

# UNIVERSAL RANK INFERENCE VIA RESIDUAL SUBSAMPLING WITH APPLICATION TO LARGE NETWORKS

BY XIAO HAN<sup>1,a</sup>, QING YANG<sup>1,\*,b</sup> AND YINGYING FAN<sup>2,\*,c</sup>

<sup>1</sup>*International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China*, <sup>a</sup>*xhan011@ustc.edu.cn*; <sup>b</sup>*yangq@ustc.edu.cn*

<sup>2</sup>*Data Sciences and Operations Department, University of Southern California, Los Angeles, CA 90089*, <sup>c</sup>*fanyingy@marshall.usc.edu*

Determining the precise rank is an important problem in many large-scale applications with matrix data exploiting low-rank plus noise models. In this paper, we suggest a universal approach to rank inference via residual subsampling (RIRS) for testing and estimating rank in a wide family of models, including many popularly used network models such as the degree corrected mixed membership model as a special case. Our procedure constructs a test statistic via subsampling entries of the residual matrix after extracting the spiked components. The test statistic converges in distribution to the standard normal under the null hypothesis, and diverges to infinity with asymptotic probability one under the alternative hypothesis. The effectiveness of RIRS procedure is justified theoretically, utilizing the asymptotic expansions of eigenvectors and eigenvalues for large random matrices recently developed in [11] and [12]. The advantages of the newly suggested procedure are demonstrated through several simulation and real data examples.

**1. Introduction.** Matrix data have been popularly encountered in various big data applications. For example, many science and social applications involve individuals with complicated interaction systems. Such systems can often be modeled using a network with nodes representing the  $n$  individuals and edges representing the connectivity among individuals. The overall connectivity can thus be recorded in an  $n \times n$  adjacency matrix whose entries are zeros and nonzeros, representing the corresponding pair of nodes unconnected or connected, respectively. Examples include the friendship network, the citation network, the predator-prey interaction network, and many others.

There has been a large literature on statistical methods and theory proposed for analyzing matrix data. In the network setting, the observed adjacency matrix is frequently modeled as the summation of a latent deterministic low rank mean matrix and a random noise matrix, where the former stores all useful information in the data and is often the interest. One popular assumption is that the rank  $K$  of the latent mean matrix is known. However, in practice, such  $K$  is generally unknown and needs to be estimated. This paper focuses on estimation and inference on the low rank  $K$  in a general model setting including many popularly used network models as special cases.

In our model, the data matrix  $\mathbf{X}$  can be roughly decomposed as a low rank mean matrix  $\mathbf{H}$  with  $K$  spiked eigenvalues and a noise matrix  $\mathbf{W}$  whose components are mostly independent. Here,  $K$  is unknown and allowed to slowly diverge with  $n$ . To infer  $K$  with quantified statistical uncertainty, we propose a universal approach for Rank Inference by Residual Subsampling (RIRS). Specifically, we consider the hypothesis test

$$(1) \quad H_0 : K = K_0 \text{ vs. } H_1 : K > K_0$$

---

\*Yingying Fan and Qing Yang serve as co-corresponding authors.

*MSC2020 subject classifications:* Primary 62F03, 62F12; secondary 60B20, 62F35.

*Keywords and phrases:* Rank inference, Robustness, Low-rank models, High dimensionality, Asymptotic expansions, Eigenvectors, Eigenvalues, Large random matrices.

with  $K_0$  some pre-specified positive integer. The spiked mean matrix with rank  $K_0$  can be estimated by eigen decomposition, subtracting which from the observed data matrix yields the residual matrix. Then by appropriately subsampling the entries of the residual matrix, we can construct a test statistic. We prove that under the null hypothesis, the test statistic converges in distribution to the standard normal, and under the alternative hypothesis, some spiked structure remains in the residual matrix and the constructed test statistic behaves very differently. Thus, the hypothesis test in (1) can be successfully conducted. Then by sequentially testing the hypothesis (1) for  $K_0 = 1, \dots, K_{\max}$  with  $K_{\max}$  some large enough positive integer, we can estimate  $K$  as the first integer that makes our test fail to reject. We provide theoretical justifications on the effectiveness of our procedure. We show that the size of our test tends to the desired level  $\alpha$  as sample size increases, and establish conditions under which the power approaches one asymptotically. We also show that the sequential procedure correctly estimates the true rank with probability tending to  $1 - \alpha$  as network size increases.

A key to RIRS’s success is the subsampling step. Although the noise matrix  $\mathbf{W}$  has mostly independent components, the residual matrix is only an estimate of  $\mathbf{W}$  and has correlated components. Intuitively speaking, if too many entries of the residual matrix are sampled, the accumulated estimation error and the correlation among sampled entries would be too large, rendering the asymptotic normality invalid. We provide both theoretical and empirical guidance on how many entries to subsample. In the special case where the diagonals of the data matrix  $\mathbf{X}$  are nonzero independent random variables (which corresponds to selfloops in network models), a special deterministic sampling scheme can be used and the RIRS test takes a simpler form.

The structure of low rank mean matrix plus noise matrix is very general and includes many popularly used network models such as the Stochastic Block Model (SBM, [14, 25, 1]), Degree Corrected SBM (DCSBM, [17]), Mixed Membership (MM) Model, and Degree Corrected Mixed Membership (DCMM) Model [3] as special cases. RIRS procedure and the theory established in this paper are applicable to all these network models and go beyond them. In network model settings, RIRS can accommodate sparse networks and allows for extreme degree heterogeneity.

Substantial efforts have been made in the literature on estimating  $K$  in some specific network models, where  $K$  is referred to as the number of communities. For example, [22] proposed an MCMC algorithm based on the allocation sampler to cluster the nodes in SBM and simultaneously estimate  $K$ . [3] developed a general variational inference algorithm to estimate the parameters in MM model with  $K$  chosen according to some BIC criterion. [15] considered testing (1) with  $K_0 = 1$  and proposed a signed polygon statistic which can accommodate the degree heterogeneity in the DCMM model. [13] proposed EZ statistics constructed by “frequencies of three-node subgraphs” to test (1) with  $K_0 = 1$  in the setting of DCSBM. [5] introduced a linear spectral statistic to test  $H_0 : K = 1$  vs.  $H_1 : K = 2$  under the SBM. [16] proposed a stepwise goodness-of-fit test for estimating  $K$  under DCSBM. [27] suggested an adjusted chi-square test to address the goodness-of-fit testing and model selection problem for DCSBM. Compared to these works, we consider more general model and general positive integer  $K_0$  that can be larger than 1.

There is also a popular line of work using likelihood based methods to estimate  $K$ . For example, [10], [18], [24], and [26], among others. [8] introduced a universal singular value thresholding procedure for the matrix estimation which can be applied to estimate  $K$ . [9] proposed a network cross-validation method for estimating  $K$  and proved the consistency of the estimator under SBM. The cross-validation idea was also explored in [21] under the widely used inhomogeneous Erdős-Renyi model with low rank mean matrix via edge sampling. [19] proposed to estimate  $K$  using the spectral properties of two graph operators – the non-backtracking matrix and the Bethe Hessian matrix. [28] proposed to sequentially extract

one community at a time by optimizing some extraction criterion, based on which they proposed a hypothesis test for testing the number of communities empirically via permutation method. [7] proposed a new test based on the asymptotic distribution of the largest eigenvalue of the appropriately rescaled adjacency matrix for testing whether a network is Erdős Renyi or not, and suggested a recursive bipartition algorithm for estimating  $K$ . [20] generalized the test in [7] for testing whether a network is SBM with some specific  $K_0$ , and proposed a sequential testing idea to estimate the true number of communities.

Among the existing literature reviewed above, the works by [7] and [20] are most closely related to ours. The main idea in both papers is that under the null hypothesis of SBM with  $K_0$  communities, the model parameters can be estimated and the residual matrix can be calculated and appropriately rescaled. The rescaled residual matrix is close to a generalized Wigner matrix whose extreme eigenvalues (after recentering and rescaling) converge in distribution to the Tracy-Widom distribution. However, under the alternative hypothesis, the extreme eigenvalues behave very differently. At a high level, this idea is related to ours in the sense that our proposal is also based on the residual matrix.

RIRS test differs from the literature in the way of using the residual matrix. Instead of investigating the spectral distribution of the residual matrix, we construct RIRS test by subsampling just a fraction of the entries in residual matrix. The subsampling idea ensures that the noise accumulation caused by estimating the mean matrix does not dominate the signal which guarantees the nice performance of our test. Compared to the existing literature, RIRS test behaves more similarly to a nonparameteric test in the sense that there is no assumption on specific structure of the low rank mean matrix. Yet, it is also simple and fast to implement. Our asymptotic theory is also new to the literature. It is built on the recent developments on random matrix theory in [11] and [12], which establishes the asymptotic expansions of the eigenvalues and eigenvectors for a very general class of random matrices. This powerful result allows us to establish the sampling properties of RIRS test in equally general setting.

The remaining of the paper is organized as follows: Section 2 presents the model setting and motivation for RIRS. We introduce our new approach and establish its asymptotic properties in Section 3. Simulations under various models are conducted to justify the performance of RIRS in Section 4. We further apply RIRS to a real data example in Section 5. Additional simulation examples and all proofs are relegated to the Appendix and the Supplementary material.

1.1. *Notations.* We introduce some notations that will be used throughout the paper. We use  $a \ll b$  to represent  $a/b \rightarrow 0$  and write  $a \lesssim b$  if there exists a positive constant  $c$  such that  $0 \leq a \leq cb$ . For a matrix  $\mathbf{A}$ , we use  $\lambda_j(\mathbf{A})$  to denote the  $j$ -th largest eigenvalue, and  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{A}\|$ , and  $\|\mathbf{A}\|_\infty$  to denote the Frobenius norm, the spectral norm, and the maximum element-wise infinity norm, respectively. In addition, denote by  $\mathbf{A}(k)$  the  $k$ th row of the matrix  $\mathbf{A}$ . For a unit vector  $\mathbf{x} = (x_1, \dots, x_n)^T$ , let  $d_x = \|\mathbf{x}\|_\infty = \max |x_i|$  represent the vector infinity norm.

## 2. Model setting and motivation.

2.1. *Model setting.* Consider an  $n \times n$  symmetric random matrix  $\tilde{\mathbf{X}}$  admitting the following decomposition

$$(2) \quad \tilde{\mathbf{X}} = \mathbf{H} + \mathbf{W},$$

where  $\mathbf{H} = \mathbb{E}(\tilde{\mathbf{X}})$  is the mean matrix with unknown rank  $K \ll n$  and  $\mathbf{W}$  is the noise matrix with bounded and independent entries on and above the diagonals. In network applications, the observed matrix  $\mathbf{X}$  is the adjacency matrix and can be either  $\tilde{\mathbf{X}}$  or  $\tilde{\mathbf{X}} - \text{diag}(\tilde{\mathbf{X}})$ , with

the former corresponding to network with selfloops and the latter corresponding to network without selfloops, respectively. An important and interesting question is inferring the unknown rank  $K$ , which corresponds to the number of communities in network models. We address the problem by testing the hypotheses (1) under the universal model (2).

We note that with some transformation, model (2) can accommodate nonsymmetric matrices. In fact, for any matrix  $\tilde{\mathbf{X}}$  that can be written as the summation of a rank  $K$  mean matrix and a noise matrix of independent components, we can define a new matrix as

$$\begin{pmatrix} \mathbf{0} & \tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}^T & \mathbf{0} \end{pmatrix}.$$

It is seen that this new matrix has the same structure as in (2) with rank  $2K$ , and our new method and theory both apply. For simplicity of presentation, hereafter we assume the symmetric matrix structure for  $\tilde{\mathbf{X}}$  and  $\mathbf{X}$ .

Write the eigen-decomposition of  $\mathbf{H}$  as  $\mathbf{VDV}^T$ , where  $\mathbf{D} = \text{diag}(d_1, \dots, d_K)$  collects the nonzero eigenvalues of  $\mathbf{H}$  in decreasing magnitude and  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_K)$  is the matrix collecting the corresponding eigenvectors. Denote by  $\hat{d}_1, \dots, \hat{d}_n$  the eigenvalues of  $\mathbf{X}$  in decreasing magnitude and  $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_n$  the corresponding eigenvectors. We next discuss the motivation of RIRS.

*2.2. Motivation.* To gain insights, consider the simple case when the observed data matrix  $\mathbf{X} = \tilde{\mathbf{X}}$  and follows model (2). It is seen that  $\mathbb{E}\mathbf{W} = \mathbf{0}$ . Intuitively, as  $n \rightarrow \infty$ , the normalized statistic  $\sum_{i=1}^n w_{ii} / \sqrt{\sum_{i=1}^n \mathbb{E}w_{ii}^2}$  converges in distribution to standard normal. Meanwhile, we expect  $\sum_{i=1}^n \mathbb{E}w_{ii}^2 / \sum_{i=1}^n w_{ii}^2$  to converge to 1 in probability as  $n \rightarrow \infty$ . These two results entail that

$$(3) \quad \frac{\sum_{i=1}^n w_{ii}}{\sqrt{\sum_{i=1}^n w_{ii}^2}}$$

is asymptotically normal as the matrix size  $n \rightarrow \infty$ .

In the ideal case where the eigenvalues  $d_1, \dots, d_K$  and eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_K$  are known, a test of the form (3) can be constructed by replacing  $w_{ii}$  with  $\tilde{w}_{ii}$  where  $\tilde{\mathbf{W}} = (\tilde{w}_{ij}) = \mathbf{X} - \sum_{k=1}^{K_0} d_k \mathbf{v}_k \mathbf{v}_k^T$ . Under the null hypothesis,  $\tilde{\mathbf{W}} = \mathbf{W}$  and the corresponding test statistic (constructed in the same way as (3)) is asymptotically normal. However, under the alternative hypothesis,  $\tilde{\mathbf{W}}$  still contains some information from the  $K - K_0$  smallest spiked eigenvalues and the corresponding eigenvectors and the test statistic is expected to exhibit different asymptotic behavior. Thus, the hypotheses in (1) can be successfully tested by using this statistic.

In practice, the eigenvalues and eigenvectors of  $\mathbf{H}$  are unavailable and need to be estimated. A natural estimate of  $\tilde{\mathbf{W}}$  takes the form

$$(4) \quad \widehat{\tilde{\mathbf{W}}} = (\hat{w}_{ij}) = \mathbf{X} - \sum_{k=1}^{K_0} \hat{d}_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T.$$

Under  $H_0$ , the residual matrix  $\widehat{\tilde{\mathbf{W}}}$  is expected to be close to  $\mathbf{W}$ , which motivates us to consider test of the form

$$(5) \quad \tilde{T}_n = \frac{\sum_{i=1}^n \hat{w}_{ii}}{\sqrt{\sum_{i=1}^n \hat{w}_{ii}^2}}.$$

Intuitively, the asymptotic behavior of the above statistic should be close to the one in (3). Thus, by examining the asymptotic behavior of  $\tilde{T}_n$  we can test the desired hypotheses. We will formalize this intuition in a later section.

The statistic in (3) is constructed using only the diagonals of  $\mathbf{W}$ . In theory, the asymptotic normality remains true if we aggregate any randomly sampled entries of the matrix  $\mathbf{W}$  (instead of just the diagonals) and normalize properly, as long as the sampling size is large enough, thanks to the independence of the entries on and above the diagonals of  $\mathbf{W}$ . However, this does not translate into the asymptotic normality of the test based on  $\widehat{\mathbf{W}}$  for at least two reasons: First, in applications absence of selfloops, the observed data matrix  $\mathbf{X}$  takes the form  $\widetilde{\mathbf{X}} - \text{diag}(\widetilde{\mathbf{X}})$  and thus  $\widehat{\mathbf{W}}$  estimates  $\mathbf{W} - \text{diag}(\widetilde{\mathbf{X}})$  which has nonrandom diagonals. Consequently, test constructed using diagonals of  $\widehat{\mathbf{W}}$  becomes invalid. Second, the entries of  $\widehat{\mathbf{W}}$  are all correlated and have errors caused by estimating the corresponding entries of  $\mathbf{W}$ . Aggregating too many entries of  $\widehat{\mathbf{W}}$  will cause too much noise accumulation. This together with the correlations among  $\hat{w}_{ij}$  makes the asymptotic normality of the corresponding test statistic invalid. This heuristic argument is formalized in a later Section 3.6. Thus to overcome these difficulties, we need to carefully choose which and how many entries to aggregate. These issues are formally addressed in the next section.

### 3. Rank inference via residual subsampling.

3.1. *A universal RIRS test.* A key ingredient of RIRS is subsampling the entries of  $\widehat{\mathbf{W}}$ . Specifically, define i.i.d. Bernoulli random variables  $Y_{ij}$  with  $\mathbb{P}(Y_{ij} = 1) = \frac{1}{m}$  for  $1 \leq i < j \leq n$ , where  $m$  is some positive integer diverging with  $n$  at a rate that will be specified later. In addition, set  $Y_{ji} = Y_{ij}$  for  $i < j$ . A universal RIRS test that works under the broad model (2) takes the following form

$$(6) \quad T_n = \frac{\sqrt{m} \sum_{i \neq j} \hat{w}_{ij} Y_{ij}}{\sqrt{2 \sum_{i \neq j} \hat{w}_{ij}^2}}.$$

The effect of  $m$  is to control on average how many entries of the residual matrix to aggregate for calculating the test statistic. It will be made clear in a moment that  $n^2/m$  needs to grow to infinity in order for the central limit theorem to take effect. However, the growth rate cannot be too fast because otherwise the noise accumulation and the correlation in  $\hat{w}_{ij}$  would make the asymptotic normality invalid.

The following conditions will be used in our theoretical analysis.

CONDITION 1.  $\mathbf{W}$  is a symmetric matrix with independent and bounded upper triangular entries (including the diagonals) and  $\mathbb{E}w_{ij} = 0$  for  $i \neq j$ .

CONDITION 2. For  $1 \leq i < j \leq K$ , if  $|d_i| \neq |d_j|$ , there exists a positive constant  $c_0$  such that  $\frac{|d_i|}{|d_j|} \geq 1 + c_0$ .

CONDITION 3. There exists a positive sequence  $\theta_n$ , which may converge to 0 as  $n \rightarrow \infty$ , such that  $\sigma_{ij}^2 = \text{var}(w_{ij}) \leq \theta_n$  and  $\max_{1 \leq i \leq n} |h_{ii}| \lesssim \theta_n$ . In addition,  $\max\{n\theta_n, \log n\} \lesssim \alpha_n^2 := \max_i \sum_{j=1}^n \sigma_{ij}^2$ ,  $|d_K| \gtrsim \frac{\alpha_n^2}{(\log \log n)^{\epsilon'}}$  and  $\frac{|d_K|}{\alpha_n} \gtrsim (\log n)^{1+\epsilon}$  for some positive constants  $\epsilon, \epsilon'$ .

CONDITION 4.  $\|\mathbf{V}\|_\infty \lesssim \frac{1}{\sqrt{n}}$ .

CONDITION 5. It holds that  $\sum_{i \neq j} \sigma_{ij}^2 \gg m$ . In addition, for some small positive constant  $\epsilon_1 < \epsilon$  with  $\epsilon$  the constant in Condition 3,

$$\sum_{i \neq j} \sigma_{ij}^2 \gtrsim (\log n)^{\epsilon_1} \left( \frac{n \sum_{k=1}^K (\mathbf{1}^T \mathbf{v}_k)^2}{m} + \alpha_n^2 (\log n)^2 + \frac{n}{m} + \frac{n^2 \alpha_n^2 (\log n)^6}{m d_K^2} \right).$$

CONDITION 6. *The true rank satisfies that  $K \leq O(\log \log n)$ .*

With the above conditions, Theorem 3.1 below provides the asymptotic null distribution of RIRS test and Theorem 3.2 establishes the asymptotic alternative distribution.

THEOREM 3.1. *Assume Conditions 1-6. Under null hypothesis in (1) we have*

$$(7) \quad T_n \xrightarrow{d} N(0, 1), \text{ as } n \rightarrow \infty.$$

THEOREM 3.2. *Assume Conditions 1-6 and the alternative hypothesis in (1). If*

$$\sum_{i \neq j} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2 \ll \sum_{i \neq j} \sigma_{ij}^2,$$

then as  $n \rightarrow \infty$ ,

$$(8) \quad T_n - \frac{\sqrt{m} \sum_{i \neq j} \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}}{\sqrt{2 \sum_{i \neq j} \hat{w}_{ij}^2}} \xrightarrow{d} N(0, 1).$$

In addition, if

$$(9) \quad \left| \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) \right| \gg \sqrt{m} \left( \sqrt{\sum_{i \neq j} \sigma_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k| \right),$$

we have

$$(10) \quad \mathbb{P}(|T_n| > C) \rightarrow 1, \text{ as } n \rightarrow \infty$$

for arbitrarily large positive constant  $C$ .

The result in (8) guarantees that if  $\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}$  is non-negligible compared with  $\sqrt{2m^{-1} \sum_{i \neq j} \hat{w}_{ij}^2}$ , then our test has non-vanishing power. If in addition, the asymptotic mean is large enough such that (9) is satisfied, then the asymptotic power can reach one. The results on asymptotic size and power of RIRS test are formally summarized in the following Corollary.

COROLLARY 1. *Under the conditions of Theorem 3.1, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n| \geq \Phi^{-1}(1 - \alpha/2) | H_0) = \alpha,$$

where  $\Phi^{-1}(t)$  is the inverse of the standard normal distribution function, and  $\alpha$  is the pre-specified significance level. Alternatively under the same conditions for ensuring (10), we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n| \geq \Phi^{-1}(1 - \alpha/2) | H_1) = 1.$$

3.2. *Remarks on the conditions.* Random matrix satisfying Condition 1 is often termed as generalized Wigner matrix in the literature. Condition 2 allows for eigenvalue multiplicity and requires that there is enough gap between distinct eigenvalues. The constant  $c_0$  can be replaced with some slowly vanishing term such as  $(\log n)^{-1}$  and our main results will still hold with relevant conditions updated accordingly.

Condition 3 constraints that the nonzero eigenvalues of the low rank mean matrix should have enough spikiness. The two constraints  $|d_K| \gtrsim \alpha_n^2 (\log \log n)^{-\epsilon'}$  and  $|d_K| \gtrsim \alpha_n (\log n)^{1+\epsilon}$  in Condition 3 are imposed for controlling the noise accumulation in our test caused by estimating  $w_{ij}$ . The second constraint is a signal strength condition commonly imposed in random matrix theory literature; see [4], [6] and [23], among others. The logarithmic factor,  $(\log n)^{1+\epsilon}$ , measures the gap between the signal (spiked) eigenvalue and the noise eigenvalue, and is hard to be removed completely because otherwise the sample eigenvector would depend on the noise matrix  $\mathbf{W}$  in a complicated way that is not useful for statistical inference. The first constraint can be satisfied by many network models with low rank structure. To see this, note that if  $X_{ij}, j \geq i \geq 1$  follows Bernoulli distribution and  $h_{ij} \sim \theta_n$  with  $\max_{i,j} h_{ij} < 1$ , then  $\alpha_n^2 \sim n\theta_n$ . Since  $h_{ij}$ 's and  $\sigma_{ij}^2$ 's are the means and variances of Bernoulli random variables, respectively, we have  $h_{ij} \sim \sigma_{ij}^2 \sim \theta_n$  and  $\|\mathbf{H}\|_F = \{\sum_{i,j} h_{ij}^2\}^{1/2} \sim n\theta_n$ . Note also that  $\|\mathbf{H}\|_F = \{\sum_{i=1}^K d_i^2\}^{1/2}$ . Assuming that  $K$  is finite and  $d_1 \sim d_K$ , these results together with  $\alpha_n^2 \sim n\theta_n$  derived earlier ensure that  $|d_K| \gtrsim \alpha_n^2 (\log \log n)^{-\epsilon'}$  is satisfied.

Condition 5 characterizes what kind of  $m$  can make RIRS succeed. More detailed discussion on the choice of  $m$  will be given in Section 3.3. Condition 6 allows the rank  $K$  to grow with network size  $n$ .

Condition 4 is a technical condition needed for proving key Lemmas 2-3. We remark that it can hold under extreme degree heterogeneity in network models. The following example is used to illustrate this point.

EXAMPLE 1. Consider DCSBM with  $K = 2$  where mean matrix takes the form

$$(11) \quad \mathbf{H} = \Theta \mathbf{\Pi} \mathbf{B} \mathbf{\Pi}^T \Theta.$$

Here,  $\mathbf{B}$  is a  $2 \times 2$  nonsingular matrix with diagonals 1 and off diagonals taking a constant value in  $[0, 1)$ ,  $\Theta$  is a diagonal matrix with the first  $n/2$  diagonal entries taking the same value  $\vartheta_1 > 0$  and the remaining diagonal entries taking the same value  $\vartheta_2 > 0$ , and  $\mathbf{\Pi} \in \mathbb{R}^{n \times 2}$  has the first  $n/2$  rows equal to  $(1, 0)$  and the remaining ones equal to  $(0, 1)$ . Here, for simplicity we assume  $n$  is an even number. It is seen that the first  $n/2$  nodes belong to community 1 and share the common degree parameter  $\vartheta_1$ , and the remaining belong to community 2 and share the common degree parameter  $\vartheta_2$ . Since the population eigenvector  $\mathbf{v}_k, k = 1, 2$  satisfies

$$\Theta \mathbf{\Pi} \mathbf{B} \mathbf{\Pi}^T \Theta \mathbf{v}_k = \lambda_k \mathbf{v}_k,$$

we see that  $\mathbf{v}_k$  takes the form  $(a_1 \mathbf{1}_{n/2}^T, a_2 \mathbf{1}_{n/2}^T)^T$  with  $\mathbf{1}_{n/2} \in \mathbb{R}^{n/2}$  a vectors of 1's and  $a_1$  and  $a_2$  two constants. Since  $\|\mathbf{v}_k\|_2 = 1$ , it follows that  $\max\{|a_1|, |a_2|\} \lesssim \frac{1}{\sqrt{n}}$  and thus Condition 4 holds regardless of the values of  $\vartheta_1$  and  $\vartheta_2$ .

3.3. Choice of  $m$ . It is seen from the previous two theorems that the tuning parameter  $m$  plays a crucial role for RIRS to achieve the desired size with power tending to one. Condition 5 provides general conditions on the choice of  $m$  for ensuring the null and alternative distributions in (7) and (8). For (10) to hold, we also need the additional assumption (9). In some special cases, these conditions boil down to simpler forms which can provide us more specific guideline on the choice of  $m$ .

As an example, we consider the special case

$$(12) \quad \begin{aligned} & (\log n)^\delta \min_{i \neq j} \sigma_{ij}^2 \gtrsim \max_{i \neq j} \sigma_{ij}^2, \\ & \text{for } K_0 < K, \sum_{i \neq j} \sigma_{ij}^2 \lesssim \left| \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) \right|, \text{ and } |d_1| \lesssim n\theta_n, \end{aligned}$$

where  $\delta < \epsilon$  is some small positive constant with  $\epsilon$  the same as in Condition 3. The first condition above is for guaranteeing both the asymptotic size and power results, while the remaining two conditions will be only used for verifying (9) which is only needed in establishing the power results. In network models, the  $(\log n)^\delta$  factor in the first condition of (12) is related to degree heterogeneity. For very extreme degree heterogeneity, we may have  $\max_{i \neq j} \sigma_{ij}^2 / \min_{i \neq j} \sigma_{ij}^2$  diverge at some polynomial rate of sample size. We remark that similar results can be derived using identical proof idea for Theorem 3.3 with some technical conditions appropriately modified.

We next discuss some more specific network models which can give us more insights on the three conditions in (12). Consider the same DCSBM in (11) except that  $\Theta = \text{diag}\{\vartheta_1, \dots, \vartheta_n\}$  with  $\vartheta_j > 0, j = 1, \dots, n$  the degree parameters. Assume  $\max_{i,j} \{\vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}\}$  is bounded away from 1 by some constant, where  $\mathcal{C}_i \in \{1, 2\}$  is the membership for node  $i$ , and  $b_{kl} \in (0, 1]$  is the  $(k, l)$  entry of matrix  $\mathbf{B}$  taking constant values. Then we have  $\sigma_{ij}^2 \sim h_{ij} = \vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}$  because the entries of  $\mathbf{X}$  have Bernoulli distributions. It is seen that the ratio  $\max_{i \neq j} \sigma_{ij}^2 / \min_{i \neq j} \sigma_{ij}^2 \sim \max_{i \neq j} (\vartheta_i \vartheta_j) / \min_{i \neq j} (\vartheta_i \vartheta_j)$ . Thus, the first condition in (12) is satisfied if  $\max_j \vartheta_j / \min_j \vartheta_j \lesssim \sqrt{(\log n)^\delta}$ . It is also straightforward to see that  $\theta_n = \max_{i,j} \{\vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}\}$  and  $|d_1| \sim \|\mathbf{H}\|_F \lesssim n \max_{i,j} \{\vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}\}$ . Thus, the last condition  $|d_1| \lesssim n\theta_n$  in (12) holds. The second condition reduces to  $\sum_{i \neq j} \sigma_{ij}^2 \lesssim \left| d_2 \sum_{i \neq j} \mathbf{v}_2(i) \mathbf{v}_2(j) \right|$  in this model setting. Since  $\sum_{i \neq j} \sigma_{ij}^2 \leq n^2 \max_{i \neq j} \{\vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}\}$ , the second condition is satisfied if  $n^2 \max_{i \neq j} \vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j} \lesssim \left| d_2 \sum_{i \neq j} \mathbf{v}_2(i) \mathbf{v}_2(j) \right|$ .

The next theorem specifies what kind of  $m$  satisfies the two inequalities in Condition 5 and (9).

**THEOREM 3.3.** *Set  $\theta_n = \max_{i \neq j} \sigma_{ij}^2$ . Assume (12) and Condition 3, and let  $\epsilon_1 \in (\delta, \epsilon)$  be some small constant. Then  $m$  satisfying the following condition*

$$(13) \quad \theta_n^{-1} (\log n)^{\delta + \epsilon_1} K + n^{-1} \theta_n^{-2} (\log n)^{6 + 2\delta + \epsilon_1} (\log \log n)^{2\epsilon'} \ll m \ll (n/K)^2 \theta_n (\log n)^{-2\delta}$$

*makes Condition 5 and inequality (9) hold. Consequently, (7) and (10) hold under Conditions 1–4 and 6. Moreover, a sufficient condition for (13) is  $n(\log n)^{-2\epsilon + \epsilon_1 + 2} (\log \log n)^{-2\epsilon'} K \ll m \ll nK^{-2} (\log n)^{2\epsilon - \delta + 2} (\log \log n)^{2\epsilon'}$  under Conditions 1–4.*

It is seen that Theorem 3.3 allows for a wide range of values for  $m$ . In theory, any  $m$  satisfying (13) guarantees the asymptotic size and power of our test. In implementation, we found smaller  $m$  in this range yields better empirical size. It is also seen from (13) that RIRS works with sparse networks. In fact, the only sparsity condition imposed by (13) is that  $\theta_n \gg n^{-1} K^{3/2} (\log n)^{2 + \delta + \epsilon_1/2}$ . Our sparsity condition is weaker than many existing ones in related work in the literature. In particular, both [7] and [20] considered dense SBM with  $\theta_n$  bounded below by some constant.

We remark that sparse models have been considered in the network literature, though mostly in estimation instead of inference problems. For example, [26] proposed a model selection criterion for estimating  $K$  under the sparse setting of SBM with  $n\theta_n / \log n \rightarrow \infty$ . [19] established the consistency of their method for estimating  $K$  under the setting  $n\theta_n = O(1)$ . We consider the statistical inference problem of hypothesis testing, which involves more delicate analyses for establishing the asymptotic distributions of the test statistic.

**3.4. A special case: networks with selfloops.** We formalize the heuristic arguments in Section 2.2 about the ratio statistic  $\tilde{T}_n$  in (5) when the network admits selfloops. In such a case, the general test (6) still works. However, the simpler one  $\tilde{T}_n$  can enjoy similar asymptotic properties without the trouble of choosing  $m$ .



**THEOREM 3.4.** *Suppose that Conditions 1-4 and 6 hold, the network contains selfloops and  $\sqrt{\sum_{i=1}^n \sigma_{ii}^2} \gg (\log n)^{2+\epsilon_2}$  for some positive constant  $\epsilon_2$ .*

(i) *Under null hypothesis we have*

$$(14) \quad \tilde{T}_n \xrightarrow{d} N(0, 1), \text{ as } n \rightarrow \infty.$$

(ii) *Under alternative hypothesis, if further  $\sum_{i=1}^n (\sum_{k=K_0+1}^K d_k v_k^2(i))^2 \ll \sum_{i=1}^n \sigma_{ii}^2$ , we have*

$$(15) \quad \tilde{T}_n - \frac{\sum_{k=K_0+1}^K d_k}{\sqrt{\sum_{i=1}^n \hat{w}_{ii}^2}} \xrightarrow{d} N(0, 1), \text{ as } n \rightarrow \infty.$$

*In addition, if  $|\sum_{k=K_0+1}^K d_k|^2 \gg \sum_{i=1}^n \sigma_{ii}^2 + \sum_{i=1}^n (\sum_{k=K_0+1}^K d_k v_k^2(i))^2$ , then*

$$(16) \quad \mathbb{P}(|\tilde{T}_n| > C) \rightarrow 1,$$

*for arbitrarily large positive constant  $C$ .*

It is seen that with the same critical value  $\Phi^{-1}(1 - \alpha/2)$ ,  $\tilde{T}_n$  enjoys the same asymptotic properties on size and power as  $T_n$ . In addition, since the construction of  $\tilde{T}_n$  does not depend on any tuning parameter, the implementation is much easier.

**3.5. Estimation of  $K$ .** RIRS naturally suggests a simple method for estimating the rank  $K$ . The idea is similar to the one in [20]. That is, we sequentially test the following hypotheses

$$H_0 : K = K_0 \quad \text{vs.} \quad H_1 : K > K_0,$$

for  $K_0 = 1, 2, \dots, K_{\max}$  at a given significance level  $\alpha \in (0, 1)$  using RIRS. Here,  $K_{\max}$  is some prespecified positive integer that should be larger than the true rank. In the application, we may select  $K_{\max} = \lfloor C \log \log n \rfloor$  with some positive constant  $C$ . Once RIRS fails to reject a value of  $K_0$ , we stop and use it as the estimate of the rank. Denote by  $\hat{K}$  our resulting estimate.

**COROLLARY 2.** *Suppose there exists some positive constant  $\delta_1$  such that*

$$(17) \quad \left| \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) \right| \gg \sqrt{m \log^{\delta_1} n} \left( \sqrt{\sum_{i \neq j} \sigma_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k| \right),$$

*holds for all  $K_0 < K$ . Under the conditions of Theorem 3.2, we have*

$$\mathbb{P}(\hat{K} = K) \rightarrow 1 - \alpha.$$

**REMARK 1.** *The additional condition (17) ensures that the estimation procedure rejects all  $K_0 < K$  with asymptotic probability 1 even when  $K$  diverges.*

**REMARK 2.** *Our numerical studies suggest that the RIRS based sequential testing approach may underestimate  $K$  when the network is very sparse or the signal strength is very low; see simulation results in the Supplementary material. To improve the estimation accuracy, we propose to add a penalty on underestimation to  $|T_n|$  as follows*

$$(18) \quad |\check{T}_n| = |T_n| + \frac{|\max_i \sum_j X_{ij}|^{1/2}}{|\hat{d}_{K_0}| - |\hat{d}_{K_0+1}|}.$$

*Then we implement our estimation procedure identically with  $|T_n|$  replaced with  $|\check{T}_n|$ . It is easy to show that Corollary 2 still holds. The intuition is that the penalty term in (18) tends to zero with asymptotic probability one when  $K_0 = K$ , and provides a positive penalty when  $K_0 < K$ . This modified test is useful for increasing the estimation accuracy, especially when  $K$  is large, as demonstrated in Table S.3 in the Supplementary file.*

3.6. *Networks without selfloops: why subsampling?*. In this section, we formalize the heuristic arguments given in Section 2.2 on why sub-sampling is necessary. We theoretically show that  $T_n$  is no longer asymptotically normal under  $H_0$  without the ingredient of sub-sampling. For technical simplicity, we constraint ourselves to the setting of finite  $K$  and no eigenvalue multiplicity in this subsection.

We start with introducing some additional notations that will be used in this subsection. For any matrices  $\mathbf{M}_1$  and  $\mathbf{M}_2$  of appropriate dimensions, let

$$(19) \quad \mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, t) = - \sum_{l=0, l \neq 1}^L \frac{\mathbf{M}_1^T \mathbb{E} \mathbf{W}^l \mathbf{M}_2}{t^{l+1}},$$

where  $L = \lfloor \log n \rfloor$ . By Lemma 6 and Theorem 1 of [11], for each  $k = 1, \dots, K$ , there exists a unique deterministic  $t_k$  such that  $\frac{t_k}{d_k} \rightarrow 1$  as  $n \rightarrow \infty$  and  $\hat{d}_k - t_k = \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\frac{\alpha_n}{|\hat{d}_k|})$ . Define

$$\begin{aligned} \mathbf{b}_{\mathbf{e}_i, k, t}^T &= \mathbf{e}_i^T - \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t) \right)^{-1} \mathbf{V}_{-k}^T, \\ \mathbf{s}_{k, i} &= \mathbf{b}_{\mathbf{e}_i, k, t_k} - \mathbf{e}_i^T \mathbf{v}_k \mathbf{v}_k, \quad \mathbf{s}_k = \sum_{i=1}^n \mathbf{s}_{k, i}, \quad \mathbf{s}_k(i) = \mathbf{e}_i^T \mathbf{s}_k, \\ \text{and } \mathbf{r}_k &= \mathbf{V}_{-k} (t_k \mathbf{D}_{-k}^{-1} - \mathbf{I})^{-1} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k, \end{aligned}$$

where  $\mathbf{V}_{-k}$  is the submatrix of  $\mathbf{V}$  by removing the  $k$ -th column, and we slightly abuse the notation and use  $\mathbf{D}_{-k}$  to denote the submatrix of  $\mathbf{D}$  by removing the  $k$ th diagonal entry.

Further define  $a_k = \sum_{i=1}^n \mathbf{v}_k(i)$ ,  $k = 1, \dots, K$  and

$$(20) \quad \begin{aligned} R(K) &= 2 \sum_{k=1}^K \frac{\mathbf{1}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k a_k}{t_k} - 2 \sum_{k=1}^K \frac{a_k^2 \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} \\ &\quad + \sum_{k=1}^K \mathbf{v}_k^T \text{diag}(\mathbf{W}) \mathbf{v}_k a_k^2 + 2 \sum_{k=1}^K a_k \mathbf{s}_k^T \text{diag}(\mathbf{W}) \mathbf{v}_k + 2 \sum_{k=1}^K a_k \frac{\mathbf{1}^T \mathbf{r}_k}{t_k}. \end{aligned}$$

We have the following theorems.

**THEOREM 3.5.** *Suppose that Conditions 1–4 hold with no eigenvalue multiplicity, and*

$$(21) \quad \sum_{i < j} \sigma_{ij}^2 \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)^2 \geq (\log n)^{6+\epsilon_1} \left( n + \frac{n^2 \alpha_n^2}{d_{K_0}^2} \right),$$

for some positive constant  $\epsilon_1$ . Under null hypothesis, as  $n \rightarrow \infty$ , we have

$$\frac{\sum_{i \neq j} \hat{w}_{ij} + R(K_0)}{2 \sqrt{\sum_{i < j} \sigma_{ij}^2 \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)^2}} \xrightarrow{d} N(0, 1).$$

**THEOREM 3.6.** *Suppose that Conditions 1–4 hold with no eigenvalue multiplicity. In addition, assume (21) holds with  $K_0$  and  $d_{K_0}$  replaced with  $K$  and  $d_K$ , respectively. Under*

alternative hypothesis, as  $n \rightarrow \infty$ , we have

$$2\sqrt{\frac{\sum_{i \neq j} \hat{w}_{ij} + R(K) - \sum_{k=K_0+1}^K d_k a_k^2}{\sum_{i < j} \sigma_{ij}^2 \left(1 - \sum_{k=1}^K a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^K a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j))\right)^2}} \xrightarrow{d} N(0, 1).$$

It is seen from Theorems 3.5 and 3.6 that aggregating all entries of the residual matrix leads to a statistic with bias and variance taking very complicated forms under both null and alternative hypotheses. The complicated forms of bias and variance limit the practical usage of the above results. In addition, and more importantly, these asymptotic normality results may even fail to hold in some cases. To understand this, note that the variance of  $\sum_{i \neq j} \hat{w}_{ij} + R(K_0)$  in Theorem 3.5 is approximately equal to

$$4 \sum_{i < j} \sigma_{ij}^2 \left(1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j))\right)^2.$$

Condition (21) is imposed to put a lower bound on the variance. Without this condition, the asymptotic normality in Theorem 3.5 will no longer hold. We next give an example where inequality (21) and the asymptotic normality both fail to hold. This justifies the necessity of the subsampling step.

**EXAMPLE 2.** Consider networks with eigenvector taking the form  $\mathbf{v}_1 = \frac{1}{\sqrt{n}} \mathbf{1}$ . Then  $a_1 = \sqrt{n}$ . Since  $\mathbf{v}_k$ ,  $k \geq 2$  are orthogonal to  $\mathbf{v}_1$ , we have  $a_k = 0$ ,  $k \geq 2$ . By Condition 4 and Theorem S.1 in the Supplementary file, we have  $\max_i |\mathbf{s}_1(i)| \lesssim \frac{\alpha_n^2}{\sqrt{n} d_1^2}$ . Combining this with Condition 3 and using the fact  $\mathbf{v}_1 = \frac{1}{\sqrt{n}} \mathbf{1}$ , we have

$$\begin{aligned} & \sum_{i < j} \sigma_{ij}^2 \left(1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j))\right)^2 \\ &= \sum_{i < j} \sigma_{ij}^2 (1 - n \mathbf{v}_1(i) \mathbf{v}_1(j) - \sqrt{n} (\mathbf{v}_1(j) \mathbf{s}_1(i) + \mathbf{v}_1(i) \mathbf{s}_1(j)))^2 \\ &= \sum_{i < j} \sigma_{ij}^2 (\mathbf{s}_1(i) + \mathbf{s}_1(j))^2 \\ &\lesssim \frac{\alpha_n^4}{n d_1^4} \sum_{i < j} \sigma_{ij}^2 \leq \frac{\alpha_n^6}{d_1^4} \lesssim \frac{n^2 \alpha_n^2}{d_1^4} \lesssim \left(1 + \frac{n^2 \alpha_n^2}{d_{K_0}^2}\right), \end{aligned}$$

where in the last line we have used  $\sum_{i < j} \sigma_{ij}^2 \leq n \alpha_n^2$  and  $\alpha_n^2 \lesssim n$ . This contradicts (21). Further, by checking the proof of Theorem 3.5, we see that the intrinsic problem is when aggregating too many terms from the residual matrix, the noise accumulation is no longer negligible, canceling the first order term  $\sum_{i \neq j} \sigma_{ij}^2$ , and consequentially makes the central limit theorem fail. Similar phenomenon happens under the alternative hypothesis as well.

**4. Simulation studies.** In this section, we use simulations to justify the performance of RIRS in testing and estimating  $K$ . Section 4.1 considers the network model and Section 4.2 considers more general low rank plus noise matrices. The nominal level is fixed to be  $\alpha = 0.05$  in all settings.

4.1. *Network models.* Consider the DCMM model (11). We simulate two types of nodes: pure node with  $\pi_i$  chosen from the set of unit vectors

$$\text{PN}(K) = \{\mathbf{e}_1, \dots, \mathbf{e}_K\},$$

and the mixed membership node with  $\pi_i$  chosen from

$$\text{MM}(K, x) = \left\{ (x, \underbrace{1-x, 0, \dots, 0}_{K-2}), (1-x, x, \underbrace{0, \dots, 0}_{K-2}), \left( \underbrace{\frac{1}{K}, \dots, \frac{1}{K}}_K \right) \right\}$$

where  $x \in (0, 1)$ . Sections 4.1.1 and 4.1.2 concern the testing performance and Section 4.1.3 focuses on the estimation performance with RIRS.

4.1.1. *SBM.* When all rows of  $\mathbf{\Pi}$  are chosen from the pure node set  $\text{PN}(K)$  and the degree heterogeneity matrix  $\mathbf{\Theta} = \rho \mathbf{I}_n$ , the DCMM (11) reduces to the SBM with the following mean matrix structure

$$(22) \quad \mathbf{H} = \rho \mathbf{\Pi} \mathbf{B} \mathbf{\Pi}^T, \quad \rho \in (0, 1), \quad \pi_i \in \text{PN}(K), \quad i = 1, \dots, n.$$

We generate 200 independent adjacency matrices each with  $n = 1000$  nodes and  $K$  equal-sized communities from the above SBM (22). We set  $\mathbf{B} = (B_{ij})_{K \times K}$  with  $B_{ij} = s^{|i-j|}$ ,  $i \neq j$  and  $B_{ii} = (K + 1 - i)/K$ . The value of  $\rho$  ranges from 0.04 to 0.9, with smaller  $\rho$  corresponding to sparser network model. For all values of  $K$ , we choose  $m = \sqrt{n}$  in calculating the RIRS test statistics  $T_n$  and  $\tilde{T}_n$  for networks without and with selfloops, respectively.

The performance of RIRS is compared with the methods in [20], where two versions of test – one with and one without bootstrap correction – were proposed when the network is absent of selfloops. The empirical sizes and powers of both methods when  $s = 0.1$  are reported in Tables 1 and 2 for  $K = 2$  and 3, respectively. The corresponding computation times are reported in Table 3. We also test the performance of both methods with different values of  $s$ , and the corresponding results are summarized in Table S.1 in the Supplementary file.

From Tables 1 and 2, we observe that when  $K = 2$ , the performance of RIRS is relatively robust to the sparsity level  $\rho$ , with size close to the nominal level and power close to 1 in almost all settings. On contrary, the method in [20] without bootstrap correction has much worse performance. It suffers from size distortion for smaller  $\rho$  (sparser setting). This phenomenon becomes even more severe when  $K = 3$ , where the sizes are close or even equal to one at all sparsity levels. With such distorted size, it is no longer meaningful to compare the power. Therefore we omit its power in Table 2. With bootstrap correction, the method in [20] performs much better and is comparable to RIRS when  $K = 2$  except for the setting of  $\rho = 0.04$ , where the size is severely distorted. When  $K = 3$ , both methods suffer from some size distortion when  $\rho \leq 0.1$ , where the problem is more severe for the method in [20]. Comparing Table 1 with Table 2, we see that the increased number of communities  $K$  makes the performance of both methods worse. This is reasonable because the network size is fixed at  $n = 1000$ . So larger  $K$  results in smaller size of each community. From Table 3 we see that the computational cost of the bootstrap method in [20] is much higher than that of RIRS. We also experimented with larger value of  $K = 5$  and the results are summarized in Table S.2 of the Supplementary file to save space.

Finally, we present in Figure 1 the histogram plots as well as the fitted density curves of our test statistics from 1000 repetitions when  $K = 2$ ,  $s = 0.1$ , and  $\rho = 0.7$  under the null hypothesis. The standard normal density curves are also plotted as reference. It visually confirms that the asymptotic null distribution is standard normal.

TABLE 1  
Empirical size and power under SBM with  $K = 2$  and  $s = 0.1$ .

$\rho$	No selfloop						Selfloops	
	RIRS ( $T_n$ )		Lei (no bootstrap)		Lei (bootstrap)		RIRS ( $\tilde{T}_n$ )	
	size	power ( $K_0=1$ )	size	power ( $K_0=1$ )	size	power ( $K_0=1$ )	size	power ( $K_0=1$ )
0.04	0.055	0.86	1	1	0.575	1	0.140	0.310
0.07	0.060	0.99	1	1	0.055	1	0.060	0.605
0.09	0.055	1	0.995	1	0.040	1	0.085	0.750
0.1	0.025	1	0.995	1	0.035	1	0.085	0.815
0.3	0.025	1	0.24	1	0.02	1	0.06	1
0.5	0.045	1	0.07	1	0.025	1	0.065	1
0.7	0.065	1	0.1	1	0.055	1	0.05	1
0.9	0.04	1	0.045	1	0.065	1	0.075	1

TABLE 2  
Empirical size and power under SBM with  $K = 3$  and  $s = 0.1$ .

$\rho$	No selfloop							Selfloops		
	RIRS ( $T_n$ )			Lei (no bootstrap)	Lei (bootstrap)			RIRS( $\tilde{T}_n$ )		
	size	power ( $K_0=1$ )	power ( $K_0=2$ )	size	size	power ( $K_0=1$ )	power ( $K_0=2$ )	size	power ( $K_0=1$ )	power ( $K_0=2$ )
0.04	0.225	0.985	0.235	1	1	1	1	0.35	0.71	0.075
0.07	0.27	1	0.26	1	1	1	1	0.365	0.925	0.09
0.09	0.225	1	0.335	1	0.985	1	1	0.21	1	0.12
0.1	0.065	1	0.36	1	0.895	1	1	0.1	0.98	0.19
0.3	0.075	1	0.795	1	0.06	1	1	0.065	1	0.625
0.5	0.045	1	0.98	0.99	0.02	1	1	0.075	1	0.94
0.7	0.045	1	0.985	0.925	0.04	1	1	0.065	1	1
0.9	0.05	1	1	0.69	0.015	1	1	0.05	1	1

TABLE 3  
Average computation time (in seconds) for test statistics in Table 1 and Table 5 in one replication under SBM with no selfloop,  $K = 2$  and  $\rho = 0.5$ .

	RIRS ( $T_n$ )		Lei (no bootstrap)		Lei (bootstrap)	
	Size	Estimation	Size	Estimation	Size	Estimation
Time	0.504	0.906	0.432	2.88	14.410	147.142

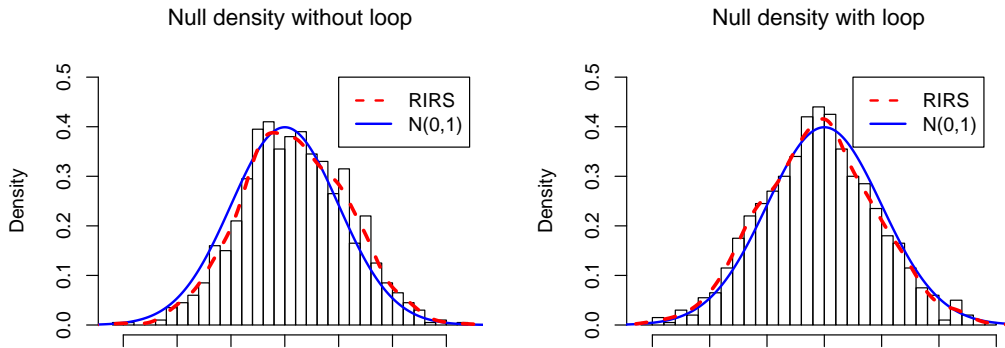


FIG 1. Histogram plots and the estimated densities (red curves) of RIRS test statistic when  $K = 2$  and  $\rho = 0.7$ . Left:  $T_n$  when no selfloop; Right:  $\tilde{T}_n$  when selfloops exist.

4.1.2. *DCMM*. Next consider the general DCMM model (11). The number of repetition is still 200. We simulate the node degree parameters  $\vartheta_j$ 's independently from the uniform distribution over  $[0.5, 1]$ . The vectors  $\pi_i$  are chosen from  $\text{PN}(K) \cup \text{MM}(K, 0.2)$ , with  $n_0$  pure nodes from each community and  $(n - Kn_0)/3$  nodes from each mixed membership probability mass vector in  $\text{MM}(K, 0.2)$ . We select  $n_0 = 0.35n$  when  $K = 2$  and  $n_0 = 0.25n$  when  $K = 3$ . The matrix  $\mathbf{B}$  is chosen to be the same as in the SBM with  $s = 0.1$ . The network size  $n$  ranges from 800 to 2000. The empirical sizes and powers are summarized in Table 4.

Since [20] only considers SBM, the tests therein are no longer applicable in this setting. RIRS performs well and similarly to the SBM setting. Figure 2 presents the histogram plots as well as the fitted density curves of RIRS under the null hypothesis from 1000 repetitions when  $K = 3$  and  $n = 1500$ . These results well justify our theoretical findings.

TABLE 4  
Empirical size and power of RIRS under DCMM model.

$n$	$K = 2$				$K = 3$					
	No Selfloop ( $T_n$ )		Selfloop ( $\tilde{T}_n$ )		No Selfloop ( $T_n$ )			Selfloop ( $\tilde{T}_n$ )		
	Size	Power ( $K_0=1$ )	Size	Power ( $K_0=1$ )	Size	Power ( $K_0=1$ )	Power ( $K_0=2$ )	Size	Power ( $K_0=1$ )	Power ( $K_0=2$ )
800	0.045	1	0.08	1	0.05	1	0.58	0.08	1	0.845
1000	0.04	1	0.05	1	0.025	1	0.68	0.06	1	0.92
1200	0.065	1	0.05	1	0.045	1	0.77	0.07	1	0.92
1500	0.045	1	0.03	1	0.075	1	0.9	0.055	1	0.98
1800	0.075	1	0.055	1	0.045	1	0.98	0.065	1	0.995
2000	0.075	1	0.065	1	0.05	1	0.965	0.045	1	1

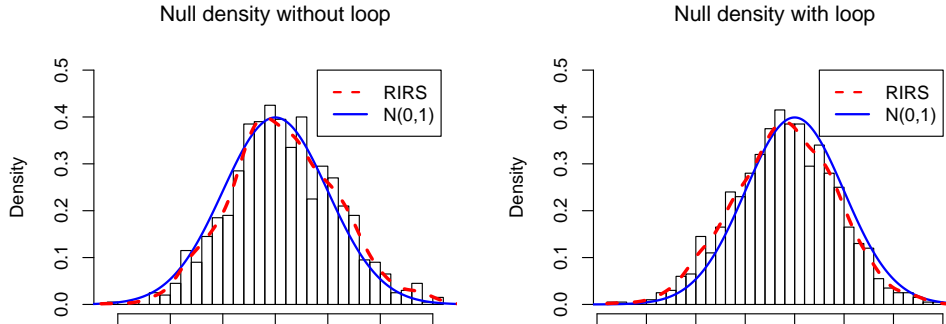


FIG 2. *DCMM*. Histogram plots and the estimated densities (red curves) of RIRS when  $K = 3$  and  $n = 1500$ . Left:  $T_n$  when no selfloop; Right:  $\tilde{T}_n$  when selfloops exist.

4.1.3. *Estimating the Number of Communities*. We use the method discussed in Section 3.5 to estimate the number of communities  $K$ . Since the approaches in [20] are not applicable to the DCMM model, we only compare the performance of RIRS with [20] in SBM setting in the absence of selfloops. The proportions of correctly estimated  $K$  are calculated over 200 replications and tabulated in Table 5 for SBM and in Table 6 for DCMM model.

Table 5 shows that RIRS is generally comparable to or is only slightly worse than Lei's method under the SBM when  $\rho$  is large, and is significantly better than the latter for very small  $\rho$ . While for DCMM model (Table 6), RIRS can also estimate the number of communities with high accuracy. In particular, the estimation accuracy gets closer and closer to the expected value of 95% as  $n$  increases, which is consistent with our theory.

TABLE 5  
Proportion of correctly estimated  $K$  under SBM.

$\rho$	$K = 2$				$K = 3$			
	No Selfloop			Selfloop	No Selfloop			Selfloop
	RIRS $T_n$	Lei (no bootstrap)	Lei (bootstrap)	RIRS $\tilde{T}_n$	RIRS $T_n$	Lei (no bootstrap)	Lei (bootstrap)	RIRS $\tilde{T}_n$
0.04	0.85	0	0.32	0.345	0.12	0	0	0.015
0.07	0.945	0	0.98	0.59	0.2	0	0	0.04
0.09	0.945	0	0.96	0.775	0.23	0	0.02	0.125
0.1	0.93	0	0.97	0.815	0.285	0	0.165	0.105
0.3	0.94	0.795	0.955	0.96	0.745	0	0.935	0.645
0.5	0.945	0.925	0.925	0.95	0.9	0.005	0.98	0.895
0.7	0.97	0.915	0.945	0.96	0.955	0.065	0.995	0.955
0.9	0.94	0.94	0.935	0.955	0.93	0.275	0.975	0.945

TABLE 6  
Proportion of correctly estimated  $K$  under DCMM.

$n$	$K = 2$						$K = 3$					
	800	1000	1200	1500	1800	2000	800	1000	1200	1500	1800	2000
	No Selfloop ( $T_n$ )											
RIRS	0.935	0.935	0.93	0.965	0.935	0.95	0.505	0.625	0.805	0.865	0.915	0.935
	Selfloop ( $\tilde{T}_n$ )											
RIRS	0.935	0.94	0.955	0.955	0.955	0.945	0.79	0.85	0.93	0.895	0.935	0.965

4.2. *Low rank data matrix.* RIRS can be applied to other low rank data matrices beyond the network model. In this section, we generate  $n \times n$  data matrix  $\mathbf{X}$  from the following model

$$\mathbf{X} = \mathbf{H} + \mathbf{W} = \mathbf{V}\mathbf{D}\mathbf{V}^T + \mathbf{W},$$

where the residual matrix  $\mathbf{W}$  is symmetric with upper triangle entries (including the diagonal ones) i.i.d from uniform distribution over  $(-1,1)$ . Let  $\mathbf{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{pmatrix}$ , where  $\mathbf{V}_1$  and  $\mathbf{V}_2$  are  $n_1 \times K$  and  $(n - n_1) \times K$  matrices respectively. We randomly generate an  $n_1 \times n_1$  Wigner matrix and collect its  $K$  eigenvectors corresponding to the largest  $K$  eigenvalues to form  $\mathbf{V}_1$ . We set  $\mathbf{V}_2 = \frac{\sqrt{K}}{\sqrt{n-n_1}} \mathbf{\Pi}$  with  $\mathbf{\Pi} = (\pi_1, \dots, \pi_{n-n_1})^T$ , where  $\pi_i \in \text{PN}(K)$  and the number of rows taking each distinct value from  $\text{PN}(K)$  is the same. The diagonal matrix  $\mathbf{D} = n \times \text{diag}(K, K-1, \dots, 1)$ . We set  $n_1 = n/2$  and range the value of  $n$  from 100 to 500. When  $K = 2$ , the empirical sizes and powers as well as the proportions of correctly estimated  $K$  over 500 repetitions are recorded in Table 7. It is seen that both  $T_n$  and  $\tilde{T}_n$  performs well, with  $\tilde{T}_n$  having slightly higher power. This higher power further translates into better estimation accuracy (closer to 95%) of estimated  $K$ .

TABLE 7  
Empirical size and power, and the proportion (Prop) of correctly estimated  $K$  over 500 replications.

$n$	No Selfloop ( $T_n$ )					Selfloop ( $\tilde{T}_n$ )				
	100	200	300	400	500	100	200	300	400	500
Size	0.048	0.042	0.05	0.054	0.052	0.05	0.05	0.032	0.062	0.052
Power	0.612	0.914	0.994	1	1	1	1	1	1	1
Prop	0.588	0.856	0.944	0.95	0.954	0.95	0.95	0.968	0.938	0.948

**5. Real data analysis.** We consider a popularly studied network of political blogs assembled by [2]. The nodes are blogs over the period of two months before the 2004 U.S. Presidential Election. The edges are the web links between the blogs. These blogs have known political divisions and were labeled into two communities ( $K = 2$ ) by [2] – the liberal and conservative communities. This blog data has been frequently used in the literature, see [17], [29] and [20] among others. It is widely believed to follow a degree corrected block model. Following the literature, we ignore the directions and study only the largest connected component, which has  $n = 1222$  nodes. Consider the following two hypothesis tests:

$$\text{(HT1)}: H_0: K = 1 \text{ vs. } H_1: K > 1.$$

$$\text{(HT2)}: H_0: K = 2 \text{ vs. } H_1: K > 2.$$

[20] considered **(HT2)** and obtained test statistic values 1172.3 and 491.5, corresponding to the test without bootstrap and with bootstrap, respectively. Both are much larger than the critical value (about 1.454) from the Tracy-Widom distribution, and thus the null hypothesis in **(HT2)** was strongly rejected. This is not surprising because the testing procedure in [20] is based on the SBM. It is possible that the model is misspecified when applying the tests therein.

RIRS does not depend on any specific network model structure and is expected to be more robust to model misspecification. Since most of the diagonal entries of  $\mathbf{X}$  are zero, we use the test statistic  $T_n$ . Noticing that the observed data matrix  $\mathbf{X}$  is non-symmetric, we consider two simple transformations:

$$(23) \quad \text{Method 1: } \tilde{\mathbf{X}}_1 = \mathbf{X} + \mathbf{X}^T; \quad \text{Method 2: } \tilde{\mathbf{X}}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{pmatrix}_{2n \times 2n}.$$

The transformation in Method 2 is general and can be applied to even non-square data matrix  $\mathbf{X}$ . After the transformations,  $\text{rank}(\mathbb{E}(\tilde{\mathbf{X}}_1)) = K$  and  $\text{rank}(\mathbb{E}(\tilde{\mathbf{X}}_2)) = 2K$ . The results of applying  $T_n$  to the two hypothesis test problems **(HT1)** and **(HT2)**, together with the estimated number of communities by the sequential testing procedure are reported in Table 8. We can see that for both transformations, RIRS consistently estimated the number of communities to be 2, which is consistent with the common belief in the literature.

TABLE 8  
*Hypothesis testing and estimation results for the political blog data.*

	Method 1		Method 2		Decision
	Test Statistic	P-value	Test Statistic	P-value	
<b>(HT1)</b>	3.3527	0.0008	2.7131	0.0067	Reject $H_0$ in <b>(HT1)</b>
<b>(HT2)</b>	-1.2424	0.2141	-0.8936	0.3716	Accept $H_0$ in <b>(HT2)</b>
Estimate	2		2		$K = 2$

## APPENDIX A: PROOF OF THE MAIN RESULTS

We introduce a definition that will be used frequently in the proof.

**DEFINITION 1.** Let  $\zeta_n(i)$  and  $\xi_n(i)$  be random (or deterministic) variables depend on  $i$ ,  $i = 1, \dots, n$ . We say  $\xi_n = O_{p_n}(\zeta_n)$  if for any positive constants  $D$  and  $\epsilon$ , there exists some positive integer  $n_0(D, \epsilon)$  depending only on  $D$  and  $\epsilon$  such that for all  $n \geq n_0(D, \epsilon)$  we have

$$\mathbb{P}[\exists 1 \leq i \leq n, \text{ s.t. } |\xi_n(i)| > (\log n)^\epsilon |\zeta_n(i)|] \leq (\log \log n)^{-D}.$$



In addition, to facilitate our proof presentation, we introduce some additional notations. Let  $\tilde{K}$  be the number of distinct nonzero eigenvalues of  $\mathbf{H} = (h_{ij})$ . Denote by  $\tilde{d}_1, \dots, \tilde{d}_{\tilde{K}}$  the distinct values in  $\{d_1, \dots, d_K\}$ , sorting in decreasing magnitudes. In addition, denote by  $K_j$  the multiplicity of  $\tilde{d}_j$ . That is, for each  $j = 1, \dots, \tilde{K}$ , the cardinality of  $\{1 \leq l \leq K : d_l = \tilde{d}_j\}$  is  $K_j$  with  $\sum_{j=1}^{\tilde{K}} K_j = K$ .

**A.1. Outline of The Proof.** Our Condition 2 and main results Theorems 3.1–3.4 allow for multiplicity in eigenvalues. When proving these main results, we consider the case with and without multiplicity separately. The proof of our main results highly depends on the asymptotic expansions of the eigenvectors  $\hat{\mathbf{v}}_k$  and eigenvalues  $\hat{d}_k$ . Briefly speaking, Lemma 1 in Section 7.1 of the Supplementary material establishes the relationship between  $\hat{w}_{ij}$  and  $(\hat{v}_k(i), \hat{d}_k)$ . This together with the asymptotic expansions of eigenvalues and eigenvectors gives us the asymptotic expansion of  $\hat{w}_{ij}$ . Substituting this asymptotic expansion into the proposed test statistics, we are able to prove our main theorems by additional analysis and calculations. In the case without eigenvalue multiplicity, the asymptotic expansions of eigenvalues and eigenvectors are established in Lemmas 2 and 3 in the Supplementary material. The corresponding results in the existence of multiplicity are established in Lemmas 4 and 5 in the Supplementary file. In the main paper we only provide the proof of Theorem 3.1. All other proofs are relegated to the Supplement.

**A.2. Proof of Theorem 3.1.** The result in Theorem 3.1 can be obtained by combing the following two results.

$$(24) \quad \text{CLT :} \quad \frac{\sqrt{m} \sum_{i \neq j} \hat{w}_{ij} Y_{ij}}{\sqrt{2 \sum_{i \neq j} \mathbb{E} w_{ij}^2}} \xrightarrow{d} N(0, 1),$$

$$(25) \quad \text{Consistency:} \quad \frac{\sum_{i \neq j} \hat{w}_{ij}^2}{\sum_{i \neq j} \mathbb{E} w_{ij}^2} = 1 + o_p(1).$$

We next proceed with proving (24) and (25).

We first verify the central limit theorem (24). By (S.44) and Corollary S.1 we have

$$(26) \quad \begin{aligned} \sum_{i \neq j} \hat{w}_{ij} Y_{ij} &= \sum_{i \neq j} w_{ij} Y_{ij} - \sum_{k=1}^{K_0} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \\ &\quad - \sum_{k=1}^{K_0} \sum_{i \neq j} Y_{ij} \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &\quad - \sum_{k=1}^{K_0} \sum_{i \neq j} Y_{ij} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} + 2 \sum_{k=1}^{K_0} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}) \\ &\quad - 2 \sum_{k=1}^{K_0} \frac{d_k}{t_k} \sum_{i \neq j} Y_{ij} \mathbf{v}_k(i) \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k - \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k Y_{ij}}{t_k^2} \\ &\quad + \sum_{i \neq j} (Y_{ij} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right)). \end{aligned}$$

Recall that  $\mathbb{E} Y_{ij} = \frac{1}{m}$ ,  $a_k = \sum_{i=1}^n \mathbf{v}_k(i)$  and  $|a_k| \leq \sqrt{n}$ . Our aim is to bound all terms on the right hand side of equation (26) except for the first term  $\sum_{i \neq j} w_{ij} Y_{ij}$ . We begin with splitting

the term

$$\sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k$$

into two parts:

$$\frac{1}{m} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \quad \text{and} \quad \sum_{i \neq j} (Y_{ij} - \frac{1}{m}) (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k.$$

For the first part, first note that since  $|w_{ij}|$  is bounded, we have  $|\mathbf{v}_k^T \mathbb{E} \mathbf{W} \mathbf{v}_k| \lesssim 1$ . Then by Corollary S.1 in the Supplementary material and Condition 1 we have

$$\begin{aligned} (27) \quad & \frac{1}{m} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = \frac{1}{m} (a_k^2 - 1) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \\ & = \frac{1}{m} (a_k^2 - 1) (\mathbf{v}_k^T (\mathbf{W} - \mathbb{E} \mathbf{W}) \mathbf{v}_k + \mathbf{v}_k^T \mathbb{E} \mathbf{W} \mathbf{v}_k) = (a_k^2 + 1) (O_{p_u}(\frac{\alpha_n}{m\sqrt{n}}) + O(\frac{1}{m})). \end{aligned}$$

For the second part, first note that  $\mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 = \text{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) + \mathbb{E}^2(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) \lesssim \alpha_n^2/n + 1$ . Since  $Y_{ij}$ ,  $i \neq j$  are i.i.d with  $\mathbb{E} Y_{ij} = \frac{1}{m}$ , Corollary S.1 and Condition 1 ensure that

$$\begin{aligned} & \text{var}(\sum_{i \neq j} (Y_{ij} - \frac{1}{m}) (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) \\ & = \mathbb{E} \left[ \text{var} \left( \sum_{i \neq j} (Y_{ij} - \frac{1}{m}) (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \mid \mathbf{W} \right) \right] \\ & \lesssim \frac{1}{m} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j))^2 \mathbb{E} (\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 \lesssim \frac{\alpha_n^2}{mn} + \frac{1}{m}, \end{aligned}$$

then

$$(28) \quad \sum_{i \neq j} (Y_{ij} - \frac{1}{m}) (\mathbf{v}_k(i) \mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = O_p(\frac{\alpha_n}{\sqrt{mn}} + \frac{1}{\sqrt{m}}).$$

Therefore,

$$(29) \quad \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = (a_k^2 + 1) O_{p_u}(\frac{\alpha_n}{m\sqrt{n}}) + O(\frac{a_k^2 + 1}{m}) + O_p(\frac{\alpha_n}{\sqrt{mn}}) + O_p(\frac{1}{\sqrt{m}}).$$

Similar to (29), we get

$$\frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}) = O_{p_u}(\frac{a_k^2}{m} + \frac{1}{\sqrt{m}}),$$

and

$$(30) \quad \sum_{i \neq j} Y_{ij} \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_{p_u}(\frac{|a_k| \sqrt{n}}{m} + \frac{1}{\sqrt{m}}).$$

Next we split the term  $\sum_{i \neq j} Y_{ij} \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}$  into the following two parts

$$\frac{1}{m} \sum_{i \neq j} \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \quad \text{and} \quad \sum_{i \neq j} (Y_{ij} - \frac{1}{m}) \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}.$$

By (S.85), we have

$$\begin{aligned} & \|\mathcal{R}(\mathbf{1}, \mathbf{V}_{-k}, t_k) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T\| \\ &= \left\| \sum_i \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T \right\| = O_{p_u}(\sqrt{n}). \end{aligned}$$

In light of (S.46), (S.85), Corollary S.1 in the Supplementary material and Condition 4, the following three results hold:

$$\begin{aligned} & \frac{d_k(\mathbf{s}_k - \mathbf{1})^T \mathbf{W} \mathbf{v}_k a_k}{t_k} \\ &= - \frac{d_k a_k \mathcal{R}(\mathbf{1}, \mathbf{V}_{-k}, t_k) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k}{t_k} \\ &\quad - \frac{d_k a_k^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k}{t_k} = O_{p_u}(\alpha_n + |a_k| \sqrt{n}) + O_{p_u}(|a_k| \alpha_n), \\ & \frac{1}{m} \sum_{i \neq j} \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = \frac{1}{m} a_k \frac{d_k \mathbf{1}^T (\mathbf{W} - \mathbb{E} \mathbf{W}) \mathbf{v}_k}{t_k} + O_{p_u} \left( \frac{(|a_k| + 1) \alpha_n}{m} \right) \\ &= O_{p_u} \left( \frac{(|a_k| + 1) \alpha_n}{m} \right) \end{aligned}$$

$$\text{and} \quad \sum_{i \neq j} \left( Y_{ij} - \frac{1}{m} \right) \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_p \left( \frac{\alpha_n}{\sqrt{m}} \right),$$

where the calculation of the variance of the second part  $\sum_{i \neq j} \left( Y_{ij} - \frac{1}{m} \right) \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}$  is similar to that of (28). Therefore,

$$(31) \quad \sum_{i \neq j} Y_{ij} \frac{d_k \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_{p_u} \left( \frac{(|a_k| + 1) \alpha_n + |a_k| \sqrt{n}}{m} \right) + O_p \left( \frac{\alpha_n}{\sqrt{m}} \right).$$

For the term  $\sum_{i \neq j} Y_{ij} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k}$ , we write

$$\begin{aligned} & \sum_{i \neq j} Y_{ij} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} = \frac{1}{m} \sum_{i \neq j} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} \\ (32) \quad & \quad + \sum_{i \neq j} \left( Y_{ij} - \frac{1}{m} \right) \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k}. \end{aligned}$$

It follows from Conditions 2–4 and 6,  $\frac{t_k}{d_k} \rightarrow 1$  in Section 3.6 and Corollary S.1 that

$$|\mathbf{1}^T \mathbf{r}_k| = |\mathbf{1}^T \mathbf{V}_{-k} (t_k \mathbf{D}_{-k}^{-1} - \mathbf{I})^{-1} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k| = O_{p_u}(\sqrt{n} \alpha_n^2),$$

$$|\mathbf{r}_k(i)| = |\mathbf{e}_i^T \mathbf{V}_{-k} (t_k \mathbf{D}_{-k}^{-1} - \mathbf{I})^{-1} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k| = O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n}} \right),$$

and thus

$$\frac{1}{m} \sum_{i \neq j} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} = \frac{2a_k \mathbf{1}^T \mathbf{r}_k}{t_k m} - \frac{2}{m} \sum_{i=1}^n \frac{\mathbf{r}_k(i) \mathbf{v}_k(i)}{t_k} = O_{p_u} \left( \frac{|a_k| \sqrt{n} + 1}{m} \right).$$

As for (28), calculating the variance of the second term on the right hand side of (32) yields

$$\sum_{i \neq j} (Y_{ij} - \frac{1}{m}) \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} = O_{p_u} \left( \frac{1}{\sqrt{m}} \right).$$

Therefore,

$$(33) \quad \sum_{i \neq j} Y_{ij} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} = O_{p_u} \left( \frac{|a_k| \sqrt{n}}{m} \right) + O_{p_u} \left( \frac{1}{\sqrt{m}} \right).$$

Now for the term  $\sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k Y_{ij}}{t_k^2}$ , similarly we write

$$\sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k Y_{ij}}{t_k^2} = \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k}{m t_k^2} + \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k (Y_{ij} - \frac{1}{m})}{t_k^2}.$$

It follows from Corollary S.1 and Theorem S.1 that the first part has order

$$\begin{aligned} \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k}{m t_k^2} &= \frac{d_k \mathbf{1}^T \mathbf{W} \mathbf{v}_k \mathbf{1}^T \mathbf{W} \mathbf{v}_k}{m t_k^2} - \sum_{i=1}^n \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k)^2}{m t_k^2} \\ &= O_{p_u} \left( \frac{\alpha_n^2 (\log n)^2}{m |d_k|} \right) = O_{p_u} \left( \frac{(\log n)^2}{m} \right). \end{aligned}$$

Moreover, calculating the variance, we have

$$\begin{aligned} &\text{var} \left( \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k (Y_{ij} - \frac{1}{m})}{t_k^2} \right) \\ &= \mathbb{E} \left( \left[ \text{var} \left( \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k (Y_{ij} - \frac{1}{m})}{t_k^2} \right) \middle| \mathbf{W} \right] \right) \\ &\lesssim \frac{\sum_{i \neq j} \mathbb{E} (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k)^2}{m d_k^2} \leq \frac{\sum_{i \neq j} \sqrt{\mathbb{E} (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k)^4 \mathbb{E} (\mathbf{e}_j^T \mathbf{W} \mathbf{v}_k)^4}}{m d_k^2} \\ &\lesssim \frac{\sum_{i \neq j} \sqrt{[\mathbb{E} (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k - \mathbb{E} \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k)^4 + (\mathbb{E} \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k)^4] [\mathbb{E} (\mathbf{e}_j^T \mathbf{W} \mathbf{v}_k - \mathbb{E} \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k)^4 + (\mathbb{E} \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k)^4]}}{m d_k^2} \\ &\lesssim \frac{\alpha_n^4}{m d_k^2} = O_{p_u} \left( \frac{1}{m} \right). \end{aligned}$$

Therefore

$$(34) \quad \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k Y_{ij}}{t_k^2} = O_{p_u} \left( \frac{(\log n)^2}{m} \right) + O_{p_u} \left( \frac{1}{\sqrt{m}} \right).$$

Finally, consider the residual term

$$\sum_{i \neq j} (Y_{ij} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right))$$

$$= \sum_{i \neq j} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right) (Y_{ij} - \frac{1}{m}) + \frac{1}{m} \sum_{i \neq j} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right).$$

Note that  $Y_{ij}$  is independent of  $O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right)$ . Calculating the variance of  $\sum_{i \neq j} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right) (Y_{ij} - \frac{1}{m})$  gives us

$$\sum_{i \neq j} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right) (Y_{ij} - \frac{1}{m}) = O_p \left( \frac{1}{\sqrt{m}} \right) \times O_{p_u} \left( \frac{\alpha_n (\log n)^3}{|d_{K_0}|} + \frac{1}{n^{1/2}} \right).$$

The ‘‘mean’’ of the residual term should be

$$\frac{1}{m} \sum_{i \neq j} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right) = O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{m |d_{K_0}|} + \frac{\sqrt{n}}{m} \right).$$

Therefore we have

$$(35) \quad \sum_{i \neq j} (Y_{ij} O_{p_u} \left( \frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}} \right)) \\ = O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{m |d_{K_0}|} + \frac{\sqrt{n}}{m} \right) + O_p \left( \frac{1}{\sqrt{m}} \right) \times O_{p_u} \left( \frac{\alpha_n (\log n)^3}{|d_{K_0}|} + \frac{1}{\sqrt{n}} \right).$$

So far we have found the orders of all other terms on the right hand side of equation (26) except for  $\sum_{i \neq j} w_{ij} Y_{ij}$ . Note that

$$(36) \quad \text{var} \left( \sum_{i \neq j} w_{ij} Y_{ij} \right) = \frac{2}{m} \sum_{i \neq j} \mathbb{E} w_{ij}^2.$$

According to the orders (29), (30), (31), (33), (34) and (35), we can conclude that as long as

$$\frac{2}{m} \sum_{i \neq j} \mathbb{E} w_{ij}^2 \geq (\log n)^{\epsilon_1} \left( \frac{\sum_{k=1}^{K_0} a_k^2 n}{m^2} + \frac{(\alpha_n \log n)^2}{m} + \frac{n}{m^2} + \frac{n^2 \alpha_n^2 (\log n)^6}{m^2 d_{K_0}^2} \right),$$

the term  $\sum_{i \neq j} w_{ij} Y_{ij}$  dominates all other terms on the right hand side of (26). Moreover, by the condition  $\sum_{i \neq j} \mathbb{E} w_{ij}^2 \gg m$ , the fact  $\mathbb{E} Y_{ij}^4 \lesssim 1/m$  and the independence between  $Y_{ij}$  and  $w_{ij}$  we have

$$\frac{m^2}{(\sum_{i \neq j} \mathbb{E} w_{ij}^2)^2} \sum_{i \neq j} \mathbb{E} w_{ij}^4 Y_{ij}^4 \lesssim \frac{m \sum_{i \neq j} \mathbb{E} w_{ij}^4}{(\sum_{i \neq j} \mathbb{E} w_{ij}^2)^2} \lesssim \frac{m}{(\sum_{i \neq j} \mathbb{E} w_{ij}^2)} \rightarrow 0.$$

Therefore, the central limit theorem (24) holds by Lyapunov CLT.

We now show the consistency of  $\sum_{i \neq j} \hat{w}_{ij}^2$  in (25). By (S.31), we have

$$\hat{w}_{ij}^2 = w_{ij}^2 + 2w_{ij} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) - 2 \sum_{k=1}^{K_0} w_{ij} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \\ + O_{p_u} \left( \frac{(\alpha_n \log n)^4}{n^2 |d_{K_0}|^2} + \frac{1}{n^2} \right) + \left( \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right)^2 \\ - 2 O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k}$$

and

$$\begin{aligned}
(37) \quad \sum_{i \neq j} \hat{w}_{ij}^2 &= \sum_{i \neq j} w_{ij}^2 + 2 \sum_{i \neq j} \left( w_{ij} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \right. \\
&\quad \left. - 2 \sum_{k=1}^{K_0} \sum_{i \neq j} \left( w_{ij} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right) \right. \\
&\quad \left. + O_{p_u} \left( \frac{\alpha_n^2 (\log n)^4}{|d_{K_0}|} + 1 \right) + \sum_{i \neq j} \left( \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right)^2 \right. \\
&\quad \left. - 2 \sum_{i \neq j} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right).
\end{aligned}$$

Combing the fact  $\text{var}(\sum_{i \neq j} w_{ij}^2) \leq \sum_{i \neq j} \mathbb{E} w_{ij}^4 \lesssim \sum_{i \neq j} \mathbb{E} w_{ij}^2$  with Conditions 3 and 5

$$(38) \quad \sum_{i \neq j} \mathbb{E} w_{ij}^2 \gg (\log n)^{2+\epsilon_1},$$

we have

$$(39) \quad \frac{\sum_{i \neq j} w_{ij}^2}{\sum_{i \neq j} \mathbb{E} w_{ij}^2} = 1 + O_{p_u} \left( \frac{1}{(\log n)^{\epsilon_1/2}} \right).$$

Then to prove (25), it suffices to show

$$(40) \quad \frac{\sum_{i \neq j} \hat{w}_{ij}^2}{\sum_{i \neq j} w_{ij}^2} = 1 + O_{p_u} \left( \frac{1}{(\log n)^{\epsilon_1/2}} \right).$$

We now check the other terms on the right hand side of (37) to verify (40). First of all,

$$\begin{aligned}
\left| \sum_{i \neq j} w_{ij} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \right| &\leq \left| O_{p_u} \left( \frac{(\alpha_n \log n)^2}{|d_{K_0}|} + 1 \right) \times \sqrt{\sum_{i \neq j} w_{ij}^2} \right| \\
&= O_{p_u} (\alpha_n (\log n) \sqrt{\sum_{i \neq j} \mathbb{E} w_{ij}^2}).
\end{aligned}$$

Condition 5 further implies that

$$(41) \quad \left| \sum_{i \neq j} w_{ij} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \right| = \left( \sum_{i \neq j} \mathbb{E} w_{ij}^2 \right) \times O_{p_u} \left( \frac{1}{(\log n)^{\epsilon_1/2}} \right).$$

Now consider the term  $\sum_{i \neq j} w_{ij} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k}$ . We will only provide detail for proving  $\sum_{i \neq j} w_{ij} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j)}{t_k}$  because the other part can be proved similarly. Write

$$\sum_{i \neq j} w_{ij} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j)}{t_k} = \sum_{i \neq j} \frac{d_k w_{ij}^2 \mathbf{v}_k^2(j)}{t_k} + \sum_{i \neq j, l \neq j} \frac{d_k w_{ij} w_{il} \mathbf{v}_k(j) \mathbf{v}_k(l)}{t_k}.$$

Direct calculations yield

$$\mathbb{E} \left| \sum_{i \neq j} \frac{d_k w_{ij}^2 \mathbf{v}_k^2(j)}{t_k} \right| \lesssim \frac{1}{n} \sum_{i \neq j} \mathbb{E} w_{ij}^2, \quad \sum_{i \neq j, l \neq j} \mathbb{E} \frac{d_k w_{ij} w_{il} \mathbf{v}_k(j) \mathbf{v}_k(l)}{t_k} = 0, \quad \text{and}$$

$$\text{var}\left(\sum_{i \neq j, l \neq j} \frac{d_k w_{ij} w_{il} \mathbf{v}_k(j) \mathbf{v}_k(l)}{t_k}\right) \lesssim \sum_{i \neq j, l \neq j} \mathbb{E} w_{ij}^2 w_{il}^2 \mathbf{v}_k^2(j) \mathbf{v}_k^2(l) \lesssim \frac{\alpha_n^2}{n^2} \sum_{i \neq j} \mathbb{E} w_{ij}^2.$$

Thus,  $\sum_{i \neq j} w_{ij} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j)}{t_k} = O_{p_u}\left(\frac{1}{\sqrt{n}}\right) \times \sum_{i \neq j} \mathbb{E} w_{ij}^2$ . And consequently,

$$(42) \quad \sum_{i \neq j} w_{ij} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} = O_{p_u}\left(\frac{1}{\sqrt{n}}\right) \times \sum_{i \neq j} \mathbb{E} w_{ij}^2.$$

Next by Theorem S.1 and Condition 5, we have

$$(43) \quad \begin{aligned} & \sum_{i \neq j} O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \\ & = O_{p_u} \left( \frac{(\alpha_n \log n)^3}{|d_{K_0}|} + \alpha_n \log n \right) = O_{p_u}(\alpha_n^2 \log^2 n) = O_{p_u} \left( \frac{1}{(\log n)^{\epsilon_1/2}} \sum_{i \neq j} \mathbb{E} w_{ij}^2 \right). \end{aligned}$$

Finally, similar to (43), it holds by Condition 5 that

$$(44) \quad \begin{aligned} & \sum_{i \neq j} \left( \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right)^2 \\ & \lesssim \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{d_k^2 (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))^2}{t_k^2} \\ & = O_{p_u}(\alpha_n^2 \log^2 n) = O_{p_u} \left( \frac{1}{(\log n)^{\epsilon_1/2}} \sum_{i \neq j} \mathbb{E} w_{ij}^2 \right). \end{aligned}$$

Substituting the arguments (39), (41), (42), (43) and (44) into equation (37), we complete the proof of (40). Thus, (25) is proved and the results in the theorem follow automatically.

**Acknowledgments.** The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

**Funding.** Qing Yang was supported by National Natural Science Foundation of China (No. 12101585). Xiao Han was partly supported by NSF of China (No.12001518). Yingying Fan was supported by NSF grant DMS-2052964.

## REFERENCES

- [1] ABBE, E. (2017). Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research* **18** 177:1–177:86.
- [2] ADAMIC, L. A. and GLANCE, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43. ACM, New York, USA. <https://doi.org/10.1145/1134271.1134277>
- [3] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of machine learning research* **9** 1981–2014.
- [4] BAI, Z. D. and YAO, J. F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **44** 447–474.
- [5] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv preprint arXiv:1705.05305*.
- [6] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics* **49** 370–392.

- [7] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 253-273.
- [8] CHATTERJEE, S. et al. (2015). Matrix estimation by universal singular value thresholding. *Annals of Statistics* **43** 177–214.
- [9] CHEN, K. and LEI, J. (2018). Network Cross-Validation for Determining the Number of Communities in Network Data. *Journal of the American Statistical Association* **113** 241-251. <https://doi.org/10.1080/01621459.2016.1246365>
- [10] DAUDIN, J. J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Statistics and Computing* **18** 173–183.
- [11] FAN, J., FAN, Y., HAN, X. and LV, J. (2020). Asymptotic Theory of Eigenvectors for Random Matrices with Diverging Spikes\*. *Journal of the American Statistical Association* **0** 1-63. <https://doi.org/10.1080/01621459.2020.1840990>
- [12] FAN, J., FAN, Y., HAN, X. and LV, J. (2021). SIMPLE: Statistical Inference on Membership Profiles in Large Networks. *arXiv preprint arXiv:1910.01734*.
- [13] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. *arXiv preprint arXiv:1710.00862*.
- [14] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109–137.
- [15] JIN, J., KE, Z. T. and LUO, S. (2019). Optimal Adaptivity of Signed-Polygon Statistics for Network Testing. *arXiv e-prints arXiv:1904.09532*.
- [16] JIN, J., KE, Z. T., LUO, S. and WANG, M. (2020). Estimating the number of communities by Stepwise Goodness-of-fit.
- [17] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107. <https://doi.org/10.1103/PhysRevE.83.016107>
- [18] LATOUCHE, P., BIRMELE, E. and AMBROISE, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling* **12** 93-115. <https://doi.org/10.1177/1471082X1001200105>
- [19] LE, C. M. and LEVINA, E. (2015). Estimating the number of communities in networks by spectral methods. *arXiv e-prints arXiv:1507.00827*.
- [20] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics* **44** 401–424.
- [21] LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276.
- [22] MCDAID, A. F., MURPHY, T. B., FRIEL, N. and HURLEY, N. J. (2013). Improved Bayesian inference for the stochastic block model with application to large networks. *Computational Statistics & Data Analysis* **60** 12-31.
- [23] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642.
- [24] SALDANA, D., YU, Y. and FENG, Y. (2017). How Many Communities Are There? *Journal of Computational and Graphical Statistics* **26** 171-181.
- [25] WANG, Y. J. and WONG, G. Y. (1987). Stochastic Blockmodels for Directed Graphs. *Journal of the American Statistical Association* **82** 8–19.
- [26] WANG, Y. X. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528. <https://doi.org/10.1214/16-AOS1457>
- [27] ZHANG, L. and AMINI, A. A. (2020). Adjusted chi-square test for degree-corrected block models. *arXiv preprint arXiv:2012.15047*.
- [28] ZHAO, Y., LEVINA, E. and ZHU, J. (2011). Community extraction for social networks. *Proceedings of the National Academy of Sciences* **108** 7321–7326. <https://doi.org/10.1073/pnas.1006642108>
- [29] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. <https://doi.org/10.1214/12-AOS1036>



# SUPPLEMENTARY MATERIAL FOR “UNIVERSAL RANK INFERENCE VIA RESIDUAL SUBSAMPLING WITH APPLICATION TO LARGE NETWORKS”

BY XIAO HAN<sup>1,a</sup>, QING YANG<sup>1\*,b</sup> AND YINGYING FAN<sup>2,\*,c</sup>

<sup>1</sup>International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China, <sup>a</sup>xhan011@ustc.edu.cn; <sup>b</sup>yangq@ustc.edu.cn

<sup>2</sup>Data Sciences and Operations Department, University of Southern California, Los Angeles, CA 90089, <sup>c</sup>fanyingy@marshall.usc.edu

This Supplementary material contains some additional simulation examples, theorems, key lemmas and their proofs as well as additional technical details. Condition 2 in the main text allows for multiplicity of eigenvalues. For the clarity of technical arguments, we will separate the proof under the assumption of no-multiplicity from the one with multiplicity. Specifically, we will first prove Lemmas 1, 2 and 3 under the assumption of no-multiplicity. Based on these results, we can prove Theorems 3.1–3.6 when there is no multiplicity. Then we will prove Lemmas 4-5, which are the counterparts of Lemmas 2-3 when multiplicity exists. We will then discuss in the proof of Theorem S.2 how Lemmas 4-5 can be used to prove our main results when multiplicity exists.

Also, let us recall some notations:  $\tilde{\mathcal{K}}$  is the number of distinct nonzero eigenvalues of  $\mathbf{H} = (h_{ij})$ . We denote by  $\tilde{d}_1, \dots, \tilde{d}_{\tilde{\mathcal{K}}}$  the distinct values in  $\{d_1, \dots, d_K\}$ , sorting in decreasing magnitudes. In addition, denote by  $K_j$  the multiplicity of  $\tilde{d}_j$ . That is, for each  $j = 1, \dots, \tilde{\mathcal{K}}$ , the cardinality of  $\{1 \leq l \leq K : d_l = \tilde{d}_j\}$  is  $K_j$  with  $\sum_{j=1}^{\tilde{\mathcal{K}}} K_j = K$ . Moreover, we say that an event  $\mathcal{E}_n$  holds with high probability if  $\mathbb{P}(\mathcal{E}_n) = 1 - O((\log \log n)^{-l})$  for some positive constant  $l > 3$  and sufficiently large  $n$ .

**6. Additional simulation results.** We consider the same model setting used for Table 1 of the main paper but vary the value of  $s$ . Simulation results over 200 repetitions are summarized in Table S.1 below. Comparing Tables 1 and 5 with Table S.1, we observe that when  $s$  is larger, that is, denser connections between communities, both  $T_n$  and  $\tilde{T}_n$  perform slightly worse, which is reasonable because the separation between nodes within communities and between communities is weaker.

Next we consider a larger value of  $K = 5$ . We use the same SBM as defined in (22) of the main paper, but set  $\mathbf{B} = (B_{ij})_{K \times K}$  with  $B_{ij} = s^{|i-j|}$ ,  $i \neq j$  and  $B_{ii} = (2K + 1 - i)/2K$ . We let  $\rho$  vary from 0.04 to 0.9, and fix  $s = 0.05$  and  $n = 2500$ . The simulation results are reported in Table S.2. It is seen that RIRS with  $T_n$  controls the size at the desired level even when  $\rho$  is small, while RIRS with  $\tilde{T}_n$  has some size distortion when  $\rho$  is small. In addition, the power is low when  $\rho < 0.1$ , and increases significantly as  $\rho$  increases.

It is seen from Table S.2 that RIRS based sequential testing approach tends to underestimate the true rank in finite samples when  $\rho$  is small. We experiment the performance of modified test proposed in Remark 2 of our main paper. Table S.3 summarizes the corresponding simulation results with the constant  $c$  in the penalty term taken as 1/2. The model setting is the same as in the above Table S.2. Compared with Table S.2, it is seen that the additional penalty greatly improves the estimation accuracy for small  $\rho$ .

\*Yingying Fan and Qing Yang serve as co-corresponding authors.

*MSC2020 subject classifications:* Primary 62F03, 62F12; secondary 60B20, 62F35.

*Keywords and phrases:* Rank inference, Robustness, Low-rank models, High dimensionality, Asymptotic expansions, Eigenvectors, Eigenvalues, Large random matrices.

TABLE S.1

Empirical size, power and proportion of correctly estimated  $K$  (Prop) under SBM when  $K = 2$ ,  $s = 0.2$  or  $0.3$ .

$\rho$	$s = 0.2$						$s = 0.3$					
	No selfloop ( $T_n$ )			Selfloop ( $\tilde{T}_n$ )			No selfloop ( $T_n$ )			Selfloop ( $\tilde{T}_n$ )		
	Size	Power ( $K_0=1$ )	Prop	Size	Power ( $K_0=1$ )	Prop	Size	Power ( $K_0=1$ )	Prop	Size	Power ( $K_0=1$ )	Prop
0.04	0.065	0.445	0.425	0.095	0.235	0.265	0.08	0.165	0.16	0.2	0.24	0.19
0.07	0.04	0.565	0.53	0.14	0.515	0.495	0.045	0.125	0.18	0.11	0.35	0.34
0.09	0.03	0.625	0.645	0.105	0.71	0.675	0.035	0.26	0.24	0.08	0.435	0.44
0.1	0.035	0.74	0.675	0.085	0.675	0.68	0.055	0.275	0.265	0.085	0.515	0.48
0.3	0.065	1	0.96	0.065	0.99	0.95	0.04	0.63	0.63	0.055	0.99	0.96
0.5	0.04	1	0.97	0.055	1	0.975	0.03	0.92	0.89	0.035	1	0.955
0.7	0.06	1	0.955	0.04	1	0.97	0.03	0.965	0.93	0.035	1	0.945
0.9	0.05	1	0.96	0.05	1	0.935	0.055	1	0.93	0.065	1	0.94

TABLE S.2

Empirical size, power and proportion of correctly estimated  $K$  (Prop) under SBM when  $K = 5$ ,  $\alpha = 0.05$ .

$n = 2500$	No selfloop, RIRS ( $T_n$ )						Selfloop, RIRS ( $\tilde{T}_n$ )					
$\rho$	size	power ( $K_0=1$ )	power ( $K_0=2$ )	power ( $K_0=3$ )	power ( $K_0=4$ )	Prop	size	power ( $K_0=1$ )	power ( $K_0=2$ )	power ( $K_0=3$ )	power ( $K_0=4$ )	Prop
0.04	0.05	1	1	0.92	0.28	0.19	0.17	1	1	0.8	0.09	0.16
0.07	0.05	1	1	1	0.27	0.34	0.09	1	1	0.97	0.32	0.17
0.09	0.02	1	1	1	0.32	0.38	0.14	1	1	0.98	0.4	0.4
0.1	0.06	1	1	1	0.48	0.42	0.07	1	1	1	0.48	0.5
0.3	0.04	1	1	1	0.86	0.8	0.05	1	1	1	0.98	0.96
0.5	0.06	1	1	1	0.99	0.93	0.03	1	1	1	1	0.92
0.7	0.06	1	1	1	1	0.94	0.06	1	1	1	1	0.98
0.9	0.04	1	1	1	1	0.94	0.04	1	1	1	1	0.98

TABLE S.3

Proportion of correctly estimated  $K$  by using  $|\tilde{T}_n|$  in Remark 2,  $K = 5$ , no selfloop.

$\rho$	0.04	0.07	0.09	0.1	0.3	0.5	0.7	0.9
Remark 2( $\alpha = 0.05$ )	0.255	0.675	0.755	0.755	0.930	0.920	0.945	0.940
Remark 2( $\alpha = 0.01$ )	0.615	0.720	0.615	0.700	0.875	0.970	0.970	0.990

## 7. Proof of other main results.

7.1. *Lemma 1 and its proof.* Recall the definition of the residual matrix  $\widehat{\mathbf{W}}$  in (4). In this section we connect its entries  $\hat{w}_{ij}$  with  $(\hat{v}_k(i), \hat{d}_k)$ , which is important for analyzing the asymptotic properties of  $\hat{w}_{ij}$ .

LEMMA 1. Let  $\Delta(d_k) = \hat{d}_k - d_k$  and  $\Delta(\mathbf{v}_k(i)) = \hat{\mathbf{v}}_k(i) - \mathbf{v}_k(i)$ . For  $K \geq K_0$ , we have

$$(S.1) \quad \hat{w}_{ij} = w_{ij} + \delta(K > K_0) \left[ \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} \left[ \Delta(d_k) \hat{\mathbf{v}}_k(i) \hat{\mathbf{v}}_k(j) \right. \right. \\ \left. \left. + d_k \Delta(\mathbf{v}_k(i)) \mathbf{v}_k(j) + d_k \Delta(\mathbf{v}_k(j)) \mathbf{v}_k(i) + d_k \Delta(\mathbf{v}_k(i)) \Delta(\mathbf{v}_k(j)) \right] \right],$$

where  $\delta(K > K_0) = \begin{cases} 1, & K > K_0, \\ 0, & \text{otherwise.} \end{cases}$

PROOF. By the definition of  $\widehat{\mathbf{W}}$ , we have

$$(S.2) \quad \widehat{\mathbf{W}} = \mathbf{X} - \sum_{k=1}^{K_0} \hat{d}_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T = \mathbf{W} + \sum_{k=1}^K d_k \mathbf{v}_k \mathbf{v}_k^T - \sum_{k=1}^{K_0} \hat{d}_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T.$$

Equation (S.1) follows directly from (S.2) by considering each entry separately.  $\square$

7.2. *Proof of Theorem 3.2.* According to Lemma 1, under the alternative hypothesis ( $K > K_0$ ), we have the following expansion

$$(S.3) \quad \hat{w}_{ij} - \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) = w_{ij} - \sum_{k=1}^{K_0} \left[ \Delta(d_k) \hat{\mathbf{v}}_k(i) \hat{\mathbf{v}}_k(j) + d_k \Delta(\mathbf{v}_k(i)) \mathbf{v}_k(j) \right. \\ \left. + d_k \Delta(\mathbf{v}_k(j)) \mathbf{v}_k(i) + d_k \Delta(\mathbf{v}_k(j)) \Delta(\mathbf{v}_k(i)) \right].$$

The only difference from the expression of  $\hat{w}_{ij}$  under the null hypothesis is the extra term  $\sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j)$  on the left hand side of (S.3). Notice that this term is non random. Define

$$\tilde{w}_{ij} = \hat{w}_{ij} - \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j).$$

Then the central limit theorem for  $\sum_{i \neq j} \tilde{w}_{ij}$  can be obtained by using the same proof as that for Theorem 3.1. Thus,

$$(S.4) \quad \left( T_n - \frac{\sqrt{m} \sum_{i \neq j} \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}}{\sqrt{2 \sum_{i \neq j} \hat{w}_{ij}^2}} \right) \frac{\sqrt{\sum_{i \neq j} \hat{w}_{ij}^2}}{\sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2}} \\ = \frac{\sqrt{m} \sum_{i \neq j} \tilde{w}_{ij} Y_{ij}}{\sqrt{2 \sum_{i \neq j} \tilde{w}_{ij}^2}} \xrightarrow{d} N(0, 1).$$

Note that  $\tilde{w}_{ij}$  is the residual term under the alternative hypothesis. Similar to (25), we also have

$$(S.5) \quad \frac{\sum_{i \neq j} \tilde{w}_{ij}^2}{\sum_{i \neq j} \mathbb{E} w_{ij}^2} = 1 + o_p(1).$$

Moreover, direct calculations show that

$$(S.6) \quad \sum_{i \neq j} \hat{w}_{ij}^2 = \sum_{i \neq j} \tilde{w}_{ij}^2 + 2 \sum_{i \neq j} \tilde{w}_{ij} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right) + \sum_{i \neq j} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2.$$

We first prove (8) in Theorem 3.2. It follows from Cauchy-Schwarz inequality, (S.5) and the condition  $\sum_{i \neq j} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2 \ll \sum_{i \neq j} \mathbb{E} w_{ij}^2$  that

$$(S.7) \quad \left| \sum_{i \neq j} \tilde{w}_{ij} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right) \right| \leq \sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2 \sum_{i \neq j} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2} \\ = o_p \left( \sum_{i \neq j} \mathbb{E} w_{ij}^2 \right).$$

Combing (S.6) with (S.7) and in view of (S.5) we arrive at

$$(S.8) \quad \frac{\sum_{i \neq j} \hat{w}_{ij}^2}{\sum_{i \neq j} \mathbb{E} w_{ij}^2} = 1 + o_p(1).$$

This together with (S.4) and (S.5) completes the proof of (8) in Theorem 3.2.

Next consider the case when condition (9) holds. The definition of  $\tilde{w}_{ij}$  entails that

$$\begin{aligned}
\text{(S.9)} \quad \sum_{i \neq j} \hat{w}_{ij}^2 &= \sum_{i \neq j} \left( \tilde{w}_{ij} + \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2 \\
&\leq 2 \left( \sum_{i \neq j} \tilde{w}_{ij}^2 + \sum_{i \neq j} \left( \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) \right)^2 \right) \\
&\leq 2 \left( \sum_{i \neq j} \tilde{w}_{ij}^2 + (K - K_0) \sum_{k=K_0+1}^K d_k^2 \sum_{i \neq j} \mathbf{v}_k^2(i) \mathbf{v}_k^2(j) \right) \\
&\leq 2 \left( \sum_{i \neq j} \tilde{w}_{ij}^2 + (K - K_0) \sum_{k=K_0+1}^K d_k^2 \right) \\
&\leq 2 \sum_{i \neq j} \tilde{w}_{ij}^2 + 2(K - K_0) \left( \sum_{k=K_0+1}^K |d_k|^2 \right) \lesssim \left( \sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k| \right)^2.
\end{aligned}$$

By (S.4), (S.5) and (S.9), in order to obtain (10), it suffices to show that

$$\text{(S.10)} \quad \frac{\sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k|}{\sqrt{m} \left| \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} \right|} = o_p(1).$$

For the term  $\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}$  in the numerator, we calculate its expectation and variance respectively as follows:

$$\mathbb{E} \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} = \frac{\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)}{m}$$

and

$$\text{var} \left( \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} \right) \lesssim \frac{(\sum_{k=K_0+1}^K |d_k|)^2}{m}$$

Therefore we have

$$\text{(S.11)} \quad \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} = \frac{\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)}{m} + O_p \left( \frac{\sum_{k=K_0+1}^K |d_k|}{\sqrt{m}} \right).$$

Combining (S.11) with condition (9) that  $\frac{|\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)|}{\sqrt{m} (\sqrt{\sum_{i \neq j} \mathbb{E} w_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k|)} \gg 1$  as well as (S.8) for the denominator, we conclude (S.10) and thus complete the proof.

**7.3. Proof of Theorem 3.3.** We start with analyzing the first inequality  $\sum_{i \neq j} \sigma_{ij}^2 \gg m$  in Condition 5. Combing the assumptions  $\theta_n = \max_{i \neq j} \sigma_{ij}^2, \min_{i \neq j} \sigma_{ij}^2 \gtrsim \theta_n (\log n)^{-\delta}$  with Condition 3, we can see that

$$\text{(S.12)} \quad n\theta_n (\log n)^{-\delta} \lesssim \alpha_n^2 \lesssim n\theta_n \quad \text{and} \quad \sum_{i \neq j} \sigma_{ij}^2 \gtrsim n^2 \theta_n (\log n)^{-\delta}.$$

Then the first inequality  $\sum_{i \neq j} \sigma_{ij}^2 \gg m$  is satisfied if

$$(S.13) \quad (n/K)^2 \theta_n (\log n)^{-2\delta} \gg m.$$

By Conditions 3 and (S.12), we have

$$(S.14) \quad \frac{|d_K|}{\alpha_n} \gtrsim \frac{\alpha_n}{(\log \log n)^{\epsilon'}} \gtrsim \sqrt{n \theta_n (\log n)^{-\delta} (\log \log n)^{-2\epsilon'}}.$$

Moreover, the assumption  $|d_1| \lesssim n \theta_n$  yields

$$(S.15) \quad \frac{|d_K|}{\alpha_n} \lesssim \frac{|d_1|}{\alpha_n} \lesssim \sqrt{n \theta_n}.$$

It follows from (S.14) and (S.15) that

$$(S.16) \quad \theta_n \gtrsim n^{-1} (\log n)^{\delta+2\epsilon+2} (\log \log n)^{2\epsilon'}$$

is a sufficient condition for ensuring  $\frac{|d_K|}{\alpha_n} \gtrsim (\log n)^{1+\epsilon}$  in Condition 3. In view of (S.13) and (S.16), any  $m$  satisfying

$$(S.17) \quad m \ll n K^{-2} (\log n)^{2\epsilon-\delta+2} (\log \log n)^{2\epsilon'}$$

makes the first inequality in Condition 5 hold.

Next, we discuss the second inequality in Condition 5, i.e.

$$\sum_{i \neq j} \sigma_{ij}^2 \gtrsim (\log n)^{\epsilon_1} \left( \frac{n \sum_{k=1}^K (\mathbf{1}^T \mathbf{v}_k)^2}{m} + \alpha_n^2 (\log n)^2 + \frac{n^2 \alpha_n^2 (\log n)^6}{m d_K^2} + \frac{n}{m} \right).$$

By assumption (12) and  $\theta_n = \max_{i \neq j} \sigma_{ij}^2$  and  $\min_{i \neq j} \sigma_{ij} \gtrsim \theta_n (\log n)^{-\delta}$ , it suffices to have

$$(S.18) \quad \sum_{i \neq j} \sigma_{ij}^2 \gtrsim n^2 \theta_n (\log n)^{-\delta} \gg \frac{n (\log n)^{\epsilon_1} \sum_{k=1}^K (\mathbf{1}^T \mathbf{v}_k)^2}{m} \\ + (\log n)^{2+\epsilon_1} \alpha_n^2 + \frac{n^2 (\log n)^{6+\epsilon_1} \alpha_n^2}{m d_K^2} + \frac{n (\log n)^{\epsilon_1}}{m}.$$

Now we compare the four terms on the very right hand side of (S.18) with  $n^2 \theta_n (\log n)^{-\delta}$  one by one. Note that the second term  $(\log n)^{2+\epsilon_1} \alpha_n^2 \lesssim n (\log n)^{2+\epsilon_1} \theta_n \ll n^2 \theta_n (\log n)^{-\delta}$ , making no contribution to the choice of  $m$ . For the fourth term, by Condition 3, we have the inequality that  $n (\log n)^{\epsilon_1} \ll n^2 \theta_n (\log n)^{-\delta}$ , and therefore this term makes no contribution to the choice of  $m$  either. For the third term, it is easy to see from (S.14) that  $n \theta_n \lesssim d_K^2 \alpha_n^{-2} (\log n)^{\delta} (\log \log n)^{2\epsilon'}$ , which guarantees that

$$\frac{n^2 (\log n)^{6+\epsilon_1} \alpha_n^2}{m d_K^2} \lesssim \frac{n (\log n)^{6+\delta+\epsilon_1} (\log \log n)^{2\epsilon'}}{m \theta_n}.$$

Therefore, any  $m$  satisfying

$$(S.19) \quad m \gg n^{-1} \theta_n^{-2} (\log n)^{6+2\delta+\epsilon_1} (\log \log n)^{2\epsilon'}$$

ensures that  $\frac{n^2 (\log n)^{6+\epsilon_1} \alpha_n^2}{m d_K^2} \ll n^2 \theta_n (\log n)^{-\delta}$ . Finally, for the first term, it suffices to have

$$(S.20) \quad m \gg n^{-1} \theta_n^{-1} (\log n)^{\delta+\epsilon_1} \sum_{k=1}^K (\mathbf{1}^T \mathbf{v}_k)^2.$$

By Cauchy-Schwarz inequality we know that  $\sum_{k=1}^K (\mathbf{1}^T \mathbf{v}_k)^2 \leq nK$ . The above two results entail that (S.20) is satisfied as long as

$$(S.21) \quad m \gg \theta_n^{-1} (\log n)^{\delta+\epsilon_1} K.$$

Combining (S.18), (S.19) and (S.21), it is easy to see that any  $m$  satisfying

$$(S.22) \quad m \gg \theta_n^{-1} (\log n)^{\delta+\epsilon_1} K + n^{-1} \theta_n^{-2} (\log n)^{6+2\delta+\epsilon_1} (\log \log n)^{2\epsilon'}$$

can make the second inequality in Condition 5 hold. Moreover, it follows from (S.16) that

$$(S.23) \quad m \gg n (\log n)^{-2\epsilon+\epsilon_1-2} (\log \log n)^{-2\epsilon'} K + n (\log n)^{\epsilon_1-4\epsilon+2} (\log \log n)^{-2\epsilon'}$$

is a sufficient condition for (S.22).

Summarizing the arguments above tells that (13) in Theorem 3.3 is sufficient for Condition 5 (see (S.13) and (S.22)), and  $n (\log n)^{-2\epsilon+\epsilon_1+2} (\log \log n)^{-2\epsilon'} K \ll m \ll nK^{-2} (\log n)^{2\epsilon-\delta+2} (\log \log n)^{2\epsilon'}$  is a sufficient condition to ensure (13) (see (S.17) and (S.23)).

To complete the proof of Theorem 3.3, it remains to verify inequality (9) in Theorem 3.2 under condition (13). By (S.12) and Condition 3 we have  $n\theta_n (\log n)^{-\delta} (\log \log n)^{-\epsilon'} \lesssim |d_k| \lesssim n\theta_n \lesssim n\sqrt{\theta_n}$  for all  $k = 1, \dots, K$ . This together with (S.12) entails that

$$n\sqrt{\theta_n (\log n)^{-\delta}} \lesssim \sqrt{\sum_{i \neq j} \sigma_{ij}^2} + \sum_{k=K_0+1}^K |d_k| \lesssim nK\sqrt{\theta_n}.$$

This implies that the inequality

$$(S.24) \quad \sqrt{m} \ll \frac{|\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)|}{nK\sqrt{\theta_n}}$$

is sufficient for (9). Furthermore, it follows from (12) and (S.12) that

$$\frac{|\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)|}{n\sqrt{\theta_n}} \gtrsim \frac{\sum_{i \neq j} \sigma_{ij}^2}{n\sqrt{\theta_n}} \gtrsim n\sqrt{\theta_n} (\log n)^{-\delta}.$$

Thus, any  $m \ll (n/K)^2 \theta_n (\log n)^{-2\delta}$  is sufficient for (9). This completes our proof.

7.4. *Proof of (14) in Theorem 3.4.* The asymptotic distribution (14) under the null hypothesis in Theorem 3.4 can be concluded from the following two results:

$$(S.25) \quad \frac{\sum_{i=1}^n \hat{w}_{ii}}{\sqrt{\sum_{i=1}^n \mathbb{E} w_{ii}^2}} \xrightarrow{d} N(0, 1) \quad \text{and}$$

$$(S.26) \quad \frac{\sum_{i=1}^n \hat{w}_{ii}^2}{\sum_{i=1}^n \mathbb{E} w_{ii}^2} = 1 + o_p(1).$$

We first prove (S.25). Under the null hypothesis ( $K = K_0$ ), Lemma 1 ensures that

$$(S.27) \quad \hat{w}_{ij} = w_{ij} - \sum_{k=1}^{K_0} \left[ \Delta(d_k) \hat{\mathbf{v}}_k(i) \hat{\mathbf{v}}_k(j) + d_k \Delta(\mathbf{v}_k(i)) \mathbf{v}_k(j) \right. \\ \left. + d_k \Delta(\mathbf{v}_k(j)) \mathbf{v}_k(i) + d_k \Delta(\mathbf{v}_k(i)) \Delta(\mathbf{v}_k(j)) \right].$$

By (S.81) in Lemma 3, Theorem S.1 and Corollary S.1 as well as Condition 4, we have

$$\begin{aligned}
\Delta(d_k) &= \hat{d}_k - d_k = \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} + \mathbf{v}_k^T \mathbb{E} \mathbf{W} \mathbf{v}_k + \mathbf{v}_k^T (\mathbf{W} - \mathbb{E} \mathbf{W}) \mathbf{v}_k + O_{p_u} \left( \frac{1}{\sqrt{n}} + \frac{\alpha_n}{|d_k|} \right) \\
&= \frac{O(\alpha_n^2)}{d_k} + O_{p_u}(\min\{1, \frac{\alpha_n}{\sqrt{n}}\}) + O_{p_u} \left( \frac{1}{\sqrt{n}} + 1 \right) \\
\text{(S.28)} \quad &= O_{p_u}(1) + O_{p_u}(\sqrt{\theta_n}),
\end{aligned}$$

where in the second step we have used  $\mathbf{v}_k^T \mathbb{E} \mathbf{W} \mathbf{v}_k = 0$  because the network has self loops. Recall the definition of  $\mathbf{r}_k$  in Section 3.6. Note that the following results can be proved by Theorem S.1, Corollary S.1 and (S.84):

$$\begin{aligned}
\text{(S.29)} \quad |\mathbf{r}_k(i)| &= |\mathbf{e}_i^T \mathbf{V}_{-k} (t_k \mathbf{D}_{-k}^{-1} - \mathbf{I})^{-1} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k| = O_{p_u} \left( \frac{1}{\sqrt{n}} \cdot \alpha_n^2 \right), \\
|\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k - \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k| &= |\mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + \mathbf{e}_i^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k| \\
&\lesssim \left\| \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \right\| \cdot \left\| \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k \right\| \\
&\quad + \left| \mathbf{e}_i^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \right| \\
&= \frac{1}{\sqrt{n}} \cdot O_{p_u}(1).
\end{aligned}$$

By Lemma 3, it is easy to see that  $d_k \sim t_k$  for all  $k = 1, \dots, K$ . Furthermore, we have

$$\begin{aligned}
\text{(S.30)} \quad \Delta(\mathbf{v}_k(i)) &= \hat{\mathbf{v}}_k(i) - \mathbf{v}_k(i) \\
&= \frac{\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k}{t_k} + \frac{\mathbf{r}_k(i)}{t_k^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k}{t_k^2} - \mathbf{v}_k(i) \frac{3\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{2t_k^2} + \frac{\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k - \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k}{t_k} \\
&\quad + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} + \frac{(\alpha_n \log n)^2}{n |t_k|^2} \right) \\
&= \frac{\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k}{t_k} + \frac{O_{p_u}(\alpha_n^2)}{\sqrt{n} t_k^2} + \frac{O_{p_u}((\alpha_n \log n)^2 / \sqrt{n}) + O(\alpha_n^2 / \sqrt{n})}{t_k^2} + O_{p_u} \left( \frac{1}{\sqrt{n} |t_k|} \right) \\
&\quad + \frac{O_{p_u}((\alpha_n \log n)^3)}{\sqrt{n} |t_k|^3} \\
&= \frac{\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k}{t_k} + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} |d_k|^2} + \frac{1}{\sqrt{n} |d_k|} \right),
\end{aligned}$$

uniformly for all  $i = 1, \dots, n$ , where the penultimate step uses Theorem S.1, Corollary S.1 and (S.29). Condition 4 together with (S.30) and Corollary S.1 ensures  $\|\hat{\mathbf{v}}_k\|_\infty \lesssim \frac{1}{\sqrt{n}} + O_{p_u} \left( \frac{\alpha_n \log n}{\sqrt{n} |d_k|} \right)$ . Using this and Condition 6 and substituting (S.28) and (S.30) into (S.27) gives us

$$\text{(S.31)} \quad \hat{w}_{ij} = w_{ij} + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n} \right) - \sum_{k=1}^{K_0} \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k}.$$

Therefore,

$$(S.32) \quad \frac{1}{n} \sum_{i=1}^n \hat{w}_{ii} = \frac{1}{n} \sum_{i=1}^n w_{ii} + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n} \right) - 2 \sum_{k=1}^{K_0} \frac{d_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k}{nt_k}.$$

Since  $\mathbf{X}$  contains selfloops,  $\mathbb{E} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = 0$ . Together with Corollary S.1 and the fact that  $\|\mathbf{v}_k\|_\infty \lesssim \frac{1}{\sqrt{n}}$ , we have

$$\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = O_{p_u} \left( \frac{\alpha_n}{\sqrt{n}} \right).$$

Then it follows from  $\alpha_n \lesssim \sqrt{n}$  and Condition 6 that (S.32) can be simplified to

$$(S.33) \quad \frac{1}{n} \sum_{i=1}^n \hat{w}_{ii} = \frac{1}{n} \sum_{i=1}^n w_{ii} + O_{p_u} \left( \frac{1}{n} + \frac{(\alpha_n \log n)^2}{n|d_{K_0}|} \right).$$

It is easy to see that  $\mathbb{E}(\frac{1}{n} \sum_{i=1}^n w_{ii}) = 0$  and  $\sqrt{\text{var}(\frac{1}{n} \sum_{i=1}^n w_{ii})} = \frac{1}{n} \sqrt{\sum_{i=1}^n \mathbb{E} w_{ii}^2} \gg \frac{1}{n} (\log n)^{2+\epsilon_2}$  by the condition of Theorem 3.4. Therefore (S.33) can be rewritten as

$$(S.34) \quad \frac{1}{n} \sum_{i=1}^n \hat{w}_{ii} = \frac{1}{n} \sum_{i=1}^n w_{ii} + o_p \left( \sqrt{\text{var}(\frac{1}{n} \sum_{i=1}^n w_{ii})} \right).$$

Moreover, according to Condition 1 it holds that

$$\frac{\sum_{i=1}^n \mathbb{E} w_{ii}^4}{(\sum_{i=1}^n \mathbb{E} w_{ii}^2)^2} \lesssim \frac{\sum_{i=1}^n \mathbb{E} w_{ii}^2}{(\sum_{i=1}^n \mathbb{E} w_{ii}^2)^2} \rightarrow 0,$$

satisfying Lyapunov's condition. Thus,  $\sum_{i=1}^n w_{ii} / \sqrt{\sum_{i=1}^n \mathbb{E} w_{ii}^2} \xrightarrow{d} N(0, 1)$ . This together with (S.34) proves (S.25).

Next we prove (S.26). Condition 1 yields

$$\text{var} \left( \sum_{i=1}^n w_{ii}^2 \right) \leq \sum_{i=1}^n \mathbb{E} w_{ii}^4 \lesssim \sum_{i=1}^n \mathbb{E} w_{ii}^2,$$

which implies that

$$\text{var} \left( \frac{\sum_{i=1}^n w_{ii}^2}{\sum_{i=1}^n \mathbb{E} w_{ii}^2} \right) \lesssim \frac{1}{\sum_{i=1}^n \mathbb{E} w_{ii}^2} \rightarrow 0.$$

Therefore we have

$$(S.35) \quad \frac{\sum_{i=1}^n w_{ii}^2}{\sum_{i=1}^n \mathbb{E} w_{ii}^2} = 1 + o_p(1).$$

Then to verify (S.26), it suffices to show that

$$\frac{\sum_{i=1}^n \hat{w}_{ii}^2}{\sum_{i=1}^n w_{ii}^2} = 1 + o_p(1).$$

By equation (S.35) and the condition  $\sum_{i=1}^n \mathbb{E} w_{ii}^2 \geq (\log n)^{4+2\epsilon_2}$  in Theorem 3.4, we only need to show

$$(S.36) \quad \sum_{i=1}^n \hat{w}_{ii}^2 = \sum_{i=1}^n w_{ii}^2 + O_{p_u}((\log n)^2).$$



In view of equation (S.31), it holds uniformly over all  $i = 1, \dots, n$  that

$$\begin{aligned} \hat{w}_{ii}^2 &= w_{ii}^2 + 2w_{ii}O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) - 4\sum_{k=1}^{K_0} w_{ii} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} + O_{p_u}\left(\frac{(\alpha_n \log n)^4}{n^2|d_{K_0}|^2} + \frac{1}{n^2}\right) \\ &\quad + 4\left(\sum_{k=1}^{K_0} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}\right)^2 - 4O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) \sum_{k=1}^{K_0} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}, \end{aligned} \quad (\text{S.37})$$

and thus

$$\begin{aligned} \sum_{i=1}^n \hat{w}_{ii}^2 &= \sum_{i=1}^n w_{ii}^2 + 2\sum_{i=1}^n \left(w_{ii}O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right)\right) - 4\sum_{k=1}^{K_0} \sum_{i=1}^n \left(w_{ii} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}\right) \\ &\quad + O_{p_u}\left(\frac{(\alpha_n \log n)^4}{n|d_{K_0}|^2} + \frac{1}{n}\right) + 4\sum_{i=1}^n \left(\sum_{k=1}^{K_0} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}\right)^2 \\ (\text{S.38}) \quad &- 4\sum_{i=1}^n \sum_{k=1}^{K_0} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k}. \end{aligned}$$

To prove (S.36), we study the terms on the right hand side of (S.38). To begin with, we know that

$$(\text{S.39}) \quad \left| \sum_{i=1}^n w_{ii} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) \right| \leq \sum_{i=1}^n |w_{ii}| O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) = O_{p_u}((\log n)^2).$$

Next, we write

$$(\text{S.40}) \quad \sum_{i=1}^n w_{ii} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = \sum_{i=1}^n \frac{d_k w_{ii}^2 \mathbf{v}_k^2(i)}{t_k} + \sum_{1 \leq l \neq i \leq n} \frac{d_k w_{ii} w_{il} \mathbf{v}_k(i) \mathbf{v}_k(l)}{t_k}.$$

For the first term on the right hand side of (S.40), it follows from  $\mathbb{E} \left| \sum_{i=1}^n \frac{d_k w_{ii}^2 \mathbf{v}_k^2(i)}{t_k} \right| \lesssim$

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} w_{ii}^2 \lesssim 1 \text{ that}$$

$$\sum_{i=1}^n \frac{d_k w_{ii}^2 \mathbf{v}_k^2(i)}{t_k} = O_{p_u}(1).$$

For the second term, it follows from the calculations

$$\mathbb{E} \left( \sum_{1 \leq l \neq i \leq n} \frac{d_k w_{ii} w_{il} \mathbf{v}_k(i) \mathbf{v}_k(l)}{t_k} \right) = 0 \quad \text{and}$$

$$\begin{aligned} \text{var} \left( \sum_{1 \leq l \neq i \leq n} \frac{d_k w_{ii} w_{il} \mathbf{v}_k(i) \mathbf{v}_k(l)}{t_k} \right) &\lesssim \sum_{1 \leq l \neq i \leq n} \mathbb{E} w_{ii}^2 \mathbb{E} w_{il}^2 \mathbf{v}_k^2(j) \mathbf{v}_k^2(l) \\ &\lesssim \frac{\alpha_n^2}{n^2} \sum_{i=1}^n \mathbb{E} w_{ii}^2 \lesssim \frac{\sum_{i=1}^n \mathbb{E} w_{ii}^2}{n} \lesssim 1 \end{aligned}$$

that

$$\sum_{1 \leq l \neq i \leq n} \frac{d_k w_{ii} w_{il} \mathbf{v}_k(i) \mathbf{v}_k(l)}{t_k} = O_{p_u}(1).$$

Therefore,

$$(S.41) \quad \sum_{i=1}^n w_{ii} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_{p_u}(1)$$

and

$$\sum_{k=1}^{K_0} \sum_{i=1}^n w_{ii} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_{p_u}(1).$$

Since  $\mathbb{E} \mathbf{W} = 0$ , it follows from  $d_k \sim t_k$ , Condition 4 and Corollary S.1 that

$$\frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} = O_{p_u}(\alpha_n/n) = O_{p_u}\left(\frac{1}{\sqrt{n}}\right).$$

By Condition 4, Condition 6 and Corollary S.1 we can see that

$$(S.42) \quad \begin{aligned} \sum_{i=1}^n \left( \sum_{k=1}^{K_0} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \right)^2 &\leq K_0 \sum_{k=1}^{K_0} \sum_{i=1}^n \frac{d_k^2 (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))^2}{t_k^2} \\ &= O_{p_u}\left(\frac{\alpha_n^2}{n}\right) = O_{p_u}(1). \end{aligned}$$

And we also have

$$(S.43) \quad \begin{aligned} &\left| \sum_{i=1}^n \sum_{k=1}^{K_0} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n}\right) \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \right| \\ &= \sum_{i=1}^n \sum_{k=1}^{K_0} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n |d_{K_0}|} + \frac{1}{n}\right) O_{p_u}\left(\frac{1}{\sqrt{n}}\right) = O_{p_u}\left(\frac{(\log n)^2}{\sqrt{n}}\right). \end{aligned}$$

Then (S.36) is concluded from combining the arguments (S.38), (S.39), (S.41), (S.42) and (S.43). Therefore, equation (S.26) holds and the proof of (14) in Theorem 3.4 is completed.

*7.5. Proof of Theorem 3.5.* Recall that Theorem 3.5 assumes the null hypothesis  $K = K_0$ . Plugging the expansions of  $\Delta(d_k)$  and  $\Delta(\mathbf{v}_k(i))$  in Lemma 3 into the expression in Lemma 1, and using results in Theorem S.1, Corollary S.1 and (S.29) we arrive at

$$(S.44) \quad \begin{aligned} \hat{w}_{ij} &= w_{ij} - \sum_{k=1}^{K_0} \mathbf{v}_k(i) \mathbf{v}_k(j) \left( -2 \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} + \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \right) \\ &\quad - \sum_{k=1}^{K_0} \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &\quad - \sum_{k=1}^{K_0} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} - \sum_{k=1}^{K_0} \frac{d_k (\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{s}_{k,j}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \\ &\quad - \sum_{k=1}^{K_0} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k}{t_k^2} + O_{p_u}\left(\frac{\alpha_n (\log n)^3}{n |d_{K_0}|} + \frac{1}{n^{3/2}}\right), \end{aligned}$$

uniformly over all  $i, j$ . Summing up (S.44) over subscripts  $i$  and  $j$  yields

(S.45)

$$\begin{aligned} \sum_{i \neq j} \hat{w}_{ij} &= \sum_{i \neq j} w_{ij} - \sum_{k=1}^{K_0} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j)) \\ &\quad - \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} - \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{\mathbf{r}_k(i) \mathbf{v}_k(j) + \mathbf{r}_k(j) \mathbf{v}_k(i)}{t_k} \\ &\quad + 2 \sum_{k=1}^{K_0} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} \sum_{i \neq j} (\mathbf{v}_k(i) \mathbf{v}_k(j)) - 2 \sum_{k=1}^{K_0} \mathbf{s}_k^T \mathbf{W} \mathbf{v}_k \sum_{i=1}^n \mathbf{v}_k(i) + O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{|d_{K_0}|} + \sqrt{n} \right), \end{aligned}$$

where we have made use of the following relationship, which is a direct consequence of Theorem S.1 and Corollary S.1

$$\begin{aligned} \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{d_k \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k}{t_k^2} &= \sum_{k=1}^{K_0} \frac{d_k \mathbf{1}^T \mathbf{W} \mathbf{v}_k \mathbf{1}^T \mathbf{W} \mathbf{v}_k}{t_k^2} - \sum_{k=1}^{K_0} \sum_{i=1}^n \frac{d_k (\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k)^2}{t_k^2} \\ &= O_{p_u} \left( \frac{\alpha_n^2 (\log n)^2}{d_{K_0}} \right). \end{aligned}$$

Recall that  $a_k = \sum_{i=1}^n \mathbf{v}_k(i)$ . By Cauchy–Schwarz inequality,  $|a_k| \leq \sqrt{n}$ . Furthermore, Corollary S.1 ensures that

$$\mathbf{1}^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k = O_{p_u} (\alpha_n^2).$$

Therefore by Theorem S.1 and Corollary S.1 we have

$$\begin{aligned} &\sum_{k=1}^{K_0} \sum_{i \neq j} \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &= \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{\mathbf{e}_i^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &\quad + \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &= 2 \sum_{k=1}^{K_0} \frac{\mathbf{1}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k a_k}{t_k} + 2 \sum_{k=1}^{K_0} \frac{\mathbf{1}^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k a_k}{t_k} \\ &\quad - 2 \sum_{k=1}^{K_0} \sum_{i=1}^n \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k \mathbf{v}_k(i)}{t_k} \\ &= 2 \sum_{k=1}^{K_0} \frac{\mathbf{1}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k a_k}{t_k} + O_{p_u} \left( \frac{n \alpha_n}{|d_{K_0}|} \right), \end{aligned}$$

where we have used the simple inequality that  $\alpha_n \lesssim \sqrt{n}$ . Then (S.45) can be written as

$$\sum_{i \neq j} \hat{w}_{ij} = \sum_{i \neq j} w_{ij} \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) \right) - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k^T \text{diag}(\mathbf{W}) \mathbf{v}_k$$

$$\begin{aligned}
& -2 \sum_{k=1}^{K_0} \frac{\mathbf{1}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k a_k}{t_k} - 2 \sum_{k=1}^{K_0} a_k \frac{\mathbf{1}^T \mathbf{r}_k}{t_k} + 2 \sum_{k=1}^{K_0} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} a_k^2 \\
& - 2 \sum_{k=1}^{K_0} a_k \mathbf{s}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{|d_{K_0}|} + \sqrt{n} \right) \\
= & 2 \sum_{i < j} w_{ij} \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right) \\
& - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k^T \text{diag}(\mathbf{W}) \mathbf{v}_k - 2 \sum_{k=1}^{K_0} \frac{\mathbf{1}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k a_k}{t_k} - 2 \sum_{k=1}^{K_0} a_k \frac{\mathbf{1}^T \mathbf{r}_k}{t_k} \\
& + 2 \sum_{k=1}^{K_0} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} a_k^2 - 2 \sum_{k=1}^{K_0} a_k \mathbf{s}_k^T \text{diag}(\mathbf{W}) \mathbf{v}_k + O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{|d_{K_0}|} + \sqrt{n} \right) \\
= & 2 \sum_{i < j} w_{ij} \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right) \\
& - \mathcal{R}(K_0) + O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{|d_{K_0}|} + \sqrt{n} \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \sum_{i \neq j} \hat{w}_{ij} + \mathcal{R}(K_0) \\
= & \underbrace{2 \sum_{i < j} w_{ij} \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)}_{S_n} \\
& + O_{p_u} \left( \frac{n \alpha_n (\log n)^3}{|d_{K_0}|} + \sqrt{n} \right),
\end{aligned}$$

where  $S_n$  is the sum of independent random variables and its variance equals to

$$\text{var}(S_n) = 4 \sum_{i < j} \sigma_{ij}^2 \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)^2.$$

In addition, by (S.85) and the definition of  $\mathbf{s}_k(j)$ , we have

$$\text{(S.46)} \quad \|\mathbf{s}_k(j) - \mathbf{e}_j\| \lesssim \frac{1}{\sqrt{n}}.$$

This together with  $|a_k| \leq \sqrt{n}$  implies that

$$\max_{1 \leq i < j \leq n} \left| 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right| \lesssim 1.$$

Combining the above result with the condition  $\max_{1 \leq i < j \leq n} |w_{i,j}| \leq C$  and (21) we have

$$[\text{var}(S_n)]^{-2} \times \left[ 2 \sum_{i < j} \mathbb{E} w_{ij}^4 \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)^4 \right]$$

$$\lesssim [\text{var}(S_n)]^{-1} \rightarrow 0.$$

Combining (21) with Lyapunov's condition, we can conclude that  $\sum_{i \neq j} \hat{w}_{ij} + R(K_0)$  converges weakly to the standard normal distribution after centralization and normalization, i.e.

$$\frac{\sum_{i \neq j} \hat{w}_{ij} + R(K_0)}{2 \sqrt{\sum_{i < j} \sigma_{ij}^2 \left( 1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i) \mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k (\mathbf{v}_k(j) \mathbf{s}_k(i) + \mathbf{v}_k(i) \mathbf{s}_k(j)) \right)^2}} \xrightarrow{d} N(0, 1).$$

This completes the proof of the theorem.

*7.6. Proof of (15) and (16) in Theorem 3.4.* Recall (S.3). The proof of (15) and (16) in Theorem 3.4 is similar to that of Theorem 3.2 and thus we omit the details.

*7.7. Proof of Theorem 3.6.* The proof of Theorem 3.6 is almost the same as Theorem 3.2 by defining  $\tilde{w}_{ij} = \hat{w}_{ij} - \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j)$  and thus we omit the details.

*7.8. Proof of Corollary 2.* The proof is similar to the one of Theorem 3.2. The only difference lies in the fact that the additional condition (17) ensures the estimation procedure rejects all  $K_0 < K$  with probability 1 even when  $K$  diverges. We below show the key steps. According to the proof of Theorem 3.2, we know that

$$\begin{aligned} \text{(S.47)} \quad & \left( T_n - \frac{\sqrt{m} \sum_{i \neq j} \sum_{k=K_0+1}^K d_k \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}}{\sqrt{2 \sum_{i \neq j} \hat{w}_{ij}^2}} \right) \frac{\sqrt{\sum_{i \neq j} \hat{w}_{ij}^2}}{\sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2}} \\ &= \frac{\sqrt{m} \sum_{i \neq j} \tilde{w}_{ij} Y_{ij}}{\sqrt{2 \sum_{i \neq j} \tilde{w}_{ij}^2}} \xrightarrow{d} N(0, 1). \end{aligned}$$

Moreover, by Bernstein inequality we have

$$\text{(S.48)} \quad \frac{\sqrt{m} \sum_{i \neq j} w_{ij} Y_{ij}}{\sqrt{2 \sum_{i \neq j} E w_{ij}^2}} = O_{p_u}(1).$$

Note that  $\tilde{w}_{ij}$  is the residual term under the alternative hypothesis. Similar to (39), we also have

$$\text{(S.49)} \quad \frac{\sum_{i \neq j} \tilde{w}_{ij}^2}{\sum_{i \neq j} E w_{ij}^2} = 1 + O_{p_u}\left(\frac{1}{(\log n)^{\epsilon_1/2}}\right).$$

Consider the case when condition (17) holds. Recall that (S.9) tells

$$\text{(S.50)} \quad \sum_{i \neq j} \hat{w}_{ij}^2 \lesssim \left( \sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k| \right)^2.$$

Without loss of generality, we assume that  $\delta_1 < \epsilon_1$ . By (S.47), (S.49) and (S.50), in order to obtain the conclusion, it suffices to show that

$$\text{(S.51)} \quad \frac{\sqrt{\sum_{i \neq j} \tilde{w}_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k|}{\sqrt{m} \left| \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} \right|} = O_{p_u}\left(\frac{1}{(\log n)^{\delta_1/3}}\right).$$

For the term  $\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij}$  in the numerator, by (S.11) we have

$$(S.52) \quad \sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j) Y_{ij} = \frac{\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)}{m} + O_{p_u} \left( \frac{\sum_{k=K_0+1}^K |d_k|}{\sqrt{m}} \right).$$

Combining (S.52) with condition (17) that  $\frac{|\sum_{k=K_0+1}^K d_k \sum_{i \neq j} \mathbf{v}_k(i) \mathbf{v}_k(j)|}{\sqrt{m \log^{\delta_1} n (\sqrt{\sum_{i \neq j} \mathbb{E} w_{ij}^2} + \sqrt{K - K_0} \sum_{k=K_0+1}^K |d_k|)}} \gg 1$  as well as (S.49) for the denominator, we conclude (S.51). This together with (S.48) and the condition  $K \leq O(\log \log n)$  completes the proof.

**8. Lemmas and their proofs.** Before presenting the key Lemmas, we first state the following Corollary, which is a consequence of Lemmas 7, 8 in [1] and Lemma 11 in [2] by carefully checking the corresponding proofs. This corollary is used to control the small term containing  $\mathbf{v}_k^T \mathbf{W}^l \mathbf{v}_t$  in this paper.

**COROLLARY S.1.** *Under Conditions 1-4, it holds that for any positive integers  $l$ , and unit vectors  $\mathbf{x}$  and  $\mathbf{y}$ ,*

$$(S.53) \quad \mathbb{E} \left[ \mathbf{x}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{y} \right]^2 \leq c(c_1 l)^{2l} (\min\{\alpha_n^{l-1}, d_{\mathbf{x}} \alpha_n^l, d_{\mathbf{y}} \alpha_n^l\})^2.$$

$$(S.54) \quad \left| \mathbb{E} \left[ \mathbf{x}^T \mathbf{W}^l \mathbf{y} \right] \right| \leq c(c_1 l)^l \alpha_n^l,$$

where  $d_x = \|\mathbf{x}\|_\infty$ ,  $d_y = \|\mathbf{y}\|_\infty$  and  $c$  and  $c_1$  are some positive constants independent of  $l$ .

**REMARK 1.** *Corollary S.1 bounded the second moment of  $|\mathbf{x}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{y}|$ . If a stronger but nontrivial tail probability of  $|\mathbf{v}_i^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k|$  can be proved, it is possible to extend our theoretical results to the case  $K = O(\log n)$ . We leave it to the future work.*

The following Theorem S.1 is a direct consequence of Lemma 5.4 in [3], which we include here for easier reference. One should notice that although Lemma 5.4 in [3] is proved under Model (1) in [3], by checking the proof of Lemma 5.4 carefully, we can see that it can be trivially extended to our model by almost the same proof.

**THEOREM S.1.** *Under Conditions 1-4, there exists a positive constant  $C_0$  such that for any positive constant  $\epsilon_0 < \epsilon$ , we have*

$$(S.55) \quad P(|\mathbf{e}_i^T \mathbf{W}^l \mathbf{y}| > (C_0 \alpha_n)^l (\log n)^{l(1+\epsilon_0)} d_y) \leq C e^{-\frac{1}{3}(\log n)^{1+\epsilon_0}},$$

where  $l$  is a positive integer not larger than  $\log n$ ,  $C$  is a positive integer independent of  $n$  and  $\mathbf{y}$  is a unit vector. Furthermore, we have

$$(S.56) \quad |\mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{y}| \leq l^l (C_0 \alpha_n)^l d_y.$$

Now we are ready to proceed to the key lemmas as well as their proofs.

8.1. *Lemma 2.* We will need the following notations for the proof of the lemma.

$$\mathcal{P}(\mathbf{M}_1, \mathbf{M}_2, t) = t\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, t).$$

$$A_{\mathbf{x}, k, t} = \mathcal{P}(\mathbf{x}, \mathbf{v}_k, t) - \mathcal{P}(\mathbf{x}, \mathbf{V}_{-k}, t) [t(\mathbf{D}_{-k})^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t)]^{-1} \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t),$$

$$\tilde{\mathcal{P}}_{k, t} = [t^2(A_{\mathbf{v}_k, k, t}/t)']^{-1},$$

where  $\mathcal{R}(\cdot)$  is defined at the beginning of Section 3.6,  $\mathbf{x}$  is a vector of appropriate dimension and the derivative in  $(\cdot)'$  is with respect to  $t$ .

LEMMA 2. *Assume there is no eigenvalue multiplicity. Under Conditions 1-4 and 6, we have the following expansion*

$$\begin{aligned} \mathbf{e}_i^T \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T \mathbf{v}_k &= \left[ \tilde{\mathcal{P}}_{k, t_k} - 2t_k^{-1} \tilde{\mathcal{P}}_{k, t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right) \right] \\ &\quad \times \left[ A_{\mathbf{e}_i, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{e}_i, k, t_k}^T \mathbf{W} \mathbf{v}_k - \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k}{t_k^2} + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right) \right] \\ (S.57) \quad &\quad \times \left[ A_{\mathbf{v}_k, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{v}_k, k, t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right) \right]. \end{aligned}$$

In addition, we have

$$(S.58) \quad \hat{d}_k = t_k + \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{1}{\sqrt{n}} \right).$$

PROOF. Proof of this lemma is similar to the one of Lemma 8 in [2] (one can also refer to Theorems 4 and 5 of [1]) and we focus on the proof of (S.57). Recall the conditions required in the Lemma 8 of [2]. Condition 1 therein is our Condition 2. And according to their proof, Conditions 2 and 4 therein are needed just for the sake of following two statements:

1). Lemma 6 of [2] holds, that is  $\max_{1 \leq k \leq K} \|\mathbf{v}_k\|_\infty = \|\mathbf{V}\|_\infty \lesssim \frac{1}{\sqrt{n}}$ . This is Condition 4 of our paper.

2).  $\frac{|d_K|}{\alpha_n} \geq n^\epsilon$  for some positive constant  $\epsilon$ . This is actually stronger than Condition 3 of our paper that  $\frac{|d_K|}{\alpha_n} \geq (\log n)^{1+\epsilon}$ . We will state how to relax the conditions in the proof of Lemma 8 in [2] to fit our model.

Actually, the crucial difference between  $\frac{|d_K|}{\alpha_n} \geq (\log n)^{1+\epsilon}$  and  $\frac{|d_K|}{\alpha_n} \geq n^\epsilon$  in the proof of Lemma 8 is to control the small order term  $\sum_{l=3}^{\infty} z^{-l-1} \mathbf{u}^T \mathbf{W}^l \mathbf{v}_k$ ,  $|z| \gtrsim |d_k|$ ,  $\mathbf{u} = \mathbf{v}_l, \mathbf{e}_i$ ,  $l = 1, \dots, K, i = 1, \dots, n$ .

Let  $L = \log n$ , without loss of generality we assume  $\log n$  is an integer. By the simple inequality  $\alpha_n \lesssim \sqrt{n}$  we have the following inequality

$$(S.59) \quad \frac{\sqrt{n} \alpha_n^{L+1} (C \log n)^{(L+1)/2}}{|d_K|^{L-3}} \leq \frac{\sqrt{n} \alpha_n^4 (C \log n)^{(L+1)/2}}{(\log n)^{(L-3)(1+\epsilon)}} \leq \frac{C^{(L+1)/2} n^2}{(\log n)^{(\log n - 7)/2}} \rightarrow 0$$

for any positive constant  $C$ . In view of  $\mathbf{u} = \mathbf{v}_l$ , it is sufficient to establish the following two equations

$$(S.60) \quad \sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{u}^T \mathbf{W}^l \mathbf{v}_k = O_{p_u} \left( \frac{1}{\sqrt{n} |z|^5} \right)$$

and

$$(S.61) \quad \sum_{l=3}^L z^{-(l+1)} \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k = O_{p_u} \left( \frac{\alpha_n^3}{\sqrt{n} |z|^4} \right)$$

for  $|z| \gtrsim d_K$ . In fact, (S.60) is a direct consequence of (S.59) and Theorem 6.2 of [4] (e.g. the proof of Lemma 1 in [2]). Moreover, for any positive constant  $C$  and sufficiently large  $n$ , we can conclude that

$$(S.62) \quad \left\{ \frac{\alpha_n^l (C \log n)^l}{|z|^l} \right\} \text{ is a decreasing sequence when } l \text{ is increasing.}$$

Then it follows from Corollary S.1,  $K \lesssim \log \log n$  and (S.62) that

$$(S.63) \quad \sum_{l=3}^9 z^{-(l+1)} \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k = O_{p_u} \left( \frac{\alpha_n^3}{\sqrt{n} |z|^4} \right), \mathbf{u} = \mathbf{v}_l,$$

$$\begin{aligned} & \left| \mathbb{E} \left[ \sum_{l=\sqrt{L}}^L z^{-(l+1)} \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k \right]^2 \right| \leq L \sum_{l=\sqrt{L}}^L |z|^{-(2l+2)} \mathbb{E} \left[ \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k \right]^2 \\ & \leq CL \sum_{l=\sqrt{L}}^L \frac{(c_1 l)^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}} \leq CL \sum_{l=\sqrt{L}}^L \frac{(c_1 \log n)^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}} \\ & \leq c(\log n)^8 \frac{(c_1)^{2\sqrt{\log n}} \alpha_n^4}{|z|^8 (\log n)^{\epsilon(2\sqrt{\log n}-6)}} \ll \frac{\alpha_n^4}{n |z|^8}, \end{aligned}$$

(S.64)

and

$$\begin{aligned} & \left| \mathbb{E} \left[ \sum_{l=10}^{\sqrt{L}} z^{-(l+1)} \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k \right]^2 \right| \leq \sqrt{L} \sum_{l=10}^{\sqrt{L}} |z|^{-(2l+2)} \mathbb{E} \left[ \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k \right]^2 \\ & \leq C \sqrt{L} \sum_{l=10}^{\sqrt{L}} \frac{(c_1 l)^{2l} \alpha_n^{2l}}{n |z|^{2l+2}} \leq C \sqrt{L} \sum_{l=10}^{\sqrt{L}} \frac{(c_1 \sqrt{\log n})^{2l} \alpha_n^{2l}}{n |z|^{2l+2}} \\ & \leq C \log n \frac{\alpha_n^4 ((c_1)^2 \log n)^{10} (\log \log n)^{8\epsilon'}}{n |z|^8 (\log n)^{14(1+\epsilon)}} \ll \frac{\alpha_n^4}{n |z|^8}. \end{aligned}$$

(S.65)

Here the constants  $C$  and  $c_1$  may vary from line to line. Combining (S.63)–(S.65) yields (S.61). For  $\mathbf{u} = \mathbf{e}_i$ , (S.60) also holds. Then it suffices to prove similar inequality as (S.61) for  $\mathbf{u} = \mathbf{e}_i$ . Actually, by Theorem S.1 we have

$$(S.66) \quad \sum_{l=3}^L z^{-(l+1)} \mathbf{u}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{v}_k = O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |z|^4} \right), \mathbf{u} = \mathbf{e}_i.$$

Therefore, by almost the same proof, Lemma 8 in [2] also holds under the conditions of our paper replacing  $O_{\prec}$  therein with  $O_{p_u}$ , which directly implies (S.58) and the following



expansion

$$\begin{aligned}
\mathbf{u}^T \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T \mathbf{v}_k &= \frac{\hat{d}_k^2 \mathbf{u}^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k \mathbf{v}_k^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k}{\hat{d}_k^2 \mathbf{v}_k^T \left[ \mathbf{G}'(\hat{d}_k) - \mathbf{F}'_k(\hat{d}_k) \right] \mathbf{v}_k} \\
&= \left[ \tilde{\mathcal{P}}_{k,t_k} - 2t_k^{-1} \tilde{\mathcal{P}}_{k,t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{nt_k^2}} \right) \right] \\
&\quad \times \left[ A_{\mathbf{u},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{u},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{nt_k^2}} \right) \right] \\
\text{(S.67)} \quad &\quad \times \left[ A_{\mathbf{v}_k,k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{v}_k,k,t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{nt_k^2}} \right) \right],
\end{aligned}$$

for  $\mathbf{u} = \mathbf{e}_i$  or  $\mathbf{v}_k$ , where

$$\text{(S.68)} \quad \mathbf{G}(z) = (\mathbf{W} - z\mathbf{I})^{-1}$$

and

$$\text{(S.69)} \quad \mathbf{F}_k(z) = \mathbf{G}(z) \mathbf{V}_{-k} [\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k}]^{-1} \mathbf{V}_{-k}^T \mathbf{G}(z).$$

Comparing (S.67) with Lemma 2, we see that to prove Lemma 2 we only need to show when  $\mathbf{u} = \mathbf{e}_i$ , the term  $\left[ A_{\mathbf{e}_i,k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{e}_i,k,t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{nt_k^2}} \right) \right]$  in (S.67) can be further expanded as  $\left[ A_{\mathbf{e}_i,k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{e}_i,k,t_k}^T \mathbf{W} \mathbf{v}_k - \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k}{t_k^2} + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right) \right]$ . In fact, by comparing these two terms we see that Lemma 2 indeed provides higher order expansion of the remainder term  $O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{nt_k^2}} \right)$  in (S.67). We next discuss how to obtain this higher order expansion.

By Theorem S.1, Corollary S.1, (S.60)-(S.61) and (S.66), we have

(S.70)

$$\begin{aligned}
\mathbf{e}_i^T \mathbf{G}(z) \mathbf{y} &= -z^{-1} \mathbf{e}_i^T \mathbf{y} - z^{-2} \mathbf{e}_i^T \mathbf{W} \mathbf{y} - \sum_{l=2}^L z^{-(l+1)} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{y} \\
&\quad - \sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{e}_i^T \mathbf{W}^l \mathbf{y} - \sum_{l=2}^L z^{-(l+1)} \mathbf{e}_i^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{y} \\
&= -z^{-1} \mathbf{e}_i^T \mathbf{y} - z^{-2} \mathbf{e}_i^T \mathbf{W} \mathbf{y} - z^{-3} \mathbf{e}_i^T \mathbf{W}^2 \mathbf{y} - \sum_{l=3}^L z^{-(l+1)} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{y} + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |z|^4} \right),
\end{aligned}$$

(S.71)

$$\begin{aligned}
\mathbf{v}_k^T \mathbf{G}(z) \mathbf{y} &= -z^{-1} \mathbf{v}_k^T \mathbf{y} - z^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{y} - \sum_{l=2}^L z^{-(l+1)} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{y} \\
&\quad - \sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{v}_k^T \mathbf{W}^l \mathbf{y} - \sum_{l=2}^L z^{-(l+1)} \mathbf{v}_k^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathbf{y} \\
&= -z^{-1} \mathbf{v}_k^T \mathbf{y} - z^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{y} - z^{-3} \mathbf{v}_k^T \mathbf{W}^2 \mathbf{y} - \sum_{l=3}^L z^{-(l+1)} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{y} + O_{p_u} \left( \frac{\alpha_n^3}{\sqrt{n} |z|^4} \right),
\end{aligned}$$

for all  $z \sim t_k$  and  $\|\mathbf{y}\|_\infty \lesssim \frac{1}{\sqrt{n}}$ . By Theorem S.1 and Corollary S.1, we have the following higher order expansions

$$(S.72) \quad \begin{aligned} \mathbf{e}_i^T \mathbf{G}(z) \mathbf{v}_k &= -z^{-1} \mathbf{e}_i^T \mathbf{v}_k - z^{-2} \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k - z^{-3} \mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k - \sum_{l=3}^L z^{-(l+1)} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k \\ &\quad + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |z|^4} \right), \end{aligned}$$

$$(S.73) \quad \mathbf{v}_k^T \mathbf{G}(z) \mathbf{v}_k = -z^{-1} - z^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \sum_{l=2}^L z^{-(l+1)} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} |z|^3} \right),$$

$$(S.74) \quad \mathbf{v}_k^T \mathbf{G}(z) \mathbf{V}_{-k} = -z^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{V}_{-k} - \sum_{l=2}^L z^{-(l+1)} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{V}_{-k} + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} |z|^3} \right),$$

$$(S.75) \quad \begin{aligned} \mathbf{e}_i^T \mathbf{G}(z) \mathbf{V}_{-k} &= -z^{-1} \mathbf{e}_i^T \mathbf{V}_{-k} - z^{-2} \mathbf{e}_i^T \mathbf{W} \mathbf{V}_{-k} - z^{-3} \mathbf{e}_i^T \mathbf{W}^2 \mathbf{V}_{-k} \\ &\quad - \sum_{l=3}^L z^{-(l+1)} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{V}_{-k} + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |z|^4} \right), \end{aligned}$$

$$(S.76) \quad \begin{aligned} \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} &= -z^{-1} - z^{-2} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{V}_{-k} - \sum_{l=2}^L z^{-(l+1)} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^l \mathbf{V}_{-k} \\ &\quad + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} |z|^3} \right). \end{aligned}$$

It follows from (S.72)-(S.75) that

$$(S.77) \quad \begin{aligned} \mathbf{e}_i^T \mathbf{F}_k(z) \mathbf{v}_k &= \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, z) \left[ \mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \\ &\quad - z^{-2} \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, z) \left[ \mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |z|^4} \right). \end{aligned}$$

Moreover, according to the proof in [1], the term  $\left[ A_{\mathbf{u},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{u},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right) \right]$  is the expansion of  $\hat{d}_k \mathbf{u}^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k$ , i.e.

$$(S.78) \quad \hat{d}_k \mathbf{u}^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k = A_{\mathbf{u},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{u},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right).$$

Therefore by (S.58), (S.72) and (S.77) we have

$$(S.79) \quad \begin{aligned} &\hat{d}_k \mathbf{e}_i^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k \\ &= \mathcal{P}(\mathbf{e}_i, \mathbf{v}_k, \hat{d}_k) - \mathcal{P}(\mathbf{e}_i, \mathbf{V}_{-k}, \hat{d}_k) \left[ \hat{d}_k \mathbf{D}_{-k}^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, \hat{d}_k) \right]^{-1} \\ &\quad \times \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, \hat{d}_k) - \hat{d}_k^{-1} \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k - \hat{d}_k^{-2} \mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k + \hat{d}_k^{-1} \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, \hat{d}_k) \\ &\quad \times \left[ \mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, \hat{d}_k) \right]^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right) \end{aligned}$$

$$\begin{aligned}
&= \mathcal{P}(\mathbf{e}_i, \mathbf{v}_k, t_k) - \mathcal{P}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) [t_k \mathbf{D}_{-k}^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \\
&\quad \times \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k) - t_k^{-1} \mathbf{e}_i^T \mathbf{W} \mathbf{v}_k - t_k^{-2} \mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k + t_k^{-1} \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) \\
&\quad \times [\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right) \\
&= \mathbf{A}_{\mathbf{e}_i, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{e}_i, k, t_k}^T \mathbf{W} \mathbf{v}_k - t_k^{-2} \mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right).
\end{aligned}$$

That is to say, we can replace (S.79) with (S.78) and this completes the proof.  $\square$

### 8.2. Lemma 3.

LEMMA 3. *Assume there is no eigenvalue multiplicity. Under Conditions 1-4 and 6, fixing the direction  $\hat{\mathbf{v}}_k$  such that  $\hat{\mathbf{v}}_k^T \mathbf{v}_k \geq 0$ , we have the following expansion*

$$\begin{aligned}
\hat{\mathbf{v}}_k(i) &= \mathbf{v}_k(i) + \frac{\mathbf{r}_k(i)}{t_k^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k}{t_k^2} - \mathbf{v}_k(i) \frac{3\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{2t_k^2} + \frac{\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k}{t_k} \\
&\quad + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} + \frac{(\alpha_n \log n)^2}{n |t_k|^2} \right) \\
\text{(S.80)} \quad &= \mathbf{v}_k(i) + \frac{\mathbf{r}_k(i)}{t_k^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k}{t_k^2} - \mathbf{v}_k(i) \frac{3\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{2t_k^2} + \frac{\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k}{t_k} + O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} d_k^2} \right),
\end{aligned}$$

where  $\mathbf{r}_k$  and  $\mathbf{s}_{k,i}$  are defined in Section 3.6. Moreover,

$$\text{(S.81)} \quad \hat{d}_k - d_k = \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} + \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{1}{\sqrt{n}} + \frac{\alpha_n}{|d_k|} \right).$$

PROOF. We prove (S.80) first. By Corollary S.1, we have

$$\text{(S.82)} \quad t_k \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, t_k) = \mathcal{P}(\mathbf{v}_k, \mathbf{v}_k, t_k) = -1 + O_{p_u} \left( \frac{\alpha_n^2}{t_k^2} \right)$$

and

$$\text{(S.83)} \quad \|t_k \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k)\| = \|\mathcal{P}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k)\| = O_{p_u} \left( \frac{\alpha_n^2}{t_k^2} \right).$$

By Theorem S.1 and Corollary S.1, we have the following inequalities

$$\text{(S.84)} \quad \mathcal{R}(\mathbf{e}_i, \mathbf{v}_k, t_k) \lesssim \frac{1}{\sqrt{n} |t_k|},$$

(S.85)

$$\|\mathbf{b}_{\mathbf{e}_i, k, t_k} - \mathbf{e}_i\| = \|\mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t) \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T\| = O_{p_u} \left( \frac{1}{\sqrt{n}} \right),$$

(S.86)

$$\|\mathbf{b}_{\mathbf{v}_k, k, t_k} - \mathbf{v}_k\| = \left\| \frac{\mathcal{P}(\mathbf{v}_k, \mathbf{V}_{-k}, t)}{t_k} \left( (\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathbf{V}_{-k}^T \right\| = O_{p_u} \left( \frac{\alpha_n^2}{t_k^2} \right),$$

$$\begin{aligned}
\text{(S.87)} \quad & \|A_{\mathbf{v}_k, k, t_k} + 1 + \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{t_k^2}\| \leq \|\mathcal{P}(\mathbf{v}_k, \mathbf{v}_k, t_k) + 1 + \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{t_k^2}\| \\
& + \|\mathcal{P}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k) \left( t(\mathbf{D}_{-k})^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k)\| \\
& = \left\| \sum_{l=3}^L \frac{1}{t_k^l} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k \right\| + O_{p_u} \left( \frac{\alpha_n^4}{t_k^4} \right) = O_{p_u} \left( \frac{\alpha_n^3}{|t_k|^3} \right),
\end{aligned}$$

$$\begin{aligned}
\text{(S.88)} \quad & t_k^2 \tilde{\mathcal{R}}'(\mathbf{v}_k, \mathbf{v}_k, t_k) = \left( 1 + \frac{2}{t_k^2} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + \sum_{l=3}^L \frac{l}{t_k^l} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k \right)^{-1} \\
& = \left( 1 + \frac{2}{t_k^2} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + O \left( \frac{\alpha_n^3}{|t_k|^3} \right) \right)^{-1} = 1 - \frac{2}{t_k^2} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + O \left( \frac{\alpha_n^3}{|t_k|^3} \right).
\end{aligned}$$

Similarly we have

$$\text{(S.89)} \quad t_k^2 \tilde{\mathcal{R}}'(\mathbf{v}_k, \mathbf{V}_{-k}, t_k) = \left( \sum_{l=2}^L \frac{l}{t_k^l} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k \right)^{-1} = O_{p_u} \left( \frac{\alpha_n^2}{|t_k|^2} \right).$$

By (83) and (A.16) of [1], we have

$$\text{(S.90)} \quad \left\| \left\{ [\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \right\}' \right\| = O_{p_u}(1),$$

and

$$\text{(S.91)} \quad \left\| [\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \right\| = O_{p_u}(|t_k|).$$

By (86) of [1], (S.82), (S.82), (S.88), (S.89), (S.90) and (S.91), we conclude that

$$\begin{aligned}
\text{(S.92)} \quad & \frac{1}{t_k^2 \tilde{\mathcal{P}}_{k, t_k}} = \left( \frac{A_{\mathbf{v}_k, k, t_k}}{t_k} \right)' \\
& = \mathcal{R}'(\mathbf{v}_k, \mathbf{v}_k, t_k) - 2\mathcal{R}'(\mathbf{v}_k, \mathbf{V}_{-k}, t_k) [\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \times \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k) \\
& \quad - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k) \left\{ [\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1} \right\}' \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k) \\
& = \frac{1}{t_k^2} - \frac{2}{t_k^4} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + O_{p_u}(\alpha_n^3/|t_k|^5).
\end{aligned}$$

Therefore

$$\text{(S.93)} \quad \tilde{\mathcal{P}}_{k, t_k} = 1 + \frac{2}{t_k^2} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + O_{p_u}(\alpha_n^3/|t_k|^3).$$

Recalling the definition of  $\mathbf{r}_k$  in Section 3.6, we have

$$\begin{aligned}
\text{(S.94)} \quad & \|A_{\mathbf{e}_i, k, t_k} + \mathbf{v}_k(i) + \frac{1}{t_k^2} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k + \mathbf{e}_i^T \mathbf{r}_k\| \\
& \leq \|\mathcal{P}(\mathbf{e}_i, \mathbf{v}_k, t_k) + \mathbf{v}_k(i) + \frac{1}{t_k^2} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k\|
\end{aligned}$$

$$\begin{aligned}
& + \|\mathcal{P}(\mathbf{e}_i, \mathbf{V}_{-k}, t_k) \left( t_k (\mathbf{D}_{-k})^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \right)^{-1} \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k) + \mathbf{e}_i^T \mathbf{r}_k\| \\
& = \left\| \sum_{l=3}^L \frac{1}{t_k^l} \mathbf{e}_i^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k \right\| + O_{p_u} \left( \frac{\alpha_n^3}{\sqrt{n} |t_k|^3} \right) = O_{p_u} \left( \frac{\alpha_n^3}{\sqrt{n} |t_k|^3} \right).
\end{aligned}$$

By Corollary S.1 and Theorem S.1 we have

$$|\mathbf{b}_{\mathbf{e}_i, k, t_k}^T (\mathbf{W} - \mathbb{E} \mathbf{W}) \mathbf{v}_k| + |\mathbf{v}_k^T (\mathbf{W} - \mathbb{E} \mathbf{W}) \mathbf{v}_k| = O_{p_u} \left( \frac{\alpha_n \log n}{\sqrt{n}} \right).$$

It follows from (S.85) and (S.86) that

$$(S.95) \quad \mathbb{E} \mathbf{b}_{\mathbf{e}_i, k, t_k}^T \mathbf{W} \mathbf{v}_k = O_{p_u} \left( \frac{\theta_n}{\sqrt{n}} \right),$$

$$(S.96) \quad \mathbb{E} \mathbf{b}_{\mathbf{v}_k, k, t_k}^T \mathbf{W} \mathbf{v}_k - \mathbb{E} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = O_{p_u} \left( \frac{\alpha_n^2 \theta_n}{t_k^2} \right).$$

By the expressions from (S.82)-(S.96) and Lemma 2, we have

$$\begin{aligned}
(S.97) \quad & \mathbf{e}_i^T \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T \mathbf{v}_k = \left[ \tilde{\mathcal{P}}_{k, t_k} - 2t_k^{-1} \tilde{\mathcal{P}}_{k, t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right) \right] \\
& \times \left[ A_{\mathbf{e}_i, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{e}_i, k, t_k}^T \mathbf{W} \mathbf{v}_k - \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k}{t_k^2} + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} \right) \right] \\
& \times \left[ A_{\mathbf{v}_k, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{v}_k, k, t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{(\alpha_n \log n)^2}{\sqrt{n} t_k^2} \right) \right] \\
& = \tilde{\mathcal{P}}_{k, t_k} A_{\mathbf{e}_i, k, t_k} A_{\mathbf{v}_k, k, t_k} + t_k^{-1} \mathbf{s}_{k, i}^T \mathbf{W} \mathbf{v}_k - \frac{\tilde{\mathcal{P}}_{k, t_k} A_{\mathbf{v}_k, k, t_k} \mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k}{t_k^2} \\
& + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} |t_k|^3} + \frac{(\alpha_n \log n)^2}{n |t_k|^2} \right). \\
& = \tilde{\mathcal{P}}_{k, t_k} A_{\mathbf{e}_i, k, t_k} A_{\mathbf{v}_k, k, t_k} + t_k^{-1} \mathbf{s}_{k, i}^T \mathbf{W} \mathbf{v}_k - \frac{\tilde{\mathcal{P}}_{k, t_k} A_{\mathbf{v}_k, k, t_k} \mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathbf{v}_k}{t_k^2} \\
& + O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} |t_k|^2} \right).
\end{aligned}$$

Choosing  $\mathbf{u} = \mathbf{v}_k$  in (S.67), similar to Lemma 8 in [2] we have

$$\begin{aligned}
(S.98) \quad & (\mathbf{v}_k^T \hat{\mathbf{v}}_k)^2 = \frac{\hat{d}_k^2 \mathbf{v}_k^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k \mathbf{v}_k^T \left[ \mathbf{G}(\hat{d}_k) - \mathbf{F}_k(\hat{d}_k) \right] \mathbf{v}_k}{\hat{d}_k^2 \mathbf{v}_k^T \left[ \mathbf{G}'(\hat{d}_k) - \mathbf{F}'_k(\hat{d}_k) \right] \mathbf{v}_k} \\
& = \left[ \tilde{\mathcal{P}}_{k, t_k} - 2t_k^{-1} \tilde{\mathcal{P}}_{k, t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} t_k^2} \right) \right] \\
& \quad \times \left[ A_{\mathbf{v}_k, k, t_k} - t_k^{-1} \mathbf{b}_{\mathbf{v}_k, k, t_k}^T \mathbf{W} \mathbf{v}_k + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} t_k^2} \right) \right]^2 \\
& = \tilde{\mathcal{P}}_{k, t_k} A_{\mathbf{v}_k, k, t_k}^2 + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} t_k^2} \right).
\end{aligned}$$

Since we fix the direction of  $\hat{\mathbf{v}}_k$  such that  $\mathbf{v}_k^T \hat{\mathbf{v}}_k \geq 0$ , we can obtain the expansion of  $\mathbf{v}_k^T \hat{\mathbf{v}}_k$  as follows

$$\mathbf{v}_k^T \hat{\mathbf{v}}_k = -\sqrt{\tilde{\mathcal{P}}_{k,t_k}} A_{\mathbf{v}_k, k, t_k} + O_{p_u}\left(\frac{\alpha_n^2}{\sqrt{n}t_k^2}\right).$$

Divide (S.97) by  $\mathbf{v}_k^T \hat{\mathbf{v}}_k$ . According to (S.87) and (S.93) we can expand the estimator  $\hat{\mathbf{v}}_k(i)$  to higher order as follows

$$\begin{aligned} \hat{\mathbf{v}}_k(i) &= \mathbf{v}_k(i) + \frac{\mathbf{r}_k(i)}{t_k^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k}{t_k^2} - \mathbf{v}_k(i) \frac{3\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{2t_k^2} + \frac{\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k}{t_k} \\ &\quad + O_{p_u}\left(\frac{(\alpha_n \log n)^3}{\sqrt{n}|t_k|^3} + \frac{(\alpha_n \log n)^2}{n|t_k|^2}\right) \\ \text{(S.99)} \quad &= \mathbf{v}_k(i) + \frac{\mathbf{r}_k(i)}{t_k^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathbf{v}_k}{t_k^2} - \mathbf{v}_k(i) \frac{3\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{2t_k^2} + \frac{\mathbf{s}_{k,i}^T \mathbf{W} \mathbf{v}_k}{t_k} + O_{p_u}\left(\frac{\alpha_n (\log n)^3}{\sqrt{n}d_k^2}\right), \end{aligned}$$

where we have used the inequality that  $\frac{\alpha_n^2}{(\log \log n)^{c'}} \lesssim |d_k|$  by Condition 3. This completes the proof of (S.80).

Now we focus on the proof of (S.81). From (S.58) we have

$$\hat{d}_k = t_k + \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u}\left(\frac{1}{\sqrt{n}}\right).$$

Combining with the definition of  $t_k$  and (S.100)

$$1 + d_k (\mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z_0) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z_0) (\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z))^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z_0)) = O_{p_u}\left(\frac{\alpha_n^3}{d_k^3}\right),$$

$z_0 = d_k + \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k}$ , we conclude that

$$t_k = z_0 + O_{p_u}\left(\frac{\alpha_n}{|d_k|}\right).$$

Hence we have

$$\hat{d}_k - d_k = \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k}{d_k} + \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_{p_u}\left(\frac{1}{\sqrt{n}} + \frac{\alpha_n}{|d_k|}\right).$$

This proves (S.81), and thus concludes the proof of the lemma.  $\square$

8.3. *Lemmas with multiplicity and their proofs.* In the following sections we will consider the case with eigenvalue multiplicity. We first introduce some notations that will be used frequently hereafter. Define

$$\mathcal{V}_j = (\mathbf{v}_{K_{j-1}+1}, \dots, \mathbf{v}_{K_j}), \quad j = 1, \dots, \tilde{\mathcal{K}},$$

$$\mathbf{A}_{\mathbf{x}, j, t} = \mathcal{P}(\mathbf{x}, \mathcal{V}_j, t) - \mathcal{P}(\mathbf{x}, \mathcal{V}_{-j}, t) [t(\mathbf{D}_{-j})^{-1} + \mathcal{P}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, t)]^{-1} \mathcal{P}(\mathcal{V}_{-j}, \mathcal{V}_j, t),$$

$$\mathbf{B}_{\mathbf{M}, j, t}^T = \mathbf{M}^T - \mathcal{R}(\mathbf{M}, \mathcal{V}_{-j}, t) \left( (\mathbf{D}_{-j})^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, t) \right)^{-1} \mathcal{V}_{-j}^T,$$

$$\mathbf{S}_{j,i} = \mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j} - \mathcal{V}_j \mathcal{V}_j^T \mathbf{e}_i, \quad \mathbf{S}_j = \sum_{i=1}^n \mathbf{S}_{j,i}, \quad \mathbf{S}_j(i) = \mathbf{e}_i^T \mathbf{S}_j,$$

$$\text{and } \mathbf{R}_j = \mathcal{V}_{-j} (\tilde{d}_j \mathbf{D}_{-j}^{-1} - \mathbf{I})^{-1} \mathcal{V}_{-j}^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j,$$

where  $\mathbf{M}$  is some matrix or vector with appropriate dimensions, and we slightly abuse the notation by using  $\mathbf{D}_{-j}$  to denote the submatrix of  $\mathbf{D}$  by removing all diagonal entries with the same value  $\tilde{d}_j$ . Moreover,  $\mathcal{V}_j$  is the matrix of spiked eigenvectors with the same eigenvalue  $\tilde{d}_j$ , and  $\mathcal{V}_{-j}$  is the matrix of spiked eigenvectors whose eigenvalues are different from  $\tilde{d}_j$ ,  $j = 1, \dots, \mathcal{K}$ .

Now we are ready to proceed to the key lemmas regarding the asymptotic expansion of eigenvectors and eigenvalues with multiplicity as well as their proofs.

### 8.3.1. Lemma 4 and its proof.

LEMMA 4. *Under Conditions 1-4 and 6, we have the following expansion*

$$(S.101) \quad \begin{aligned} \mathbf{e}_i^T \widehat{\mathcal{V}}_j \widehat{\mathcal{V}}_j^T \mathcal{V}_j &= \mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} + \frac{\mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathcal{V}_j^T \mathbf{W} \mathbf{B}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} - \frac{\mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{W} \mathcal{V}_j \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} \\ &+ O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} \tilde{d}_j^2} \right) + \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{\tilde{d}_j^2} - \frac{3 \mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} + \frac{\mathbf{e}_i^T \mathbf{R}_j}{\tilde{d}_j^2}. \end{aligned}$$

PROOF. There are four major steps in our proof. First, we write  $\mathbf{e}_i^T \widehat{\mathcal{V}}_j \widehat{\mathcal{V}}_j^T \mathcal{V}_j$  as an integral which is a functional of  $\mathbf{X} = \mathbf{H} + \mathbf{W}$ . By doing so we can deal with the matrix  $\mathbf{H} + \mathbf{W}$  instead of the eigenvectors. Second, for the functional of  $\mathbf{H} + \mathbf{W}$  obtained in the previous step we extract the  $\mathbf{H}$  part from  $\mathbf{H} + \mathbf{W}$  and further obtain a functional of  $\mathbf{W}$ . Roughly speaking, we can get an explicit function of form  $f((\mathbf{W} - \tilde{d}_j \mathbf{I})^{-1})$ . Third, by the matrix series expansion  $(\mathbf{W} - \tilde{d}_j \mathbf{I})^{-1} = -\sum_{l=0}^{\infty} \tilde{d}_j^{-(l+1)} \mathbf{W}^l$ , the function  $f((\mathbf{W} - \tilde{d}_j \mathbf{I})^{-1})$  can be approximated by  $f(-\sum_{l=0}^L \tilde{d}_j^{-(l+1)} \mathbf{W}^l)$  for some positive integer  $L$ . Fourth, we can then calculate the first (second or higher) order expansion of  $f(-\sum_{l=0}^L \tilde{d}_j^{-(l+1)} \mathbf{W}^l)$  since we have an explicit expression of the function  $f$ .

In various parts of this proof, we will use the Sherman–Morrison–Woodbury formula which we give a brief review here. For any matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{F}$  of appropriate dimensions, it holds that

$$(S.102) \quad (\mathbf{A} + \mathbf{BFC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{F}^{-1} + \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1},$$

when the corresponding matrices for matrix inversion are nonsingular.

Let  $\Omega$  be a contour centered at  $\tilde{d}_j$  with radius  $\min \left\{ |\tilde{d}_j| - |\tilde{d}_{j-1}|, |\tilde{d}_j| - |\tilde{d}_{j+1}| \right\} / 2$ . Since we consider  $z$  on the contour  $\Omega$  enclosing  $\tilde{d}_j$ , it follows that  $|z| \geq c |\tilde{d}_j|$  for some positive constant  $c$  by Condition 2. Similar to the argument of (S.59), let  $L = \log n$ . Thus, with probability tending to 1, we have

$$(S.103) \quad \left\| \sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{W}^l \right\| \leq \sum_{l=L+1}^{\infty} \frac{(C \log n)^{l/2} \alpha_n^l}{|z|^{l+1}} \leq \frac{1}{\sqrt{n} |z|^4},$$

where  $C > 0$  is a constant.

By Theorem 6.2 of [4], Condition 3, Corollary S.1, and Weyl's inequality we have

$$(S.104) \quad |\hat{d}_k - \tilde{d}_j| \leq \|\mathbf{W}\| = O(\alpha_n \sqrt{\log n}), \quad k \in \{K_{j-1} + 1, \dots, K_j\},$$

$$|\hat{d}_k - \tilde{d}_j| \geq |d_k - \tilde{d}_j| - |\hat{d}_k - d_k| \gg \alpha_n \sqrt{\log n}, \quad k \notin \{K_{j-1} + 1, \dots, K_j\}.$$

with high probability. In other words, the contour  $\Omega$  encloses only eigenvalues  $\{\hat{d}_k, k = K_{j-1} + 1, \dots, K_j\}$  with probability tending to 1. By Cauchy's residue theorem we have with asymptotic probability one,

$$-\frac{1}{2\pi i} \oint_{\Omega} \frac{1}{\hat{d}_k - z} dz = \begin{cases} 1, & \text{if } k \in \{K_{j-1} + 1, \dots, K_j\} \\ 0, & \text{otherwise.} \end{cases}$$

Note that  $(\mathbf{X} - z\mathbf{I})^{-1} = \sum_{i=1}^n (\hat{d}_i - z)^{-1} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^T$ . Thus, it follows from Cauchy's residue theorem that for  $k \in \{K_{j-1} + 1, \dots, K_j\}$ , with asymptotic probability one,

$$\begin{aligned} \mathbf{e}_i^T \hat{\mathcal{V}}_j \hat{\mathcal{V}}_j^T \mathcal{V}_j &= \sum_{k=K_{j-1}+1}^{K_j} \mathbf{e}_i^T \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T \mathcal{V}_j = -\frac{1}{2\pi i} \oint_{\Omega} \mathbf{e}_i^T \sum_{k=1}^n \frac{\hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T}{\hat{d}_k - z} \mathcal{V}_j dz \\ &= -\frac{1}{2\pi i} \oint_{\Omega} \mathbf{e}_i^T (\mathbf{X} - z\mathbf{I})^{-1} \mathcal{V}_j dz. \end{aligned}$$

Since  $\mathbf{X} - z\mathbf{I} = \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T + \tilde{\mathbf{d}}_j \mathcal{V}_j \mathcal{V}_j^T$ , it follows from (S.102) that the above bilinear form can be further written as

$$\begin{aligned} \mathbf{e}_i^T \hat{\mathcal{V}}_j \hat{\mathcal{V}}_j^T \mathcal{V}_j &= -\frac{1}{2\pi i} \oint_{\Omega} \mathbf{e}_i^T \left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T + \tilde{\mathbf{d}}_j \mathcal{V}_j \mathcal{V}_j^T \right)^{-1} \mathcal{V}_j dz \\ &= \frac{1}{2\pi i} \oint_{\Omega} \tilde{\mathbf{d}}_j \mathbf{e}_i^T \left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T \right)^{-1} \mathcal{V}_j \left\{ \mathbf{I} + \tilde{\mathbf{d}}_j \mathcal{V}_j^T \left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T \right)^{-1} \mathcal{V}_j \right\}^{-1} \\ &\quad \times \mathcal{V}_j^T \left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T \right)^{-1} \mathcal{V}_j dz, \end{aligned} \tag{S.105}$$

where in the last step we have used  $\oint_{\Omega} \mathbf{e}_i^T \left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T \right)^{-1} \mathcal{V}_j dz = 0$  by the fact that  $(\mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T)^{-1}$  is analytic inside the circle  $\Omega$  (following from Condition 2).

Another application of (S.102) leads to

$$\left( \mathbf{W} - z\mathbf{I} + \mathcal{V}_{-j} \mathbf{D}_{-j} \mathcal{V}_{-j}^T \right)^{-1} = \mathbf{G}(z) - \mathbf{G}(z) \mathcal{V}_{-j} \left[ \mathbf{D}_{-j}^{-1} + \mathcal{V}_{-j}^T \mathbf{G}(z) \mathcal{V}_{-j} \right]^{-1} \mathcal{V}_{-j}^T \mathbf{G}(z), \tag{S.106}$$

where the Green function  $\mathbf{G}(z)$  associated with only the noise part  $\mathbf{W}$  is defined in (S.68). To simplify the expression, let

$$\mathbf{F}(z) = \mathbf{G}(z) \mathcal{V}_{-j} \left[ \mathbf{D}_{-j}^{-1} + \mathcal{V}_{-j}^T \mathbf{G}(z) \mathcal{V}_{-j} \right]^{-1} \mathcal{V}_{-j}^T \mathbf{G}(z). \tag{S.107}$$

Then in view of (S.106), the last line integral in (S.105) can be further represented as

$$\begin{aligned} \mathbf{e}_i^T \hat{\mathcal{V}}_j \hat{\mathcal{V}}_j^T \mathcal{V}_j &= \frac{1}{2\pi i} \oint_{\Omega} \tilde{\mathbf{d}}_j \mathbf{e}_i^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j \\ &\quad \times \left( \mathbf{I} + \tilde{\mathbf{d}}_j \mathcal{V}_j^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j \right)^{-1} \mathcal{V}_j^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j dz. \end{aligned} \tag{S.108}$$

It is challenging to analyze the terms in (S.108) directly since the expression of  $\mathbf{F}(z)$  is complicated and we need to study the asymptotic expansion of  $\mathbf{F}(z)$  carefully. We will use the truncation idea. Recall  $L = \log n$ . Similar to (S.60) we have  $\|\sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{e}_i^T \mathbf{W}^l \mathcal{V}_j\| = O_{p_u}(n^{-1/2} \alpha_n |z|^{-4})$  for  $z$  on the contour  $\Omega$ . Similar to the proof of (S.61)-(S.66), it follows from Corollary S.1 and Conditions 3-4 that

$$\left\| \sum_{l=s}^L z^{-(l+1)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E} \mathbf{W}^l) \mathcal{V}_j \right\| = n^{-1/2} O_{p_u}((\alpha_n \log n)^s |z|^{-(s+1)}).$$



Moreover, since for  $z \in \Omega$  we have  $|z|^{-(s+2)} \leq \alpha_n |z|^{-(s+1)}$  by Condition 3. Similar to (S.70) we can further obtain

$$\begin{aligned}
\mathbf{x}^T \mathbf{G}(z) \mathcal{V}_j &= -z^{-1} \mathbf{x}^T \mathcal{V}_j - z^{-2} \mathbf{x}^T \mathbf{W} \mathcal{V}_j - z^{-3} \mathbf{x}^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j \\
&\quad - \sum_{l=2}^L z^{-(l+1)} \mathbf{x}^T \mathbb{E} \mathbf{W}^l \mathcal{V}_j + O_{p_u}(n^{-1/2}(\alpha_n \log n)^3 |z|^{-4}) \\
\text{(S.109)} \quad &= -z^{-1} \mathbf{x}^T \mathcal{V}_j - z^{-2} \mathbf{x}^T \mathbf{W} \mathcal{V}_j - \sum_{l=2}^L z^{-(l+1)} \mathbf{x}^T \mathbb{E} \mathbf{W}^l \mathcal{V}_j + O_{p_u}(n^{-1/2}(\alpha_n \log n)^2 |z|^{-3}).
\end{aligned}$$

In fact, the probabilistic event associated with the small order term  $O_{p_u}(n^{-1/2}(\alpha_n \log n)^2 |z|^{-3})$  in (S.109) holds uniformly over  $z$  since the term  $O_{p_u}(n^{-1/2}(\alpha_n \log n)^2 |z|^{-3})$  is simply  $n^{-1/2} |z|^{-3} O_{p_u}((\alpha_n \log n)^2)$ .

Using similar idea as for proving (S.77), it follows from (S.107) and (S.109) that

$$\begin{aligned}
\mathbf{e}_j^T \mathbf{F}(z) \mathcal{V}_j &= \mathcal{R}(\mathbf{e}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_j, z) \\
&\quad - z^{-2} \mathcal{R}(\mathbf{e}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{V}_{-j}^T \mathbf{W} \mathcal{V}_j \\
&\quad - z^{-2} \mathbf{e}_j^T \mathbf{W} \mathcal{V}_{-j} \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_j, z) \\
&\quad + z^{-2} \mathcal{R}(\mathbf{e}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{V}_{-j}^T \mathbf{W} \mathcal{V}_{-j} \\
&\quad \times \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}, z) \\
&\quad + O_{p_u}(n^{-1/2}(\alpha_n \log n)^3 \tilde{\mathbf{d}}_j^{-4} + n^{-1}(\alpha_n \log n)^2 |\tilde{\mathbf{d}}_j|^{-3}), \\
\text{(S.110)} \quad &= \mathcal{R}(\mathbf{e}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_j, z) \\
&\quad - z^{-2} \mathcal{R}(\mathbf{e}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{V}_{-j}^T \mathbf{W} \mathcal{V}_j \\
&\quad + O_{p_u}(n^{-1/2}(\alpha_n \log n)^3 |\tilde{\mathbf{d}}_j|^{-4} + n^{-1}(\alpha_n \log n)^2 |\tilde{\mathbf{d}}_j|^{-3})
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{V}_j^T \mathbf{F}(z) \mathcal{V}_j &= \mathcal{V}_j^T \mathbf{G}(z) \mathcal{V}_{-j} \left[ \mathbf{D}_{-j}^{-1} + \mathcal{V}_{-j}^T \mathbf{G}(z) \mathcal{V}_{-j} \right]^{-1} \mathcal{V}_{-j}^T \mathbf{G}(z) \mathcal{V}_j \\
&= \mathcal{R}(\mathcal{V}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_j, z) \\
&\quad - 2z^{-2} \mathcal{R}(\mathcal{V}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{V}_{-j}^T \mathbf{W} \mathcal{V}_j + O_{p_u}(n^{-1/2} \alpha_n^3 \tilde{\mathbf{d}}_j^{-4}) \\
&= \mathcal{R}(\mathcal{V}_j, \mathcal{V}_{-j}, z) \left[ \mathbf{D}_{-j}^{-1} + \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_{-j}, z) \right]^{-1} \mathcal{R}(\mathcal{V}_{-j}, \mathcal{V}_j, z) \\
\text{(S.111)} \quad &+ O_{p_u}(n^{-1/2}(\alpha_n \log n)^3 \tilde{\mathbf{d}}_j^{-4}),
\end{aligned}$$

where  $\mathbf{F}(z)$  is defined in (S.107) and  $\mathcal{R}$  is defined in (19).

Now we are ready to consider the right hand side of (S.108). By (S.109)–(S.111) we have

$$(S.112) \quad \mathbf{e}_i^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j = \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} - \frac{\mathbf{B}_{\mathbf{e}_i, j, z}^T \mathbf{W} \mathcal{V}_j}{z^2} - \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{z^3} \\ + O_{p_u} \left( \frac{(\alpha_n \log n)^3}{\sqrt{n} \tilde{d}_j^4} + \frac{(\alpha_n \log n)^2}{n |\tilde{d}_j|^3} \right),$$

$$(S.113) \quad \mathcal{V}_j^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j = \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} - \frac{\mathcal{V}_j^T \mathbf{W} \mathbf{B}_{\mathcal{V}_j, j, z}}{z^2} + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} |\tilde{d}_j|^3} \right),$$

and

$$(S.114) \quad (\mathbf{I} + \tilde{d}_j \mathcal{V}_j^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j)^{-1} \\ = (1 - \frac{\tilde{d}_j}{z})^{-1} \mathbf{I} + (1 - \frac{\tilde{d}_j}{z})^{-2} \left( \frac{\tilde{d}_j \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j}{z^2} + \frac{\tilde{d}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{z^3} \right) + O_{p_u} \left( \frac{\alpha_n^4}{\tilde{d}_j^4} \right).$$

Plugging the expansion (S.114) into (S.108) will result in two main terms and a smaller order term, where the first term can be further written as

$$(S.115) \quad \frac{1}{2\pi i} \oint_{\Omega} \tilde{d}_j (1 - \frac{\tilde{d}_j}{z})^{-1} \mathbf{e}_i^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j \mathcal{V}_j^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j dz \\ = \mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{\mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathcal{V}_j^T \mathbf{W} \mathbf{B}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} - \frac{\mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{W} \mathcal{V}_j \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} \\ + \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{\tilde{d}_j^2} + \frac{\mathbf{e}_i^T \mathbf{R}_j}{\tilde{d}_j^2} + O_{p_u} \left( \frac{(\alpha_n (\log n))^3}{\sqrt{n} \tilde{d}_j^2} \right).$$

For the second term, by the equation that  $\frac{d_i}{z(z-d_i)} = \frac{1}{z-d_i} - \frac{1}{z}$ , we have

$$(S.116) \quad \frac{1}{2\pi i} \oint_{\Omega} \tilde{d}_j (1 - \frac{\tilde{d}_j}{z})^{-2} \mathbf{e}_i^T [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j \left( \frac{\tilde{d}_j \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j}{z^2} + \frac{\tilde{d}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{z^3} \right) \mathcal{V}_j^T \\ \times [\mathbf{G}(z) - \mathbf{F}(z)] \mathcal{V}_j dz \\ = \tilde{d}_j^2 \left( \left( \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \right)^T \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} + \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \left( \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} \right)' \right) \Big|_{z=\tilde{d}_j} \\ - \frac{3\mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} + O_{p_u} \left( \frac{(\alpha_n \log n)^4}{\sqrt{n} \tilde{d}_j^4} \right).$$

Combining the above two results, the right hand side of (S.108) can be further written as

$$\mathbf{e}_i^T \widehat{\mathcal{V}}_j \widehat{\mathcal{V}}_j^T \mathcal{V}_j = \mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{\mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathcal{V}_j^T \mathbf{W} \mathbf{B}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} - \frac{\mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{W} \mathcal{V}_j \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} + \frac{\mathbf{e}_i^T \mathbf{R}_j}{\tilde{d}_j^2} \\ + O_{p_u} \left( \frac{(\alpha_n (\log n))^3}{\sqrt{n} \tilde{d}_j^2} \right) + \tilde{d}_j^2 \left( \left( \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \right)^T \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} + \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \left( \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} \right)' \right) \Big|_{z=\tilde{d}_j} \\ - \frac{3\mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} + \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{\tilde{d}_j^2}$$

$$\begin{aligned}
&= \mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} + \frac{\mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathcal{V}_j^T \mathbf{W} \mathbf{B} \mathcal{V}_j}{\tilde{d}_j} - \frac{\mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{W} \mathcal{V}_j \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} + \frac{\mathbf{e}_i^T \mathbf{R}_j}{\tilde{d}_j^2} \\
&+ O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} \tilde{d}_j^2} \right) + \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{\tilde{d}_j^2} - \frac{3 \mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2},
\end{aligned} \tag{S.117}$$

where  $\tilde{d}_j^2 \left( \left( \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \right)^T \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} + \frac{\mathbf{A}_{\mathbf{e}_i, j, z}}{z} \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j \left( \frac{\mathbf{A}_{\mathcal{V}_j, j, z}}{z} \right)' \right) \Big|_{z=\tilde{d}_j} = -\frac{2 \mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbf{W} \mathcal{V}_j}{\tilde{d}_j} + O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} \tilde{d}_j^2} \right)$ . Therefore, we complete our proof.  $\square$

### 8.3.2. Lemma 5 and its proof.

LEMMA 5. Let  $\mathcal{V}_j = (\mathbf{v}_{K_{j-1}+1}, \dots, \mathbf{v}_{K_j})$ . Under Conditions 1-4 and 6, There exists an orthogonal matrix  $\mathbf{U}_n$  such that

$$\widehat{\mathcal{V}}_j(i) \mathbf{U}_n = \mathcal{V}_j(i) + \frac{\mathbf{R}_j(i)}{\tilde{d}_j^2} + \frac{\mathbf{e}_i^T \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} - \mathcal{V}_j(i) \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{2 \tilde{d}_j^2} + \frac{\mathbf{S}_{j,i}^T \mathbf{W} \mathcal{V}_j}{\tilde{d}_j} + O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} \tilde{d}_j^2} \right), \tag{S.118}$$

where  $\mathbf{R}_j$  and  $\mathbf{S}_{j,i}$  are defined at the beginning of Section 8.3. Moreover,

$$\hat{d}_k - \tilde{d}_j = O \left( \frac{\alpha_n^2}{|\tilde{d}_j|} \right) + O_{p_u}(\theta) + O_{p_u} \left( \frac{1}{\sqrt{n}} + \frac{\alpha_n}{|\tilde{d}_j|} \right), \quad k \in \{K_{j-1} + 1, \dots, K_j\}. \tag{S.119}$$

PROOF. Replacing  $\mathbf{e}_i$  by  $\mathcal{V}_j$ , following the same lines in the proof of Lemma 4 we have

$$\mathcal{V}_j^T \widehat{\mathcal{V}}_j \widehat{\mathcal{V}}_j^T \mathcal{V}_j = \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j^2} \right). \tag{S.120}$$

By Theorem S.1 and Corollary S.1, we have  $\| \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} - \mathbf{I} \|_\infty = o(1)$ , which means that  $\mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2}$  is invertible positive definite matrix for large  $n$ . Define a matrix (maybe random)  $\tilde{\mathbf{U}}_n$  as

$$\tilde{\mathbf{U}}_n = \left( \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} \right)^{-1/2} \mathcal{V}_j^T \widehat{\mathcal{V}}_j.$$

By (S.120), we have

$$\tilde{\mathbf{U}}_n \tilde{\mathbf{U}}_n^T = \mathbf{I} + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j^2} \right).$$

Therefore, there exists an orthogonal matrix  $\mathbf{U}_n$  such that

$$\| \mathbf{U}_n - \tilde{\mathbf{U}}_n \| = O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j^2} \right).$$

Combining the above three results, we obtain that

$$\mathcal{V}_j^T \widehat{\mathcal{V}}_j = \left( \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} \right)^{1/2} \mathbf{U}_n + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j^2} \right).$$

This together with Lemma 4 ensures that

$$\begin{aligned}
\mathbf{e}_i^T \widehat{\mathcal{V}}_j &= \left[ \mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} + \frac{\mathbf{A}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathcal{V}_j^T \mathbf{W} \mathbf{B}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} - \frac{\mathbf{B}_{\mathbf{e}_i, j, \tilde{d}_j}^T \mathbf{W} \mathcal{V}_j \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}}{\tilde{d}_j} \right. \\
&\quad \left. + O_{p_u} \left( \frac{\alpha_n (\log n)^3}{\sqrt{n} \tilde{d}_j^2} \right) + \frac{\mathbf{e}_i^T (\mathbf{W}^2 - \mathbb{E} \mathbf{W}^2) \mathcal{V}_j}{\tilde{d}_j^2} - \frac{3 \mathbf{e}_i^T \mathcal{V}_j \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} + \frac{\mathbf{e}_i^T \mathbf{R}_j}{\tilde{d}_j^2} \right] \\
\text{(S.121)} \quad &\times \left[ \left( \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j}^T \mathbf{A}_{\mathcal{V}_j, j, \tilde{d}_j} - \frac{3 \mathcal{V}_j^T \mathbb{E} \mathbf{W}^2 \mathcal{V}_j}{\tilde{d}_j^2} \right)^{-1/2} \mathbf{U}_n + O_{p_u} \left( \frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j^2} \right) \right],
\end{aligned}$$

which ensures (S.118) by trivial calculation.

Now we prove (S.119). By Condition 3 and (S.104),  $(\mathbf{W} - \hat{d}_k \mathbf{I})$  is invertible with high probability and is in fact equal to  $\mathbf{G}^{-1}(\hat{d}_k)$ . It follows from the definition of the eigenvalue and the representation  $\mathbf{X} = \mathbf{H} + \mathbf{W} = \mathbf{V} \mathbf{D} \mathbf{V}^T + \mathbf{W}$  that

$$\begin{aligned}
0 &= \det(\mathbf{X} - \hat{d}_k \mathbf{I}) = \det(\mathbf{W} - \hat{d}_k \mathbf{I} + \mathbf{V} \mathbf{D} \mathbf{V}^T) = \det[\mathbf{G}^{-1}(\hat{d}_k) + \mathbf{V} \mathbf{D} \mathbf{V}^T] \\
&= \det[\mathbf{G}^{-1}(\hat{d}_k)] \det[\mathbf{I} + \mathbf{G}(\hat{d}_k) \mathbf{V} \mathbf{D} \mathbf{V}^T].
\end{aligned}$$

Therefore  $\det[\mathbf{I} + \mathbf{G}(\hat{d}_k) \mathbf{V} \mathbf{D} \mathbf{V}^T] = 0$ . Using the fact  $\det(\mathbf{I} + \mathbf{A} \mathbf{B}) = \det(\mathbf{I} + \mathbf{B} \mathbf{A})$  for matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we obtain for each  $1 \leq k \leq K$ ,

$$\text{(S.122)} \quad 0 = \det[\mathbf{I} + \mathbf{G}(\hat{d}_k) \mathbf{V} \mathbf{D} \mathbf{V}^T] = \det[\mathbf{I} + \mathbf{D} \mathbf{V}^T \mathbf{G}(\hat{d}_k) \mathbf{V}],$$

where the second  $\mathbf{I}$  represents an identity matrix of size  $K$  and we slightly abuse the notation for simplicity. Since the diagonal matrix  $\mathbf{D}$  is nonsingular by assumption, it follows from (S.122) that

$$\text{(S.123)} \quad \det[\tilde{d}_j \mathbf{V}^T \mathbf{G}(\hat{d}_k) \mathbf{V} + \tilde{d}_j \mathbf{D}^{-1}] = \tilde{d}_j \det(\mathbf{D}^{-1}) \det[\mathbf{I} + \mathbf{D} \mathbf{V}^T \mathbf{G}(\hat{d}_k) \mathbf{V}] = 0$$

for each  $1 \leq k \leq K$ . We next show that  $\hat{d}_k$ , which is a solution to (S.123) is asymptotically close to its population counterpart  $\tilde{d}_j$ . We will also provide its asymptotic expansion.

By (S.104) and (S.74), for  $k \in \{K_{j-1} + 1, \dots, K_j\}$ , we have  $\tilde{d}_j \mathbf{v}_{\ell_1}^T \mathbf{G}(\hat{d}_k) \mathbf{v}_{\ell_2} = -\tilde{d}_j O_{p_u}(\hat{d}_k^{-2}) = O_{p_u}(1/|\tilde{d}_j|)$  when  $\ell_1 \neq \ell_2$ . Thus, we can see that all off diagonal entries of matrix  $\tilde{d}_j \mathbf{V}^T \mathbf{G}(\hat{d}_k) \mathbf{V} + \tilde{d}_j \mathbf{D}^{-1}$  in (S.123) are of order  $O_{p_u}(1/|\tilde{d}_j|)$ . For  $\ell \notin \{K_{j-1} + 1, \dots, K_j\}$ , the  $\ell$ -th diagonal entry of  $\tilde{d}_j \mathbf{V}^T \mathbf{G}(\hat{d}_k) \mathbf{V} + \tilde{d}_j \mathbf{D}^{-1}$  equals  $\tilde{d}_j \mathbf{v}_{\ell}^T \mathbf{G}(\hat{d}_k) \mathbf{v}_{\ell} + \tilde{d}_j/d_{\ell}$ . By (S.73), Corollary S.1 and Theorem S.1, we have  $\tilde{d}_j \mathbf{v}_{\ell}^T \mathbf{G}(\hat{d}_k) \mathbf{v}_{\ell} + 1 = o(1) + O_{p_u}(\frac{\alpha_n}{\sqrt{n}|\tilde{d}_j|})$ ,  $k \in \{K_{j-1} + 1, \dots, K_j\}$ ,  $\ell \in \{1, \dots, K\}$ . Moreover, by Condition 3,  $|\tilde{d}_j/\tilde{d}_{\ell} - 1| \geq c$  for some positive constant  $c$ ,  $j \neq \ell$ . Hence, all these diagonal entries except the ones with diagonal indices in  $\{K_{j-1} + 1, \dots, K_j\}$  are of order at least  $O_p(1)$  uniformly. Thus the matrix  $(\tilde{d}_j \mathbf{v}_{\ell_1}^T \mathbf{G}(\hat{d}_k) \mathbf{v}_{\ell_2} + \delta_{\ell_1 \ell_2} \tilde{d}_j/d_{\ell_1})_{\ell_1, \ell_2 \notin \{K_{j-1} + 1, \dots, K_j\}}$  is invertible with significant probability, where  $\delta_{\ell_1 \ell_2} = 1$  when  $\ell_1 = \ell_2$  and 0 otherwise. Recall the determinant identity for block matrices from linear algebra

$$\det \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \det(\mathbf{A}_{22}) \det(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})$$

when the lower right block matrix  $\mathbf{A}_{22}$  is nonsingular. Treating the sub matrix  $(\tilde{d}_j \mathbf{v}_{\ell_1}^T \mathbf{G}(z) \mathbf{v}_{\ell_2} + \delta_{\ell_1 \ell_2} \tilde{d}_j/d_{\ell_1})_{\ell_1, \ell_2 \in \{K_{j-1} + 1, \dots, K_j\}}$  as the first block  $\mathbf{A}_{11}$  and  $(\tilde{d}_j \mathbf{v}_{\ell_1}^T \mathbf{G}(z) \mathbf{v}_{\ell_2} + \delta_{\ell_1 \ell_2} \tilde{d}_j/d_{\ell_1})_{\ell_1, \ell_2 \notin \{K_{j-1} + 1, \dots, K_j\}}$  as the second block  $\mathbf{A}_{22}$ , summarizing the arguments above, we have with high probability that

$$\text{(S.124)} \quad \det[\tilde{d}_j \mathbf{V}^T \mathbf{G}(z) \mathbf{V} + \tilde{d}_j \mathbf{D}^{-1}] = \det(\mathbf{A}_{22}) \det(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}),$$

for  $z \in [\tilde{d}_j - 2\Delta, \tilde{d}_j + 2\Delta]$ , where  $\Delta = \log n (\|\frac{\mathbf{V}^T \mathbb{E} \mathbf{W}^2 \mathbf{V}}{\tilde{d}_j}\|_\infty + \|\mathbf{V}^T \mathbf{W} \mathbf{V}\|_\infty) + O_{p_u}(\frac{\alpha_n^2}{\sqrt{n} \tilde{d}_j})$ .

Here  $\Delta$  is chosen to ensure the invertibility of  $\mathbf{A}_{22}$  and we will show that the solution  $\hat{d}_k$  belongs to the interval  $[\tilde{d}_j - 2\Delta, \tilde{d}_j + 2\Delta]$ . It is easy to see that for all  $z \in [\tilde{d}_j - \alpha_n \log n, \tilde{d}_j + \alpha_n \log n]$ , the sign of  $\det(\mathbf{A}_{22})$  keeps the same with high probability. Moreover, with high probability we have

$$\text{sign}(\det(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})) = (-1)^{K_j}, z \in [\tilde{d}_j - \alpha_n \log n, \tilde{d}_j - 2\Delta],$$

and

$$\text{sign}(\det(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})) = 1, z \in [\tilde{d}_j + 2\Delta, \tilde{d}_j + \alpha_n \log n].$$

In view of (S.124), the sign of the determinant  $\det[d_k \mathbf{V}^T \mathbf{G}(z) \mathbf{V} + d_k \mathbf{D}^{-1}]$  does not change in  $[\tilde{d}_j - \alpha_n \log n, \tilde{d}_j - 2\Delta]$  and  $[\tilde{d}_j + 2\Delta, \tilde{d}_j + \alpha_n \log n]$  with high probability. Hence,  $\hat{d}_k \notin [\tilde{d}_j - \alpha_n \log n, \tilde{d}_j - 2\Delta] \cup [\tilde{d}_j + 2\Delta, \tilde{d}_j + \alpha_n \log n]$ . Finally, by Theorem 6.2 of [4],  $\hat{d}_k$  must be in the interval  $(\tilde{d}_j - \alpha_n \log n, \tilde{d}_j + \alpha_n \log n)$  with high probability. The proof is completed.  $\square$

#### 8.4. Theorems 3.1–3.4 under multiplicity.

**THEOREM S.2.** *Under Conditions 1-6, the conclusions in Theorems 3.1–3.4 hold under their corresponding assumptions.*

**PROOF.** The proof of Theorems 3.1–3.4 in the case of no-multiplicity is essentially based on Theorem S.1, Corollary S.1, Lemmas 2 and 3. Lemmas 4-5 are similar to Lemmas 2-3 except we do not have the explicit asymptotic expansion of  $\hat{d}_k$  and the denominators in the expansions of eigenvectors are polynomials of  $\tilde{d}_j$ . For the denominators, the difference is negligible by (S.119). Moreover, the explicit asymptotic expansion of  $\hat{d}_k$  does not play an important role in the proof of Theorems 3.1–3.4, while we only need the order of the difference between  $\hat{d}_k$  and  $\tilde{d}_j$ . Finally, whether the eigenvalues are distinct or not does not affect Theorem S.1 and Corollary S.1. Theorem S.2 is proved by combining the arguments above.  $\square$

## REFERENCES

- [1] FAN, J., FAN, Y., HAN, X. and LV, J. (2020). Asymptotic Theory of Eigenvectors for Random Matrices with Diverging Spikes\*. *Journal of the American Statistical Association* **0** 1-63. <https://doi.org/10.1080/01621459.2020.1840990>
- [2] FAN, J., FAN, Y., HAN, X. and LV, J. (2021). SIMPLE: Statistical Inference on Membership Profiles in Large Networks. *arXiv preprint arXiv:1910.01734*.
- [3] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2020). Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association* 1–13.
- [4] TROPP, J. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434.