Strong Consistency of Spectral Clustering for Stochastic Block Models

Liangjun Su* Wuyi Wang† Yichong Zhang‡

May 14, 2019

Abstract

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.

Key words and phrases: Community detection, degree-corrected stochastic block model, K-means, regularization, strong consistency.

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 Holland, Laskey and Leinhardt (1983) 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 method of moments (Bickel, Chen and Levina, 2011), modularity maximization (Newman and Girvan, 2004), semidefinite programming (Abbe, Bandeira and Hall, 2016; Cai and Li, 2015), spectral clustering (Joseph and Yu, 2016; Lei and Rinaldo, 2015; Qin and Rohe, 2013; Rohe, Chatterjee and Yu, 2011; Sarkar and Bickel, 2015; Vu, 2018; Yun and Proutiere, 2014, 2016), likelihood methods (Amini, Chen, Bickel and Levina, 2013; Bickel and Chen, 2009; Choi, Wolfe and Airoldi, 2012; Zhao, Levina and Zhu, 2012), and spectral embedding (Lyzinski, Sussman, Tang, Athreya and Priebe, 2014; Sussman, Tang, Fishkind and Priebe, 2012). Abbe (2018) provides an excellent survey on recent developments on community detection and stochastic block models. Among the methods

*Singapore Management University. E-mail address: ljsu@smu.edu.sg. Su acknowledges the funding support provided by the Lee Kong Chian Fund for Excellence.

†Jinan University. E-mail address: wangwuyi@live.com.

‡Singapore Management University. E-mail address: yczhang@smu.edu.sg. The corresponding author.

mentioned above, spectral clustering is arguably one of the most widely used methods due to its computational tractability.

Bickel and Chen (2009) 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 large samples. Based on the parameters of the block model, properties of the modularities, and expected degree of the graph (λ_n), Bickel and Chen (2009) give the sufficient conditions for strong consistency, which is $\lambda_n/\log(n) \to \infty$. Zhao et al. (2012) define weak consistency of community detection, which essentially means that the number of misclassified nodes is of smaller order than the number of nodes. Bickel and Chen (2012) find that weak consistency requires that $\lambda_n \to \infty$ for the SBM. Similarly, under the conditions that $\lambda_n/\log(n) \to \infty$ ($\lambda_n \to \infty$), Zhao et al. (2012) establish the strong (weak) consistency under both standard SBMs and degree-corrected SBMs.

If the community detection method is strongly consistent, then it means that the communities are exactly recoverable. From an information-theory perspective, Abbe and Sandon (2015), Abbe et al. (2016), Mossel, Neeman and Sly (2014), and Vu (2018) study the phase transition threshold for exact recovery, which requires $\lambda_n = \Omega(\log(n))$. It is well known that some methods like the modularity maximization of Newman and Girvan (2004) and the likelihood method of Bickel and Chen (2009) yield strongly consistent community recovery, 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. Abbe et al. (2016) show that semidefinite programming can achieve exact recovery when there are two equal-sized communities. Yun and Proutiere (2014), Yun and Proutiere (2016), and Vu (2018) establish strong consistency for the variants of spectral method, which involve graph splitting, trimming, and a final improvement step. The pure spectral clustering method has been shown to enjoy weak consistency under standard or degree-corrected SBMs by various researchers; see Joseph and Yu (2016), Lei and Rinaldo (2015), Qin and Rohe (2013), and Rohe et al. (2011). Weak consistency here means that the fraction of misclassified nodes decreases to zero as n grows. Because the decrease rates established in above papers are usually slower than n, the above weak consistency results imply that the number of misclassified nodes still increases to infinity as n grows. On the contrary, strong consistency implies that the number of misclassified nodes is zero for sufficiently large n, which greatly improves upon weak consistency.

The aim of this paper is to formally establish the strong consistency of spectral clustering for standard/regular SBMs without any extra refinement steps, 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 $\mu_n/\log(n)$ being sufficiently large can ensure strong consistency. In other words, the spectral clustering method achieves the optimal rate for exact recovery, as pointed out in Abbe et al. (2016) and Abbe and Sandon (2015).

As demonstrated by Amini et al. (2013), the performance of spectral clustering can be considerably improved via regularization. Joseph and Yu (2016) 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.

[1] Bickel and Chen (2009) use the terminology "asymptotic consistency" in place of strong consistency.

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, Karrer and Newman (2011) 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.

Our paper is mostly related to Abbe, Fan, Wang and Zhong (2017). Abbe et al. (2017) derive the L_∞ bound for the entrywise eigenvector of random matrices with low expected rank. Then they apply their general results to SBM with two communities, where both within- and cross-community probabilities are of order $\log(n)/n$ and show that classifying nodes based on the sign of the entries in the second eigenvector can achieve exact recovery. Our paper complements theirs in the following three aspects. First, we consider the eigenvectors of normalized graph Laplacian L rather than the adjacency matrix A. Therefore, the entrywise bound of the eigenvectors derived in Abbe et al. (2017) cannot be directly used in our case. Our proof relies on the construction of a contraction mapping for the entrywise bound, via which we can iteratively refine the bound. Such strategy is different from that in Abbe et al. (2017).

Second, we consider SBM with a general block probability matrix whereas Abbe et al. (2017) consider a 2×2 block probability matrix. Even though Abbe et al. (2017) establish general theories of L_∞ bound for the entrywise eigenvector of random matrices, when applying their theory to SBMs, they only study the model with the following block probability matrix:

$$\begin{pmatrix} \frac{a\log(n)}{n} & \frac{b\log(n)}{n} \\ \frac{b\log(n)}{n} & \frac{a\log(n)}{n} \end{pmatrix}. \tag{1.1}$$

Their block probability matrix assumes that there are two groups, the connection probability within groups are the same for the two groups, and the within- and cross-group connection probabilities are of the same order of $\log(n)/n$. In contrast, our paper studies the general SBM with generic K groups, where K is allowed to diverge to infinity at a slow rate and the decay rates for different elements in the block probability matrix can be different. When there are two communities, Abbe et al. (2017) use the sign of the eigenvector associated with the second largest eigenvalue (in absolute value) to identify the node's membership. When $K > 2$, just checking the sign is not sufficient to identify all K groups. Our paper shows that applying the K-means algorithm to the first K eigenvectors can achieve strong consistency.

Third, we consider SBM with both regularization and degree correction. We show that, by regularization, the strong consistency is still possible even when the minimal degree does not diverge at all. For the DC-SBM with regularization, we also derive the conditions for strong consistency. Neither regularization nor degree-corrected SBM is discussed in Abbe et al. (2017).

In the simulation, we consider both standard SBMs and DC-SBMs. For standard SBMs, we adopt Joseph and Yu (2016)'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 ((Amini et al., 2013)). 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 describes the proof strategy of the key theorem in our paper. Section 6 concludes. The proofs of the main results are relegated to the mathematical appendix.

Notation. Throughout the paper, we use $[M]_{ij}$ and $[M]_{i\cdot}$ to denote the (i,j)-th entry and i-th row of matrix M, respectively. Without confusion, we sometimes simplify $[M]_{ij}$ as M_{ij}. $\|M\|$ and $\|M\|_F$ denote the spectral norm and Frobenius norm of M, respectively. Note that $\|M\| = \|M\|_F$ when M is a vector. In addition, let $\|M\|_{2\to\infty} = \sup_i \|[M]_{i\cdot}\|$. We use $\mathbf{1}\{\cdot\}$ to denote the indicator function which takes value 1 when $\cdot$ holds and 0 otherwise. C_1 and c_1 denote specific absolute constants that remain the same throughout the paper.

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 Lei's 2016 sequential goodness-of-fit testing procedure, the likelihood-based model selection method proposed by Wang and Bickel (2017), or the network cross-validation method proposed by Chen and Lei (2017). 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 membership 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 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 $\Omega_n = diag(\omega_{1n}, \ldots, \omega_{Kn})$ such that $|\omega_{1n}| \geq \cdots \geq |\omega_{Kn}| > 0$.*

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 $B_0 = \Pi_n^{-1/2} S_n \Omega_n S_n^T \Pi_n^{-1/2}$. The full-rank assumption is also made in Rohe et al. (2011), Lei and Rinaldo (2015), and Joseph and Yu (2016) and can be relaxed at the cost of more complicated notation.[2] In addition, we allows for the possibility that $K \to \infty$ and/or $\omega_{Kn} \to 0$ as $n \to \infty$ below. This also mitigates concern of the full-rank condition. Assumption 1 implies that $\mathcal{L}$ has rank K and the following spectral decomposition:

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

where $\Sigma_n = \text{diag}(\sigma_{1n}, \ldots, \sigma_{Kn}, 0, \ldots, 0)$ is a $n \times n$ matrix that contains the eigenvalues of $\mathcal{L}$ such that $|\sigma_{1n}| \geq |\sigma_{2n}| \geq \cdots \geq |\sigma_{Kn}| > 0$, $\Sigma_{1n} = \text{diag}(\sigma_{1n}, \ldots, \sigma_{Kn})$, 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_1 and c_1 such that*

$$\infty > C_1 \geq \limsup_n \sup_k n_k K/n \geq \liminf_n \inf_k n_k K/n \geq c_1 > 0.$$

Assumption 2 implies that the network has balanced communities. It is commonly assumed in the literature on strong consistency of community detection; see, e.g., Bickel and Chen (2009), Zhao et al. (2012), Amini et al. (2013), and Abbe and Sandon (2015).

Theorem 2.1. *Let $z_i^T = [Z]_{i\cdot}$, the i-th row of Z. If Assumptions 1 and 2 hold, then $\Omega_n = \Sigma_{1n}$, $U_{1n} = Z(Z^T Z)^{-1/2} S_n$ and*

$$\sup_{1 \leq i \leq n} (n/K)^{1/2} \| z_i^T (Z^T Z)^{-1/2} S_n \| \leq c_1^{-1/2}.$$

In addition, for n sufficiently large, if $z_i \neq z_j$, then

$$(n/K)^{1/2} \| (z_i^T - z_j^T)(Z^T Z)^{-1/2} S_n \| \geq C_1^{-1/2} \sqrt{2} > 0.$$

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 K 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] The first version of our paper only requires that B_0 has distinct rows and rank K^*, which can be less than K. Then, researchers need to apply K-means algorithm to the first K^* eigenvectors. By modifying the corresponding assumptions accordingly, the strong consistency result in this paper still holds. We stick to the full rank case mainly for notation simplicity.

2.3 Uniform bound for the estimated eigenvectors

To study the upper bound of the eigenvectors of L associated with its K largest eigenvalues, we add the following assumption.

Assumption 3. *Let* $\mu_n = \min_i d_i$ *and* $\rho_n = \max(\sup_{k_1k_2}[B_0]_{k_1k_2}, 1)$. *Then, for* n *being sufficiently large,*

$$\frac{K\rho_n \log^{1/2}(n)}{\mu_n^{1/2}\sigma_{Kn}^2}\left(1+\rho_n+\left(\frac{1}{K}+\frac{\log(5)}{\log(n)}\right)^{1/2}\rho_n^{1/2}\right) \le 10^{-8}C_1^{-1}c_1^{1/2}.$$

Several remarks are in order. First, 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

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

we have $\limsup_n \rho_n \le c_1^{-1}K$. Therefore, if the number of blocks is fixed, then ρ_n is also bounded.

Second, if K is fixed and $\liminf_n |\sigma_{Kn}|$ is bounded away from zero, then Assumption 3 reduces to the requirement that $\mu_n \ge \underline{C}\log(n)$ for some constant $\underline{C}$. Therefore, Assumption 3 allows for $\mu_n = \Omega(\log(n))$. Such condition is the minimal requirement for strong consistency (exact recovery), as established in Abbe et al. (2016) and Abbe and Sandon (2015). Our results in Theorem 2.3 based on Assumption 3 imply that, in the baseline case, the spectral clustering method achieve strong consistency under this minimal rate requirement.

Third, to provide a more detailed comparison between Assumption 3 and the phase transition threshold, let us consider the special case that there are two equal sized communities and the block probability matrix is

$$B = \begin{pmatrix} \frac{a\log(n)}{n} & \frac{b\log(n)}{n} \\ \frac{b\log(n)}{n} & \frac{a\log(n)}{n} \end{pmatrix},$$

where $a > b$. In this case, $K = 2$, $\Pi_n = \mathrm{diag}(0.5, 0.5)$, $\mathcal{D}_B = \mathrm{diag}(\frac{(a+b)\log(n)}{2n}, \frac{(a+b)\log(n)}{2n})$, and

$$B_0 = \mathcal{D}_B^{-1/2}B\mathcal{D}_B^{-1/2} = \begin{pmatrix} \frac{2a}{a+b} & \frac{2b}{a+b} \\ \frac{2b}{a+b} & \frac{2a}{a+b} \end{pmatrix}.$$

Note that $\mu_n = \frac{(a+b)\log(n)}{2}$, $\rho_n = \frac{2a}{a+b} \in (1,2)$, and σ_{2n}, the second eigenvalue of $\Pi_n^{1/2}B_0\Pi_n^{1/2}$, is $\frac{a-b}{a+b}$. Then, Assumption 3 boils down to

$$\left(\frac{2a}{a+b}\right)^2\sqrt{\frac{2}{a+b}}\left(\frac{a+b}{a-b}\right)^2 \le \underline{c}$$

for some small constant $0.0001 > \underline{c} > 0$. Since $\frac{2a}{a+b} \ge 1$ and $\frac{a+b}{a-b} > 1$, the above condition implies that

$$\underline{c} \ge \left(\frac{2a}{a+b}\right)^2\sqrt{\frac{2}{a+b}}\left(\frac{a+b}{a-b}\right)^2 \ge \frac{\sqrt{2(a+b)}}{a-b} \ge \frac{\sqrt{a}+\sqrt{b}}{a-b} = \frac{1}{\sqrt{a}-\sqrt{b}},$$

or equivalently,

$$\sqrt{a}-\sqrt{b} \ge \underline{c}^{-1} > \sqrt{2}.$$

Because $\sqrt{2}$ is the information-theoretic threshold for exact recovery established in Abbe et al. (2016), Assumption 3 ensures that the SBM under our consideration is in the region that exact recovery is solvable.

Fourth, the constants in Assumption 3, and thus, $\underline{c}$ in the above remark, are not optimal. We choose these constants purely for their technical ease. We conjecture that more sophisticated arguments such as those in Abbe and Sandon (2015), Abbe et al. (2016), and Abbe et al. (2017) are needed to establish the optimal constant for the exact recovery of spectral clustering method. On the other hand, although our method cannot show the exact recovery all the way down to the information-theoretic threshold, it can be easily extended to handle degree-corrected and/or regularized SBM, as shown in Section 3.

Consider the spectral decomposition

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

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

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

The following lemma indicates that L and $\hat{U}_{1n}$ are close to their population counterparts, and up to an orthogonal matrix in the latter case.

Lemma 2.1. *If Assumptions 1–3 hold, then there exists a $K \times K$ orthogonal (random) matrix $\hat{O}_n$ such that*

$$\|\mathcal{L} - L\| \leq 7 \log^{1/2}(n) \mu_n^{-1/2} \quad a.s.$$

and

$$\|\hat{U}_{1n} \hat{O}_n - U_{1n}\| \leq 10 \log^{1/2}(n) \mu_n^{-1/2} |\sigma_{Kn}^{-1}| \quad a.s.$$

Two variants of Lemma 2.1 have been derived in Joseph and Yu (2016) and Qin and Rohe (2013) as special cases. The main difference is that we obtain the almost sure bound for the objects of interest instead of the probability bound in those papers. As illustrated in Abbe et al. (2017),

$$\hat{O}_n = \bar{U} \bar{V}^T,$$

where $\bar{U} \bar{\Sigma} \bar{V}^T$ is the singular value decomposition of $\hat{U}_{1n}^T U_{1n}$. Apparently, $\hat{O}_n$ is random.

In order to study the strong consistency, we have to derive the uniform bound for $\|\hat{u}_{1i}^T \hat{O}_n - u_{1i}^T\|$, where $\hat{u}_{1i}^T$ and u_{1i}^T are the i-th rows of $\hat{U}_{1n}$ and U_{1n}, respectively.

Theorem 2.2. *If Assumptions 1–3 hold, then*

$$\sup_i \sqrt{n/K} \|\hat{u}_{1i}^T \hat{O}_n - u_{1i}^T\| \leq C^* \frac{\rho_n \log^{1/2}(n)}{\mu_n^{1/2} \sigma_{Kn}^2} \left(1 + \rho_n + \left(\frac{1}{K} + \frac{\log(5)}{\log(n)} \right)^{1/2} \rho_n^{1/2} \right) \quad a.s.,$$

where C^* is the same absolute constant as in Theorem 3.5.

We consider the four-parameter SBM studied in Rohe et al. (2011) to illustrate the upper bound in Theorem 2.2.

Example 2.1. *The SBM 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, $\rho_n = \frac{(p+r)K}{p+rK}$, $\sigma_{Kn} = \frac{p}{Kr+p}$, and $\mu_n = \frac{n(p+rK)}{K} - (p+r)$. Therefore, the probability bound of $\sup_i \sqrt{n/K}\|\hat{u}_{1i} - O_n^T u_{1i}\|$ is of order*

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

The above display is small if $K^5\log(n)/(np)$ is small and $rK/p \to c \in (0,\infty)$, or if $K^4\log(n)/(nr)$ is small 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 displayed item in (2.1) becomes small as long as $K^4\log(n)/n$ is small.

Since both U_{1n} and $\hat{U}_{1n}$ 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 $\hat{U}_{1n}$ also contain the community membership information. Let $\hat{\beta}_{in} = (n/K)^{1/2}\hat{u}_{1i}^T$. Let $g_i^0 \in \{1,\ldots,K\}$ denote the true community that node i belongs to. Theorems 2.1-2.2 and the fact that $\hat{O}_n\hat{O}_n^T = I_K$ imply that there exist $\beta_{kn} = (K\pi_{kn})^{-1/2}[S_n\hat{O}_n^T]_{k\cdot}$, $k = 1,\cdots,K$ such that

$$(n/K)^{1/2}u_{1i}^T\hat{O}_n^T = \beta_{g_i^0 n}, \quad ||\beta_{kn}|| \le c_1^{-1/2},$$

and

$$\sup_i \|\hat{\beta}_{in} - \beta_{g_i^0 n}\| \le C^* \frac{\rho_n \log^{1/2}(n)}{\mu_n^{1/2}\sigma_{Kn}^2}\left(1 + \rho_n + \left(\frac{1}{K} + \frac{\log(5)}{\log(n)}\right)^{1/2}\rho_n^{1/2}\right) \quad a.s.$$

If the distance between $\hat{\beta}_{in}$ and $\beta_{g_i^0 n}$ is much smaller than that among distinctive $\{\beta_{kn}\}_{k=1}^K$, then K-means algorithm applying to $\{\hat{\beta}_{in}\}_{i=1}^n$ are expected to recover the true community memberships. The statistical properties of K-means method are studied in the next two sections.

2.4 Strong consistency of the K-means algorithm

By abuse of notation, let $\hat{\beta}_{in} \in \Re^K$ be a generic estimator of $\beta_{g_i^0 n} \in \Re^K$ 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 we compute the estimated cluster identity as

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

where 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 vectors $\{\beta_{kn}\}_{k=1}^K$ satisfy the following restrictions.

Assumption 4. *1. There exists a constant M such that*

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

2. There exist some deterministic sequences c_{1n} and c_{2n} such that $\sup_i \|\hat\beta_{in} - \beta_{g_i^0 n}\| \le c_{2n} \le M$ a.s. and $\inf_{1\le k<k'\le K} \|\beta_{kn} - \beta_{k'n}\| \ge c_{1n} > 0$.

3. $(2c_{2n}c_1^{1/2} + 16K^{3/4}M^{1/2}c_{2n}^{1/2})^2 \le c_1 c_{1n}^2$.

Assumption 4.1 requires that the centroids are uniformly bounded. Assumption 4.2 requires that the centroids are well-separated and the vectors to be classified (i.e., $\{\hat\beta_{in}\}$) are sufficiently close to one of the centroids. Assumption 4.3 requires that the distance between the estimated vector and the corresponding centroid is smaller than that among any of the two distinctive centroids. When the number of clusters K is fixed and the gap c_{1n} between the centroids is bounded away from zero, Assumption 4.3 holds as long as c_{2n} is sufficiently small. Note here, we do not necessarily need $c_{2n} = o(1)$, i.e., $\hat\beta_{in}$ is not necessarily consistent.

Let $H(\cdot,\cdot)$ denote the Hausdorff distance between two sets and $\mathcal{B}_n = \{\beta_{1n},\dots,\beta_{Kn}\}$. The following lemma shows that the K-means algorithm can estimate the true centroids $\{\beta_{kn}\}_{k=1}^K$ up to the rate $O_{a.s.}(c_{2n}^{1/2}K^{3/4})$.

Lemma 2.2. *Suppose that Assumptions 2 and 4 hold. Then*

$$H(\widehat{\mathcal{A}}_n, \mathcal{B}_n) \le (15M/c_1)^{1/2} c_{2n}^{1/2} K^{3/4} \quad a.s.$$

Theorem 2.3. *Suppose that Assumptions 2 and 4 hold. Then for sufficiently large n, we have*

$$\sup_{1\le i\le n} \mathbf{1}\{\hat g_i \ne g_i^0\} = 0 \quad a.s.$$

Theorem 2.3 establishes that, under the given conditions, the K-means algorithm yields perfect classification in large samples. Intuitively, as long as the estimated vectors $\{\hat\beta_{in}\}_{i=1}^n$ are uniformly much closer to the true centroid $\beta_{g_i^0 n}$ rather than others, the K-means algorithm can divide each individual into the right group. To achieve strong consistency for our SBM, we need the following condition.

Assumption 5. *For n sufficiently large,*

$$C^* \frac{K^{3/2}\rho_n \log^{1/2}(n)}{\mu_n^{1/2}\sigma_{Kn}^2}\left(1 + \rho_n + \left(\frac{1}{K} + \frac{\log(5)}{\log(n)}\right)^{1/2}\rho_n^{1/2}\right) \le \frac{2c_1^{3/2}C_1^{-1}}{257},$$

where C^* is the absolute constant in Theorem 2.2.

Corollary 2.1. *Suppose that Assumptions 1–3 and 5 hold and the K-means algorithm is applied to $\hat\beta_{in} = (n/K)^{1/2}\hat u_{1i}$ and $\beta_{g_i^0 n} = (n/K)^{1/2}\hat O_n u_{1i}$ Then,*

$$\sup_{1\le i\le n} \mathbf{1}\{\hat g_i \ne g_i^0\} = 0 \quad a.s.$$

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 3 is equivalent to*

$$\frac{(p+r)^4 K^8 \log(n)}{p^4 n(p+rK)} \tag{2.2}$$

being sufficiently small. If rK/p is bounded, then the above display further reduces to $K^8 \log(n)/(np)$, which allows $K = O((np/\log(n))^{1/8})$. As long as p decays to zero no faster than $\log(n)/n$, Assumption 3 holds even when K grows slowly to infinity. On the other hand, if $r/p \to c \in (0, \infty)$, (2.2) reduces to $K^7 \log(n)/(nr)$. In addition, if both p and r are bounded away from zero, then (2.2) requires that $K^7 \log(n)/n$ is sufficiently small. In contrast, Rohe et al. (2011) 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 $o\left(K^3 \log^2(n)\right) = o\left(n^{3/4}\right)$.

2.5 Strong consistency of the modified K-means algorithm

It is possible to improve the rate requirement for the number of communities in Assumption 5 by considering a modified K-means algorithm:

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

and $\widetilde{\mathcal{A}}_n = \arg\min_{\mathcal{A}} \widetilde{Q}_n(\mathcal{A})$, where $||\cdot||$ still denote the Euclidean distance. Denote $\widetilde{\mathcal{A}}$ as $\{\widetilde{\alpha}_1, \cdots, \widetilde{\alpha}_K\}$. Then, we compute the estimated cluster identity as

$$\tilde{g}_i = \arg\min_{1 \le l \le K} \|\hat{\beta}_{in} - \widetilde{\alpha}_l\|,$$

where if there are multiple l's that achieve the minimum, $\tilde{g}_i$ takes value of the smallest one.

Assumption 6. *1. There exist some deterministic sequences c_{1n} and c_{2n} such that $\sup_i \|\hat{\beta}_{in} - \beta_{g_i^0 n}\| \le c_{2n}$ a.s. and $\inf_{1 \le k < k' \le K} \|\beta_{kn} - \beta_{k'n}\| \ge c_{1n} > 0$.*

2. $15 K c_{2n} \le c_1 c_{1n}$.

The following two results parallel Lemma 2.2 and Theorem 2.3.

Lemma 2.3. *Suppose that Assumptions 2 and 6 hold. Then*

$$H(\widetilde{\mathcal{A}}_n, \mathcal{B}_n) \le 3 K c_1^{-1} c_{2n} \quad a.s.$$

Theorem 2.4. *Suppose that Assumptions 2 and 6 hold. Then for sufficiently large n, we have*

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

In order to apply the modified K-means algorithm in spectral clustering, we only need to verify conditions in Assumption 6.

Assumption 7. *Suppose there exists some constant c^* such that, for n sufficiently large,*

$$15 C^* \frac{K \rho_n \log^{1/2}(n)}{\mu_n^{1/2} \sigma_{Kn}^2} \left(1 + \rho_n + \left(\frac{1}{K} + \frac{\log(5)}{\log(n)}\right)^{1/2} \rho_n^{1/2}\right) \le c_1 C_1^{-1/2} \sqrt{2},$$

where C^* is the absolute constant in Theorem 2.2.

Corollary 2.2. *Suppose that Assumptions 1–3 and 7 hold and the K-means algorithm is applied to* $\hat{\beta}_{in} = (n/K)^{1/2}\hat{u}_{1i}$ *and* $\beta_{g_i^0 n} = (n/K)^{1/2}\hat{O}_n u_{1i}$ *Then,*

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

Corollary 2.2 implies that the community memberships estimated by the modified K-means can recover the truth. Assumption 7 implies a weaker requirement in the rate of K than Assumption 5, as the exponent for K is reduced from 1.5 in Assumption 5 to 1 in Assumption 7. To derive the optimal rate for K may be much more difficult. We leave it as one topic for future research. We investigate the performance of the K-means algorithm in Section 4.

Like spectral clustering, semidefinite programming (SDP) has also become very popular in the community detection literature. Numerically, SDP relaxation enjoys the computational feasibility that spectral clustering has, and various efficient algorithms have been proposed to solve different types of SDP. Theoretically, under the ordinary SBM, SDP methods have been shown to be capable in detecting communities; see, Abbe et al. (2016), Ames (2014), Bandeira, Boumal and Voroninski (2016), Chen, Sanghavi and Xu (2012), Chen, Jalali, Sanghavi and Xu (2014), Cai and Li (2015), Hajek, Wu and Xu (2016a), and Hajek, Wu and Xu (2016b), among others, and Li, Chen and Xu (2018) for an excellent survey. In particular, Abbe et al. (2016) propose an efficient SDP algorithm to solve a standard SBM with two communities, and show that it succeeds in recovering the true communities with high probability when certain threshold conditions are satisfied; Cai and Li (2015) propose a new SDP-based convex optimization method for a generalized SBM and show that a SDP relaxation followed by a K-means clustering can accurately detect the communities with small misclassification rate and the method is both computationally fast and robust to different kinds of outliers. In contrast, Cai and Li (2015) and Joseph and Yu (2016) show that the standard spectral clustering applied to the graph Laplacian may not work due to the existence of small and weak clusters. The possible presence of weak clusters in SBMs motivates the use of regularization to be studied in the following section.

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 Amini et al. (2013) and Joseph and Yu (2016), 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 $D_\tau = \mathrm{diag}(\hat{d}_1 + \tau, \ldots, \hat{d}_n + \tau)$. The regularized version of P and $\mathcal{D}$ are denoted as P_τ and $\mathcal{D}_\tau$ and defined as

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

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 $W_k^\tau = [B^\tau]_{k\cdot}Z^T\iota_n/n = \sum_{l=1}^K [B^\tau]_{kl}\pi_{ln}$, $\mathcal{D}_B^\tau = \mathrm{diag}(W_1^\tau, \ldots, W_K^\tau)$, 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 Assumption 1 as follows.

Assumption 8. *Suppose B_0^τ has rank 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 $\Omega_n^\tau = diag(\omega_{1n}^\tau, \ldots, \omega_{Kn}^\tau)$ such that $|\omega_{1n}^\tau| \geq \cdots \geq |\omega_{Kn}^\tau| > 0$.*

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 $\Sigma_n^\tau = \mathrm{diag}(\sigma_{1n}^\tau, \ldots, \sigma_{Kn}^\tau, 0, \ldots, 0)$ is an $n \times n$ matrix that contains the eigenvalues of $\mathcal{L}_\tau$ such that $|\sigma_{1n}^\tau| \geq |\sigma_{2n}^\tau| \geq \cdots \geq |\sigma_{Kn}^\tau| > 0$, $\Sigma_{1n}^\tau = \mathrm{diag}(\sigma_{1n}^\tau, \ldots, \sigma_{Kn}^\tau)$, 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 8 hold, then $\Omega_n^\tau = \Sigma_n^\tau$, $U_{1n}^\tau = Z(Z^TZ)^{-1/2}S_n^\tau$ and*

$$\sup_{1\leq i\leq n}(n/K)^{1/2}\|z_i^T(Z^TZ)^{-1/2}S_n^\tau\| \leq c_1^{-1/2}.$$

In addition, there exists a constant c independent of n such that if $z_i \neq z_j$,

$$(n/K)^{1/2}\|(z_i^T - z_j^T)(Z^TZ)^{-1/2}S_n^\tau\| \geq C_1^{-1/2}\sqrt{2} > 0.$$

Since $\mathcal{L}_\tau = n^{-1}ZB_0^\tau Z$, 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 :

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

where $\hat{\Sigma}_n^\tau = \mathrm{diag}(\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{nn}^\tau) = \mathrm{diag}(\hat{\Sigma}_{1n}^\tau, \hat{\Sigma}_{2n}^\tau)$ with $|\hat{\sigma}_{1n}^\tau| \geq |\hat{\sigma}_{2n}^\tau| \geq \cdots \geq |\hat{\sigma}_{nn}^\tau| \geq 0$, $\hat{\Sigma}_{1n}^\tau = \mathrm{diag}(\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{Kn}^\tau)$, and $\hat{\Sigma}_{2n}^\tau = \mathrm{diag}(\hat{\sigma}_{K+1,n}^\tau, \ldots, \hat{\sigma}_{nn}^\tau)$; $\hat{U}_n^\tau = (\hat{U}_{1n}^\tau, \hat{U}_{2n}^\tau)$ is the corresponding eigenvectors such that $(\hat{U}_{1n}^\tau)^T\hat{U}_{1n} = I_K$ and $\hat{U}_{2n}^T\hat{U}_{1n} = 0$. Note that $\hat{U}_{1n}^\tau$ contains the eigenvectors associated with eigenvalues $\hat{\sigma}_{1n}^\tau, \ldots, \hat{\sigma}_{Kn}^\tau$. To study the asymptotic properties of $\hat{U}_{1n}^\tau$, we modify Assumption 3 as follows.

Assumption 9. *Denote $\mu_n^\tau = \min_i d_i + \tau$ and $\rho_n^\tau = \max(\sup_{k_1k_2}[B_0^\tau]_{k_1k_2}, 1)$. Then, for n sufficiently large,*

$$\frac{K\rho_n^\tau\log^{1/2}(n)}{(\mu_n^\tau)^{1/2}(\sigma_{Kn}^\tau)^2}\left(1 + \rho_n^\tau + \left(\frac{1}{K} + \frac{\log(5)}{\log(n)}\right)^{1/2}(\rho_n^\tau)^{1/2}\right) \leq 10^{-8}C_1^{-1}c_1^{1/2}.$$

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

Let $(u_{1i}^\tau)^T$ and $(\hat{u}_{1i}^\tau)^T$ be the i-th row of U_{1n}^τ and $\hat{U}_{1n}^\tau$, respectively.

Theorem 3.2. *Suppose that Assumptions 2, 8, and 9 hold. Then there exists a $K \times K$ orthonormal matrix $\hat{O}_n^\tau$ such that*

$$\sup_{1\le i\le n} \sqrt{n/K}\|(\hat{u}_{1i}^\tau)^T\hat{O}_n^\tau - (u_{1i}^\tau)^T\| \le C^* \frac{\rho_n^\tau \log^{1/2}(n)}{(\mu_n^\tau)^{1/2}(\sigma_{Kn}^\tau)^2}\left(1+\rho_n^\tau + \left(\frac{1}{K}+\frac{\log(5)}{\log(n)}\right)^{1/2}(\rho_n^\tau)^{1/2}\right) \quad a.s.,$$

where C^* is the same absolute constant defined in Theorem 2.2.

The following assumption parallels Assumptions 5 and 7. The following theorem parallels Theorem 2.2.

Assumption 10. *1. For n sufficiently large,*

$$C^* \frac{K^{3/2}\rho_n^\tau \log^{1/2}(n)}{(\mu_n^\tau)^{1/2}(\sigma_{Kn}^\tau)^2}\left(1+\rho_n^\tau + \left(\frac{1}{K}+\frac{\log(5)}{\log(n)}\right)^{1/2}(\rho_n^\tau)^{1/2}\right) \le \frac{2c_1^{3/2}C_1^{-1}}{257},$$

where C^* is the absolute constant in Theorem 3.2.

2. For n sufficiently large,

$$15C^* \frac{K\rho_n^\tau \log^{1/2}(n)}{(\mu_n^\tau)^{1/2}(\sigma_{Kn}^\tau)^2}\left(1+\rho_n^\tau + \left(\frac{1}{K}+\frac{\log(5)}{\log(n)}\right)^{1/2}(\rho_n^\tau)^{1/2}\right) \le c_1 C_1^{-1/2}\sqrt{2},$$

where C^* is the absolute constant in Theorem 3.2.

The following theorem parallels Corollaries 2.1 and 2.2 in Section 2.3.

Theorem 3.3. *Suppose that Assumptions 2, 8, and 9 hold. If Assumption 10.1 holds and the K-means algorithm defined in Section 2.4 is applied to $\hat{\beta}_{in} = \sqrt{n/K}(\hat{u}_{1i}^\tau)^T$ and $\beta_{g_i^0 n} = (n/K)^{1/2}\hat{O}_n^\tau u_{1i}^\tau$. Denote the estimated community identities as $\{\hat{g}_i\}_{i=1}^n$. Then for sufficiently large n, we have*

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

If Assumption 10.2 holds and the modified K-means algorithm defined in Section 2.5 is applied to $\hat{\beta}_{in} = \sqrt{n/K}(\hat{u}_{1i}^\tau)^T$ and $\beta_{g_i^0 n} = (n/K)^{1/2}\hat{O}_n^\tau u_{1i}^\tau$. Denote the estimated community identities as $\{\tilde{g}_i\}_{i=1}^n$. Then, for sufficiently large n, we have

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

As in the standard SBM case, $\hat{O}_n^\tau = \bar{U}^\tau(\bar{V}^\tau)^T$, where $\bar{U}^\tau\bar{\Sigma}^\tau(\bar{V}^\tau)^T$ is the singular value decomposition of $(\hat{U}_{1n}^\tau)^T U_{1n}^\tau$.Theorem 3.3 indicates that the regularized spectral clustering, in conjunction with the standard or modified K-means algorithm, consistently recovers the community membership for all nodes almost surely in large samples.

To see the effect of regularization, let K be fixed and $|\sigma_{Kn}^\tau|$ be bounded away from zero. Then, Assumption 9 boils down to $\log(n)/\mu_n^\tau \le \underline{c}$ for some sufficiently small $\underline{c}$. 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 $\tau/\log(n) = \Omega(1)$ such that Assumption 9 holds. This implies that we can obtain strong consistency for some SBMs in which some nodes have very limited number of links.

In addition, regularization introduces a trade-off between $|\sigma^\tau_{Kn}|$ and μ^τ_n. As τ increases, μ^τ_n increases and the rows of B^τ_0 become more similar, which means that $|\sigma^\tau_{Kn}|$ decreases. Rohe et al. (2011) and Joseph and Yu (2016) explore such intuition to choose the regularizer. Following their leads, we choose over a grid of τ and find the one that minimizes

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

where $\hat{\mathcal{L}}_\tau$ is an estimator of $\mathcal{L}_\tau$. We refer to our Section 4 for more details.

The following is a non-trivial SBM which does not satisfy Assumption 3 but satisfies Assumption 9.

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 3 does not hold. However, for some τ such that $\tau = \Omega(\log(n))$, we have

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

and $d^\tau_i = 0.2n + 0.6 + \tau(1-n^{-1})$ for node i in cluster 1 and $d^\tau_i = 3 - 4n^{-1} + \tau(1-n^{-1})$ for node i in cluster 2. In addition, it is easy to see that

$$B^\tau_0 = \begin{pmatrix} \frac{0.4+\tau n^{-1}}{0.2+(1+\tau)n^{-1}} & \frac{2+\tau}{[0.2n+(1+\tau)]^{1/2}(3+\tau)^{1/2}} \\ \frac{2+\tau}{[0.2n+(1+\tau)]^{1/2}(3+\tau)^{1/2}} & \frac{4+\tau}{3+\tau} \end{pmatrix} \to \begin{pmatrix} \frac{0.4+c_0}{0.2+c_0} & \sqrt{\frac{c_0}{0.2+c_0}} \\ \sqrt{\frac{c_0}{0.2+c_0}} & 1 \end{pmatrix},$$

when $c_0 = \lim_{n\to\infty} \tau/n \in [0,1)$. Apparently, B^τ_0 has full rank and Assumption 9 holds. Therefore, the strong consistency of the regularized spectral clustering still holds.

Let $\sigma^\tau_{2,n}$ denote the second eigenvalue of $\Pi_n^{1/2} B^\tau_0 \Pi_n^{1/2}$. Then as $n \to \infty$,

$$\sigma^\tau_{2,n} \to \frac{0.3 + c_0 - \sqrt{c_0^2 + 0.2c_0 + 0.01}}{2(c_0 + 0.2)} = \frac{0.2}{0.3 + c_0 + \sqrt{c_0^2 + 0.2c_0 + 0.01}},$$

where $c_0 \in [0,1)$. The minimal degree $\mu^\tau_n \asymp \tau$. Then, $Q(\tau) = O(\frac{1}{\sigma^\tau_{2,n}(\mu^\tau_n)^{1/2}})$ where

$$\frac{1}{\sigma^\tau_{2,n}(\mu^\tau_n)^{1/2}} \asymp \frac{0.3 + c_0 + \sqrt{c_0^2 + 0.2c_0 + 0.01}}{0.2\tau^{1/2}}.$$

In order to achieve maximal convergence rate, we need $c_0 \neq 0$. For simplicity, we just assume $\tau = c_0 n$. Then, the constant $\frac{0.3+c_0+\sqrt{c_0^2+0.2c_0+0.01}}{c_0^{1/2}}$ achieves minimum on $(0,1)$ at $c_0 = 0.2$.

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.

Example 3.2. *Consider the above example with B replaced by*

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

and $\tau/\log(n) = \Omega(1)$. 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 9 cannot be satisfied in this case because $\mu_n^\tau |\sigma_{Kn}^\tau|^4/\log(n)$ is converging to zero at rate $1/(\tau^3 \log(n))$. Consequently, we cannot show that $\sup_i \sqrt{n}\|(\hat{O}_n^\tau)^T \hat{u}_{1i}^\tau - u_{1i}^\tau\|$ is sufficiently small or prove strong consistency in this case.

The above example shows that 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 Karrer and Newman (2011), 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 $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$. 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., (Qin and Rohe, 2013; Zhao et al., 2012)) 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 when $\theta_i = 1$ 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 an $n \times n$ 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 Qin and Rohe (2013) 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 Joseph and Yu (2016), 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 $\mathcal{D} = \mathrm{diag}(d_1, \ldots, d_n)$ 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 assume the following:

Assumption 11. *1. There exists a sequence ρ_n such that $\rho_n \geq 1$ and $B_0 \leq \rho_n$ element-wise.*

2. B_0 has full rank K.

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

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

where $\Sigma_n = \mathrm{diag}(\sigma_{1n}, \ldots, \sigma_{Kn})$ 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 $\Theta_\tau = \mathrm{diag}(\theta_1^\tau, \ldots, \theta_n^\tau)$ 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 3.4. *Suppose Assumptions 11 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, (1) there exists a $K \times K$ matrix S_n^τ such that $U_{1n} = \Theta_\tau^{1/2} Z (Z^T \Theta_\tau Z)^{-1/2} S_n^\tau$, (2) $(n_{g_i^0}^\tau)^{1/2} (\theta_i^\tau)^{-1/2} \|u_i^T\| = 1$, and (3) if $z_i = z_j$, then $\| \frac{u_i}{\|u_i\|} - \frac{u_j}{\|u_j\|} \| = 0$; if $z_i \neq z_j$, then $\| \frac{u_i^T}{\|u_i^T\|} - \frac{u_j^T}{\|u_j^T\|} \| = \sqrt{2}$.*

Theorem 3.4 follows Qin and Rohe (2013, Lemma 3.3). In particular, Theorem 3.4(3) 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 12. *There exist two constants C_1 and c_1 such that*

$$\infty > C_1 \geq \limsup_n \sup_{1\leq i\leq n} n_{g_i^0}^\tau d_i^\tau K/(nd_i) \geq \liminf_n \inf_{1\leq i\leq n} n_{g_i^0}^\tau d_i^\tau K/(nd_i) \geq c_1 > 0.$$

Assumption 12 holds for the simplest case in which the degrees are homogeneous within the same cluster. Note that in this case, $n_{g_i^0}^\tau = n_{g_i^0} d_i/d_i^\tau$, which may be of smaller order of magnitude of n/K if $d_i/\tau \to 0$. However, Assumption 12 still holds because the factor d_i/d_i^τ is removed. In general, Assumption 12 holds if d_i is of the same order of magnitude for all i in the same cluster.

Assumption 13. *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, for n sufficiently large,*

1. $\frac{\bar{\theta}^{1/2}\log^{1/2}(n)}{\underline{\theta}^{1/2}(\mu_n^\tau)^{1/2}\rho_n} \leq 10^{-4}$,
2. $$\left(K\frac{\rho_n \log^{1/2}(n)}{(\mu_n^\tau)^{1/2}\,\sigma_{Kn}^2}\right)\left(\frac{\left(\frac{1}{K}+\frac{\log(5)}{\log(n)}\right)^{1/2}\rho_n^{1/2}\overline{\theta}^{1/4}}{\underline{\theta}^{1/4}}+\rho_n+1\right) \leq 10^{-8}C_1^{-1}c_1^{1/2}, \quad \textit{and}$$
3. *there exists a positive constant c such that $\underline{\theta} \geq n^{-c}$.*

Assumption 13 specifies conditions on d_i, θ_i, and σ_{Kn}. The same remarks after Assumption 3 apply. Admittedly, the constants in Assumption 13 are not optimal. We choose them purely for technical ease. If $0 < \underline{\theta} \leq \overline{\theta} < \infty$, then Assumption 13.1 is nested by Assumption 13.2, which is similar to Assumption 3. If in addition, K is fixed and $\liminf_n |\sigma_{Kn}| > 0$, then Assumption 13.2 further boils down to $\log(n)/\mu_n^\tau \leq \underline{c}$ for some sufficiently small $\underline{c}$. This indicates that even if the minimal degree μ_n is bounded, Assumption 13.2 still holds if $\tau = \Omega(\log(n))$.

Consider the spectral decomposition of L_τ', the sample counterpart of $\mathcal{L}_\tau'$, as

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

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

The following lemma parallels Lemma 2.1.

Lemma 3.1. *If Assumptions 11–13 hold, then*

$$\|\mathcal{L}_\tau' - L_\tau'\| \leq 7(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.$$

and

$$\|\hat{U}_{1n}\hat{O}_n - U_{1n}\| \leq 10(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1} \quad a.s.,$$

where $\hat{O}_n = \bar{U}\bar{V}^T$ is a $K\times K$ orthogonal matrix and $\bar{U}\bar{\Sigma}\bar{V}^T$ for some diagonal matrix $\bar{\Sigma}$ is the singular value decomposition of $\hat{U}_{1n}^T U_{1n}$.

In order to obtain the strong consistency, we need to derive the uniform bound for $\|\hat{u}_i^T \hat{O}_n - u_i^T\|$, where $\hat{u}_i^T$ and u_i^T are the i-th rows of $\hat{U}_{1n}$ and U_{1n}, respectively.

Theorem 3.5. *If Assumptions 11–13 hold, then*

$$\sup_i (n_{g_i^0}^\tau)^{1/2} (\theta_i^\tau)^{-1/2} \|\hat{u}_i^T \hat{O}_n - u_i^T\| \leq C^* \eta_n \quad a.s.,$$

where C^* is an absolute constant specified in the proof and

$$\eta_n = \left(\frac{\rho_n \log^{1/2}(n)}{(\mu_n^\tau)^{1/2} \sigma_{Kn}^2} \right) \left(\frac{\left(\frac{1}{K} + \frac{\log(5)}{\log(n)} \right)^{1/2} \rho_n^{1/2} \overline{\theta}^{1/4}}{\underline{\theta}^{1/4}} + \rho_n + 1 \right).$$

Theorem 3.5 is essential to establish the strong consistency result. The following Assumption specifies the rate requirement for strong consistency depending on whether the standard or modified K-means algorithm is used.

Assumption 14. *Let C^* denote the absolute constant in Theorem 3.5. For n sufficiently large we have*

1. $C^* K^{3/2} \eta_n \leq \frac{c_1}{257}$,
2. $30 C^* K \eta_n \leq c_1 \sqrt{2}$.

Corollary 3.1. *If Assumptions 11–13 hold, then*

$$\sup_i \left\| \frac{\hat{u}_i}{\|\hat{u}_i\|} - \frac{\hat{O}_n u_i}{\|\hat{O}_n u_i\|} \right\| \leq 2 C^* \eta_n \quad a.s. \tag{3.2}$$

If Assumption 14.1 holds and the K-means algorithm is applied to $\hat{\beta}_{in} = \hat{u}_{1i}/\|\hat{u}_{1i}\|$ and $\beta_{g_i^0 n} = \hat{O}_n u_{1i}/\|u_{1i}\|$. Denote the obtained community memberships as $\{\hat{g}_i\}_{i=1}^n$. Then,

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

If Assumption 14.2 holds and the modified K-means algorithm is applied to $\hat{\beta}_{in} = \hat{u}_{1i}/\|\hat{u}_{1i}\|$ and $\beta_{g_i^0 n} = \hat{O}_n u_{1i}/\|u_{1i}\|$. Denote the obtained community memberships as $\{\tilde{g}_i\}_{i=1}^n$. Then,

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

Corollary 3.1 justifies the use of standard and modified K-means algorithms on $\hat{u}_{in}/\|\hat{u}_{in}\|$ provided the bound on the right hand side of (3.2) is $O\left(1/K^{3/2}\right)$ and $O(K)$, respectively, which is ensured by Assumptions 14.1 and 14.2, respectively.

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 $D''_\tau = \text{diag}(A''_\tau \iota_n)$. 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 $\mathcal{D}''_\tau = \text{diag}(P''_\tau \iota_n) = D + \tau \Theta$.

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 Wilson, Stevens and Woodall (2016) and estimate Θ by $\hat{\Theta} = \text{diag}(\hat{\theta}_1, \cdots, \hat{\theta}_n)$, where

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

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 15. *1.* $\limsup_n \overline{\theta} < \infty$.

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

Assumption 15.1 requires that the degree of heterogeneity is bounded, which is common in practical applications. Assumption 15.2 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 Amini et al. (2013).

Let $m_k = \sum_{j=1}^n \theta_j B_{k g_j^0}$ 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 3.6. *If Assumption 15 holds, then* $\sup_{1 \le i \le n} |\hat{\theta}_i - \theta_i| = O_{a.s.}(\log(n)/\underline{m}_n)$.

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. $L'_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}$ 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 and 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 Amini et al. (2013) to detect the communities in the SBM and the DC-SBM, respectively.[3] To evaluate the classification performance, we consider two criteria: the Correct Classification Proportion (CCP) and the Normalized Mutual Information (NMI). All the simulation results below are computed using the modified K-means algorithm. The simulation results for the standard K-means algorithm can be found in previous versions of this paper. When the regularizer τ is small, the modified K-means algorithm can produce slightly more accurate classification while at the optimal τ selected by our data-driven method explained below, the classification results in terms of CCP and NMI for the two algorithms are basically the same.

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.

[3] As Amini et al. (2013) 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.

DGP 2: Let $K = 3$. Each community has $n/3$ nodes. The matrix B is set 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 Joseph and Yu (2016) 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 $\hat{P} = \hat{Z}\hat{B}\hat{Z}^T$, $\hat{P}_\tau = \hat{P} + \tau J_n$, $\hat{\mathcal{D}}_\tau = \text{diag}(\hat{P}_\tau \iota_n)$, $\hat{\mathcal{L}}_\tau = \hat{\mathcal{D}}_\tau^{-1/2}\hat{P}_\tau\hat{\mathcal{D}}_\tau^{-1/2}$, 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 $j = 1, \ldots, 18$. Qin and Rohe (2013) 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 3 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 Amini et al. (2013). Both results are in line with the theoretical analysis by Joseph and Yu (2016).

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.

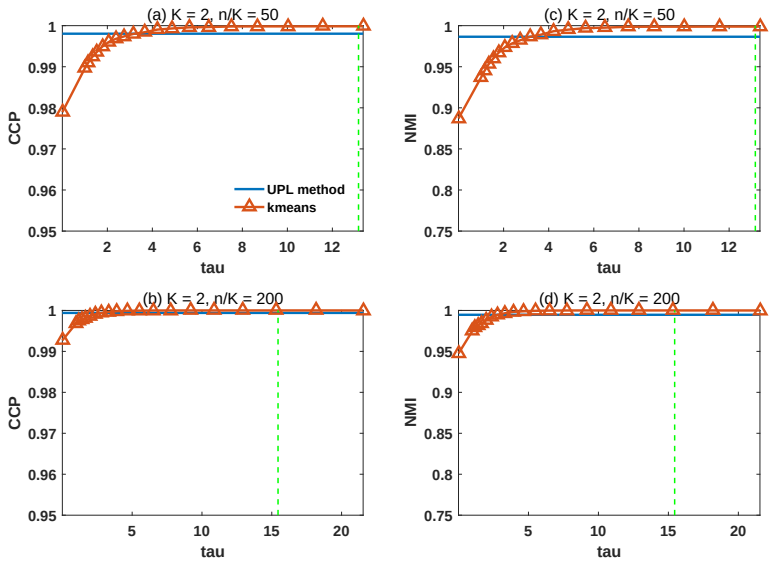

Figure 1: Classification results for K-means for DGP 1 ($K = 2$) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$ and for UPL method. The x-axis marks τ values, and the y-axis is either CCP (left column) or NMI (right column). The green vertical line in each subplot indicates the estimated τ value by using the method of Joseph and Yu (2016). The first and second rows correspond to $n/K = 50$ and 200, respectively.

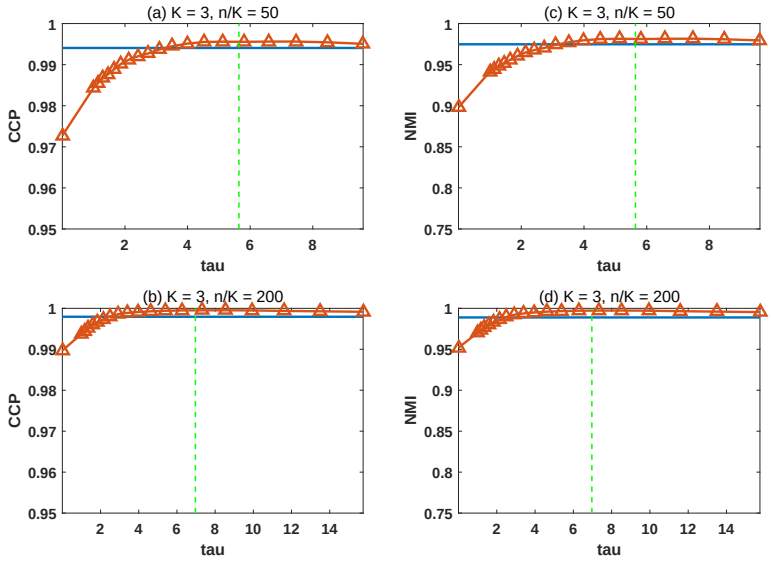

Figure 2: Classification results for DGP 2 ($K = 3$) based on $L_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. (See the explanations in Figure 1.)

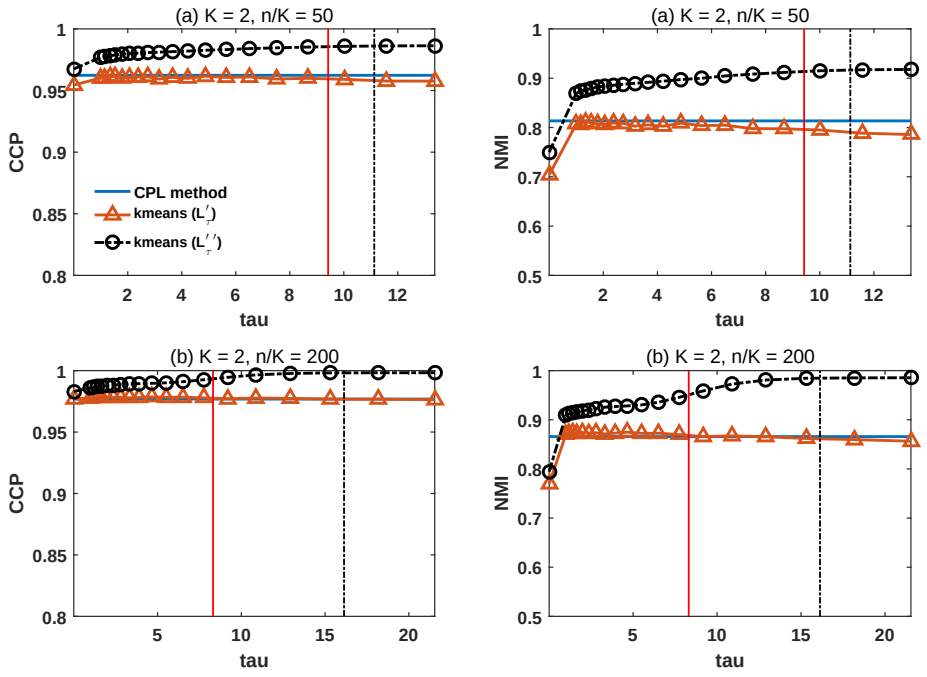

Figure 3: Classification results for DGP 3 ($K = 2$, degree-corrected) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ and $L''_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. The red and black vertical lines correspond to the optimal regularizers τ'^{JY} and τ''^{JY}, respectively. (See Figure 1 for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

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

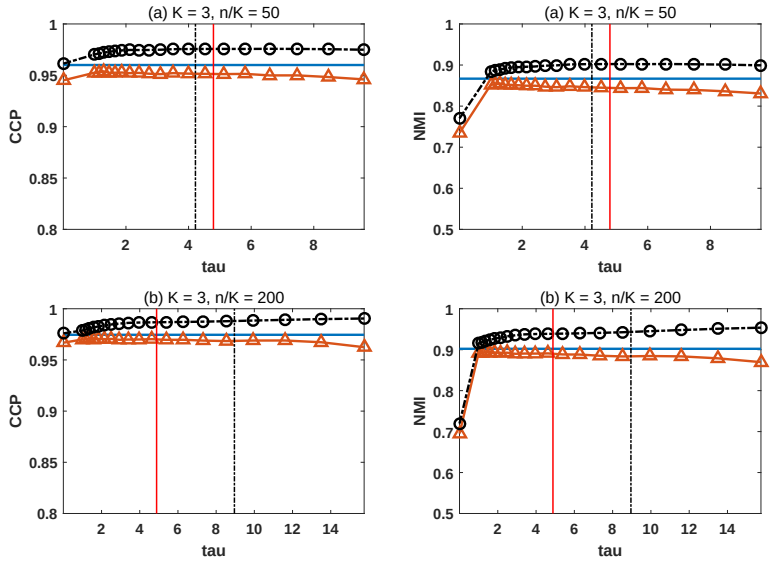

Figure 4: Classification results for DGP 4 ($K = 3$, degree-corrected) based on $L'_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}$ and $L''_\tau = D_\tau^{-1/2} A_\tau D_\tau^{-1/2}$. The red and black vertical lines corresponds to the optimal regularizers τ'^{JY} and τ''^{JY}, respectively. (See Figure 1 for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

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.

Table 1: Comparison of classification results

			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	0.9998	0.9998	0.9980	0.9989	0.9989	0.9865
	2	200	1.0000	1.0000	0.9994	1.0000	1.0000	0.9947
2	3	50	0.9951	0.9956	0.9941	0.9795	0.9812	0.9748
	3	200	0.9992	0.9995	0.9979	0.9954	0.9972	0.9889
3	2	50	0.9576	0.9596	0.9623	0.7857	0.7964	0.8134
	2	200	0.9764	0.9777	0.9769	0.8564	0.8689	0.8658
4	3	50	0.9460	0.9513	0.9600	0.8308	0.8444	0.8668
	3	200	0.9624	0.9701	0.9745	0.8696	0.8902	0.9022

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.

5 Proof strategy

In this section we outline the proof strategies for the main results in Section 3.2. First, noting that the regularized spectral clustering for the DC-SBM nests standard SBM without regularization by setting $\tau = 0$ and $\theta_i = 1 \ \forall \ i = 1, \cdots, n$, all the main results in Section 2 follow that in Section 3.2. Second, based on the results in Section 2, the results for the standard SBM with regularization in Section 3.1 can be derived by replacing B_0, μ_n, ρ_n, and σ_{Kn} by their counterparts with regularization, i.e., B_0^τ, μ_n^τ, ρ_n^τ, and σ_{Kn}^τ, respectively.

Section 3.2 contains Theorems 3.4, 3.5 and 3.6, Lemma 3.1 and Corollary 3.1. Since the proofs of Theorems 3.4 and 3.6, Lemma 3.1 and Corollary 3.1 are relatively simple, below we focus on the proof strategy for Theorem 3.5.

Theorem 3.5 aims to establish a uniform upper bound for each row of the gap between sample and population eigenvectors (up to some rotation), i.e., $\sup_i ||\hat{u}_i^T \hat{O}_n - u_i^T||$, where $\hat{u}_i^T$ and u_i^T are the i-th rows of $\hat{U}_{1n}$ and U_{1n}, respectively. Let $\hat{\Lambda} = L_\tau' \hat{U}_{1n} \hat{O}_n = \hat{U}_{1n} \hat{\Sigma}_n \hat{O}_n$, $\Lambda = \mathcal{L}_\tau' U_{1n} = U_{1n} \Sigma_n$, $\hat{\Lambda}_i = \hat{u}_i^T \hat{\Sigma}_n \hat{O}_n$, and $\Lambda_i = u_i^T \Sigma_n$. Our proof strategy is to obtain the upper and lower bounds for $(n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}||\hat{\Lambda}_i - \Lambda_i||$, both of which involve $(n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}||\hat{u}_i^T \hat{O}_n - u_i^T||$. The two bounds produce a contraction mapping for $\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}||\hat{u}_i^T \hat{O}_n - u_i^T||$. By iterating the contraction mapping sufficiently many times, we obtain the desired bound.

Lower bound. In order to derive the lower bound for $(n_{g_i^0}^\tau)^{1/2}||\hat{\Lambda}_i - \Lambda_i||$, we note that

$$\begin{aligned}
(n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\| &= (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{u}_i^T \hat{\Sigma}_n \hat{O}_n - u_i^T \Sigma_n\| \\
&\geq (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|(\hat{u}_i^T \hat{O}_n - u_i^T)\hat{\Sigma}_n\| - (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|u_i^T(\hat{\Sigma}_n - \Sigma_n)\| \\
&\quad - (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{u}_i^T(\hat{\Sigma}_n \hat{O}_n - \hat{O}_n \hat{\Sigma}_n)\| \\
&\equiv I_i - II_i - III_i. \qquad (5.1)
\end{aligned}$$

Clearly, by Assumption 13 and Lemma 3.1, $|\hat{\sigma}_{Kn}| \geq 0.999|\sigma_{Kn}|$ *a.s.*, and thus,

$$\sup_i I_i \geq 0.999|\sigma_{Kn}|\Gamma_n \quad a.s.,$$

where $\Gamma_n = \sup_i |(n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{u}_i^T \hat{O}_n - u_i^T\|$. It is the leading term of the lower bound involving Γ_n. In the online Appendix B, we show that $\sup_i II_i \leq 7(\log(n)/\mu_n^\tau)^{1/2} \quad a.s.$ and $\sup_i III_i \leq 34(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}|^{-1}(\Gamma_n + 1) \quad a.s.$ It follows that

$$\begin{aligned}
\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\| &\geq (0.999|\sigma_{Kn}| - 34(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-1}|)\Gamma_n - 41(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-1}| \\
&\geq 0.99|\sigma_{Kn}|\Gamma_n - 41(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-1}|, \qquad (5.2)
\end{aligned}$$

where we use the fact that $34(\log(n)/\mu_n^\tau)^{1/2}|\sigma_{Kn}^{-2}| \leq 0.09$.

Upper bound. To derive the upper bound for $\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\|$, we first denote $\tilde{\Lambda} = D_\tau^{-1/2} P D_\tau^{-1/2} U_{1n}$ and $\tilde{\Lambda}_i = (\hat{d}_i^\tau)^{-1/2}[P]_{i\cdot} D_\tau^{-1/2} U_{1n}$ as the i-th row of $\tilde{\Lambda}$. Then, we have

$$\begin{aligned}
\sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\| &\leq \sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\Lambda_i - \tilde{\Lambda}_i\| + \sup_i (n_{g_i^0}^\tau)^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \tilde{\Lambda}_i\| \\
&\equiv T_1 + T_2. \qquad (5.3)
\end{aligned}$$

For T_2, we have

$$\begin{aligned}
T_2 &= \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|(\hat{d}_i^\tau)^{-1/2}[A]_{i\cdot}D_\tau^{-1/2}\hat{U}_{1n}\hat{O}_n - (\hat{d}_i^\tau)^{-1/2}[P]_{i\cdot}D_\tau^{-1/2}U_{1n}\| \\
&\leq \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|[P]_{i\cdot}D_\tau^{-1/2}(\hat{U}_{1n}\hat{O}_n - U_{1n})\| \\
&+ \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})(D_\tau^{-1/2} - \mathcal{D}_\tau^{-1/2})\hat{U}_{1n}\hat{O}_n\| \\
&+ \sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}(\hat{d}_i^\tau)^{-1/2}\|([A]_{i\cdot} - [P]_{i\cdot})\mathcal{D}_\tau^{-1/2}\hat{U}_{1n}\hat{O}_n\| \\
&\equiv T_{2,1} + T_{2,2} + T_{2,3}. \qquad (5.4)
\end{aligned}$$

Lemma C.5 in the online Appendix C provides the upper bounds for T_1, $T_{2,1}$, $T_{2,2}$, and $T_{2,3}$. Taking $T_{2,3}$ as an example, we note that

$$T_{2,3} = \sup_i \sup_{h=\hat{U}_{1n}\hat{O}_n f, f\in S^{K-1}} (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\sum_{j=1}^n (A_{ij} - P_{ij})(\hat{d}_i^\tau d_j^\tau)^{-1/2}h_j.$$

Here, h_j denotes the jth element of h. Lemma C.4 builds a Bernstein-type concentration inequality to upper bound $T_{2,3}$, which involves the l_∞ and l_2 norms of h, In particular, $||h||_\infty$ depends on the rough upper bound $\delta_n^{(0)}$ for Γ_n.[4] One of the technical difficulties is that, due to the correlation between the sample graph Laplacian and its eigenvectors, the sequence of random variables $A_{ij} : j = 1, \cdots, n$ are not independent of $h = \hat{U}_{1n}\hat{O}_n f$ for some $f \in S^{K-1}$. To deal with it, we rely on the "leave-one-out" technique used in Abbe et al. (2017), Bean, Bickel, El Karoui and Yu (2013), Javanmard and Montanari (2015), and Zhong and Boumal (2018). The idea is to approximate the eigenvector by a vector which is independent of one particular row of the sample graph Laplacian. This helps to restore the independence. Then, the approximation errors are bounded in Lemma C.7, which further calls upon Lemmas C.6 and C.8.

At the end, Lemma C.5 establishes that

$$\begin{aligned}
&\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\| \\
\leq& 3450C_1c_1^{-1/2}\rho_n\log^{1/2}(n)(\mu_n^\tau)^{-1/2}|\sigma_{Kn}^{-1}|\left[\delta_n^{(0)} + 1 + \rho_n + \frac{\left(\frac{1}{K} + \frac{\log(5)}{\log(n)}\right)^{1/2}\rho_n^{1/2}\overline{\theta}^{1/4}}{\underline{\theta}^{1/4}}\right], \quad a.s., \qquad (5.5)
\end{aligned}$$

where we can choose $\delta_n^{(0)} = n^{1/2}\underline{\theta}^{-1/2}$. Combining the lower and upper bounds in (5.2) and (5.5) for $\sup_i (n^\tau_{g_i^0})^{1/2}(\theta_i^\tau)^{-1/2}\|\hat{\Lambda}_i - \Lambda_i\|$ and applying Assumption 13, we have

$$0.001\delta_n^{(0)} + 3527C_1c_1^{-1/2}\eta_n \geq \Gamma_n, \qquad (5.6)$$

where η_n is defined in Theorem 3.5.

Iteration. (5.6) suggests that the initial rough upper bound $\delta_n^{(0)}$ for Γ_n can be refined to $\delta_n^{(1)} \equiv 0.001\delta_n^{(0)} + 3527C_1c_1^{-1/2}\eta_n$. Then we can take this new upper bound into the previous calculations to obtain

$$0.001\delta_n^{(1)} + 3527C_1c_1^{-1/2}\eta_n \geq \Gamma_n.$$

[4] In fact, the upper bound for $||h||_\infty$ in the proof, which is denoted as ψ_n, is $\delta_n^{(0)} + 1$.

Therefore, we have constructed a contraction mapping, through which we can refine our upper bound for Γ_n via iterations. We iterate the above calculation t times for some arbitrary integer t, and obtain that

$$\Gamma_n \leq \delta_n^{(t)}, \quad \delta_n^{(t)} = 0.001\delta_n^{(t-1)} + 3527C_1c_1^{-1/2}\eta_n.$$

This implies

$$\delta_n^{(t)} = (0.001)^t \left[\delta_n^{(0)} - 3527C_1c_1^{-1/2}\eta_n\right] + 3527C_1c_1^{-1/2}\eta_n.$$

Letting $t = n$, we have

$$\Gamma_n \leq \delta_n^{(n)} \leq 1000^{-n}n^{1/2}\underline{\theta}^{-1/2} + 3527C_1c_1^{-1/2}\eta_n \leq 3528C_1c_1^{-1/2}\eta_n,$$

where we denote C^* in Theorem 3.5 as $3528C_1c_1^{-1/2}$ and we use the fact that it is possible to choose $\delta_n^{(0)} = n^{1/2}\underline{\theta}^{-1/2}$ as the initial rough bound for Γ_n.

6 Conclusion

In this paper, we show that under suitable conditions, 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 the number of communities is fixed and the probability block matrix has minimal eigenvalue bounded away from zero, the strong consistency essentially requires that the minimal degree diverges to infinity at least as fast as $\log(n)$, which is the minimal rate requirement for the strong consistency discussed in Abbe (2018). Similar results are also established for the regularized DC-SBMs. The simulations confirm our theoretical findings and indicate that an adaptive procedure can improve the finite sample performance of the regularized spectral clustering for DC-SBMs.

References

Abbe, E., 2018. Community detection and stochastic block models: Recent developments. Journal of Machine Learning Research 18, 1–86.

Abbe, E., Bandeira, A.S., Hall, G., 2016. Exact recovery in the stochastic block model. IEEE Transactions on Information Theory 62.

Abbe, E., Fan, J., Wang, K., Zhong, Y., 2017. Entrywise eigenvector analysis of random matrices with low expected rank. arXiv preprint arXiv:1709.09565 .

Abbe, E., Sandon, C., 2015. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery, in: Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on, IEEE. pp. 670–688.

Ames, B.P., 2014. Guaranteed clustering and biclustering via semidefinite programming. Mathematical Programming 147, 429–465.

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, 2097–2122.

Bandeira, A.S., Boumal, N., Voroninski, V., 2016. On the low-rank approach for semidefinite programs arising in synchronization and community detection, in: Conference on learning theory, pp. 361–382.

Bean, D., Bickel, P.J., El Karoui, N., Yu, B., 2013. Optimal m-estimation in high-dimensional regression. Proceedings of the National Academy of Sciences 110, 14563–14568.

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, 21068–21073.

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

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

Cai, T.T., Li, X., 2015. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. The Annals of Statistics 43, 1027–1059.

Chen, K., Lei, J., 2017. Network cross-validation for determining the number of communities in network data. Journal of the American Statistical Association 0, 1–11. doi:10.1080/01621459.2016.1246365, arXiv:https://doi.org/10.1080/01621459.2016.1246365.

Chen, Y., Jalali, A., Sanghavi, S., Xu, H., 2014. Clustering partially observed graphs via convex optimization. The Journal of Machine Learning Research 15, 2213–2238.

Chen, Y., Sanghavi, S., Xu, H., 2012. Clustering sparse graphs, in: Advances in neural information processing systems, pp. 2204–2212.

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

Hajek, B., Wu, Y., Xu, J., 2016a. Achieving exact cluster recovery threshold via semidefinite programming. IEEE Transactions on Information Theory 62, 2788–2797.

Hajek, B., Wu, Y., Xu, J., 2016b. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. IEEE Transactions on Information Theory 62, 5918–5937.

Holland, P.W., Laskey, K.B., Leinhardt, S., 1983. Stochastic blockmodels: First steps. Social networks 5, 109–137.

Javanmard, A., Montanari, A., 2015. De-biasing the lasso: Optimal sample size for gaussian designs. arXiv preprint arXiv:1508.02757 .

Joseph, A., Yu, B., 2016. Impact of regularization on spectral clustering. The Annals of Statistics 44, 1765–1791.

Karrer, B., Newman, M.E., 2011. Stochastic blockmodels and community structure in networks. Physical Review E 83, 016107.

Lei, J., 2016. A goodness-of-fit test for stochastic block models. The Annals of Statistics 44, 401–424.

Lei, J., Rinaldo, A., 2015. Consistency of spectral clustering in stochastic block models. The Annals of Statistics 43, 215–237.

Li, X., Chen, Y., Xu, J., 2018. Convex relaxation methods for community detection. arXiv preprint arXiv:1810.00315 .

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, 2905–2922.

Mossel, E., Neeman, J., Sly, A., 2014. Consistency thresholds for binary symmetric block models. arXiv preprint arXiv:1407.1591 In proc. of STOC15.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical review E 69, 026113.

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. Curran Associates, Inc.. volume 26, pp. 3120–3128.

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

Sarkar, P., Bickel, P.J., 2015. Role of normalization in spectral clustering for stochastic blockmodels. The Annals of Statistics 43, 962–990.

Sussman, D.L., Tang, M., Fishkind, D.E., Priebe, C.E., 2012. A consistent adjacency spectral embedding for stochastic blockmodel graphs. Journal of the American Statistical Association 107, 1119–1128.

Vu, V., 2018. A simple svd algorithm for finding hidden partitions. Combinatorics, Probability and Computing 27, 124–140.

Wang, Y., Bickel, P.J., 2017. Likelihood-based model selection for stochastic block models. The Annals of Statistics 45, 500–528.

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 .

Yun, S.Y., Proutiere, A., 2014. Accurate community detection in the stochastic block model via spectral algorithms. arXiv preprint arXiv:1412.7335 .

Yun, S.Y., Proutiere, A., 2016. Optimal cluster recovery in the labeled stochastic block model, in: Advances in Neural Information Processing Systems, pp. 965–973.

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

Zhong, Y., Boumal, N., 2018. Near-optimal bounds for phase synchronization. SIAM Journal on Optimization 28, 989–1016.