbf{1}_n)^2\cdot\|\theta\|^2}.$$

We apply Lemma 47 to obtain that

$$\widehat{H}_{kk}/H_{kk} \xrightarrow{p} 1, \qquad 1 \le k \le K. \tag{3.C.107}$$

We plug (3.C.103), (3.C.104), (3.C.105), (3.C.106) and (3.C.107) into (3.C.102). It follows from elementary probability that $B_n^{(K)}/b_n \to 1$. This gives the second claim. □

3.C.5 Proof of Lemma 40

Recall $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, ..., \mathcal{N}_m^{(m,0)}$ are "fake" communities associated with Π_0, and we decompose the vector $\mathbf{1}_n \in \mathbb{R}^n$ as follows

$$\mathbf{1}_n = \sum_{k=1}^m \mathbf{1}_k^{(m,0)}, \qquad \text{where } \mathbf{1}_k^{(m,0)}(j) = 1 \text{ if } j \in \mathcal{N}_k^{(m,0)} \text{ and 0 otherwise.} \tag{3.C.108}$$

Notice for $\Pi_0 \in \mathcal{G}_m$ defined in (3.4.26), there exists an $K \times m$ matrix L_0 such that $\Pi_0 = \Pi L_0$.

By definitions, $\Omega^{(m,0)} = \Theta^{(m,0)}\Pi_0 P^{(m,0)}\Pi_0'\Theta^{(m,0)}$. Here $\Theta^{(m,0)}$ and $P^{(m,0)}$ are obtained by replacing $(d_i, \widehat{\mathbf{1}}_k, A)$ by $(d_i^*, \mathbf{1}_k^{(m,0)}, \Omega)$ in the definition (3.2.6). It yields that, for $1 \le k, \ell \le m$ and $i \in \mathcal{N}_k^{(m,0)}$,

$$\theta_i^{(m,0)} = \frac{d_i^*}{(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_n}\cdot\sqrt{(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_k^{(m,0)}}, \qquad P_{k\ell}^{(m,0)} = \frac{(\mathbf{1}_k^{(m,0)})'\Omega(\mathbf{1}_\ell^{(m,0)})}{\sqrt{(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_k^{(m,0)}}\sqrt{(\mathbf{1}_\ell^{(m,0)})'\Omega\mathbf{1}_\ell^{(m,0)}}}.$$

As a result, for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\Omega_{ij}^{(m,0)} = \theta_i^{(m,0)}\theta_j^{(m,0)}P_{k\ell}^{(m,0)} = d_i^* d_j^* \cdot \frac{(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_\ell^{(m,0)}}{[(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_n]\cdot[(\mathbf{1}_\ell^{(m,0)})'\Omega\mathbf{1}_n]}. \tag{3.C.109}$$

Note that $(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_\ell^{(m,0)} = (\Pi_0'\Omega\Pi_0)_{k\ell}$. Since $\Omega = \Theta\Pi P\Pi'\Theta$ and $D_0 = \Pi_0'\Theta\Pi$, we immediately have $\Pi_0'\Omega\Pi_0 = \Pi_0'\Theta\Pi P\Pi'\Theta'\Pi_0 = D_0 P D_0'$. It follows that

$$(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_\ell^{(m,0)} = (D_0 P D_0')_{k\ell}, \qquad 1 \le k, \ell \le m.$$

Similarly, $(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_n = (e_k'\Pi_0')\Omega(\Pi\mathbf{1}_K) = e_k'\Pi_0'\Theta\Pi P\Pi'\Theta\Pi\mathbf{1}_K = e_k'D_0PD\mathbf{1}_K$. This gives

$$(\mathbf{1}_k^{(m,0)})'\Omega\mathbf{1}_n = \mathrm{d}iag(D_0PD\mathbf{1}_K)_{kk}, \qquad 1 \le k, \ell \le m.$$

We plug the above equalities into (3.C.109). It follows that, for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\Omega_{ij}^{(m,0)} = d_i^* d_j^* \cdot \left[(\mathrm{d}iag(D_0PD\mathbf{1}_K))^{-1}D_0PD_0'(\mathrm{d}iag(D_0PD\mathbf{1}_K))^{-1}\right]_{k\ell}. \tag{3.C.110}$$

Write for short

$$M = [\mathrm{d}iag(D_0PD\mathbf{1}_K)]^{-1}(D_0PD_0')[\mathrm{d}iag(D_0PD\mathbf{1}_K)]^{-1}. \tag{3.C.111}$$

Then, (3.C.110) can be written equivalently as

$$\Omega_{ij}^{(m,0)} = d_i^* d_j^* \cdot \sum_{k,\ell=1}^{m} M_{k\ell} \cdot 1\{i \in \mathcal{N}_k^{(m,0)}\} \cdot 1\{j \in \mathcal{N}_\ell^{(m,0)}\}.$$

By definition, $L_0(u,k) = 1\{\mathcal{N}_u \subset \mathcal{N}_k^{(m,0)}\}$, for $1 \le u \le K$ and $1 \le k \le m$. Therefore, we have the equalities: $1\{i \in \mathcal{N}_k^{(m,0)}\} = \sum_{u=1}^K L_0(u,k) \cdot 1\{i \in \mathcal{N}_u\}$ and $1\{j \in \mathcal{N}_\ell^{(m,0)}\} = \sum_{v=1}^K L_0(v,\ell) \cdot 1\{j \in \mathcal{N}_v\}$. Combining them with the above equation gives

$$\begin{aligned}\Omega_{ij}^{(m,0)} &= d_i^* d_j^* \cdot \sum_{u,v=1}^{K} 1\{i \in \mathcal{N}_u\} \cdot 1\{j \in \mathcal{N}_v\} \sum_{k,\ell=1}^{m} L_0(u,k)L_0(v,\ell)M_{k\ell} \\ &= d_i^* d_j^* \cdot \sum_{u,v=1}^{K} 1\{i \in \mathcal{N}_u\} \cdot 1\{j \in \mathcal{N}_v\} \cdot (L_0ML_0')_{uv}. \end{aligned} \tag{3.C.112}$$

By definition, $d^* = \Omega\mathbf{1}_n = \Omega(\Pi\mathbf{1}_K)$. Since $\Omega = \Theta\Pi P\Pi'\Theta$, we immediately have

$$d_i^* = \theta_i \cdot \pi_i'P\Pi'\Theta\Pi\mathbf{1}_K = \theta_i \cdot \pi_i'PD\mathbf{1}_K = \theta_i \cdot \sum_{u=1}^{K} \mathrm{d}iag(PD\mathbf{1}_K)_{uu} \cdot 1\{i \in \mathcal{N}_u\}.$$

Similarly, we have $d_j^* = \theta_i \cdot \sum_{v=1}^K \mathrm{d}iag(PD\mathbf{1}_K)_{vv} \cdot 1\{j \in \mathcal{N}_v\}$. Plugging the expressions of (d_i^*, d_j^*) into (3.C.112) gives

$$\begin{aligned}\Omega_{ij}^{(m,0)} &= \theta_i\theta_j \sum_{u,v=1}^{K} 1\{i \in \mathcal{N}_u\}\, 1\{j \in \mathcal{N}_v\}\, \mathrm{d}iag(PD\mathbf{1}_K)_{uu}(L_0ML_0')_{uv}\mathrm{d}iag(PD\mathbf{1}_K)_{vv} \\ &= \theta_i\theta_j \cdot \pi_i'\big[\mathrm{d}iag(PD\mathbf{1}_K)L_0ML_0'\mathrm{d}iag(PD\mathbf{1}_K)\big]\pi_j. \end{aligned} \tag{3.C.113}$$

Combining it with the expression of M in (3.C.111) gives the claim. □

3.C.6 Proof of Lemma 41

The claim of $c_n \asymp \|\theta\|^8$ is proved in Lemma 36. To prove the claim of $\lambda_1 \asymp \|\theta\|^2$, we note that by Lemma 30, $\lambda_k = \|\theta\|^2 \cdot \lambda_k(HPH)$, where H is the diagonal matrix such that $H_{kk} = \|\theta^{(k)}\|^2/\|\theta\|^2$. By the condition (3.2.2), all the diagonal entries of H are between $[c, 1]$, for a constant $c \in (0,1)$. It follows that $\lambda_1(HPH) \asymp \lambda_1(P)$. Since $\lambda_1 \ge P_{11} = 1$ and $\lambda_1 \le \|P\| \le C$, we have $\lambda_1(P) \asymp 1$. Combining the above gives

$$\lambda_1 \asymp \|\theta\|^2\lambda_1(P) \asymp \|\theta\|^2.$$

We then prove the claims related to the matrix $\widetilde{\Omega}$. First, we show the upper bound of $|\widetilde{\Omega}_{ij}|$ and the lower bound of $\mathrm{tr}(\widetilde{\Omega}^4)$. Recall that $\widetilde{\Omega} = \Omega - \Omega^{(m,0)}$. By Lemma 40, $\Omega^{(m,0)} = \Theta\Pi P_0 \Pi'\Theta$ for a rank-m matrix P_0. It follows that

$$\widetilde{\Omega} = \Theta\Pi(P - P_0)\Pi'\Theta. \tag{3.C.114}$$

Let H be the same diagonal matrix as above. It can be easily verified that $\|\theta\|^2 \cdot H^2 = \Pi'\Theta^2\Pi$. This means that the matrix $U = \|\theta\|^{-1}\Theta\Pi H^{-1}$ satisfies the equality $U'U = I_K$. As a result, we can write $\widetilde{\Omega} = U \cdot (\|\theta\|^2 \cdot H(P - P_0)H) \cdot U'$. Since U contains orthonormal columns, the nonzero eigenvalues of $\widetilde{\Omega}$ are the same as the nonzero eigenvalues of $\|\theta\|^2 \cdot H(P - P_0)H$, i.e.,

$$\widetilde{\lambda}_k = \|\theta\|^2 \cdot \lambda_k(H(P - P_0)H), \qquad 1 \le k \le m.$$

In particular, $|\widetilde{\lambda}_1| = \|\theta\|^2 \cdot \|H(P - P_0)H\| \asymp \|\theta\|^2 \cdot \|P - P_0\| \asymp \lambda_1\|P - P_0\|$, where we have used $\|H\| \asymp \|H^{-1}\| \asymp 1$, and $\lambda_1 \asymp \|\theta\|^2$. Combining it with the definition of τ gives

$$\tau \asymp \|P - P_0\|. \tag{3.C.115}$$

Consider $|\widetilde{\Omega}_{ij}|$. By (3.C.114), $|\widetilde{\Omega}_{ij}| = \theta_i\theta_j \cdot |\pi_i'(P - P_0)\pi_j| \le \theta_i\theta_j \cdot C\|P - P_0\|$. We plug in (3.C.115) to get $|\widetilde{\Omega}_{ij}| \le C\tau\theta_i\theta_j$, for $1 \le i, j \le n$. Consider $\mathrm{tr}(\widetilde{\Omega}^4)$. We have seen that $|\widetilde{\lambda}_1| \asymp \|\theta\|^2 \cdot \|P - P_0\| \asymp \tau\|\theta\|^2$. As a result, $\mathrm{tr}(\widetilde{\Omega}^4) \ge \widetilde{\lambda}_1^4 \ge C\tau^4\|\theta\|^8$.

Next, we study the order of τ. Note that $\Omega = \Omega^{(m,0)} + \widetilde{\Omega}$. We aim to apply Weyl's inequality. In our notation, $\lambda_k(\cdot)$ refers to the kth largest eigenvalue (in magnitude) of a symmetric matrix. As a result, $|\lambda_k(\cdot)|$ is the kth singular value. By Weyl's inequality for singular values (equation (7.3.13) of Horn and Johnson (1985)), we have

$$|\lambda_{r+s-1}(\Omega)| \le |\lambda_r(\Omega^{(m,0)})| + |\lambda_s(\widetilde{\Omega})|, \quad \text{for } 1 \le r, s \le n-1.$$

Since $\Omega^{(m,0)}$ only has m nonzero eigenvalues, by taking $r = m+1$ and $s = k$ in the above, we immediately have

$$|\lambda_{m+k}(\Omega)| \le |\lambda_k(\widetilde{\Omega})| = |\widetilde{\lambda}_k|, \qquad 1 \le k \le K - m. \tag{3.C.116}$$

In particular, $|\widetilde{\lambda}_1| \ge |\lambda_{m+1}| \ge |\lambda_K|$. At the same time, $\lambda_1 \asymp \|\theta\|^2$ and by definition, $\tau = |\widetilde{\lambda}_1|/\lambda_1$. It follows that

$$\tau\|\theta\| \ge (|\lambda_K|/\lambda_1) \cdot \|\theta\| \ge C\big(|\lambda_K|/\sqrt{\lambda_1}\big) \to \infty.$$

This gives $\tau\|\theta\| \to \infty$. We then prove $\tau \le C$. In light of (3.C.115), it suffices to show $\|P_0\| \le C$. Consider the expression of P_0 in Lemma 40. It is easy to see that $\|L_0\| \le C$, $\|D_0PD_0'\| \le C\|\theta\|_1^2$, and $\|\mathrm{diag}(PD\mathbf{1}_K)\| \le C\|\theta\|_1$. As a result,

$$\|P_0\| \le C\|\theta\|_1^4 \cdot \|diag(D_0PD\mathbf{1}_K)^{-1}\|^2. \tag{3.C.117}$$

Since $D_0 = \Pi_0'\Theta\Pi$ and $D = \Pi'\Theta\Pi$, it is true that $D_0PD\mathbf{1}_K = \Pi_0'\Theta\Pi P\Pi'\Theta\Pi\mathbf{1}_K = \Pi_0'\Theta\Pi P\Pi'\Theta\mathbf{1}_n = \Pi_0'\Omega\mathbf{1}_n$. Then, for each $1 \le k \le m$,

$$[diag(D_0PD\mathbf{1}_K)]_{kk} = (\Pi_0'\Omega\mathbf{1}_n)_k = \sum_{i \in \mathcal{N}_k^{(m,0)}} d_i^*, \qquad \text{where} \quad d^* = \Omega\mathbf{1}_n.$$

Here $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, ..., \mathcal{N}_m^{(m,0)}$ are the pseudo-communities defined by Π_0. Suppose $i \in \mathcal{N}_\ell$ for some true community $\mathcal{N}_\ell$. Then, $d_i^* \ge \sum_{j \in \mathcal{N}_\ell} \theta_i\theta_j P_{\ell\ell} = \theta_i\|\theta^{(\ell)}\|_1 \ge C\theta_i\|\theta\|_1$. Moreover, for any $\Pi_0 \in \mathcal{G}_m$, each pseudo-community $\mathcal{N}_k^{(m,0)}$ is the union of one or more true community. It yields that $\sum_{i \in \mathcal{N}_k^{(m,0)}} \theta_i \ge \min_{1 \le \ell \le K}\{\|\theta^{(\ell)}\|_1\} \ge C\|\theta\|_1$. Combining these results gives $\sum_{i \in \mathcal{N}_k^{(m,0)}} d_i^* \ge C\|\theta\|_1^2$. This shows that each diagonal entry of $diag(D_0PD\mathbf{1}_K)$ is lower bounded by $C\|\theta\|_1^2$. We immediately have

$$\|\mathrm{d}iag(D_0PD\mathbf{1}_K)^{-1}\| \leq C\|\theta\|_1^{-2}. \tag{3.C.118}$$

Combining (3.C.117) and (3.C.118) gives $\|P_0\| \leq C$. The claim $\tau \leq C$ then follows from (3.C.115). □

3.C.7 Proof of Lemma 42

Recall that $W = A - \Omega$. Given any $n \times n$ symmetric matrix T, we can define a random variable as follows:

$$\mathcal{Q}_W(T) = \sum_{i_1,i_2,i_3,i_4(dist)} (W_{i_1i_2} + T_{i_1i_2})(W_{i_2i_3} + T_{i_2i_3})(W_{i_3i_4} + T_{i_3i_4})(W_{i_4i_1} + T_{i_4i_1}). \tag{3.C.119}$$

Then, $\widetilde{Q}_n^{(m,0)}$ is a special case with $T = \widetilde{\Omega}^{(m,0)}$, where $\widetilde{\Omega}^{(m,0)}$ is defined in (3.4.31). We study the general form of $\mathcal{Q}_W(T)$. By an expansion of each summand, we can write $\mathcal{Q}_W(T)$ as the sum of 2^4 post-expansion sums. Each post-expansion sum takes a form

$$X = \sum_{i_1,i_2,i_3,i_4(dist)} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \tag{3.C.120}$$

where each of $a_{ij}, b_{ij}, c_{ij}, d_{ij}$ may take value in $\{W_{ij}, T_{ij}\}$. We divide the post-expansion sums into 6 common types and compute the mean and variance of each of them (see Table 3.2 for the special case of $T = \widetilde{\Omega}^{(m,0)}$). For example, the post-expansion sum $\sum_{i_1,i_2,i_3,i_4(dist)} T_{i_1i_2}T_{i_2i_3}T_{i_3i_4}T_{i_4}T_{i_1}$ is non-stochastic and has a zero variance. Its mean equals to $\mathrm{tr}(T^4) - \Delta$, where Δ contains the sum of $T_{i_1i_2}T_{i_2i_3}T_{i_3i_4}T_{i_4}T_{i_1}$ when some of the indices (i_1, i_2, i_3, i_4) are equal. As another example, the post-expansion sum $\sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$ has a zero mean, and since the summands are mutually uncorrelated, its variance is $\sum_{i_1,i_2,i_3,i_4(dist)} \Omega^*_{i_1i_2}\Omega^*_{i_2i_3}\Omega^*_{i_3i_1}\Omega^*_{i_4i_1}$, where $\Omega^*_{ij} = \Omega_{ij}(1 - \Omega_{ij})$.

Table 3.2: The 6 different types of the 16 post-expansion sums of $\widetilde{Q}_n^{(m,0)}$. In our setting, $\tau = \widetilde{\lambda}_1^{(m,0)}/\lambda_1$ and $\|\theta\|^{-1} \ll \tau \leq C$, and $\|\theta\|_3^3 \ll \|\theta\|^2 \ll \|\theta\|_1$.

Type	#	$(N_{\widetilde{\Omega}}, N_W)$	Examples	Mean	Variance
I	1	(0, 4)	$X_1 = \sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$	0	$\asymp \|\theta\|^8$
II	4	(1, 3)	$X_2 = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{\Omega}_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$	0	$\leq C\tau^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
IIIa	4	(2, 2)	$X_3 = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{\Omega}_{i_1i_2}\widetilde{\Omega}_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$	0	$\leq C\tau^4\|\theta\|^6\|\theta\|_3^6 = o(\tau^6\|\theta\|^8\|\theta\|_3^6)$
IIIb	2	(2, 2)	$X_4 = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{\Omega}_{i_1i_2}W_{i_2i_3}\widetilde{\Omega}_{i_3i_4}W_{i_4i_1}$	0	$\leq C\tau^4\|\theta\|_3^{12} = o(\|\theta\|^8)$
IV	4	(3, 1)	$X_5 = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{\Omega}_{i_1i_2}\widetilde{\Omega}_{i_2i_3}\widetilde{\Omega}_{i_3i_4}W_{i_4i_1}$	0	$\leq \tau^6\|\theta\|^8\|\theta\|_3^6$
V	1	(4, 0)	$X_6 = \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{\Omega}_{i_1i_2}\widetilde{\Omega}_{i_2i_3}\widetilde{\Omega}_{i_3i_4}\widetilde{\Omega}_{i_4i_1}$	$\sim \mathrm{tr}(\widetilde{\Omega}^4)$	0

Here we omit the calculation details, because similar calculations were done in Jin et al. (2019). In their Theorem 4.4, they analyzed $\mathcal{Q}_W(T)$ for T equal to a rank-1 matrix (denoted by $\widetilde{\Omega}$ there). However, their proof does not reply on the condition that $\widetilde{\Omega}$ is rank-1 and applies to any symmetric matrix. They actually proved the following lemma:

Lemma 48. *Consider a DCBM model where (3.2.1)-(3.2.2) and (3.2.4) hold. Let $W = A - \Omega$ and let $\mathcal{Q}_W(T)$ be the random variable defined in* (3.C.119). *As $n \to \infty$, suppose there is a constant $C > 0$ and a scalar $\alpha_n > 0$ such that $\alpha_n \leq C$, $\alpha_n\|\theta\| \to \infty$, and $|T_{ij}| \leq C\alpha_n\theta_i\theta_j$*

for all $1 \le i,j \le n$*. Then,* $\mathbb{E}[\mathcal{Q}_W(T)] = \mathrm{tr}(T^4) + o(\|\theta\|^4)$ *and* $\mathrm{Var}(\mathcal{Q}_W(T)) \le C(\|\theta\|^8 + \alpha_n^6\|\theta\|^8\|\theta\|_3^6)$.

We now set $T = \widetilde{\Omega}^{(m,0)}$ and verify the conditions of Lemma 48. Recall that $\tau = \widetilde{\lambda}_1/\lambda_1$, where $\widetilde{\lambda}_1$ and λ_1 are the respective largest (in magnitude) eigenvalue of $\widetilde{\Omega}^{(m,0)}$ and Ω. By Lemma 41,

$$\tau \le C, \qquad \tau\|\theta\| \to \infty, \qquad |\widetilde{\Omega}_{ij}^{(m,0)}| \le C\tau\theta_i\theta_j, \qquad \text{for all } 1 \le i,j \le n.$$

Therefore, we can apply Lemma 48 with $\alpha_n = \tau$. The claim follows immediately. $\square$

3.C.8 Proof of Lemma 43

Before proceed, recall (3.4.32) that

$$\widetilde{Q}_n^{(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} (W_{i_1i_2} + \widetilde{\Omega}_{i_1i_2}^{(m,0)})(W_{i_2i_3} + \widetilde{\Omega}_{i_2i_3}^{(m,0)})(W_{i_3i_4} + \widetilde{\Omega}_{i_3i_4}^{(m,0)})(W_{i_4i_1} + \widetilde{\Omega}_{i_4i_1}^{(m,0)}).$$

Here $\widetilde{\Omega}^{(m,0)} = \Omega - \Omega^{(m,0)}$ and $\Omega^{(m,0)}$ is as in (3.4.29). By Lemma 40, $\Omega^{(m,0)} = \Theta\Pi P_0 \Pi'\Theta$, for a rank-$m$ matrix P_0. If $m = K$ and $\Pi_0 = \Pi$, it can be verified that $P_0 = P$. Therefore, $\Omega^{(m,0)} = \Omega$, and $\widetilde{\Omega}^{(m,0)}$ reduces to a zero matrix. In this case, $\widetilde{Q}_n^{(m,0)}$ reduces to $\widetilde{Q}_n$ in (3.4.20). It means that we can treat Lemma 38 as a "special case" of Lemma 43, with $\widetilde{\Omega}^{(m,0)}$ being a zero matrix. We thus combine the proofs of two lemmas.

We now show the claim. First, we introduce two proxies of $Q_n^{(m,0)}$. By definition,

$$Q_n^{(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} (A_{i_1i_2} - \widehat{\Omega}_{i_1i_2}^{(m,0)})(A_{i_2i_3} - \widehat{\Omega}_{i_2i_3}^{(m,0)})(A_{i_3i_4} - \widehat{\Omega}_{i_3i_4}^{(m,0)})(A_{i_4i_1} - \widehat{\Omega}_{i_4i_1}^{(m,0)}).$$

By (3.4.27), $\widehat{\Omega}^{(m,0)}$ is defined by $\widehat{\theta}$, Π_0, and $\widehat{P}$. For $1 \le k \le m$, let $\mathcal{N}_k^{(m,0)}$ and $\mathbf{1}_k^{(m,0)}$ be the same as in (3.C.108). Then, $(\widehat{\theta}, \widehat{P})$ are obtained by replacing $\widehat{\mathbf{1}}_k$ with $\mathbf{1}_k^{(m,0)}$ in (3.2.6). For the rest of the proof, we write $\mathbf{1}_k = \mathbf{1}_k^{(m,0)}$ for short. It follows that, for $1 \le k, \ell \le K$ and $i \in \mathcal{N}_k^{(m,0)}$,

$$\widehat{\theta}_i^{(m,0)} = d_i \frac{\sqrt{\mathbf{1}_k' A \mathbf{1}_k}}{\mathbf{1}_k' A \mathbf{1}_n}, \quad \widehat{P}_{k\ell}^{(m,0)} = \frac{\mathbf{1}_k' A \mathbf{1}_\ell}{\sqrt{(\mathbf{1}_k' A \mathbf{1}_k)(\mathbf{1}_\ell' A \mathbf{1}_\ell)}}, \quad \text{with } \mathbf{1}_k = \mathbf{1}_k^{(m,0)} \text{ (for short)}.$$

We plug it into (3.4.27) and note that $d = A\mathbf{1}_n$. It yields that, for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\widehat{\Omega}_{ij}^{(m,0)} = d_i d_j \cdot \widehat{U}_{k\ell}^{(m,0)}, \qquad \text{where} \quad \widehat{U}_{k\ell}^{(m,0)} = \frac{\mathbf{1}_k' A \mathbf{1}_\ell}{(\mathbf{1}_k' d)(\mathbf{1}_\ell' d)}. \tag{3.C.121}$$

At the same time, in (3.C.109), we have seen that (recall: $d^* = \Omega\mathbf{1}_n$)

$$\Omega_{ij}^{(m,0)} = d_i^* d_j^* \cdot U_{k\ell}^{*(m,0)}, \qquad \text{where} \quad U_{k\ell}^{*(m,0)} = \frac{\mathbf{1}_k' \Omega \mathbf{1}_\ell}{(\mathbf{1}_k' d^*)(\mathbf{1}_\ell' d^*)}. \tag{3.C.122}$$

Note that (Ω, d^*) are approximately $(\mathbb{E}[A], \mathbb{E}[d])$ but there is subtle difference. We thus introduce an intermediate quantity:

$$U_{k\ell}^{(m,0)} = \frac{\mathbf{1}_k' \mathbb{E}[A] \mathbf{1}_\ell}{(\mathbf{1}_k' \mathbb{E}[d])(\mathbf{1}_\ell' \mathbb{E}[d])}. \tag{3.C.123}$$

We now use (3.C.121)-(3.C.123) to decompose $(A_{ij} - \widehat{\Omega}_{ij}^{(m,0)})$. Recall that $\widetilde{\Omega}_{ij}^{(m,0)} = \Omega_{ij} -$

$\Omega_{ij}^{(m,0)}$. We immediately have

$$A_{ij} - \widehat{\Omega}_{ij}^{(m,0)} = W_{ij} + \widetilde{\Omega}_{ij}^{(m,0)} + (\Omega_{ij}^{(m,0)} - \widehat{\Omega}_{ij}^{(m,0)}). \tag{3.C.124}$$

From now on, we omit the superscript "$(m,0)$" in $\widehat{U}_{k\ell}^{(m,0)}$, $U_{k\ell}^{*(m,0)}$ and $U_{k\ell}^{(m,0)}$, and rewrite them as $\widehat{U}_{k\ell}$, $U_{k\ell}^*$, and $U_{k\ell}$, respectively. By (3.C.121)-(3.C.123), $\Omega_{ij}^{(m,0)} - \widehat{\Omega}_{ij}^{(m,0)} = d_i^* d_j^* U_{k\ell}^* - d_i d_j \widehat{U}_{k\ell} = [d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell}] + U_{k\ell}[(\mathbb{E}d_i)(\mathbb{E}d_j) - d_i d_j] + (U_{k\ell} - \widehat{U}_{k\ell})d_i d_j$. It turns our that the term $U_{k\ell}[(\mathbb{E}d_i)(\mathbb{E}d_j) - d_i d_j]$ is the "dominating" term. This term does not have an exactly zero mean, and so we introduce a proxy to this term as

$$\delta_{ij}^{(m,0)} = U_{k\ell}\big[(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)\big]. \tag{3.C.125}$$

Note that $U_{k\ell}[(\mathbb{E}d_i)(\mathbb{E}d_j) - d_i d_j] = \delta_{ij}^{(m,0)} - U_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j)$. We then have

$$\begin{aligned}
\Omega_{ij}^{(m,0)} - \widehat{\Omega}_{ij}^{(m,0)} &= [d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell}] + [\delta_{ij}^{(m,0)} - U_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j)] + (U_{k\ell} - \widehat{U}_{k\ell})d_i d_j \\
&= \delta_{ij}^{(m,0)} + [d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell}] - U_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j) \\
&\quad + (U_{k\ell} - \widehat{U}_{k\ell})(\mathbb{E}d_i)(\mathbb{E}d_j) + (U_{k\ell} - \widehat{U}_{k\ell})[(\mathbb{E}d_i)(d_j - \mathbb{E}d_j) + (\mathbb{E}d_j)(d_i - \mathbb{E}d_i)] \\
&\quad + (U_{k\ell} - \widehat{U}_{k\ell})(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j) \\
&= \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)} + \epsilon_{ij}^{(m,0)},
\end{aligned}$$

where

$$\widetilde{r}_{ij}^{(m,0)} = -\widehat{U}_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j) \tag{3.C.126}$$

and

$$\begin{aligned}
\epsilon_{ij}^{(m,0)} = [d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell}] + (U_{k\ell} - \widehat{U}_{k\ell})(\mathbb{E}d_i)(\mathbb{E}d_j) \\
+ (U_{k\ell} - \widehat{U}_{k\ell})[(\mathbb{E}d_i)(d_j - \mathbb{E}d_j) + (\mathbb{E}d_j)(d_i - \mathbb{E}d_i)].
\end{aligned} \tag{3.C.127}$$

We plug the above results into (3.C.124) to get

$$A_{ij} - \widehat{\Omega}_{ij}^{(m,0)} = \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)} + \epsilon_{ij}^{(m,0)}. \tag{3.C.128}$$

We then use (3.C.128) to define two proxies of $Q_n^{(m,0)}$. For any $1 \leq i \neq j \leq n$, let

$$\begin{aligned}
X_{ij} &= \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)} + \varepsilon_{ij}^{(m,0)}, \\
\widetilde{X}_{ij}^* &= \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)}, \\
X_{ij}^* &= \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)}, \\
\widetilde{X}_{ij} &= \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij}.
\end{aligned} \tag{3.C.129}$$

Correspondingly, we introduce

$$\begin{aligned}
Q_n^{(m,0)} &= \sum_{i_1,i_2,i_3,i_4(dist)} X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}X_{i_4i_1} \\
\widetilde{Q}_n^{*(m,0)} &= \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{X}_{i_1i_2}^*\widetilde{X}_{i_2i_3}^*\widetilde{X}_{i_3i_4}^*\widetilde{X}_{i_4i_1}^*, \\
Q_n^{*(m,0)} &= \sum_{i_1,i_2,i_3,i_4(dist)} X_{i_1i_2}^*X_{i_2i_3}^*X_{i_3i_4}^*X_{i_4i_1}^*, \\
\widetilde{Q}_n^{(m,0)} &= \sum_{i_1,i_2,i_3,i_4(dist)} \widetilde{X}_{i_1i_2}\widetilde{X}_{i_2i_3}\widetilde{X}_{i_3i_4}\widetilde{X}_{i_4i_1}.
\end{aligned} \tag{3.C.130}$$

By comparing it with (3.4.32), we can see that the above expression of $\widetilde{Q}_n^{(m,0)}$ is the same as before. Additionally, by (3.C.128), the above expression of $Q_n^{(m,0)}$ is also equivalent to the definition. The other two quantities, $Q_n^{*(m,0)}$ and $\widetilde{Q}_n^{*(m,0)}$, are the two proxies we introduce here.

Next, we decompose

$$Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)} = (Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}) + (\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)}) + (Q_n^{(m,0)} - \widetilde{Q}_n^{(*,m,0)}).$$

For any random variables X, Y, Z, we know that $\mathbb{E}[X+Y+Z] = \mathbb{E}X + \mathbb{E}Y + \mathbb{E}Z$ and $\mathrm{Var}(X+Y+Z) \leq 3\mathrm{Var}(X) + 3\mathrm{Var} + 3\mathrm{Var}(Z)$. Therefore, to show the claim, we only need to study the mean and variance of each term in the above equation. The next three lemmas are proved in Sections 3.D.3-3.D.5.

Lemma 49. *Let $b_n = 2\|\theta\|^4 \cdot [g'V^{-1}(PH^2P \circ PH^2P)V^{-1}g]$ be the same as in (3.4.21). Under conditions of Lemma 38, it is true that*

$$\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}] = b_n + o(\|\theta\|^4), \quad \textit{and} \quad \mathrm{Var}\big(Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}\big) = o(\|\theta\|^8),$$

Let $\tau = \widetilde{\lambda}_1/\lambda_1$ be the same as in (3.4.33). Under conditions of Lemma 43, it is true that

$$\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}] = o(\tau^4\|\theta\|^8), \quad \textit{and} \quad \mathrm{Var}\big(Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}\big) \leq C\tau^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8).$$

Lemma 50. *Under conditions of Lemma 38, it is true that*

$$\mathbb{E}[\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)}] = o(\|\theta\|^4), \quad \textit{and} \quad \mathrm{Var}\big(\widetilde{Q}_n^{*(m,0)} - Q_n^{(*,m,0)}\big) = o(\|\theta\|^8).$$

Under conditions of Lemma 43, it is true that

$$\mathbb{E}[\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)}] = o\big(\|\theta\|^4 + \tau^4\|\theta\|^8\big), \quad \textit{and} \quad \mathrm{Var}\big(\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)}\big) = o\big(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6\big).$$

Lemma 51. *Under conditions of Lemma 38, it is true that*

$$\mathbb{E}[Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)}] = o(\|\theta\|^4), \quad \textit{and} \quad \mathrm{Var}\big(Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)}\big) = o(\|\theta\|^8).$$

Under conditions of Lemma 43, it is true that

$$\mathbb{E}[Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)}] = o\big(\|\theta\|^4 + \tau^4\|\theta\|^8\big), \quad \textit{and} \quad \mathrm{Var}\big(Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)}\big) = o\big(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6\big).$$

We now prove Lemma 38 and Lemma 43. By Lemma 49-Lemma 51, under the conditions of Lemma 38,

$$\mathbb{E}[Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}] = b_n + o(\|\theta\|^4), \quad \text{and} \quad \mathrm{Var}(Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}) = o(\|\theta\|^8),$$

which implies $\mathbb{E}(Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)} - b_n)^2 = o(\|\theta\|^8)$ and completes the proof of Lemma 38. Under the conditions of Lemma 43, it follows from Lemma 49-Lemma 51 that

$$\mathbb{E}[Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}] = o(\tau^4\|\theta\|^8) \quad \text{and} \quad \mathrm{Var}(Q_n^{(m,0)} - \widetilde{Q}_n^{(m,0)}) \leq C\tau^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8),$$

which completes the proof of Lemma 43. □

3.C.9 Proof of Lemma 44

Let $\mathcal{G}_m$ be the class of $n \times m$ membership matrices that satisfy NSP (the definition of $\mathcal{G}_m$ is in Section 3.4.2). By Theorem 3.2.2, $\widehat{\Pi}^{(m)} \in \mathcal{G}_m$ with probability $1 - O(n^{-3})$. Given any $\Pi_0 \in \mathcal{G}_m$, Let $B_n^{(m)}(\Pi_0)$ be defined in the same way as in (3.2.11), except that $(\widehat{\theta}, \widehat{g}, \widehat{V}, \widehat{P}, \widehat{H})$

are defined based on Π_0 instead of $\widehat{\Pi}^{(m)}$. Then, with probability $1 - O(n^{-3})$,

$$B_n^{(m)} \leq \max_{\Pi_0 \in \mathcal{G}_m} B_n(\Pi_0).$$

It follows from the probability union bound that

$$\mathbb{P}\big(B_n^{(m)} > C\|\theta\|^4\big) \leq \sum_{\Pi_0 \in \mathcal{G}_m} \mathbb{P}\big(B_n(\Pi_0) > C\|\theta\|^4\big) + O(n^{-3}).$$

Since $m < K$ and K is finite, $\mathcal{G}_m$ has only a bounded number of elements. Therefore, it suffices to show that

$$\mathbb{P}\big(B_n(\Pi_0) > C\|\theta\|^4\big) = o(1), \qquad \text{for each } \Pi_0 \in \mathcal{G}_m. \tag{3.C.131}$$

We now show (3.C.131). From now on, we fix $\Pi_0 \in \mathcal{G}_m$ and write $B_n(\Pi_0) = B_n$ for short. By (3.2.11) and direct calculations,

$$B_n = 2\|\widehat{\theta}\|^4 \cdot \widehat{g}'\widehat{V}^{-1}(\widehat{P}\widehat{H}^2\widehat{P} \circ \widehat{P}\widehat{H}^2\widehat{P})\widehat{V}^{-1}\widehat{g} = 2\|\widehat{\theta}\|^4 \cdot \sum_{1 \leq k, \ell \leq m} \frac{\widehat{g}_k\widehat{g}_\ell[(\widehat{P}\widehat{H}^2\widehat{P})_{k,\ell}]^2}{(\widehat{P}_k'\widehat{g}) \cdot (\widehat{P}_\ell'\widehat{g})},$$

where $\widehat{P}_k$ denotes the kth column of $\widehat{P}$. We have mis-used the notations $(\widehat{\theta}, \widehat{g}, \widehat{V}, \widehat{P}, \widehat{H})$, using them to refer to the counterparts of original definitions with $\widehat{\Pi}^{(m)}$ replaced by Π_0. Denote by $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, \ldots, \mathcal{N}_m^{(m,0)}$ the pseudo-communities defined by Π_0. Let $\mathbf{1}_k^{(m,0)} \in \mathbb{R}^n$ be such that $\mathbf{1}_k^{(m,0)}(i) = 1\{i \in \mathcal{N}_k^{(m,0)}\}$. We write $\mathbf{1}_k = \mathbf{1}_k^{(m,0)}$ when there is no confusion. By (3.2.10),

$$\widehat{g} = (\mathbf{1}_k'\widehat{\theta})/\|\widehat{\theta}\|_1, \qquad \widehat{h}_k^2 = (\mathbf{1}_k'\widehat{\Theta}^2\mathbf{1}_k)/\|\widehat{\theta}\|^2, \qquad 1 \leq k \leq m.$$

Note that $\widehat{g}$, $\widehat{h}$ and $\widehat{P}$ all have non-negative entries, with all entries of $\widehat{g}$ and $\widehat{h}$ are further bounded by 1. Moreover, the diagonals of $\widehat{P}$ are all equal to 1. It follows that, for all $1 \leq k, \ell \leq m$,

$$0 \leq \widehat{g}_k \leq \widehat{P}_k'\widehat{g}, \quad \text{and} \quad 0 \leq (\widehat{P}\widehat{H}^2\widehat{P})_{k\ell} \leq (\widehat{P}^2)_{k\ell}.$$

As a result,

$$B_n \;\leq\; 2\|\widehat{\theta}\|^4 \sum_{k,\ell=1}^{m} [(\widehat{P}^2)_{k\ell}]^2 \;\leq\; 2\|\widehat{\theta}\|^4 \cdot m^4\|\widehat{P}\|_{\max}^4, \tag{3.C.132}$$

where $\|\cdot\|_{\max}$ is the element-wise maximum norm. Below, we study $\|\widehat{P}\|_{\max}$ and $\|\widehat{\theta}\|$ separately.

First, we bound $\|\widehat{P}\|_{\max}$. By (3.2.6),

$$\widehat{P}_{k\ell} = (\mathbf{1}_k'A\mathbf{1}_\ell)/\sqrt{(\mathbf{1}_k'A\mathbf{1}_k)(\mathbf{1}_\ell'A\mathbf{1}_\ell)}.$$

Write $\mathbf{1}_k'A\mathbf{1}_\ell = \sum_{i \in \mathcal{N}_k^{(m,0)}, j \in \mathcal{N}_\ell^{(m,0)}} A_{ij}$, where $\mathbb{E}[A_{ij}] = \Omega_{ij}$, and $\sum_{i \in \mathcal{N}_k^{(m,0)}, j \in \mathcal{N}_\ell^{(m,0)}} \mathrm{Var}(A_{ij}) \leq \sum_{i \in \mathcal{N}_k^{(m,0)}, j \in \mathcal{N}_\ell^{(m,0)}} C\theta_i\theta_j \leq C(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)$. We apply the Bernstein's inequality Shorack and Wellner (1986) to get

$$\mathbb{P}\big(|\mathbf{1}_k'A\mathbf{1}_\ell - \mathbf{1}_k'\Omega\mathbf{1}_\ell| > t\big) \leq 2\exp\Big(-\frac{t^2/2}{C(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta) + t/3}\Big), \quad \text{for all } t > 0.$$

By NSP, each pseudo-community $\mathcal{N}_k^{(m,0)}$ contains at least one true community, say, $\mathcal{N}_{k^*}$. Combining it with the condition (3.2.2) gives $\mathbf{1}_k'\theta \geq \sum_{i \in \mathcal{N}_{k^*}} \theta_i \geq C\|\theta\|_1$. At the same time, $\mathbf{1}_k'\theta \leq \|\theta\|_1$. We thus have $\mathbf{1}_k'\theta \asymp \|\theta\|_1 \gg \sqrt{\log(n)}$. Similarly, we can show that $\mathbf{1}_k\Omega\mathbf{1}_\ell \asymp \|\theta\|_1^2$. In the above equation, if we choose $t = C_1\|\theta\|_1\sqrt{\log(n)}$ for a properly large constant $C_1 > 0$, then the right hand side is $O(n^{-3})$. In other words, with probability $1 - O(n^{-3})$,

$$|\mathbf{1}_k' A \mathbf{1}_\ell - \mathbf{1}_k' \Omega \mathbf{1}_\ell| \leq C\|\theta\|_1 \sqrt{\log(n)}.$$

Since $\mathbf{1}_k'\Omega\mathbf{1}_\ell \asymp \|\theta\|_1^2 \gg \|\theta\|_1\sqrt{\log(n)}$, the above implies $\mathbf{1}_k' A\mathbf{1}_\ell \asymp \|\theta\|_1^2$. We combine this result with the probability union bound. It follows that there exists a constant $C_2 > 1$ such that with probability $1 - O(n^{-3})$,

$$C_2^{-1}\|\theta\|_1^2 \leq \min_{1\leq k,\ell\leq m}\{\mathbf{1}_k' A\mathbf{1}_\ell\} \leq \max_{1\leq k,\ell\leq m}\{\mathbf{1}_k' A\mathbf{1}_\ell\} \leq C_2\|\theta\|_1^2 \tag{3.C.133}$$

We plug it into the expression of $\widehat{P}_{k\ell}$ above and can easily see that

$$\|\widehat{P}\|_{\max} \leq C, \qquad \text{with probability } 1 - O(n^{-3}). \tag{3.C.134}$$

Second, we bound $\|\widehat{\theta}\|$. By (3.2.6), $\widehat{\theta}_i = d_i\sqrt{\mathbf{1}_k' A\mathbf{1}_k}/(\mathbf{1}_k' A\mathbf{1}_n)$ for $i \in \mathcal{N}_k^{(m,0)}$. It follows that

$$\|\widehat{\theta}\|^2 = \sum_{k=1}^m \frac{(\mathbf{1}_k' D^2\mathbf{1}_k)(\mathbf{1}_k' A\mathbf{1}_k)}{(\mathbf{1}_k' A\mathbf{1}_n)^2}, \qquad \text{where} \quad D = \mathrm{d}iag(d_1, d_2, \ldots, d_n).$$

Note that $\mathbf{1}_k' A\mathbf{1}_n = \sum_{\ell=1}^m \mathbf{1}_k' A\mathbf{1}_\ell$. It follows from (3.C.133) that $\mathbf{1}_k' A\mathbf{1}_k \asymp \|\theta\|_1^2$ and $\mathbf{1}_k' A\mathbf{1}_n \asymp \|\theta\|_1^2$. As a result, $\|\widehat{\theta}\|^2 \leq C\|\theta\|_1^{-2}\sum_{k=1}^m(\mathbf{1}_k' D^2\mathbf{1}_k)$. Since $\sum_{k=1}^m(\mathbf{1}_k' D^2\mathbf{1}_k) = \|d\|^2$, we immediately have

$$\|\widehat{\theta}\|^2 \leq C\|\theta\|_1^{-2}\|d\|^2, \qquad \text{with probability } 1 - O(n^{-3}). \tag{3.C.135}$$

Recall that $d_i = \sum_{j:j\neq i} A_{ij} = \sum_{j:j\neq i}(\Omega_{ij} + W_{ij})$. Then,

$$\begin{aligned}
\|d\|^2 &= \sum_{i=1}^n \sum_{j,s:j\neq i,s\neq i} (\Omega_{ij} + W_{ij})(\Omega_{is} + W_{is}) \\
&= \sum_{i,j,s:j\neq i,s\neq i} \Omega_{ij}\Omega_{is} + 2\underbrace{\sum_{i\neq j}\Big(\sum_{s\notin\{i,j\}}\Omega_{is}\Big)W_{ij}}_{\equiv X_1} + \underbrace{\sum_{i\neq j} W_{ij}^2}_{\equiv X_2} + \underbrace{\sum_{i,j,s(dist)} W_{ij}W_{is}}_{\equiv X_3}.
\end{aligned}$$

Since $\sum_{s\notin\{i,j\}}\Omega_{is} \leq C\theta_i\|\theta\|_1$, we have $\mathbb{E}[X_1^2] \leq \sum_{i\neq j} C\theta_i^2\|\theta\|_1^2 \cdot \mathbb{E}[W_{ij}^2] \leq C\|\theta\|_3^3\|\theta\|_1^3$. Moreover, $X_2 \geq 0$ and $\mathbb{E}[X_2] = \sum_{i\neq j}\mathbb{E}[W_{ij}^2] \leq C\|\theta\|_1^2$. Last, $\mathbb{E}[X_3^2] = 2\sum_{i,j,s(dist)}\mathrm{Var}(W_{ij}W_{is}) \leq C\sum_{i,j,s}\theta_i^2\theta_j\theta_s \leq C\|\theta\|^2\|\theta\|_1^2$. By Markov's inequality, for any sequence $\epsilon_n \to 0$,

$$|X_1| \leq C\sqrt{\epsilon_n^{-1}\|\theta\|_3^3\|\theta\|_1^3}, \qquad |X_2| \leq C\epsilon_n^{-1}\|\theta\|_1^2, \qquad |X_3| \leq C\sqrt{\epsilon_n^{-1}\|\theta\|^2\|\theta\|_1^2}.$$

It is not hard to see that we can choose a property $\epsilon_n \to 0$ so that all the right hand sides are $o(\|\theta\|_1^2\|\theta\|^2)$. Then, with probability $1 - \epsilon_n$,

$$\|d\|^2 = \sum_{i,j,s:j\neq i,s\neq i}\Omega_{ij}\Omega_{is} + o(\|\theta\|_1^2\|\theta\|^2) \leq C\|\theta\|^2\|\theta\|_1^2.$$

We plug it into (3.C.135) to get

$$\|\widehat{\theta}\|^2 \leq C\|\theta\|^2, \qquad \text{with probability } 1 - o(1). \tag{3.C.136}$$

Then, (3.C.131) follows from plugging (3.C.134) and (3.C.136) into (3.C.132). This proves the claim. $\square$

3.D Proof of secondary lemmas

3.D.1 Proof of Lemma 45

Note that for any set $M \subset \{1, 2, \dots, n\}$ and $z \in \mathbb{R}^d$,

$$\begin{aligned}\sum_{i\in M} \|y_i - z\|^2 &= \sum_{i\in M} \|(y_i - \bar{y}_M) + (\bar{y}_M - z)\|^2 \\ &= \sum_{i\in M} \|y_i - \bar{y}_M\|^2 + 2(\bar{y}_M - z)' \sum_{i\in M}(y_i - \bar{y}_M) + |M|\|\bar{y}_M - z\|^2 \\ &= \sum_{i\in M} \|y_i - \bar{y}_M\|^2 + |M|\|\bar{y}_M - z\|^2.\end{aligned}$$

The clusters associated with RSS are $A = \widetilde{A} \cup C$ and B, and the clusters associated with $\widetilde{RSS}$ are $\widetilde{A}$ and $\widetilde{B} = C \cup B$. By direct calculations,

$$\begin{aligned}RSS &= \sum_{i\in \widetilde{A}} \|y_i - \bar{y}_A\|^2 + \sum_{i\in C} \|y_i - \bar{y}_A\|^2 + \sum_{i\in B} \|y_i - \bar{y}_B\|^2 \\ &= \Big(\sum_{i\in \widetilde{A}} \|y_i - \bar{y}_{\widetilde{A}}\|^2 + |\widetilde{A}|\|\bar{y}_{\widetilde{A}} - \bar{y}_A\|^2\Big) + \Big(\sum_{i\in C} (y_i - \bar{y}_C)^2 + |C|\|\bar{y}_C - \bar{y}_A\|^2\Big) + \sum_{i\in B} \|y_i - \bar{y}_B\|^2, \\ \widetilde{RSS} &= \sum_{i\in \widetilde{A}} \|y_i - \bar{y}_{\widetilde{A}}\|^2 + \sum_{i\in C} \|y_i - \bar{y}_{\widetilde{B}}\|^2 + \sum_{i\in B} \|y_i - \bar{y}_{\widetilde{B}}\|^2 \\ &= \sum_{i\in \widetilde{A}} \|y_i - \bar{y}_{\widetilde{A}}\|^2 + \Big(\sum_{i\in C} \|y_i - \bar{y}_C\|^2 + |C|\|\bar{y}_C - \bar{y}_{\widetilde{B}}\|^2\Big) + \Big(\sum_{i\in B} \|y_i - \bar{y}_B\|^2 + |B|\|\bar{y}_B - \bar{y}_{\widetilde{B}}\|\Big).\end{aligned}$$

It follows that

$$\widetilde{RSS} - RSS = \big(|B|\|\bar{y}_B - \bar{y}_{\widetilde{B}}\|^2 + |C|\|\bar{y}_C - \bar{y}_{\widetilde{B}}\|^2\big) - \big(|\widetilde{A}|\|\bar{y}_{\widetilde{A}} - \bar{y}_A\|^2 + |C|\|\bar{y}_C - \bar{y}_A\|^2\big). \tag{3.D.137}$$

By definition,

$$\bar{y}_A = \frac{|A| - |C|}{|A|}\bar{y}_{\widetilde{A}} + \frac{|C|}{|A|}\bar{y}_C, \qquad \bar{y}_{\widetilde{B}} = \frac{|B|}{|B| + |C|}\bar{y}_B + \frac{|C|}{|B| + |C|}\bar{y}_C.$$

Re-arranging the terms, we have

$$\bar{y}_{\widetilde{A}} - \bar{y}_A = \frac{|C|}{|A| - |C|}(\bar{y}_A - \bar{y}_C), \quad \bar{y}_{\widetilde{B}} - \bar{y}_B = \frac{|C|}{|B| + |C|}(\bar{y}_C - \bar{y}_B), \quad \bar{y}_C - \bar{y}_{\widetilde{B}} = \frac{|B|}{|B| + |C|}(\bar{y}_C - \bar{y}_B). \tag{3.D.138}$$

We plug (3.D.138) into (3.D.137) to get

$$\begin{aligned}\widetilde{RSS} - RSS &= \left(|B| \cdot \frac{|C|^2}{(|B| + |C|)^2} + |C| \cdot \frac{|B|^2}{(|B| + |C|)^2}\right)\|\bar{y}_C - \bar{y}_B\|^2 \\ &\quad - \left(|\widetilde{A}| \cdot \frac{|C|^2}{(|A| - |C|)^2} + |C|\right)\|\bar{y}_C - \bar{y}_A\|^2 \\ &= \frac{|B||C|}{|B| + |C|}\|\bar{y}_C - \bar{y}_B\|^2 - \frac{|A||C|}{|A| - |C|}\|\bar{y}_C - \bar{y}_A\|^2.\end{aligned}$$

This proves the claim. ☐

3.D.2 Proof of Lemma 47

Recall that $\mathbf{1}_k \in \mathbb{R}^n$ is such that $\mathbf{1}_k(i) = \{i \in \mathcal{N}_k\}$, $D = \mathrm{d}iag(d_1, d_2, \ldots, d_n)$, and $d^* = \Omega \mathbf{1}_n$. We re-state the claims as

$$\frac{\mathbf{1}_n' A \mathbf{1}_n}{\mathbf{1}_n' \Omega \mathbf{1}_n} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' A \mathbf{1}_n}{\mathbf{1}_k' \Omega \mathbf{1}_n} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' A \mathbf{1}_k}{\mathbf{1}_k' \Omega \mathbf{1}_k} \xrightarrow{p} 1. \tag{3.D.139}$$

and

$$\frac{\|\widehat{\theta}\|_1}{\|\theta\|_1} \xrightarrow{p} 1, \qquad \frac{\|\widehat{\theta}\|}{\|\theta\|} \xrightarrow{p} 1, \qquad \frac{\mathbf{1}_k' D^2 \mathbf{1}_k}{\mathbf{1}_k' (D^*)^2 \mathbf{1}_k} \xrightarrow{p} 1. \tag{3.D.140}$$

We note that convergence in ℓ^2-norm implies convergence in probability. Hence, to show $X \xrightarrow{p} 1$ for a random variable X, it is sufficient to show $\mathbb{E}[(X-1)^2] \to 0$. Using the equality $\mathbb{E}[(X-1)^2] = (\mathbb{E}X - 1)^2 + \mathrm{Var}(X)$, we only need to prove that $\mathbb{E}[X] \to 1$ and $\mathrm{Var}(X) \to 0$, for each variable X on the left hand sides of (3.D.139)-(3.D.140).

First, we prove the three claims in (3.D.139). Since the proofs are similar, we only show the proof of the first claim. Note that $\mathbf{1}_n' \Omega \mathbf{1}_n = \sum_{k,\ell} (\mathbf{1}_k' \theta)(\mathbf{1}_\ell' \theta) P_{k\ell}$. Under the conditions (3.2.1)-(3.2.2), $\mathbf{1}_n' \Omega \mathbf{1}_n \asymp \|\theta\|_1^2$. Additionally, $\mathbf{1}_n' \mathrm{d}iag(\Omega) \mathbf{1}_n = \|\theta\|^2$. It follows that

$$\left| \frac{\mathbb{E}[\mathbf{1}_n' A \mathbf{1}_n]}{\mathbf{1}_n' \Omega \mathbf{1}_n} - 1 \right| = \frac{\mathbf{1}_n' \mathrm{d}iag(\Omega) \mathbf{1}_n}{\mathbf{1}_n' \Omega \mathbf{1}_n} \asymp \frac{\|\theta\|^2}{\|\theta\|_1^2} = o(1),$$

where the last inequality is because $\|\theta\|^2 \leq \theta_{\max} \|\theta\|_1 \leq C\|\theta\|_1$ and $\|\theta\|_1 \to \infty$. Also, since the upper triangular entries of A are independent, $\mathrm{Var}(\mathbf{1}_n' A \mathbf{1}_n) = 4\mathrm{Var}(\sum_{i<j} A_{ij}) \leq 4 \sum_{i<j} \Omega_{ij} \leq C\|\theta\|_1^2$. It follows that

$$\frac{\mathrm{Var}(\mathbf{1}_n' A \mathbf{1}_n)}{(\mathbf{1}_n' \Omega \mathbf{1}_n)^2} \leq \frac{C\|\theta\|_1^2}{\|\theta\|_1^4} = o(1).$$

Combining the above gives $(\mathbf{1}_n' A \mathbf{1}_n)/(\mathbf{1}_n' \Omega \mathbf{1}_n) \xrightarrow{p} 1$.

Second, we show the first claim in (3.D.140). By Theorem 3.2.2, $\widehat{\Pi}^{(K)} = \Pi$, with a probability of $1 - O(n^{-3})$. It is sufficient to consider the re-defined $\widehat{\theta}$ where $\widehat{\Pi}^{(K)}$ is replaced with Π. Combining it with the definition in (3.2.6), we have $\widehat{\theta}_i = d_i \sqrt{\mathbf{1}_k' A \mathbf{1}_k} / (\mathbf{1}_k' A \mathbf{1}_n)$. It follows that

$$\|\widehat{\theta}\|_1 = \sum_{k=1}^K \frac{(\mathbf{1}_k' d)\sqrt{\mathbf{1}_k' A \mathbf{1}_k}}{\mathbf{1}_k' A \mathbf{1}_n} = \sum_{k=1}^K \sqrt{\mathbf{1}_k' A \mathbf{1}_k},$$

where the last equality is because of $d = A\mathbf{1}_n$. At the same time, it is easy to see that $\mathbf{1}_k' \Omega \mathbf{1}_k = (\mathbf{1}_k' \theta) P_{kk} (\mathbf{1}_k' \theta) = (\mathbf{1}_k' \theta)^2$, which implies $\|\theta\|_1 = \sum_{k=1}^K \sqrt{\mathbf{1}_k' \Omega \mathbf{1}_k}$. We thus have

$$\frac{\|\widehat{\theta}\|_1}{\|\theta\|_1} = \sum_{k=1}^K \delta_k X_k, \qquad \text{where} \quad \delta_k = \frac{\sqrt{\mathbf{1}_k' \Omega \mathbf{1}_k}}{\sum_{\ell=1}^K \sqrt{\mathbf{1}_\ell' \Omega \mathbf{1}_\ell}}, \quad X_k = \sqrt{\frac{\mathbf{1}_k' A \mathbf{1}_k}{\mathbf{1}_k' \Omega \mathbf{1}_k}}.$$

By the last claim in (3.D.139) and the continuous mapping theorem, $X_k \xrightarrow{p} 1$ for each $1 \leq k \leq K$. Also, $\sum_{k=1}^K \delta_k = 1$. It follows immediately that $\sum_{k=1}^K \delta_k X_k \xrightarrow{p} 1$. This proves $\|\widehat{\theta}\|_1 / \|\theta\|_1 \xrightarrow{p} 1$.

Next, we show the last claim in (3.D.140). Recall that $d^* = \Omega \mathbf{1}_n$ and $D^* = \mathrm{d}iag(d^*)$. Then, for $i \in \mathcal{N}_k$, $\sum_{i \in \mathcal{N}_k} (d_i^*)^2 \leq C \sum_{i \in \mathcal{N}_k} (\theta_i \|\theta\|_1)^2 \leq C\|\theta\|^2 \|\theta\|_1^2$. At the same time, $d_i^* \geq \theta_i P_{kk} (\mathbf{1}_k' \theta) \geq C\theta_i \|\theta\|_1$, where we have used the condition (3.2.2). As a result, $\sum_{i \in \mathcal{N}_k} (d_i^*)^2 \geq$

$C\|\theta\|_1^2\sum_{i\in\mathcal{N}_k}\theta_i^2 \geq C\|\theta\|^2\|\theta\|_1^2$, where we have used (3.2.2) again. Combining the above gives

$$\mathbf{1}_k'(D^*)^2\mathbf{1}_k \asymp \|\theta\|^2\|\theta\|_1^2. \tag{3.D.141}$$

Note that $\mathbf{1}_k'D^2\mathbf{1}_k = \sum_{t\in\mathcal{N}_k}(\sum_{i:i\neq t}A_{it})^2 = \sum_{i,j}\sum_{t\in\mathcal{N}_k\setminus\{i,j\}}A_{it}A_{jt}$. Similarly, $\mathbf{1}_k'(D^*)^2\mathbf{1}_k = \sum_{i,j}\sum_{t\in\mathcal{N}_k}\Omega_{it}\Omega_{jt}$. We now write

$$\mathbf{1}_k'D^2\mathbf{1}_k = \sum_i\sum_{t\in\mathcal{N}_k\setminus\{i\}}A_{it}^2 + 2\sum_{i<j}\sum_{t\in\mathcal{N}_k\setminus\{i,j\}}A_{it}A_{jt},$$
$$\mathbf{1}_k'(D^*)^2\mathbf{1}_k = \sum_i\sum_{t\in\mathcal{N}_k}\Omega_{it}^2 + 2\sum_{i<j}\sum_{t\in\mathcal{N}_k}\Omega_{it}\Omega_{jt}.$$

Note that $\mathbb{E}[A_{it}^2] = \mathbb{E}[A_{it}] = \Omega_{it}$ and $\mathbb{E}[A_{it}A_{jt}] = \Omega_{it}\Omega_{jt}$. As a result,

$$\begin{aligned}\left|\mathbb{E}[\mathbf{1}_k'D^2\mathbf{1}_k] - \mathbf{1}_k'(D^*)^2\mathbf{1}_k\right| &\leq \sum_i\sum_{t\in\mathcal{N}_k\setminus\{i\}}(\Omega_{it}-\Omega_{it}^2) + \sum_i\Omega_{ii}^2 + 2\sum_{i<j}(\Omega_{ii}\Omega_{ji}+\Omega_{ij}\Omega_{jj})\\ &\leq C\sum_i\sum_{t\in\mathcal{N}_k}\theta_i\theta_t + \|\theta\|^2 + C\sum_{i,j}\theta_i^3\theta_j\\ &\leq C\big(\|\theta\|_1^2 + \|\theta\|^2 + \|\theta\|_3^3\|\theta\|_1\big)\\ &\leq C\|\theta\|_1^2,\end{aligned}$$

where the last line is because $\|\theta\|_3^3 \leq \theta_{\max}^2\|\theta\|_1 \leq C\|\theta\|_1$. Combining it with (3.D.141) gives

$$\left|\frac{\mathbb{E}[\mathbf{1}_k'D^2\mathbf{1}_k]}{\mathbf{1}_k'(D^*)^2\mathbf{1}_k} - 1\right| \leq \frac{C\|\theta\|_1^2}{\|\theta\|^2\|\theta\|_1^2} = o(1). \tag{3.D.142}$$

We then compute the variance. Write for short $X = \sum_{i<j}\sum_{t\in\mathcal{N}_k\setminus\{i,j\}}A_{it}A_{jt}$. Note that

$$\begin{aligned}\mathrm{Var}(\mathbf{1}_k'D^2\mathbf{1}_k) &\leq 2\mathrm{Var}\Big(\sum_i\sum_{t\in\mathcal{N}_k\setminus\{i\}}A_{it}^2\Big) + 2\mathrm{Var}(2X)\\ &\leq C\sum_i\sum_{t\in\mathcal{N}_k}\Omega_{it} + 8\mathrm{Var}(X)\\ &\leq C\|\theta\|_1^2 + 8\mathrm{Var}(X).\end{aligned}$$

Since $A_{it}A_{jt} = (\Omega_{it}+W_{it})(\Omega_{jt}+W_{jt})$, we write

$$\begin{aligned}X &= \sum_{i<j}\sum_{t\in\mathcal{N}_k\setminus\{i,j\}}\Omega_{it}\Omega_{jt} + 2\sum_j\sum_{t\in\mathcal{N}_k\setminus\{j\}}\Big(\sum_{i:i\neq t,i<j}\Omega_{it}\Big)W_{jt} + \sum_{i<j}\sum_{t\in\mathcal{N}_k\setminus\{i,j\}}W_{it}W_{jt}\\ &\equiv X_0 + 2X_1 + X_2.\end{aligned}$$

Here, X_0 is non-stochastic. Therefore, $\mathrm{Var}(X) = \mathrm{Var}(2X_1+X_2) \leq 8\mathrm{Var}(X_1) + 2\mathrm{Var}(X_2)$. It is seen that $\mathrm{Var}(X_1) \leq \sum_j\sum_{t\in\mathcal{N}_k}(\sum_i\Omega_{it})^2\cdot\Omega_{jt} \leq C\sum_j\sum_{t\in\mathcal{N}_k}(\theta_t\|\theta\|_1)^2\cdot\theta_j\theta_t \leq C\|\theta\|_3^3\|\theta\|_1^3$. Additionally, the summands in X_2 are mutually uncorrelated, so $\mathrm{Var}(X_3) \leq \sum_{i<j}\sum_{t\in\mathcal{N}_k}\Omega_{it}\Omega_{jt} \leq C\sum_{i,j,t}\theta_i\theta_j\theta_t^2 \leq C\|\theta\|_1^2\|\theta\|^2$. Combining the above gives

$$\mathrm{Var}(X) \leq C\big(\|\theta\|_3^3\|\theta\|_1^3 + \|\theta\|_1^2\|\theta\|^2\big) \leq C\|\theta\|_3^3\|\theta\|_1^3,$$

where in the second inequality we have used $\|\theta\|^2 \leq \|\theta\|_1\|\theta\|_3^3$, which is a direct consequence of the Cauchy-Schwarz inequality. We combine the above to get

$$\mathrm{Var}(\mathbf{1}_k'D^2\mathbf{1}_k) \leq C\big(\|\theta\|_1^2 + \|\theta\|_3^3\|\theta\|_1^3\big) \leq C(\|\theta\|_1^2 + \theta_{\max}\|\theta\|^2\|\theta\|_1^3),$$

where in the second inequality we have used $\|\theta\|_3^3 \leq \theta_{\max}\|\theta\|^2$. Combining it with (3.D.141) gives

$$\frac{\mathrm{Var}(\mathbf{1}_k' D^2 \mathbf{1}_k)}{[\mathbf{1}_k'(D^*)^2\mathbf{1}_k]^2} \leq \frac{C\|\theta\|_1^2}{\|\theta\|^4\|\theta\|_1^4} + \frac{C\theta_{\max}\|\theta\|^2\|\theta\|_1^3}{\|\theta\|^4\|\theta\|_1^4} = o(1). \tag{3.D.143}$$

By (3.D.142) and (3.D.143), we have $(\mathbf{1}_k' D^2\mathbf{1}_k)/[\mathbf{1}_k'(D^*)^2\mathbf{1}_k] \xrightarrow{p} 1$.

Last, we show the second claim in (3.D.140). Since $\widehat{\theta}_i = d_i\sqrt{\mathbf{1}_k' A\mathbf{1}_k}/(\mathbf{1}_k' A\mathbf{1}_n)$, we have

$$\|\widehat{\theta}\|^2 = \sum_{k=1}^K \frac{(\mathbf{1}_k' D^2\mathbf{1}_k)(\mathbf{1}_k' A\mathbf{1}_k)}{(\mathbf{1}_k' A\mathbf{1}_n)^2}.$$

At the same time, $\mathbf{1}_k'\Omega\mathbf{1}_k = (\mathbf{1}_k'\theta)^2$ and $\mathbf{1}_k'\Omega\mathbf{1}_n = (\mathbf{1}_k'\theta)[\sum_{\ell=1}^K P_{k\ell}(\mathbf{1}_\ell'\theta)]$. Furthermore, for $i\in\mathcal{N}_k$, $d_i^* = (\Omega\mathbf{1}_n)_i = \theta_i[\sum_{\ell=1}^K P_{k\ell}(\mathbf{1}_\ell'\theta)]$, and so $\mathbf{1}_k'(D^*)^2\mathbf{1}_n = (\mathbf{1}_k'\Theta^2\mathbf{1}_k)[\sum_{\ell=1}^K P_{k\ell}(\mathbf{1}_\ell'\theta)]^2$. Combining these equalities gives

$$\|\theta\|^2 = \sum_{k=1}^K \mathbf{1}_k'\Theta^2\mathbf{1}_k = \sum_{k=1}^K \frac{[\mathbf{1}_k'(D^*)^2\mathbf{1}_k](\mathbf{1}_k'\Omega\mathbf{1}_k)}{(\mathbf{1}_k'\Omega\mathbf{1}_n)^2}.$$

It follows that

$$\frac{\|\widehat{\theta}\|^2}{\|\theta\|^2} = \sum_{k=1}^K \widetilde{\delta}_k\widetilde{X}_k, \quad \text{where } \widetilde{\delta}_k = \frac{\frac{[\mathbf{1}_k'(D^*)^2\mathbf{1}_k](\mathbf{1}_k'\Omega\mathbf{1}_k)}{(\mathbf{1}_k'\Omega\mathbf{1}_n)^2}}{\sum_{\ell=1}^K\frac{[\mathbf{1}_\ell'(D^*)^2\mathbf{1}_\ell](\mathbf{1}_\ell'\Omega\mathbf{1}_\ell)}{(\mathbf{1}_\ell'\Omega\mathbf{1}_n)^2}},\quad \widetilde{X}_k = \frac{\mathbf{1}_k'D^2\mathbf{1}_k}{\mathbf{1}_k'(D^*)^2\mathbf{1}_k}\frac{\mathbf{1}_k'A\mathbf{1}_k}{\mathbf{1}_k'\Omega\mathbf{1}_k}\frac{(\mathbf{1}_k'\Omega\mathbf{1}_n)^2}{(\mathbf{1}_k'A\mathbf{1}_n)^2}.$$

By the claims in (3.D.139) and the last claim in (3.D.140), as well as the continuous mapping theorem, we have $\widetilde{X}_k \xrightarrow{p} 1$ for each $1\leq k\leq K$. Since $\sum_{k=1}^K\widetilde{\delta}_k = 1$, it follows that $\sum_{k=1}^K\widetilde{\delta}_k\widetilde{X}_k \xrightarrow{p} 1$. This proves that $\|\widehat{\theta}\|^2/\|\theta\|^2 \xrightarrow{p} 1$. By the continuous mapping theorem again, $\|\widehat{\theta}\|/\|\theta\| \xrightarrow{p} 1$. □

3.D.3 Proof of Lemma 49

We introduce a notation $M_{ijk\ell}(X) = X_{ij}X_{jk}X_{k\ell}X_{\ell i}$, for any symmetric $n\times n$ matrix X and distinct indices (i,j,k,ℓ). Using the definition in (3.C.130), we can write

$$\begin{aligned} &Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)} \\ &= \sum_{i_1,i_2,i_3,i_4(dist)}[M_{i_1i_2i_3i_4}(X^*) - M_{i_1i_2i_3i_4}(\widetilde{X})], \qquad \text{where } \begin{cases} X_{ij}^* = \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)}, \\ \widetilde{X}_{ij} = \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij}. \end{cases}\end{aligned}$$

For the rest of the proof, we omit superscripts in $\widetilde{\Omega}_{ij}^{(m,0)}$ and $\delta_{ij}^{(m,0)}$ to simplify notations. From the expression of X_{ij}^* and $\widetilde{X}_{ij}$, we notice that $[M_{i_1i_2i_3i_4}(X^*) - M_{i_1i_2i_3i_4}(\widetilde{X})]$ expands to $3^4 - 2^4 = 65$ terms. Consequently, there are 65 post-expansion sums in $Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}$, each with the form

$$\sum_{i_1,i_2,i_3,i_4(dist)} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \qquad \text{where} \quad a,b,c,d\in\{\widetilde{\Omega}, W, \delta\}.$$

In the first 4 columns of Table 3.3, we group these post-expansion sums into 15 distinct terms, where the second column shows the counts of each distinct term. For example, in the setting of Lemma 38, $\widetilde{\Omega}$ reduces to a zero matrix. Therefore, any post-expansion sum that involves $\widetilde{\Omega}$ is zero. Then, it follows from Table 3.3 that

$$Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)} = 4Y_1 + 4Z_1 + 2Z_2 + 4T_1 + F, \tag{3.D.144}$$

where the expression of (Y_1, Z_1, Z_2, T_1, F) are given in the fourth column of Table 3.3. Similarly, in the setting of Lemma 43, we have $Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)} = 4Y_1 + 8Y_2 + 4Y_3 + \cdots + 4T_2 + F$. These are elementary calculations.

Table 3.3: The 10 types of the post-expansion sums for $(Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)})$. Notations: same as in Table 3.2.

Type	#	Name	Formula	Abs. Mean	Variance
Ia	4	Y_1	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} W_{i_2i_3} W_{i_3i_4} W_{i_4i_1}$	0	$\leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
Ib	8	Y_2	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{i_2i_3} W_{i_3i_4} W_{i_4i_1}$	0	$\leq C\tau^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4	Y_3	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} W_{i_2i_3} \widetilde{\Omega}_{i_3i_4} W_{i_4i_1}$	0	$\leq C\tau^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
Ic	8	Y_4	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{i_2i_3} \widetilde{\Omega}_{i_3i_4} W_{i_4i_1}$	$\leq C\tau^2\|\theta\|^6 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\tau^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = o(\tau^6\|\theta\|^8\|\theta\|_3^6)$
	4	Y_5	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{i_2i_3} W_{i_3i_4} \widetilde{\Omega}_{i_4i_1}$	0	$\leq \frac{C\tau^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1} = o(\|\theta\|^8)$
Id	4	Y_6	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{i_2i_3} \widetilde{\Omega}_{i_3i_4} \widetilde{\Omega}_{i_4i_1}$	0	$\leq \frac{C\tau^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1} = O(\tau^6\|\theta\|^8\|\theta\|_3^6)$
IIa	4	Z_1	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} W_{i_3i_4} W_{i_4i_1}$	$\leq C\|\theta\|^4 = o(\tau^4\|\theta\|^8)$	$\leq C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
	2	Z_2	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} W_{i_2i_3} \delta_{i_3i_4} W_{i_4i_1}$	$\leq C\|\theta\|^4 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIb	8	Z_3	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} \widetilde{\Omega}_{i_3i_4} W_{i_4i_1}$	0	$\leq C\tau^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4	Z_4	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{jk} \delta_{i_3i_4} W_{i_4i_1}$	$\leq C\tau\|\theta\|^4 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\tau^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIc	4	Z_5	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} \widetilde{\Omega}_{i_3i_4} \widetilde{\Omega}_{i_4i_1}$	$\leq C\tau^2\|\theta\|^6 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\tau^4\|\theta\|^{14}}{\|\theta\|_1^2} = o(\tau^6\|\theta\|^8\|\theta\|_3^6)$
	2	Z_6	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \widetilde{\Omega}_{i_2i_3} \delta_{i_3i_4} \widetilde{\Omega}_{i_4i_1}$	$\frac{C\tau^2\|\theta\|^8}{\|\theta\|_1^2} = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\tau^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8)$
IIIa	4	T_1	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} \delta_{i_3i_4} W_{i_4i_1}$	$\leq C\|\theta\|^4 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIIb	4	T_2	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} \delta_{i_3i_4} \widetilde{\Omega}_{i_4i_1}$	$\leq \frac{C\tau\|\theta\|^6}{\|\theta\|_1^3} = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\tau^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IV	1	F	$\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} \delta_{i_1i_2} \delta_{i_2i_3} \delta_{i_3i_4} \delta_{i_4i_1}$	$\leq C\|\theta\|^4 = o(\tau^4\|\theta\|^8)$	$\leq \frac{C\|\theta\|^{10}}{\|\theta\|_1^2} = o(\|\theta\|^8)$

To show the claim, we need to study the mean and variance of each post-expansion sum. We take Y_1 for example. Let $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, \ldots, \mathcal{N}_m^{(m,0)}$ be the pseudo-communities defined by Π_0. For each $1 \leq i \leq n$, let $\tau(i) \in \{1, 2, \ldots, m\}$ be the index of the pseudo-community that contains node i. By (3.C.125),

$$\begin{aligned}
\delta_{i_1i_2} &= U_{\tau(i_1)\tau(i_2)}\big[(\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2}) + (\mathbb{E}d_{i_2})(\mathbb{E}d_{i_1} - d_{i_1})\big] \\
&= U_{\tau(i_1)\tau(i_2)} \cdot \mathbb{E}d_{i_1} \cdot \Big(-\sum_{j:j\neq i_2} W_{ji_2}\Big) + U_{\tau(i_1)\tau(i_2)} \cdot \mathbb{E}d_{i_2} \cdot \Big(-\sum_{\ell:\ell\neq i_1} W_{\ell i_1}\Big) \\
&= -2\sum_{j:j\neq i_2} U_{\tau(i_1)\tau(i_2)} \cdot \mathbb{E}d_{i_1} \cdot W_{ji_2}. \qquad (3.D.145)
\end{aligned}$$

It follows that

$$Y_1 = -2\sum_{i_2,i_3,i_4,j}\Big(\sum_{i_1} U_{\tau(i_1)\tau(i_2)} \cdot \mathbb{E}d_{i_1}\Big) \cdot W_{ji_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1},$$

where we note that the indices $\{i_1, i_2, i_3, i_4, j\}$ have to satisfy the constraint that i_1, i_2, i_3, i_4 are distinct and that $j \neq i_2$. We can see that Y_1 is a weighted sum of $W_{ji_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$, where the summands have zero mean and are mutually uncorrelated. The mean and variance of Y_1 can be calculated easily. We will use the same strategy to analyze each term in Table 3.3— we use the expansion of δ_{ij} in (3.D.145) to write each post-expansion sum as a weighted sum of monomials of W, and then we calculate the mean and variance. The calculations can become very tedious for some terms (e.g., T_1, T_2 and F), because of combinatorics. Fortunately, similar calculations were done in the proof of Theorem 4.4 in Jin et al. (2019), where they analyzed a special case with $U_{k\ell} \equiv 1/v$ for all $1 \le k, \ell \le m$. However, their proof does not rely on that $U_{k\ell}$'s are equal but only require that $U_{k\ell}$'s have a uniform upper bound. Essentially, they have proved the following lemma:

Lemma 52. *Consider a DCBM model where (3.2.1)-(3.2.2) and (3.2.4) hold. Let $W = A-\Omega$ and $\Delta = \sum_{i_1,i_2,i_3,i_4(dist)}\big[M_{i_1i_2i_3i_4}\big(\widetilde{\Omega}+W+\delta\big)-M_{i_1i_2i_3i_4}\big(\widetilde{\Omega}+W\big)\big]$, where $\widetilde{\Omega}$ is a non-stochastic symmetric matrix, $\delta_{ij} = v_{ij}\cdot[(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$, $\{v_{ij}\}_{1\le i\ne j\le n}$ are non-stochastic scalars, d_i is the degree of node i, and $M_{i_1i_2i_3i_4}(\cdot)$ is as defined above. As $n\to\infty$, suppose there is a constant $C>0$ and a scalar $\alpha_n>0$ such that $\alpha_n\le C$, $\alpha_n\|\theta\|\to\infty$, $|\widetilde{\Omega}_{ij}|\le C\alpha_n\theta_i\theta_j$ and $|v_{ij}|\le C\|\theta\|_1^{-1}$ for all $1\le i,j\le n$. Then, $|\mathbb{E}[\Delta]| = o(\alpha_n^4\|\theta\|^8)$ and $\mathrm{Var}(\Delta)\le C\alpha_n^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8)$. Furthermore, if $\widetilde{\Omega}$ is a zero matrix, then $|\mathbb{E}[\Delta]|\le C\|\theta\|^4$ and $\mathrm{Var}(\Delta) = o(\|\theta\|^8)$.*

We check the conditions of Lemma 52. By Lemma 41, $\tau\le C$, $\tau\|\theta\|\to\infty$, and $|\widetilde{\Omega}_{ij}|\le C\tau\theta_i\theta_j$. We now verify that $U_{k\ell}$ has a uniform upper bound for all $1\le k,\ell\le m$. By (3.C.123),

$$U_{k\ell} = (\mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell)/[(\mathbf{1}_k'\mathbb{E}[d])(\mathbf{1}_\ell'\mathbb{E}[d])].$$

where $\mathbf{1}_k = \mathbf{1}_k^{(m,0)}$ is the same as in (3.C.108). Since $\mathbb{E}[A_{ij}] = \Omega_{ij}\le C\theta_i\theta_j$, we have $0\le \mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell\le C\|\theta\|_1^2$. At the same time, by the NSP of SCORE, for each $1\le k\le m$, there is at least one true community $\mathcal{N}_{k^*}$ such that $\mathcal{N}_{k^*}\subset\mathcal{N}_k^{(m,0)}$. It follows that $\mathbf{1}_k'\mathbb{E}[d] = \sum_{i\in\mathcal{N}_k^{(m,0)}}\sum_{j:j\ne i}\Omega_{ij}\ge\sum_{\{i,j\}\subset\mathcal{N}_{k^*},i\ne j}\theta_i\theta_jP_{kk} = \|\theta^{(k)}\|_1^2[1+o(1)]\ge C\|\theta\|_1^2$, where the last inequality is from the condition (3.2.2). We plug these results into $U_{k\ell}$ to get

$$0\le U_{k\ell}\le C\|\theta\|_1^{-2}. \tag{3.D.146}$$

Then, the conditions of Lemma 52 are satisfied. We apply this lemma with $\alpha_n=\tau$ and $v_{ij} = U_{k\ell}$ for $i\in\mathcal{N}_k^{(m,0)}$ and $j\in\mathcal{N}_\ell^{(m,0)}$. It yields that, under the conditions of Lemma 43,

$$\big|\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}]\big| = o(\tau^4\|\theta\|^8), \qquad \mathrm{Var}\big(Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}\big)\le C\tau^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8),$$

and that under the conditions of Lemma 38 (where $\widetilde{\Omega}$ is a zero matrix)

$$\big|\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}]\big|\le C\|\theta\|^4, \qquad \mathrm{Var}\big(Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}\big)\le o(\|\theta\|^8).$$

This proves all the desirable claims except for the following one: Under conditions of Lemma 38. It remains to show that, under the conditions of Lemma 38,

$$\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}] = b_n + o(\|\theta\|^4). \tag{3.D.147}$$

We now show (3.D.147). By (3.D.144), we only need to calculate the expectations of Y_1, Z_1, Z_2, T_1 and F. From Table 3.3, $\mathbb{E}[Y_1] = 0$. We now study $\mathbb{E}[Z_1]$. Recall that $\delta_{ij} = U_{\tau(i)\tau(j)}[(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$, where $\tau(i)$ is the index of pseudo-community defined by Π_0 that contains node i. We plug δ_{ij} into Z_1, by elementary calculations,

$$
\begin{aligned}
Z_1 = & \sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2})^2 (\mathbb{E}d_{i_3}) W_{i_3 i_4} W_{i_4 i_1} \\
& + 2 \sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2})(\mathbb{E}d_{i_2})(\mathbb{E}d_{i_3} - d_{i_3}) W_{i_3 i_4} W_{i_4 i_1} \\
& + \sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} (\mathbb{E}d_{i_1} - d_{i_1})(\mathbb{E}d_{i_2})^2 (\mathbb{E}d_{i_3} - d_{i_3}) W_{i_3 i_4} W_{i_4 i_1}.
\end{aligned}
$$

We write it as $Z_1 = Z_{11} + 2Z_{12} + Z_{13}$. For Z_{1k}, we can further replace $\mathbb{E}d_i - d_i$ by $\sum_{j:j\neq i} W_{ji}$ and write Z_{1k} as a weighted sum of monomials of W. Then, $\mathbb{E}[Z_{1k}] \neq 0$ if some of the monomials are $W_{i_3 i_4}^2 W_{i_4 i_1}^2$. This will not happen in Z_{11} and Z_{12}, and so only Z_{13} has a nonzero mean. It is seen that

$$
\begin{aligned}
\mathbb{E}[Z_{13}] &= \mathbb{E}\Big[\sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} \Big(\sum_{j:j\neq i_1} W_{j i_1} \Big) (\mathbb{E}d_{i_2})^2 \Big(\sum_{k:k\neq i_3} W_{i_3 k} \Big) W_{i_3 i_4} W_{i_4 i_1} \Big] \\
&= \mathbb{E}\Big[\sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} \big(W_{i_4 i_1} \big) (\mathbb{E}d_{i_2})^2 (W_{i_3 i_4}) \cdot W_{i_3 i_4} W_{i_4 i_1} \Big] \\
&= \sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} (\mathbb{E}d_{i_2})^2 \cdot \mathbb{E}[W_{i_3 i_4}^2 W_{i_4 i_1}^2] \\
&= \sum_{k_1,k_2,k_3,k_4} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1 k_2} U_{k_2 k_3} (\mathbb{E}d_{i_2})^2 \cdot \mathbb{E}[W_{i_3 i_4}^2 W_{i_4 i_1}^2].
\end{aligned}
\tag{3.D.148}
$$

Here, in the second line, we only keep $(j,k) = (i_4, i_4)$, because other (j,k) only contribute zero means. Recall that we are considering the setting of Lemma 38, where $m = K$ and $\Pi_0 = \Pi$. In (3.C.122), we introduce a proxy of $U_{k\ell}$ as $U_{k\ell}^* = (\mathbf{1}_k' \Omega \mathbf{1}_\ell)/[(\mathbf{1}_k' \Omega \mathbf{1}_n)(\mathbf{1}_k' \Omega \mathbf{1}_n)]$, for all $1 \leq k, \ell \leq K$. Note that $\Omega_{ij} = \theta_i \theta_j P_{k\ell}$ for $i \in \mathcal{N}_k$ and $j \in \mathcal{N}_\ell$. At the same time, by (3.4.21), $g_k = (\mathbf{1}_k'\theta)/\|\theta\|_1$, and $V_{kk} = (\mathrm{d}iag(Pg))_{kk} = [\sum_\ell P_{k\ell}(\mathbf{1}_\ell'\theta)]/\|\theta\|_1$. It follows that

$$
U_{k\ell}^* = \frac{P_{k\ell}(\mathbf{1}_k'\theta)(\mathbf{1}_\ell'\theta)}{(\mathbf{1}_k'\theta)[\sum_{k_1} P_{kk_1}(\mathbf{1}_{k_1}'\theta)] \cdot (\mathbf{1}_\ell'\theta)[\sum_{\ell_1} P_{\ell\ell_1}(\mathbf{1}_{\ell_1}'\theta)]} = \frac{P_{k\ell}}{V_{kk} V_{\ell\ell} \|\theta\|_1^2}.
$$

Comparing $U_{k\ell}$ with $U_{k\ell}^*$ (see (3.C.122)-(3.C.123)), the difference is negligible. (We can rigorously justify this by directly computing the difference caused by replacing $U_{k\ell}$ with $U_{k\ell}^*$, similarly as in the proof of $c_n = \mathrm{tr}(\widetilde{\Omega}^4) + o(\|\theta\|^8)$ in Section 3.C.1; see details therein. Such calculations are too elementary and so omitted.) We thus have

$$
U_{k\ell} = [1 + o(1)] \cdot \frac{P_{k\ell}}{V_{kk} V_{\ell\ell} \|\theta\|_1^2}. \tag{3.D.149}
$$

Furthermore, for $i \in \mathcal{N}_k$,

$$
\mathbb{E}[d_i] = [1 + o(1)] \sum_{j=1}^{n} \Omega_{ij} = [1 + o(1)] \cdot \theta_i \Big[\sum_{\ell=1}^{K} P_{k\ell}(\mathbf{1}_\ell'\theta) \Big] = [1 + o(1)] \cdot \theta_i \|\theta\|_1 V_{kk}. \tag{3.D.150}
$$

Also, $\mathbb{E}[W_{ij}^2] = \Omega_{ij}(1-\Omega_{ij}) = \Omega_{ij}[1+o(1)]$. We plug these results into (3.D.148) to get

$$\begin{aligned}
\mathbb{E}[Z_{13}] &= [1+o(1)] \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} \frac{P_{k_1k_2}P_{k_2k_3}}{V_{k_1k_1}V_{k_2k_2}^2 V_{k_3k_3}\|\theta\|_1^4} \cdot \left(\theta_{i_2}^2 \|\theta\|_1^2 V_{k_2k_2}^2\right) \cdot \Omega_{i_3i_4}\Omega_{i_4i_1} \\
&= [1+o(1)] \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \frac{P_{k_1k_2}P_{k_2k_3}P_{k_3k_4}P_{k_4k_1}}{V_{k_1k_1}V_{k_3k_3}\|\theta\|_1^2} \Big(\sum_{i_j \in \mathcal{N}_{k_j}} \sum_{j=1}^{4} \theta_{i_1}\theta_{i_2}^2\theta_{i_3}\theta_{i_4}^2 \Big) \\
&= [1+o(1)] \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \frac{P_{k_1k_2}P_{k_2k_3}P_{k_3k_4}P_{k_4k_1}}{V_{k_1k_1}V_{k_3k_3}\|\theta\|_1^2} \left(\|\theta\|^4 \|\theta\|_1^2 \cdot g_{k_1} g_{k_3} H_{k_2k_2}^2 H_{k_4k_4}^2 \right) \\
&= [1+o(1)]\|\theta\|^4 \sum_{k_1,k_3} \frac{g_{k_1}}{V_{k_1k_1}} \Big(\sum_{k_2} P_{k_1k_2} H_{k_2k_2}^2 P_{k_2k_3} \Big) \Big(\sum_{k_4} P_{k_3k_4} H_{k_4k_4}^2 P_{k_4k_1} \Big) \frac{g_{k_3}}{V_{k_3k_3}} \\
&= [1+o(1)]\|\theta\|^4 \sum_{k_1,k_3} (V^{-1}g)_{k_1} (PH^2P)_{k_1k_3} (PH^2P)_{k_3k_1} (V^{-1}g)_{k_3} \\
&= [1+o(1)]\|\theta\|^4 \cdot g'V^{-1}[(PH^2P)\circ(PH^2P)]V^{-1}g \\
&= [1+o(1)] \cdot b_n/2,
\end{aligned}$$

where in the third line we have used the definition of H which gives $H_{kk} = (\mathbf{1}_k'\Theta^2\mathbf{1}_k)^{1/2}/\|\theta\|$. It follows that

$$\mathbb{E}[Z_1] = \mathbb{E}[Z_{13}] = [1+o(1)] \cdot b_n/2. \tag{3.D.151}$$

We then study $\mathbb{E}[Z_2]$. Similarly, we first plug in $\delta_{ij} = U_{\tau(i)\tau(j)}[(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$ and then plug in $d_i - \mathbb{E}d_i = \sum_{j\neq i} W_{ij}$. This allows us to write Z_2 as a weighted sum of monomials of W. When calculating $\mathbb{E}[Z_2]$, we only keep monomials of the form $W_{i_1i_4}^2 W_{i_2i_3}^2$. It follows that

$$\begin{aligned}
\mathbb{E}[Z_2] &= \mathbb{E}\Bigg[2 \sum_{i_1,i_2,i_3,i_4 (dist)} U_{\tau(i_1)\tau(i_2)} (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2}) W_{i_2i_3} U_{\tau(i_3)\tau(i_4)} (\mathbb{E}d_{i_3})(\mathbb{E}d_{i_4} - d_{i_4}) W_{i_4i_1} \Bigg] \\
&= \mathbb{E}\Bigg[2 \sum_{i_1,i_2,i_3,i_4 (dist)} U_{\tau(i_1)\tau(i_2)} (\mathbb{E}d_{i_1}) W_{i_2i_3}^2 U_{\tau(i_3)\tau(i_4)} (\mathbb{E}d_{i_3}) W_{i_4i_1}^2 \Bigg] \\
&= 2 \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_j} U_{k_1k_2} U_{k_3k_4} (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_3}) W_{i_2i_3}^2 W_{i_1i_4}^2 \\
&= 2\,[1+o(1)] \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_j} \frac{P_{k_1k_2}P_{k_3k_4}}{V_{k_1k_1}V_{k_2k_2}V_{k_3k_3}V_{k_4k_4}\|\theta\|_1^4} \left(\theta_{i_1}\theta_{i_3} \|\theta\|_1^2 V_{k_1k_1} V_{k_3k_3} \right) \cdot \Omega_{i_2i_3}\Omega_{i_1i_4} \\
&= 2\,[1+o(1)] \sum_{\substack{k_1,k_2,\\ k_3,k_4}} \frac{P_{k_1k_2}P_{k_3k_4}P_{k_2k_3}P_{k_1k_4}}{V_{k_2k_2}V_{k_4k_4}\|\theta\|_1^2} \Big(\sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_j} \theta_{i_1}^2 \theta_{i_2} \theta_{i_3}^2 \theta_{i_4} \Big) \\
&= [1+o(1)] \cdot 2\|\theta\|^4 g'V^{-1}[(PH^2P)\circ(PH^2P)]V^{-1}g.
\end{aligned}$$

Here, the first two lines come from discarding terms with mean zero, the fourth line is because of (3.D.149)-(3.D.150), and the last line is obtained similarly as in the equation above (3.D.151). Hence,

$$\mathbb{E}[Z_2] = b_n \cdot [1 + o(1)]. \tag{3.D.152}$$

We then study $\mathbb{E}[T_1]$. We plug in $\delta_{ij} = U_{\tau(i)\tau(j)}[(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$ to get

$$\begin{aligned} T_1 = 2 \sum_{\substack{i_1,i_2,i_3,i_4 \\ (dist)}} & U_{\tau(i_1)\tau(i_2)} U_{\tau(i_2)\tau(i_3)} U_{\tau(i_3)\tau(i_4)} \times \\ & (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2})^2 (\mathbb{E}d_{i_3})^2 (\mathbb{E}d_{i_4} - d_{i_4}) W_{i_4 i_1} + rem \\ \equiv 2T_{11} + rem. & \end{aligned}$$

We claim that

$$|\mathbb{E}[rem]| = o(\|\theta\|^4).$$

The calculations here are similar to those in Equation (E.176) of Jin et al. (2019), where T_1 there (with a slightly different meaning) is decomposed into $2T_{1a} + 2T_{1b} + 2T_{1c} + 2T_{1d}$. Here, T_{11} is analogous to T_{1d}, and the remainder term is analogous to $2T_{1a} + 2T_{1b} + 2T_{1c}$. In Jin et al. (2019), it was shown that $|\mathbb{E}[T_{1a}]| + |\mathbb{E}[T_{1b}]| + |\mathbb{E}[T_{1c}]| = o(\|\theta\|^4)$; see Equations (E.179)-(E.181) in Jin et al. (2019). We can adapt their proof to show $|\mathbb{E}[rem]| = o(\|\theta\|^4)$. Since the calculations are elementary, we omit the details to save space. We then compute $\mathbb{E}[T_{11}]$. Since $\mathbb{E}d_i - d_i = -\sum_{j:j\neq i} W_{ji}$, it follows that

$$\begin{aligned} \mathbb{E}[T_{11}] &= -\mathbb{E}\Big[\sum_{\substack{k_1,k_2 \\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}(\mathbb{E}d_{i_1})\Big(\sum_{i_5:i_5\neq i_2} W_{i_2i_5}\Big)^2 (\mathbb{E}d_{i_3})^2 \Big(\sum_{i_6:i_6\neq i_4} W_{i_4i_6}\Big) W_{i_4i_1}\Big] \\ &= -\mathbb{E}\Big[\sum_{\substack{k_1,k_2 \\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}(\mathbb{E}d_{i_1})\Big(\sum_{i_5:i_5\neq i_2} W_{i_2i_5}^2\Big)(\mathbb{E}d_{i_3})^2 W_{i_4i_1}^2\Big] \\ &= -\sum_{\substack{k_1,k_2 \\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}(\mathbb{E}d_{i_1})(\mathbb{E}d_{i_3})^2 \mathbb{E}[W_{i_4i_1}^2]\Big(\sum_{i_5:i_5\neq i_2} \mathbb{E}[W_{i_2i_5}^2]\Big) \\ &= -\sum_{\substack{k_1,k_2 \\ k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}(\mathbb{E}d_{i_1})(\mathbb{E}d_{i_3})^2 \mathbb{E}[W_{i_4i_1}^2] \cdot [1+o(1)]\Big(\theta_{i_2}\|\theta\|_1 \underbrace{\sum_{k_5} P_{k_2k_5} g_{k_5}}_{V_{k_2k_2}}\Big) \\ &= -[1+o(1)] \sum_{\substack{k_1,k_2 \\ k_3,k_4}} \frac{P_{k_1k_2}P_{k_2k_3}P_{k_3k_4}P_{k_1k_4}}{V_{k_2k_2}V_{k_4k_4}\|\theta\|_1^2} \Big(\sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} \theta_{i_1}^2 \theta_{i_2} \theta_{i_3}^2 \theta_{i_4}\Big) \\ &= -[1+o(1)] \cdot \|\theta\|^4 g' V^{-1}[(PH^2P) \circ (PH^2P)]V^{-1} g, \end{aligned}$$

where we have plugged in (3.D.149)-(3.D.150) in the second last line, and the last line can be derived similarly as in the equation above (3.D.151). We have proved $\mathbb{E}[T_{11}] = -[1 + o(1)] \cdot b_n/2$. Then,

$$\mathbb{E}[T_1] = 2\mathbb{E}[T_{11}] + o(\|\theta\|^4) = -b_n \cdot [1 + o(1)]. \tag{3.D.153}$$

We then study $\mathbb{E}[F]$. Similar to the analysis of T_1, after plugging in $\delta_{ij} = U_{\tau(i)\tau(j)}[(\mathbb{E}d_i)(\mathbb{E}d_j -$

$d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$, we can obtain that

$$
\begin{aligned}
F = rem + 2 \sum_{i_1,i_2,i_3,i_4(dist)} & U_{\tau(i_1)\tau(i_2)}U_{\tau(i_2)\tau(i_3)}U_{\tau(i_3)\tau(i_4)}U_{\tau(i_4)\tau(i_1)}\times \\
& (\mathbb{E}d_{i_1})(\mathbb{E}d_{i_2} - d_{i_2})^2(\mathbb{E}d_{i_3})^2(\mathbb{E}d_{i_4} - d_{i_4})^2(\mathbb{E}d_{i_1}), \\
\equiv rem + 2F_1, \qquad & \text{where} \quad |\mathbb{E}[rem]| = o(\|\theta\|^4).
\end{aligned}
$$

The proof of $|\mathbb{E}[rem]| = o(\|\theta\|^4)$ is similar to the proof of (E.188)-(E.189) in Jin et al. (2019). There they analyzed a quantity F, which bears some similarity to the F here, and decomposed $F = 2F_a + 12F_b + 2F_c$, where $2F_a + 12F_b$ is analogous to rem here. They proved that $|\mathbb{E}[F_a]| + |\mathbb{E}[F_b]| = o(\|\theta\|^4)$. We can mimic their proof to show $|\mathbb{E}[rem]| = o(\|\theta\|^4)$. By direct calculations,

$$
\begin{aligned}
\mathbb{E}[F_1] &= \mathbb{E}\Big[\sum_{\substack{k_1,k_2\\k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}U_{k_4k_1}(\mathbb{E}d_{i_1})^2(\mathbb{E}d_{i_3})^2(\mathbb{E}d_{i_2} - d_{i_2})^2(\mathbb{E}d_{i_4} - d_{i_4})^2\Big] \\
&= \mathbb{E}\Big[\sum_{\substack{k_1,k_2\\k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} U_{k_1k_2}U_{k_2k_3}U_{k_3k_4}U_{k_4k_1}(\mathbb{E}d_{i_1})^2(\mathbb{E}d_{i_3})^2\Big(\sum_{i_5: i_5 \neq i_2} W_{i_2i_5}^2\Big)\Big(\sum_{i_6: i_6 \neq i_4} W_{i_4i_6}^2\Big)\Big] \\
&= [1+o(1)] \sum_{\substack{k_1,k_2\\k_3,k_4}} \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} \frac{P_{k_1k_2}P_{k_2k_3}P_{k_3k_4}P_{k_4k_1}\theta_{i_1}^2\theta_{i_3}^2}{V_{k_2k_2}^2V_{k_4k_4}^2\|\theta\|_1^4}\Big(\theta_{i_2}\|\theta\|_1 \underbrace{\sum_{k_5} P_{k_2k_5}g_{k_5}}_{V_{k_2k_2}}\Big)\Big(\theta_{i_4}\|\theta\|_1 \underbrace{\sum_{k_6} P_{k_4k_6}g_{k_6}}_{V_{k_4k_4}}\Big) \\
&= [1+o(1)] \sum_{\substack{k_1,k_2\\k_3,k_4}} \frac{P_{k_1k_2}P_{k_2k_3}P_{k_3k_4}P_{k_4k_1}}{V_{k_2k_2}V_{k_4k_4}\|\theta\|_1^2}\Big(\sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}} \theta_{i_1}^2\theta_{i_2}\theta_{i_3}^2\theta_{i_4}\Big) \\
&= [1+o(1)] \cdot \|\theta\|^4 g'V^{-1}[(PH^2P) \circ (PH^2P)]V^{-1}g,
\end{aligned}
$$

where in the second line we discard terms with mean zero, in the third line we plug in (3.D.149)-(3.D.150), and in the last line we use elementary calculations similar to those in the equation above (3.D.151). It follows that $\mathbb{E}[F_1] = [1+o(1)] \cdot b_n/2$ and that

$$
\mathbb{E}[F] = 2\mathbb{E}[F_1] + o(\|\theta\|^4) = [1+o(1)] \cdot b_n. \tag{3.D.154}
$$

We now plug (3.D.151), (3.D.152), (3.D.153), and (3.D.154) into (3.D.144) to get

$$
\begin{aligned}
\mathbb{E}[Q_n^{*(m,0)} - \widetilde{Q}_n^{(m,0)}] &= 4\mathbb{E}[Z_1] + 2\mathbb{E}[Z_2] + 4\mathbb{E}[T_1] + \mathbb{E}[F] \\
&= [1+o(1)] \cdot [4(b_n/2) + 2b_n - 4b_n + b_n] \\
&= [1+o(1)] \cdot b_n.
\end{aligned}
$$

Since $b_n \asymp \|\theta\|^4$, (3.D.147) follows immediately. $\square$

3.D.4 Proof of Lemma 50

Similar to the proof of Lemma 49, we use the notation $M_{ijk\ell}(X) = X_{ij}X_{jk}X_{k\ell}X_{\ell i}$. By (3.C.130),

$$\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} [M_{i_1i_2i_3i_4}(\widetilde{X}^*) - M_{i_1i_2i_3i_4}(X^*)],$$

$$\text{where} \quad \begin{cases} \widetilde{X}^*_{ij} = \widetilde{\Omega}^{(m,0)}_{ij} + W_{ij} + \delta^{(m,0)}_{ij} + \widetilde{r}^{(m,0)}_{ij}, \\ X^*_{ij} = \widetilde{\Omega}^{(m,0)}_{ij} + W_{ij} + \delta^{(m,0)}_{ij}. \end{cases}$$

For the rest of the proof, we omit the superscripts $(m,0)$ in $(\widetilde{\Omega}, \delta, \widetilde{r})$. There are $4^4 - 3^4 = 175$ post-expansion sums in $\widetilde{Q}_n^{*(m,0)} - Q_n^{*(m,0)}$, each with the form

$$S \equiv \sum_{i_1,i_2,i_3,i_4(dist)} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \qquad \text{where} \quad a,b,c,d \in \{\widetilde{\Omega}, W, \delta, \widetilde{r}\}. \tag{3.D.155}$$

Here we use S as a generic notation for any post-expansion sum. To show the claim, it suffices to bound $|\mathbb{E}[S]|$ and $\mathrm{Var}(S)$ for each post-expansion sum S.

We now study S. Let $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, \ldots, \mathcal{N}_m^{(m,0)}$ be the pseudo-communities defined by Π_0. By (3.C.125) and (3.C.126), for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\delta_{ij} = U_{k\ell}\big[(\mathbb{E}d_i)(d_j - \mathbb{E}d_j) + (\mathbb{E}d_j)(d_i - \mathbb{E}d_i)\big], \qquad \widetilde{r}_{ij} = -\widehat{U}_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j).$$

The term $\widehat{U}_{k\ell}$ has a complicated correlation with each summand, so we want to "replace" it with $U_{k\ell}$. Introduce a proxy of $\widetilde{r}_{ij}$ as

$$r_{ij} = -U_{k\ell}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j) \tag{3.D.156}$$

We define a proxy of S as

$$T \equiv \sum_{i_1,i_2,i_3,i_4(dist)} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \qquad \text{where} \quad a,b,c,d \in \{\widetilde{\Omega}, W, \delta, r\}. \tag{3.D.157}$$

We note that T is also a generic notation, and it has a one-to-one correspondence with S. For example, if $S = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}W_{i_2i_3}\widetilde{\Omega}_{i_3i_4}\widetilde{r}_{i_4i_1}$, then $T = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}W_{i_2i_3}\widetilde{\Omega}_{i_3i_4}r_{i_4i_1}$; if $S = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}\widetilde{r}_{i_2i_3}\widetilde{r}_{i_3i_4}W_{i_4i_1}$, then $T = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}r_{i_2i_3}r_{i_3i_4}W_{i_4i_1}$. Therefore, to bound the mean and variance of S, we only need to study T and $S - T$ separately.

First, we study the mean and variance of T. Since $d_i - \mathbb{E}d_i = \sum_{j:j\neq i} W_{ij}$, we can write δ_{ij} as a linear form of W and r_{ij} as a quadratic form of W. We then plug them into the expression of T and write T as a weighted sum of monomials of W. Take $T = \sum_{i_1,i_2,i_3,i_4(dist)} r_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$ for example. It can be re-written as (note: $\tau(i)$ is the index of pseudo-community that contains node i)

$$\begin{aligned} T &= -\sum_{i_1,i_2,i_3,i_4(dist)} U_{\tau(i_1)\tau(i_2)}\Big(\sum_{j_1:j_1\neq i_1} W_{i_1j_1}\Big)\Big(\sum_{j_2:j_2\neq i_2} W_{i_2j_2}\Big)W_{i_2i_3}W_{i_3i_4}W_{i_4i_1} \\ &= -\sum_{\substack{i_1,i_2,i_3,i_4(dist) \\ j_1,j_2:j_1\neq i_1,j_2\neq i_2}} U_{\tau(i_1)\tau(i_2)}W_{i_1j_1}W_{i_2j_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}. \end{aligned}$$

Then, we can compute the mean and variance of T directly. We use the same strategy to analyze each of the 175 post-expansion sums of the form (3.D.157). Similar calculations were conducted in the proof of Lemma E.11 of Jin et al. (2019). The setting of Lemma E.11

is a special case where $U_{k\ell} \equiv 1/v$ for a scalar v. However, their proof does not rely on that $U_{k\ell}$'s are equal to each other. Instead, their proof only requires a universal upper bound on $U_{k\ell}$. In fact, they have proved the following lemma:

Lemma 53. *Consider a DCBM model where (3.2.1)-(3.2.2) and (3.2.4) hold. Let* $W = A - \Omega$ *and* $\Delta = \sum_{i_1,i_2,i_3,i_4(dist)} \big[M_{i_1i_2i_3i_4}\big(\widetilde{\Omega} + W + \delta + r\big) - M_{i_1i_2i_3i_4}\big(\widetilde{\Omega} + W + \delta\big)\big]$, *where* $\widetilde{\Omega}$ *is a non-stochastic symmetric matrix,* $\delta_{ij} = v_{ij} \cdot [(\mathbb{E}d_i)(\mathbb{E}d_j - d_j) + (\mathbb{E}d_j)(\mathbb{E}d_i - d_i)]$, $r_{ij} = -u_{ij}(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j)$, $\{v_{ij}, u_{ij}\}_{1\le i\ne j\le n}$ *are non-stochastic scalars,* d_i *is the degree of node* i, *and* $M_{i_1i_2i_3i_4}(\cdot)$ *is as defined above. As* $n \to \infty$, *suppose there is a constant* $C > 0$ *and a scalar* $\alpha_n > 0$ *such that* $\alpha_n \le C$, $\alpha_n\|\theta\| \to \infty$, $|\widetilde{\Omega}_{ij}| \le C\alpha_n\theta_i\theta_j$, $|v_{ij}| \le C\|\theta\|_1^{-1}$, *and* $|u_{ij}| \le C\|\theta\|_1^{-1}$ *for* $1 \le i, j \le n$. *Let* T *be an arbitrary post-expansion sum of* Δ. *Then,* $|\mathbb{E}[T]| \le C\alpha_n^2\|\theta\|^6 + o(\|\theta\|^4)$ *and* $\mathrm{Var}(T) = o\big(\alpha_n^6\|\theta\|^8\|\theta\|_3^6 + \|\theta\|^8\big)$.

We apply Lemma 53 for $\alpha_n = \tau$ and $v_{ij} = u_{ij} = U_{\tau(i)\tau(j)}$. By Lemma 41, $\tau \le C$, $\tau\|\theta\| \to \infty$, and $|\widetilde{\Omega}_{ij}| \le C\tau\theta_i\theta_j$. In (3.D.146), we have seen that $|U_{k\ell}| \le C\|\theta\|_1^{-1}$. The conditions of Lemma 53 are satisfied. We immediately have: Under the conditions of Lemma 43 (note: $\tau\|\theta\| \to \infty$)

$$|\mathbb{E}[T]| \le C\tau^2\|\theta\|^6 + o(\|\theta\|^4) = o(\tau^4\|\theta\|^8), \qquad \mathrm{Var}(T) = o\big(\tau^6\|\theta\|^8\|\theta\|_3^6 + \|\theta\|^8\big), \tag{3.D.158}$$

and under the conditions of Lemma 38 (i.e., $\widetilde{\Omega}$ is a zero matrix and $\tau = 0$),

$$|\mathbb{E}[T]| = o(\|\theta\|^4), \qquad \mathrm{Var}(T) = o(\|\theta\|^8). \tag{3.D.159}$$

Next, we study the variable $(S - T)$. In (3.D.155) and (3.D.157), if we group the summands based on pseudo-communities of (i_1, i_2, i_3, i_4), then we have

$$S = \sum_{1\le k_1,k_2,k_3,k_4\le m} S_{k_1k_2k_3k_4} \qquad \text{and} \qquad T = \sum_{1\le k_1,k_2,k_3,k_4\le m} T_{k_1k_2k_3k_4},$$

where $S_{k_1k_2k_3k_4}$ contains all the summands such that $i_s \in \mathcal{N}_{k_s}^{(m,0)}$ for $s = 1, 2, 3, 4$. By straightforward calculations and definitions of $(r_{ij}, \widetilde{r}_{ij})$, we have

$$S_{k_1k_2k_3k_4} = \widehat{U}_{k_1k_2}^{\ell_a}\widehat{U}_{k_2k_3}^{\ell_b}\widehat{U}_{k_3k_4}^{\ell_c}\widehat{U}_{k_4k_1}^{\ell_d} \sum_{s=1}^{4} \sum_{i_s\in\mathcal{N}_{k_s}^{(m,0)}} \widetilde{a}_{i_1i_2}\widetilde{b}_{i_2i_3}\widetilde{c}_{i_3i_4}\widetilde{d}_{i_4i_1},$$

$$T_{k_1k_2k_3k_4} = U_{k_1k_2}^{\ell_a}U_{k_2k_3}^{\ell_b}U_{k_3k_4}^{\ell_c}U_{k_4k_1}^{\ell_d} \sum_{s=1}^{4} \sum_{i_s\in\mathcal{N}_{k_s}^{(m,0)}} \widetilde{a}_{i_1i_2}\widetilde{b}_{i_2i_3}\widetilde{c}_{i_3i_4}\widetilde{d}_{i_4i_1},$$

$$\text{where} \quad \widetilde{a}_{ij}, \widetilde{b}_{ij}, \widetilde{c}_{ij}, \widetilde{d}_{ij} \in \big\{\widetilde{\Omega}_{ij},\, W_{ij},\, \delta_{ij},\, -(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j)\big\}.$$

Here $\ell_a \in \{0, 1\}$ is an indicator about whether a_{ij} takes the value of $\widetilde{r}_{ij}$ in S, and (ℓ_b, ℓ_c, ℓ_d) are similar. For example, if $S = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}W_{i_2i_3}\widetilde{\Omega}_{i_3i_4}\widetilde{r}_{i_4i_1}$, then $(\ell_a, \ell_b, \ell_c, \ell_d) = (0, 0, 0, 1)$; if $S = \sum_{i_1,i_2,i_3,i_4(dist)} \delta_{i_1i_2}\widetilde{r}_{i_2i_3}\widetilde{r}_{i_3i_4}W_{i_4i_1}$, then $(\ell_a, \ell_b, \ell_c, \ell_d) = (0, 1, 1, 0)$. For any post-expansion sum S considered here, $1 \le \ell_a + \ell_b + \ell_c + \ell_d \le 4$. To study the difference between $S_{k_1k_2k_3k_4}$ and $T_{k_1k_2k_3k_4}$, we introduce an intermediate term

$$R_{k_1k_2k_3k_4} = \Big(\frac{1}{\|\theta\|_1^2}\Big)^{\ell_a+\ell_b+\ell_c+\ell_d} \sum_{s=1}^{4} \sum_{i_s\in\mathcal{N}_{k_s}^{(m,0)}} \widetilde{a}_{i_1i_2}\widetilde{b}_{i_2i_3}\widetilde{c}_{i_3i_4}\widetilde{d}_{i_4i_1}.$$

In fact, $R_{k_1k_2k_3k_4}$ has a similar form as $T_{k_1k_2k_3k_4}$ except that the scalar $U_{k\ell}$ in the definition of r_{ij} (see (3.D.156)) is replaced by $1/\|\theta\|_1^2$. We apply Lemma 53 with $u_{ij} \equiv 1/\|\theta\|_1^2$. It yields that, under conditions of Lemma 38,

$$|\mathbb{E}[R_{k_1k_2k_3k_4}]| = o(\|\theta\|^4), \qquad \mathrm{Var}(R_{k_1k_2k_3k_4}) = o(\|\theta\|^8),$$

and under conditions of Lemma 43,

$$|\mathbb{E}[R_{k_1k_2k_3k_4}]| \leq C\tau^2\|\theta\|^6 + o(\|\theta\|^4), \qquad \mathrm{Var}(R_{k_1k_2k_3k_4}) = o\big(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6\big).$$

Particularly, since $\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \mathrm{Var}(X)$ for any variable X, we have

$$\begin{aligned}\|\theta\|^{-4}\,\mathbb{E}[R^2_{k_1k_2k_3k_4}] &\leq \begin{cases} o(\|\theta\|^4), & \text{for setting of Lemma 38,}\\ C\tau^4\|\theta\|^8 + o\big(\|\theta\|^4 + \tau^6\|\theta\|^4\|\theta\|_3^6\big), & \text{for setting of Lemma 43,}\end{cases}\\ &= \begin{cases} o(\|\theta\|^4), & \text{for setting of Lemma 38,}\\ C\|\theta\|^8, & \text{for setting of Lemma 43.}\end{cases}\end{aligned} \tag{3.D.160}$$

Note that in deriving (3.D.160) we have used $\tau \leq C$ and $\tau^6\|\theta\|^4\|\theta\|_3^6 \leq \tau^6\|\theta\|^4 \cdot \theta_{\max}^2\|\theta\|^4 \leq C\|\theta\|^8$.

We now investigate $(S_{k_1k_2k_3k_4} - T_{k_1k_2k_3k_4})$. By condition (3.2.1), $\sqrt{\log(n)} \ll \|\theta\|_1/\|\theta\|^2$. Hence, we can take a sequence of x_n, such that $\sqrt{\log(n)} \ll x_n \ll \|\theta\|_1/\|\theta\|^2$, and define the event E_n:

$$E_n = \Big\{|U_{k\ell} - \widehat{U}_{k\ell}| \leq \frac{C_0 x_n}{\|\theta\|_1^3}, \quad \text{for all } 1 \leq k,\ell \leq m\Big\}, \tag{3.D.161}$$

where $C_0 > 0$ is a constant to be decided. To bound the probability of E_n^c, we recall that (by definitions in (3.C.121) and (3.C.123))

$$\widehat{U}_{k\ell} = \frac{\mathbf{1}_k' A \mathbf{1}_\ell}{(\mathbf{1}_k' d)(\mathbf{1}_\ell' d)}, \qquad \text{and} \qquad U_{k\ell} = \frac{\mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell}{(\mathbf{1}_k'\mathbb{E}[d])(\mathbf{1}_\ell'\mathbb{E}[d])},$$

where $\mathbf{1}_k$ is a shorthand notation for $\mathbf{1}_k^{(m,0)}$ in (3.C.108). Using Bernstein's inequality and mimicking the argument from (E.299)-(E.300) of Jin et al. (2019), we can easily show that, there is a constant $C_1 > 0$ such that, for any $1 \leq k,\ell \leq m$,

$$\mathbb{P}\Big(\big|\mathbf{1}_k' A\mathbf{1}_\ell - \mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell\big| > x_n\|\theta\|_1\Big) \leq 2\exp(-C_1x_n^2). \tag{3.D.162}$$

By probability union bound, with probability $1 - 2m^2\exp(-C_1x_n^2)$,

$$\max_{1\leq k,\ell\leq m}\big\{\big|\mathbf{1}_k' A\mathbf{1}_\ell - \mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell\big|\big\} \leq x_n\|\theta\|_1.$$

Furthermore, $\mathbf{1}_k'd - \mathbf{1}_k'\mathbb{E}[d] = \sum_{\ell=1}^m(\mathbf{1}_k'A\mathbf{1}_\ell - \mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell)$. So, with probability $1 - 2m^2\exp(-C_1x_n^2)$,

$$\max_{1\leq k\leq m}\big\{\big|\mathbf{1}_k'd - \mathbf{1}_k'\mathbb{E}[d]\big|\big\} \leq m \cdot x_n\|\theta\|_1.$$

At the same time, we know that $\mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell \asymp \|\theta\|_1^2$ and $\mathbf{1}_k'\mathbb{E}[d] \asymp \|\theta\|_1^2$. We plug the above results into the expressions of $U_{k\ell}$ and $\widehat{U}_{k\ell}$ and can easily find that, with probability $1 - 2m^2\exp(-C_1x_n^2)$,

$$\max_{1\leq k,\ell\leq m}|\widehat{U}_{k\ell} - U_{k\ell}| \leq C_0x_n/\|\theta\|_1^3,$$

for some constant $C_0 > 0$ (C_0 still depends on m, but m is bounded here). We use the same

C_0 to define E_n. Then,

$$\mathbb{P}(E_n^c) \leq 2m^2 \exp(-C_1 x_n^2) = o(n^{-L}), \qquad \text{for any fixed } L > 0, \tag{3.D.163}$$

where the last equality is due to $x_n^2 \gg \log(n)$. We aim to use (3.D.163) to bound $\mathbb{E}[(S_{k_1k_2k_3k_4} - T_{k_1k_2k_3k_4}) \cdot I_{E_n^c}]$. It is easy to see the trivial bound $|\widehat{U}_{k\ell}| \leq 1$ and $|U_{k\ell}| \leq 1$. Also, recall that $\widetilde{a}_{ij}$ takes value in $\{\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, -(d_i - \mathbb{E}d_i)(d_j - \mathbb{E}d_j)\}$, and so $|a_{ij}| \leq n^2$; we have the same bound for $|\widetilde{b}_{ij}|, |\widetilde{c}_{ij}|, |\widetilde{d}_{ij}|$. This gives a trivial bound

$$(S_{k_1k_2k_3k_4} - T_{k_1k_2k_3k_4})^2 \leq 2S^2_{k_1k_2k_3k_4} + 2T^2_{k_1k_2k_3k_4} \leq 2(n^4 \cdot n^8)^2 + 2(n^4 \cdot n^8)^2 = 4n^{24}.$$

Combining it with (3.D.163), we have

$$\mathbb{E}[(T_{k_1k_2k_3k_4} - S_{k_1k_2k_3k_4})^2 \cdot I_{E_n^c}] \leq 4n^{24} \cdot 2m^2 \exp(-C_1 x_n^2) = o(1). \tag{3.D.164}$$

At the same time, on the event E_n,

$$\begin{aligned}
&|S_{k_1k_2k_3k_4} - T_{k_1k_2k_3k_4}| \\
= \; & \big|\widehat{U}^{\ell_a}_{k_1k_2}\widehat{U}^{\ell_b}_{k_2k_3}\widehat{U}^{\ell_c}_{k_3k_4}\widehat{U}^{\ell_d}_{k_4k_1} - U^{\ell_a}_{k_1k_2}U^{\ell_b}_{k_2k_3}U^{\ell_c}_{k_3k_4}U^{\ell_d}_{k_4k_1}\big| \cdot \|\theta\|_1^{2(\ell_a+\ell_b+\ell_c+\ell_d)}|R_{k_1k_2k_3k_4}| \\
\leq \; & C\Big(|U^{\ell_a}_{k_1k_2}U^{\ell_b}_{k_2k_3}U^{\ell_c}_{k_3k_4}U^{\ell_d}_{k_4k_1}| \max_{1\leq k,\ell\leq m}\big|\widehat{U}_{k\ell}/U_{k\ell} - 1\big|\Big) \cdot \|\theta\|_1^{2(\ell_a+\ell_b+\ell_c+\ell_d)}|R_{k_1k_2k_3k_4}| \\
\leq \; & C\|\theta\|_1^2 \cdot \max_{1\leq k,\ell\leq m} |\widehat{U}_{k\ell} - U_{k\ell}| \cdot |R_{k_1k_2k_3k_4}| \\
\leq \; & Cx_n\|\theta\|_1^{-1} \cdot |R_{k_1k_2k_3k_4}| \\
= \; & o(\|\theta\|^{-2}) \cdot |R_{k_1k_2k_3k_4}|,
\end{aligned}$$

where the fourth line is because $\|\theta\|_1^{-2} \leq |U_{k\ell}| \leq C\|\theta\|_1^{-2}$ (e.g., see (3.D.146)) and the last line is because $x_n \ll \|\theta\|_1/\|\theta\|^2$. It follows that

$$\mathbb{E}[(T_{k_1k_2k_3k_4} - S_{k_1k_2k_3k_4})^2 \cdot I_{E_n}] = o(\|\theta\|^{-4}) \cdot \mathbb{E}[R^2_{k_1k_2k_3k_4}]. \tag{3.D.165}$$

We combine (3.D.164) and (3.D.165) and plug in (3.D.160). It follows that

$$\begin{aligned}
\mathbb{E}[(T_{k_1k_2k_3k_4} - S_{k_1k_2k_3k_4})^2] &= o(\|\theta\|^{-4}) \cdot \mathbb{E}[R^2_{k_1k_2k_3k_4}] + o(1) \\
&= \begin{cases} o(\|\theta\|^4), & \text{under conditions of Lemma 38,} \\ o(\|\theta\|^8), & \text{under conditions of Lemma 43.} \end{cases}
\end{aligned}$$

Since m is bound, we immediately know that

$$\mathbb{E}[(S-T)^2] = \begin{cases} o(\|\theta\|^4), & \text{under conditions of Lemma 38,} \\ o(\|\theta\|^8), & \text{under conditions of Lemma 43.} \end{cases} \tag{3.D.166}$$

Last, we combine the results on T and the results on $(S - T)$. By (3.D.158)-(3.D.159) and (3.D.166),

$$\begin{aligned}
|\mathbb{E}[S]| &\leq |\mathbb{E}[T]| + |\mathbb{E}[S-T]|| \\
&\leq |\mathbb{E}[T]| + \sqrt{\mathbb{E}[(S-T)^2]} \\
&= \begin{cases} o(\|\theta\|^4) + o(\|\theta\|^2) = o(\|\theta\|^4), & \text{for setting of Lemma 38,} \\ o(\tau^4\|\theta\|^8) + o(\|\theta\|^4) = o(\tau^4\|\theta\|^8), & \text{for setting of Lemma 43.} \end{cases}
\end{aligned}$$

Additionally,

$$\begin{aligned}\mathrm{Var}(S) &\le 2\mathrm{Var}(T) + 2\mathrm{Var}(S-T)\\ &\le 2\mathrm{Var}(T) + 2\mathbb{E}[(S-T)^2]\\ &\le \begin{cases} o(\|\theta\|^8) + o(\|\theta\|^4) = o(\|\theta\|^8), & \text{for setting of Lemma 38,}\\ o\big(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6\big) + o(\|\theta\|^8) = o\big(\|\theta\|^8 + \tau^6\|\theta\|^8\|\theta\|_3^6\big), & \text{for setting of Lemma 43.}\end{cases}\end{aligned}$$

This gives the desirable claim. □

3.D.5 Proof of Lemma 51

Similar to the proof of Lemma 49, we use the notation $M_{ijk\ell}(X) = X_{ij}X_{jk}X_{k\ell}X_{\ell i}$. By (3.C.130),

$$Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)} = \sum_{i_1,i_2,i_3,i_4(dist)} [M_{i_1i_2i_3i_4}(X) - M_{i_1i_2i_3i_4}(\widetilde{X}^*)],$$

$$\text{where} \quad \begin{cases} X_{ij} = \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)} + \epsilon_{ij}^{(m,0)},\\ \widetilde{X}_{ij}^* = \widetilde{\Omega}_{ij}^{(m,0)} + W_{ij} + \delta_{ij}^{(m,0)} + \widetilde{r}_{ij}^{(m,0)}.\end{cases}$$

We shall omit the superscripts $(m,0)$ in $(\widetilde{\Omega}, \delta, \widetilde{r}, \epsilon)$. Let $\mathcal{N}_1^{(m,0)}, \mathcal{N}_2^{(m,0)}, \ldots, \mathcal{N}_m^{(m,0)}$ be the pseudo-communities defined by Π_0. By (3.C.127), $\epsilon_{ij} = \widetilde{\alpha}_{ij} + \widetilde{\beta}_{ij} + \widetilde{\gamma}_{ij}$, where for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\begin{aligned}\widetilde{\alpha}_{ij} &= d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell},\\ \widetilde{\beta}_{ij} &= (U_{k\ell} - \widehat{U}_{k\ell})(\mathbb{E}d_i)(\mathbb{E}d_j),\\ \widetilde{\gamma}_{ij} &= (U_{k\ell} - \widehat{U}_{k\ell})[(\mathbb{E}d_i)(d_j - \mathbb{E}d_j) + (\mathbb{E}d_j)(d_i - \mathbb{E}d_i)].\end{aligned} \tag{3.D.167}$$

Therefore, we can write

$$\begin{aligned}&Q_n^{(m,0)} - \widetilde{Q}_n^{*(m,0)}\\ &= \sum_{i_1,i_2,i_3,i_4(dist)} [M_{i_1i_2i_3i_4}(\widetilde{\Omega} + W + \delta + \widetilde{r} + \widetilde{\alpha} + \widetilde{\beta} + \widetilde{\gamma}) - M_{i_1i_2i_3i_4}(\widetilde{\Omega} + W + \delta + \widetilde{r})].\end{aligned}$$

There are $7^4 - 4^4 = 2145$ post-expansion sums. Let S be the generic notation for any such post-expansion sum. Similarly as in the proof of Lemma 50, we group the summands according to which pseudo-communities (i_1, i_2, i_3, i_4) belong to, i.e., we write $S = \sum_{1\le k_1,k_2,k_3,k_4\le m} S_{k_1k_2k_3k_4}$, where

$$S_{k_1k_2k_3k_4} = \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}^{(m,0)}} a_{i_1i_2} b_{i_2i_3} c_{i_3i_4} d_{i_4i_1}, \quad \text{where } a, b, c, d \in \{\widetilde{\Omega}, W, \delta, \widetilde{r}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}\}. \tag{3.D.168}$$

It suffices to study the mean and variance of each $S_{k_1k_2k_3k_4}$.

Let τ and r_{ij} be the same as in (3.4.33) and (3.D.156). Define

$$\begin{aligned}\alpha_{ij} &= \frac{\tau\|\theta\|_1}{\theta_{\max}}\big[d_i^* d_j^* U_{k\ell}^* - (\mathbb{E}d_i)(\mathbb{E}d_j)U_{k\ell}\big],\\ \beta_{ij} &= \tau U_{k\ell}(\mathbb{E}d_i)(\mathbb{E}d_j),\\ \gamma_{ij} &= U_{k\ell}[(\mathbb{E}d_i)(d_j - \mathbb{E}d_j) + (\mathbb{E}d_j)(d_i - \mathbb{E}d_i)].\end{aligned} \tag{3.D.169}$$

We introduce a proxy of $S_{k_1k_2k_3k_4}$ as

$$S^*_{k_1k_2k_3k_4} = \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}^{(m,0)}} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \quad \text{where } a,b,c,d \in \{\widetilde{\Omega}, W, \delta, r, \alpha, \beta, \gamma\}. \tag{3.D.170}$$

Reviewing the expressions of $(\widetilde{\Omega}_{ij}, W_{ij}, \delta_{ij}, r_{ij}, \alpha_{ij}, \beta_{ij}, \gamma_{ij})$, we know that $S^*_{k_1k_2k_3k_4}$ can always be written as a weighted sum of monomials of W, and so we can calculate the mean and variance of $S^*_{k_1k_2k_3k_4}$ (the straightforward calculations are still tedious, but later we will introduce a simple trick to do that). Comparing (3.D.169) with (3.D.167) and r_{ij} with $\widetilde{r}_{ij}$, we observe that, for $i \in \mathcal{N}_k^{(m,0)}$ and $j \in \mathcal{N}_\ell^{(m,0)}$,

$$\widetilde{r}_{ij} = \frac{\widehat{U}_{k\ell}}{U_{k\ell}} r_{ij}, \qquad \widetilde{\alpha}_{ij} = \frac{\theta_{\max}}{\tau\|\theta\|_1}\alpha_{ij}, \qquad \widetilde{\beta}_{ij} = \frac{U_{k\ell} - \widehat{U}_{k\ell}}{\tau U_{k\ell}}\beta_{ij}, \qquad \widetilde{\gamma}_{ij} = \frac{U_{k\ell} - \widehat{U}_{k\ell}}{U_{k\ell}}\gamma_{ij}.$$

We plug them into (3.D.168) to get

$$S_{k_1k_2k_3k_4} = \Big(\frac{\widehat{U}_{k\ell}}{U_{k\ell}}\Big)^{N_{\widetilde{r}}}\Big(\frac{\theta_{\max}}{\tau\|\theta\|_1^2}\Big)^{N_{\widetilde{\alpha}}}\Big(\frac{U_{k\ell} - \widehat{U}_{k\ell}}{\tau U_{k\ell}}\Big)^{N_{\widetilde{\beta}}}\Big(\frac{U_{k\ell} - \widehat{U}_{k\ell}}{U_{k\ell}}\Big)^{N_{\widetilde{\gamma}}} S^*_{k_1k_2k_3k_4}, \tag{3.D.171}$$

where $N_{\widetilde{r}}$ is the count of $\{a,b,c,d\}$ in (3.D.168) taking the value of $\widetilde{r}$, and $(N_{\widetilde{\alpha}}, N_{\widetilde{\beta}}, N_{\widetilde{\gamma}})$ are similar. For any post-expansion sum considered here, $1 \le N_{\widetilde{\alpha}} + N_{\widetilde{\beta}} + N_{\widetilde{\gamma}} \le 4$. The notation $(\frac{\widehat{U}_{k\ell}}{U_{k\ell}})^{N_{\widetilde{r}}}$ is interpreted in this way: For example, if in (3.D.168) only a takes the value of $\widetilde{r}$, then $N_{\widetilde{r}} = 1$ and $(\frac{\widehat{U}_{k\ell}}{U_{k\ell}})^{N_{\widetilde{r}}} = \frac{\widehat{U}_{k_1k_2}}{U_{k_1k_2}}$; if (a,b,c) take the value of $\widetilde{r}$, then $N_{\widetilde{r}} = 3$ and $(\frac{\widehat{U}_{k\ell}}{U_{k\ell}})^{N_{\widetilde{r}}} = \frac{\widehat{U}_{k_1k_2}}{U_{k_1k_2}}\frac{\widehat{U}_{k_2k_3}}{U_{k_2k_3}}\frac{\widehat{U}_{k_3k_4}}{U_{k_3k_4}}$. In (3.D.171), $S^*_{k_1k_2k_3k_4}$ is a random variable whose mean and variance are relatively easy to calculate. The factor in front of $S^*_{k_1k_2k_3k_4}$ has a complicated correlation with the summands in $S^*_{k_1k_2k_3k_4}$, but fortunately we can apply a simple bound on this factor. Consider the event E_n as in (3.D.161). We have shown in (3.D.163) that $\mathbb{P}(E_n^c) = o(n^{-L})$ for any fixed $L > 0$. Therefore, the event E_n^c has a negligible effect on the mean and variance of $S_{k_1k_2k_3k_4}$, i.e.,

$$\mathbb{E}[S^2_{k_1k_2k_3k_4} \cdot I_{E_n^c}] = o(1).$$

On the event E_n, we have $\max_{k,\ell}\{|\widehat{U}_{k\ell} - U_{k\ell}|/U_{k\ell}\} \le C_0 x_n/\|\theta\|_1$. It follows that

$$\begin{aligned}|S_{k_1k_2k_3k_4}| &\le \Big(\max_{k,\ell}\frac{|\widehat{U}_{k\ell}|}{U_{k\ell}}\Big)^{N_{\widetilde{\gamma}}}\Big(\frac{\theta_{\max}}{\tau\|\theta\|_1^2}\Big)^{N_{\widetilde{\alpha}}}\Big(\max_{k,\ell}\frac{|U_{k\ell} - \widehat{U}_{k\ell}|}{\tau U_{k\ell}}\Big)^{N_{\widetilde{\beta}}}\Big(\max_{k,\ell}\frac{|U_{k\ell} - \widehat{U}_{k\ell}|}{U_{k\ell}}\Big)^{N_{\widetilde{\gamma}}}|S^*_{k_1k_2k_3k_4}|\\ &\le C\Big(\frac{\theta_{\max}}{\tau\|\theta\|_1}\Big)^{N_{\widetilde{\alpha}}}\Big(\frac{x_n}{\tau\|\theta\|_1}\Big)^{N_{\widetilde{\beta}}}\Big(\frac{x_n}{\|\theta\|_1}\Big)^{N_{\widetilde{\gamma}}}|S^*_{k_1k_2k_3k_4}|.\end{aligned}$$

Since $x_n \ll \frac{\|\theta\|_1}{\|\theta\|^2}$ and $\tau\|\theta\| \to \infty$, we immediately have $\frac{x_n}{\|\theta\|_1} = o(\frac{1}{\|\theta\|^2})$, $\frac{x_n}{\tau\|\theta\|_1} = o(\frac{1}{\tau\|\theta\|^2}) = o(\frac{1}{\|\theta\|})$ and $\frac{\theta_{\max}}{\tau\|\theta\|_1} \le \frac{\theta^2_{\max}}{\tau\|\theta\|^2} = o(\frac{1}{\|\theta\|})$. It follows that

$$|S_{k_1k_2k_3k_4}| = o(1) \cdot \|\theta\|^{-(N_{\widetilde{\alpha}} + N_{\widetilde{\beta}} + 2N_{\widetilde{\gamma}})} \cdot |S^*_{k_1k_2k_3k_4}|, \qquad \text{on the event } E_n.$$

Combining the above gives

$$\mathbb{E}[S^2_{k_1k_2k_3k_4}] = \mathbb{E}[S^2_{k_1k_2k_3k_4} \cdot I_{E_n}] + \mathbb{E}[S^2_{k_1k_2k_3k_4} \cdot I_{E_n^c}]$$
$$= o(1) \cdot \|\theta\|^{-(2N_{\widetilde{\alpha}}+2N_{\widetilde{\beta}}+4N_{\widetilde{\gamma}})} \cdot \mathbb{E}\big[(S^*_{k_1k_2k_3k_4})^2\big] + o(1). \qquad (3.D.172)$$

It remains to bound $\mathbb{E}\big[(S^*_{k_1k_2k_3k_4})^2\big]$. As we mentioned, we can write $S^*_{k_1k_2k_3k_4}$ as a weighted sum of monomials of W and calculate its mean and variance directly. However, given that there are 2145 types of $S^*_{k_1k_2k_3k_4}$, the calculation is still very tedious. We now use a simple trick to relate the $S^*_{k_1k_2k_3k_4}$ to the post-expansion sums we have analyzed in Lemmas 49-50. We first bound $|\alpha_{ij}|$ in (3.D.169). Since $d_i^* = \mathbb{E}[d_i] + \Omega_{ii}$,

$$|\alpha_{ij}| \leq \frac{\tau\|\theta\|_1}{\theta_{\max}}\Big(\mathbb{E}[d_i]\mathbb{E}[d_j]|U^*_{k\ell} - U_{k\ell}| + (\Omega_{ii}\mathbb{E}[d_j] + \Omega_{jj}\mathbb{E}[d_i])U^*_{k\ell} + \Omega_{ii}\Omega_{jj}U^*_{k\ell}\Big).$$

By basic algebra, $|(x_1+x_2)/(y_1+y_2) - x_1/y_1| \leq |x_2|/(y_1+y_2) + |x_1y_2|/[(y_1+y_2)y_1]$. We apply it on (3.C.122)-(3.C.123) and note that $\mathbf{1}_k'(\Omega - \mathbb{E}[A])\mathbf{1}_\ell = \mathbf{1}_k' \mathrm{d}iag(\Omega)\mathbf{1}_\ell = O(\|\theta\|^2)$ and $\mathbf{1}_k'(d^* - \mathbb{E}[d]) = \mathbf{1}_k'\mathrm{d}iag(\Omega)\mathbf{1}_n = O(\|\theta\|^2)$. It yields

$$\begin{aligned}
&|U^*_{k\ell} - U_{k\ell}| \\
\leq\ & \frac{|\mathbf{1}_k'\Omega\mathbf{1}_\ell - \mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell|}{(\mathbf{1}_k'd^*)(\mathbf{1}_\ell'd^*)} + \frac{(\mathbf{1}_k'\mathbb{E}[A]\mathbf{1}_\ell)\,|(\mathbf{1}_k'd^*)(\mathbf{1}_\ell'd^*) - (\mathbf{1}_k'\mathbb{E}[d])(\mathbf{1}_\ell'\mathbb{E}[d])|}{(\mathbf{1}_k'd^*)(\mathbf{1}_\ell'd^*)(\mathbf{1}_k'\mathbb{E}[d])(\mathbf{1}_\ell'\mathbb{E}[d])} \\
\leq\ & C\|\theta\|_1^{-4} \cdot \mathbf{1}_k'\mathrm{d}iag(\Omega)\mathbf{1}_n + C\|\theta\|_1^{-6} \cdot |(\mathbf{1}_k'd^*)(\mathbf{1}_\ell'd^*) - (\mathbf{1}_k'\mathbb{E}[d])(\mathbf{1}_\ell'\mathbb{E}[d])| \\
\leq\ & C\|\theta\|_1^{-4} \cdot \|\theta\|^2 + C\|\theta\|_1^{-6} \cdot \|\theta\|_1^2\|\theta\|^2 \\
\leq\ & C\|\theta\|_1^{-3}\theta_{\max},
\end{aligned}$$

where in the last line we have used $\|\theta\|^2 \leq \theta_{\max}\|\theta\|_1$. Combining the above gives

$$\begin{aligned}
|\alpha_{ij}| &\leq \frac{C\tau\|\theta\|_1}{\theta_{\max}}\Big[\theta_i\theta_j\|\theta\|_1^2 \cdot \|\theta\|_1^{-3}\theta_{\max} + (\theta_i^2\theta_j\|\theta\|_1 + \theta_j^2\theta_i\|\theta\|_1) \cdot \|\theta\|_1^{-2} + \theta_i^2\theta_j^2\|\theta\|_1^{-2}\Big] \\
&\leq \frac{C\tau\|\theta\|_1}{\theta_{\max}} \cdot \frac{\theta_i\theta_j\theta_{\max}}{\|\theta\|_1} \leq C\tau\theta_i\theta_j.
\end{aligned}$$

Additionally, in (3.D.169), we observe that $\gamma_{ij} = \delta_{ij}$. Since $|U_{k\ell}| \leq C\|\theta\|^{-1}$ and $\mathbb{E}[d_i] \leq C\theta_i\|\theta\|_1$, it is true that $|\beta_{ij}| \leq C\tau\theta_i\theta_j$. We summarize the results as

$$|\alpha_{ij}| \leq C\tau\theta_i\theta_j, \qquad |\beta_{ij}| \leq C\tau\theta_i\theta_j, \qquad \gamma_{ij} = \delta_{ij}. \qquad (3.D.173)$$

It says that γ is the same as δ, and (α, β) behave similarly as $\widetilde{\Omega}$. Consequently, the calculation of mean and variance of $S^*_{k_1k_2k_3k_4}$ in (3.D.170) can be carried out by replacing (α, β, γ) with $(\widetilde{\Omega}, \widetilde{\Omega}, \delta)$. In other words, we only need to study a sum like

$$S^{**}_{k_1k_2k_3k_4} = \sum_{j=1}^{4} \sum_{i_j \in \mathcal{N}_{k_j}^{(m,0)}} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}, \quad \text{where } \ a, b, c, d \in \{\widetilde{\Omega}, W, \delta, r\}.$$

Let $(N_{\widetilde{\Omega}}, N_W, N_\delta, N_r, N_\alpha, N_\beta, N_\gamma)$ be the count of different terms in $\{a, b, c, d\}$ determined by $S^*_{k_1k_2k_3k_4}$, where these counts sum to 4. In $S^{**}_{k_1k_2k_3k_4}$, the counts become $N^*_{\widetilde{\Omega}} = N_{\widetilde{\Omega}} + N_\alpha + N_\beta$, $N^*_W = N_W$, $N^*_\delta = N_\delta + N_\gamma$ and $N^*_r = N_r$. Luckily, anything like $S^{**}_{k_1k_2k_3k_4}$ has been analyzed in Lemmas 49-50. Especially, in light of (3.D.172), the mean and variance contributed by any post-expansion sum considered here must be dominated by the mean and variance of some post-expansion sum considered in Lemmas 49-50. We thus immediately obtain the claim, without any extra calculation. □

Bibliography

Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, 9:1981–2014.

Bai, Z. and Silverstein, J. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Banerjee, D. (2018). Contiguity and non-reconstruction results for planted partition models: the dense case. *Electron. J. Probab.*, 23:485–512.

Banerjee, D. and Ma, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv:1705.05305*.

Banks, J., Moore, C., Neeman, J., and Netrapalli, P. (2016). Information-theoretic thresholds for community detection in sparse networks. In *COLT*, pages 383–416.

Béjar, J., Álvarez, S., García, D., Gómez, I., Oliva, L., Tejeda, A., and Vázquez-Salceda, J. (2016). Discovery of spatio-temporal patterns from location-based social networks. *J. Exp. Theor. Artif. Intell.*, 28(1-2):313–329.

Brualdi, R. (1974). The dad theorem for arbitrary row sums. *Proc. Amer. Math. Soc*, 45(2):189–194.

Bubeck, S., Ding, J., Eldan, R., and Rácz, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Struct. Algor.*, 49(3):503–532.

Chen, K. and Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.*, 113(521):241–251.

Chen, Y., Li, X., Xu, J., et al. (2018). Convexified modularity maximization for degree-corrected stochastic block models. *The Annals of Statistics*, 46(4):1573–1602.

Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and computing*, 18(2):173–183.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM J. Numer. Anal.*, 7(1):1–46.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32:962–994.

Du, H. and Yang, S. J. (2011). Discovering collaborative cyber attack patterns using social network analysis. In *Proceedings of the 4th International Conference on Social Computing, Behavioral-cultural Modeling and Prediction*, SBP'11, pages 129–136. Springer-Verlag.

Efron, B. (2004). Large-scale simultaneous hypothesis testing. *J. Amer. Statist. Soc.*, 99(465):96–104.

Fu, Y.-H., Huang, C.-Y., and Sun, C.-T. (2015). Using global diversity and local topology features to identify influential network spreaders. *Physica A.*, 433:344 – 355.

Gao, C. and Lafferty, J. (2017). Testing for global network structure using small subgraph statistics. *arXiv:1710.00862*.

Gao, C., Ma, Z., Zhang, A. Y., Zhou, H. H., et al. (2018). Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*. Cambridge University Press.

Hu, J., Zhang, J., Qin, H., Yan, T., and Zhu, J. (2020). Using maximum entry-wise deviation to test the goodness-of-fit for stochastic block models. *J. Amer. Statist. Assoc. to appear*.

Ingster, Y. I., Tsybakov, A. B., and Verzelen, N. (2010). Detection boundary in sparse regression. *Electron. J. Statist.*, 4:1476–1526.

Ji, P. and Jin, J. (2016). Coauthorship and citation networks for statisticians. *Ann. Appl. Statist., to appear*.

Ji, P., Jin, J., Ke, Z. T., and Li, W. (2020). Statistics about statisticians. *Manuscript*.

Jin, J. (2015). Fast community detection by score. *Ann. Statist.*, 43(1):57–89.

Jin, J., Ke, Z. T., and Luo, S. (2017). Estimating network memberships by simplex vertices hunting. *arXiv:1708.07852*.

Jin, J., Ke, Z. T., and Luo, S. (2018). Network global testing by counting graphlets. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 2338–2346.

Jin, J., Ke, Z. T., and Luo, S. (2019). Optimal adaptivity of signed-polygon statistics for network testing. *arXiv:1904.09532*.

Jin, J., Ke, Z. T., and Luo, S. (2021a). Improvements on score, especially for weak signals. *Sankhya A*.

Jin, J., Ke, Z. T., Luo, S., and Ma, Y. (2021b). Optimal network pairwise comparison. *Manuscript*.

Jin, J., Ke, Z. T., Luo, S., and Wang, M. (2020). Estimating the number of communities by stepwise goodness-of-fit. *arxiv:2009.09177*.

Johnson, C. and Reams, R. (2009). Scaling of symmetric matrices by positive diagonal congruence. *Linear Multilinear A*, 57:123–140.

Karrer, B. and Newman, M. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83(1):016107.

Ke, Z. T. (2019). Comparison of different network testing procedures. *Manuscript*.

Latouche, P., Birmele, E., and Ambroise, C. (2012). Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115.

Le, C. M. and Levina, E. (2015). Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, 44(1):401–424.

Lei, L., Li, X., and Lou, X. (2020). Consistency of spectral clustering on hierarchical stochastic block models. *arXiv:2004.14531*.

Li, T., Lei, L., Bhattacharyya, S., Sarkar, P., Bickel, P. J., and Levina, E. (2018). Hierarchical community detection by recursive partitioning. *arXiv:1810.01509*.

Li, T., Levina, E., and Zhu, J. (2020). Network cross-validation by edge sampling. *Biometrika*, 107(2):257–276.

Liu, F., Choi, D., Xie, L., and Roeder, K. (2017). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci.*, 115(5):927–932.

Liu, W., Jiang, X., Pellegrini, M., and Wang, X. (2016). Discovering communities in complex networks by edge label propagation. *Scientific Reports*, 6:22470.

Liu, Y., Hou, Z., Yao, Z., Bai, Z., Hu, J., and Zheng, S. (2019). Community detection based on the l_∞ convergence of eigenvectors in dcbm. *arXiv:1906.06713*.

Ma, S., Su, L., and Zhang, Y. (2018). Determining the number of communities in degree-corrected stochastic block models. *arXiv:1809.01028*.

Ma, Z., Ma, Z., and Yuan, H. (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.*, 21(4):1–67.

Ma, Z. and Wu, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.*, 43(3):1089–1116.

Marshall, A. and Olkin, I. (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12(1):83–90.

Mossel, E., Neeman, J., and Sly, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Relat. Fields*, 162(3-4):431–461.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Stat. Sin.*, 17(4):1617–1642.

Saldana, D., Yu, Y., and Feng, Y. (2017). How many communities are there? *J. Comput. Graph Stat.*, 26(1):171–181.

Saldaña, D. F., Yu, Y., and Feng, Y. (2017). How many communities are there? *Journal of Computational and Graphical Statistics*, 26(1):171–181.

Shorack, G. and Wellner, J. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons.

Sinkhorn, R. (1974). Diagonal equivalence to matrices with prescribed row and column sums. ii. *Proc. Amer. Math. Soc*, 45(2):195–198.

Tropp, J. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434.

Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.

Wang, Y. R., Bickel, P. J., et al. (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2):500–528.

Xu, M., Jog, V., and Loh, P.-L. (2020). Optimal rates for community estimation in the weighted stochastic block model. *Ann. Statist.*, 48(1):183–204.

Zhang, Y., Levina, E., and Zhu, J. (2014). Detecting overlapping communities in networks with spectral methods. *arXiv:1412.3432*.

Zhao, Y., Levina, E., and Zhu, J. (2011). Community extraction for social networks. *Proc. Natl. Acad. Sci.*, 108(18):7321–7326.