UNIVERSAL RANK INFERENCE VIA RESIDUAL SUBSAMPLING WITH APPLICATION TO LARGE NETWORKS

By Xiao Han[1,a], Qing Yang[1,b] and Yingying Fan[2,c]

[1]*International Institute of Finance, School of Management, University of Science and Technology of China,*
[a]*xhan011@ustc.edu.cn*, [b]*yangq@ustc.edu.cn*

[2]*Data Sciences and Operations Department, University of Southern California,* [c]*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 (*J. Amer. Statist. Assoc.* **117** (2022) 996–1009) and (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** (2022) 630–653). 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

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

with K_0 some prespecified positive integer. The spiked mean matrix with rank K_0 can be estimated by eigendecomposition, 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 that 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 $K_0 = 1, \ldots, K_{\max}$ 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 α 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 self-loops 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, [1, 15, 26]), Degree Corrected SBM (DCSBM, [18]), 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, [23] proposed an MCMC algorithm based on the allocation sampler to cluster the nodes in SBM and simultaneously estimate K. Airoldi et al. [3] developed a general variational inference algorithm to estimate the parameters in MM model with K chosen according to some BIC criterion. Jin et al. [16] considered testing (1) with $K_0 = 1$ and proposed a signed polygon statistic, which can accommodate the degree heterogeneity in the DCMM model. Gao and Lafferty [13] proposed EZ statistics constructed by "frequencies of three-node subgraphs" to test (1) with $K_0 = 1$ in the setting of DCSBM. Banerjee and Ma [5] introduced a linear spectral statistic to test $H_0 : K = 1$ versus $H_1 : K = 2$ under the SBM. Jin et al. [17] proposed a stepwise goodness-of-fit test for estimating K under DCSBM. Zhang and Amini [28] 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, 19, 25] and [27], among others. Chatterjee [8] introduced a universal singular value thresholding procedure for the matrix estimation, which can be applied to estimate K. Chen and Lei [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 [22] under the widely used inhomogeneous Erdős–Renyi model with low rank mean matrix via edge sampling. Le and Levina [20] proposed to estimate K using the spectral properties

of two graph operators—the nonbacktracking matrix and the Bethe Hessian matrix. Zhao et al. [29] 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. Bickel and Sarkar [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. Lei [21] 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 [21] 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 similar 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 [14].

1.1. *Notation.* We introduce some notation that will be used throughout the paper. We use $a \ll b$ to represent $a/b \to 0$ and write $a \lesssim b$ if there exists a positive constant c such that $0 \le a \le cb$. For a matrix $\mathbf{A}$, we use $\lambda_j(\mathbf{A})$ to denote the jth largest eigenvalue, and $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_\infty$ to denote the Frobenius norm, the spectral norm and the maximum elementwise infinity norm, respectively. In addition, denote by $\mathbf{A}(k)$ the kth 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 $\widetilde{\mathbf{X}}$ admitting the following decomposition:

$$(2) \qquad \widetilde{\mathbf{X}} = \mathbf{H} + \mathbf{W},$$

where $\mathbf{H} = \mathbb{E}(\widetilde{\mathbf{X}})$ 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 $\widetilde{\mathbf{X}}$ or $\widetilde{\mathbf{X}} - \text{diag}(\widetilde{\mathbf{X}})$, with the former corresponding to network with self-loops and the latter corresponding to network without self-loops, 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 $\widetilde{\mathbf{X}}$ 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

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

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 $\widetilde{\mathbf{X}}$ and $\mathbf{X}$.

Write the eigendecomposition of $\mathbf{H}$ as $\mathbf{VDV}^T$, where $\mathbf{D} = \text{diag}(d_1, \ldots, d_K)$ collects the nonzero eigenvalues of $\mathbf{H}$ in decreasing magnitude and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_K)$ is the matrix collecting the corresponding eigenvectors. Denote by $\hat{d}_1, \ldots, \hat{d}_n$ the eigenvalues of $\mathbf{X}$ in decreasing magnitude and $\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_n$ 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 $\mathbf{X} = \widetilde{\mathbf{X}}$ and follows model (2). It is seen that $\mathbb{E}\mathbf{W} = \mathbf{0}$. Intuitively, as $n \to \infty$, 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 $n \to \infty$. These two results entail that

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

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

In the ideal case where the eigenvalues $d_1, \ldots, d_K$ and eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_K$ are known, a test of the form (3) can be constructed by replacing w_{ii} with $\tilde{w}_{ii}$ where $\widetilde{\mathbf{W}} = (\tilde{w}_{ij}) = \mathbf{X} - \sum_{k=1}^{K_0} d_k \mathbf{v}_k \mathbf{v}_k^T$. Under the null hypothesis, $\widetilde{\mathbf{W}} = \mathbf{W}$ and the corresponding test statistic (constructed in the same way as (3)) is asymptotically normal. However, under the alternative hypothesis, $\widetilde{\mathbf{W}}$ 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 $\widetilde{\mathbf{W}}$ takes the form

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

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

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

Intuitively, the asymptotic behavior of the above statistic should be close to the one in (3). Thus, by examining the asymptotic behavior of $\widetilde{T}_n$ 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 self-loops, 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, the 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:

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

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* θ_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* ϵ, ϵ'.

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 ϵ 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) $$T_n \xrightarrow{d} N(0, 1) \quad as \ n \to \infty.$$

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 $n \to \infty$,

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

In addition, if

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

we have

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

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 nonnegligible compared with $\sqrt{2m^{-1} \sum_{i \neq j} \hat{w}_{ij}^2}$, then our test has nonvanishing 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*

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

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

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

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 restricts 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 [24], 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 σ_{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 in the Supplementary Material. 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

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

Here, $\mathbf{B}$ is a 2×2 nonsingular matrix with diagonals 1 and off diagonals taking a constant value in $[0, 1)$, $\boldsymbol{\Theta}$ 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 $\boldsymbol{\Pi} \in \mathbb{R}^{n \times 2}$ 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 ϑ_1, and the remaining belong to community 2 and share the common degree parameter ϑ_2. Since the population eigenvector $\mathbf{v}_k$, $k = 1, 2$ satisfies

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

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 ϑ_1 and ϑ_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 guidelines on the choice of m.

As an example, we consider the special case

$$(\log n)^{\delta} \min_{i \neq j} \sigma_{ij}^2 \gtrsim \max_{i \neq j} \sigma_{ij}^2$$

(12)

$$\text{for } K_0 < K, \sum_{i \neq j} \sigma_{ij}^2 \lesssim \left| \sum_{k=K_0+1}^{K} d_k \sum_{i \neq j} \mathbf{v}_k(i)\mathbf{v}_k(j) \right| \text{ and } |d_1| \lesssim n\theta_n,$$

where $\delta < \epsilon$ is some small positive constant with ϵ 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 $\mathbf{\Theta} = \text{diag}\{\vartheta_1, \ldots, \vartheta_n\}$ with $\vartheta_j > 0$, $j = 1, \ldots, n$ the degree parameters. Assume $\max_{i,j}\{\vartheta_j \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j}\}$ 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 $\sum_{i \neq j} \sigma_{ij}^2 \lesssim |d_2 \sum_{i \neq j} \mathbf{v}_2(i)\mathbf{v}_2(j)|$ 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 $n^2 \max_{i \neq j} \vartheta_i \vartheta_j b_{\mathcal{C}_i \mathcal{C}_j} \lesssim |d_2 \sum_{i \neq j} \mathbf{v}_2(i)\mathbf{v}_2(j)|$.

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 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 [21] considered dense SBM with θ_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, [27] proposed a model selection criterion for estimating K under the sparse setting of SBM with $n\theta_n / \log n \to \infty$. Le

and Levina [20] 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 self-loops*. We formalize the heuristic arguments in Section 2.2 about the ratio statistic $\widetilde{T}_n$ in (5) when the network admits self-loops. In such a case, the general test (6) still works. However, the simpler one $\widetilde{T}_n$ 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 self-loops and* $\sqrt{\sum_{i=1}^n \sigma_{ii}^2} \gg (\log n)^{2+\epsilon_2}$ *for some positive constant* ϵ_2.

(i) *Under null hypothesis, we have*

$$(14) \qquad \widetilde{T}_n \xrightarrow{d} N(0,1) \quad as \ n \to \infty.$$

(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*

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

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*

$$(16) \qquad \mathbb{P}(|\widetilde{T}_n| > C) \to 1,$$

for arbitrarily large positive constant C.

It is seen that with the same critical value $\Phi^{-1}(1 - \alpha/2)$, $\widetilde{T}_n$ enjoys the same asymptotic properties on size and power as T_n. In addition, since the construction of $\widetilde{T}_n$ 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 [21]. That is, we sequentially test the following hypotheses:

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

for $K_0 = 1, 2, \ldots, K_{\max}$ 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 $\widehat{K}$ our resulting estimate.

COROLLARY 2. *Suppose there exists some positive constant* δ_1 *such that*

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

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

$$\mathbb{P}(\widehat{K} = K) \to 1 - \alpha.$$

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:

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

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 Material.

3.6. *Networks without self-loops*: *Why subsampling?*   In this section, we formalize the heuristic arguments given in Section 2.2 on why subsampling is necessary. We theoretically show that T_n is no longer asymptotically normal under H_0 without the ingredient of subsampling. 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 notation that will be used in this subsection. For any matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of appropriate dimensions, let

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

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

$$\mathbf{b}_{\mathbf{e}_i,k,t}^T = \mathbf{e}_i^T - \mathcal{R}(\mathbf{e}_i, \mathbf{V}_{-k}, t)\big((\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t)\big)^{-1} \mathbf{V}_{-k}^T,$$

$$\mathbf{s}_{k,i} = \mathbf{b}_{\mathbf{e}_i,k,t_k} - \mathbf{e}_i^T \mathbf{v}_k \mathbf{v}_k, \qquad \mathbf{s}_k = \sum_{i=1}^{n} \mathbf{s}_{k,i}, \ \mathbf{s}_k(i) = \mathbf{e}_i^T \mathbf{s}_k, \quad \text{and}$$

$$\mathbf{r}_k = \mathbf{V}_{-k}\big(t_k \mathbf{D}_{-k}^{-1} - \mathbf{I}\big)^{-1} \mathbf{V}_{-k}^T \mathbb{E} \mathbf{W}^2 \mathbf{v}_k,$$

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

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

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

We have the following theorems.

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

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

for some positive constant ϵ_1. Under null hypothesis, as $n \to \infty$, we have

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

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 $n \to \infty$, we have*

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

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

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

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 $\mathbf{v}_1$ and we have $a_k = 0$, $k \geq 2$. By Condition 4 and Theorem S.1 in the Supplementary Material, 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

$$\sum_{i<j}\sigma_{ij}^2\left(1 - \sum_{k=1}^{K_0} a_k^2 \mathbf{v}_k(i)\mathbf{v}_k(j) - \sum_{k=1}^{K_0} a_k\big(\mathbf{v}_k(j)\mathbf{s}_k(i) + \mathbf{v}_k(i)\mathbf{s}_k(j)\big)\right)^2$$
$$= \sum_{i<j}\sigma_{ij}^2\big(1 - n\mathbf{v}_1(i)\mathbf{v}_1(j) - \sqrt{n}\big(\mathbf{v}_1(j)\mathbf{s}_1(i) + \mathbf{v}_1(i)\mathbf{s}_1(j)\big)\big)^2$$
$$= \sum_{i<j}\sigma_{ij}^2\big(\mathbf{s}_1(i) + \mathbf{s}_1(j)\big)^2$$
$$\lesssim \frac{\alpha_n^4}{nd_1^4}\sum_{i<j}\sigma_{ij}^2 \leq \frac{\alpha_n^6}{d_1^4} \lesssim \frac{n^2\alpha_n^2}{d_1^4} \lesssim \left(1 + \frac{n^2\alpha_n^2}{d_{K_0}^2}\right),$$

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 $\boldsymbol{\pi}_i$ chosen from the set of unit vectors

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

and the mixed membership node with $\boldsymbol{\pi}_i$ chosen from

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

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

$$(22) \qquad \mathbf{H} = \rho \boldsymbol{\Pi} \mathbf{B} \boldsymbol{\Pi}^T, \quad \rho \in (0, 1), \boldsymbol{\pi}_i \in \text{PN}(K), i = 1, \ldots, n.$$

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 ρ ranges from 0.04 to 0.9, with smaller ρ corresponding to sparser network model. For all values of K, we choose $m = \sqrt{n}$ in calculating the RIRS test statistics T_n and $\widetilde{T}_n$ for networks without and with self-loops, respectively.

The performance of RIRS is compared with the methods in [21], where two versions of the test—one with and one without bootstrap correction—were proposed when the network is absent of self-loops. 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 Material.

From Tables 1 and 2, we observe that when $K = 2$ the performance of RIRS is relatively robust to the sparsity level ρ, with size close to the nominal level and power close to 1 in almost all settings. On the contrary, the method in [21] without bootstrap correction has much worse performance. It suffers from size distortion for smaller ρ (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 [21] 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 [21].

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

| | No self-loop | | | | | | Self-loops | |
| | RIRS (T_n) | | Lei (no bootstrap) | | Lei (bootstrap) | | RIRS ($\widetilde{T}_n$) | |
ρ	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

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 a smaller size of each community. From Table 3, we see that the computational cost of the bootstrap method in [21] 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 Material 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 a reference. It visually confirms that the asymptotic null distribution is standard normal.

4.1.2. *DCMM*. Next, consider the general DCMM model (11). The number of repetitions is still 200. We simulate the node degree parameters ϑ_j's independently from the uniform distribution over [0.5, 1]. The vectors $\boldsymbol{\pi}_i$ are chosen from $\mathrm{PN}(K) \cup \mathrm{MM}(K, 0.2)$, with n_0 pure nodes from each community and $(n - Kn_0)/3$ nodes from each mixed membership probability mass vector in $\mathrm{MM}(K, 0.2)$. We select $n_0 = 0.35n$ when $K = 2$ and $n_0 = 0.25n$

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

| | No self-loop | | | | | | | Self-loops | | |
| | RIRS (T_n) | | | Lei (no bootstrap) | Lei (bootstrap) | | | RIRS ($\widetilde{T}_n$) | | |
ρ	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 self-loop, $K = 2$ 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

when $K = 3$. The matrix **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 [21] only considers SBM, the tests therein are no longer applicable in this setting. RIRS performs well and similar 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.

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 [21] are not applicable to the DCMM model, we only compare the performance of RIRS with [21] in SBM setting in the absence of self-loops. 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 ρ is large, and is significantly better than the latter for very small ρ. 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.

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 **X** from the following model:

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

where the residual matrix **W** is symmetric with upper triangle entries (including the diagonal ones) i.i.d. from uniform distribution over $(-1, 1)$. Let $\mathbf{V} = \frac{1}{\sqrt{2}}\binom{\mathbf{V}_1}{\mathbf{V}_2}$, 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
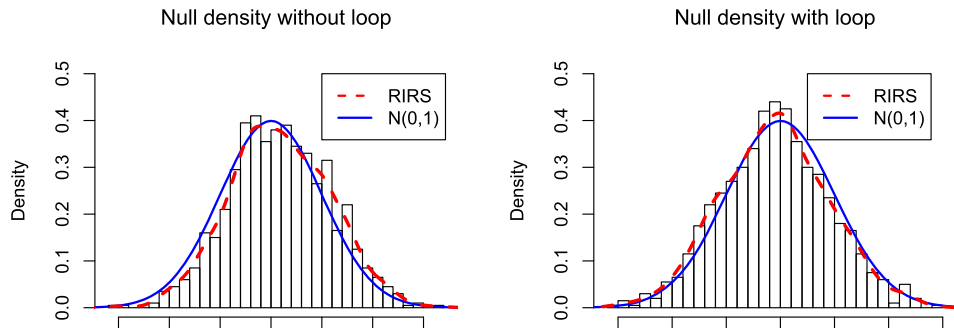

FIG. 1. *Histogram plots and the estimated densities* (*red curves*) *of RIRS test statistic when $K = 2$ and $\rho = 0.7$. Left: T_n when no self-loop; Right: $\tilde{T}_n$ when self-loops exist.*

TABLE 4
Empirical size and power of RIRS under DCMM model

| | K = 2 | | | | K = 3 | | | | | | |
| | No self-loop (T_n) | | Self-loop ($\widetilde{T}_n$) | | No self-loop (T_n) | | | Self-loop ($\widetilde{T}_n$) | | | |
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

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 $\mathbf{\Pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{n-n_1})^T$, where $\boldsymbol{\pi}_i \in \mathrm{PN}(K)$ and the number of rows taking each distinct value from $\mathrm{PN}(K)$ is the same. The diagonal matrix $\mathbf{D} = n \times \mathrm{diag}(K, K-1, \ldots, 1)$. 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 $\widetilde{T}_n$ perform well, with $\widetilde{T}_n$ having slightly higher power. This higher power further translates into better estimation accuracy (closer to 95%) of estimated K.

5. Real data analysis. We consider a popularly studied network of political blogs assembled by [2]. The nodes are blogs over the period of 2 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 [18, 30] and [21] 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:

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

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

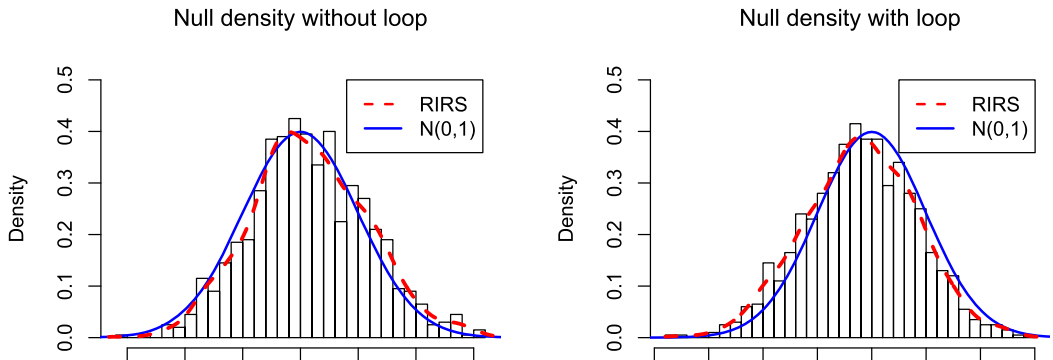

FIG. 2. *DCMM. Histogram plots and the estimated densities (red curves) of RIRS when $K = 3$ and $n = 1500$. Left: T_n when no self-loop; Right: $\widetilde{T}_n$ when self-loops exist.*

TABLE 5
Proportion of correctly estimated K under SBM

	$K = 2$				$K = 3$			
	No self-loop			Self-loop	No self-loop			Self-loop
	RIRS	Lei	Lei	RIRS	RIRS	Lei	Lei	RIRS
ρ	T_n	(no bootstrap)	(bootstrap)	$\widetilde{T}_n$	T_n	(no bootstrap)	(bootstrap)	$\widetilde{T}_n$
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

	$K = 2$						$K = 3$					
n	800	1000	1200	1500	1800	2000	800	1000	1200	1500	1800	2000
	No self-loop (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
	Self-loop ($\widetilde{T}_n$)											
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

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

	No self-loop (T_n)					Self-loop ($\widetilde{T}_n$)				
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

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

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

Lei [21] 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 [21] 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 nonsymmetric, we consider two simple transformations:

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

The transformation in Method 2 is general and can be applied to even nonsquare 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.

APPENDIX: 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, $i = 1, \ldots, n$. We say $\xi_n = O_{p_u}(\zeta_n)$ if for any positive constants D and ϵ, there exists some positive integer $n_0(D, \epsilon)$ depending only on D and ϵ such that for all $n \geq n_0(D, \epsilon)$ we have

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

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

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 give 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 Material. In the main paper, we only provide the proof of Theorem 3.1. All other proofs are relegated to the Supplementary Material.

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

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

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

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

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

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:

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

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

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

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 $Y_{ij}, i \le j$ are i.i.d. with $\mathbb{E}Y_{ij} = \frac{1}{m}$, Corollary S.1 and Condition 1 ensure that

$$\text{var}\left(\sum_{i \ne j} \left(Y_{ij} - \frac{1}{m} \right) (\mathbf{v}_k(i)\mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k \right)$$

$$= \mathbb{E}\left[\text{var}\left(\sum_{i \ne j} \left(Y_{ij} - \frac{1}{m} \right) (\mathbf{v}_k(i)\mathbf{v}_k(j)) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k | \mathbf{W} \right) \right]$$

$$\lesssim \frac{1}{m} \sum_{i \ne j} (\mathbf{v}_k(i)\mathbf{v}_k(j))^2 \mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 \lesssim \frac{\alpha_n^2}{mn} + \frac{1}{m},$$

then

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

Therefore,

$$\sum_{i \ne j} (\mathbf{v}_k(i)\mathbf{v}_k(j)Y_{ij}) \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = (a_k^2 + 1) O_{p_u}\left(\frac{\alpha_n}{m\sqrt{n}} \right)$$

$$(29)$$

$$+ O\left(\frac{a_k^2 + 1}{m} \right) + O_p\left(\frac{\alpha_n}{\sqrt{mn}} \right) + O_p\left(\frac{1}{\sqrt{m}} \right).$$

Similar to (29), we get

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

and

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

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

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

By (S.85), we have

$$\| \mathcal{R}(\mathbf{1}, \mathbf{V}_{-k}, t_k)((\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k))^{-1} \mathbf{V}_{-k}^T \|$$

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

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

$$\frac{d_k (\mathbf{s}_k - \mathbf{1})^T \mathbf{W} \mathbf{v}_k a_k}{t_k} = -\frac{d_k a_k \mathcal{R}(\mathbf{1}, \mathbf{V}_{-k}, t_k)((\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k))^{-1} \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k}{t_k}$$

$$- \frac{d_k a_k^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k}{t_k} = O_{p_u}(\alpha_n + |a_k|\sqrt{n}) + O_{p_u}(|a_k|\alpha_n),$$

$$\frac{1}{m}\sum_{i\neq j}\frac{d_k\mathbf{s}_{k,j}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k(i)}{t_k} = \frac{1}{m}a_k\frac{d_k\mathbf{1}^T(\mathbf{W}-\mathbb{E}\mathbf{W})\mathbf{v}_k}{t_k} + O_{p_u}\left(\frac{(|a_k|+1)\alpha_n}{m}\right)$$

$$= O_{p_u}\left(\frac{(|a_k|+1)\alpha_n}{m}\right)$$

and

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

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

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

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

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

It follows from Conditions 2–4 and 6, $\frac{t_k}{d_k}\to 1$ 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

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

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

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

Therefore,

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

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

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

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

$$\sum_{i\neq j}\frac{d_k\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k}{mt_k^2}=\frac{d_k\mathbf{1}^T\mathbf{W}\mathbf{v}_k\mathbf{1}^T\mathbf{W}\mathbf{v}_k}{mt_k^2}-\sum_{i=1}^n\frac{d_k(\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k)^2}{mt_k^2}$$

$$=O_{p_u}\left(\frac{\alpha_n^2(\log n)^2}{m|d_k|}\right)=O_{p_u}\left(\frac{(\log n)^2}{m}\right).$$

Moreover, calculating the variance, we have

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

$$=\mathbb{E}\left(\left[\text{var}\left(\sum_{i\neq j}\frac{d_k\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k(Y_{ij}-\frac{1}{m})}{t_k^2}\right)\Big|\mathbf{W}\right]\right)$$

$$\lesssim\frac{\sum_{i\neq j}\mathbb{E}(\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k)^2}{md_k^2}\leq\frac{\sum_{i\neq j}\sqrt{\mathbb{E}(\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k)^4\mathbb{E}(\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k)^4}}{md_k^2}$$

$$\lesssim\sum_{i\neq j}([\mathbb{E}(\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k-\mathbb{E}\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k)^4+(\mathbb{E}\mathbf{e}_i^T\mathbf{W}\mathbf{v}_k)^4][\mathbb{E}(\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k-\mathbb{E}\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k)^4$$

$$+(\mathbb{E}\mathbf{e}_j^T\mathbf{W}\mathbf{v}_k)^4])^{1/2}/(md_k^2)$$

$$\lesssim\frac{\alpha_n^4}{md_k^2}=O_{p_u}\left(\frac{1}{m}\right).$$

Therefore,

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

Finally, consider the residual term

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

we have

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

Then to prove (25), it suffices to show

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

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

$$\left| \sum_{i \neq j} w_{ij} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) \right| \leq \left| O_{p_u}\left(\frac{(\alpha_n \log n)^2}{|d_{K_0}|} + 1\right) \times \sqrt{\sum_{i \neq j} w_{ij}^2} \right|$$

$$= O_{p_u}\left(\alpha_n (\log n) \sqrt{\sum_{i \neq j} \mathbb{E} w_{ij}^2}\right).$$

Condition 5 further implies that

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

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

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

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

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

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

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

$$\tag{43} \sum_{i \neq j} O_{p_u}\left(\frac{(\alpha_n \log n)^2}{n|d_{K_0}|} + \frac{1}{n}\right) \sum_{k=1}^{K_0} \frac{d_k(\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k}$$

$$= O_{p_u}\left(\frac{(\alpha_n \log n)^3}{|d_{K_0}|} + \alpha_n \log n\right) = O_{p_u}(\alpha_n^2 \log^2 n) = O_{p_u}\left(\frac{1}{(\log n)^{\epsilon_1/2}} \sum_{i \neq j} \mathbb{E} w_{ij}^2\right).$$

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

$$
\sum_{i \neq j} \left(\sum_{k=1}^{K_0} \frac{d_k(\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))}{t_k} \right)^2
$$

(44)
$$
\lesssim \sum_{k=1}^{K_0} \sum_{i \neq j} \frac{d_k^2(\mathbf{e}_i^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(j) + \mathbf{e}_j^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k(i))^2}{t_k^2}
$$

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

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.

SUPPLEMENTARY MATERIAL

Supplement to "Universal rank inference via residual subsampling with application to large networks" (DOI: 10.1214/23-AOS2282SUPP; .pdf). This Supplementary Material contains some additional simulation examples, theorems, key lemmas and their proofs as well as additional technical details.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[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. *J. Mach. Learn. Res.* **9** 1981–2014.

[4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

[5] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. Available at arXiv:1705.05305.

[6] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

[7] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

[8] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

[9] CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 https://doi.org/10.1080/01621459.2016.1246365

[10] DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817 https://doi.org/10.1007/s11222-007-9046-7

[11] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990

[12] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505

[13] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. Available at arXiv:1710.00862.

[14] HAN, X., YANG, Q. and FAN, Y. (2023). Supplement to "Universal rank inference via residual subsampling with application to large networks." https://doi.org/10.1214/23-AOS2282SUPP

[15] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Comput. Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[16] JIN, J., KE, Z. T. and LUO, S. (2021). Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.* **49** 3408–3433. MR4352535 https://doi.org/10.1214/21-aos2089

[17] JIN, J., KE, Z. T., LUO, S. and WANG, M. (2020). Estimating the number of communities by stepwise goodness-of-fit. Available at arXiv:2009.09177.

[18] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E (3)* **83** 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[19] LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **12** 93–115. MR2953099 https://doi.org/10.1177/1471082X1001200105

[20] LE, C. M. and LEVINA, E. (2022). Estimating the number of communities in networks by spectral methods. Available at arXiv:1507.00827.

[21] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

[22] LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. MR4108931 https://doi.org/10.1093/biomet/asaa006

[23] 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. *Comput. Statist. Data Anal.* **60** 12–31. MR3007016 https://doi.org/10.1016/j.csda.2012.10.021

[24] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[25] SALDAÑA, D. F., YU, Y. and FENG, Y. (2017). How many communities are there? *J. Comput. Graph. Statist.* **26** 171–181. MR3610418 https://doi.org/10.1080/10618600.2015.1096790

[26] WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

[27] WANG, Y. X. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528. MR3650391 https://doi.org/10.1214/16-AOS1457

[28] ZHANG, L. and AMINI, A. A. (2020). Adjusted chi-square test for degree-corrected block models. Available at arXiv:2012.15047.

[29] ZHAO, Y., LEVINA, E. and ZHU, J. (2011). Community extraction for social networks. *Proc. Natl. Acad. Sci. USA* **108** 7321–7326. https://doi.org/10.1073/pnas.1006642108

[30] 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. MR3059083 https://doi.org/10.1214/12-AOS1036