10-2017

Strong consistency of spectral clustering for stochastic block models

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Wuyi WANG

Yichong ZHANG

STRONG CONSISTENCY OF SPECTRAL CLUSTERING FOR STOCHASTIC BLOCK MODELS

By Liangjun Su* Wuyi Wang and Yichong Zhang

Singapore Management University

In this paper we prove the strong consistency of several methods based on the spectral clustering techniques that are widely used to study the community detection problem in stochastic block models (SBMs). We show that under some weak conditions on the minimal degree, the number of communities, and the eigenvalues of the probability block matrix, the K-means algorithm applied to the eigenvectors of the graph Laplacian associated with its first few largest eigenvalues can classify all individuals into the true community uniformly correctly almost surely. Extensions to both regularized spectral clustering and degree-corrected SBMs are also considered. We illustrate the performance of different methods on simulated networks.

1. Introduction. Community detection is one of the fundamental problems in network analysis, where communities are groups of nodes that are, in some sense, more similar to each other than to the other nodes. The stochastic block model (SBM) that was first proposed by [8] is a common tool for model-based community detection that has been widely studied in the statistics literature. Within the SBM framework, the most essential task is to recover the community membership of the nodes from a single observation of the network. Various procedures have been proposed to solve this problem in the last decade or so. These include modularity maximization [15], likelihood methods [1, 2, 6, 23], method of moments [4], spectral clustering [9, 12, 16, 17, 18], and spectral embedding [13, 19]. Among them, spectral clustering is arguably one of the most widely used methods due to its computational tractability.

[2] introduce the notion of strong consistency of community detection as the number of nodes, n, grows.[1] By strong consistency, they mean that one can identify the members of the block model communities perfectly in

*Su acknowledges the funding support provided by the Lee Kong Chian Fund for Excellence.

MSC 2010 subject classifications: Primary 62E10; secondary 62G05

Keywords and phrases: Clustering, community detection, degree-corrected stochastic block model, k-means, regularization, strong Consistency

[1] [2] use the terminology "asymptotic consistency" in place of strong consistency.

large samples. They give the necessary and sufficient conditions for strong consistency based on the parameters of the block model, properties of the modularities, and expected degree of the graph (λ_n). In particular, they find that $\lambda_n / \log(n) \to \infty$ is necessary for strong consistency. [23] define weak consistency of community detection, which essentially means that the number of misclassified nodes is of smaller order than the number of nodes. [3] find that weak consistency requires that $\lambda_n \to \infty$ for the SBM. Similarly, under the conditions that $\lambda_n / \log(n) \to \infty$ ($\lambda_n \to \infty$), [23] establish the strong (weak) consistency under both standard SBMs and degree-corrected SBMs.

It is well known that some methods like the modularity maximization of [15] and the likelihood method of [2] yield strongly consistent community recovery under some mild conditions on the growth of the node degrees, but they either rely on combinatorial methods that are computationally demanding or are guaranteed to be successful only when the starting values are well-chosen. Spectral clustering has been shown to enjoy weak consistency under standard or degree-corrected SBMs by various researchers; see [9], [12], [16], and [17]. For example, [12] establish the weak consistency for spectral clustering in SBMs with expected degree as small as $\log(n)$. But to the best of our knowledge, the strong consistency of spectral clustering has not been formally established.

The aim of this paper is to formally establish the strong consistency of spectral clustering for standard/regular SBMs under a set of conditions on the minimal degree of nodes (μ_n), the number of communities (K), the minimal value of the nonzero eigenvalue of the normalized block probability matrix, and some other parameters of the block model. In the special case where K is fixed and the normalized block probability matrix has minimal eigenvalue bounded away from zero in absolute value, we show that $\mu_n / \log(n) \to \infty$ is sufficient to ensure strong consistency.

As demonstrated by [1], the performance of spectral clustering can be considerably improved via regularization. [9] provide an attempt at quantifying this improvement through theoretical analysis and find that the typical minimal degree assumption for the consistency of spectral clustering can potentially be removed with suitable regularization. In this paper we also establish the strong consistency of regularized spectral clustering.

The SBM is limited by its assumption that all nodes within a community are stochastically equivalent and thus provides a poor fit to real-world networks with hubs or highly varying node degrees within communities. For this reason, [10] propose a degree-corrected SBM (DC-SBM) to allow variation in node degrees within a community while preserving the overall block community structure. The DC-SBM greatly enhances the flexibility of modeling degree heterogeneity and enables us to fit network data with varying degree distributions. We also prove the strong consistency of spectral clustering for regularized DC-SBMs.

In the simulation, we consider both standard SBMs and DC-SBMs. For standard SBMs, we adopt [9]'s regularization method and choose the tuning parameter τ according to their recommendation. The results show that in terms of classification, spectral clustering tends to outperform the unconditional pseudo-likelihood (UPL) method, which also has the strong consistency property ([1]). In contrast, for the DC-SBMs our simulations suggest that the regularized spectral clustering tends to slightly underperform the conditional pseudo-likelihood (CPL) method even though both are strongly consistent under some conditions. We also show that an adaptive procedure helps the regularized spectral clustering to achieve much better performance than the CPL method.

The rest of the paper is organized as follows. We study the strong consistency of spectral clustering for the basic SBMs in Section 2. We consider the extensions to regularized spectral clustering and degree-corrected SBMs in Section 3. Section 4 reports the numerical performance of various spectral-clustering-based methods for a range of simulated networks. Section 5 concludes. All proofs of the main results are relegated to the mathematical appendix.

Notation. Throughout the paper, we write $[M]_{ij}$ as the (i,j)-th entry of matrix M. Without confusion, we sometimes simplify $[M]_{ij}$ as M_{ij}. In addition, we write $[M]_{i\cdot}$ as the i-th row of M. $\|M\|$ and $\|M\|_F$ denote the spectral norm and Frobenius norm of M, respectively. Note that $\|M\| = \|M\|_F$ when M is a vector. We use $\mathbf{1}\{\cdot\}$ to denote the indicator function which takes value 1 when $\cdot$ holds and 0 otherwise. C, c, and c' denote arbitrary positive constants that are independent of n, but may not be the same in different contexts.

2. Strong consistency of spectral clustering.

2.1. *Basic setup.* Let $A \in \{0,1\}^{n\times n}$ be the adjacency matrix. By convention, we do not allow self-connection, i.e., $A_{ii} = 0$. Let $\hat{d}_i = \sum_{j=1}^n A_{ij}$ denote the degree of node i, $D = \text{diag}(\hat{d}_1, \ldots, \hat{d}_n)$, and $L = D^{-1/2}AD^{-1/2}$ be the graph Laplacian. The graph is generated from a SBM with K communities. We assume that K is known and potentially depends on the number of nodes n. We omit the dependence of K on n for notation simplicity. If K is unknown, it can be determined by either [11]'s sequential goodness-of-fit testing procedure or the likelihood-based model selection method proposed by [20]. The communities, which represent a partition of the n nodes, are assumed to be fixed beforehand. Denote these by $C_1, \ldots, C_K$. Let n_k, for $k = 1, \ldots, K$, be the number of nodes belonging to each of the clusters.

Given the communities, the edge between nodes i and j are chosen independently with probability depending on the communities i and j belong to. In particular, for nodes i and j belonging to cluster C_{k_1} and C_{k_2}, respectively, the probability of edge between i and j is given by $P_{ij} = B_{k_1k_2}$, where the *block probability matrix* $B = \{B_{k_1k_2}\}$, $k_1, k_2 = 1, \ldots, K$, is a symmetric matrix with each entry between $[0, 1]$. The $n \times n$ edge probability matrix $P = \{P_{ij}\}$ represents the population counterpart of the adjacency matrix A. Frequently we suppress the dependence of matrices and their elements on n.

Denote $Z = \{Z_{ik}\}$ as the $n \times K$ binary matrix providing the cluster memberships of each node, i.e., $Z_{ik} = 1$ if node i is in C_k and $Z_{ik} = 0$ otherwise. Then we have $P = ZBZ^T$. Let $\mathcal{D} = \text{diag}(d_1, \ldots, d_n)$ where $d_i = \sum_{j=1}^n P_{ij}$. The population version of the graph Laplacian is $\mathcal{L} = \mathcal{D}^{-1/2} P \mathcal{D}^{-1/2}$. The standard spectral clustering corresponds to classifying the eigenvectors of L by K-means algorithm. In this paper, we focus on the strong consistency of both the standard spectral clustering and its variant.

2.2. *Identification of the group membership.* Let $\pi_{kn} = n_k/n$, $W_k = [B]_{k\cdot} Z^T \iota_n / n = \sum_{l=1}^K B_{kl}\pi_{ln}$, $\mathcal{D}_B = \text{diag}(W_1, \ldots, W_K)$, and $B_0 = \mathcal{D}_B^{-1/2} B \mathcal{D}_B^{-1/2}$, where ι_n is a vector of ones in $\Re^n$. We can view W_k as the weighted average of the k-th row of B with weights given by π_{kn}. Similarly, B_0 is a normalized version of B. Note that B_0 is symmetric as B is. Let $\Pi_n = \text{diag}(\pi_{1n}, \ldots, \pi_{Kn})$. Throughout the paper, we allow for the elements in the block probability matrix B to depend on n and decay to zero as n grows, which leads to a sparse graph.

ASSUMPTION 1. *B_0 has rank $K^* \leq K$ and the spectral decomposition of $\Pi_n^{1/2} B_0 \Pi_n^{1/2}$ is $S_n \Omega_n S_n^T$, in which S_n is a $K \times K^*$ matrix such that $S_n^T S_n = I_{K^*}$ and $\Omega_n = diag(\omega_{1n}, \ldots, \omega_{K^*n})$ such that $|\omega_{1n}| \geq \cdots \geq |\omega_{K^*n}|$.*

Assumption 1 implies that $B = \mathcal{D}_B^{1/2} \Pi_n^{-1/2} S_n \Omega_n S_n^T \Pi_n^{-1/2} \mathcal{D}_B^{1/2}$ and $B_0 = \Pi_n^{-1/2} S_n \Omega_n S_n^T \Pi_n^{-1/2}$. It is weaker than the full-rank assumption of [9] who also assume that K is fixed as $n \to \infty$. It also implies that $\mathcal{L}$ has rank K^* and has the spectral decomposition:

$$\mathcal{L} = U_n \Sigma_n U_n^T = U_{1n} \Sigma_{1n} U_{1n}^T,$$

where $\Sigma_n = \text{diag}(\sigma_{1n}, \ldots, \sigma_{K^*n}, 0, \ldots, 0)$ is a $n \times n$ matrix that contains the eigenvalues of $\mathcal{L}$ such that $|\sigma_{1n}| \geq |\sigma_{2n}| \geq \cdots \geq |\sigma_{K^*n}| > 0$, $\Sigma_{1n} =$

$\text{diag}(\sigma_{1n}, \ldots, \sigma_{K^*n})$, the columns of U_n contain the eigenvectors of $\mathcal{L}$ associated with the eigenvalues in Σ_n, $U_n = (U_{1n}, U_{2n})$, and $U_n^T U_n = I_n$. As shown in Theorem 2.1 below, $\sigma_{kn} = \omega_{kn}$ for $k = 1, \ldots, K^*$.

ASSUMPTION 2. *There exist some constants C and c such that*

$$\infty > C \geq \limsup_n \sup_k n_k K/n \geq \liminf_n \inf_k n_k K/n \geq c > 0.$$

Assumption 2 implies that the network has balanced communities. It is commonly assumed in the literature but can be relaxed at the cost of more complicated notation.

ASSUMPTION 3. *Let $z_i^T = [Z]_{i\cdot}$, the i-th row of Z. There exists a deterministic sequence $\{\xi_n\}_{n\geq 1}$ such that if $z_i \neq z_j$,*

$$(n/K)^{1/2} \|(z_i^T - z_j^T)(Z^T Z)^{-1/2} S_n\| \geq \xi_n > 0.$$

Assumption 3 imposes a condition on the matrix S_n. It is weak because we have not specified any explicit condition on $\{\xi_n\}$. In fact, as Theorem 2.1 below shows, ξ_n is bounded below from zero if $K^* = K$ and the above assumption can be ensured by imposing some restrictions on the rows of B_0 when $K^* < K$.

THEOREM 2.1. *If Assumptions 1 and 2 hold, then $\Omega_n = \Sigma_n$, $U_{1n} = Z(Z^T Z)^{-1/2} S_n$ and*

$$\sup_{1\leq i\leq n} (n/K)^{1/2} \|z_i^T (Z^T Z)^{-1/2} S_n\| = O(1).$$

In addition, if (i) $K^* = K$, then Assumption 3 holds with $\liminf_n \xi_n > 0$; (ii) if $K^* < K$ and there exists a deterministic sequence $\{\xi_n'\}_{n\geq 1}$ such that

$$\inf_{1\leq k_1 < k_2 \leq K} K^{-1} \|[B_0]_{k_1\cdot} - [B_0]_{k_2\cdot}\| \geq \xi_n' > 0,$$

then Assumption 3 holds with $\xi_n \geq c\xi_n'$ for some constant $c > 0$.

Noting that the ith row of U_{1n} is given by $z_i^T (Z^T Z)^{-1/2} S_n$, Theorem 2.1 indicates that the rows of U_{1n} contain the same community information as Z for all nodes in the network. Therefore, we can infer each node's community membership based on the eigenvector matrix U_{1n} if $\mathcal{L}$ is observed.

In practice, $\mathcal{L}$ is not observed. But we can estimate it by L. We show below that the eigenvectors of L associated with its few largest eigenvalues in absolute value consistently estimate those of $\mathcal{L}$ up to an orthogonal matrix so that the rows of the eigenvector matrix of L also contains the useful community information.

2.3. *Uniform consistency of the estimated eigenvectors.* To study the consistency of the eigenvectors of L associated with its K^* largest eigenvalues, we add the following assumption.

ASSUMPTION 4. *Let $\mu_n = \min_i d_i$. Then $\frac{\log(n)KK^*}{\mu_n} \to 0$, $\frac{\log(n)K^*}{\mu_n|\sigma_{K^*n}|} \to 0$, and $\frac{\log(n)}{\mu_n\sigma_{K^*n}^2} \to 0$.*

If K is fixed and $\liminf_n |\sigma_{K^*n}|$ is bounded away from zero, then Assumption 4 reduces to the requirement that $\log(n)/\mu_n \to 0$. That is, the minimal expected degree should grow faster than $\log(n)$. As pointed out by [1], such a condition is almost the minimal requirement for establishing the strong consistency of the standard spectral clustering. [5], [9], [16], and [19] consider the regularization of the graph Laplacian, which was proposed by [1], and establish weak consistency under conditions weaker than Assumption 4. We will come back to the strong consistency of the regularized spectral decomposition in Section 3.

Consider the spectral decomposition

$$L = \widehat{U}_n\widehat{\Sigma}_n\widehat{U}_n^T,$$

where $\widehat{\Sigma}_n = \mathrm{diag}(\hat{\sigma}_{1n}, \ldots, \hat{\sigma}_{nn})$ with $|\hat{\sigma}_{1n}| \geq |\hat{\sigma}_{2n}| \geq \cdots \geq |\hat{\sigma}_{nn}| \geq 0$, and $\widehat{U}_n$ is the corresponding eigenvectors. Let $\widehat{\Sigma}_{1n} = \mathrm{diag}(\hat{\sigma}_{1n}, \ldots, \hat{\sigma}_{K^*n})$, $\widehat{\Sigma}_{2n} = \mathrm{diag}(\hat{\sigma}_{K^*+1n}, \ldots, \hat{\sigma}_{nn})$, and $\widehat{U}_n = (\widehat{U}_{1n}, \widehat{U}_{2n})$, where $\widehat{U}_{1n}$ contains the eigenvectors associated with eigenvalues $\hat{\sigma}_{1n}, \ldots, \hat{\sigma}_{K^*n}$. Then, $\widehat{U}_{1n}^T\widehat{U}_{1n} = I_{K^*}$, $\widehat{U}_{2n}^T\widehat{U}_{1n} = 0$, and

$$L = \widehat{U}_{1n}\widehat{\Sigma}_{1n}\widehat{U}_{1n}^T + \widehat{U}_{2n}\widehat{\Sigma}_{2n}\widehat{U}_{2n}^T.$$

The following lemma indicates that L and $\widehat{U}_{1n}$ are consistent estimates of $\mathcal{L}$ and U_{1n} in terms of spectral norm, respectively, and up to an orthogonal matrix in the latter case.

LEMMA 2.1. *If Assumptions 1–4 hold, then there exist a positive constant C sufficiently large and a $K^* \times K^*$ orthogonal matrix O_n such that*

$$\|\mathcal{L} - L\| \leq C \log^{1/2}(n)\mu_n^{-1/2} \quad a.s.$$

and

$$\|\widehat{U}_{1n} - U_{1n}O_n\| \leq C \log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|(K^*)^{1/2} \quad a.s.$$

Two variants of Lemma 2.1 have been derived in [9] and [16] as special cases. The key differences are two-fold. First, we obtain the almost sure bound for the objects of interest instead of the probability bound. Second, U_{1n} and $\widehat{U}_{1n}$ are $n \times K^*$ instead of $n \times K$.

Let $\widehat{\Lambda} = L\widehat{U}_{1n} = \widehat{U}_{1n}\widehat{\Sigma}_{1n}$, $\Lambda = \mathcal{L}U_{1n}O_n = U_{1n}\Sigma_{1n}O_n$, $\widehat{\Lambda}_i = \widehat{u}_{1i}^T\widehat{\Sigma}_{1n}$, and $\Lambda_i = u_{1i}^T\Sigma_{1n}O_n$, where $\widehat{u}_{1i}^T$ and u_{1i}^T are the i-th rows of $\widehat{U}_{1n}$ and U_{1n}, respectively. In order to study the strong consistency, we have to derive the uniform bound for $\|\widehat{u}_{1i}^T - u_{1i}^T O_n\|$.

THEOREM 2.2. *Denote $\rho_n = \max(\sup_{k_1k_2}[B_0]_{k_1k_2}, 1)$. If Assumptions 1–4 hold, then*

$$\sup_i \sqrt{n/K}\|\widehat{u}_{1i} - O_n^T u_{1i}\| \leq C\left(\frac{\log(n)K^*}{\mu_n K}\right)^{1/2}\left(\frac{\rho_n + K^{1/2}}{\sigma_{K^*n}^2}\right) \quad a.s.$$

Note that ρ_n is a measure of heterogeneity of the normalized block probability matrix B_0. If all the entries in B are of the same order of magnitude, then ρ_n is bounded. In addition, by Assumption 2 and the fact that

$$(\pi_{k_1n}\pi_{k_2n})^{1/2}[B_0]_{k_1k_2} = \frac{(\pi_{k_1n}\pi_{k_2n})^{1/2}B_{k_1k_2}}{(\sum_{l=1}^K \pi_{ln}B_{k_1l})^{1/2}(\sum_{l=1}^K \pi_{ln}B_{k_2l})^{1/2}} \leq 1,$$

we have $\rho_n \leq CK$ for some constant $C > 0$. Therefore, if the number of blocks is fixed, then ρ_n is also bounded.

Since both U_{1n} and $\widehat{U}_{1n}$ have orthonormal columns, they have a typical element of order $(n/K)^{-1/2}$. This explains why we need the normalization constant $(n/K)^{1/2}$ in Theorem 2.2. An important implication of Theorem 2.2 is that like U_{1n}, the rows of $\widehat{U}_{1n}$ also contain the community membership information. Let $\tilde{\beta}_{in} = (n/K)^{1/2}\widehat{u}_{1i}$. Let $g_i^0 \in \{1, \ldots, K\}$ denote the true community that node i belongs to. Theorems 2.1 and 2.2 imply that there exists $\beta_{g_i^0 n} = (n/K)^{1/2}O_n^T u_{1i}$ such that

$$\|\tilde{\beta}_{in} - \beta_{g_i^0 n}\| = O_{a.s.}\left(\left(\frac{\log(n)K^*}{\mu_n K}\right)^{1/2}\left(\frac{\rho_n + K^{1/2}}{\sigma_{K^*n}^2}\right)\right) = o_{a.s.}(1)$$

uniformly in i provided that $\left(\frac{\log(n)K^*}{\mu_n K}\right)^{1/2}\left(\frac{\rho_n + K^{1/2}}{\sigma_{K^*n}^2}\right) = o(1)$. Then $\tilde{\beta}_{in}$ serves as a consistent estimator of $\beta_{g_i^0 n}$.

Example 2.1. *Consider the four-parameter SBM studied in [17]. The model is parametrized by K, s, r and p, where the K communities contain s nodes each, and r and $r+p$ denote the probability of a connection between two nodes in two separate blocks and in the same block, respectively. For this* *model,* $K^* = K$*,* $\rho_n = \frac{(p+r)K}{p+rK}$*,* $\sigma_{Kn} = \frac{p}{Kr+p}$*, and* $\mu_n = \frac{n(p+rK)}{K} - (p+r)$*. Therefore, the probability bound of* $\sup_i \sqrt{n/K}\|\widehat{u}_{1i} - O_n^T u_{1i}\|$ *is of order*

$$\left(\frac{K\log(n)(p+rK)}{n}\right)^{1/2}\left(\frac{(p+r)K + K^{1/2}(p+rK)}{p^2}\right).$$

The last term is of $o(1)$ *if* $K^3\log(n)/(np) \to 0$ *and* $rK/p \to c \in (0,\infty)$*, or if* $K^5\log(n)/(nr) \to 0$ *and* $r/p \to c \in (0,\infty)$*. If we further restrict our attention to the dense SBM with both* r *and* p *bounded away from zero, then the last displayed item becomes* $o(1)$ *as long as* $K^5\log(n)/n \to 0$*.*

2.4. *Strong consistency of the K-means algorithm.* Let $\hat{\beta}_{in}$ be a generic estimator of $\beta_{g_i^0 n}$ for $i = 1, \ldots, n$. To recover the community membership structure (i.e., to estimate g_i^0), it is natural to apply the K-means clustering algorithm to $\{\hat{\beta}_{in}\}$. Specifically, let $\mathcal{A} = \{\alpha_1, \ldots, \alpha_K\}$ be a set of K arbitrary $K^* \times 1$ vectors: $\alpha_1, \ldots, \alpha_K$. Define

$$\widehat{Q}_n(\mathcal{A}) = \frac{1}{n}\sum_{i=1}^{n} \min_{1\le l\le K} \|\hat{\beta}_{in} - \alpha_l\|^2$$

and $\widehat{\mathcal{A}}_n = \{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K\}$, where $\widehat{\mathcal{A}}_n = \arg\min_{\mathcal{A}} \widehat{Q}_n(\mathcal{A})$. Then, $\hat{g}_i$, the estimated cluster identity, is computed as

$$\hat{g}_i = \operatorname*{arg\,min}_{1\le l\le K} \|\hat{\beta}_{in} - \widehat{\alpha}_l\|,$$

in which if there are multiple l's that achieve the minimum, $\hat{g}_i$ takes value of the smallest one. Next, we consider the case in which the estimates $\{\hat{\beta}_{in}\}_{i=1}^n$ and the true vector $\{\beta_{kn}\}_{k=1}^K$ satisfy the following restrictions.

ASSUMPTION 5. *(i) There exists a constant* C *such that*

$$\limsup_n \sup_{1\le k\le K} \|\beta_{kn}\| \le C < \infty.$$

(ii) There exist some deterministic sequences c_{1n} *and* c_{2n} *such that* $\sup_i \|\hat{\beta}_{in} - \beta_{g_i^0 n}\| = O_{a.s.}(c_{2n})$ *and* $\liminf_n \inf_{1\le k<k'\le K} \|\beta_{kn} - \beta_{k'n}\| \ge c_{1n} > 0$*. (iii)* $\limsup_n c_{2n}c_{1n}^{-1} = 0$ *and* $\limsup_n c_{2n}c_{1n}^{-2}K(K^*)^{1/2} = 0$*.*

Let $H(\cdot,\cdot)$ denote the Hausdorff distance between two sets. Let $\mathcal{B}_n = \{\beta_{1n}, \ldots, \beta_{Kn}\}$. The following lemma shows that the K-means algorithm can estimate the true centroids $\{\beta_{kn}\}_{k=1}^K$ up to the rate $O_{a.s.}(c_{2n}^{1/2}K^{1/2}(K^*)^{1/4})$.

Lemma 2.2. *Suppose that Assumptions 2 and 5 hold. Then*

$$H(\widehat{\mathcal{A}}_n, \mathcal{B}_n) = O_{a.s.}(c_{2n}^{1/2} K^{1/2} (K^*)^{1/4}).$$

With Lemma 2.2, we can show the K-means classification is strongly consistent.

Theorem 2.3. *Suppose that Assumptions 2 and 5 hold. Then for sufficiently large n we have*

$$\sup_{1\le i\le n} \mathbf{1}\{\widehat{g}_i \neq g_i^0\} = 0 \quad a.s.$$

That is, we can classify all nodes into the true community a.s. in large samples. To apply the above theorem to $\tilde{\beta}_{in}$, we add the following condition.

Assumption 6. $\left(\frac{\log(n)K^*K}{\mu_n}\right)^{1/2}\left(\frac{(K^*)^{1/2}\rho_n+(K^*K)^{1/2}}{\sigma^2_{K^*n}\xi_n^2}\right) \to 0$ *as* $n \to \infty$.

Assumption 6 imposes the conditions for the strong consistency of the spectral clustering under fairly general conditions. If both $|\sigma_{K^*n}|$ and ξ_n are bounded away from zero, and $K^* = K$ is fixed, then the above assumption reduces to the requirement that $\mu_n/\log(n) \to \infty$ as $n \to \infty$, which is the minimal condition for the strong consistency discussed in [2].

Corollary 2.1. *Suppose that Assumptions 1–4 and 6 hold and the K-means algorithm is applied to $\hat{\beta}_{in} = \tilde{\beta}_{in} = (n/K)^{1/2}\widehat{u}_{1i}$. Then,*

$$\sup_{1\le i\le n} \mathbf{1}\{\widehat{g}_i \neq g_i^0\} = 0 \quad a.s.$$

Corollary 2.1 shows that the spectral-clustering-based K-means algorithm consistently recovers the community membership for all nodes almost surely in large samples.

Example 2.1 (cont.) *For the four-parameter model in Example 2.1, Assumption 6 is equivalent to*

$$\left(\frac{K^4\log(n)(p+rK)}{n}\right)^{1/2}\left(\frac{(p+r)K + K^{1/2}(p+rK)}{p^2}\right) \to 0, \tag{2.1}$$

If rK/p is bounded, then the above rate further reduces to $K^6\log(n)/(np) \to 0$, which allows $K = o((np/\log(n))^{1/6})$. As long as p decays to zero faster than $\log(n)/n$, Assumption 6 holds even when K grows slowly to infinity. On *the other hand, if $r/p \to c \in (0,\infty)$, (2.1) reduces to $K^8 \log(n)/(nr) \to 0$. In addition, if both p and r are bounded away from zero, then (2.1) requires that $K^8 \log(n)/n \to 0$. In contrast, [2] find that when $K = O\left(n^{1/4}/\log(n)\right)$ and p is bounded away from 0, the number of misclassified nodes from the K-means algorithm in the four-parameter SBM is of order $o\left(K^3 \log^2(n)\right) = o\left(n^{3/4}\right)$.*

3. Extensions. In this section we consider two extensions of the above results: regularized spectral clustering of the standard and degree-corrected SBMs.

3.1. *Regularized spectral clustering analysis for standard SBMs.* The SBM is the same as considered in the previous section. Following [1] and [9], we regularize the adjacency matrix A to be $A_\tau = A + \tau n^{-1}\iota_n\iota_n^T$, where $\tau \le n$ is the regularization parameter and ι_n is the $n\times 1$ vector of ones. Given the regularized adjacency matrix, we can compute the regularized degree for each node as $\hat{d}_i^\tau = \hat{d}_i + \tau$ and $D_\tau = \operatorname{diag}(\hat{d}_1+\tau,\ldots,\hat{d}_n+\tau)$. The regularized version of P and $\mathcal{D}$ are denoted as P_τ and $\mathcal{D}_\tau$ and defined as

$$P_\tau = P + \tau n^{-1}\iota_n\iota_n^T \quad \text{and} \quad \mathcal{D}_\tau = \operatorname{diag}(d_1+\tau,\ldots,d_n+\tau),$$

respectively. Consequently, the regularized graph Laplacian and its population counterpart are denoted as L_τ and $\mathcal{L}_\tau$ and written as

$$L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2} \quad \text{and} \quad \mathcal{L}_\tau = \mathcal{D}_\tau^{-1/2} P_\tau \mathcal{D}_\tau^{-1/2},$$

respectively. Noting that $\iota_n = Z\iota_K$, we have

$$P_\tau = P + \tau n^{-1}\iota_n\iota_n^T = ZBZ^T + \tau n^{-1} Z\iota_K\iota_K^T Z^T = ZB^\tau Z^T,$$

where $B^\tau = B + \tau n^{-1}\iota_K\iota_K^T$. Apparently, the block model structure is preserved after regularization. Given B^τ, we can define B_0^τ, the normalized version of B^τ as in the previous section. Let $W_k^\tau = [B^\tau]_{k\cdot} Z^T\iota_n/n = \sum_{l=1}^K [B^\tau]_{kl}\pi_{ln}$, $\mathcal{D}_B^\tau = \operatorname{diag}(W_1^\tau,\ldots,W_K^\tau)$, and $B_0^\tau = (\mathcal{D}_B^\tau)^{-1/2} B^\tau (\mathcal{D}_B^\tau)^{-1/2}$.

In order to follow the identification analysis in the previous section, we need to modify Assumptions 1 and 3.

ASSUMPTION 7. *(i) B_0^τ has rank $K^* \le K$ and the spectral decomposition of $\Pi_n^{1/2} B_0^\tau \Pi_n^{1/2}$ is $S_n^\tau \Omega_n^\tau (S_n^\tau)^T$, in which S_n^τ is a $K\times K^*$ matrix such that $(S_n^\tau)^T S_n^\tau = I_{K^*}$ and $\Omega_n^\tau = diag(\omega_{1n}^\tau,\ldots,\omega_{K^*n}^\tau)$ such that $|\omega_{1n}^\tau| \ge \cdots \ge |\omega_{K^*n}^\tau|$. (ii) There exists a deterministic sequence $\{\xi_n^\tau\}_{n\ge 1}$ such that if $z_i \ne z_j$,*

$$(n/K)^{1/2}\|(z_i^T - z_j^T)(Z^T Z)^{-1/2} S_n^\tau\| \ge \xi_n^\tau > 0.$$

We consider the eigenvalue decomposition of $\mathcal{L}_\tau$ as

$$\mathcal{L}_\tau = U_n^\tau \Sigma_n^\tau (U_n^\tau)^T = U_{1n}^\tau \Sigma_{1n}^\tau (U_{1n}^\tau)^T$$

where $\Sigma_n^\tau = \mathrm{diag}(\sigma_{1n}^\tau, \ldots, \sigma_{K^*n}^\tau, 0, \ldots, 0)$ is a $n \times n$ matrix that contains the eigenvalues of $\mathcal{L}_\tau$ such that $|\sigma_{1n}^\tau| \geq |\sigma_{2n}^\tau| \geq \cdots \geq |\sigma_{K^*n}^\tau| > 0$, $\Sigma_{1n}^\tau = \mathrm{diag}(\sigma_{1n}^\tau, \ldots, \sigma_{K^*n}^\tau)$, the columns of U_n^τ contain the eigenvectors of $\mathcal{L}_\tau$ associated with the eigenvalues in Σ_n^τ, $U_n^\tau = (U_{1n}^\tau, U_{2n}^\tau)$, and $(U_n^\tau)^T U_n^\tau = I_n$.

The following theorem parallels Theorem 2.1 in Section 2.2.

Theorem 3.1. *If Assumptions 2 and 7 hold, then* $\Omega_n^\tau = \Sigma_n^\tau$, $U_{1n}^\tau = Z(Z^TZ)^{-1/2}S_n^\tau$ *and*

$$\sup_{1 \leq i \leq n} (n/K)^{1/2} \|z_i^T (Z^TZ)^{-1/2} S_n^\tau\| = O(1).$$

In addition, if (i) $K^* = K$*, then Assumption 7(ii) holds with* $\liminf_n \xi_n^\tau > 0$*; (ii) if* $K^* < K$ *but there exists a deterministic sequence* $\{\xi_n'\}_{n \geq 1}$ *such that*

$$\inf_{1 \leq k_1 < k_2 \leq K} K^{-1} \|[B_0^\tau]_{k_1\cdot} - [B_0^\tau]_{k_2\cdot}\| \geq \xi_n' > 0,$$

then Assumption 7(ii) holds with $\xi_n^\tau \geq c\xi_n'$ *for some constant* $c > 0$.

Since $\mathcal{L}_\tau = n^{-1} Z B_0^\tau Z$, the proof of Theorem 3.1 is exactly the same as that of Theorem 2.1 with obvious modifications. Theorem 3.1 indicates that we can infer each node's community membership based on the eigenvector matrix U_{1n}^τ if $\mathcal{L}_\tau$ is observed.

As before, we consider the spectral decomposition of L_τ :

$$L_\tau = \widehat{U}_n^\tau \widehat{\Sigma}_n^\tau (\widehat{U}_n^\tau)^T = \widehat{U}_{1n}^\tau \widehat{\Sigma}_{1n}^\tau (\widehat{U}_{1n}^\tau)^T + \widehat{U}_{2n}^\tau \widehat{\Sigma}_{2n}^\tau (\widehat{U}_{2n}^\tau)^T.$$

where $\widehat{\Sigma}_n^\tau = \mathrm{diag}(\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{nn}^\tau) = \mathrm{diag}(\widehat{\Sigma}_{1n}^\tau, \widehat{\Sigma}_{2n}^\tau)$ with $|\hat{\sigma}_{1n}^\tau| \geq |\hat{\sigma}_{2n}^\tau| \geq \cdots \geq |\hat{\sigma}_{nn}^\tau| \geq 0$, $\widehat{\Sigma}_{1n}^\tau = \mathrm{diag}(\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{K^*n}^\tau)$, and $\widehat{\Sigma}_{2n}^\tau = \mathrm{diag}(\hat{\sigma}_{K^*+1,n}^\tau, \ldots, \hat{\sigma}_{nn}^\tau)$; $\widehat{U}_n^\tau = (\widehat{U}_{1n}^\tau, \widehat{U}_{2n}^\tau)$ is the corresponding eigenvectors such that $(\widehat{U}_{1n}^\tau)^T \widehat{U}_{1n} = I_{K^*}$ and $\widehat{U}_{2n}^T \widehat{U}_{1n} = 0$. Note that $\widehat{U}_{1n}^\tau$ contains the eigenvectors associated with eigenvalues $\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{K^*n}^\tau$. To study the asymptotic properties of $\widehat{U}_{1n}^\tau$, we modify Assumption 4 and 6 as follows.

Assumption 8. *Denote* $\mu_n^\tau = \min_i d_i + \tau$*. Then* $\frac{\log(n)KK^*}{\mu_n^\tau} \to 0$, $\frac{\log(n)K^*}{\mu_n^\tau |\sigma_{K^*n}^\tau|} \to 0$, $\frac{\log(n)}{\mu_n^\tau (\sigma_{K^*n}^\tau)^2} \to 0$, *and*

$$\left(\frac{\log(n)K^*K}{\mu_n^\tau}\right)^{1/2} \left(\frac{(K^*)^{1/2}\rho_n^\tau + (K^*K)^{1/2}}{(\sigma_{K^*n}^\tau \xi_n^\tau)^2}\right) \to 0.$$

The above modification is natural because node i's degree becomes $d_i+\tau$ after regularization. μ_n^τ can be interpreted as the effective minimum expected degree after regularization.

Let $(u_{1i}^\tau)^T$ and $(\widehat{u}_{1i}^\tau)^T$ be the i-th row of U_{1n}^τ and $\widehat{U}_{1n}^\tau$, respectively. The following theorem parallels Theorem 2.2 and Corollary 2.1 in Section 2.3.

THEOREM 3.2. *Denote $\rho_n^\tau = \max(\sup_{k_1k_2}[B_0^\tau]_{k_1k_2}, 1)$. Suppose that Assumptions 2, 7, and 8 hold. Then there exists a $K^* \times K^*$ orthonormal matrix O_n^τ such that*

$$\sup_{1\le i\le n} \sqrt{n/K}\|\widehat{u}_{1i}^\tau - (O_n^\tau)^T u_{1i}^\tau\| \le C\left(\frac{\log(n)}{\mu_n^\tau K}\right)^{1/2}\left(\frac{(K^*)^{1/2}\rho_n^\tau + (K^*K)^{1/2}}{(\sigma_{K^*n}^\tau)^2}\right) \quad a.s.$$

If K-means algorithm defined in Section 2.4 is applied to $\hat{\beta}_{in} = \sqrt{n/K}\widehat{u}_{1i}^\tau$. Then for sufficiently large n, we have

$$\sup_{1\le i\le n} 1\{\widehat{g}_i \neq g_i^0\} = 0 \quad a.s.$$

Theorem 3.2 indicates that the regularized spectral clustering, in conjunction with the K-means algorithm, consistently recovers the community membership for all nodes almost surely in large samples.

To see the effect of regularization, in the simplified case when $K^* = K$, K is fixed and σ_{Kn}^τ is bounded away from zero, Assumption 8 boils down to $\log(n)/\mu_n^\tau \to 0$. Even if $\min_i d_i$ grows slower than $\log(n)$ or does not grow to infinity at all, we can still choose τ with $\tau/\log(n) \to \infty$ such that Assumption 8 holds. This implies we can obtain strong consistency for some SBMs in which some nodes have very limited number of links. The following is a non-trivial SBM which does not satisfy Assumption 4 but satisfies Assumption 8.

Example 3.1. *Consider a SBM with two groups such that $n_1 = n_2 = n/2$ and*

$$B = \begin{pmatrix} 0.4 & 2/n \\ 2/n & 4/n \end{pmatrix}.$$

In this case, $d_i = 0.4(\frac{n}{2}-1) + \frac{2}{n}\cdot\frac{n}{2} = 0.2n + 0.6$ for node i in cluster 1 and $d_i = \frac{2}{n}\cdot\frac{n}{2} + \frac{4}{n}(\frac{n}{2}-1) = 3 - \frac{4}{n}$ for node i in cluster 2. Therefore, Assumption 4 does not hold. However, for some τ such that $\tau/\log n \to \infty$, we have

$$B^\tau = \begin{pmatrix} 0.4+\tau/n & (2+\tau)/n \\ (2+\tau)/n & (4+\tau)/n \end{pmatrix}$$

and $d_i^\tau = 0.2n + 0.6 + \tau(1 - n^{-1})$ *for node* i *in cluster 1 and* $d_i^\tau = 3 - 4n^{-1} + \tau(1 - n^{-1})$ *for node* i *in cluster 2. Thus* $\log(n)/\mu_n^\tau \to 0$*. In addition, it is easy to see that*

$$B_0^\tau = \begin{pmatrix} \frac{0.4+\tau n^{-1}}{0.2+(1+\tau)n^{-1}} & \frac{2+\tau}{[0.2n+(1+\tau)]^{1/2}(3+\tau)^{1/2}} \\ \frac{2+\tau}{[0.2n+(1+\tau)]^{1/2}(3+\tau)^{1/2}} & \frac{4+\tau}{3+\tau} \end{pmatrix} \to \begin{pmatrix} \frac{0.4+c_0}{0.2+c_0} & \sqrt{\frac{c_0}{0.2+c_0}} \\ \sqrt{\frac{c_0}{0.2+c_0}} & 1 \end{pmatrix},$$

when $c_0 = \lim_{n\to\infty} \tau/n \in [0, 1)$*. Apparently,* B_0^τ *has full rank and Assumption 8 holds. Therefore, the strong consistency of the regularized spectral clustering still holds.*

The previous example illustrates that the regularization works for the case when one cluster has strong links and the other one has weak links. However, if both clusters have weak links, it is hard to separate them. Consider the above example with B replaced by

$$B = \begin{pmatrix} 4/n & 2/n \\ 2/n & 4/n \end{pmatrix},$$

and $\tau/\log(n) \to \infty$. Then we can verify that

$$B_0^\tau = \begin{pmatrix} (4+\tau)/(3+\tau) & (2+\tau)/(3+\tau) \\ (2+\tau)/(3+\tau) & (4+\tau)/(3+\tau) \end{pmatrix}$$

such that B_0^τ has two eigenvalues given by 2 and $2/(3+\tau)$. But Assumption 8 cannot be satisfied in this case because $\mu_n^\tau |\sigma_{K^*n}^\tau|^4/\log(n)$ is converging to zero at rate $1/(\tau^3\log(n))$. Consequently, we cannot show that $\sup_i \sqrt{n}\|\widehat{u}_{1i}^\tau - (O_n^\tau)^T u_{1i}^\tau\| = o_{a.s.}(1)$ or prove strong consistency in this case. In general, the regularization may not work for the case in which we have multiple clusters with weak links.

3.2. *Regularized spectral clustering analysis for degree-corrected SBMs.* In this subsection, we extend our early analyses to the spectral clustering for a degree-corrected stochastic block model (DC-SBM).

3.2.1. *Degree-corrected SBMs.* Since [10], degree-corrected SBMs have become widely used in communication detection. The major advantage of a DC-SBM lies in the fact that it allows variation in node degrees within a community while preserving the overall block community structure. Given the K communities, the edge between nodes i and j are chosen independently with probability depending on the communities that nodes i and j

belong to. In particular, for nodes i and j belonging to clusters C_{k_1} and C_{k_2}, respectively, the probability of edge between i and j is given by

$$P_{ij} = \theta_i \theta_j B_{k_1 k_2},$$

where the block probability matrix $B = \{B_{k_1 k_2}\}$, $k_1, k_2 = 1, \ldots, K$, is a symmetric matrix with each entry between $[0, 1]$. The $n \times n$ edge probability matrix $P = \{P_{ij}\}$ represents the population counterpart of the adjacency matrix A. We continue to use $Z = \{Z_{ik}\}$ to denote the cluster membership matrix for all n nodes. Let $\Theta = \text{diag}(\theta_1, \ldots, \theta_n)$. Then we have

$$P = \Theta Z B Z^T \Theta^T.$$

Note Θ and B are only identifiable up to scale. We adopt the following normalization rule:

$$\sum_{i \in C_k} \theta_i = n_k, \quad k = 1, \ldots, K. \tag{3.1}$$

Alternatively, one can follow the literature (e.g., [16, 23]) and apply the following normalization $\sum_{i \in C_k} \theta_i = 1, \quad k = 1, \ldots, K$. We use the normalization in (3.1) because it nests the standard SBM as a special case in which $\theta_i = 1$ for $i = 1, \ldots, n$.

We first observe that, if we regularize both the adjacency matrix A and the degree matrix D, we are unable to preserve the DC-SBM structure unless Θ is homogeneous. To see this, note that when A is regularized to $A_\tau = A + \tau n^{-1} \iota_n \iota_n^T$, its population counterpart is

$$P_\tau = P + \tau n^{-1} \iota_n \iota_n^T = \Theta Z B Z^T \Theta + \tau n^{-1} Z \iota_k \iota_k^T Z.$$

Since Θ does not have the block structure, we are unable to find a $K \times K$ matrix B^τ and a $n \times n$ diagonal matrix Θ^τ such that

$$P_\tau = \Theta^\tau Z B^\tau Z^T \Theta^\tau.$$

For this reason, we follow the lead of [16] and only regularize the degree matrix D as $D_\tau = D + \tau I_n$. To differentiate from the regularized graph Laplacian L_τ considered in [9], we denote the new regularized graph Laplacian as

$$L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2},$$

and its population counterpart as

$$\mathcal{L}'_\tau = \mathcal{D}_\tau^{-1/2} P \mathcal{D}_\tau^{-1/2},$$

where $P = \Theta Z B Z^T \Theta$, $\mathcal{D}_\tau = \mathcal{D} + \tau I_n$, and $\mathcal{D} = \text{diag}(d_1, \ldots, d_n)$ with $d_i = \sum_{j=1}^n P_{ij}$.

3.2.2. *Identification of the group membership.* Let π_{kn}, W_k, $\mathcal{D}_B$ and B_0 be as defined in Section 2.2. To facilitate the asymptotic study, we now strengthen Assumption 1 to:

ASSUMPTION 9. *(i) There exists some sequence ρ_n such that $\rho_n \geq 1$ and $B_0 \leq \rho_n$ element-wise. (ii) B_0 has full rank K.*

It is possible to relax Assumption 9(ii) as we have done in the previous section. Here, we assume full rank of B_0 for simplicity and clarity.

As before, we consider the spectral decomposition of $\mathcal{L}'_\tau$:

$$\mathcal{L}'_\tau = U_{1n}\Sigma_n U_{1n}^T,$$

where $\Sigma_n = \text{diag}(\sigma_{1n}, \ldots, \sigma_{Kn})$ is a $K \times K$ matrix that contains the eigenvalues of $\mathcal{L}_\tau$ such that $|\sigma_{1n}| \geq |\sigma_{2n}| \geq \cdots \geq |\sigma_{Kn}| > 0$ and $U_{1n}^T U_{1n} = I_K$. Note that we suppress the dependence of U_{1n} and Σ_n on τ. Let $\Theta_\tau = \text{diag}(\theta_1^\tau, \ldots, \theta_n^\tau)$ where $\theta_i^\tau = \theta_i d_i/(d_i+\tau)$ for $i = 1, \ldots, n$. Let $n_k^\tau = \sum_{i \in C_k} \theta_i^\tau$.

THEOREM 3.3. *Suppose Assumptions 9 holds and let g_i^0 and u_i^T be the node i's true community identity and the i-th row of U_{1n}, respectively. Then (i) there exists a $K \times K$ matrix S_n^τ such that $U_{1n} = \Theta_\tau^{1/2} Z(Z^T\Theta_\tau Z)^{-1/2} S_n^\tau$, (ii) $(n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|u_i^T\| = 1$, and (iii) if $z_i = z_j$, then $\|\frac{u_i}{\|u_i\|} - \frac{u_j}{\|u_j\|}\| = 0$; if $z_i \neq z_j$, then $\|\frac{u_i^T}{\|u_i^T\|} - \frac{u_j^T}{\|u_j^T\|}\| = \sqrt{2}$.*

Like Lemma 3.3 in [16], Theorem 3.3(iii) provides useful facts about the rows of U_{1n}. First, if two nodes i and j belong to the same cluster, then the corresponding rows of U_{1n} point to the same direction so that $u_i/\|u_i\| = u_j/\|u_j\|$. Second, if two nodes i and j belong to the different clusters, then the corresponding rows of U_{1n} are orthogonal to each other. As a result, we can detect the community membership based on a feasible version of $\{u_i/\|u_i\|\}$.

3.2.3. *Uniform consistency of the estimated eigenvectors and strong consistency of the spectral clustering.* To proceed, we add the following assumptions.

ASSUMPTION 10. *There exist two constants C and c such that*

$$\infty > C \geq \limsup_n \sup_{1\leq i\leq n} n_{g_i^0}^\tau d_i^\tau K/(nd_i) \geq \liminf_n \inf_{1\leq i\leq n} n_{g_i^0}^\tau d_i^\tau K/(nd_i) \geq c > 0.$$

ASSUMPTION 11. *Denote* $\mu_n = \min_i d_i$, $\mu_n^\tau = \mu_n + \tau$, $\overline{\theta} = \max_i \theta_i$, *and* $\underline{\theta} = \min_i \theta_i$. *Then (i)*

$$\left(\frac{\log^{1/2}(n) K^{3/2}}{(\mu_n^\tau)^{1/2}\,|\sigma_{Kn}|}\right)\left(\frac{\overline{\theta}^{1/4}\rho_n^{1/2}}{\underline{\theta}^{1/4}} + \frac{\rho_n + K^{1/2}}{|\sigma_{Kn}|}\right) \to 0,$$

and (ii) there exists a positive constant c *such that* $\underline{\theta} \geq n^{-c}$.

Assumption 10 holds for the simplest case in which the degrees are homogeneous within the same cluster. Note in this case, $n^\tau_{g_i^0} = n_{g_i^0} d_i / d_i^\tau$, which may be of smaller order of magnitude of n/K if $d_i/\tau \to 0$. However, Assumption 10 still holds because the factor d_i/d_i^τ is removed. In general, Assumption 10 holds if d_i is of the same order of magnitude for all i in the same cluster.

Assumption 11 specifies conditions on d_i, θ_i, and σ_{Kn}. If $0 < \underline{\theta} \leq \overline{\theta} < \infty$, then Assumption 11(i) reduces to Assumption 6 with $K^* = K$ and $\xi_n = \sqrt{2}$. If in addition, K is fixed and $\liminf_n |\sigma_{Kn}| > 0$, then Assumption 11(i) further boils down to $\log(n)/\mu_n^\tau \to 0$. This indicates that even if the minimal degree μ_n is bounded, Assumption 11(i) still holds if $\tau/\log(n) \to \infty$.

Consider the spectral decomposition

$$L'_\tau = \widehat{U}_n \widehat{\Sigma}_n \widehat{U}_n^T = \widehat{U}_{1n} \widehat{\Sigma}_{1n} \widehat{U}_{1n}^T + \widehat{U}_{2n} \widehat{\Sigma}_{2n} \widehat{U}_{2n}^T,$$

where $\widehat{\Sigma}_n = \operatorname{diag}(\hat{\sigma}_n, \ldots, \hat{\sigma}_{nn}) = \operatorname{diag}(\widehat{\Sigma}_{1n}, \widehat{\Sigma}_{2n})$ with $|\hat{\sigma}_{1n}| \geq |\hat{\sigma}_{2n}| \geq \cdots \geq |\hat{\sigma}_{nn}| \geq 0$, $\widehat{\Sigma}_{1n} = \operatorname{diag}(\hat{\sigma}_{1n}, \ldots, \hat{\sigma}_{Kn})$, $\widehat{\Sigma}_{2n} = \operatorname{diag}(\hat{\sigma}_{K+1,n}, \ldots, \hat{\sigma}_{nn})$, and $\widehat{U}_n = (\widehat{U}_{1n}, \widehat{U}_{2n})$ is the corresponding eigenvectors such that $\widehat{U}_{1n}^T \widehat{U}_{1n} = I_K$ and $\widehat{U}_{2n}^T \widehat{U}_{1n} = 0$.

The following lemma parallels Lemma 2.1.

LEMMA 3.1. *If Assumptions 9–11 hold, then there exist a positive constant* C *sufficiently large and a* $K \times K$ *orthogonal matrix* O_n *such that*

$$\|\mathcal{L}'_\tau - L'_\tau\| \leq C(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.$$

and

$$\|\widehat{U}_{1n} - U_{1n} O_n\| \leq C(\log(n)/\mu_n^\tau)^{1/2} |\sigma_{Kn}|^{-1} K^{1/2} \quad a.s.$$

Let $\widehat{\Lambda} = L_\tau \widehat{U}_{1n} = \widehat{U}_{1n} \widehat{\Sigma}_n$, $\Lambda = \mathcal{L}_\tau U_{1n} O_n = U_{1n} \Sigma_n O_n$, $\widehat{\Lambda}_i = \widehat{u}_i^T \widehat{\Sigma}_n$, and $\Lambda_i = u_i^T \Sigma_n O_n$, where $\widehat{u}_i^T$ and u_i^T are the i-th rows of $\widehat{U}_{1n}$ and U_{1n}, respectively. In order to obtain the strong consistency, we need to derive the uniform bound for $\|\widehat{u}_i^T - u_i^T O_n\|$.

THEOREM 3.4. *If Assumptions 9–11 hold, then*

$$\sup_i (n^{\tau}_{g_i^0})^{1/2} (\theta_i^{\tau})^{-1/2} \|\widehat{u}_i - O_n^T u_i\| \leq C\eta_n \quad a.s.,$$

where $\eta_n = \left(\frac{\log^{1/2}(n)}{(\mu_n^{\tau})^{1/2}|\sigma_{Kn}|}\right)\left(\frac{\overline{\theta}^{1/4}\rho_n^{1/2}}{\underline{\theta}^{1/4}} + \frac{\rho_n + K^{1/2}}{|\sigma_{Kn}|}\right)$.

Theorem 3.4 is essential to establish the strong consistency result. As a corollary, we have:

COROLLARY 3.1. *If Assumptions 9–11 hold, then*

(3.2)
$$\sup_i \left\| \frac{\widehat{u}_i}{\|\widehat{u}_i\|} - \frac{O_n^T u_i}{\|O_n^T u_i\|} \right\| \leq \left(\frac{\log^{1/2}(n)}{(\mu_n^{\tau})^{1/2}|\sigma_{Kn}|}\right)\left(\frac{\overline{\theta}^{1/4}\rho_n^{1/2}}{\underline{\theta}^{1/4}} + \frac{\rho_n + K^{1/2}}{|\sigma_{Kn}|}\right) \quad a.s.$$

In addition, if the K-means algorithm is applied to $\hat{\beta}_{in} = \widehat{u}_{1i}/\|\widehat{u}_{1i}\|$, *then*

$$\sup_{1\leq i\leq n} \mathbf{1}\{\widehat{g}_i \neq g_i^0\} = 0 \quad a.s.$$

Corollary 3.1 justifies the use of K-means algorithm on $\widehat{u}_{in}/\|\widehat{u}_{in}\|$ provided the bound on the right hand side of (3.2) is $o\left(1/K^{3/2}\right)$, which is ensured by Assumption 11(i).

3.2.4. *An adaptive procedure.* Given the strong consistency of the spectral clustering, it is possible to consistently estimate Θ by some estimator, namely $\hat{\Theta}$. Built upon $\hat{\Theta}$, we propose an adaptive procedure by spectral clustering a new regularized graph Laplacian denoted as L''_{τ}, which is defined as

$$L''_{\tau} = (D''_{\tau})^{-1/2} A''_{\tau} (D''_{\tau})^{-1/2},$$

where $A''_{\tau} = A + \tau n^{-1}\hat{\Theta}\iota_n\iota_n^T\hat{\Theta}$ and $D''_{\tau} = \mathrm{diag}(A''_{\tau}\iota_n)$. The population counterpart of L''_{τ} is denoted as $\mathcal{L}''_{\tau}$ and defined as

$$\mathcal{L}''_{\tau} = (\mathcal{D}''_{\tau})^{-1/2} P''_{\tau} (\mathcal{D}''_{\tau})^{-1/2},$$

where $P''_{\tau} = P + \tau n^{-1}\Theta\iota_n\iota_n^T\Theta = \Theta Z B''_{\tau} Z^T \Theta$, $B''_{\tau} = B + \tau n^{-1}\iota_k\iota_k^T$, and $\mathcal{D}''_{\tau} = \mathrm{diag}(P''_{\tau}\iota_n) = D + \tau\Theta$.

Provided $\hat{\Theta}$ is consistent, we conjecture that one can show the adaptive procedure is strongly consistent by applying the same proof strategy as used in the derivation of strong consistency of the spectral clustering based on L_{τ} and L'_{τ}. We leave this important extension for future research. In the following, we focus on establishing the consistency of $\hat{\Theta}$.

Given the estimated group membership $\{\hat{g}_i\}_{i=1}^n$, we follow [21] and estimate Θ by $\hat{\Theta} = \text{diag}(\hat{\theta}_1, \cdots, \hat{\theta}_n)$, where

$$\hat{\theta}_i = \hat{n}_{\hat{g}_i}(\sum\nolimits_{j=1}^n A_{ij})/(\sum\nolimits_{i':\hat{g}_{i'}=\hat{g}_i} \sum\nolimits_{j=1}^n A_{i'j}) \tag{3.3}$$

and $\hat{n}_k = \#\{i : \hat{g}_i = k\}$. Next, we show $\hat{\theta}_i \to \theta_i$ a.s. uniformly in $i = 1, \cdots, n$.

ASSUMPTION 12. *(i)* $\limsup_n \overline{\theta} < \infty$. *(ii)* $\sup_{1\leq i\leq n} \mathbf{1}\{\widehat{g}_i \neq g_i^0\} = 0$ *a.s.*

Assumption 12(i) requires that the degree of heterogeneity is bounded, which is common in practical applications. Assumption 12(ii) requires the preliminary clustering is strongly consistent. For instance, this assumption can be verified by Corollary 3.1. However, we also allow for any other strongly consistent clustering methods, such as the conditional pseudo likelihood method proposed by [1].

Let $m_k = \sum_{j=1}^n \theta_j B_{kg_j^0}$ and $\underline{m}_n = \inf_k m_k$. Note $m_k = \sum_{i'\in C_k} d_{i'}/n_k$ is the average degree of nodes in community k and $\underline{m}_n$ is the minimal average degree.

THEOREM 3.5. *If Assumption 12 holds, then* $\sup_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| = O_{a.s.}(\log(n)/\underline{m}_n)$.

In order for $\hat{\Theta}$ to be consistent, we need the average degree for each community to grow faster than $\log(n)$. In some cases, the average degree and the minimal degree are of the same order of magnitude. Then we basically need $\mu_n/\log(n) \to \infty$ for the consistency of $\hat{\Theta}$. In our simulation designs, $\mu_n/\log(n) \to 0$, which is, in some sense, the worst case for the adaptive procedure. However, even in this case, the performance of the adaptive procedure improves upon that of the spectral clustering based on L'_τ.

4. Numerical Examples on Simulated Networks.

In this section, we consider the finite sample performance of spectral clustering with two and three communities, i.e., $K = 2$ and $K = 3$. The corresponding numbers of community members have ratio $1 : 1$ and $1 : 1 : 1$ for these two cases, respectively. The number of nodes is given by 50 and 200 for each community, which indicates $n = 100$ and 400 for the case of $K = 2$ and 150 and 600 for the case of $K = 3$. We use four variants of graph Laplacian to conduct the spectral clustering, namely, L, L_τ, L'_τ, and L''_τ defined in Sections 2 and 3.

1. $L = D^{-1/2}AD^{-1/2}$ where $D = \text{diag}(A\iota_n)$. It is possible that for some realizations, the minimum degree is 0, yielding singular D.

2. $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$ where $A_\tau = A + \tau J_n$, $D_\tau = \text{diag}(A_\tau \iota_n)$, and $J_n = n^{-1}\iota_n \iota_n^T$.
3. $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ where $D_\tau = D + \tau I_n$ and I_n is an $n \times n$ identity matrix.
4. $L''_\tau = (D''_\tau)^{-1/2} A''_\tau (D''_\tau)^{-1/2}$ where $A''_\tau = A + \tau n^{-1} \hat{\Theta} \iota_n \iota_n^T \hat{\Theta}$ and $D''_\tau = \text{diag}(A''_\tau \iota_n)$.

The theoretical results in Sections 2 and 3 suggest the strong consistency of the spectral clustering with L_τ and L'_τ for the standard SBM and DC-SBM, respectively under some conditions. In Sections 4.1 and 4.2, we consider these two cases. In addition, for the DC-SBM, we will also consider the adaptive procedure introduced in Subsection 3.2.4. Additional simulation results of spectral clustering with L and L'_τ for the standard SBM and L and L_τ for the DC-SBM can be found in the supplementary Appendix D.

For the standard SBM, after obtaining the eigenvectors corresponding to the largest K eigenvalues of the graph Laplacian (L, L_τ, L'_τ, or L''_τ), we classify them based on K-means algorithm (Matlab "kmedoids" function, which is more robust to noise and outliers than "kmeans" function, with default options). For the DC-SBM, before classification, we normalize each row of the $n \times K$ eigenvectors so that its L_2 norm equals 1. For comparison, we apply the unconditional pseudo-likelihood method (UPL) and conditional pseudo-likelihood method (CPL) proposed by [1] to detect the communities in the SBM and the DC-SBM, respectively.[2] To evaluate the classification performance, we consider two criteria: the Correct Classification Proportion (CCP) and the Normalized Mutual Information (NMI).

4.1. *The standard SBM.* We consider two data generating processes (DGPs).
DGP 1: Let $K = 2$. Each community has $n/2$ nodes. The matrix B is set as

$$B = \frac{2}{n} \begin{pmatrix} \log^2(n) & 0.2\log(n) \\ 0.2\log(n) & 0.8\log(n) \end{pmatrix}.$$

The expected degrees are of order $\log^2(n)$ and $\log(n)$ respectively for communities 1 and 2.
DGP 2: Let $K = 3$. Each community has $n/3$ nodes. The matrix B is set

[2]As [1] remark, the UPL and CPL are correctly fitting the SBM and the DC-SBM, respectively. In both UPL and CPL, the initial classification is generated by spectral clustering with perturbations (SCP). The SCP is spectral clustering based on L_τ with $\tau = \bar{d}/4$ and $\bar{d}$ being the average degree.

as

$$B=\frac{3}{n}\begin{pmatrix} n^{1/2} & 0.1\log^{5/6}(n) & 0.1\log^{5/6}(n)\\ 0.1\log^{5/6}(n) & \log^{3/2}(n) & 0.1\log^{5/6}(n)\\ 0.1\log^{5/6}(n) & 0.1\log^{5/6}(n) & 0.8\log^{5/6}(n)\end{pmatrix}.$$

The expected degrees are of order $n^{1/2}$, $\log^{3/2}(n)$ and $\log^{5/6}(n)$ respectively for communities 1, 2 and 3.

We follow [9] and select the regularizer τ that minimizes a feasible version of

$$\|L_\tau-\mathcal{L}_\tau\|/|\sigma^\tau_{Kn}|.$$

In particular, for a given τ, we can obtain the community identities $\hat{Z}$ based on the spectral clustering of L_τ. Given $\hat{Z}$, we can estimate the block probability matrix B by the fraction of links between the estimated communities, which is denoted as $\hat{B}$. Let $\hat{P}=\hat{Z}\hat{B}\hat{Z}^T$, $\hat{P}_\tau=\hat{P}+\tau J_n$, $\hat{\mathcal{D}}_\tau=\operatorname{diag}(\hat{P}_\tau\iota_n)$, $\hat{\mathcal{L}}_\tau=\hat{\mathcal{D}}_\tau^{-1/2}\hat{P}_\tau\hat{\mathcal{D}}_\tau^{-1/2}$, and $\hat{\sigma}^\tau_{Kn}$ be the K-th largest in absolute value eigenvalue of $\hat{\mathcal{L}}_\tau$. Then we can compute

$$Q(\tau)=\|L_\tau-\hat{\mathcal{L}}_\tau\|/\hat{\sigma}^\tau_{Kn}.$$

We search for some τ^{JY} that minimizes $Q(\tau)$ over a grid of 20 points, τ_j, on the interval $[\tau_{\min},\tau_{\max}]$, where $j=1,\ldots,20$, $\tau_{\min}=10^{-4}$ and $\tau_{\max}$ is set to be the expected average degree. We set $\tau_1=\tau_{\min}$, $\tau_2=1$, and $\tau_{j+2}=(\tau_{\max})^{j/18}$ for $j=1,\ldots,18$. [16] suggested choosing τ as the average degree of nodes, which is approximately equal to the expected average degree.

All results reported here are based on 500 replications. For DGPs 1 and 2, we report the classification results based on $L_\tau=D_\tau^{-1/2}A_\tau D_\tau^{-1/2}$ in Figures 1 and 2. The results based on L and L'_τ are relegated to the supplementary Appendix D.

In Figures 1 and 2, the first and second rows correspond to the results with $n=100$ and $n=400$, respectively. For each replication, we can compute the feasible τ^{JY} as mentioned above. Their averages across all replications are reported in each subplot of Figures 1 and 2. In particular, the green dashed line represents τ^{JY}, which can be easily compared with the expected average degree, the rightmost vertical border.

We summarize our findings from Figures 1 and 2. First, despite the fact that the minimal degrees for neither DGP satisfies Assumption 4 so that the standard spectral clustering may not be consistent, the regularized spectral clustering performs quite well in both DGPs. This confirms our theoretical finding that the regularization can help to relax the requirement on the minimal degree and to achieve the strong consistency. In addition, when a proper τ is used, the spectral clustering based on L_τ outperforms the UPL method of [1]. Both results are in line with the theoretical analysis by [9].

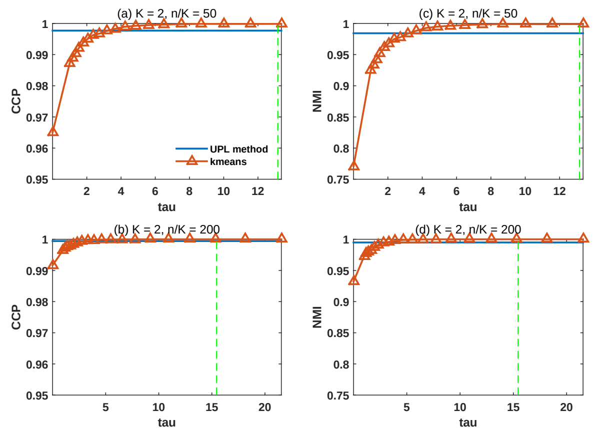

Fig 1: Classification results for K-means for DGP 1 ($K = 2$) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$ and for UPL method. The x-axis marks τ values, and the y-axis is either CCP (left column) or NMI (right column). The green vertical line in each subplot indicates the estimated τ value by using the method of [9]. The first and second rows correspond to $n/K = 50$ and 200, respectively.

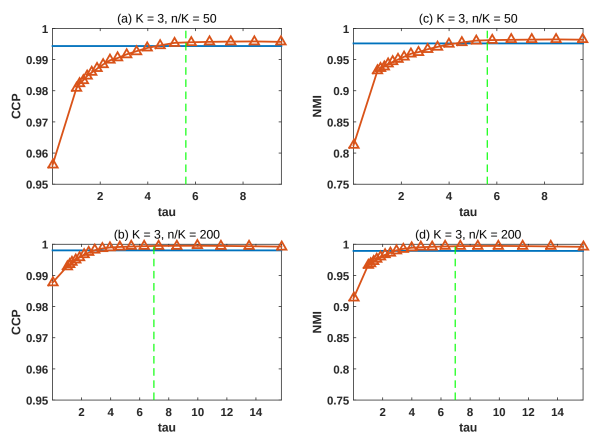

Fig 2: Classification results for DGP 2 ($K = 3$) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. (See the explanations in Figure 1.)

4.2. *The DC-SBM.* The next two DGPs consider the degree-corrected SBM.

DGP 3: This DGP is the same as DGP 1 except that here $P = \Theta Z B Z^T \Theta^T$, where Θ is a diagonal matrix with each diagonal element taking a value from $\{0.5, 1.5\}$ with equal probability.

DGP 4: This one is the same as DGP 2 except that here $P = \Theta Z B Z^T \Theta^T$ and Θ is generated as in DGP 3.

To compute the feasible regularizer for the DC-SBM, we modify the previous procedure to incorporate the degree heterogeneity. In particular, given τ, by spectral clustering L'_τ, we can obtain a classification $\hat{Z} = (\hat{Z}_1, \ldots, \hat{Z}_n)^T$, where $\hat{Z}_i$ is a K by 1 vector with its $\hat{g}_i$th entry being 1 and the rest being 0 and $\hat{g}_i$ is an estimator of node i's community membership. Let $\hat{n}_k = \#\{i : \hat{g}_i = k\}$. Then we can estimate the block probability matrix B and Θ by $\hat{B} = [\hat{B}_{kl}]_{1 \le k,l \le K}$ and $\hat{\Theta} = \text{diag}(\hat{\theta}_1, \ldots, \hat{\theta}_n)$, where $\hat{\theta}_i$ is defined in (3.3) and $\hat{B}_{kl} = (\sum_{(i,j):\hat{g}_i = k, \hat{g}_j = l} A_{ij})/(\hat{n}_k \hat{n}_l)$. Let $\hat{P} = \hat{\Theta}\hat{Z}\hat{B}\hat{Z}^T\hat{\Theta}^T$, $\hat{\mathcal{D}}_\tau = \text{diag}(\hat{P}\iota_n) + \tau I_n$, and $\hat{\mathcal{L}}'_\tau = \hat{\mathcal{D}}_\tau^{-1/2}\hat{P}\hat{\mathcal{D}}_\tau^{-1/2}$. Let $\hat{\sigma}'^{\tau}_{Kn}$ denote the K-th largest eigenvalue of $\hat{\mathcal{L}}'_\tau$ (in absolute value). Let

$$Q'(\tau) = \|L'_\tau - \hat{\mathcal{L}}'_\tau\| / \hat{\sigma}'^{\tau}_{Kn}.$$

We search for some τ'^{JY} that minimizes $Q'(\tau)$ over the same aforementioned grid.

For DGPs 3 and 4, we report the classification results based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ as the orange lines in Figures 3 and 4. For each subplot, the rightmost border line and the red vertical line represent the averages of $\bar{d}$ and τ'^{JY}, respectively. Figures 3 and 4 show the regularized spectral clustering based on L'_τ is slightly outperformed by CPL in DC-SBMs. However, τ'^{JY} has the close-to-optimal performance in terms of both CCP and NMI over a range of values for τ.

Table 1 reports the classification results for the spectral clustering with $\tau = \tau^{\text{JY}}$ for DGPs 1–2 (or τ'^{JY} for DGPs 3–4) and $\bar{d}$ in comparison with those for the UPL (or CPL for DGPs 3–4) method over 500 replications. In general, the spectral clustering with $\tau = \tau^{\text{JY}}$ outperforms the UPL method in DGPs 1–2 but slightly underperforms the CPL method for DGPs 3 and 4. In all cases, we observe that the increase of the probability of correct classification as n increases. This is consistent with the theory because both the UPL/CPL method and our regularized spectral clustering method are strongly consistent.

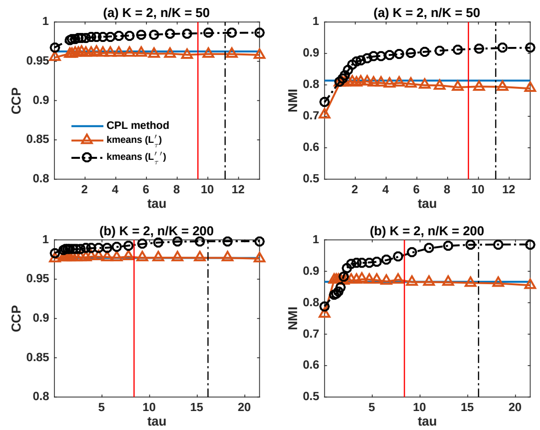

Fig 3: Classification results for DGP 3 ($K = 2$, degree-corrected) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ and $L''_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. The red and black vertical lines correspond to the optimal regularizers τ'^{JY} and τ''^{JY}, respectively. (See Figure 1 for the explanation of other features of the figure.)

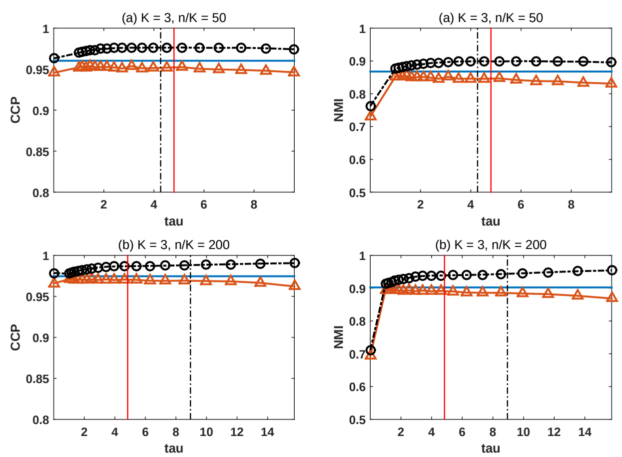

Fig 4: Classification results for DGP 4 ($K = 3$, degree-corrected) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ and $L''_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. The red and black vertical lines correspond to the optimal regularizers τ'^{JY} and τ''^{JY}, respectively. (See Figure 1 for the explanation of other features of the figure.)

TABLE 1
Comparison of classification results based on spectral clustering with τ set to τ^{JY} or $\bar{d}$, and the UPL/CPL method

			CCP			NMI		
			Spectral clustering		UPL/CPL	Spectral clustering		UPL/CPL
DGP	K	n/K	$\bar{d}$	$\tau^{\mathrm{JY}}/\tau'^{\mathrm{JY}}$		$\bar{d}$	$\tau^{\mathrm{JY}}/\tau'^{\mathrm{JY}}$	
1	2	50	0.9999	0.9999	0.9977	0.9990	0.9990	0.9843
	2	200	1.0000	1.0000	0.9994	1.0000	1.0000	0.9949
2	3	50	0.9905	0.9957	0.9943	0.9693	0.9820	0.9760
	3	200	0.9945	0.9994	0.9980	0.9847	0.9968	0.9893
3	2	50	0.9592	0.9619	0.9625	0.7918	0.7942	0.8136
	2	200	0.9756	0.9779	0.9770	0.8531	0.8690	0.8667
4	3	50	0.9468	0.9534	0.9603	0.8320	0.8414	0.8678
	3	200	0.9640	0.9702	0.9745	0.8710	0.8898	0.9023

4.3. *The adaptive procedure.* Next we consider the adaptive procedure for DC-SBM. The procedure contains two steps:

1. Following [1], we adopt the SCP to get a preliminary estimate of the community structure. Based on this preliminary estimate, we estimate Θ by $\hat{\Theta}$ as in (3.3).

2. With the estimated $\hat{\Theta}$ from the previous step, we construct $A''_\tau = A + \tau\hat{\Theta}J_n\hat{\Theta}^T$, $D''_\tau = \mathrm{diag}(A''_\tau \iota_n)$, and $L''_\tau = (D''_\tau)^{-1/2}A''_\tau(D''_\tau)^{-1/2}$.[3] For a given τ, by spectral clustering L''_τ, we can obtain a new classification $\hat{Z} = (\hat{Z}_1, \dots, \hat{Z}_n)^T$, where $\hat{Z}_i$ is a K by 1 vector with its $\hat{g}_i$th entry being 1 and the rest being 0 and $\hat{g}_i$ is an estimator of node i's community membership. Let $\hat{n}_k = \#\{i : \hat{g}_i = k\}$. Then, based on the new classification, we can estimate the block probability matrix B and Θ by $\hat{B}_{kl} = (\sum_{(i,j):\hat{g}_i=k,\hat{g}_j=l} A_{ij})/(\hat{n}_k\hat{n}_l)$ and $\hat{\theta}_i = \hat{n}_{\hat{g}_i}(\sum_{j=1}^n A_{ij})/(\sum_{i':\hat{g}_{i'}=\hat{g}_i}\sum_{j=1}^n A_{i'j})$. Given $\hat{B} = [\hat{B}_{kl}]_{1\le k,l\le K}$ and $\hat{\Theta}_1 = \mathrm{diag}(\hat{\theta}_1, \dots, \hat{\theta}_n)$, we compute $\hat{P} = \hat{\Theta}_1\hat{Z}\hat{B}\hat{Z}^T\hat{\Theta}_1^T$, $\hat{P}''_\tau = \hat{P} + \tau\hat{\Theta}J_n\hat{\Theta}^T$, $\hat{\mathcal{D}}_\tau = \mathrm{diag}(\hat{P}''_\tau\iota_n)$, and $\hat{\mathcal{L}}''_\tau = \hat{\mathcal{D}}_\tau^{-1/2}\hat{P}''_\tau\hat{\mathcal{D}}_\tau^{-1/2}$. Let $\hat{\sigma}''^{\tau}_{Kn}$ denote the K-th largest eigenvalue of $\hat{\mathcal{L}}''_\tau$ (in absolute value). Let

$$Q''(\tau) = \|L''_\tau - \hat{\mathcal{L}}''_\tau\|/\hat{\sigma}''^{\tau}_{Kn}.$$

We search for some τ''^{JY} that minimizes $Q''(\tau)$ over the aforementioned grid.

Figures 3 and 4 also report the classification results based on L''_τ, which are shown as the dark lines. We find the performance of spectral clustering based on L''_τ is better than those using the CPL method. In addition, our choice of τ''^{JY}, marked as the dark vertical line in each subplot, performs well in both DGPs 3 and 4.

[3]In practice, we let $A''_\tau = A + \tau\hat{\Theta}J_n\hat{\Theta}^T + 10^{-8}J_n$ to overcome the computation problem incurred by 0 degree nodes.

5. Conclusion. In this paper we study the strong consistency of spectral clustering for stochastic block models (SBMs). We first consider the standard SBMs and show that under some conditions on the minimal degree (μ_n), the number of communities (K), and the eigenvalues of the probability block matrix, the K-means algorithm applied to the eigenvectors of the graph Laplacian associated with its first few largest eigenvalues can classify all individuals into the true community uniformly correctly almost surely in large samples. In the special case where K is fixed and the probability block matrix has minimal eigenvalue bounded away from zero, the strong consistency essentially requires that $\mu_n/\log(n) \to \infty$ as $n \to \infty$, which is the minimal condition for the strong consistency discussed in [2]. We also consider the regularized spectral clustering for the standard SBMs and show that the regularization can greatly relax the above conditions and it only requires $(\mu_n + \tau)/\log(n) \to \infty$ as $n \to \infty$ in the aforementioned special case, where τ is the regularization parameter. The extension to the regularized spectral clustering for a degree-corrected SBM is also studied and we show that strong clustering can also be achieved in this case. Our simulations indicate that an adaptive procedure helps to improve the finite sample performance of the regularized spectral clustering for a degree-corrected SBM.

APPENDIX A: PROOFS OF THE RESULTS IN SECTION 2

In this appendix we prove the results in Section 2. We will apply some technical lemmas whose proofs are relegated to the supplement.

Proof of Theorem 2.1. By the proof of Rohe et al. [17, Lemma 3.1], we have $\mathcal{L} = n^{-1}ZB_0Z^T$. Therefore, $\mathcal{L}^2 = n^{-1}ZB_0(Z^TZ/n)B_0Z^T$. Let $\Pi_n = Z^TZ/n = \mathrm{diag}(\pi_{1n}, \ldots, \pi_{Kn})$. By the spectral decomposition in Assumption 1, we have

$$\Pi_n^{1/2}B_0\Pi_nB_0\Pi_n^{1/2} = S_n\Omega_n^2S_n^T, \tag{A.1}$$

where $\Omega_n = \mathrm{diag}(\omega_{1n}, \ldots, \omega_{K^*n})$ such that $|\omega_{1n}| \geq |\omega_{2n}| \geq \cdots \geq |\omega_{K^*n}| > 0$ and S_n is a $K \times K^*$ matrix such that $S_n^TS_n = I_{K^*}$. Let $U_{1n}^* = Z(Z^TZ)^{-1/2}S_n$. Then, we have

$$U_{1n}^*\Omega_n^2U_{1n}^{*T} = \mathcal{L}^2 = U_{1n}\Sigma_{1n}^2U_{1n}^T. \tag{A.2}$$

In addition, $U_{1n}^{*T}U_{1n}^* = S_n^TS_n = I_{K^*}$. Therefore the columns of U_{1n}^* are the eigenvectors of $\mathcal{L}$ associated with eigenvalues $\sigma_{1n}, \ldots, \sigma_{K^*n}$, up to sign normalization. W.l.o.g., we can take $U_{1n}^* = U_{1n}$ and $\Omega_n = \Sigma_{1n}$.

In addition, if node i is in cluster C_{k_1}, then $z_i^T(Z^TZ)^{-1/2}S_n = n_{k_1}^{-1/2}[S_n]_{k_1\cdot}$, where $[S_n]_{k\cdot}$ denotes the k-th row of S_n. Therefore, by Assumption 2 and the fact that $\|[S_n]_{k_1\cdot}\| \leq 1$,

$$(n/K)^{1/2}\|z_i^T(Z^TZ)^{-1/2}S_n\| \leq c^{-1/2}\|[S_n]_{k_1\cdot}\| \leq c^{-1/2}.$$

Taking $\sup_i$ on both sides establishes the first desired result.

Now, let $\xi_n = \min_{1\le i<j\le n}(\frac{n}{K})^{1/2}\|(z_i - z_j)(Z^TZ)^{-1/2}S_n\|$. Let nodes i and j be in communities k_1 and k_2, respectively, where $k_1 \neq k_2$. If $K^* = K$, $\|[S_n]_{k\cdot}\| = 1$ for $k = 1, \dots, K$ and $[S_n]_{k_1\cdot}^T[S_n]_{k_2\cdot} = 0$. Then noting that

$$\frac{n}{K}\|(z_i - z_j)(Z^TZ)^{-1/2}S_n\|^2 = \frac{n}{Kn_{k_1}} + \frac{n}{Kn_{k_2}} \ge \frac{2}{C},$$

we have $\xi_n \ge \sqrt{\frac{2}{C}} > 0$. This concludes part (i) in the second claim. For part (ii), by (A.1), $(Z^TZ)^{-1/2}S_n\Omega_n^2S_n^T(Z^TZ)^{-1/2} = n^{-1}B_0\pi_nB_0$. Noting that $z_i^T(Z^TZ)^{-1/2}S_n = n_{k_1}^{-1/2}[S_n]_{k_1\cdot}$ and $z_j^T(Z^TZ)^{-1/2}S_n = n_{k_2}^{-1/2}[S_n]_{k_2\cdot}$, this implies that

$$\begin{aligned}\sum_{k^*=1}^{K^*}&\omega_{k^*n}^2\left(n_{k_1}^{-1/2}[S_n]_{k_1k^*}\right)\left(n_{k_2}^{-1/2}[S_n]_{k_2k^*}\right)\\ &= z_i^T(Z^TZ)^{-1/2}S_n\Omega_nS_n^T(Z^TZ)^{-1/2}z_j = n^{-1}\sum_{k_3=1}^{K}\pi_{k_3n}[B_0]_{k_1k_3}[B_0]_{k_2k_3},\end{aligned}$$

$$\sum_{k^*=1}^{K^*}\omega_{k^*n}^2\left(n_{k_1}^{-1/2}[S_n]_{k_1k^*}\right)^2 = n^{-1}\sum_{k_3=1}^{K}\pi_{k_3n}[B_0]_{k_1k_3}^2,$$

and

$$\sum_{k^*=1}^{K^*}\omega_{k^*n}^2\left(n_{k_2}^{-1/2}[S_n]_{k_2k^*}\right)^2 = n^{-1}\sum_{k_3=1}^{K}\pi_{k_3n}[B_0]_{k_2k_3}^2.$$

It follows that

(A.3)

$$\sum_{k^*=1}^{K^*}\omega_{k^*n}^2\left(n_{k_1}^{-1/2}[S_n]_{k_1k^*} - n_{k_2}^{-1/2}[S_n]_{k_2k^*}\right)^2 = n^{-1}\sum_{k_3=1}^{K}\pi_{k_3n}([B_0]_{k_1k_3} - [B_0]_{k_2k_3})^2.$$

By (A.2), $\omega_{1n}^2 = \|\mathcal{L}^2\|$. In addition, $\|L\| \le 1$, and by Lemma 2.1 below which does not use the result in this theorem, $\|\mathcal{L} - L\| = o_{a.s.}(1)$. This implies $|\omega_{1n}| \le 2$. Then, by Assumption 1(ii) and (A.3),

$$\begin{aligned}4\xi_n^2 &= 4\min_{1\le k_1<k_2\le n}\frac{n}{K}\sum_{k^*=1}^{K^*}\left(n_{k_1}^{-1/2}[S_n]_{k_1k^*} - n_{k_2}^{-1/2}[S_n]_{k_2k^*}\right)^2\\ &\ge \min_{1\le k_1<k_2\le n}\frac{n}{K}\sum_{k^*=1}^{K^*}\omega_{k^*n}^2\left(n_{k_1}^{-1/2}[S_n]_{k_1k^*} - n_{k_2}^{-1/2}[S_n]_{k_2k^*}\right)^2\\ &= \min_{1\le k_1<k_2\le n}\frac{1}{K^2}\sum_{k_3=1}^{K}(\pi_{k_3n}K)([B_0]_{k_1k_3} - [B_0]_{k_2k_3})^2\\ &\ge \min_{1\le k_1<k_2\le n}\frac{c}{K^2}\|[B_0]_{k_1\cdot} - [B_0]_{k_2\cdot}\|^2 \ge c(\xi_n')^2,\end{aligned}$$

Then $\xi_n \ge c'\xi_n'$ with $c' = \sqrt{c}/2$ and the conclusion in part (ii) of the second claim follows. □

The proof of Lemma 2.1 is similar to the ones in [9] and [16]. We include it just for completeness. We will use the following Bernstein inequality of the matrix derived by [14].

LEMMA A.1. *Consider an independent sequence* $(Y_k)_{k\geq 1}$ *of real symmetric* $d\times d$ *random matrices that satisfy* $\mathbb{E}Y_k = 0$ *and* $\|Y_k\| \leq R$ *for each index* k. *Then, for all* $t \geq 0$ *and* $\sigma^2 = \|\sum_{k\geq 1}\mathbb{E}Y_k^2\|$,

$$P(\|\sum_{k\geq 1} Y_k\| \geq t) \leq d\exp\left(\frac{-t^2}{3\sigma^2 + 2Rt}\right).$$

PROOF OF LEMMA 2.1. Let $\tilde{L} = \mathcal{D}^{-1/2}A\mathcal{D}^{-1/2}$. Then

$$\|\mathcal{L} - L\| \leq \|\mathcal{L} - \tilde{L}\| + \|L - \tilde{L}\| =: I + II.$$

Let $Y_{ij} = (d_i d_j)^{-1/2}(A_{ij} - P_{ij})(e_i e_j^T + e_j e_i^T)$ for $1 \leq i < j \leq n$ and $Y_{ii} = -d_i^{-1}P_{ii}e_i e_i^T$, where e_i is the $n\times 1$ vector with its i-th coordinate being 1 and the rest being 0. Then, $\{Y_{ij}\}_{1\leq i\leq j\leq n}$ is a sequence of independent symmetric random matrices such that $\mathbb{E}Y_{ij} = 0$,

$$\tilde{L} - \mathcal{L} + \mathrm{diag}(\mathcal{L}) = \sum_{1\leq i<j\leq n} Y_{ij} \text{ and } \mathrm{diag}(\mathcal{L}) = \sum_{i=1}^{n} Y_{ii}.$$

In addition, we note that $\sup_{1\leq i<j\leq n}\|Y_{ij}\| \leq \sqrt{2}/\mu_n$ and

$$\begin{aligned}\sigma^2 = \|\sum_{1\leq i<j\leq n}\mathbb{E}Y_{ij}^2\| &= \|\mathrm{diag}(\sum_{j\neq 1} p_{1j}(1-p_{1j})/(d_1 d_j), \ldots, \sum_{j\neq n} p_{nj}(1-p_{nj})/(d_n d_j))\| \\ &\leq \mu_n^{-1}\max_{1\leq i\leq n}\sum_{j=1}^{n} p_{ij}(1-p_{ij})/d_i \leq \mu_n^{-1}.\end{aligned}$$

Therefore, by Lemma A.1 and Assumption 4, there exist some constant $C > 2$ and some integer n_0 sufficiently large, such that for $n > n_0$

$$\begin{aligned}P(\|\tilde{L} - \mathcal{L} + \mathrm{diag}(\mathcal{L})\| \geq C(\log(n)/\mu_n)^{1/2}) &= P(\|\sum_{1\leq i<j\leq n} Y_{ij}\| \geq C(\log(n)/\mu_n)^{1/2}) \\ &\leq n\exp\left(\frac{-C^2\log(n)/\mu_n}{3\mu_n^{-1} + 2\sqrt{2}C(\log(n)/\mu_n)^{1/2}\mu_n^{-1}}\right) \\ &\leq Cn^{1-C}. \qquad \text{(A.4)}\end{aligned}$$

This implies that

$$\sum_{n=1}^{\infty} P(\|\tilde{L} - \mathcal{L} + \mathrm{diag}(\mathcal{L})\| \geq C\log^{1/2}(n)\mu_n^{-1/2}) < \infty,$$

or equivalently, $\|\tilde{L}-\mathcal{L}+\text{diag}(\mathcal{L})\| \le C\log^{1/2}(n)\mu_n^{-1/2} \quad a.s.$ In addition, $\|\text{diag}(\mathcal{L})\| \le \mu_n^{-1} = o(\log^{1/2}(n)\mu_n^{-1/2})$. Therefore,

$$I \le \|\tilde{L}-\mathcal{L}+\text{diag}(\mathcal{L})\| + \|\text{diag}(\mathcal{L})\| \le C\log^{1/2}(n)\mu_n^{-1/2} \quad a.s.$$

Now we turn to II. By Bernstein inequality for independent bounded random variables, for some $C > 2$, we have,

$$\begin{aligned} P(\sup_i |\hat{d}_i - d_i|/d_i \ge C(\log(n)/\mu_n)^{1/2}) &\le 2\sum_{i=1}^n \exp\left(\frac{-C^2 d_i^2 \log(n)/\mu_n}{2d_i + 2C(\log(n)/\mu_n)^{1/2} d_i/3}\right) \\ &\le Cn^{1-C}, \end{aligned}$$

where the second line follows from the fact that $\min_i d_i = \mu_n$ and $\log(n)/\mu_n \to 0$ under Assumption 4. Therefore, $\sup_i |\hat{d}_i - d_i|/d_i \le C(\log(n)/\mu_n)^{1/2} \quad a.s.$, and we have

$$\|\mathcal{D}^{-1/2}D^{1/2} - I\| = \max_i |(\hat{d}_i/d_i)^{1/2} - 1| \le \max_i |(\hat{d}_i/d_i) - 1| \le C(\log(n)/\mu_n)^{1/2} \quad a.s.$$

Then by the triangle inequality and the fact that $\|L\| \le 1$,

$$\begin{aligned} \|\tilde{L} - L\| &= \|L - \mathcal{D}^{-1/2}D^{1/2}LD^{1/2}\mathcal{D}^{-1/2}\| \\ &\le \|\mathcal{D}^{-1/2}D^{1/2}L - \mathcal{D}^{-1/2}D^{1/2}LD^{1/2}\mathcal{D}^{-1/2}\| + \|L - \mathcal{D}^{-1/2}D^{1/2}L\| \\ &\le \|\mathcal{D}^{-1/2}D^{1/2} - I\|\|\mathcal{D}^{-1/2}D^{1/2}\| + \|\mathcal{D}^{-1/2}D^{1/2} - I\| \le C(\log(n)/\mu_n)^{1/2} \quad a.s. \end{aligned}$$

This concludes the first part of the proof. For the second part, by the Davis-Kahan Theorem (e.g., Yu et al. [22, Theorem 2]), we have

$$\|\widehat{U}_{1n} - U_{1n}O_n\| \le \frac{C(K^*)^{1/2}\|L - \mathcal{L}\|}{|\sigma_{K^*n}|} \le C\log^{1/2}(n)(K^*)^{1/2}\mu_n^{-1/2}|\sigma_{K^*n}^{-1}| \quad a.s.$$

□

To prove Theorem 2.2, we need the following three lemmas.

Lemma A.2. *Let $\rho_n = \max(\sup_{k_1,k_2}[B_0]_{k_1k_2}, 1)$. Then $P_{ij} \le \rho_n n^{-1}(d_i d_j)^{1/2}$.*

Lemma A.3. *Let V_{1n} be some $n\times K^*$ (random) matrix and v_{1i}^T be the i-th row of V_{1n}. Assume there exist two deterministic sequences $\{\phi_{1n}\}_{n\ge 1}$ and $\{\phi_{2n}\}_{n\ge 1}$ such that $\|V_{1n}\| \le \phi_{1n}$ and $\sup_i \|v_{1i}\| \le \phi_{2n}$ almost surely. In addition, if Assumptions 1–4 hold, then there exists some positive constant C sufficiently large such that*

$$\sup_i\left(d_i^{-1/2}\|([A]_{i\cdot} - [\mathcal{P}]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}\|\right) \le 2C\left[\frac{\phi_{2n}\log(n)K^*}{\mu_n} \vee \left(\frac{\log(n)\phi_{1n}^2\rho_n K^*}{n\mu_n}\right)^{1/2}\right] \quad a.s.$$

LEMMA A.4. *Assume there exists a deterministic sequence $\{\psi_n\}_{n\geq 1}$ such that $\sup_j \|\widehat{u}_{1j}\| \leq \psi_n$ almost surely. If Assumptions 1–4 hold, then there exists some positive constant C sufficiently large such that*

$$\|\widehat{\Lambda} - \Lambda\| \leq C \log^{1/2}(n)(K^*)^{1/2}\mu_n^{-1/2}|\sigma_{K^*n}^{-1}| \quad a.s.$$

and

$$\sup_i \|\widehat{\Lambda}_i - \Lambda_i\| \leq C\log(n)K^*\mu_n^{-1}\psi_n + C(\log(n)K^*)^{1/2}\rho_n(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}| \quad a.s.$$

Proof of Theorem 2.2. First, by the Weilandt-Hoffman inequality and Lemma 2.1,

$$\|\widehat{\Sigma}_{1n} - \Sigma_{1n}\| \leq \|L - \mathcal{L}\| \leq C\log^{1/2}(n)\mu_n^{-1/2} \quad a.s. \tag{A.5}$$

Then, by Lemmas A.4 and 2.1,

$$\begin{aligned}
&C(\log(n)K^*)^{1/2}\mu_n^{-1/2}|\sigma_{K^*n}^{-1}| \\
\geq&\|\widehat{\Lambda} - \Lambda\| \\
=&\|\widehat{U}_{1n}\widehat{\Sigma}_{1n} - U_{1n}\Sigma_{1n}O_n\| \\
\geq&\|U_{1n}(O_n\Sigma_{1n} - \Sigma_{1n}O_n)\| - \|(\widehat{U}_{1n} - U_{1n}O_n)\Sigma_{1n}\| - \|\widehat{U}_{1n}(\widehat{\Sigma}_{1n} - \Sigma_{1n})\| \\
=&\|O_n\Sigma_{1n} - \Sigma_{1n}O_n\| - C'\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|(K^*)^{1/2} \quad a.s.
\end{aligned}$$

Therefore,

$$\|O_n\Sigma_{1n} - \Sigma_{1n}O_n\| \leq C\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|(K^*)^{1/2} \quad a.s. \tag{A.6}$$

In addition,

$$\begin{aligned}
\|\widehat{\Lambda}_i - \Lambda_i\| &= \|\widehat{u}_{1i}^T\widehat{\Sigma}_{1n} - u_{1i}^T\Sigma_{1n}O_n\| \\
&\geq \|(\widehat{u}_{1i}^T - u_{1i}^TO_n)\Sigma_{1n}\| - \|\widehat{u}_{1i}^T(\widehat{\Sigma}_{1n} - \Sigma_{1n})\| - \|u_{1i}^T(\Sigma_{1n}O_n - O_n\Sigma_{1n})\| \\
&=: I - II - III.
\end{aligned}$$

Next, we bound the three terms on the right hand side (RHS) of the above display. First, since $\Sigma_{1n} = \text{diag}(\sigma_{1n}, \ldots, \sigma_{K^*n})$,

$$I \geq |\sigma_{K^*n}|\|\widehat{u}_{1i}^T - u_{1i}^TO_n\| \quad a.s.$$

Denote $\Gamma_n = \sup_i \|\widehat{u}_{1i}^T - u_{1i}^TO_n\|$. By Theorem 2.1 and (A.5)

$$\begin{aligned}
II &\leq (\sup_i \|\widehat{u}_{1i}^T - u_{1i}^TO_n\| + \sup_i \|u_{1i}^T\|)\|\widehat{\Sigma}_{1n} - \Sigma_{1n}\| \\
&\leq C\log^{1/2}(n)\mu_n^{-1/2}\Gamma_n + C\log^{1/2}(n)(n\mu_n)^{-1/2}K^{1/2} \quad a.s.
\end{aligned}$$

Similarly, by (A.6) and Theorem 2.1,

$$III \leq C\log^{1/2}(n)(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}|(KK^*)^{1/2} \quad a.s.$$

Therefore, we have

$$\sup_i \|\widehat{\Lambda}_i - \Lambda_i\| \geq (|\sigma_{K^*n}| - C\log^{1/2}(n)\mu_n^{-1/2})\Gamma_n - C\log^{1/2}(n)(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}|(KK^*)^{1/2}.$$

On the other hand, if $\Gamma_n \leq \delta_n^{(0)}$ a.s. for some deterministic sequence $\{\delta_n^{(0)}\}_{n\geq 1}$, then

$$\sup_i \|\widehat{u}_{1i}\| \leq C(\delta_n^{(0)} + (n/K)^{-1/2}) \quad a.s.$$

Applying Lemma A.4 with $\psi_n = C(\delta_n^{(0)} + (n/K)^{-1/2})$ and Assumption 4, we have

$$\begin{aligned} C\mu_n^{-1}\log(n)\delta_n^{(0)}K^* &+ C(\log(n)K^*)^{1/2}\rho_n(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}| \\ &\geq \sup_i \|\widehat{\Lambda}_i - \Lambda_i\| \\ &\geq (|\sigma_{K^*n}| - C\log^{1/2}(n)\mu_n^{-1/2})\Gamma_n - C\log^{1/2}(n)(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}|(KK^*)^{1/2}. \end{aligned}$$

Since $\log(n)\mu_n^{-1}|\sigma_{K^*n}^{-2}| \to 0$ under Assumption 4, we can choose n_0 sufficiently large such that for $n \geq n_0$,

$$|\sigma_{K^*n}| > C\log^{1/2}(n)\mu_n^{-1/2}.$$

Then, by combining and rearranging terms, we have,

$$\Gamma_n \leq \frac{C\mu_n^{-1}|\sigma_{K^*n}^{-1}|\log(n)K^*}{(1 - C\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|)}\delta_n^{(0)} + \frac{C(\log(n)K^*)^{1/2}(n\mu_n)^{-1/2}\sigma_{K^*n}^{-2}(\rho_n + K^{1/2})}{(1 - C\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|)}.$$

Again, since $\frac{\log(n)K^*}{\mu_n\sigma_{K^*n}} \to 0$ under Assumption 4, we can choose $n_1 > n_0$ sufficiently large such that for any $n \geq n_1$,

$$\frac{C\mu_n^{-1}\log(n)|\sigma_{K^*n}^{-1}|K^*}{(1 - C\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|)} \leq \frac{1}{2}$$

and

$$\frac{C}{(1 - C\log^{1/2}(n)\mu_n^{-1/2}|\sigma_{K^*n}^{-1}|)} \leq 2C.$$

Therefore, for $n \geq n_1$,

$$\delta_n^{(1)} =: \frac{1}{2}\delta_n^{(0)} + \eta_n \geq \Gamma_n,$$

where $\eta_n = 2C(\log(n)K^*)^{1/2}(n\mu_n)^{-1/2}\sigma_{K^*n}^{-2}(\rho_n + K^{1/2})$. We iterate the above calculation $t \geq 1$ times for some arbitrary integer t, and obtain that, for $n \geq n_1$,

$$\Gamma_n \leq \delta_n^{(t)}, \quad \delta_n^{(t)} =: \frac{1}{2}\delta_n^{(t-1)} + \eta_n.$$

This implies $\delta_n^{(t)} = (\frac{1}{2})^t\left[\delta_n^{(0)} - 2\eta_n\right] + 2\eta_n$.

Because $\sup_i \|\widehat{u}_{1i}\|^2 \leq \|\widehat{U}_{1n}\|_F^2 = K^*$, we have $\sup_i \|\widehat{u}_{1i}\| \leq (K^*)^{1/2}$. Then, we set $\delta_n^{(0)} = (K^*)^{1/2}$ and choose $n_2 > n_1$ sufficiently large and $t = n$ such that for $n \geq n_2$,

$$\Gamma_n \leq \delta_n^{(n)} \leq 2^{-n}(K^*)^{1/2} + 2\eta_n \leq 3\eta_n.$$

This concludes the proof. □

Proof of Lemma 2.2. Let $Q_n(\mathcal{A}) = \sum_{k=1}^K \min_{1\le l\le K} \|\beta_{kn} - \alpha_l\|^2 \pi_{kn}$. We first derive the convergence rate of $\widehat{Q}_n(\mathcal{A}) - Q_n(\mathcal{A})$ uniformly over $\mathcal{A} \in \mathcal{M} = \{(\alpha_1, \ldots, \alpha_K) : \sup_{1\le k\le K} \|\alpha_k\| \le M\}$ for some constant M independent of n. Let $R_n = \sup_i \|\hat{\beta}_{in} - \beta_{g_i^0 n}\|$. Then, by Assumption 5(iii),

$$R_n = O_{a.s.}(c_{2n}). \tag{A.7}$$

In addition,

$$\begin{aligned}
\|\hat{\beta}_{in} - \alpha_l\|^2 &\le \|\beta_{g_i^0 n} - \alpha_l\|^2 + 2|(\beta_{g_i^0 n} - \hat{\beta}_{in})^T(\beta_{g_i^0 n} - \alpha_l)| + \|\beta_{g_i^0 n} - \hat{\beta}_{in}\|^2 \\
&\le \|\beta_{g_i^0 n} - \alpha_l\|^2 + 2\|\beta_{g_i^0 n} - \hat{\beta}_{in}\|_1 \|\beta_{g_i^0 n} - \alpha_l\|_\infty + R_n^2 \\
&\le \|\beta_{g_i^0 n} - \alpha_l\|^2 + 2\sqrt{K^*} R_n \|\beta_{g_i^0 n} - \alpha_l\| + R_n^2 \\
&\le \|\beta_{g_i^0 n} - \alpha_l\|^2 + 2\sqrt{K^*} R_n (\|\beta_{g_i^0 n}\| + \|\alpha_l\|) + R_n^2.
\end{aligned}$$

Taking $\min_{1\le l\le K}$ on both sides and averaging over i, we have

$$\widehat{Q}_n(\mathcal{A}) \le Q_n(\mathcal{A}) + C\sqrt{K^*} R_n.$$

Similarly, we have $\widehat{Q}_n(\mathcal{A}) \ge Q_n(\mathcal{A}) - C\sqrt{K^*} R_n$. By (A.7),

$$\breve{R}_n \equiv \sup_{\mathcal{A}\in\mathcal{M}} |\widehat{Q}_n(\mathcal{A}) - Q_n(\mathcal{A})| = O_{a.s.}(c_{2n}\sqrt{K^*}).$$

Next, we show $\widehat{\mathcal{A}}_n \in \mathcal{M}$. Denote $\widehat{\mathcal{A}}_n = \{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K\}$. By Assumption 5(i), we can choose $\infty > M > 0$ such that

$$\sup_i \|\hat{\beta}_{in}\| \le R_n + \sup_{1\le k\le K} \|\beta_{kn}\| \le M.$$

Denote $I_n(k) = \{i : k = \arg\min_{1\le l\le K} \|\hat{\beta}_{in} - \widehat{\alpha}_l\|\}$ for some $k \le K$. If $\|\widehat{\alpha}_k\| > M$ and $I_n(k) = \emptyset$, then we can choose

$$\widehat{\mathcal{A}}_n' = \{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{k-1}, \widehat{\alpha}_k', \widehat{\alpha}_{k+1}, \ldots, \widehat{\alpha}_K\},$$

where $\widehat{\alpha}_k' = \hat{\beta}_{in}$ for some arbitrary $i \le n$. Therefore, we have $\|\widehat{\alpha}_k'\| \le M < \|\widehat{\alpha}_k\|$ and $\widehat{Q}_n(\widehat{\mathcal{A}}_n') < \widehat{Q}_n(\widehat{\mathcal{A}}_n)$, which is a contradiction. On the other hand, if $\|\widehat{\alpha}_k\| > M$ and $I_n(k) \ne \emptyset$, then we can choose

$$\widehat{\mathcal{A}}_n' = \{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{k-1}, \widehat{\alpha}_k', \widehat{\alpha}_{k+1}, \ldots, \widehat{\alpha}_K\},$$

where $\widehat{\alpha}_k' = \frac{1}{|I_n(k)|}\sum_{i\in I_n} \hat{\beta}_{in}$ and $|I_n(k)|$ is the cardinality of $I_n(k)$. This means $\|\widehat{\alpha}_k'\| \le M < \|\widehat{\alpha}_k\|$ and $\widehat{Q}_n(\widehat{\mathcal{A}}_n') < \widehat{Q}_n(\widehat{\mathcal{A}}_n)$, which is a contradiction too. Therefore, $\|\widehat{\alpha}_k\| \le M$. Since k is arbitrary, $\widehat{\mathcal{A}}_n \in \mathcal{M}$.

Third, we show for any $C_1 > 0$, there exists a constant C independent of n such that,

$$\inf_{\mathcal{A}:H(\mathcal{A},\mathcal{B}_n)>C_1} Q_n(\mathcal{A}) \ge C\min(C_1^2/K, c_{1n}^2/K),$$

where $\mathcal{B}_n = \{\beta_{1n}, \ldots, \beta_{Kn}\}$. If there exist some $l_0 \in \{1, \ldots, K\}$ and two indices k_1 and k_2 such that

$$l_0 = \arg\min_{1\le l\le K} \|\beta_{k_1 n} - \alpha_l\| = \arg\min_{1\le l\le K} \|\beta_{k_2 n} - \alpha_l\|,$$

then by Assumption 5(ii)

$$\begin{aligned} Q_n(\mathcal{A}) &\ge \pi_{k_1 n}\|\beta_{k_1 n} - \alpha_{l_0}\|^2 + \pi_{k_2 n}\|\beta_{k_2 n} - \alpha_{l_0}\|^2 \\ &\ge \frac{C}{K}(\|\beta_{k_1 n} - \alpha_{l_0}\| + \|\beta_{k_2 n} - \alpha_{l_0}\|)^2 \ge \frac{C}{K}\|\beta_{k_1 n} - \beta_{k_2, n}\|^2 \ge C c_{1n}^2/K. \end{aligned}$$

On the other hand, if there does not exist such an l_0, then there is a one-to-one mapping $h : \{1, \ldots, K\} \mapsto \{1, \ldots, K\}$ such that

$$h(k) = \arg\min_{1\le l\le K} \|\beta_{kn} - \alpha_l\|.$$

Therefore,

$$Q_n(\mathcal{A}) = \sum_{k=1}^{K} \pi_{kn}\|\beta_{kn} - \alpha_{h(k)}\|^2 \ge (\inf_k \pi_{kn}) H^2(\mathcal{A}, \mathcal{B}_n) \ge C C_1^2/K.$$

Last, we show $H(\widehat{\mathcal{A}}_n, \mathcal{B}_n) = O_{a.s.}(c_{2n}^{1/2} K^{1/2}(K^*)^{1/4})$. For any $\varepsilon > 0$ and sufficiently large C_2,

$$\begin{aligned} &P(H(\widehat{\mathcal{A}}_n, \mathcal{B}_n) \ge C_2 c_{2n}^{1/2} K^{1/2}(K^*)^{1/4} \quad i.o.) \\ &\quad = P(H(\widehat{\mathcal{A}}_n, \mathcal{B}_n) \ge C_2 c_{2n}^{1/2} K^{1/2}(K^*)^{1/4}, Q_n(\widehat{\mathcal{A}}_n) \ge Q_n(\mathcal{B}_n) \\ &\quad + C\min(C_2^2 c_{2n}(K^*)^{1/2}, c_{1n}^2/K) \quad i.o.) \\ &\quad \le P(\widehat{Q}_n(\widehat{\mathcal{A}}_n) + \breve{R}_n \ge \widehat{Q}_n(\mathcal{B}_n) - \breve{R}_n + C\min(C_2^2 c_{2n}(K^*)^{1/2}, c_{1n}^2/K) \quad i.o.) \\ &\quad = P(2\breve{R}_n \ge \widehat{Q}_n(\mathcal{B}_n) - \widehat{Q}_n(\widehat{\mathcal{A}}_n) + C\min(C_2^2 c_{2n}(K^*)^{1/2}, c_{1n}^2/K) \quad i.o.) \\ &\quad \le P(2\breve{R}_n \ge C\min(C_2^2 c_{2n}(K^*)^{1/2}, c_{1n}^2/K) \quad i.o.) \to 0, \end{aligned}$$

where the last inequality because $\widehat{Q}_n(\mathcal{B}_n) - \widehat{Q}_n(\widehat{\mathcal{A}}_n) \ge 0$ and the last equality holds because $\breve{R}_n = O_{a.s.}(c_{2n}(K^*)^{1/2})$ and $\limsup_n c_{2n}c_{1n}^{-2}(K^*)^{1/2}K \to 0$. This concludes the proof. □

Proof of Theorem 2.3. By Lemma 2.2 and Assumption 5(ii) and (iii), for each n, there is a one-to-one mapping $\mu_n : \{1, \ldots, K\} \mapsto \{1, \ldots, K\}$, such that

$$\sup_k \|\widehat{\alpha}_{kn} - \beta_{\mu_n(k)n}\| = O_{a.s.}(c_{2n}^{1/2} K^{1/2}(K^*)^{1/4}).$$

W.l.o.g., we can assume $\mu_n(k) = k$ such that

$$\tilde{R}_n \equiv \sup_k \|\widehat{\alpha}_{kn} - \beta_{kn}\| = O_{a.s.}(c_{2n}^{1/2} K^{1/2}(K^*)^{1/4}). \tag{A.8}$$

If $\widehat{g}_i \neq g_i^0$, then $\|\hat{\beta}_{in} - \widehat{\alpha}_{\widehat{g}_i n}\| \leq \|\hat{\beta}_{in} - \widehat{\alpha}_{g_i^0 n}\|$. This, in conjunction with the triangle inequality, implies that

$$\|\widehat{\alpha}_{\widehat{g}_i n} - \widehat{\alpha}_{g_i^0 n}\| - \|\hat{\beta}_{in} - \widehat{\alpha}_{g_i^0 n}\| \leq \|\hat{\beta}_{in} - \widehat{\alpha}_{\widehat{g}_i n}\| \leq \|\hat{\beta}_{in} - \widehat{\alpha}_{g_i^0 n}\|.$$

It follows that

$$\|\hat{\beta}_{in} - \widehat{\alpha}_{g_i^0 n}\| \geq \frac{1}{2}\|\widehat{\alpha}_{\widehat{g}_i n} - \widehat{\alpha}_{g_i^0 n}\|.$$

By (A.7), (A.8), and the repeated use of the triangle inequality, we have

$$\begin{aligned}
R_n + \tilde{R}_n &\geq \|\hat{\beta}_{in} - \beta_{g_i^0 n}\| + \|\beta_{g_i^0 n} - \widehat{\alpha}_{g_i^0 n}\| \\
&\geq \|\hat{\beta}_{in} - \widehat{\alpha}_{g_i^0 n}\| \geq \frac{1}{2}\|\widehat{\alpha}_{\widehat{g}_i n} - \widehat{\alpha}_{g_i^0 n}\| \\
&= \frac{1}{2}\|(\beta_{\widehat{g}_i n} - \beta_{g_i^0 n}) + (\widehat{\alpha}_{\widehat{g}_i n} - \beta_{\widehat{g}_i n}) + (\beta_{g_i^0 n} - \widehat{\alpha}_{g_i^0 n})\| \\
&\geq \frac{1}{2}\|\beta_{\widehat{g}_i n} - \beta_{g_i^0 n}\| - \tilde{R}_n \geq c_{1n}/2 - \tilde{R}_n.
\end{aligned}$$

This implies

$$1\{\widehat{g}_i \neq g_i^0\} \leq 1\{R_n + 2\tilde{R}_n \geq c_{1n}/2\}.$$

Noting that the RHS of the above display is independent of i, we have

$$\begin{aligned}
P(\sup_i 1\{\widehat{g}_i \neq g_i^0\} > 0 \quad i.o.) &\leq P(R_n + 2\tilde{R}_n \geq c_{1n}/2 \quad i.o.) \\
&= P(O_{a.s.}(c_{2n}) + O_{a.s.}(c_{2n}^{1/2} K^{1/2} (K^*)^{1/4}) \geq c_{1n}/2 \quad i.o.) \\
&= 0 \text{ under Assumption 5(iii)}.
\end{aligned}$$

This concludes the proof. □

Proof of Corollary 2.1. We note that Theorems 2.1-2.2 and Assumption 6 verify Assumptions 5(i) and (ii) and Assumption 5(iii), with $\beta_{kn} = \pi_{kn}^{-1/2}[S_n O_n]_{k\cdot}$, $\hat{\beta}_{in} = n^{1/2}\widehat{u}_{1i}^T$, $c_{2n} = \xi_n^2$, and $c_{1n} = \xi_n$. Then the corollary follows from Theorem 2.3. □

APPENDIX B: PROOFS OF THE TECHNICAL LEMMAS IN APPENDIX A

Proof of Lemma A.2. Definition of ρ_n, $[B_0]_{k_1 k_2} \leq \rho_n$ for any $k_1, k_2 = 1, \ldots, K$. Consider the case nodes i and j are in C_{k_1} and C_{k_2}, respectively. Then

$$\begin{aligned}
P_{ij} &= B_{k_1 k_2} = n^{-1}(nW_{k_1})^{1/2}[B_0]_{k_1 k_2}(nW_{k_2})^{1/2} \\
&= n^{-1}[B_0]_{k_1 k_2}(d_i d_j)^{1/2} \leq \rho_n n^{-1}(d_i d_j)^{1/2}.
\end{aligned}$$

□

Proof of Lemma A.3. Let $r_n = \left[\frac{\phi_{2n}\log(n)K^*}{\mu_n} \vee \left(\frac{\log(n)\phi_{1n}^2\rho_n K^*}{n\mu_n}\right)^{1/2}\right]$, $S^{K^*-1} = \{g \in \Re^{K^*} : \|g\| = 1\}$, and $\mathcal{G}_n = \{g_1, \ldots, g_{M_n}\} \subset S^{K^*-1}$ such that for any $g' \in S^{K^*-1}$,

$$\|g' - f(g')\| \leq n^{-1},$$

where $f(g') = \arg\min_{g\in\mathcal{G}_n} \|g' - g\|$. Then, we have $M_n \leq Cn^{K^*-1}$ and

$$\begin{aligned}
&P\left(\sup_i d_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}\| \geq 2Cr_n\right) \\
&\quad \leq \sum_{i=1}^n P\left(\sup_{g\in S^{K^*-1}} d_i^{-1/2}|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}(g - f(g))| \geq Cr_n\right) \\
&\quad\quad + \sum_{i=1}^n P\left(\sup_{g\in\mathcal{G}_n} d_i^{-1/2}|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}g| \geq Cr_n\right) \\
&\quad := I_n + II_n. \qquad \text{(B.1)}
\end{aligned}$$

Note that $d_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}\| \leq n^{1/2}\mu_n^{-1}$ almost surely, $\sup_{g\in S^{K^*-1}} \|V_{1n}(g - f(g))\| \leq \phi_{1n}n^{-1}$ and $r_n \geq C\frac{\phi_{1n}}{(n\mu_n)^{1/2}}$. Therefore,

$$\begin{aligned}
I_n &\leq \sum_{i=1}^n P\left(\phi_{1n}n^{-1}\|d_i^{-1/2}([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}\| \geq C\phi_{1n}(n\mu_n)^{-1/2}\right) \\
&\leq \sum_{i=1}^n P\left(1 \geq C(\mu_n)^{1/2}\right) = 0. \qquad \text{(B.2)}
\end{aligned}$$

Now we turn to II_n. Let

$$\mathcal{H} = \{h \in \Re^n : \|h\| \leq \phi_{1n} \text{ and } \sup_j |h_j| \leq \phi_{2n}\},$$

where h_j is the j-th element of h. Note, for any $g \in S^{K^*-1}$, $\|V_{1n}g\| = \|V_{1n}\| \leq \phi_{1n}$ and $|[V_{1n}g]_{j\cdot}| \leq \|v_{1j}\| \leq \phi_{2n}$ almost surely. Thus, $\{V_{1n}g : g \in S^{K^*-1}\} \subset \mathcal{H}$. In addition, for any $h \in \mathcal{H}$, $|(A_{ij} - P_{ij})(d_id_j)^{-1/2}h_j| \leq \phi_{2n}/\mu_n$ and by Lemma A.2,

$$\sigma^2 := \sum_{j\neq i} \mathbb{E}(A_{ij} - P_{ij})^2(d_id_j)^{-1}h_j^2 \leq \sum_{j=1}^n P_{ij}(d_id_j)^{-1}h_j^2 \leq \rho_n(n\mu_n)^{-1}\phi_{1n}^2$$

and

$$|A_{ii} - P_{ii}|d_i^{-1}|h_i| = P_{ii}(d_i)^{-1}|h_i| \leq \phi_{2n}\mu_n^{-1} \leq Cr_n/2.$$

Then, by the Bernstein inequality,

$$II_n \le \sum_{i=1}^{n} Cn^{K^*-1} \sup_{g\in\mathcal{S}^{K^*-1}} P\bigg(d_i^{-1/2}|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}g| \ge Cr_n\bigg)$$

$$\le \sum_{i=1}^{n} Cn^{K^*-1} \sup_{h\in\mathcal{H}} P\bigg(Cr_n/2 + |\sum_{j\ne i}(A_{ij} - P_{ij})(d_i d_j)^{-1/2}h_j| \ge Cr_n\bigg)$$

$$\le Cn^{K^*} \exp\bigg(\frac{-(C/2)^2 r_n^2}{\frac{Cr_n\phi_{2n}}{3\mu_n} + \frac{2\rho_n\phi_{1n}^2}{n\mu_n}}\bigg). \tag{B.3}$$

Note that $\phi_{2n}/\mu_n \le r_n/(\log(n)K^*)$ and $\rho_n\phi_{1n}^2/(n\mu_n) \le r_n^2/(\log(n)K^*)$ by the definition of r_n Therefore for a sufficiently large constant C,

$$\text{the RHS of (B.3)} \le Cn^{[K^* - \frac{(C/2)^2K^*}{C/3+2}]} \le Cn^{-2}.$$

Combining (B.1), (B.2), and (B.3), we have,

$$\sum_{n=1}^{\infty} P\bigg(\sup_i\bigg(d_i^{-1/2}\|([A]_{i\cdot} - [\mathcal{P}]_{i\cdot})\mathcal{D}^{-1/2}V_{1n}\|\bigg) \ge 2Cr_n\bigg) < \infty.$$

This leads to the desired result by the Borel-Cantelli lemma. □

Proof of Lemma A.4. Note $\|L\| \le 1$. Then, by Lemma 2.1,

$$\begin{aligned}\|\widehat{\Lambda} - \Lambda\| &= \|L\widehat{U}_{1n} - \mathcal{L}U_{1n}O_n\| \\ &\le \|L(\widehat{U}_{1n} - U_{1n}O_n)\| + \|(L-\mathcal{L})U_{1n}O_n\| \\ &\le \|\widehat{U}_{1n} - U_{1n}O_n\| + \|L - \mathcal{L}\| \\ &\le C\log^{1/2}(n)(K^*)^{1/2}\mu_n^{-1/2}|\sigma_{K^*n}^{-1}| \quad a.s.\end{aligned}$$

For the second result, denote $\widetilde{\Lambda} = D^{-1/2}PD^{-1/2}U_{1n}O_n$ and $\widetilde{\Lambda}_i = \hat{d}_i^{-1/2}[P]_{i\cdot}D^{-1/2}U_{1n}O_n$ as the i-th row of $\widetilde{\Lambda}$. Then we have

$$\sup_i\|\widehat{\Lambda}_i - \Lambda_i\| \le \sup_i\|\Lambda_i - \widetilde{\Lambda}_i\| + \sup_i\|\widehat{\Lambda}_i - \widetilde{\Lambda}_i\| := I + II. \tag{B.4}$$

For I, we have

$$\begin{aligned}I &= \sup_i\|(\hat{d}_i^{-1/2}[P]_{i\cdot}D^{-1/2} - d_i^{-1/2}[P]_{i\cdot}\mathcal{D}^{-1/2})U_{1n}O_n\| \\ &\le \sup_i\|\hat{d}_i^{-1/2}[P]_{i\cdot}D^{-1/2} - d_i^{-1/2}[P]_{i\cdot}\mathcal{D}^{-1/2}\| \\ &\le \sup_i\|\hat{d}_i^{-1/2}[P]_{i\cdot}\mathcal{D}^{-1/2}(\mathcal{D}^{1/2}D^{-1/2} - I)\| + \|(\hat{d}_i^{-1/2} - d_i^{-1/2})[P]_{i\cdot}\mathcal{D}^{-1/2}\| \\ &\le \sup_i\hat{d}_i^{-1/2}\|[P]_{i\cdot}\mathcal{D}^{-1/2}\|\|\mathcal{D}^{1/2}D^{-1/2} - I\| + \sup_i|\hat{d}_i^{-1/2} - d_i^{-1/2}|\|[P]_{i\cdot}\mathcal{D}^{-1/2}\| \\ &:= I_1 + I_2.\end{aligned}$$

By Lemma A.2,

$$\|[P]_{i\cdot}\mathcal{D}^{-1/2}\| = (\sum_{j=1}^{n} P_{ij}^2 d_j^{-1})^{1/2} \le \rho_n (d_i/n)^{1/2}. \tag{B.5}$$

In addition, by the proof of Lemma 2.1,

$$\sup_i |\hat{d}_i^{-1/2} d_i^{1/2} - 1| \le C(\log(n)/\mu_n)^{1/2} \quad a.s. \tag{B.6}$$

and

$$\|\mathcal{D}^{1/2} D^{-1/2} - I\| \le C(\log(n)/\mu_n)^{1/2} \quad a.s. \tag{B.7}$$

Therefore,

$$I_1 \le C \sup_i d_i^{-1/2}(d_i/n)^{1/2}(\log(n)/\mu_n)^{1/2}\rho_n \le C\log^{1/2}(n)(n\mu_n)^{-1/2}\rho_n \quad a.s.,$$

$$I_2 \le C \sup_i d_i^{-1/2}(\log(n)/\mu_n)^{1/2}(d_i/n)^{1/2}\rho_n \le C\log^{1/2}(n)(n\mu_n)^{-1/2}\rho_n \quad a.s.,$$

and

$$I \le C\log^{1/2}(n)(n\mu_n)^{-1/2}\rho_n \quad a.s. \tag{B.8}$$

For II, we have

$$\begin{aligned}
&\sup_i \|\widehat{\Lambda}_i - \widetilde{\Lambda}_i\| \\
=&\sup_i \|\hat{d}_i^{-1/2}[A]_{i\cdot}D^{-1/2}\widehat{U}_{1n} - \hat{d}_i^{-1/2}[P]_{i\cdot}D^{-1/2}U_{1n}O_n\| \\
\le&\sup_i \hat{d}_i^{-1/2}\|[P]_{i\cdot}D^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\| + \sup_i \hat{d}_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})D^{-1/2}U_{1n}O_n\| \\
&+ \sup_i \hat{d}_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})D^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\| \\
:=& II_1 + II_2 + II_3.
\end{aligned} \tag{B.9}$$

By Lemma 2.1, (B.5), and (B.6),

$$\begin{aligned}
II_1 &\le C\sup_i d_i^{-1/2}\rho_n(d_i/n)^{1/2}(\log(n)/\mu_n)^{1/2}|\sigma_{K^*n}^{-1}|(K^*)^{1/2} \\
&\le C\rho_n(\log(n)K^*)^{1/2}(n\mu_n)^{-1/2}|\sigma_{K^*n}^{-1}| \quad a.s.
\end{aligned} \tag{B.10}$$

By (B.6), we have

$$II_2 \le C\sup_i d_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}\mathcal{D}^{1/2}D^{-1/2}U_{1n}O_n\|.$$

Denote $V_{1n} = \mathcal{D}^{1/2}D^{-1/2}U_{1n}O_n$. Then, by (B.7), Theorem 2.1, and the fact that $O_n = O_{a.s.}(1)$, we have $\phi_{1n} = \|V_{1n}\| \le C$ and $\phi_{2n} = \sup_{1\le i\le n}\|v_{1i}\| \le C(n/K)^{-1/2}$

almost surely. Therefore, by Lemma A.3 and the fact that $\rho_n \geq 1$ and $K^*K\log(n)/\mu_n \to 0$ under Assumption 4, we have

$$II_2 \leq C\log^{1/2}(n)(n\mu_n)^{-1/2}(\rho_n K^*)^{1/2} \quad a.s. \tag{B.11}$$

Similarly, we have

$$II_3 \leq Cd_i^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}^{-1/2}\mathcal{D}^{1/2}D^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\|.$$

Let $V_{1n} = \mathcal{D}^{1/2}D^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)$. Then, by (B.7), Lemma 2.1, and Theorem 2.1, we have

$$\phi_{1n} = \|V_{1n}\| \leq C\|\widehat{U}_{1n} - U_{1n}O_n\| \leq C(\log(n)K^*/\mu_n)^{1/2}|\sigma_{K^*n}^{-1}| \quad a.s.$$

and

$$\phi_{2n} = \sup_{1\leq i\leq n}\|v_{1i}\| \leq C\sup_{1\leq i\leq n}\|\widehat{u}_{1i}^T - u_{1i}^T O_n\| \leq \psi_n + C(n/K)^{-1/2} \quad a.s.$$

Then, by Lemma A.3, we have

$$II_3 \leq C\log(n)\mu_n^{-1}\psi_n K^* + C\log(n)\mu_n^{-1}n^{-1/2}K^*(K^{1/2} + \rho_n^{1/2}|\sigma_{K^*n}^{-1}|). \tag{B.12}$$

Combining (B.4) and (B.8)–(B.12) with the fact that $\frac{\log(n)KK^*}{\mu_n} \to 0$, $|\sigma_{K^*n}| \leq |\sigma_{1n}| \leq 1$ and $\rho_n \geq 1$, we find the bound of II_1 and the first term of the bound of II_3 dominate the rest, which leads to the desired result. □

APPENDIX C: PROOFS OF THE RESULTS IN SECTION 3

Proof of Theorem 3.1. Since $\mathcal{L}_\tau = n^{-1}ZB_0^\tau Z$, the proof follows that of Theorem 2.1 with A, B_0, and S_n replaced by A_τ, B_0^τ, and S_n^τ, respectively. □

Proof of Theorem 3.2. The proof of part (i) is analogous to that of Theorem 2.2. The main difference is that we need to use Theorem 3.1 in place of Theorem 2.1.

Theorem 3.1 and the first part of Theorem 3.2 verify Assumptions 5(i) and (ii) and Assumption 5(iii), respectively, with $\beta_{kn} = \pi_{kn}^{-1/2}[S_n^\tau O_n^\tau]_{k\cdot}$ and $\hat{\beta}_{in} = n^{1/2}(\widehat{u}_{1i}^\tau)^T$. Assumption 2 is maintained. Then part (ii) follows from Theorem 2.3. □

Proof of Theorem 3.3. Let $g_i^0 \in \{1,\ldots,K\}$ denote node i's membership. Similar to Qin and Rohe [16, Lemma 3.2], we have by (3.1)

(C.1)

$$d_i = \sum_{j=1}^n P_{ij} = \theta_i\sum_{j=1}^n \theta_j B_{g_i^0 g_j^0} = \theta_i\sum_{k=1}^K\sum_{j\in C_k}\theta_j B_{g_i^0 k} = n\theta_i\sum_{k=1}^K \pi_{kn}B_{g_i^0 k} = n\theta_i W_{g_i^0}.$$

Therefore,

$$
\begin{aligned}
[\mathcal{L}'_\tau]_{ij} = P_{ij}((d_i+\tau)(d_j+\tau))^{-1/2} &= B_{g_i^0 g_j^0}(\theta_i\theta_j)((d_i+\tau)(d_j+\tau))^{-1/2} \\
&= B_{g_i^0 g_j^0}(\theta_i^\tau\theta_j^\tau)^{1/2}(\theta_i\theta_j)^{1/2}(d_i d_j)^{-1/2} \\
&= n^{-1}B_{g_i^0 g_j^0}(\theta_i^\tau\theta_j^\tau)^{1/2}(W_{g_i^0}W_{g_j^0})^{-1/2} \\
&= n^{-1}[\Theta_\tau^{1/2} Z\mathcal{D}_B^{-1/2} B \mathcal{D}_B^{-1/2} Z^T \Theta_\tau^{1/2}]_{ij} \\
&= n^{-1}[\Theta_\tau^{1/2} Z B_0 Z^T \Theta_\tau^{1/2}]_{ij}.
\end{aligned}
$$

That is, $\mathcal{L}'_\tau = n^{-1}\Theta_\tau^{1/2} Z B_0 Z^T \Theta_\tau^{1/2}$. Then

$$
(\mathcal{L}'_\tau)^2 = n^{-1}\Theta_\tau^{1/2} Z B_0 (Z^T\Theta_\tau Z/n) B_0 Z^T \Theta_\tau^{1/2} = n^{-1}\Theta_\tau^{1/2} Z B_0 \Pi_n^\tau B_0 Z^T \Theta_\tau^{1/2},
$$

where $\Pi_n^\tau = Z^T\Theta_\tau Z/n = \mathrm{diag}(\pi_{1n}^\tau, \ldots, \pi_{Kn}^\tau)$, and $\pi_{kn}^\tau = n_k^\tau/n = \sum_{i\in C_k}\theta_i^\tau/n$. By the spectral decomposition, we have

$$
(\Pi_n^\tau)^{1/2} B_0 \Pi_n^\tau B_0 (\Pi_n^\tau)^{1/2} = S_n^\tau \Omega_n (S_n^\tau)^T, \tag{C.2}
$$

where $\Omega_n = \mathrm{diag}(\omega_n, \ldots, \omega_{Kn})$ such that $\omega_n \geq \omega_{2n} \geq \cdots \geq \omega_{Kn} > 0$ and S_n is a $K\times K$ matrix such that $(S_n^\tau)^T S_n^\tau = I_K$. Let $U_{1n}^* = \Theta_\tau^{1/2} Z(Z^T\Theta_\tau Z)^{-1/2} S_n^\tau$. Then, we have

$$
U_{1n}^*\Omega_n U_{1n}^{*T} = (\mathcal{L}'_\tau)^2 = U_{1n}\Sigma_n^2 U_{1n}^T.
$$

In addition, $U_{1n}^{*T}U_{1n}^* = (S_n^\tau)^T S_n^\tau = I_K$. Therefore the columns of U_{1n}^* are the eigenvectors of $\mathcal{L}'_\tau$ associated with eigenvalues $\sigma_n, \ldots, \sigma_{Kn}$, up to sign normalization. W.l.o.g., we can take $U_{1n} = U_{1n}^*$ to obtain the first result.

Now we turn to the second result. If node i is in cluster C_{k_1}, then

$$
u_i^T = (\theta_i^\tau)^{1/2} z_i^T (Z^T\Theta_\tau Z)^{-1/2} S_n^\tau = (\theta_i^\tau)^{1/2}(n_{k_1}^\tau)^{-1/2}[S_n^\tau]_{k_1\cdot},
$$

where $[S_n^\tau]_{k\cdot}$ denotes the k-th row of S_n^τ. Therefore,

$$
(n_{k_1}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|u_i^T\| = \|[S_n^\tau]_{k_1\cdot}\| = 1.
$$

Last, we note that $\frac{u_i^T}{\|u_i^T\|} = [S_n^\tau]_{g_i^0\cdot}$. Therefore, if $z_i \neq z_j$, then $g_i^0 \neq g_j^0$ and

$$
\left\|\frac{u_i^T}{\|u_i^T\|} - \frac{u_{1j}^T}{\|u_{1j}^T\|}\right\| = \|[S_n^\tau]_{g_i^0\cdot} - [S_n^\tau]_{g_j^0\cdot}\| = \sqrt{2}.
$$

Similarly, if $z_i = z_j$, then $g_i^0 = g_j^0$ and $\frac{u_i^T}{\|u_i^T\|} = \frac{u_{1j}^T}{\|u_{1j}^T\|}$. □

Proof of Lemma 3.1. Let $\tilde{L}_\tau = \mathcal{D}_\tau^{-1/2} A \mathcal{D}_\tau^{-1/2}$. Then

$$
\|\mathcal{L}'_\tau - L'_\tau\| \leq \|\mathcal{L}'_\tau - \tilde{L}_\tau\| + \|L'_\tau - \tilde{L}_\tau\| := I + II.
$$

Let $d_i^\tau = d_i + \tau$, $Y_{ij} = (d_i^\tau d_j^\tau)^{-1/2}(A_{ij} - P_{ij})(e_i e_j^T + e_j e_i^T)$ for $1 \leq i < j \leq n$, and $Y_{ii} = -(d_i^\tau)^{-1}P_{ii}e_i e_i^T$, where e_i is the $n\times 1$ vector with its i-th coordinate being 1 and the rest being 0. Then $\{Y_{ij}\}_{1\le i<j\le n}$ is a sequence of independent symmetric random matrices such that $\mathbb{E}Y_{ij}=0$,

$$\tilde{L}_\tau - \mathcal{L}'_\tau + \mathrm{diag}(\mathcal{L}'_\tau) = \sum_{1\le i<j\le n} Y_{ij}, \text{ and } \mathrm{diag}(\mathcal{L}'_\tau) = \sum_{i=1}^n Y_{ii}.$$

In addition, we note that $\sup_{1\le i<j\le n}\|Y_{ij}\| \le \sqrt{2}/\mu_n^\tau$ and

$$\begin{aligned}\sigma^2 = \|\sum_{1\le i<j\le n}\mathbb{E}Y_{ij}^2\| &= \|\mathrm{diag}(\sum_{j\ne 1} p_{1j}(1-p_{1j})/(d_1^\tau d_j^\tau),\ldots,\sum_{j\ne n} p_{nj}(1-p_{nj})/(d_n^\tau d_j^\tau))\| \\ &\le (\mu_n^\tau)^{-1}\max_{1\le i\le n}\sum_{j=1}^n p_{ij}(1-p_{ij})/d_i^\tau \le (\mu_n^\tau)^{-1}.\end{aligned}$$

By Lemma A.1, there exist some constant $C>2$ and some integer n_0 sufficiently large, such that for $n>n_0$

$$\begin{aligned}&P(\|\tilde{L}_\tau - \mathcal{L}'_\tau + \mathrm{diag}(\mathcal{L}'_\tau)\| \ge C(\log(n)/\mu_n^\tau)^{1/2}) \\ &= P(\|\sum_{1\le i<j\le n} Y_{ij}\| \ge C(\log(n)/\mu_n^\tau)^{1/2}) \\ &\le n\exp\left(\frac{-C^2\log(n)/\mu_n^\tau}{3(\mu_n^\tau)^{-1} + 2\sqrt{2}C(\log(n)/\mu_n^\tau)^{1/2}(\mu_n^\tau)^{-1}}\right) \\ &\le Cn^{1-C}. \qquad \text{(C.3)}\end{aligned}$$

This implies

$$\sum_{n=1}^\infty P(\|\tilde{L}_\tau - \mathcal{L}'_\tau + \mathrm{diag}(\mathcal{L}'_\tau)\| \ge C(\log(n)/\mu_n^\tau)^{1/2}) < \infty,$$

or equivalently, $\|\tilde{L}_\tau - \mathcal{L}'_\tau + \mathrm{diag}(\mathcal{L}'_\tau)\| \le C(\log(n)/\mu_n^\tau)^{1/2}\quad a.s.$ In addition, Assumption 11 implies $\log(n)/\mu_n^\tau \to 0$. Therefore,

$$\|\mathrm{diag}(\mathcal{L}'_\tau)\| \le (\mu_n^\tau)^{-1} \le C(\log(n)/\mu_n^\tau)^{1/2}.$$

Therefore,

$$I \le \|\tilde{L}_\tau - \mathcal{L}'_\tau + \mathrm{diag}(\mathcal{L}'_\tau)\| + \|\mathrm{diag}(\mathcal{L}'_\tau)\| \le C(\log n/\mu_n^\tau)^{1/2}\quad a.s.$$

Now we turn to II. Let $\hat{d}_i^\tau = \hat{d}_i + \tau$. By Bernstein inequality, for some $C>2$, we have,

$$\begin{aligned}P(\sup_i |\hat{d}_i^\tau - d_i^\tau|/d_i^\tau \ge C(\log(n)/\mu_n^\tau)^{1/2}) &\le 2\sum_{i=1}^n \exp\left(\frac{-C^2(d_i^\tau)^2\log(n)/\mu_n^\tau}{2d_i^\tau + 2C(\log n/\mu_n^\tau)^{1/2}d_i^\tau/3}\right) \\ &\le Cn^{1-C}.\end{aligned}$$

Therefore, $\sup_i |\hat{d}_i^\tau - d_i^\tau|/d_i^\tau \leq C(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.$, and thus,

$$\|\mathcal{D}_\tau^{-1/2} D_\tau^{1/2} - I\| = \max_i |(\hat{d}_i^\tau/d_i^\tau)^{1/2} - 1| \leq \max_i |(\hat{d}_i^\tau/d_i^\tau) - 1| \leq C(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.$$

In addition, by Chung [7, Lemma 1.7], $\|L'_\tau\| \leq \|L\| \leq 1$. Therefore,

$$\begin{aligned}
\|\tilde{L}_\tau - L'_\tau\| &= \|L'_\tau - \mathcal{D}_\tau^{-1/2} D_\tau^{1/2} L_\tau D_\tau^{1/2} \mathcal{D}_\tau^{-1/2}\| \\
&\leq \|\mathcal{D}_\tau^{-1/2} D_\tau^{1/2} L'_\tau - \mathcal{D}_\tau^{-1/2} D_\tau^{1/2} L'_\tau D_\tau^{1/2} \mathcal{D}_\tau^{-1/2}\| + \|L'_\tau - \mathcal{D}_\tau^{-1/2} D_\tau^{1/2} L'_\tau\| \\
&\leq \|\mathcal{D}_\tau^{-1/2} D_\tau^{1/2} - I\| \|\mathcal{D}_\tau^{-1/2} D_\tau^{1/2}\| + \|\mathcal{D}_\tau^{-1/2} D_\tau^{1/2} - I\| \leq C(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.
\end{aligned}$$

This concludes the first part of the proof. Then by the Davis-Kahan Theorem (e.g., Yu et al. [22, Theorem 2]),

$$\|\widehat{U}_{1n} - U_{1n} O_n\| \leq \frac{CK^{1/2} \|L'_\tau - \mathcal{L}'_\tau\|}{|\sigma_{Kn}|} \leq C(\log(n)/\mu_n^\tau)^{1/2} |\sigma_{Kn}|^{-1} K^{1/2} \quad a.s.$$

□

To prove Theorem 3.4, we need the following three lemmas.

LEMMA C.1. *If Assumption 9 holds, then* $P_{ij} \leq \rho_n n^{-1} (\theta_i \theta_j)^{1/2} (d_i d_j)^{1/2}$.

PROOF. Consider the case in which nodes i and j are in C_{k_1} and C_{k_2}, respectively. Then by the definition of B_0 and (C.1)

$$\begin{aligned}
P_{ij} = \theta_i \theta_j B_{k_1 k_2} &= n^{-1} \theta_i \theta_j (nW_{k_1})^{1/2} [B_0]_{k_1 k_2} (nW_{k_2})^{1/2} \\
&= n^{-1} (\theta_i \theta_j)^{1/2} [B_0]_{k_1 k_2} (d_i d_j)^{1/2} \leq \rho_n n^{-1} (\theta_i \theta_j)^{1/2} (d_i d_j)^{1/2}.
\end{aligned}$$

□

LEMMA C.2. *Let V_n be some $n \times K$ (random) matrix and v_i^T be the i-th row of V_n. Assume there exist two deterministic sequences $\{\phi_{1n}\}_{n \geq 1}$ and $\{\phi_{2n}\}_{n \geq 1}$ such that $\|V_n\| \leq \phi_{1n}$ and $\sup_i \|v_i\| \leq \phi_{2n}$ almost surely. In addition, if Assumptions 9–11 hold, then there exists some positive constant C sufficiently large such that*

$$\begin{aligned}
&\sup_i \Big((n_{g_i^0}^\tau)^{1/2} (\theta_i^\tau)^{-1/2} (d_i^\tau)^{-1/2} \|([A]_{i\cdot} - [\mathcal{P}]_{i\cdot}) \mathcal{D}_\tau^{-1/2} V_n\| \Big) \\
&\qquad \leq 2C \left[\frac{\phi_{2n} \log(n) (nK)^{1/2}}{\mu_n^\tau \underline{\theta}^{1/2}} \vee \left(\frac{\log(n) \rho_n \phi_{1n}^2 \overline{\theta}^{1/2}}{\mu_n^\tau \underline{\theta}^{1/2}} \right)^{1/2} \right] \quad a.s.
\end{aligned}$$

PROOF. The proof is analogous to that of Lemma A.3. We include it for completeness. Let $r_n = \left[\frac{\phi_{2n} \log(n)(nK)^{1/2}}{\mu_n^\tau \underline{\theta}^{1/2}} \vee \left(\frac{\log(n) \rho_n \phi_{1n}^2 \overline{\theta}^{1/2}}{\mu_n^\tau \underline{\theta}^{1/2}} \right)^{1/2} \right]$, $S^{K-1} = \{g \in \Re^K : \|g\| = 1\}$, and $\mathcal{G}_n = \{g_1, \ldots, g_{M_n}\} \subset S^{K-1}$ such that for any $g' \in S^{K-1}$,

$$\|g' - f(g')\| \leq n^{-1} \underline{\theta}^{1/4},$$

where $f(g') = \arg\min_{g\in\mathcal{G}_n} \|g' - g\|$. Then, we have $M_n \leq Cn^{C'K}$ for some constant $C' > 0$ and

$$
\begin{aligned}
&P\left(\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}V_n\| \geq 2Cr_n\right)\\
&\quad\leq \sum_{i=1}^n P\left(\sup_{g\in S^{K-1}} (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}V_n(g - f(g))| \geq Cr_n\right)\\
&\qquad+ \sum_{i=1}^n P\left(\sup_{g\in\mathcal{G}_n} (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}V_n g| \geq Cr_n\right)
\end{aligned}
$$

(C.4)

$$:= I_n + II_n.$$

By Assumption 10,

$$(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}\| \leq n\underline{\theta}^{-1/2}(\mu_n^\tau)^{-1} \quad a.s.$$

In addition, $\sup_{g\in S^{K-1}} \|V_n(g - f(g))\| \leq \phi_{1n}n^{-1}\underline{\theta}^{1/4}$ and $Cr_n \geq C\frac{\phi_{1n}\log^{1/2}(n)}{(\mu_n^\tau)^{1/2}\underline{\theta}^{1/4}}$. Therefore,

$$I_n \leq \sum_{i=1}^n P\left(\phi_{1n}n^{-1}\underline{\theta}^{1/4}\|(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}\| \geq Cr_n\right)$$

(C.5)

$$\leq \sum_{i=1}^n P\left(1 \geq C\log^{1/2}(n)(\mu_n^\tau)^{1/2}\right) = 0.$$

Now we turn to II_n. Let

$$\mathcal{H} = \{h \in \Re^n : \|h\| \leq \phi_{1n} \text{ and } \sup_j |h_j| \leq \phi_{2n}\},$$

where h_j is the j-th element of h. Note that for any $g \in S^{K-1}$, $\|V_n g\| = \|V_n\| \leq \phi_{1n}$ and $|[V_n g]_{j\cdot}| \leq \|v_j\| \leq \phi_{2n}$ almost surely. Thus, $\{V_n g : g \in S^{K-1}\} \subset \mathcal{H}$. For any $h \in \mathcal{H}$,

$$(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}|(A_{ij} - P_{ij})(d_i^\tau d_j^\tau)^{-1/2}h_j| \leq \phi_{2n}n^{1/2}(\underline{\theta}K)^{-1/2}(\mu_n^\tau)^{-1}.$$

In addition, by Lemma C.1,

$$\sum_{j\neq i} n^\tau_{g_i^0}(\theta_i^\tau)^{-1}\mathbb{E}(A_{ij}-P_{ij})^2(d_i^\tau d_j^\tau)^{-1}h_j^2 \leq \sum_{j=1}^n n^\tau_{g_i^0}(\theta_i^\tau)^{-1}P_{ij}(d_i^\tau d_j^\tau)^{-1}h_j^2 \leq \rho_n\overline{\theta}^{1/2}\underline{\theta}^{-1/2}K^{-1}(\mu_n^\tau)^{-1}\phi_{1n}^2$$

and

$$(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}|A_{ii} - P_{ii}|(d_i^\tau)^{-1}|h_i| \leq \phi_{2n}n^{1/2}(K\underline{\theta})^{-1/2}(\mu_n^\tau)^{-1} \leq Cr_n/2.$$

Then, by the Bernstein inequality with C sufficiently large,

$$
\begin{aligned}
II_n &\le \sum_{i=1}^{n} Cn^{C'K} \sup_{g\in\mathcal{S}^{K-1}} P\bigg((n^{\tau}_{g_i^0})^{1/2}(\theta_i^{\tau})^{-1/2}(d_i^{\tau})^{-1/2}|([A]_{i\cdot}-[P]_{i\cdot})\mathcal{D}_{\tau}^{-1/2}V_n g| \ge Cr_n\bigg)\\
&\le \sum_{i=1}^{n} Cn^{C'K} \sup_{h\in\mathcal{H}} P\bigg(Cr_n/2 + (n^{\tau}_{g_i^0})^{1/2}(\theta_i^{\tau})^{-1/2}|\sum_{j\neq i}(A_{ij}-P_{ij})(d_i d_j)^{-1/2}h_j| \ge Cr_n\bigg)\\
&\le Cn^{C'K+1}\exp\bigg(\frac{-(C/2)^2 r_n^2}{\frac{Cr_n\phi_{2n}n^{1/2}}{3\mu_n^{\tau}(K\underline{\theta})^{1/2}} + \frac{2\rho_n\phi_{1n}^2\overline{\theta}^{1/2}}{K\mu_n^{\tau}\underline{\theta}^{1/2}}}\bigg)
\end{aligned}
$$

(C.6)

$$
\le Cn^{C'K+1}\exp\bigg(\frac{-(C/2)^2 r_n^2}{(C/3+2)r_n^2/(\log(n)K)}\bigg) \le Cn^{-2}.
$$

Combining (C.4), (C.5), and (C.6), we have,

$$
\sum_{n=1}^{\infty} P\bigg(\sup_i (n^{\tau}_{g_i^0})^{1/2}(\theta_i^{\tau})^{-1/2}\bigg((d_i^{\tau})^{-1/2}\|([A]_{i\cdot}-[\mathcal{P}]_{i\cdot})\mathcal{D}_{\tau}^{-1/2}V_n\|\bigg) \ge 2Cr_n\bigg) < \infty.
$$

This leads to the desired result by the Borel-Cantelli lemma. □

LEMMA C.3. *Assume there exists a deterministic sequence $\{\psi_n\}_{n\ge1}$ such that*

$$
\sup_j (n^{\tau}_{g_j^0})^{1/2}(\theta_j^{\tau})^{-1/2}\|\widehat{u}_j\| \le \psi_n
$$

almost surely. If Assumptions 9–11 hold, then there exists some positive constant C sufficiently large such that

$$
\|\widehat{\Lambda}-\Lambda\| \le C(\log(n)K/\mu_n^{\tau})^{1/2}|\sigma_{Kn}|^{-1} \quad a.s.
$$

and

$$
\sup_i (n^{\tau}_{g_i^0})^{1/2}(\theta_i^{\tau})^{-1/2}\|\widehat{\Lambda}_i-\Lambda_i\| \le C\left[\frac{\log(n)\overline{\theta}^{1/2}K}{\mu_n^{\tau}\underline{\theta}^{1/2}}\psi_n + \frac{\rho_n\log^{1/2}(n)}{(\mu_n^{\tau})^{1/2}|\sigma_{Kn}|} + \frac{\log^{1/2}(n)\overline{\theta}^{1/4}\rho_n^{1/2}}{(\mu_n^{\tau})^{1/2}\underline{\theta}^{1/4}}\right] \quad a.s.
$$

PROOF. By Chung [7, Lemma 1.7], $\|L'_{\tau}\| \le \|L\| \le 1$. Then, by Lemma 3.1

$$
\begin{aligned}
\|\widehat{\Lambda}-\Lambda\| &= \|L'_{\tau}\widehat{U}_{1n} - \mathcal{L}'_{\tau}U_{1n}O_n\|\\
&\le \|L'_{\tau}(\widehat{U}_{1n}-U_{1n}O_n)\| + \|(L'_{\tau}-\mathcal{L}'_{\tau})U_{1n}O_n\|\\
&\le \|\widehat{U}_{1n}-U_{1n}O_n\| + \|L'_{\tau}-\mathcal{L}'_{\tau}\|\\
&\le C(\log(n)/\mu_n^{\tau})^{1/2}|\sigma_{Kn}|^{-1}K^{1/2} \quad a.s.
\end{aligned}
$$

This proves the first result.

For the second result, denote $\widetilde{\Lambda} = D_\tau^{-1/2} P D_\tau^{-1/2} U_{1n} O_n$ and $\widetilde{\Lambda}_i = (\hat{d}_i^\tau)^{-1/2} [P]_{i\cdot} D_\tau^{-1/2} U_{1n} O_n$ as the i-th row of $\widetilde{\Lambda}$. Then we have

$$\sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} \|\widehat{\Lambda}_i - \Lambda_i\| \leq \sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} \|\Lambda_i - \widetilde{\Lambda}_i\| + \sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} \|\widehat{\Lambda}_i - \widetilde{\Lambda}_i\|$$

$$:= I + II. \tag{C.7}$$

For I, we have

$$\begin{aligned}
I &\leq \sup_i \|(n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} ((\hat{d}_i^\tau)^{-1/2} [P]_{i\cdot} D_\tau^{-1/2} - (d_i^\tau)^{-1/2} [P]_{i\cdot} \mathcal{D}_\tau^{-1/2}) U_{1n} O_n\| \\
&\leq \sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} \|(\hat{d}_i^\tau)^{-1/2} [P]_{i\cdot} D_\tau^{-1/2} - (d_i^\tau)^{-1/2} [P]_{i\cdot} \mathcal{D}_\tau^{-1/2}\| \\
&\leq \sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} (\hat{d}_i^\tau)^{-1/2} \|[P]_{i\cdot} \mathcal{D}_\tau^{-1/2}\| \|\mathcal{D}^{1/2} D_\tau^{-1/2} - I\| \\
&\quad + \sup_i (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} |(\hat{d}_i^\tau)^{-1/2} - (d_i^\tau)^{-1/2}| \|[P]_{i\cdot} \mathcal{D}_\tau^{-1/2}\| \\
&:= I_1 + I_2.
\end{aligned}$$

By Assumption 10 and Lemma C.1,

$$(n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} \|[P]_{i\cdot} \mathcal{D}_\tau^{-1/2}\| = (n^\tau_{g_i^0})^{1/2} (\theta_i^\tau)^{-1/2} (\sum_{j=1}^n P_{ij}^2 (d_j^\tau)^{-1})^{1/2} \leq C \rho_n (d_i/K)^{1/2}. \tag{C.8}$$

In addition, by the proof of Lemma 3.1,

$$\sup_i |(\hat{d}_i^\tau)^{-1/2} (d_i^\tau)^{1/2} - 1| \leq C (\log(n)/\mu_n^\tau)^{1/2} \quad a.s. \tag{C.9}$$

and

$$\|\mathcal{D}_\tau^{1/2} D_\tau^{-1/2} - I\| \leq C (\log(n)/\mu_n^\tau)^{1/2} \quad a.s. \tag{C.10}$$

Therefore,

$$I_1 \leq C \sup_i (d_i^\tau)^{-1/2} (d_i/K)^{1/2} \rho_n (\log(n)/\mu_n^\tau)^{1/2} \leq C \log^{1/2}(n) (K \mu_n^\tau)^{-1/2} \rho_n \quad a.s.,$$

$$I_2 \leq C \sup_i (d_i^\tau)^{-1/2} (\log(n)/\mu_n^\tau)^{1/2} (d_i/K)^{1/2} \rho_n \leq C \log^{1/2}(n) (K \mu_n^\tau)^{-1/2} \rho_n \quad a.s.,$$

and

$$I \leq C \log^{1/2}(n) (K \mu_n^\tau)^{-1/2} \rho_n \quad a.s. \tag{C.11}$$

For II, we have

$$
\begin{aligned}
&\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{\Lambda}_i - \widetilde{\Lambda}_i\| \\
&\quad= \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|(\hat{d}_i^\tau)^{-1/2}[A]_{i\cdot}D_\tau^{-1/2}\widehat{U}_{1n} - (\hat{d}_i^\tau)^{-1/2}[P]_{i\cdot}D_\tau^{-1/2}U_{1n}O_n\| \\
&\quad\le \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|[P]_{i\cdot}D_\tau^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\| \\
&\qquad+ \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})D_\tau^{-1/2}U_{1n}O_n\| \\
&\qquad+ \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})D_\tau^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\| \\
&\quad:= II_1 + II_2 + II_3. \qquad \text{(C.12)}
\end{aligned}
$$

By Lemma 3.1, (C.8), and (C.9),

$$II_1 \le C\sup_i (d_i^\tau)^{-1/2}(d_i/K)^{1/2}\rho_n(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \le C\rho_n\log^{1/2}(n)(\mu_n^\tau)^{-1/2}|\sigma_{Kn}|^{-1} \quad a.s. \qquad \text{(C.13)}$$

By (B.6), we have

$$II_2 \le C\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}\mathcal{D}_\tau^{1/2}D_\tau^{-1/2}U_{1n}O_n\|.$$

Denote $V_n = \mathcal{D}_\tau^{1/2}D_\tau^{-1/2}U_{1n}O_n$ with i-th row given by v_i^T. Then, by (C.10), Theorem 3.3, and the fact that $O_n = O_{a.s.}(1)$, we have $\|V_n\| \le C$ and

$$\sup_{1\le i\le n}\|v_i^T\| = \sup_{1\le i\le n}\|(\theta_i^\tau)^{1/2}(d_i^\tau)^{1/2}(\hat{d}_i^\tau)^{-1/2}(n^\tau_{g_i^0})^{-1/2}[S_n^\tau]_{g_i^0\cdot}O_n\| \le C(\overline{\theta}K)^{1/2}n^{-1/2} \quad a.s.$$

In addition, Assumption 11(i) implies $\frac{\overline{\theta}^{1/2}\log(n)K^2}{\mu_n^\tau\underline{\theta}^{1/2}\rho_n} \to 0$. Therefore, by Lemma C.2, we have

$$II_2 \le C(\log(n)\overline{\theta}^{1/2}\rho_n)^{1/2}(\mu_n^\tau\underline{\theta}^{1/2})^{-1/2} \quad a.s. \qquad \text{(C.14)}$$

Similarly, we have

$$II_3 \le C\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(d_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}\mathcal{D}^{1/2}D_\tau^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)\|.$$

Let $V_n = \mathcal{D}_\tau^{1/2}D_\tau^{-1/2}(\widehat{U}_{1n} - U_{1n}O_n)$ with i-th row given by v_i^T. Then, by (B.7), Lemma 3.1, and Theorem 3.3, we have

$$
\begin{aligned}
\sup_i \|v_i^T\| &= \sup_{1\le i\le n}\|(d_i^\tau)^{1/2}(\hat{d}_i^\tau)^{-1/2}(\widehat{u}_i^T - u_i^T O_n)\| \\
&\le C\sup_{1\le i\le n}\|(n^\tau_{g_i^0})^{-1/2}(\theta_i^\tau)^{1/2}[(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\widehat{u}_i^T - u_i^T O_n)]\| \le C(K\overline{\theta}/n)^{1/2}(\psi_n + 1)
\end{aligned}
$$

and

$$\|V_n\| \le C\|\widehat{U}_{1n} - U_{1n}O_n\| \le C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \quad a.s.$$

Then, by Lemma C.2, we have

$$II_3 \le C\Big[\frac{\log(n)\overline{\theta}^{1/2}K}{\mu_n^\tau \underline{\theta}^{1/2}}\psi_n + \frac{\log(n)\overline{\theta}^{1/2}K}{\mu_n^\tau \underline{\theta}^{1/2}} + \frac{\log(n)\overline{\theta}^{1/4}(K\rho_n)^{1/2}}{\mu_n^\tau |\sigma_{Kn}|\underline{\theta}^{1/4}}\Big]. \tag{C.15}$$

Again, by Assumption 11(i), $\frac{\overline{\theta}^{1/2}\log(n)K^2}{\mu_n^\tau \underline{\theta}^{1/2}\rho_n} \to 0$. Therefore, the bounds for II_1 and II_2 dominate the bounds for the second and third term of II_3. In addition, the bound for term I is dominated by the bound for term II_1. Combining (C.7), (C.11)–(C.15), we have the desired result. □

Proof of Theorem 3.4. First, by Weilandt-Hoffman inequality and Lemma 3.1

$$\|\widehat{\Sigma}_n - \Sigma_n\| \le \|L'_\tau - \mathcal{L}'_\tau\| \le C(\log(n)/\mu_n^\tau)^{1/2} \quad a.s. \tag{C.16}$$

Then, by Lemmas C.3 and 3.1,

$$\begin{aligned}
C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} &\ge \|\widehat{\Lambda} - \Lambda\| \\
&= \|\widehat{U}_{1n}\widehat{\Sigma}_n - U_{1n}\Sigma_n O_n\| \\
&\ge \|U_{1n}(O_n\Sigma_n - \Sigma_n O_n)\| - \|(\widehat{U}_{1n} - U_{1n}O_n)\Sigma_n\| - \|\widehat{U}_{1n}(\widehat{\Sigma}_n - \Sigma_n)\| \\
&= \|O_n\Sigma_n - \Sigma_n O_n\| - C'(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \quad a.s.
\end{aligned}$$

Therefore,

$$\|O_n\Sigma_n - \Sigma_n O_n\| \le C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \quad a.s. \tag{C.17}$$

In addition,

$$\begin{aligned}
(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{\Lambda}_i - \Lambda_i\| &= (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i^T\widehat{\Sigma}_n - u_i^T\Sigma_n O_n\| \\
&\ge (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|(\widehat{u}_i^T - u_i^T O_n)\Sigma_n\| - (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i^T(\widehat{\Sigma}_n - \Sigma_n)\| \\
&\quad - (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|u_i^T(\Sigma_n O_n - O_n\Sigma_n)\| \\
&:= I - II - III.
\end{aligned}$$

Next, we bound the three terms on the RHS of the above display. For the first term, we have

$$I \ge |\sigma_{Kn}|(n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i^T - u_i^T O_n\| \quad a.s.$$

Denote $\Gamma_n = \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i^T - u_i^T O_n\|$. By Theorem 3.3 and (C.16),

$$\begin{aligned}
II &\le (\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i^T - u_i^T O_n\| + \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|u_i^T\|)\|\widehat{\Sigma}_n - \Sigma_n\| \\
&\le C(\log(n)/\mu_n^\tau)^{1/2}(\Gamma_n + 1) \quad a.s.
\end{aligned}$$

Similarly, by (C.17) and Theorem 3.3,

$$III \leq C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \quad a.s.$$

Therefore, we have

$$\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{\Lambda}_i - \Lambda_i\| \geq (|\sigma_{Kn}| - C(\log(n)/\mu_n^\tau)^{1/2})\Gamma_n - C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-1}|.$$

On the other hand, if $\Gamma_n \leq \delta_n^{(0)}$ a.s. for some deterministic sequence $\{\delta_n^{(0)}\}_{n\geq 1}$, then by Theorem 3.3

$$\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{u}_i\| \leq C(\delta_n^{(0)} + 1) \quad a.s.$$

Applying Lemma C.3 with $\psi_n = C(\delta_n^{(0)} + 1)$, we have

$$\begin{aligned} C\Big[&\frac{\log(n)\overline{\theta}^{1/2}K}{\mu_n^\tau \underline{\theta}^{1/2}}\delta_n^{(0)} + \frac{\rho_n \log^{1/2}(n)}{(\mu_n^\tau)^{1/2}|\sigma_{Kn}|} + \frac{\log^{1/2}(n)\overline{\theta}^{1/4}\rho_n^{1/2}}{(\mu_n^\tau)^{1/2}\underline{\theta}^{1/4}}\Big] \\ &\geq \sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\widehat{\Lambda}_i - \Lambda_i\| \\ &\geq (|\sigma_{Kn}| - C(\log(n)/\mu_n^\tau)^{1/2})\Gamma_n - C(\log(n)K/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-1}|. \end{aligned}$$

Since $\log(n)/(\mu_n^\tau \sigma_{Kn}^2) \to 0$ by Assumption 11(i), we can choose n_0 sufficiently large such that for $n \geq n_0$,

$$|\sigma_{Kn}| > C(\log(n)/\mu_n^\tau)^{1/2}.$$

Then, by combining and rearranging terms, we have,

$$\frac{C\log(n)K\overline{\theta}^{1/2}(\mu_n^\tau)^{-1}\underline{\theta}^{-1/2}|\sigma_{Kn}|^{-1}}{(1 - C(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1})}\delta_n^{(0)} + \frac{C\eta_n}{(1 - C(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1})} \geq \Gamma_n,$$

where

$$\eta_n = \left(\frac{\log^{1/2}(n)}{(\mu_n^\tau)^{1/2}|\sigma_{Kn}|}\right)\left(\frac{\overline{\theta}^{1/4}\rho_n^{1/2}}{\underline{\theta}^{1/4}} + \frac{\rho_n + K^{1/2}}{|\sigma_{Kn}|}\right).$$

Again, since $C\log(n)K\overline{\theta}^{1/2}(\mu_n^\tau)^{-1}\underline{\theta}^{-1/2}|\sigma_{Kn}|^{-1} \to 0$ by Assumption 11(i), we can choose $n_1 > n_0$ sufficiently large such that for any $n \geq n_1$,

$$\frac{C\log(n)K\overline{\theta}^{1/2}(\mu_n^\tau)^{-1}\underline{\theta}^{-1/2}|\sigma_{Kn}|^{-1}}{(1 - C(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1})} \leq \frac{1}{2}$$

and

$$\frac{C}{1 - C(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1}} \leq 2C.$$

Then, for $n \geq n_1$,

$$\delta_n^{(1)} \equiv \frac{1}{2}\delta_n^{(0)} + 2C\eta_n \geq \Gamma_n.$$

We iterate the above calculation t times for some arbitrary integer t, and obtain that, for $n \geq n_1$,

$$\Gamma_n \leq \delta_n^{(t)}, \quad \delta_n^{(t)} = \frac{1}{2}\delta_n^{(t-1)} + 2C\eta_n.$$

This implies

$$\delta_n^{(t)} = (\frac{1}{2})^t \left[\delta_n^{(0)} - 4C\eta_n\right] + 4C\eta_n.$$

In addition, because $\sup_i n^{\tau}_{g_i^0}(\theta_i^{\tau})^{-1}\|\widehat{u}_i\|^2 \leq n\underline{\theta}^{-1}\|\widehat{U}_{1n}\|_F^2/K = n\underline{\theta}^{-1}$, we have

$$\sup_i (n^{\tau}_{g_i^0})^{1/2}(\theta_i^{\tau})^{-1/2}\|\widehat{u}_i\| \leq n^{1/2}\underline{\theta}^{-1/2}.$$

Therefore, we can set $\delta_n^{(0)} = n^{1/2}\underline{\theta}^{-1/2}$ and choose $n_2 > n_1$ sufficiently large and $t = n$ such that for $n \geq n_2$,

$$\Gamma_n \leq \delta_n^{(n)} \leq 2^{-n}n^{1/2}\underline{\theta}^{-1/2} + 4C\eta_n \leq 5C\eta_n,$$

where the last inequality holds because η_n decays at most polynomially. This concludes the proof. □

Proof of Corollary 3.1. By the triangle inequality and Theorem 3.4,

$$\begin{aligned}
\sup_i \left\| \frac{\widehat{u}_i^T}{\|\widehat{u}_i^T\|} - \frac{u_i^T O_n}{\|u_i^T O_n\|} \right\| &\leq \sup_i \left\| \frac{\widehat{u}_i^T}{\|\widehat{u}_i^T\|} - \frac{\widehat{u}_i^T}{\|u_i^T O_n\|} \right\| + C \sup_i n_{g_i^0}^{1/2}(\theta_i^{\tau})^{-1/2}\|\widehat{u}_i^T - u_i^T O_n\| \\
&\leq 2C \sup_i n_{g_i^0}^{1/2}(\theta_i^{\tau})^{-1/2}\|\widehat{u}_i^T - u_i^T O_n\| \\
&\leq \left(\frac{\log^{1/2}(n)}{(\mu_n^{\tau})^{1/4}\,|\sigma_{Kn}|} \right) \left(\frac{\overline{\theta}^{1/4}\rho_n^{1/2}}{\mu_n^{1/4}\underline{\theta}^{1/4}} + \frac{\rho_n + K^{1/2}}{|\sigma_{Kn}|\,(\mu_n^{\tau})^{1/4}} \right) \quad a.s.
\end{aligned}$$

The second result follows Theorem 2.3 with $K^* = K$. □

Proof of Theorem 3.5. Let $\varepsilon_n = C\log(n)/\underline{m}_n$, for some positive constant C which is sufficiently large.

$$\begin{aligned}
&P\left(\sup_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| \geq \varepsilon_n \quad i.o. \right) \\
\leq &P\left(\sup_{1\leq i\leq n} |\hat{\theta}_i - \theta_i| \geq \varepsilon_n \quad i.o., \ \sup_{1\leq i\leq n} 1\{\hat{g}_i \neq g_i^0\} = 0 \right) + P\left(\sup_{1\leq i\leq n} 1\{\hat{g}_i \neq g_i^0\} > 0 \quad i.o. \right) \\
\leq &P\left(\sup_{1\leq i\leq n} |n_{g_i^0}(\textstyle\sum_{j=1}^n A_{ij})/(\sum_{i':g_{i'}^0=g_i^0}\sum_{j=1}^n A_{i'j}) - \theta_i| \geq \varepsilon_n \quad i.o. \right).
\end{aligned}$$

where the last inequality holds by Assumption 12(ii). In order to show the RHS of the above equation is zero, it suffices to show

$$\sum_{n=1}^{\infty}\sum_{i=1}^{n} P\left(|n_{g_i^0}(\textstyle\sum_{j=1}^n A_{ij})/(\sum_{i':g_{i'}^0=g_i^0}\sum_{j=1}^n A_{i'j}) - \theta_i| \geq \varepsilon_n \right) < \infty. \tag{C.18}$$

For the simplicity of notation, from now on, we assume $g_i^0 = k$. Then, we have

$$|n_{g_i^0}(\sum_{j=1}^n A_{ij})/(\sum_{i':g_{i'}^0=g_i^0}\sum_{j=1}^n A_{i'j}) - \theta_i| = \frac{\sum_{j=1}^n (A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i)}{\sum_{j=1}^n \sum_{i'\in C_k} A_{i'j}}.$$

For the denominator, note that $\mathbb{E}\sum_{j=1}^n \sum_{i'\in C_k} A_{i'j} = m_k n_k$, $\mathbb{E}\sum_{j=1}^n \sum_{i'\in C_k} A_{i'j}^2 \le m_k n_k$. Then, by Bernstein inequality, for any $\lambda > 0$,

$$P\left(|\frac{\sum_{j=1}^n \sum_{i'\in C_k} A_{i'j}}{m_k n_k} - 1| \ge \lambda\right) \le 2\exp\left(-\frac{\frac{1}{2}\lambda^2 m_k^2 n_k^2}{m_k n_k + \frac{1}{3}\lambda m_k n_k}\right) = 2\exp(-C_\lambda m_k n_k),$$

where $C_\lambda = \frac{3\lambda^2}{6+2\lambda}$. Similarly, for the numerator, we note that $|A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i| \le n_k(\theta_i + 1)$ and $\sum_{j=1}^n \mathbb{E}(A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i)^2 \le n_k^2 - \theta_i^2 m_k n_k$. Then, by Assumption 12(i) and Bernstein inequality,

$$\begin{aligned} P(|\frac{\sum_{j=1}^n (A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i)}{m_k n_k}| \ge \varepsilon_n) \le & 2\exp\left(-\frac{\frac{1}{2}\varepsilon_n^2 m_k^2 n_k^2}{n_k^2 - \theta_i^2 m_k n_k + \frac{1}{3}\varepsilon_n m_k n_k^2(\theta_i + 1)}\right) \\ \le & C\exp(-C'\varepsilon_n m_k). \end{aligned}$$

Therefore,

$$\begin{aligned} & P\left(|\frac{\sum_{j=1}^n (A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i)}{\sum_{j=1}^n \sum_{i'\in C_k} A_{i'j}}| \ge \varepsilon_n\right) \\ \le & P\left(|\frac{\sum_{j=1}^n (A_{ij}n_k - \sum_{i'\in C_k} A_{i'j}\theta_i)}{m_k n_k}| \ge \varepsilon_n(1-\lambda)\right) + 2\exp(-C_\lambda m_k n_k) \\ \le & C\exp(-C'\varepsilon_n(1-\lambda)\underline{m}_n) + 2\exp(-C_\lambda \underline{m}_n n_k). \end{aligned}$$

By construction, $\varepsilon_n \underline{m}_n = C\log(n)$ for C sufficiently large. Therefore, (C.18) holds, which concludes the proof. □

APPENDIX D: ADDITIONAL SIMULATION RESULTS

In this section, we report some additional simulation results for DGPs 1-4 studied in the paper.

Table 2 reports the classification results based on the eigenvectors corresponding to the largest K eigenvalues of $L = D^{-1/2}AD^{-1/2}$. Given an adjacency matrix A, D is not invertible when there exists a node which has degree 0. We also report the percentage of replications which generate A with strictly positive degrees for each node in the table, denoted as Ratio. For these realizations, we report the classification results. In Table 2, "CCP" indicates the Correct Classification Proportion criterion; "NMI" means the Normalized Mutual Information criterion, and "kmeans" correspond to the classification methods K-means with default options (Matlab "kmedoids"). We summarize some important findings from Table 2. First, we have a fair large probability to obtain zero degree for some nodes in DGPs 1–4 because we allow the minimum degree to diverge to infinity at a very slow rate, namely at rate-$\log(n)$ in DGPs 1 and 3 and rate-$\log^{5/6}(n)$ in DGPs 2 and 4. Second, the performance of the spectral classification based on L is not as satisfactory as that based on its regularized version studied in the paper. This is especially true when n/K is small.

TABLE 2
Classification results based on $L = D^{-1/2}AD^{-1/2}$

DGP	K	n/K	Ratio	CCP	NMI
1	2	50	0.650	0.9696	0.7863
	2	200	0.632	0.9917	0.9361
2	3	50	0.368	0.9600	0.8330
	3	200	0.170	0.9870	0.9245
3	2	50	0.102	0.9647	0.7543
	2	200	0.002	0.9773	0.8042
4	3	50	0.040	0.9810	0.7482
	3	200	0.000	–	–

Figures 5–8 report the classification results based on $L'_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}$ and $L_\tau = D_\tau^{-1/2}A_\tau D_\tau^{-1/2}$ for DGPs 1–2 and DGPs 3–4, respectively. As in the paper, the left column uses the CCP criterion and the right column uses the NMI criterion to evaluate the classification performance. The x-axis marks the τ values, i.e., $[10^{-4}, (\tau_{\max})^0, (\tau_{\max})^{1/18}, \ldots, (\tau_{\max})^{18/18}]$, where $\tau_{\max}$ is the expected average degree. There are two curves in each subplots. As marked in the legend and explained in the paper, they represent classification results by using different classification methods. In each subplot, the green dashed line is the pseudo τ value as defined in [9]. We summarize some findings from Figures 5–8. First, the spectral classification results first improve and then deteriorate as τ increases. Second, as Figures 5 and 6 suggest, the spectral clustering based on $L'_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}$ based on either $\tau = \bar{d}$ or τ^{JY} is slightly worse than the UPL method. Third, as Figures 7 and 8 suggest, the method of [9] tends to select too large a regularization parameter, but still yields classification results that are much better than those of CPL.

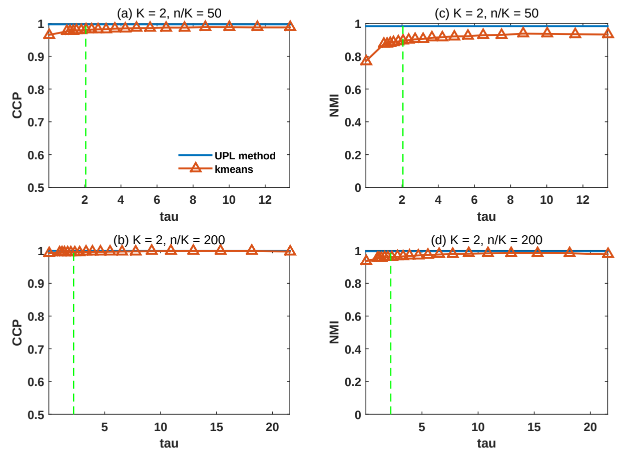

Fig 5: Classification results for CPL and K-means for DGP 1 ($K = 2$) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$. The the x-axis marks the τ values and the y-axis is either CCP (left column) or NMI (right column). The green dashed vertical line in each subplot indicated the estimated τ^{JY} value by using the method of [9]. The first and second rows correspond to $n/K = 50$ and 200, respectively.

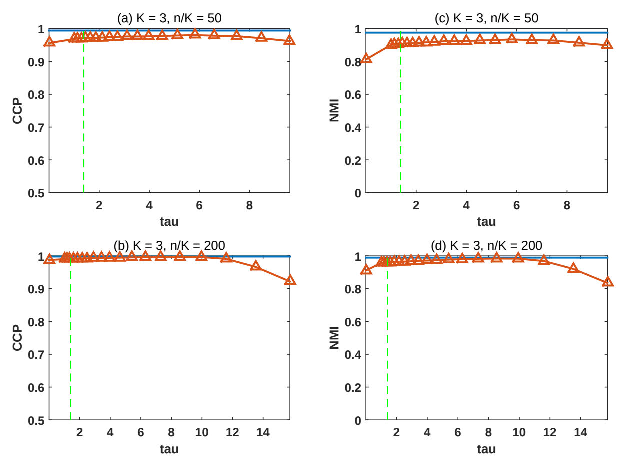

Fig 6: Classification results for DGP 2 ($K = 3$) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$. (See Figure 5 for explanations.)

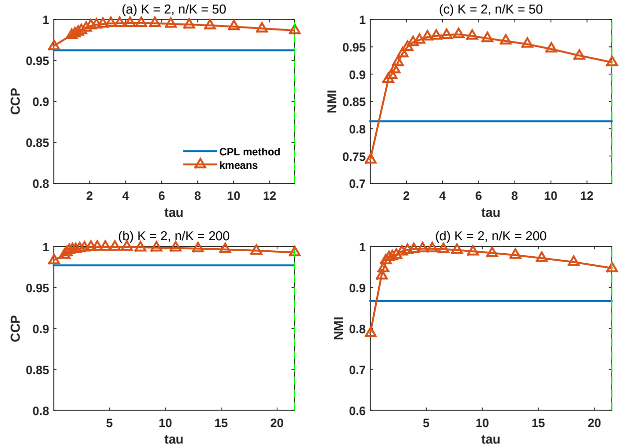

Fig 7: Classification results for DGP 3 ($K = 2$, degree-corrected) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. (See Figure 5 for explanations.)

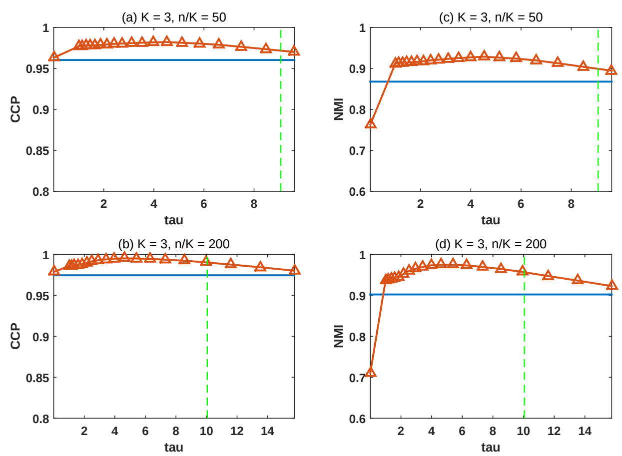

Fig 8: Classification results for DGP 4 ($K = 3$, degree-corrected) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. (See Figure 5 for explanations.)

REFERENCES

[1] Amini, A. A., Chen, A., Bickel, P. J., Levina, E., 2013. Pseudo-likelihood methods for community detection in large sparse networks. The Annals of Statistics 41 (4), 2097–2122.

[2] Bickel, P. J., Chen, A., 2009. A nonparametric view of network models and Newman–Girvan and other modularities. Proceedings of the National Academy of Sciences 106 (50), 21068–21073.

[3] Bickel, P. J., Chen, A., 2012. Weak consistency of community detection criteria under the stochastic block model. Preprint.

[4] Bickel, P. J., Chen, A., Levina, E., 2011. The method of moments and degree distributions for network models. The Annals of Statistics 39 (5), 2280–2301.

[5] Chaudhuri, K., Chung, F., Tsiatas, A., 2012. Spectral clustering of graphs with general degrees in the extended planted partition model. Journal of Machine Learning Research Workshop and Conference Proceedings 23 (35), 1–23.

[6] Choi, D. S., Wolfe, P. J., Airoldi, E. M., 2012. Stochastic blockmodels with a growing number of classes. Biometrika 99 (2), 273–284.

[7] Chung, F. R., 1997. Spectral graph theory. Vol. 92. American Mathematical Soc.

[8] Holland, P. W., Laskey, K. B., Leinhardt, S., 1983. Stochastic blockmodels: First steps. Social networks 5 (2), 109–137.

[9] Joseph, A., Yu, B., 2016. Impact of regularization on spectral clustering. The Annals of Statistics 44 (4), 1765–1791.

[10] Karrer, B., Newman, M. E., 2011. Stochastic blockmodels and community structure in networks. Physical Review E 83 (1), 016107.

[11] Lei, J., 2016. A goodness-of-fit test for stochastic block models. The Annals of Statistics 44 (1), 401–424.

[12] Lei, J., Rinaldo, A., 2015. Consistency of spectral clustering in stochastic block models. The Annals of Statistics 43 (1), 215–237.

[13] Lyzinski, V., Sussman, D., Tang, M., Athreya, A., Priebe, C., 2014. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. Electronic Journal of Statistics 8 (2), 2905–2922.

[14] Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., Tropp, J. A., 2014. Matrix concentration inequalities via the method of exchangeable pairs. The Annals of Probability 42 (3), 906–945.

[15] Newman, M. E., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical review E 69 (2), 026113.

[16] Qin, T., Rohe, K., 2013. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In: Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. Q. (Eds.), Advances in Neural Information Processing Systems. Vol. 26. Curran Associates, Inc., pp. 3120–3128.

[17] Rohe, K., Chatterjee, S., Yu, B., 2011. Spectral clustering and the high-dimensional stochastic blockmodel. The Annals of Statistics 39 (4), 1878–1915.

[18] Sarkar, P., Bickel, P. J., 2015. Role of normalization in spectral clustering for stochastic blockmodels. The Annals of Statistics 43 (3), 962–990.

[19] Sussman, D. L., Tang, M., Fishkind, D. E., Priebe, C. E., 2012. A consistent adja-

cency spectral embedding for stochastic blockmodel graphs. Journal of the American Statistical Association 107 (499), 1119–1128.

[20] Wang, Y., Bickel, P. J., 2016. Likelihood-based model selection for stochastic block models. arXiv preprint arXiv:1502.02069.

[21] Wilson, J. D., Stevens, N. T., Woodall, W. H., 2016. Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model. arXiv preprint arXiv:1605.04049.

[22] Yu, Y., Wang, T., Samworth, R. J., 2015. A useful variant of the Davis–Kahan theorem for statisticians. Biometrika 102 (2), 315–323.

[23] Zhao, Y., Levina, E., Zhu, J., 2012. Consistency of community detection in networks under degree-corrected stochastic block models. The Annals of Statistics 40 (4), 2266–2292.

90 Stamford Road, Singapore 178903
E-mail: ljsu@smu.edu.sg; wuyi.wang.2013@phdecons.smu.edu.sg; yczhang@smu.edu.sg