

Article

A Useful Criterion on Studying Consistent Estimation in Community Detection

Huan Qing 

School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China; qinghuan@cumt.edu.cn

Abstract: In network analysis, developing a unified theoretical framework that can compare methods under different models is an interesting problem. This paper proposes a partial solution to this problem. We summarize the idea of using a separation condition for a standard network and sharp threshold of the Erdős–Rényi random graph to study consistent estimation, and compare theoretical error rates and requirements on the network sparsity of spectral methods under models that can degenerate to a stochastic block model as a four-step criterion SCSTC. Using SCSTC, we find some inconsistent phenomena on separation condition and sharp threshold in community detection. In particular, we find that the original theoretical results of the SPACL algorithm introduced to estimate network memberships under the mixed membership stochastic blockmodel are sub-optimal. To find the formation mechanism of inconsistencies, we re-establish the theoretical convergence rate of this algorithm by applying recent techniques on row-wise eigenvector deviation. The results are further extended to the degree-corrected mixed membership model. By comparison, our results enjoy smaller error rates, lesser dependence on the number of communities, weaker requirements on network sparsity, and so forth. The separation condition and sharp threshold obtained from our theoretical results match the classical results, so the usefulness of this criterion on studying consistent estimation is guaranteed. Numerical results for computer-generated networks support our finding that spectral methods considered in this paper achieve the threshold of separation condition.

Keywords: community detection; consistency; mixed membership network; separation condition; sharp threshold



Citation: Qing, H. A Useful Criterion on Studying Consistent Estimation in Community Detection. *Entropy* **2022**, *24*, 1098. <https://doi.org/10.3390/e24081098>

Academic Editors: Liang-Jian Deng, Minyu Feng and Feng Chen

Received: 10 July 2022

Accepted: 8 August 2022

Published: 9 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Networks with latent structure are ubiquitous in our daily life, for example, social networks from social platforms, protein–protein interaction networks, co-citation networks and co-authorship networks [1–15]. Community detection is a powerful tool to learn the latent community structure in networks and graphs in social science, computer science, machine learning, statistical science and complex networks [16–22]. The goal of community detection is to infer a node's community information from the network.

Many models have been proposed to model networks with latent community structure; see [23] for a survey. The stochastic blockmodel (SBM) [24] stands out for its simplicity, and it has received increasing attention in recent years [25–35]. However, the SBM only models a non-overlapping network in which each node belongs to a single community. Estimating mixed memberships of the network whose node may belong to multiple communities has received a lot of attention [36–44]. To capture the structure of the network with mixed memberships, Ref. [36] proposed the popular mixed membership stochastic blockmodel (MMSB), which is an extension of SBM from non-overlapping networks to overlapping networks. It is well known that the degree-corrected stochastic blockmodel (DCSBM) [45] is an extension of SBM by considering the degree heterogeneity of nodes to fit real-world networks with various node degree. Similarly, Ref. [41] proposed a model named the degree-corrected mixed membership (DCMM) model as an extension of MMSB

by considering the degree heterogeneity of nodes. There are alternative models based on MMSB, such as the overlapping continuous community assignment model (OCCAM) of [40] and the stochastic blockmodel with overlap (SBMO) proposed by [46], which can also model networks with mixed memberships. As discussed in Section 5, OCCAM equals DCMM, while SBMO is a special case of DCMM.

1.1. Spectral Clustering Approaches

For the four models SBM, DCSBM, MMSB and DCMM, many researchers focus on designing algorithms with provable consistent theoretical guarantees. Spectral clustering [47] is one of the most widely applied methods with guarantees of consistency for community detection.

Within the SBM and DCSBM frameworks for a non-overlapping network, spectral clustering has two steps. It first conducts the eigen-decomposition of the adjacency matrix or the Laplacian matrix [26,48,49]. Then it runs a clustering algorithm (typically, k-means) on some leading eigenvectors or their variants to infer the community membership. For example, Ref. [26] showed the consistency of spectral clustering designed based on Laplacian matrix under SBM. Ref. [48] proposed a regularized spectral clustering (RSC) algorithm designed based on regularized Laplacian matrix and shows its theoretical consistency under DCSBM. Ref. [30] studied the consistencies of two spectral clustering algorithms based on the adjacency matrix under SBM and DCSBM. Ref. [50] designed the spectral clustering on the ratios-of-eigenvectors (SCORE) algorithm with a theoretical guarantee under DCSBM. Ref. [49] studied the impact of regularization on a Laplacian spectral clustering under SBM.

Within the MMSB and DCMM frameworks for the overlapping network, broadly speaking, spectral clustering has the following three steps. One first conducts an eigen-decomposition of the adjacency matrix or the graph Laplacian, then hunts corners (also known as vertexes) using a convex hull algorithm, and finally has a membership reconstruction step by projection. The convex hull algorithms suggested in [41] differ in the k-means algorithm a lot. For example, Ref. [44] designed the sequential projection after cleaning (SPACL) algorithm based on the finding that there exists a simplex structure in the eigen-decomposition of the population adjacency matrix and studies the SPACL theoretical properties under MMSB. Meanwhile, SPACL uses the successive projection algorithm proposed in [51] to find the corners for its simplex structure. To fit DCMM, Ref. [41] designs the Mixed-SCORE algorithm based on the finding that there exists a simplex structure in the entry-wise ratio matrix obtained from the eigen-decomposition of the population adjacency matrix under DCMM. Ref. [41] also introduces several choices for convex hull algorithms to find corners for the simplex structure and show the estimation consistency of the Mixed-SCORE under DCMM. Ref. [43] finds the cone structure inherent in the normalization of eigenvectors of the population adjacency matrix under DCMM as well as OCCAM, and develops an algorithm to hunt corners in the cone structure.

1.2. Separation Condition, Alternative Separation Condition and Sharp Threshold

SBM with n nodes belonging to K equal (or nearly equal) size communities and vertices connect with probability p_{in} within clusters and p_{out} across clusters, denoted by $SBM(n, K, p_{in}, p_{out})$, has been well studied in recent years, especially for the case when $K = 2$; see [21] and the references therein. In this paper, we call the network generated from $SBM(n, K, p_{in}, p_{out})$ the standard network for convenience. Without causing confusion, we also call $SBM(n, K, p_{in}, p_{out})$ the standard network, occasionally. Let $p_{in} = \alpha_{in} \frac{\log(n)}{n}$, $p_{out} = \alpha_{out} \frac{\log(n)}{n}$. Ref. [21,52] found that exact recovery in $SBM(n, 2, \alpha_{in} \frac{\log(n)}{n}, \alpha_{out} \frac{\log(n)}{n})$ is solvable, and efficiently so, if $|\sqrt{\alpha_{in}} - \sqrt{\alpha_{out}}| > \sqrt{2}$ (i.e., $|\sqrt{p_{in}} - \sqrt{p_{out}}| > \sqrt{\frac{2\log(n)}{n}}$) and unsolvable if $|\sqrt{\alpha_{in}} - \sqrt{\alpha_{out}}| < \sqrt{2}$ as summarized in Theorem 13 of [53]. This threshold can be achieved by semidefinite relaxations [21,54–56] and spectral methods with local refinements [57,58]. Unlike semidefinite relaxations, spectral methods have

a different threshold, which was particularly pointed out by [21,52]: one highlight for $SBM(n, 2, p_{in}, p_{out})$ is a theorem by [59] which says that when $p_{in} > p_{out}$, if

$$\frac{p_{in} - p_{out}}{\sqrt{p_{in}}} \gg \sqrt{\frac{\log(n)}{n}}$$

then spectral methods can exactly recover node labels with high probability as n goes to infinity (also known as consistent estimation [30,40,41,43,44,48,50]).

Consider a more general case $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$; this paper finds that the above threshold can be extended as

$$\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}} \gg \sqrt{\frac{\log(n)}{n}}, \quad (1)$$

which can be alternatively written as

$$\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}} \gg 1. \quad (2)$$

In this paper, when $K = O(1)$, the lower bound requirement on $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}}$ (and $\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}}$) for the consistent estimation of spectral methods is called the separation condition (alternative separation condition). The network generated from $SBM(n, K, p_{in}, p_{out})$ with $p_{in} > p_{out}$ is an assortative network in which nodes within the community have more edges than across communities [60]. The network generated from $SBM(n, K, p_{in}, p_{out})$ with $p_{in} < p_{out}$ is a dis-assortative network in which nodes within the community have fewer edges than across communities [60]. Therefore, Equation (2) holds for both assortative and dis-assortative networks.

Meanwhile, when $K = 1$ such that $p = p_{in} = p_{out}$, $SBM(n, K, p_{in}, p_{out}) = SBM(n, 1, p, p)$ degenerates to Erdős–Rényi (ER) random graph $G(n, p)$ [53,61,62]. Ref. [61] finds that the ER random graph is connected with high probability if

$$p \geq \frac{\log(n)}{n}. \quad (3)$$

We call the lower bound requirement on p for generating a connected ER random graph the sharp threshold in this paper.

1.3. Inconsistencies on Separation Condition in Some Previous Works

In this paper, we focus on the consistency of spectral method in community detection. The study of consistency is developed by obtaining the theoretical upper bound of error rate for a spectral method through analyzing the properties of the population adjacency matrix under the statistical model. To compare the consistencies of the theoretical results under different models, it is meaningful to study whether the separation condition and sharp threshold obtained from upper bounds of theoretical error rates for different methods under different models are consistent or not. Meanwhile, the separation condition and sharp threshold can also be seen as alternative unified theoretical frameworks to compare all methods and model parameters mentioned in the concluding remarks of [30].

Based on the separation condition and sharp threshold, here we describe some phenomena of the inconsistency in the community detection area. We find that the separation conditions of $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$ obtained from the error rates developed in [41,43,44] under DCMM or MMSB are not consistent with those obtained from the main results of [30] under SBM, and the sharp threshold obtained from the main results of [43,44] do not match the classical results. A summary of these inconsistencies is provided in Tables 1 and 2. Furthermore, after delicate analysis, we find that the requirement on the

network sparsity of [43,44] is stronger than that of [30,41], and [63] also finds that the requirement of Ref. [44] of network sparsity is sub-optimal.

Table 1. Comparison of separation condition and sharp threshold. Details of this table are given in Section 4. The classical result on separation condition given in Corollary 1 of [59] is $\sqrt{\frac{\log(n)}{n}}$ (i.e., Equation (1)). The classical result on sharp threshold is $\frac{\log(n)}{n}$ (i.e., Equation (3)) given in [61], Theorem 4.6 [62] and the first bullet in Section 2.5 [53]. In this paper, n is the number of nodes in a network, A is the adjacency matrix, Ω is the expectation of A under some models, A_{re} is a regularization of A , ρ is the sparsity parameter such that $\rho \geq \max_{i,j} \Omega(i, j)$ and it controls the overall sparsity of a network, $\|\cdot\|$ denotes spectral norm, and $\xi > 1$.

	Model	Separation Condition	Sharp Threshold
Ours using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$	MMSB&DCMM	$\sqrt{\frac{\log(n)}{n}}$	$\frac{\log(n)}{n}$
Ours using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	MMSB&DCMM	$\sqrt{\frac{\log(n)}{n}}$	$\frac{\log(n)}{n}$
Ref. [41] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	DCMM	$\sqrt{\frac{\log(n)}{n}}$	$\frac{\log(n)}{n}$
Ref. [41] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	DCMM	$\sqrt{\frac{\log(n)}{n}}$	$\frac{\log(n)}{n}$
Ref. [43,44] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	MMSB&DCMM	$\frac{\log^\xi(n)}{\sqrt{n}}$	$\frac{\log^{2\xi}(n)}{n}$
Ref. [43,44] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	MMSB&DCMM	$\frac{\log^{\xi+0.5}(n)}{\sqrt{n}}$	$\frac{\log^{2\xi+1}(n)}{n}$
Ref. [30] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	SBM&DCSBM	$\sqrt{\frac{1}{n}}$	$\frac{1}{n}$
Ref. [30] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n) \log(n)}$	SBM&DCSBM	$\sqrt{\frac{\log(n)}{n}}$	$\frac{\log(n)}{n}$

Table 2. Comparison of alternative separation condition, where the classical result on alternative separation condition is 1 (i.e., Equation (2)).

	Model	Alternative Separation Condition
Ours using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$	MMSB&DCMM	1
Ours using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	MMSB&DCMM	1
Ref. [41] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	DCMM	1
Ref. [41] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	DCMM	1
Ref. [43,44] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	MMSB&DCMM	$\log^{\xi-0.5}(n)$
Ref. [43,44] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n)}$	MMSB&DCMM	$\log^\xi(n)$
Ref. [30] using $\ A_{re} - \Omega\ \leq C\sqrt{\rho n}$ (original)	SBM&DCSBM	$\sqrt{\frac{1}{\log(n)}}$
Ref. [30] using $\ A - \Omega\ \leq C\sqrt{\rho n \log(n) \log(n)}$	SBM&DCSBM	1

1.4. Our Findings

Recall that we reviewed several spectral clustering methods under SBM, DCSBM, MMSB and DCMM introduced in [26,30,41,43,44,48–50] and DCSBM, MMSB and DCMM are extensions of SBM (i.e., $SBM(n, K, p_{in}, p_{out})$ is a special case of DCSBM, MMSB and DCMM). We have the following question:

Can these spectral clustering methods achieve the threshold in Equation (1) (or Equation (2)) for $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$ and the threshold in Equation (3) for the Erdős–Rényi (ER) random graph $G(n, p)$?

The answer is yes. In fact, spectral methods for network with mixed memberships still achieve thresholds in Equations (1) and (2) for $MMSB(n, K, \Pi, p_{in}, p_{out})$ defined in Definition 2 when $K = O(1)$, where $MMSB(n, K, \Pi, p_{in}, p_{out})$ can be seen as a generalization of $SBM(n, K, p_{in}, p_{out})$ such that there exist nodes belonging to multiple communities. Explanations for why these spectral clustering methods achieve thresholds in Equations (1)–(3) will be provided in Sections 3–5 via re-establishing theoretical guarantee for SPACL under MMSB and its extension under DCMM because we find that the main theoretical results of [43,44] are sub-optimal. Meanwhile, we can obtain (and cannot obtain) the separation condition and sharp threshold from the theoretical bounds of error rates for

spectral methods analyzed in [30,41,43,44] ([26,30,48–50]) directly. Instead of re-establishing the theoretical guarantee for all spectral methods reviewed in this paper to show that they achieve thresholds in Equations (1) and (3) for $SBM(n, K, P_{in}, P_{out})$ with $K = O(1)$, we mainly focus on the SPACL algorithm under MMSB and its extension under DCMM since MMSB and DCMM are more complex than SBM and DCSBM.

We then summarize the idea of using the separation condition and sharp threshold to study the consistencies, and compare the error rates and requirements on network sparsity of different spectral methods under different models as a four-step criterion, which we call the separation condition and sharp threshold criterion (SCSTC for short). With an application of this criterion, this paper provides an attempt to answer the questions of how the above inconsistency phenomena occur, and how to obtain consistent results with weaker requirements on the network sparsity of [43,44]. To answer the two questions, we use the recent techniques on row-wise eigenvector deviation developed in [64,65] to obtain consistent theoretical results directly related with model parameters for the SPACL and the SVM-cone-DCMMSB algorithm of [43]. The two questions are then answered by delicate analysis with an application of SCSTC to the theoretical upper bounds of error rates in this paper and some previous spectral methods. Using SCSTC for the spectral methods introduced and studied in [26,30,48–50] and some other spectral methods fitting models that can reduce to $SBM(n, K, P_{in}, P_{out})$ with $K = O(1)$, one can prove that these spectral methods achieve thresholds in Equations (1)–(3). The main contributions in this paper are as follows:

- (i) We summarize the idea of using the separation condition of a standard network and sharp threshold of the ER random graph $G(n, p)$ to study the consistent estimations of different spectral methods designed via eigen-decomposition or singular value decomposition of the adjacency matrix or its variants under different models that can degenerate to SBM under mild conditions as a four-step criterion, SCSTC. The separation condition is used to study the consistency of the theoretical upper bound for the spectral method, and the sharp threshold can be used to study the network sparsity. The theoretical results of upper bounds for different spectral methods can be compared by SCSTC. Using this criterion, a few inconsistent phenomenons of some previous works are found.
- (ii) Under MMSB and DCMM, we study the consistencies of the SPACL algorithm proposed in [44] and its extended version using the recent techniques on row-wise eigenvector deviation developed in [64,65]. Compared with the original results of [43,44], our main theoretical results enjoy smaller error rates by lesser dependence on K and $\log(n)$. Meanwhile, our main theoretical results have weaker requirements on the network sparsity and the lower bound of the smallest nonzero singular value of the population adjacency matrix. For details, see Tables 3 and 4.
- (iii) Our results for DCMM are consistent with those for MMSB when DCMM degenerates to MMSB under mild conditions. Using SCSTC, under mild conditions, our main theoretical results under DCMM are consistent with those of [41]. This answers the question that the phenomenon that the main results of [43,44] do not match those of [41] occurs due to the fact that in Ref. [43,44], the theoretical results of error rates are sub-optimal. We also find that our theoretical results (as well as those of [41]) under both MMSB and DCMM match the classical results on the separation condition and sharp threshold, i.e., achieve thresholds in Equations (1)–(3). Using the bound of $\|A - \Omega\|$ instead of $\|A_{re} - \Omega\|$ to establish the upper bound of error rate under SBM in [30], the two spectral methods studied in [30] achieve thresholds in Equations (1)–(3), which answers the question of why the separation condition obtained from error rate of [41] does not match that obtained from the error rate of [30]. Using $\|A_{re} - \Omega\|$ or $\|A - \Omega\|$ influences the row-wise eigenvector deviations in Theorem 3.1 of [44] and Theorem I.3 of [43], and thus using $\|A_{re} - \Omega\|$ or $\|A - \Omega\|$ influences the separation conditions and sharp thresholds of [43,44]. For comparison, our bound on row-wise eigenvector deviation is obtained by using the techniques developed in [64,65] and

that of [41] is obtained by applying the modified Theorem 2.1 of [66]; therefore, using $\|A_{re} - \Omega\|$ or $\|A - \Omega\|$ has no influence on the separation conditions and sharp thresholds of ours and that of [41]. For details, see Tables 1 and 2. In a word, using SCSTC, the spectral methods proposed and studied in [26,30,41,43,44,48–50,67,68] or some other spectral methods fitting models that can reduce to $SBM(n, K, p_{in}, p_{out})$ achieve thresholds in Equations (1)–(3).

- (iv) We verify our threshold in Equation (2) by some computer-generated networks in Section 6. The numerical results for networks generated under $MMSB(n, K, \Pi, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ show that SPACL and its extended version achieve a threshold in Equation (2), and results for networks generated from $SBM(n, K, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ show that the spectral methods considered in [26,30,48,50] achieve the threshold in Equation (2).

Table 3. Comparison of error rates between our Theorem 1 and Theorem 3.2 [44] under $MMSB_n(K, \tilde{P}, \Pi, \rho)$. The dependence on K is obtained when $\kappa(\Pi'\Pi) = O(1)$. For comparison, we have adjusted the l_2 error rates of Theorem 3.2 [44] into l_1 error rates. Note that as analyzed in the first bullet given after Lemma 2, whether using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ or $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ does not change our ω , and has no influence on bound in Theorem 1. For [44], using $\|A_{re} - \Omega\| \leq \sqrt{\rho n}$, the power of $\log(n)$ in their Theorem 3.2 is ξ ; using $\|A - \Omega\| \leq \sqrt{\rho n \log(n)}$, the power of $\log(n)$ in their Theorem 3.2 is $\xi + 0.5$.

	ρn	$\sigma_K(\Omega)$	$\lambda_K(\Pi'\Pi)$	Dependence on K	Dependence on $\log(n)$
Ours	$\geq \log(n)$	$\geq \sqrt{\rho n \log(n)}$	> 0	K^2	$\log^{0.5}(n)$
[44]	$\geq \log^{2\xi}(n)$	$\geq \sqrt{\rho n} \log^\xi(n)$	$\geq 1/\rho$	$K^{2.5}$	$\log^\xi(n)$

Table 4. Comparison of error rates between our Theorem 2 and Theorem 3.2 [43] under $DCMM_n(K, P, \Pi, \Theta)$. The dependence on K is obtained when $\kappa(\Pi'\Pi) = O(1)$. For comparison, we adjusted the l_2 error rates of Theorem 3.2 [43] into l_1 error rates. Since Theorem 2 enjoys the same separation condition and sharp threshold as Theorem 1, and Theorem 3.2 [43] enjoys the same separation condition and sharp threshold as Theorem 3.2 [44], we do not report them in this table. Note that as analyzed in Remark 11, whether using $\|A - \Omega\| \leq C\sqrt{\theta_{\max}\|\theta\|_1 \log(n)}$ or $\|A_{re} - \Omega\| \leq C\sqrt{\theta_{\max}\|\theta\|_1}$ does not change our ω under DCMM, and has no influence on the results in Theorem 2. For [43], using $\|A_{re} - \Omega\| \sqrt{\theta_{\max}\|\theta\|_1}$, the power of $\log(n)$ in their Theorem 3.2 is ξ ; using $\|A - \Omega\| \sqrt{\theta_{\max}\|\theta\|_1 \log(n)}$, the power of $\log(n)$ in their Theorem 3.2 is $\xi + 0.5$.

	$\Pi(i, :)$	$\theta_{\max}\ \theta\ _1$	$\sigma_K(\Omega)$	$\kappa(\Pi'\Theta^2\Pi)$	Dependence on K	Dependence on $\log(n)$
Ours	arbitrary	$\geq \log(n)$	$\geq \theta_{\max}\sqrt{n \log(n)}$	≥ 1	K^6	$\log^{0.5}(n)$
[43]	<i>iid</i> from Dirichlet	$\geq \log^{2\xi}(n)$	$\geq \theta_{\max}\sqrt{n} \log^\xi(n)$	$= O(1)$	$K^{6.5}$	$\log^\xi(n)$

The article is organized as follows. In Section 2, we give the formal introduction to the mixed membership stochastic blockmodel and review the algorithm SPACL considered in this paper. The theoretical results of consistency for the mixed membership stochastic blockmodel are presented and compared to related works in Section 3. After delicate analysis, the separation condition and sharp threshold criterion is presented in Section 4. Based on an application of this criterion, the improvement consistent estimation results for the extended version of SPACL under the degree corrected mixed membership model are provided in Section 5. Several computer-generated networks under MMSB and SBM are conducted to show that some spectral clustering methods achieve the threshold in Equation (2) in Section 6. The conclusion is given in Section 7.

Notations. We take the following general notations in this paper. Write $[m] := \{1, 2, \dots, m\}$ for any positive integer m . For a vector x and fixed $q > 0$, $\|x\|_q$ denotes its l_q -norm. We drop the subscript if $q = 2$ occasionally. For a matrix M , M' denotes

the transpose of the matrix M , $\|M\|$ denotes the spectral norm, $\|M\|_F$ denotes the Frobenius norm, $\|M\|_{2 \rightarrow \infty}$ denotes the maximum l_2 -norm of all the rows of M , and $\|M\|_\infty := \max_i \sum_j |M(i, j)|$ denotes the maximum absolute row sum of M . Let $\text{rank}(M)$ denote the rank of matrix M . Let $\sigma_i(M)$ be the i -th largest singular value of matrix M , $\lambda_i(M)$ denote the i -th largest eigenvalue of the matrix M ordered by the magnitude, and $\kappa(M)$ denote the condition number of M . $M(i, :)$ and $M(:, j)$ denote the i -th row and the j -th column of matrix M , respectively. $M(S_r, :)$ and $M(:, S_c)$ denote the rows and columns in the index sets S_r and S_c of matrix M , respectively. For any matrix M , we simply use $Y = \max(0, M)$ to represent $Y_{ij} = \max(0, M_{ij})$ for any i, j . For any matrix $M \in \mathbb{R}^{m \times m}$, let $\text{diag}(M)$ be the $m \times m$ diagonal matrix whose i -th diagonal entry is $M(i, i)$. $\mathbf{1}$ and $\mathbf{0}$ are column vectors with all entries being ones and zeros, respectively. e_i is a column vector whose i -th entry is 1, while other entries are zero. In this paper, C is a positive constant which may vary occasionally. $f(n) = O(g(n))$ means that there exists a constant $c > 0$ such that $|f(n)| \leq c|g(n)|$ holds for all sufficiently large n . $x \succeq y$ means there exists a constant $c > 0$ such that $|x| \geq c|y|$. $f(n) = o(g(n))$ indicates that $\frac{f(n)}{g(n)} \rightarrow 0$ as $n \rightarrow \infty$.

2. Mixed Membership Stochastic Blockmodel

Let $A \in \{0, 1\}^{n \times n}$ be a symmetric adjacency matrix such that $A(i, j) = 1$ if there is an edge between node i to node j , and $A(i, j) = 0$ otherwise. The mixed membership stochastic blockmodel (MMSB) [36] for generating A is as follows.

$$\Omega := \rho \Pi \tilde{P} \Pi' \quad A(i, j) \sim \text{Bernoulli}(\Omega(i, j)) \quad i, j \in [n], \tag{4}$$

where $\Pi \in \mathbb{R}^{n \times K}$ is called the membership matrix with $\Pi(i, k) \geq 0$ and $\sum_{k=1}^K \Pi(i, k) = 1$ for $i \in [n]$ and $k \in [K]$, $\tilde{P} \in \mathbb{R}^{K \times K}$ is a non-negative symmetric matrix with $\max_{k, l \in [K]} \tilde{P}(k, l) = 1$ for model identifiability under MMSB, ρ is called the sparsity parameter which controls the sparsity of the network, and $\Omega \in \mathbb{R}^{n \times n}$ is called the population adjacency matrix since $\mathbb{E}[A] = \Omega$. As mentioned in [41,44], $\sigma_K(\tilde{P})$ is a measure of the separation between communities, and we call it the separation parameter in this paper. ρ and $\sigma_K(\tilde{P})$ are two important model parameters directly related with the separation condition and sharp threshold, and they will be considered throughout this paper.

Definition 1. Call model (4) the mixed membership stochastic blockmodel (MMSB), and denote it by $MMSB_n(K, \tilde{P}, \Pi, \rho)$.

Definition 2. Let $MMSB(n, K, \Pi, p_{in}, p_{out})$ be a special case of $MMSB_n(K, \tilde{P}, \Pi, \rho)$ when $\rho \tilde{P}$ has diagonal entries p_{in} and non-diagonal entries p_{out} , and $\kappa(\Pi' \Pi) = O(1)$.

Call node i ‘pure’ if $\Pi(i, :)$ is degenerate (i.e., one entry is 1, all others $K - 1$ entries are 0) and ‘mixed’ otherwise. When all nodes are pure in Π , we see that $MMSB(n, K, \Pi, p_{in}, p_{out})$ exactly reduces to $SBM(n, K, p_{in}, p_{out})$. Thus, $MMSB(n, K, \Pi, p_{in}, p_{out})$ is a generalization of $SBM(n, K, p_{in}, p_{out})$ with mixed nodes in each community. In this paper, we show that SPACL [44] fitting MMSB and SVM-cone-DCMMSB [43] and Mixed-SCORE [41] fitting DCMM also achieve thresholds in Equations (1)–(3) for $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$. By Theorems 2.1 and 2.2 [44], the following conditions are sufficient for the identifiability of MMSB, when $\rho \tilde{P}(k, l) \in [0, 1]$ for all $k, l \in [K]$,

- (I1) $\text{rank}(\tilde{P}) = K$.
- (I2) There is at least one pure node for each of the K communities.

Unless specified, we treat conditions (I1) and (I2) as the default from now on.

For $k \in [K]$, let $\mathcal{I}^{(k)}$ be the set of pure nodes in community k such that $\mathcal{I}^{(k)} = \{i \in [n] : \Pi(i, k) = 1\}$. For $k \in [K]$, select one node from $\mathcal{I}^{(k)}$ to construct the index set \mathcal{I} , i.e., \mathcal{I} is the index of nodes corresponding to K pure nodes, one from each community. Without loss of generality, let $\Pi(\mathcal{I}, :) = I_K$ where I_K is the $K \times K$ identity matrix. Recall that $\text{rank}(\Omega) = K$. Let $\Omega = U \Lambda U'$ be the compact eigen-decomposition of Ω such that $U \in \mathbb{R}^{n \times K}$, $\Lambda \in \mathbb{R}^{K \times K}$,

and $U'U = I_K$. Lemma 2.1 [44] gives that $U = \Pi U(\mathcal{I}, :)$, and such a form is called ideal simplex (IS for short) [41,44] since all rows of U form a K -simplex in \mathbb{R}^K and the K rows of $U(\mathcal{I}, :)$ are the vertices of the K -simplex. Given Ω and K , as long as we know $U(\mathcal{I}, :)$, we can exactly recover Π by $\Pi = UU^{-1}(\mathcal{I}, :)$ since $U(\mathcal{I}, :) \in \mathbb{R}^{K \times K}$ is a full rank matrix. As mentioned in [41,44], for such IS, the successive projection (SP) algorithm [51] (i.e., Algorithm A1) can be applied to U with K communities to exactly find the corner matrix $U(\mathcal{I}, :)$. For convenience, set $Z = UU^{-1}(\mathcal{I}, :)$. Since $\Pi = Z$, we have $\Pi(i, :) = \frac{Z(i,:)}{\|Z(i,:)\|_1}$ for $i \in [n]$.

Based on the above analysis, we are now ready to give the ideal SPACL algorithm with input Ω, K and output Π .

- Let $\Omega = U\Lambda U'$ be the top- K eigen-decomposition of Ω such that $U \in \mathbb{R}^{n \times K}, \Lambda \in \mathbb{R}^{K \times K}, U'U = I$.
- Run SP algorithm on the rows of U assuming that there are K communities to obtain \mathcal{I} .
- Set $Z = UU^{-1}(\mathcal{I}, :)$.
- Recover Π by setting $\Pi(i, :) = \frac{Z(i,:)}{\|Z(i,:)\|_1}$ for $i \in [n]$.

With the given U and K , since the SP algorithm returns $U(\mathcal{I}, :)$, we see that the ideal SPACL exactly (for detail, see Appendix B) returns Π .

Now, we review the SPACL algorithm of [44]. Set $\tilde{A} = \hat{U}\hat{\Lambda}\hat{U}'$ to be the top K eigen-decomposition of A such that $\hat{U} \in \mathbb{R}^{n \times K}, \hat{\Lambda} \in \mathbb{R}^{K \times K}, \hat{U}'\hat{U} = I_K$, and $\hat{\Lambda}$ contains the top K eigenvalues of A . For the real case, use $\hat{Z}, \hat{\Pi}$ given in Algorithm 1 to estimate Z, Π , respectively. Algorithm 1 is the SPACL algorithm [44] where we only care about the estimation of the membership matrix Π , and omit the estimation of P and ρ . Meanwhile, Algorithm 1 is a direct extension of the ideal SPACL algorithm from the oracle case to the real case, and we omit the prune step in the original SPACL algorithm of [44].

Algorithm 1 SPACL [44].

Require: The adjacency matrix $A \in \mathbb{R}^{n \times n}$ and the number of communities K .

Ensure: The estimated $n \times K$ membership matrix $\hat{\Pi}$.

- 1: Obtain $\tilde{A} = \hat{U}\hat{\Lambda}\hat{U}'$, the top K eigen-decomposition of A .
 - 2: Apply SP algorithm (i.e., Algorithm A1) on the rows of \tilde{U} assuming there are K communities to obtain $\hat{\mathcal{I}}$, the index set returned by SP algorithm.
 - 3: Set $\hat{Z} = \hat{U}\hat{U}^{-1}(\hat{\mathcal{I}}, :)$. Then set $\hat{Z} = \max(0, \hat{Z})$.
 - 4: Estimate $\Pi(i, :)$ by $\hat{\Pi}(i, :) = \hat{Z}(i, :)/\|\hat{Z}(i, :)\|_1, i \in [n]$.
-

3. Consistency under MMSB

Our main result under MMSB provides an upper bound on the estimation error of each node’s membership in terms of several model parameters. Throughout this paper, K is a known positive integer. Assume that

(A1) $\rho n \geq \log(n)$.

Assumption (A1) provides a requirement on the lower bound of sparsity parameter ρ such that it should be at least $\log(n)/n$. Then we have the following lemma.

Lemma 1. Under $MMSB_n(K, \tilde{P}, \Pi, \rho)$, when Assumption (A1) holds, with probability at least $1 - o(n^{-\alpha})$ for any $\alpha > 0$, we have

$$\|A - \Omega\| \leq \frac{\alpha + 1 + \sqrt{(\alpha + 1)(\alpha + 19)}}{3} \sqrt{\rho n \log(n)}.$$

In Lemma 1, instead of simply using a constant C_α to denote $\frac{\alpha + 1 + \sqrt{(\alpha + 1)(\alpha + 19)}}{3}$, we keep the explicit form here.

Remark 1. When Assumption (A1) holds, the upper bound of $\|A - \Omega\|$ in Lemma 1 is consistent with Corollary 6.5 in [69] since $\text{Var}(A(i, j)) \leq \rho$ under $\text{MMSB}_n(K, P, \Pi, \rho)$.

Lemma 1 is obtained via Theorem 1.4 (Bernstein inequality) in [70]. For comparison, Ref. [44] applies Theorem 5.2 [30] to bound $\|A - \Omega\|$ (see, for example, Equation (14) of [44]) and obtains a bound as $C\sqrt{\rho n}$ for some $C > 0$. However, $C\sqrt{\rho n}$ is the bound between a regularization of A and Ω as stated in the proof of Theorem 5.2 [30], where such regularization of A is obtained from A with some constraints in Lemmas 4.1 and 4.2 of the supplemental material [30]. Meanwhile, Theorem 2 [71] also gives that the bound between a regularization of A and Ω is $C\sqrt{\rho n}$, where such a regularization of A should also satisfy few constraints on A ; see Theorem 2 [71] for detail. Instead of bounding the difference between a regularization of A and Ω , we are interested in bounding $\|A - \Omega\|$ by the Bernstein inequality, which has no constraints on A . For convenience, use A_{re} to denote the regularization of A in this paper. Hence, $\|A_{\text{re}} - \Omega\| \leq C\sqrt{\rho n}$ with high probability, and this bound is model independent as shown by Theorem 5.2 [30] and Theorem 2 [71] as long as $\rho \geq \max_{i,j} \Omega(i, j)$ (here, let $\Omega = \mathbb{E}[A]$ without considering models, a ρ satisfying $\rho \geq \max_{i,j} \Omega(i, j)$ is also the sparsity parameter which controls the overall sparsity of a network). Note that A_{re} is not \tilde{A} , where $\tilde{A} = \hat{U}\Lambda\hat{U}'$ is obtained by the top K eigen-decomposition of A , while A_{re} is obtained by adding constraints on degrees of A ; see Theorem 2 [71] for detail.

In [41,43,44], the main theoretical results for their proposed membership estimating methods hinge on a row-wise deviation bound for the eigenvectors of the adjacency matrix, whether under MMSB or DCMM. Different from the theoretical technique applied in Theorem 3.1 [44], which provides sup-optimal dependencies on $\log(n)$ and K , and needs sub-optimal requirements on the sparsity parameter ρ and the lower bound of $\sigma_K(\Omega)$, to obtain row-wise deviation bound for the singular eigenvector of Ω , we use Theorem 4.2 [64] and Theorem 4.2 [65].

Lemma 2. (Row-wise eigenspace error) Under $\text{MMSB}_n(K, \tilde{P}, \Pi, \rho)$, when Assumption (A1) holds, suppose $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$, with probability at least $1 - o(n^{-\alpha})$,

- When we apply Theorem 4.2 of [64], we have

$$\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O\left(\frac{\sqrt{K}(\kappa(\Omega)\sqrt{\frac{n}{K\lambda_K(\Pi\Pi)}} + \sqrt{\log(n)})}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K(\Pi\Pi)}\right),$$

- When we apply Theorem 4.2 of [65], we have

$$\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O\left(\frac{\sqrt{n \log(n)}}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K^{1.5}(\Pi\Pi)}\right).$$

For convenience, set $\omega = \|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty}$, and let ω_1, ω_2 denote the upper bound in Lemma 2 when applying Theorem 4.2 of [64] and Theorem 4.2 of [65], respectively. Note that when $\lambda_K(\Pi\Pi) = O(\frac{n}{K})$, we have $\omega_1 = \omega_2 = O(\frac{K^{1.5}}{\sigma_K(\tilde{P})} \frac{1}{\sqrt{n}} \sqrt{\frac{\log(n)}{\rho n}})$, and therefore we simply let ω_2 be the bound since its form is slightly simpler than ω_1 .

Compared with Theorem 3.1 of [44], since we apply Theorem 4.2 of [64] and Theorem 4.2 of [65] to obtain the bound of row-wise eigenspace error under MMSB, our bounds do not rely on $\min(K^2, \kappa^2(\Omega))$ while Theorem 3.1 [44] does. Meanwhile, our bound in Lemma 2 is sharper with lesser dependence on K and $\log(n)$, has weaker requirements on the lower bounds of $\sigma_K(\Omega)$, $\lambda_K(\Pi\Pi)$ and the sparsity parameter ρ . The details are given below:

- We emphasize that the bound of Theorem 3.1 of [44] should be $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O(\frac{\psi(\Omega)\sqrt{Kn \log^\xi(n)}}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K^{1.5}(\Pi\Pi)})$ instead of $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O(\frac{\psi(\Omega)\sqrt{Kn}}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K^{1.5}(\Pi\Pi)})$ for $\xi > 1$ where the function ψ is defined in Equation (7) of [44], and this is also pointed out by

Table 2 of [63]. The reason is that in the proof part of Theorem 3.1 [44], from step (iii) to step (iv), they should keep the term $\log^\xi(n)$ since this term is much larger than 1. We can also find that the bound in Theorem 3.1 [44] should multiply $\log^\xi(n)$ from Theorem VI.1 [44] directly. For comparison, this bound $O(\frac{\psi(\Omega)\sqrt{Kn}\log^\xi(n)}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K^{1.5}(\Pi'\Pi)})$ is $K^{0.5}\log^{\xi-0.5}(n)$ times our bound in Lemma 2. Meanwhile, by the proof of the bound in Theorem 3.1 of [44], we see that the bound depends on the upper bound of $\|A - \Omega\|$, and [44] applies Theorem 5.2 of [30] such that $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ with high probability. Since $C\sqrt{\rho n}$ is the upper bound of the difference between a regularization of A and Ω . Therefore, if we are only interested in bounding $\|A - \Omega\|$ instead of $\|A_{re} - \Omega\|$, the upper bound of Theorem 3.1 [44] should be $O(\frac{\psi(\Omega)\sqrt{Kn}\log^{\xi+0.5}(n)}{\sigma_K(\tilde{P})\lambda_K^{1.5}(\Pi'\Pi)})$, which is at least $K^{0.5}\log^\xi(n)$ times our bound in Lemma 2. Furthermore, the upper bound of the row-wise eigenspace error in Lemma 2 does not rely on the upper bound of $\|A - \Omega\|$ as long as $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$ holds. Therefore, whether using $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ or $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ does not change the bound in Lemma 2.

- Our Lemma 2 requires $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$, while Theorem 3.1 [44] requires $\sigma_K(\Omega) \geq 4\sqrt{\rho n} \log^\xi(n)$ by their Assumption 3.1. Therefore, our Lemma 2 has a weaker requirement on the lower bound of $\sigma_K(\Omega)$ than that of Theorem 3.1 [44]. Meanwhile, Theorem 3.1 [44] requires $\lambda_K(\Pi'\Pi) \geq \frac{1}{\rho}$ while our Lemma 2 has no lower bound requirement on $\lambda_K(\Pi'\Pi)$ as long as it is positive.
- Since $\|\Omega\| = \|\rho\Pi\tilde{P}\Pi'\| \leq C\rho n$ by basic algebra, the lower bound requirement on $\sigma_K(\Omega)$ in Assumption 3.1 of [44] gives that $4\sqrt{\rho n} \log^\xi(n) \leq \sigma_K(\Omega) \leq \|\Omega\| \leq C\rho n$, which suggests that Theorem 3.1 [44] requires $\rho n \geq C \log^{2\xi}(n)$, and this also matches with the requirement on ρn in Theorem VI.1 of [44] (and this is also pointed out by Table 1 of [63]). For comparison, our requirement on sparsity given in Assumption (A1) is $\rho n \geq \log(n)$, which is weaker than $\rho n \geq C \log^{2\xi}(n)$. Similarly, in our Lemma 2, the requirement $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$ gives $C\sqrt{\rho n \log(n)} \leq \sigma_K(\Omega) \leq \|\Omega\| \leq C\rho n$, thus we have $\log(n) \leq C\rho n$ which is consistent with Assumption (A1).

If we further assume that $K = O(1)$, $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$ (i.e., $\kappa(\Pi'\Pi) = O(1)$) and $\sigma_K(\tilde{P}) = O(1)$, the row-wise eigenspace error is of order $\frac{1}{\sqrt{n}}\sqrt{\frac{\log(n)}{\rho n}}$, which is consistent with the row-wise eigenvector deviation of the result of [63], shown in their Table 2. The next theorem gives the theoretical bounds on the estimations of memberships under MMSB.

Theorem 1. Under $MMSB_n(K, \tilde{P}, \Pi, \rho)$, let $\hat{\Pi}$ be obtained from Algorithm 1, and suppose the conditions in Lemma 2 hold; there exists a permutation matrix $\mathcal{P} \in \mathbb{R}^{K \times K}$ such that, with probability at least $1 - o(n^{-\alpha})$, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi} - \Pi\mathcal{P})\|_1 = O(\omega K \kappa(\Pi'\Pi) \sqrt{\lambda_1(\Pi'\Pi)}).$$

Remark 2. (Comparison to Theorem 3.2 [44]) Consider a special case by setting $\kappa(\Pi'\Pi) = O(1)$, i.e., $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$ and $\lambda_1(\Pi'\Pi) = O(\frac{n}{K})$. We focus on comparing the dependencies on K in bounds of our Theorem 1 and Theorem 3.2 [44]. Under this case, the bound of our Theorem 1 is proportional to K^2 by basic algebra; since $\min(K^2, \kappa^2(\Omega)) = \min(K^2, O(1)) = O(1)$ and the bound in Theorem 3.2 [44] should multiply \sqrt{K} because (in [44]'s language) $\|\hat{V}_p^{-1}\|_F \leq \frac{\sqrt{K}}{\sigma_K(\hat{V}_p)}$ instead of $\|\hat{V}_p^{-1}\|_F = \frac{1}{\lambda_K(\hat{V}_p)}$ in Equation (45) [44], the power of K is 2 by checking the bound of Theorem 3.2 [44]. Meanwhile, note that our bound in Theorem 2 is l_1 bound, while the bound in Theorem 3.2 [44] is l_2 bound. When we translate the l_2 bound of Theorem 3.2 [44] into l_1 bound, the power of K is 2.5 for Theorem 3.2 [44]. Hence, our bound in Theorem 1 has less dependence on K than that of Theorem 3.2 [44], and this is also consistent with the first bullet given after Lemma 2.

Table 3 summarizes the necessary conditions and dependence on the model parameters of the rates in Theorem 1 and Theorem 3.2 [44] for comparison. The following corollary is obtained by adding conditions on the model parameters similar to Corollary 3.1 in [44].

Corollary 1. Under $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$, when the conditions of Lemma 2 hold, with probability at least $1 - o(n^{-\alpha})$, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi} - \Pi\mathcal{P})\|_1 = O\left(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}}\right).$$

Remark 3. Consider a special case in Corollary 1 by setting $\sigma_K(\tilde{P})$ as a constant, we see that the error bound $O(\sqrt{\frac{\log(n)}{\rho n}})$ in Corollary 1 is directly related to Assumption (A1), and for consistent estimation, ρ should shrink slower than $\frac{\log(n)}{n}$.

Remark 4. Under the setting of Corollary 1, the requirement $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$ in Lemma 2 holds naturally. By Lemma II.4 [44], we know that $\sigma_K(\Omega) \geq \rho\sigma_K(\tilde{P})\lambda_K(\Pi'\Pi) = C\rho n\sigma_K(\tilde{P})$. To make the requirement $\sigma_K(\Omega) \geq C\sqrt{\rho n \log(n)}$ always hold, we just need $C\rho n\sigma_K(\tilde{P}) \geq C\sqrt{\rho n \log(n)}$, which gives that $\sigma_K(\tilde{P}) \geq C\sqrt{\frac{\log(n)}{\rho n}}$, and it just matches with the requirement of the consistent estimation of memberships in Corollary 1.

Remark 5. (Comparison to Theorem 3.2 [44]) When $K = O(1)$ and $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$, by the first bullet in the analysis given after Lemma 2, the row-wise eigenspace error of Theorem 3.1 [44] is $O(\frac{\log^{\frac{5}{2}}(n)}{\sigma_K(\tilde{P})\sqrt{\rho n}})$, and it gives that their error bound on estimation membership given in their Equation (3) is $O(\frac{\log^{\frac{5}{2}}(n)}{\sigma_K(\tilde{P})\sqrt{\rho n}})$, which is $\log^{\frac{5}{2}-0.5}(n)$ times of the bound in our Lemma 1.

Remark 6. (Comparison to Theorem 2.2 [41]) Replacing the Θ in [41] by $\Theta = \sqrt{\rho}I$, their DCMM model degenerates to MMSB. Then their conditions in Theorem 2.2 are our Assumption (A1) and $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$ for MMSB. When $K = O(1)$, the error bound in Theorem 2.2 in [41] is $O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}})$, which is consistent with ours.

4. Separation Condition and Sharp Threshold Criterion

After obtaining Corollary 1 under MMSB, now we are ready to give our criterion after introducing the separation condition of $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$ and the sharp threshold of ER random graph $G(n, p)$ in this section.

Separation condition. Let $P = \rho\tilde{P}$ be the probability matrix for $MMSB(n, K, \Pi, p_{in}, p_{out})$ when $K = O(1)$, so P has diagonal (and non-diagonal) entries p_{in} (and p_{out}) and $\sigma_K(P) = \rho\sigma_K(\tilde{P}) \equiv |p_{in} - p_{out}|$. Recall that $\max_{k,l \in [K]} \tilde{P}(k, l) = 1$ under $MMSB_n(K, \tilde{P}, \Pi, \rho)$, we have $\max_{k,l \in [K]} P(k, l) = \rho \equiv \max(p_{in}, p_{out})$. So, we have the separation condition $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}} \equiv \sqrt{\rho}\sigma_K(\tilde{P})$ (also known as the relative edge probability gap in [44]) and the alternative separation condition $\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}} \equiv \sqrt{\frac{\rho n}{\log(n)}}\sigma_K(\tilde{P})$. Now, we are ready to compare the thresholds of the (alternative) separation condition obtained from different theoretical results.

- (a) By Corollary 1, we know that $\sigma_K(\tilde{P})$ should shrink slower than $\sqrt{\frac{\log(n)}{\rho n}}$ for consistent estimation. Therefore, the separation condition $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}} \equiv \sqrt{\rho}\sigma_K(\tilde{P})$ should shrink slower than $\sqrt{\frac{\log(n)}{n}}$ (i.e., Equation (1)), and this threshold is consistent with Corollary 1 of [59] and Equation (17) of [49]. The alternative separation condition $\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}} \equiv \sqrt{\frac{\rho n}{\log(n)}}\sigma_K(\tilde{P})$ should shrink slower than 1 (i.e., Equation (2)).

- (b) Undoubtedly, the (alternative) separation condition in (a) is consistent with that of [41], since Theorem 2.2 [41] shares the same error rate $O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}})$ for $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$.
- (c) By Remark 5, using $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$, we know that in Ref. [44], Equation (3) is $O(\frac{\log^{\xi}(n)}{\sigma_K(\tilde{P})\sqrt{\rho n}})$, so $\sqrt{\rho}\sigma_K(\tilde{P})$ should shrink slower than $\frac{\log^{\xi}(n)}{\sqrt{n}}$. Thus, for [44], the separation condition is $\frac{\log^{\xi}(n)}{\sqrt{n}}$, and the alternative separation condition is $\log^{\xi-0.5}(n)$, which are sub-optimal compared with ours in (a). Using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$, and Equation (3) in Ref. [44], which is $O(\frac{\log^{\xi+0.5}(n)}{\sigma_K(\tilde{P})\sqrt{\rho n}})$, we see that for [44], now the separation condition is $\frac{\log^{\xi+0.5}(n)}{\sqrt{n}}$ and the alternative separation condition is $\log^{\xi}(n)$.
- (d) For comparison, the error bound of Corollary 3.2 [30] built under SBM for community detection is $O(\frac{1}{\sigma_K^2(\tilde{P})\rho n})$ for $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$, so $\sqrt{\rho}\sigma_K(\tilde{P})$ should shrink slower than $\frac{1}{\sqrt{n}}$. Thus the separation condition for [30] is $\frac{1}{\sqrt{n}}$. However, as we analyzed in the first bullet given after Lemma 2, [30] applied $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ to build their consistency results. Instead, we apply $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ to the built theoretical results of [30], and the error bound of Corollary 3.2 [30] is $O(\frac{\log(n)}{\sigma_K^2(\tilde{P})\rho n})$, which returns the same separation condition as our Corollary 1 and Theorem 2.2 of [41] now. Following a similar analysis to (a)–(c), we can obtain an alternative separation condition for [30] immediately, and the results are provided in Table 2. Meanwhile, as analyzed in the first bullet given after Lemma 2, whether using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ or $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ does not change our error rates. By carefully analyzing the proof of 2.1 of [41], we see that whether using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ or $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$ also does not change their row-wise large deviation, hence it does not influence their upper bound of the error rate for their Mixed-SCORE.

Sharp threshold. Consider the Erdős–Rényi (ER) random graph $G(n, p)$ [61]. To construct the ER random graph $G(n, p)$, let $K = 1$ and Π be an $n \times 1$ vector with all entries being ones. Since $K = 1$ and the maximum entry of \tilde{P} is assumed to be 1, we have $\tilde{P} = 1$ in $G(n, p)$ and hence $\sigma_K(\tilde{P}) = 1$. Then we have $\Omega = \Pi\rho\tilde{P}\Pi' = \Pi\rho\Pi' = \Pi p\Pi'$, i.e., $p = \rho$. Since the error rate is $O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}}) = O(\sqrt{\frac{\log(n)}{pn}})$, for consistent estimation, we see that p should shrink slower than $\frac{\log(n)}{n}$ (i.e., Equation (3)), which is just the sharp threshold in [61], Theorem 4.6 [62], strongly consistent with [72], and the first bullet in Section 2.5 [53] (called the lower bound requirement of p for the ER random graph to enjoy consistent estimation as the sharp threshold). Since the sharp threshold is obtained when $K = 1$, which means a connected ER random graph $G(n, p)$, this is also consistent with the connectivity in Table 2 of [21]. Meanwhile, since our Assumption (A1) requires $\rho n \geq \log(n)$, it gives that p should shrink slower than $\frac{\log(n)}{n}$ since $p = \rho$ under $G(n, p)$, which is consistent with the sharp threshold. Since Theorem 2.2 of Ref. [41] enjoys the same error rate as ours under the settings in Corollary 1, [41] also reaches the sharp threshold as $\frac{\log(n)}{n}$. Furthermore, Remark 5 says that the bound for the error rate in Equation (3) [44] should be $O(\frac{\log^{\xi}(n)}{\sigma_K(\tilde{P})\sqrt{\rho n}})$ when using $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$; following a similar analysis, we see that the sharp threshold for [44] is $\frac{\log^{2\xi}(n)}{n}$, which is sub-optimal compared with ours. When using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$, the sharp threshold for [44] is $\frac{\log^{2\xi+1}(n)}{n}$. Similarly, the error bound of Corollary 3.2 [30] is $O(\frac{1}{\sigma_K^2(\tilde{P})\rho n}) \equiv O(\frac{1}{pn})$ under ER $G(n, p)$ since $p = \rho, \sigma_K(\tilde{P}) = 1$ and $K = 1$. Hence, the sharp threshold obtained from the theoretical upper bound for error rates of [30] is $\frac{1}{n}$, which does not match the classical result. Instead, we apply $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ with a high probability to build the theoretical results of [30], and the error bound of Corollary 3.2 [30] is $O(\frac{\log(n)}{pn})$, which returns the classical sharp threshold $\frac{\log(n)}{n}$ now.

Table 1 summarizes the comparisons of the separation condition and sharp threshold. Table 2 records the respective alternative separation condition. The delicate analysis given above supports our statement that the separation condition of a standard network (i.e., $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$ or $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$) and the sharp threshold of ER random graph $G(n, p)$ can be seen as unified criteria to compare the theoretical results of spectral methods under different models. To conclude the above analysis, here, we summarize the main steps to apply the separation condition and sharp threshold criterion (SCSTC for short) to check the consistency of the theoretical results or compare the results of spectral methods under different models, where spectral methods mean methods developed based on the application of the eigenvectors or singular vectors of the adjacency matrix or its variants for community detection. The four-stage SCSTC is given below:

- step*₁ Check whether the theoretical upper bound of the error rate contains $\sigma_K(\tilde{P})$ (note that $P = \rho\tilde{P}$ is probability matrix and maximum entries of \tilde{P} should be set as 1), where the separation parameter $\sigma_K(\tilde{P})$ always appears when considering the lower bound of $\sigma_K(\Omega)$. If it contains $\sigma_K(\tilde{P})$, move to the next step. Otherwise, it suggests possible improvements for the consistency by considering $\sigma_K(\tilde{P})$ in the proofs.
- step*₂ Let $K = O(1)$ and network degenerate to the standard network whose numbers of nodes in each community are in the same order and can be seen as $O(\frac{n}{K})$ (i.e., a $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$ in the case of a non-overlapping network or a $MMSB(n, K, \Pi, p_{in}, p_{out})$ with $K = O(1)$ in the case of an overlapping network, and we will mainly focus on $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$ for convenience.). Let the model degenerate to $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$, and then we obtain the new theoretical upper bound of the error rate. Note that if the model does consider degree heterogeneity, sparsity parameter ρ should be considered in the theoretical upper bound of error rate in *step*₁. If the model considers degree heterogeneity, when it degenerates to $SBM(n, K, p_{in}, p_{out})$ with $K = O(1)$, ρ appears at this step. Meanwhile, if ρ is not contained in the error rate of *step*₁ when the model does not consider degree heterogeneity, it suggests possible improvements by considering ρ .
- step*₃ Let $P = \rho\tilde{P}$ be the probability matrix when the model degenerates to $SBM(n, K, p_{in}, p_{out})$ such that P has diagonal entries p_{in} and non-diagonal entries p_{out} . So, $\sigma_K(P) = |p_{in} - p_{out}| = \rho\sigma_K(\tilde{P})$ and separation condition $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}} \equiv \sqrt{\rho}\sigma_K(\tilde{P})$ since the maximum entry of \tilde{P} is assumed to be 1. Compute the lower bound requirement of $\sigma_K(\tilde{P})$ for consistency estimation through analyzing the new bound obtained in *step*₂. Compute separation condition $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}} \equiv \sqrt{\rho}\sigma_K(\tilde{P})$ using the lower bound requirement for $\sigma_K(\tilde{P})$. The sharp threshold for the ER random graph $G(n, p)$ is obtained from the lower bound requirement on ρ for the consistency estimation under the setting that $K = 1, \sigma_K(\tilde{P}) = 1$ and $p = \rho$.
- step*₄ Compare the separation condition and the sharp threshold obtained in *step*₃ with Equations (1) and (3), respectively. If the sharp threshold $\gg \frac{\log(n)}{n}$ or the separation condition $\gg \sqrt{\frac{\log(n)}{n}}$, then this leaves improvements on the requirement of the network sparsity or theoretical upper bound of the error rate. If the sharp threshold is $\frac{\log(n)}{n}$ and the separation condition is $\sqrt{\frac{\log(n)}{n}}$, the optimality of the theoretical results on both error rates and the requirement of network sparsity is guaranteed. Finally, if the sharp threshold $\ll \frac{\log(n)}{n}$ or separation condition $\ll \sqrt{\frac{\log(n)}{n}}$, this suggests that the theoretical result is obtained based on $\|A_{re} - \Omega\|$ instead of $\|A - \Omega\|$.

Remark 7. This remark provides some explanations on the four steps of SCSTC.

- In *step*₁, we give a few examples. When applying SCSTC to the main results of [40,48,67], we stop at *step*₁ as analyzed in Remark 8, suggesting possible improvements by considering $\sigma_K(\tilde{P})$ for these works. Meanwhile, for the theoretical result without considering $\sigma_K(\tilde{P})$, we

can also move to step₂ to obtain the new theoretical upper bound of the error rate, which is related with ρ and n . Discussions on the theoretical upper bounds of error rates of [50,68] given in Remark 8 are examples of this case.

- In step₂, letting $K = O(1)$ and the model reduce to $SBM(n, K, p_{in}, p_{out})$ for the non-overlapping network or $MMSB(n, K, \Pi, p_{in}, p_{out})$ for the overlapping network can always simplify the theoretical upper bound of error rate, as shown by our Corollaries 1 and 2. Here, we provide some examples about how to make a model degenerate to SBM. For $MMSB_n(K, \tilde{P}, \Pi, \rho)$ in this paper, when all nodes are pure, MMSB degenerates to SBM; for the $DCMM_n(K, \tilde{P}, \Pi, \Theta)$ model introduced in Section 5 or DCSBM considered in [30,48,50], setting $\Theta = \sqrt{\rho}I$ makes DCMM and DCSBM degenerates to SBM when all nodes are pure, similar to the ScBM and DCScBM considered in [67,68,71], the OCCAM model of [40], the stochastic blockmodel with the overlap proposed in [46], the extensions of SBM and DCSBM for hypergraph networks considered in [73–75], and so forth.
- In step₃ and step₄, the separation condition can be replaced by an alternative separation condition.
- When using SCSTC to build and compare theoretical results for the spectral clustering method, the key point is computing the lower bound for $\frac{|p_{in} - p_{out}|}{\sqrt{\max(p_{in}, p_{out})}}$ when the probability matrix P has diagonal entries p_{in} and non-diagonal entries p_{out} from the theoretical upper bound of the error rate for a certain spectral method. If this lower bound is consistent with that of Equation (1), this suggests theoretical optimality, and otherwise it suggests possible improvements by following the four steps of SCSTC.

The above analysis shows that SCSTC can be used to study the consistent estimation of model-based spectral methods. Use SCSTC, the following remark lists a few works whose main theoretical results leave possible improvements.

Remark 8. The unknown separation condition, or sub-optimal error rates, or a lack of requirement of network sparsity of some previous works, suggest possible improvements of their theoretical results. Here, we list a few works whose main results can be possibly improved until considering the separation condition.

- Theorem 4.4 of [48] proposes the upper bound of the error rate for their regularized spectral clustering (RSC) algorithm, designed based on a regularized Laplacian matrix under DCSBM. However, since [48] does not study the lower bound (in the [48] language) of λ_K and m , we cannot directly obtain the separation condition from their main theorem. Meanwhile, the main result of [48] does not consider the requirement on the network sparsity, which leaves some improvements. Ref. [48] does not study the theoretical optimal choice for the RSC regularizer τ . After considering $\sigma_K(\tilde{P})$ and sparsity parameter ρ , one can obtain the theoretical optimal choice for τ , and this is helpful for explaining and choosing the empirical optimal choice for τ . Therefore, the feasible network implementation of SCSTC is obtaining the theoretical optimal choices for some tuning parameters, such as regularizer τ of the RSC algorithm. By using SCSTC, we can find that RSC achieves thresholds in Equations (1)–(3), and we omit proofs for it in this paper.
- Refs. [26,49] study two algorithms designed based on the Laplacian matrix and its regularized version under SBM. They obtain meaningful results, but do not consider the network sparsity parameter ρ and separation parameter $\sigma_K(\tilde{P})$. After obtaining improved error bounds which are consistent with separation condition $\sqrt{\frac{\log(n)}{n}}$ using SCSTC, one can also obtain the theoretical optimal choice for regularizer τ of the RSC- τ algorithm considered in [49] and find that the two algorithms considered in [26,49] achieve thresholds in Equations (1)–(3).
- Theorem 2.2 of [50] provides an upper bound of their SCORE algorithm under DCSBM. However, since they do not consider the influence of $\sigma_K(\tilde{P})$, we cannot directly obtain the separation condition from their main result. Meanwhile, by setting their $\Theta = \sqrt{\rho}I$, DCSBM degenerates to SBM, which gives that their $err_n = \frac{1}{\rho^2 n} (1 + \frac{\log(n)}{\rho n}) = O(\frac{1}{\rho^2 n})$ by their assumption Equation (2.9). Hence, when $\Theta = \sqrt{\rho}I$, the upper bound of Theorem 2.2 in [50]

is $O(\frac{\log^3(n)}{\rho^2 n})$. The upper bound of error rate in Corollary 3.2 of [30] is $O(\frac{\log(n)}{\rho n})$ when using $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ under the setting that $\kappa(\Pi) = O(1), K = O(1)$ and $\sigma_K(\tilde{P}) = O(1)$. We see that $\frac{\log^3(n)}{\rho^2 n}$ grows faster than $\frac{\log(n)}{\rho n}$, which suggests that there is space to improve the main result of [50] in the aspects of the separation condition and error rates. Furthermore, using SCSTC, we can find that SCORE achieves thresholds in Equations (1)–(3) because its extension mixed-SCORE [41] achieves thresholds in Equations (1)–(3).

- Ref. [67] proposes two models, ScBM and DCScBM, to model the directed networks and an algorithm DI-SIM based on the directed regularized Laplacian matrix to fit DCScBM. However, similar to [48], their main theoretical result in their Theorem C.1 does not consider the lower bound of (in the language of Ref. [67]) σ_K, m_y, m_z and γ_z , which causes that we cannot obtain the separation condition when DCScBM degenerates to SBM. Meanwhile, their Theorem C.1 also lacks a lower bound requirement on network sparsity. Hence, there is space to improve the theoretical guarantees of [67]. Similar to [48,49], we can also obtain the theoretical optimal choices for regularizer τ of the DI-SIM algorithm and prove that DI-SIM achieves the thresholds in Equations (1)–(3) since it is the directed version of RSC [48].
- Ref. [68] mainly studies the theoretical guarantee for the D-SCORE algorithm proposed by [14] to fit a special case of the DCScBM model for directed networks. By setting their $\theta(i) = \sqrt{\rho}, \delta(j) = \sqrt{\rho}$ for $i, j \in [n]$, their directed-DCBM degenerates to SBM. Meanwhile, since their $err_n = \frac{1}{\rho}$, their mis-clustering rate is $O(\frac{T_n^2 \log(n)}{\rho n})$, which matches that of [30] under SBM when setting T_n as a constant. However, if setting T_n as $\log(n)$, then the error rate is $O(\frac{\log^3(n)}{\rho n})$, which is sub-optimal compared with that of [30]. Meanwhile, similar to [50,68], the main result does not consider the influences of K and $\sigma_K(\tilde{P})$, causing a lack of a separation condition. Hence, the main results of [68] can be improved by considering $K, \sigma_K(P)$, or a more optimal choice of T_n to make their main results comparable with those of [30] when directed-DCBM degenerates to SBM. Using SCSTC, we can find that the D-SCORE also achieves thresholds in Equations (1)–(3) since it is the directed version of SCORE [50].

5. Degree Corrected Mixed Membership Model

Using SCSTC to Theorem 3.2 of [43], as shown in Tables 1 and 2 results in Theorem 3.2 [43] being sub-optimal. To obtain the improvement theoretical results, we give a formal introduction of the degree corrected mixed membership (DCMM) model proposed in [41] first, then we review the SVM-cone-DCMMSB algorithm of [43] and provide the improvement theoretical results. A DCMM for generating A is as follows.

$$\Omega := \Theta \Pi \tilde{P} \Pi' \Theta \quad A(i, j) \sim \text{Bernoulli}(\Omega(i, j)) \quad i, j \in [n], \tag{5}$$

where $\Theta \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose i -th diagonal entry is the degree heterogeneity of node i for $i \in [n]$. Let $\theta \in \mathbb{R}^{n \times 1}$ with $\theta(i) = \Theta(i, i)$ for $i \in [n]$. Set $\theta_{\max} = \max_{i \in [n]} \theta(i), \theta_{\min} = \min_{i \in [n]} \theta(i)$ and $\tilde{P}_{\max} = \max_{k, l \in [K]} \tilde{P}(k, l), \tilde{P}_{\min} = \min_{k, l \in [K]} \tilde{P}(k, l)$.

Definition 3. Call model (5) the degree corrected mixed membership (DCMM) model, and denote it by $DCMM_n(K, \tilde{P}, \Pi, \Theta)$.

Note that if we set $\tilde{\Pi} = \Theta \Pi$ and choose Θ such that $\tilde{\Pi} \in \{0, 1\}^{n \times K}$, then we have $\Omega = \tilde{\Pi} \tilde{P} \tilde{\Pi}'$, which means that the stochastic blockmodel with overlap (SBMO) proposed in [46] is just a special case of DCMM. Meanwhile, if we write Θ as $\Theta = \tilde{\Theta} D_o$, where $\tilde{\Theta}, D_o$ are two positive diagonal matrices and let $\Pi_o = D_o \Pi$, then we can choose D_0 such that $\|\Pi_o(i, :)\|_F = 1$. By $\Omega = \Theta \Pi \tilde{P} \Pi' \Theta = \tilde{\Theta} \Pi_o \tilde{P} \Pi_o' \tilde{\Theta}$, we see that the OCCAM model proposed in [40] equals the DCMM model. By Equation (1.3) and Proposition 1.1 of [41], the following conditions are sufficient for the identifiability of DCMM when $\theta_{\max} \tilde{P}_{\max} \leq 1$:

- (II1) $\text{rank}(\tilde{P}) = K$ and \tilde{P} has unit diagonals.
- (II2) There is at least one pure node for each of the K communities.

Note that though the diagonal entries of \tilde{P} are ones, \tilde{P}_{\max} may be larger than 1 as long as $\theta_{\max} \tilde{P}_{\max} \leq 1$ under DCMM, and this is slightly different from the setting that $\max_{k,l \in [K]} \tilde{P}(k,l) = 1$ under MMSB.

Without causing confusion, under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, we still let $\Omega = U\Lambda U'$ be the top- K eigen-decomposition of Ω such that $U \in \mathbb{R}^{n \times K}, \Lambda \in \mathbb{R}^{K \times K}$ and $U'U = I_K$. Set $U_* \in \mathbb{R}^{n \times K}$ by $U_*(i,:) = \frac{U(i,:)}{\|U(i,:)\|_F}$ and let $N_U \in \mathbb{R}^{n \times n}$ be a diagonal matrix such that $N_U(i,i) = \frac{1}{\|U(i,:)\|_F}$ for $i \in [n]$. Then U_* can be rewritten as $U_* = N_U U$. The existence of the ideal cone (IC for short) structure inherent in U_* mentioned in [43] is guaranteed by the following lemma.

Lemma 3. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, $U_* = YU_*(\mathcal{I}, :)$ where $Y = N_M \Pi \Theta^{-1}(\mathcal{I}, \mathcal{I}) N_U^{-1}(\mathcal{I}, \mathcal{I})$ with N_M being an $n \times n$ diagonal matrix whose diagonal entries are positive.

Lemma 3 gives $Y = U_* U_*^{-1}(\mathcal{I}, :)$. Since $U_* = N_U U$ and $Y = N_M \Pi \Theta^{-1}(\mathcal{I}, \mathcal{I}) N_U^{-1}(\mathcal{I}, \mathcal{I})$, we have

$$N_U^{-1} N_M \Pi = U U_*^{-1}(\mathcal{I}, :) N_U(\mathcal{I}, \mathcal{I}) \Theta(\mathcal{I}, \mathcal{I}). \tag{6}$$

Since $\Omega(\mathcal{I}, \mathcal{I}) = \Theta(\mathcal{I}, \mathcal{I}) \Pi(\mathcal{I}, :) \tilde{P} \Pi'(\mathcal{I}, :) = \Theta(\mathcal{I}, \mathcal{I}) \tilde{P} \Theta(\mathcal{I}, \mathcal{I}) = U(\mathcal{I}, :) \Lambda U'(\mathcal{I}, :)$, we have $\Theta(\mathcal{I}, \mathcal{I}) \tilde{P} \Theta(\mathcal{I}, \mathcal{I}) = U(\mathcal{I}, :) \Lambda U'(\mathcal{I}, :)$. Then we have $\Theta(\mathcal{I}, \mathcal{I}) = \sqrt{\text{diag}(U(\mathcal{I}, :) \Lambda U'(\mathcal{I}, :))}$ when Condition (II1) holds such that \tilde{P} has unit-diagonals. Set $J_* = N_U(\mathcal{I}, \mathcal{I}) \Theta(\mathcal{I}, \mathcal{I}) \equiv \sqrt{\text{diag}(U_*(\mathcal{I}, :) \Lambda U'_*(\mathcal{I}, :))}$, $Z_* = N_U^{-1} N_M \Pi$, $Y_* = U U_*^{-1}(\mathcal{I}, :)$. By Equation (6), we have

$$Z_* = Y_* J_* \equiv U U_*^{-1}(\mathcal{I}, :) \text{diag}(U_*(\mathcal{I}, :) \Lambda U'_*(\mathcal{I}, :)). \tag{7}$$

Meanwhile, since $N_U^{-1} N_M$ is an $n \times n$ positive diagonal matrix, we have

$$\Pi(i,:) = \frac{Z_*(i,:)}{\|Z_*(i,:)\|_1}, i \in [n]. \tag{8}$$

With given Ω and K , we can obtain U, U_* and Λ . The above analysis shows that once $U_*(\mathcal{I}, :)$ is known, we can exactly recover Π by Equations (7) and (8). From Lemma 3, we know that $U_* = YU_*(\mathcal{I}, :)$ forms the IC structure. Ref. [43] proposes the SVM-cone algorithm (i.e., Algorithm A2) which can exactly obtain $U_*(\mathcal{I}, :)$ from the ideal cone $U_* = YU_*(\mathcal{I}, :)$ with inputs U_* and K .

Based on the above analysis, we are now ready to give the ideal SVM-cone-DCMMSB algorithm. Input Ω, K . Output: Π .

- Let $\Omega = U\Lambda U'$ be the top- K eigen-decomposition of Ω such that $U \in \mathbb{R}^{n \times K}, \Lambda \in \mathbb{R}^{K \times K}, U'U = I$. Let $U_* = N_U U$, where N_U is a $n \times n$ diagonal matrix whose i -th diagonal entry is $\frac{1}{\|U(i,:)\|_F}$ for $i \in [n]$.
- Run SVM-cone algorithm on U_* assuming that there are K communities to obtain \mathcal{I} .
- Set $J_* = \sqrt{\text{diag}(U_*(\mathcal{I}, :) \Lambda U'_*(\mathcal{I}, :))}$, $Y_* = U U_*^{-1}(\mathcal{I}, :)$, $Z_* = Y_* J_*$.
- Recover Π by setting $\Pi(i,:) = \frac{Z_*(i,:)}{\|Z_*(i,:)\|_1}$ for $i \in [n]$.

With given U_* and K , since the SVM-cone algorithm returns $U_*(\mathcal{I}, :)$, the ideal SVM-cone-DCMMSB exactly (for detail, see Appendix B) returns Π .

Now, we review the SVM-cone-DCMMSB algorithm of [43], where this algorithm can be seen as an extension of SPACL designed under MMSB to fit DCMM. For the real case, use $\hat{Y}_*, \hat{J}_*, \hat{Z}_*, \hat{\Pi}_*$ given in Algorithm 2 to estimate Y_*, J_*, Z_*, Π , respectively.

Algorithm 2 SVM-cone-DCMMSB [43].

Require: The adjacency matrix $A \in \mathbb{R}^{n \times n}$ and the number of communities K .

Ensure: The estimated $n \times K$ membership matrix $\hat{\Pi}_*$.

- 1: Obtain $\hat{A} = \hat{U}\hat{\Lambda}\hat{U}'$, the top K eigen-decomposition of A . Let $\hat{U}_* \in \mathbb{R}^{n \times K}$ such that $\hat{U}_*(i, :) = \frac{\hat{U}(i, :)}{\|\hat{U}(i, :)\|_F}$ for $i \in [n]$.
 - 2: Apply SVM-cone algorithm (i.e., Algorithm A2) on the rows of \hat{U}_* assuming there are K communities to obtain $\hat{\mathcal{I}}_*$, the index set returned by SVM-cone algorithm.
 - 3: Set $\hat{J}_* = \sqrt{\text{diag}(\hat{U}_*(\hat{\mathcal{I}}_*, :) \hat{\Lambda} \hat{U}'_*(\hat{\mathcal{I}}_*, :))}$, $\hat{Y}_* = \hat{U} \hat{U}_*^{-1}(\hat{\mathcal{I}}_*, :)$, $\hat{Z}_* = \hat{Y}_* \hat{J}_*$. Then set $\hat{Z}_* = \max(0, \hat{Z}_*)$.
 - 4: Estimate $\Pi(i, :)$ by $\hat{\Pi}_*(i, :) = \hat{Z}_*(i, :)/\|\hat{Z}_*(i, :)\|_1, i \in [n]$.
-

Consistency under DCMM

Assume that

(A2) $\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \geq \log(n)$.

Since we let $\tilde{P}_{\max} \leq C$, Assumption (A2) equals $\theta_{\max} \|\theta\|_1 \geq \log(n)/C$. The following lemma bounds $\|A - \Omega\|$ under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$ when Assumption (A2) holds.

Lemma 4. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, when Assumption (A2) holds, with probability at least $1 - o(n^{-\alpha})$, we have

$$\|A - \Omega\| \leq \frac{\alpha + 1 + \sqrt{(\alpha + 1)(\alpha + 19)}}{3} \sqrt{\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \log(n)}.$$

Remark 9. Consider a special case when $\Theta = \sqrt{\rho}I$ such that DCMM degenerates to MMSB, since \tilde{P}_{\max} is assumed to be 1 under MMSB, Assumption (A2) and the upper bound of $\|A - \Omega\|$ in Lemma 4 are consistent with that of Lemma 1. When all nodes are pure, DCMM degenerates to DCSBM [45], then the upper bound of $\|A - \Omega\|$ in Lemma 4 is also consistent with Lemma 2.2 of [50]. Meanwhile, this bound is also consistent with Equation (6.34) in the first version of [41], which also applies the Bernstein inequality to bound $\|A - \Omega\|$. However, the bound is $C\sqrt{\theta_{\max} \|\theta\|_1}$ in Equation (C.53) of the latest version for [41], which applies Corollary 3.12 and Remark 3.13 of [76] to obtain the bound. Though the bound in Equation (C.53) of the latest version for [41] is sharper by a $\sqrt{\log(n)}$ term, Corollary 3.12 of [76] has constraints on $W(i, j)$ (here, $W = A - \Omega$) such that $W(i, j)$ can be written as $W(i, j) = \xi_{ij} b_{ij}$, where $\{\xi_{i,j} : i \geq j\}$ are independent symmetric random variables with unit variance, and $\{b_{i,j} : i \geq j\}$ are given scalars; see the proof of Corollary 3.12 [76] for detail. Therefore, without causing confusion, we also use A_{re} to denote the constraint A used in [41] such that $\|A_{re} - \Omega\| \leq C\sqrt{\theta_{\max} \|\theta\|_1}$. Furthermore, if we set $\rho \geq \max_{i,j} \Omega(i, j)$ such that $\rho \geq \theta_{\max}^2$, the bound in Lemma 4 also equals $\|A - \Omega\| \leq C\sqrt{\rho n \log(n)}$ and the assumption (A2) reads $\tilde{P}_{\max} \rho n \geq \log(n)$. The bound $\|A_{re} - \Omega\| \leq C\sqrt{\theta_{\max} \|\theta\|_1}$ in Equation (C.53) of [41] reads $\|A_{re} - \Omega\| \leq C\sqrt{\rho n}$.

Lemma 5. (Row-wise eigenspace error) Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, when Assumption (A2) holds, suppose $\sigma_K(\Omega) \geq C\theta_{\max} \sqrt{\tilde{P}_{\max} n \log(n)}$, with probability at least $1 - o(n^{-\alpha})$.

- When we apply Theorem 4.2 of [64], we have

$$\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O\left(\frac{\theta_{\max} \sqrt{\tilde{P}_{\max}} K \left(\frac{\theta_{\max} \kappa(\Omega)}{\theta_{\min}} \sqrt{\frac{n}{K \lambda_K(\Pi' \Pi)}} + \sqrt{\log(n)}\right)}{\theta_{\min}^2 \sigma_K(\tilde{P}) \lambda_K(\Pi' \Pi)}\right).$$

- When we apply Theorem 4.2 of [65], we have

$$\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O\left(\frac{\theta_{\max} \sqrt{\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \log(n)}}{\theta_{\min}^3 \sigma_K(\tilde{P}) \lambda_K^{1.5}(\Pi' \Pi)}\right).$$

Without causing confusion, we also use $\omega, \omega_1, \omega_2$ under DCMM as Lemma 2 for notation convenience.

Remark 10. When $\Theta = \sqrt{\rho}I$ such that DCMM degenerates to MMSB, bounds in Lemma 5 are consistent with those of Lemma 2.

Remark 11. (Comparison to Theorem I.3 [43]) Note that the ρ in [43] is θ_{\max}^2 , which gives that the row-wise eigenspace concentration in Theorem I.3 [43] is $O(\frac{\theta_{\max}\sqrt{Kn}\|U\|_{2\rightarrow\infty}\log^\xi(n)}{\sigma_K(\Omega)})$ when using $\|A_{\text{re}} - \Omega\| \leq C\sqrt{\rho n}$ and this value is at least $O(\frac{\sqrt{\theta_{\max}\|\theta\|_1 K\|U\|_{2\rightarrow\infty}\log^\xi(n)}}{\sigma_K(\Omega)})$. Since $\|U\|_{2\rightarrow\infty} \leq \frac{\theta_{\max}}{\theta_{\min}\sqrt{\lambda_K(\Pi\Pi)}}$ by Lemma II.1 of [43] and $\sigma_K(\Omega) \geq \theta_{\min}^2 \sigma_K(\tilde{P})\lambda_K(\Pi\Pi)$ by the proof of Lemma 5, we see that the upper bound of Theorem I.3 [43] is $O(\frac{\theta_{\max}\sqrt{K\theta_{\max}\|\theta\|_1\log^\xi(n)}}{\theta_{\min}^3\sigma_K(\tilde{P})\lambda_K^{1.5}(\Pi\Pi)})$, which is $\sqrt{K}\log^{\xi-0.5}(n)$ (recall that $\xi > 1$) times than our ω_2 . Again, Theorem I.3 [43] has stronger requirements on the sparsity of $\theta_{\max}\|\theta\|_1$ and the lower bound of $\sigma_K(\Omega)$ than our Lemma 5. When using the bound of $\|A - \Omega\|$ in our Lemma 4 to obtain the row-wise eigenspace concentration in Theorem I.3 [43], their upper bound is $\sqrt{K}\log^\xi(n)$ times than our ω_2 . Similar to the first bullet given after Lemma 2, whether using $\|A - \Omega\| \leq C\sqrt{\theta_{\max}\|\theta\|_1\log(n)}$ or $\|A_{\text{re}} - \Omega\| \leq C\sqrt{\theta_{\max}\|\theta\|_1}$ does not change our ω under DCMM.

Remark 12. (Comparison to Lemma 2.1 [41]) The fourth bullet of Lemma 2.1 [41] is the row-wise deviation bound for the eigenvectors of the adjacency matrix under some assumptions translated to our $\kappa(\Pi\Pi) = O(1)$, Assumption (A2) and lower bound requirement on $\sigma_K(\Omega)$ since they apply Lemma C.2 [41]. The row-wise deviation bound in the fourth bullet of Lemma 2.1 [41] reads $O(\frac{\theta_{\max}K^{1.5}\sqrt{\theta_{\max}\|\theta\|_1\log(n)}}{\sigma_K(\tilde{P})\|\theta\|_F^3})$, where the denominator is $\sigma_K(\tilde{P})\|\theta\|_F^3$ instead of our $\theta_{\min}^3\sigma_K(\tilde{P})\lambda_K^{1.5}(\Pi\Pi)$ due to the fact that [41] uses $\frac{\sigma_K(\tilde{P})\|\theta\|_F^2}{K}$ to roughly estimate $\sigma_K(\Omega)$ while we apply $\theta_{\min}^2\sigma_K(\tilde{P})\lambda_K(\Pi\Pi)$ to strictly control the lower bound of $\sigma_K(\Omega)$. Therefore, we see that the row-wise deviation bound in the fourth bullet of Lemma 2.1 [41] is consistent with our bounds in Lemma 5 when $\kappa(\Pi\Pi) = O(1)$, while our row-wise eigenspace errors in Lemma 5 are more applicable than those of [41] since we do not need to add a constraint on $\Pi\Pi$ such that $\kappa(\Pi\Pi) = O(1)$. The upper bound of $\|A - \Omega\|$ of [41] is $C\sqrt{\theta_{\max}\|\theta\|_1}$ given in their Equation (C.53) under $\text{DCMM}_n(K, \tilde{P}, \Pi, \Theta)$, while ours is $C\sqrt{\theta_{\max}\|\theta\|_1\log(n)}$ in Lemma 4, since our bound of the row-wise eigenspace error in Lemma 5 is consistent with the fourth bullet of Lemma 2.1 [41], this supports the statement that the row-wise eigenspace error does not rely on $\|A - \Omega\|$ given in the first bullet after Lemma 2.

Let $\pi_{\min} = \min_{1 \leq k \leq K} \mathbf{1}'\Pi e_k$, where π_{\min} measures the minimum summation of nodes belonging to a certain community. Increasing π_{\min} makes the network tend to be more balanced, and vice versa. Meanwhile, the term π_{\min} appears when we propose a lower bound of η defined in Lemma A1 to keep track of the model parameters in our main theorem under $\text{DCMM}_n(K, \tilde{P}, \Pi, \Theta)$. The next theorem gives the theoretical bounds on estimations of memberships under DCMM.

Theorem 2. Under $\text{DCMM}_n(K, \tilde{P}, \Pi, \Theta)$, let $\hat{\Pi}$ be obtained from Algorithm 2, suppose conditions in Lemma 5 hold, and there exists a permutation matrix $\mathcal{P}_* \in \mathbb{R}^{K \times K}$ such that with probability at least $1 - o(n^{-\alpha})$, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi}_* - \Pi\mathcal{P}_*)\|_1 = O\left(\frac{\theta_{\max}^{15}K^5\omega\kappa^{4.5}(\Pi\Pi)\lambda_1^{1.5}(\Pi\Pi)}{\theta_{\min}^{15}\pi_{\min}}\right).$$

For comparison, Table 4 summarizes the necessary conditions and dependence on model parameters of rates for Theorem 2 and Theorem 3.2 [43], where the dependence on K and $\log(n)$ are analyzed in Remark 13 given below.

Remark 13. (Comparison to Theorem 3.2 [43]) Our bound in Theorem 2 is written as combinations of model parameters and Π can follow any distribution as long as Condition (I2) holds, where such model parameters' related form of estimation bound is convenient for further theoretical analysis (see Corollary 2), while the bound in Theorem 3.2 [43] is built when Π follows a Dirichlet distribution and $\kappa(\Pi'\Theta^2\Pi) = O(1)$. Meanwhile, since Theorem 3.2 [43] applies Theorem I.3 [43] to obtain the row-wise eigenspace error, the bound in Theorem 3.2 [43] should multiply $\log^{\xi}(n)$ by Remark 11, and this is also supported by the fact that in the proof of Theorem 3.1 [43], when computing bound of ϵ_0 (in the language in Ref. [43]) [43] ignores the $\log^{\xi}(n)$ term.

Consider a special case by setting $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$, $\pi_{\min} = O(\frac{n}{K})$ and $\frac{\theta_{\max}}{\theta_{\min}} = O(1)$ with $\theta_{\max} = \sqrt{\rho}$, where such case matches the setting $\kappa(\Pi'\Theta^2\Pi) = O(1)$ in Theorem 3.2 [43]. Now we focus on analyzing the powers of K in our Theorem 2 and Theorem 3.2 [43]. Under this case, the power of K in the estimation bound of Theorem 2 is 6 by basic algebra; since $\min(K^2, \kappa^2(\Omega)) = \min(K^2, O(1)) = O(1)$, $\frac{1}{\lambda_K^2(\Pi'\Theta^2\Pi)} = O(\frac{K^2}{\rho^2 n^2})$, $\frac{1}{\eta} = O(K)$ by Lemma A1 where η in Lemma A1 follows the same definition as that of Theorem 3.2 [43], and the bound in Theorem 3.2 [43] should multiply \sqrt{K} because (in the language of Ref. [43]) $\|(\hat{Y}_C \hat{Y}'_C)^{-1}\|_F$ should be no larger than $\frac{\sqrt{K}}{\lambda_K(\hat{Y}_C \hat{Y}'_C)}$ instead of $\frac{1}{\lambda_K(\hat{Y}_C \hat{Y}'_C)}$ in the proof of Theorem 2.8 [43], the power of K is 6 by checking the bound of Theorem 3.2 [43]. Meanwhile, note that our bound in Theorem 2 is l_1 bound, while the bound in Theorem 3.2 [43] is l_2 bound, and when we translate the l_2 bound of Theorem 3.2 [43] into l_1 bound, the power of K is 6.5 for Theorem 3.2 [43], suggesting that our bound in Theorem 2 has less dependence on K than that of Theorem 3.2 [43].

The following corollary is obtained by adding some conditions on the model parameters.

Corollary 2. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, when conditions of Lemma 5 hold, suppose $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$, $\pi_{\min} = O(\frac{n}{K})$ and $K = O(1)$, with probability at least $1 - o(n^{-\alpha})$, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi}_* - \Pi \mathcal{P}_*)\|_1 = O\left(\frac{\theta_{\max}^{16} \sqrt{\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \log(n)}}{\theta_{\min}^{18} \sigma_K(\tilde{P}) n}\right).$$

Meanwhile, when $\theta_{\max} = O(\sqrt{\rho})$, $\theta_{\min} = O(\sqrt{\rho})$ (i.e., $\frac{\theta_{\min}}{\theta_{\max}} = O(1)$), we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi}_* - \Pi \mathcal{P}_*)\|_1 = O\left(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\tilde{P}_{\max} \log(n)}{\rho n}}\right).$$

Remark 14. When $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$, $K = O(1)$, $\theta_{\max} = O(\sqrt{\rho})$ and $\theta_{\min} = O(\sqrt{\rho})$, the requirement $\sigma_K(\Omega) \geq C \theta_{\max} \sqrt{\tilde{P}_{\max} n \log(n)}$ in Lemma 5 holds naturally. By the proof of Lemma 5, $\sigma_K(\Omega)$ has a lower bound $\theta_{\min}^2 \sigma_K(\tilde{P}) \lambda_K(\Pi'\Pi) = O(\theta_{\min}^2 \sigma_K(P) n)$. To make the requirement $\sigma_K(\Omega) \geq C \theta_{\max} \sqrt{\tilde{P}_{\max} n \log(n)}$ always hold, we just need $\theta_{\min}^2 \sigma_K(\tilde{P}) n \geq C \theta_{\max} \sqrt{\tilde{P}_{\max} n \log(n)}$, and it gives $\sigma_K(\tilde{P}) \geq C \sqrt{\frac{\tilde{P}_{\max} \log(n)}{\rho n}}$, which matches the requirement of consistent estimation in Corollary 2.

Using SCSTC to Corollary 2, let $\Theta = \sqrt{\rho} I$ such that DCMM degenerates to MMSB, and it is easy to see that the bound in Lemma 2 is consistent with that of Lemma 1. Therefore, the separation condition, alternative separation condition and sharp threshold obtained from Corollary 2 for the extended version of SPACL under DCMM are consistent with classical results, as shown in Tables 1 and 2 (detailed analysis will be provided in next paragraph). Meanwhile, when $\theta_{\max} = O(\sqrt{\rho})$, $\theta_{\min} = O(\sqrt{\rho})$ and settings in Corollary 2 hold, the bound in Theorem 2.2 [41] is of order $\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}}$, which is consistent with our bound in Corollary 2.

Consider a mixed membership network under the settings of Corollary 2 when $\Theta = \sqrt{\rho}I$ such that DCMM degenerates to MMSB. By Corollary 2, $\frac{\sigma_K(\tilde{P})}{\sqrt{\tilde{P}_{\max}}}$ should shrink slower than $\sqrt{\frac{\log(n)}{\rho n}}$. We further assume that $\tilde{P} = (2 - \beta)I_K + (\beta - 1)\mathbf{1}\mathbf{1}'$ for $\beta \in [1, 2) \cup (2, \infty)$; we see that this \tilde{P} with unit diagonals and $\beta - 1$ as non-diagonal entries still satisfies Condition (III). Meanwhile, $\sigma_K(\tilde{P}) = |\beta - 2| = \tilde{P}_{\max} - \tilde{P}_{\min}$ and $\tilde{P}_{\max} = \max(1, \beta - 1)$, so $\frac{\sigma_K(\tilde{P})}{\sqrt{\tilde{P}_{\max}}} = \frac{|\beta - 2|}{\sqrt{\max(1, \beta - 1)}}$ should shrink slower than $\sqrt{\frac{\log(n)}{\rho n}}$. Setting $P = \rho\tilde{P}$ as the probability matrix for such \tilde{P} , we have $p_{\text{out}} = \rho(\beta - 1)$, $p_{\text{in}} = \rho$, and $\max(p_{\text{in}}, p_{\text{out}}) = \rho\max(1, \beta - 1)$. Sure, the separation condition $\frac{|p_{\text{in}} - p_{\text{out}}|}{\sqrt{\max(p_{\text{in}}, p_{\text{out}})}} \equiv \frac{\sqrt{\rho}|\beta - 2|}{\sqrt{\max(1, \beta - 1)}}$ should shrink slower than $\sqrt{\frac{\log(n)}{n}}$, which satisfies Equation (1). For an alternative separation condition and sharp threshold, just follow a similar analysis as that of MMSB, and we obtain the results in Tables 1 and 2.

6. Numerical Results

In this section, we present the experimental results for an overlapping network by plotting the phase transition behaviors for both SPACL and SVM-cone-DCMMSB to show that the two methods achieve the threshold in Equation (2) under $MMSB(n, K, \Pi, p_{\text{in}}, p_{\text{out}})$ when $K = 2$ and $K = 3$. We also use some experiments to show that the spectral methods studied in [26,30,48,50] achieve the threshold in Equation (2) under $SBM(n, K, p_{\text{in}}, p_{\text{out}})$ when $K = 2$ and $K = 3$ for the non-overlapping network. To measure the performance of different algorithms, we use the error rate defined below:

$$\min_{P \in \{K \times K \text{ permutation matrix}\}} \frac{1}{n} \|\hat{\Pi} - \Pi P\|_1. \tag{9}$$

For all simulations, let $p_{\text{in}} = \alpha_{\text{in}} \frac{\log(n)}{n}$ and $p_{\text{out}} = \alpha_{\text{out}} \frac{\log(n)}{n}$ be diagonal and non-diagonal entries of P , respectively. Since P is the probability matrix, α_{in} and α_{out} should be located in $(0, \frac{n}{\log(n)}]$. After setting P and Π , each simulation experiment has the following steps:

- (a) Set $\Omega = \Pi P \Pi'$.
- (b) Let W be an $n \times n$ symmetric matrix such that all diagonal entries of W are 0, and $W(i, j)$ are independent centered Bernoulli with parameters $\Omega(i, j)$. Let $A = \Omega - \text{diag}(\Omega) + W$ be the adjacency matrix of a simulated network with mixed memberships under MMSB (so there are no loops).
- (c) Apply spectral clustering method to A with K communities. Record the error rate.
- (d) Repeat (b)–(c) 50 times, and report the mean of the error rates over the 50 times.

Experiment 1: Set $n = 600, K = 2$, and $n_0 = 250$, where n_0 is the number of pure nodes in each community. Let all mixed nodes have mixed membership $(1/2, 1/2)$. Since α_{in} and α_{out} should be set to less than $\frac{n}{\log(n)} = \frac{600}{\log(600)} \approx 93.795$, we let α_{in} and α_{out} be in the range of $\{5, 10, 15, \dots, 90\}$. For each pair of $(\alpha_{\text{in}}, \alpha_{\text{out}})$, we generate P and then run steps (a)–(d) for this experiment. So, this experiment generates an adjacency matrix of network with mixed memberships under $MMSB(n, 2, \Pi, p_{\text{in}}, p_{\text{out}})$. The numerical results are displayed in panels (a) and (b) of Figure 1. We can see that our theoretical bounds (red lines) are quite tight, and the threshold regions obtained from the boundaries of light white areas in panels (a) and (b) are close to our theoretical bounds. Meanwhile, both methods perform better when $\frac{|\alpha_{\text{in}} - \alpha_{\text{out}}|}{\sqrt{\max(\alpha_{\text{in}}, \alpha_{\text{out}})}}$ increases and SVM-cone-DCMMSB outperforms SPACL for this experiment since panel (b) is darker than panel (a). Note that the network generated here is an assortative network when $\alpha_{\text{in}} > \alpha_{\text{out}}$, and the network is a dis-assortative network when $\alpha_{\text{in}} < \alpha_{\text{out}}$. So, the results of this experiment support our finding that SPACL and SVM-cone-DCMMSB achieves the threshold in Equation (2) for both assortative and dis-assortative networks.

Experiment 2: Set $n = 600, K = 3$, and $n_0 = 150$. Let all mixed nodes have mixed membership $(1/3, 1/3, 1/3)$. Let α_{in} and α_{out} range in $\{5, 10, 15, \dots, 90\}$. Sure, this experiment is under $MMSB(n, 3, \Pi, p_{in}, p_{out})$. The numerical results are displayed in panels (c) and (d) of Figure 1. We see that both methods perform poorly in the region between the two red lines, and the analysis of the numerical results for this experiment is similar to that of Experiment 1.

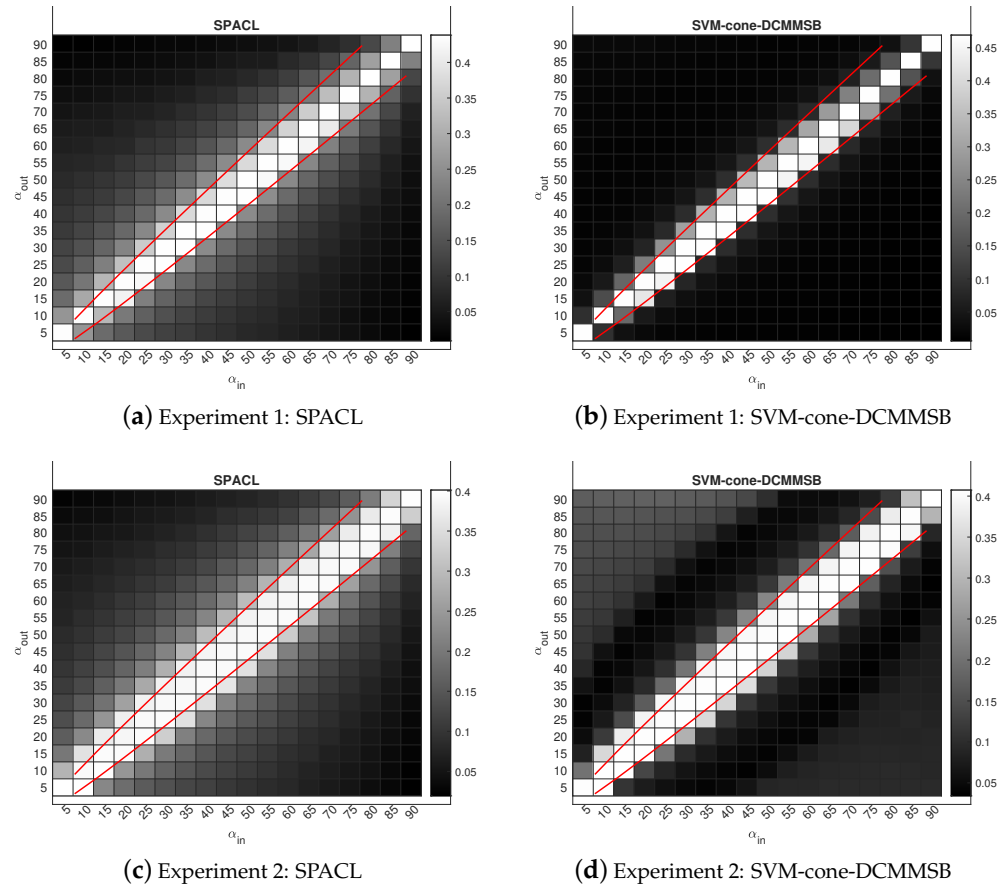


Figure 1. Phase transition for SPACL and SVM-cone-DCMMSB under MMSB: darker pixels represent lower error rates. The red lines represent $\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}} = 1$.

For visualization, we plot two networks generated from $MMSB(n, K, \Pi, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ in Figure 2. We also plot two dis-assortative networks generated from $MMSB(n, K, \Pi, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ in Figure A1 in Appendix A. In Experiments 1 and 2, there exist some mixed nodes for network generated under MMSB. The following two experiments only focus on network under SBM such that all nodes are pure. Meanwhile, we only consider four spectral algorithms studied in [26,30,48,50] for the non-overlapping network. For convenience, we call the spectral clustering method studied in [26] normalized principle component analysis (nPCA), and Algorithm 1 studied in [30] ordinary principle component analysis (oPCA), where nPCA and oPCA are also considered in [50]. Next, we briefly review nPCA, oPCA, RSC and SCORE.

The nPCA algorithm is as follows with input A, K and output $\hat{\Pi}$.

- Obtain the graph Laplacian $L = D^{-1/2}AD^{-1/2}$, where D is a diagonal matrix with $D(i, i) = \sum_{j=1}^n A(i, j)$ for $i \in [n]$.
- Obtain $\hat{U}\hat{\Lambda}\hat{U}'$, the top K eigen-decomposition of L .
- Apply k-means algorithm to \hat{U} to obtain $\hat{\Pi}$.

The oPCA algorithm is as follows with input A, K and output $\hat{\Pi}$.

- Obtain $\hat{U}\hat{\Lambda}\hat{U}'$, the top K eigen-decomposition of A .
- Apply k-means algorithm to \hat{U} to obtain $\hat{\Gamma}$.

The RSC algorithm is as follows with input A, K , regularizer τ , and output $\hat{\Gamma}$.

- Obtain the regularized graph Laplacian $L_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}$, where $D_\tau = D + \tau I$, and the default τ is the average node degree.
- Obtain $\hat{U}\hat{\Lambda}\hat{U}'$, the top K eigen-decomposition of L_τ . Let \hat{U}_* be the row-normalized version of \hat{U} .
- Apply k-means algorithm to \hat{U}_* to obtain $\hat{\Gamma}$.

The SCORE algorithm is as follows with input A, K , threshold T_n and output $\hat{\Gamma}$.

- Obtain the K (unit-norm) leading eigenvectors of A : $\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_K$.
- Obtain an $n \times (K - 1)$ matrix \hat{R}_* such that for $i \in [n], k \in [K - 1]$,

$$\hat{R}_*(i, k) = \begin{cases} \hat{R}(i, k), & \text{if } |\hat{R}(i, k)| \leq T_n, \\ T_n, & \text{if } \hat{R}(i, k) > T_n, \\ -T_n, & \text{if } \hat{R}(i, k) < -T_n, \end{cases}$$

where $\hat{R}(i, k) = \frac{\hat{\eta}_{k+1}(i)}{\hat{\eta}_1(i)}$, and the default T_n is $\log(n)$.

- Apply k-means algorithm to \hat{R}_* to obtain $\hat{\Gamma}$.

We now describe Experiments 3 and 4 under $SBM(n, K, p_{in}, p_{out})$ when $K = 2$ and $K = 3$.

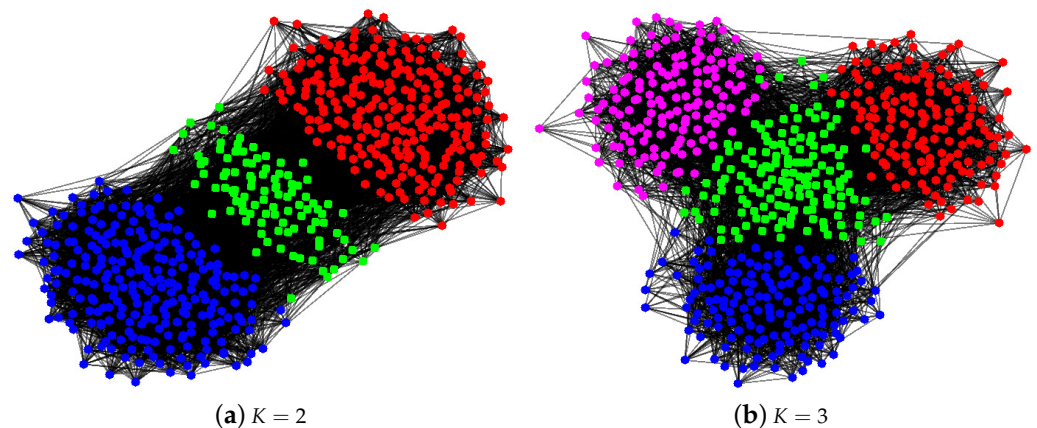


Figure 2. Panel (a): a graph generated from the mixed membership stochastic blockmodel with $n = 600$ nodes and 2 communities. Among the 600 nodes, each community has 250 pure nodes. For the 100 mixed nodes, they have mixed membership $(1/2, 1/2)$. Panel (b): a graph generated from MMSB with $n = 600$ nodes and 3 communities. Among the 600 nodes, each community has 150 pure nodes. For the 150 mixed nodes, they have mixed membership $(1/3, 1/3, 1/3)$. Nodes in panels (a,b) connect with probability $p_{in} = 60/600$ and $p_{out} = 1/600$, so the two networks in both panels are assortative networks. For panel (a), error rates for SPACL and SVM-cone-DCMMSB are 0.0285 and 0.0175, respectively, where error rate is defined in Equation (9). For panel (b), error rates for SPACL and SVM-cone-DCMMSB are 0.0709 and 0.0436, respectively. For both panels, dots in the same color are pure nodes in the same community and green square nodes are mixed.

Experiment 3: Set $n = 600, K = 2$, and $n_0 = 300$, i.e., all nodes are pure and each community has 300 nodes. So this experiment generates the adjacency matrix of the network under $SBM(n, 2, p_{in}, p_{out})$. Numerical results are displayed in panels (a)–(d) of Figure 3. We can see that these spectral clustering methods achieve the threshold in Equation (2).

Experiment 4: Set $n = 600, K = 3$, and $n_0 = 200$, i.e., all nodes are pure and each community has 200 nodes. So, this experiment is under $SBM(n, 3, p_{in}, p_{out})$. The numerical results are displayed in panels (e)–(h) of Figure 3. The results show that these methods achieve threshold in Equation (2).

For visualization, we plot two assortative networks generated from $SBM(n, K, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ in Figure 4. We also plot two dis-assortative networks generated from $SBM(n, K, p_{in}, p_{out})$ when $K = 2$ and $K = 3$ in Figure A2 in Appendix A.

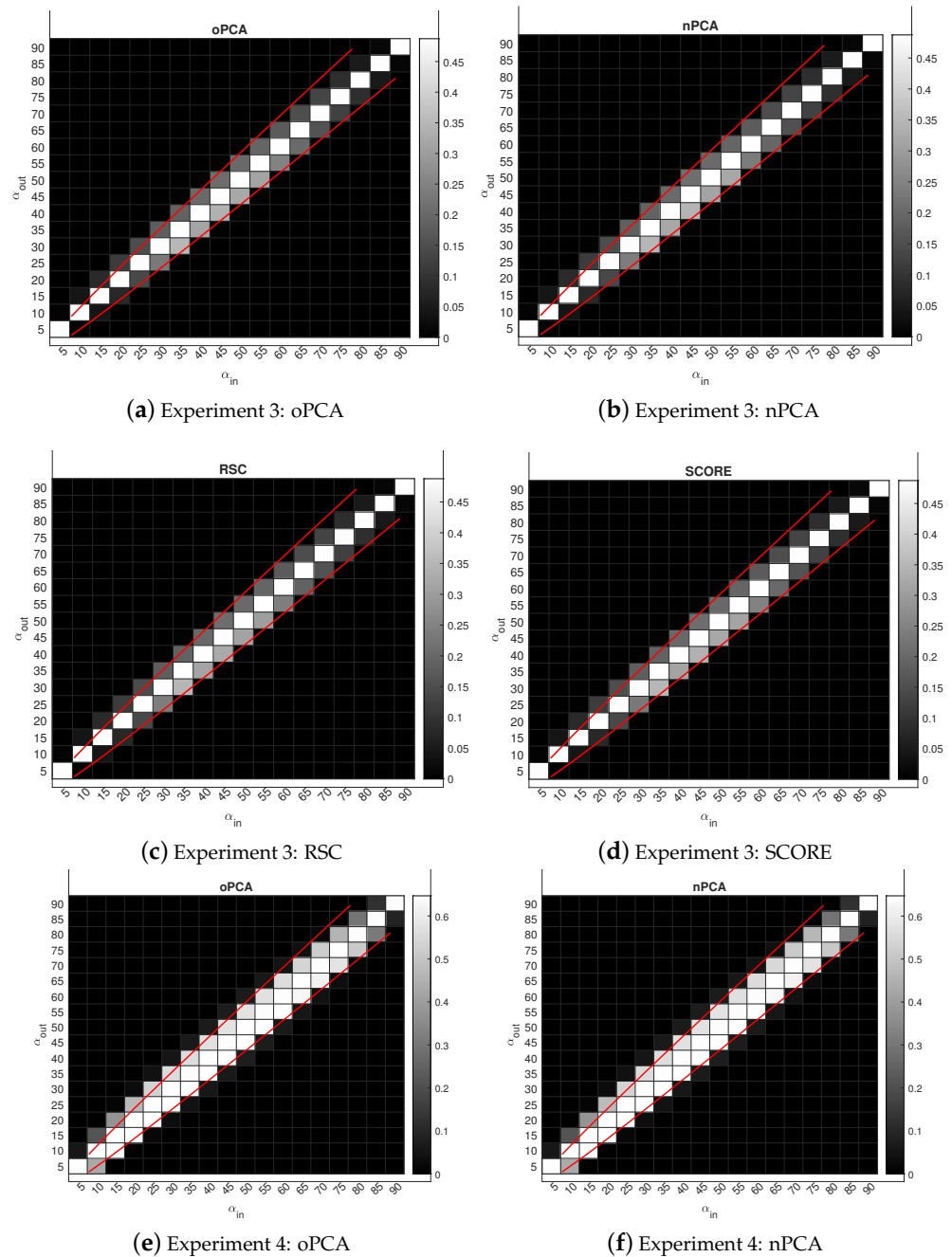
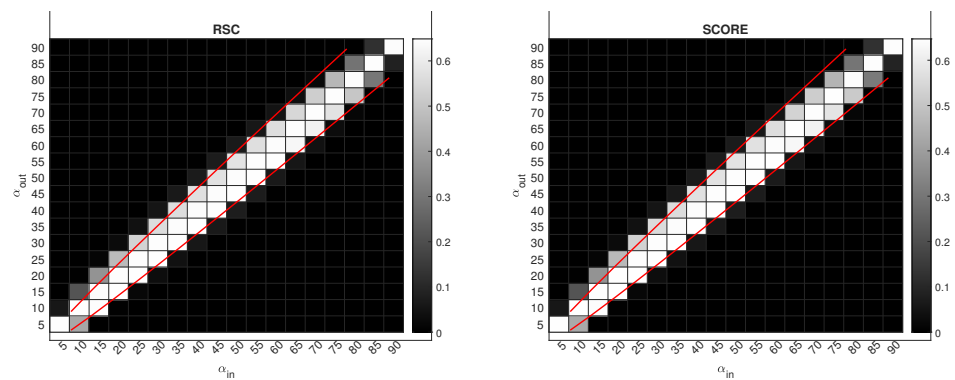


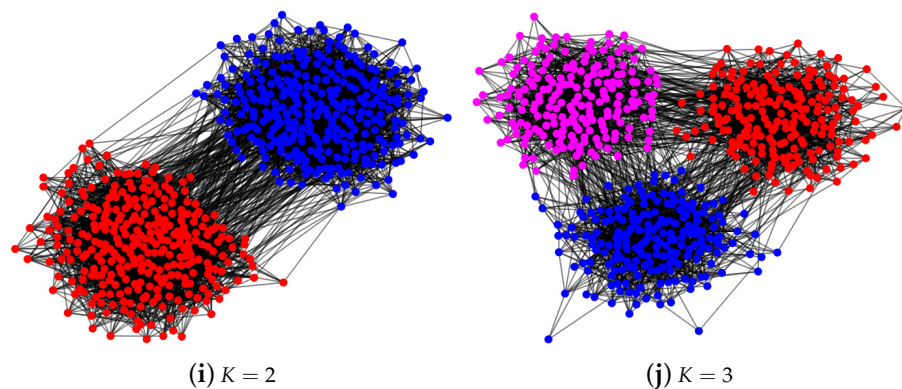
Figure 3. Cont.



(g) Experiment 4: RSC

(h) Experiment 4: SCORE

Figure 3. Phase transition for oPCA, nPCA, RSC and SCORE under SBM: darker pixels represent lower error rates. The red lines represent $\frac{|\alpha_{in} - \alpha_{out}|}{\sqrt{\max(\alpha_{in}, \alpha_{out})}} = 1$.



(i) $K = 2$

(j) $K = 3$

Figure 4. Panel (a): a graph generated from $SBM(600, 2, 30/600, 2/600)$. Panel (b): a graph generated from $SBM(600, 3, 30/600, 2/600)$. So, in panel (a), there are 2 communities and each community has 300 nodes; in panel (b), there are 3 communities and each community has 200 nodes. Networks in panels (a,b) are assortative networks since $p_{in} > p_{out}$. For both panels, error rates for oPCA, nPCA, RSC and SCORE are 0. Colors indicate clusters.

7. Conclusions

In this paper, the four-step separation condition and sharp threshold criterion SCSTC is summarized as a unified framework to study the consistencies and compare the theoretical error rates of spectral methods under models that can degenerate to SBM in a community detection area. With an application of this criterion, we find some inconsistent phenomena of a few previous works. In particular, using SCSTC, we find that the original theoretical upper bounds on error rates of the SPACL algorithm under MMSB and its extended version under DCMM are sub-optimal for the error rates and requirements on network sparsity. To find how the inconsistent phenomena occur, we re-establish theoretical upper bounds of error rates for both SPACL and its extended version by using recent techniques on row-wise eigenvector deviation. The resulting error bounds explicitly keep track of seven independent model parameters $(K, \rho, \sigma_K(\hat{P}), \lambda_K(\Pi' \Pi), \lambda_1(\Pi' \Pi), \theta_{min}, \theta_{max})$, which allow us to have a further delicate analysis. Compared with the original theoretical results, ours have smaller error rates with lesser dependence on K and $\log(n)$, weaker requirements on the network sparsity and the lower bound of the smallest nonzero singular value of population adjacency matrix under both MMSB and DCMM. For DCMM, we have no constraint on the distribution of the membership matrix as long as it satisfies the identifiability condition. When considering the separation condition of a standard network and the probability to generate a connected Erdős–Rényi (ER) random graph by using

SCSTC, our theoretical results match the classical results. Meanwhile, our theoretical results also match those of Theorem 2.2 [41] under mild conditions, and when DCMM degenerates to MMSB, the theoretical results under DCMM are consistent with those under MMSB. Using the SCSTC criterion, we find that the reasons behind the inconsistent phenomena are the sup-optimality of the original theoretical upper bounds on error rates for SPACL as well as its extended version, and the usage of a regularization version of the adjacency matrix when building theoretical results for spectral methods designed to detect nodes labels for a non-mixed network. The processes of finding these inconsistent phenomena, sub-optimality theoretical results on error rates and the formation mechanism of these inconsistent phenomena guarantee the usefulness of the SCSTC criterion. As shown by Remark 8, the theoretical results of some previous works can be improved by applying this criterion. Using SCSTC, we find that spectral methods considered in [26,41,43,44,48–50,67,68] achieve thresholds in Equations (1)–(3), and this conclusion is verified by both theoretical analysis and the numerical results in this paper. A limitation of this criterion is that it is only used for studying the consistency of spectral methods for a standard network with a constant number of communities. It would be interesting to develop a more general criterion that can study the consistency of all methods besides spectral methods, and models besides those can degenerate to SBM for a non-standard network with large K . Finally, we hope that the SCSTC criterion developed in this paper can be widely applied to build and compare theoretical results for spectral methods in the community detection area and that the thresholds in Equations (1)–(3) can be seen as benchmark thresholds for spectral methods.

Funding: This research was funded by Scientific research start-up fund of China University of Mining and Technology NO. 102520253, the High level personal project of Jiangsu Province NO.JSSCBS20211218.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SCSTC	separation condition and sharp threshold criterion
SBM	stochastic blockmodels
DCSBM	degree corrected stochastic blockmodel
MMSB	mixed membership stochastic blockmodel
DCMM	degree corrected mixed membership model
SBMO	stochastic blockmodel with overlap
OCCAM	overlapping continuous community assignment model
RSC	regularized spectral clustering
SCORE	spectral clustering on ratios-of-eigenvectors
SPACL	sequential projection after cleaning
ER	Erdős–Rényi
IS	ideal simplex
IC	ideal cone
SP	successive projection algorithm
oPCA	ordinary principle component analysis
nPCA	normalized principle component analysis

Appendix A. Additional Experiments

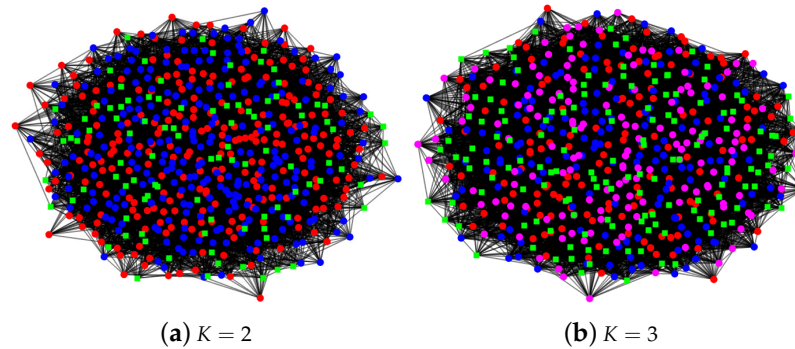


Figure A1. Panel (a): a graph generated from MMSB with $n = 600$ and $K = 2$. Each community has 250 pure nodes. For the 100 mixed nodes, they have mixed membership $(1/2, 1/2)$. Panel (b): a graph generated from MMSB with $n = 600$ and $K = 3$. Each community has 150 pure nodes. For the 150 mixed nodes, they have mixed membership $(1/3, 1/3, 1/3)$. Nodes in panels (a,b) connect with probability $p_{in} = 1/600$ and $p_{out} = 60/600$, so the two networks in both panels are dis-assortative networks. For panel (a), error rates for SPACL and SVM-cone-DCMMSB are 0.0298 and 0.0180, respectively. For panel (b), error rates for SPACL and SVM-cone-DCMMSB are 0.1286 and 0.0896, respectively. For both panels, dots in the same color are pure nodes in the same community, and green square nodes are mixed.

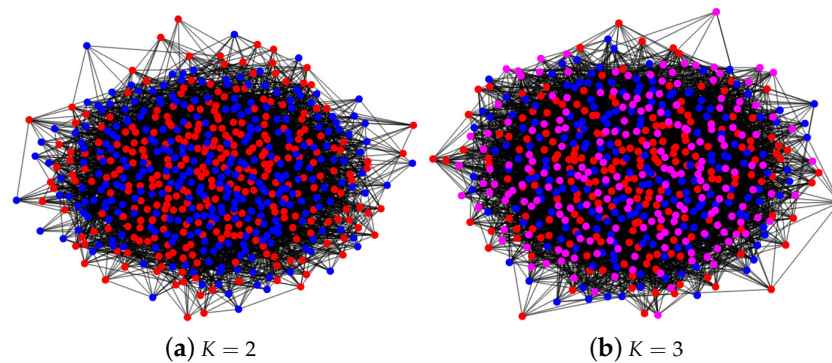


Figure A2. Panel (a): a graph generated from $SBM(600, 2, 2/600, 30/600)$. Panel (b): a graph generated from $SBM(600, 3, 2/600, 30/600)$. Networks in panels (a,b) are dis-assortative networks since $p_{in} < p_{out}$. For panel (a), error rates for oPCA, nPCA, RSC and SCORE are 0. For panel (b), error rates for oPCA, nPCA, RSC and SCORE are 0.0067. Colors indicate clusters.

Appendix B. Vertex Hunting Algorithms

The SP algorithm is written as below.

Algorithm A1 Successive projection (SP) [51].

Require: Near-separable matrix $Y_{sp} = S_{sp}M_{sp} + Z_{sp} \in \mathbb{R}_+^{m \times n}$, where S_{sp}, M_{sp} should satisfy Assumption 1 [51], the number r of columns to be extracted.

Ensure: Set of indices \mathcal{K} such that $Y(\mathcal{K}, :) \approx S$ (up to permutation)

- 1: Let $R = Y_{sp}, \mathcal{K} = \{ \}, k = 1$.
 - 2: **While** $R \neq 0$ and $k \leq r$ **do**
 - 3: $k_* = \operatorname{argmax}_k \|R(k, :)\|_F$.
 - 4: $u_k = R(k_*, :)$.
 - 5: $R \leftarrow (I - \frac{u_k u_k'}{\|u_k\|_F^2})R$.
 - 6: $\mathcal{K} = \mathcal{K} \cup \{k_*\}$.
 - 7: $k = k + 1$.
 - 8: **end while**
-

Based on Algorithm A1, the following theorem is Theorem 1.1 in [51], and it is also the Lemma VII.1 in [44]. This theorem provides the bound between the corner matrix S_{sp} and its estimated version returned by letting Y_{sp} be the input of the SP algorithm when $M'_{sp}S'_{sp}$ enjoys the ideal simplex structure.

Theorem A1. Fix $m \geq r$ and $n \geq r$. Consider a matrix $Y_{sp} = S_{sp}M_{sp} + Z_{sp}$, where $S_{sp} \in \mathbb{R}^{m \times r}$ has a full column rank, $M_{sp} \in \mathbb{R}^{r \times n}$ is a non-negative matrix such that the sum of each column is at most 1, and $Z_{sp} = [Z_{sp,1}, \dots, Z_{sp,n}] \in \mathbb{R}^{m \times n}$. Suppose that M_{sp} has a submatrix equal to I_r . Write $\epsilon \leq \max_{1 \leq i \leq n} \|Z_{sp,i}\|_F$. Suppose $\epsilon = O(\frac{\sigma_{\min}(S_{sp})}{\sqrt{r}\kappa^2(S_{sp})})$, where $\sigma_{\min}(S_{sp})$ and $\kappa(S_{sp})$ are the minimum singular value and condition number of S_{sp} , respectively. If we apply the SP algorithm to columns of Y_{sp} , then it outputs an index set $\mathcal{K} \subset \{1, 2, \dots, n\}$ such that $|\mathcal{K}| = r$ and $\max_{1 \leq k \leq r} \min_{j \in \mathcal{K}} \|S_{sp}(:, k) - Y_{sp}(:, j)\|_F = O(\epsilon\kappa^2(S_{sp}))$, where $S_{sp}(:, k)$ is the k -th column of S_{sp} .

For the ideal SPACL algorithm, since inputs of the ideal SPACL are Ω and K , we see that the inputs of SP algorithm are U and K . Let $m = K, r = K, Y_{sp} = U', Z_{sp} = U' - U' \equiv 0, S_{sp} = U'(\mathcal{I}, :)$, and $M_{sp} = \Pi'$. Then, we have $\max_{i \in [n]} \|U(i, :) - U(i, :)\|_F = 0$. By Theorem A1, the SP algorithm returns \mathcal{I} up to permutation when the input is U , assuming there are K communities. Since $U = \Pi U(\mathcal{I}, :)$ under $MMSB_n(K, P, \Pi, \rho)$, we see that $U(i, :) = U(j, :)$ as long as $\Pi(i, :) = \Pi(j, :)$. Therefore, though \mathcal{I} may be different up to the permutation, $U(\mathcal{I}, :)$ is unchanged. Therefore, following the four steps of the ideal SPACL algorithm, we see that it exactly returns Π .

Algorithm A2 below is the SVM-cone algorithm provided in [43].

Algorithm A2 SVM-cone [43].

Require: $\hat{S} \in \mathbb{R}^{n \times m}$ with rows have unit l_2 norm, number of corners K , estimated distance corners from hyperplane γ .

Ensure: The near-corner index set $\hat{\mathcal{I}}$.

- 1: Run one-class SVM on $\hat{S}(i, :)$ to get \hat{w} and \hat{b}
 - 2: Run k-means algorithm to the set $\{\hat{S}(i, :)|\hat{S}(i, :)\hat{w} \leq \hat{b} + \gamma\}$ that are close to the hyperplane into K clusters
 - 3: Pick one point from each cluster to get the near-corner set $\hat{\mathcal{I}}$
-

As suggested in [43], we can start $\gamma = 0$ and incrementally increase it until K distinct clusters are found. Meanwhile, for the ideal SVM-cone-DCMMSB algorithm, when setting U_* and K as the inputs of the SVM-cone algorithms, since $\|U_* - U_*\|_{2 \rightarrow \infty} = 0$, Lemma F.1. [43] guarantees that SVM-cone algorithm returns \mathcal{I} up to permutation. Since $U_* = YU_*(\mathcal{I}, :)$ by Lemma 3 under $DCMM_n(K, P, \Pi, \Theta)$, we have $U_*(i, :) = U_*(j, :)$ when $\Pi(i, :) = \Pi(j, :)$ by basic algebra, which gives that $U_*(\mathcal{I}, :)$ is unchanged though \mathcal{I} may be different up to permutation. Therefore, the ideal SVM-cone-DCMMSB exactly recovers Π .

Appendix C. Proof of Consistency under MMSB

Appendix C.1. Proof of Lemma 1

Proof. We apply Theorem 1.4 (Bernstein inequality) in [70] to bound $\|A - \Omega\|$, and this theorem is written as shown below.

Theorem A2. Consider a finite sequence $\{X_k\}$ of independent, random, self-adjoint matrices with dimension d . Assume that each random matrix satisfies

$$\mathbb{E}[X_k] = 0, \text{ and } \lambda_{\max}(X_k) \leq R \text{ almost surely.}$$

Then, for all $t \geq 0$,

$$\mathbb{P}(\lambda_{\max}(\sum_k X_k) \geq t) \leq d \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right),$$

where $\sigma^2 := \|\sum_k \mathbb{E}[X_k^2]\|$.

Let e_i be an $n \times 1$ vector, where $e_i(i) = 1$ and 0 elsewhere, for $i \in [n]$. For convenience, set $W = A - \Omega$. Then we can write W as $W = \sum_{i=1}^n \sum_{j=1}^n W(i, j)e_i e_j'$. Set $W^{(i, j)}$ as the $n \times n$ matrix such that $W^{(i, j)} = W(i, j)(e_i e_j' + e_j e_i')$, which gives $W = \sum_{1 \leq i < j \leq n} W^{(i, j)}$ where $\mathbb{E}[W^{(i, j)}] = 0$ and

$$\|W^{(i, j)}\| = \|W(i, j)(e_i e_j' + e_j e_i')\| = |W(i, j)| \|e_i e_j' + e_j e_i'\| = |W(i, j)| = |A(i, j) - \Omega(i, j)| \leq 1.$$

For the variance parameter $\sigma^2 := \|\sum_{1 \leq i < j \leq n} \mathbb{E}[(W^{(i, j)})^2]\|$. We bound $\mathbb{E}(W^2(i, j))$ as shown below:

$$\begin{aligned} \mathbb{E}(W^2(i, j)) &= \mathbb{E}((A(i, j) - \Omega(i, j))^2) = \text{Var}(A(i, j)) = \Omega(i, j)(1 - \Omega(i, j)) \leq \Omega(i, j) \\ &= \rho \Pi(i, :) \tilde{P} \Pi'(j, :) \leq \rho. \end{aligned}$$

Next we bound σ^2 as shown below:

$$\begin{aligned} \sigma^2 &= \left\| \sum_{1 \leq i < j \leq n} \mathbb{E}(W^2(i, j))(e_i e_j' + e_j e_i')(e_i e_j' + e_j e_i') \right\| = \left\| \sum_{1 \leq i < j \leq n} \mathbb{E}[W^2(i, j)(e_i e_i' + e_j e_j')] \right\| \\ &\leq \max_{1 \leq i \leq n} \left| \sum_{j=1}^n \mathbb{E}(W^2(i, j)) \right| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n \rho = \rho n. \end{aligned}$$

Set $t = \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3} \sqrt{\rho n \log(n)}$ for any $\alpha > 0$, combine Theorem A2 with $\sigma^2 \leq \rho n, R = 1, d = n$, and we have

$$\begin{aligned} \mathbb{P}(\|W\| \geq t) &= \mathbb{P}\left(\left\| \sum_{1 \leq i < j \leq n} W^{(i, j)} \right\| \geq t\right) \leq n \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right) \\ &\leq n \cdot \exp\left(\frac{-(\alpha+1)\log(n)}{\frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} \sqrt{\frac{\log(n)}{\rho n}}}\right) \leq \frac{1}{n^\alpha}, \end{aligned}$$

where we use Assumption (A1) such that $\frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} \sqrt{\frac{\log(n)}{\rho n}} \leq \frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} = 1$. \square

Appendix C.2. Proof of Lemma 2

Proof. Let $H = \hat{U}'U$, and $H = U_H \Sigma_H V_H'$ be the SVD decomposition of $H_{\hat{U}}$ with $U_H, V_H \in \mathbb{R}^{n \times K}$, where U_H and V_H represent respectively the left and right singular matrices of H . Define $\text{sgn}(H) = U_H V_H'$. Since $\mathbb{E}(A(i, j) - \Omega(i, j)) = 0, \mathbb{E}[(A(i, j) - \Omega(i, j))^2] \leq \rho$ by the proof of Lemma 1, $\frac{1}{\sqrt{\rho n / (\mu \log(n))}} \leq O(1)$ holds by Assumption (A1) where μ is the incoherence parameter defined as $\mu = \frac{n\|U\|_2^2}{K}$. By Theorem 4.2 [64], with high probability, we have

$$\|\hat{U} \text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq C \frac{\sqrt{K\rho}(\kappa(\Omega)\sqrt{\mu} + \sqrt{\log(n)})}{\sigma_K(\Omega)},$$

provided that $c_1\sigma_K(\Omega) \geq \sqrt{\rho n \log(n)}$ for some sufficiently small constant c_1 . By Lemma 3.1 of [44], we know that $\|U\|_{2 \rightarrow \infty}^2 \leq \frac{1}{\lambda_K(\Pi'\Pi)}$, which gives

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq C \frac{\sqrt{K\rho}(\kappa(\Omega)) \sqrt{\frac{n}{K\lambda_K(\Pi'\Pi)} + \sqrt{\log(n)}}}{\sigma_K(\Omega)}. \tag{A1}$$

Remark A1. By Theorem 4.2 of [65], when $\sigma_K(\Omega) \geq 4\|A - \Omega\|_\infty$, we have

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq 14 \frac{\|A - \Omega\|_\infty}{\sigma_K(\Omega)} \|U\|_{2 \rightarrow \infty}.$$

By Lemma 3.1 [44], we have

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq 14 \frac{\|A - \Omega\|_\infty}{\sigma_K(\Omega)} \sqrt{\frac{1}{\lambda_K(\Pi'\Pi)}}.$$

Unlike Lemma V.1 [44] which bounds $\|A - \Omega\|_\infty$ via the Chernoff bound and obtains $\|A - \Omega\|_\infty \leq C\rho n$ with high probability, we bound $\|A - \Omega\|_\infty$ by the Bernstein inequality using a similar idea as Equation (C.67) of [41]. Let $y = (y_1, y_2, \dots, y_n)'$ be any $n \times 1$ vector; by Equation (C.67) [41], we know that with an application of the Bernstein inequality, for any $t \geq 0$ and $i \in [n]$, we have

$$\mathbb{P}\left(\left|\sum_{j=1}^n (A(i, j) - \Omega(i, j))y(j)\right| > t\right) \leq 2\exp\left(-\frac{t^2/2}{\sum_{j=1}^n \Omega(i, j)y^2(j) + \frac{t\|y\|_\infty}{3}}\right).$$

By the proof of Lemma 1, we have $\Omega(i, j) \leq \rho$. Set $y(j)$ as 1 or -1 such that $(A(i, j) - \Omega(i, j))y(j) = |A(i, j) - \Omega(i, j)|$, we have

$$\mathbb{P}(\|A - \Omega\|_\infty > t) \leq 2\exp\left(-\frac{t^2/2}{\rho n + \frac{t}{3}}\right).$$

Set $t = \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3} \sqrt{\rho n \log(n)}$ for any $\alpha > 0$, by Assumption (A1), we have

$$\mathbb{P}(\|A - \Omega\|_\infty > t) \leq 2\exp\left(-\frac{t^2/2}{\rho n + \frac{t}{3}}\right) \leq n^{-\alpha}.$$

Hence, when $\sigma_K(\Omega) \geq C_0 \sqrt{\rho n \log(n)}$ where $C_0 = 4 \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3}$, with probability at least $1 - o(n^{-\alpha})$,

$$\|\hat{U}\text{sgn}(H_{\hat{U}}) - U\|_{2 \rightarrow \infty} \leq C \frac{\sqrt{\rho n \log(n)}}{\sigma_K(\Omega)} \sqrt{\frac{1}{\lambda_K(\Pi'\Pi)}}.$$

Note that when $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$, the above bound turns to be $C \frac{\sqrt{\rho K \log(n)}}{\sigma_K(\Omega)}$, which is consistent with that of Equation (A1). Also note that this bound $\frac{\sqrt{\rho n \log(n)}}{\sigma_K(\Omega)} \sqrt{\frac{1}{\lambda_K(\Pi'\Pi)}}$ is sharper than the $\frac{\rho n}{\sigma_K(\Omega)} \sqrt{\frac{1}{\lambda_K(\Pi'\Pi)}}$ of Lemma V.1 [44] by Assumption (A1).

Since \hat{U} and U have orthonormal columns, now we are ready to bound $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty}$:

$$\begin{aligned} \|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} &= \max_{i \in [n]} \|e'_i(UU' - \hat{U}\hat{U}')\|_F \\ &= \max_{i \in [n]} \|e'_i(UU' - \hat{U}\text{sgn}(H)U' + \hat{U}\text{sgn}(H)U' - \hat{U}\hat{U}')\|_F \end{aligned}$$

$$\begin{aligned}
 &\leq \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))U'\|_F + \max_{i \in [n]} \|e'_i \hat{U}(\text{sgn}(H)U' - \hat{U}')\|_F \\
 &= \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F + \max_{i \in [n]} \|\hat{U}(\text{sgn}(H)U' - \hat{U}')e_i\|_F \\
 &= \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F + \max_{i \in [n]} \|(\text{sgn}(H)U' - \hat{U}')e_i\|_F \\
 &= \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F + \max_{i \in [n]} \|e'_i(U(\text{sgn}(H))' - \hat{U}')\|_F \\
 &= \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F + \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F \\
 &= 2 \max_{i \in [n]} \|e'_i(U - \hat{U} \text{sgn}(H))\|_F = 2 \|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} \\
 &\leq C \frac{\sqrt{K}(\kappa(\Omega) \sqrt{\frac{n}{K\lambda_K(\Pi'\Pi)}} + \sqrt{\log(n)})}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K(\Pi'\Pi)},
 \end{aligned}$$

where the last inequality holds since $\sigma_K(\Omega) \geq \sigma_K(\tilde{P})\rho\lambda_K(\Pi'\Pi)$ under $MMSB_n(K, \tilde{P}, \Pi, \rho)$ by Lemma II.4 [44] This bound is $C \frac{\sqrt{n \log(n)}}{\sigma_K(\tilde{P})\sqrt{\rho}\lambda_K^1(\Pi'\Pi)}$ if we use Theorem 4.2 of [65].

Remark A2. By Theorem 4.5 [77], we have $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} \leq \sqrt{n}(\|\hat{U}\|_{2 \rightarrow \infty} + \|U\|_{2 \rightarrow \infty})\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} \leq \sqrt{n}(\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} + 2\|U\|_{2 \rightarrow \infty})\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} \leq \sqrt{n}(\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} + \frac{2}{\sqrt{\lambda_K(\Pi'\Pi)}})\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty} = O(\frac{2\sqrt{n}}{\sqrt{\lambda_K(\Pi'\Pi)}}\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty})$. Sure our bound $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} \leq 2\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty}$ enjoys concise form. In particular, when $\lambda_K(\Pi'\Pi) = O(\frac{n}{K})$ and $K = O(1)$, the two bounds give that $\|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} = O(\|U - \hat{U} \text{sgn}(H)\|_{2 \rightarrow \infty})$, which provides same error bound of the estimated memberships given in Corollary 1.

□

Appendix C.3. Proof of Theorem 1

Proof. Follow almost the same proof as Equation (3) of [44]. For $i \in [n]$, there exists a permutation matrix $\mathcal{P} \in \mathbb{R}^{K \times K}$ such that

$$\|e'_i(\hat{Z} - Z\mathcal{P})\|_F = O(\omega\kappa(\Pi'\Pi)\sqrt{K\lambda_1(\Pi'\Pi)}). \tag{A2}$$

Note that the bound in Equation (A2) is \sqrt{K} times the bound in Equation (3) of [44], and this is because in Equation (3) of [44], (in the language of Ref. [44]) $\|\hat{V}_p^{-1}\|$ denotes the Frobenius norm of \hat{V}_p^{-1} instead of the spectral norm. Since $\|\hat{V}_p^{-1}\|_F \leq \frac{\sqrt{K}}{\sigma_K(\hat{V}_p)}$, the bound in Equation (3) [44] should multiply \sqrt{K} .

Recall that $Z = \Pi, \Pi(i, :) = \frac{Z(i, :)}{\|Z(i, :)\|_1}, \hat{\Pi}(i, :) = \frac{\hat{Z}(j, :)}{\|\hat{Z}(j, :)\|_1}$, for $i \in [n]$, since

$$\begin{aligned}
 \|e'_i(\hat{\Pi} - \Pi\mathcal{P})\|_1 &= \left\| \frac{e'_i \hat{Z}}{\|e'_i \hat{Z}\|_1} - \frac{e'_i Z\mathcal{P}}{\|e'_i Z\mathcal{P}\|_1} \right\|_1 = \left\| \frac{e'_i \hat{Z} \|e'_i Z\|_1 - e'_i Z\mathcal{P} \|e'_i \hat{Z}\|_1}{\|e'_i \hat{Z}\|_1 \|e'_i Z\|_1} \right\|_1 \\
 &\leq \frac{\|e'_i \hat{Z} \|e'_i Z\|_1 - e'_i \hat{Z} \|e'_i \hat{Z}\|_1\|_1 + \|e'_i \hat{Z} \|e'_i \hat{Z}\|_1 - e'_i Z\mathcal{P} \|e'_i \hat{Z}\|_1\|_1}{\|e'_i \hat{Z}\|_1 \|e'_i Z\|_1} \\
 &= \frac{|\|e'_i Z\|_1 - \|e'_i \hat{Z}\|_1| + \|e'_i \hat{Z} - e'_i Z\mathcal{P}\|_1}{\|e'_i Z\|_1} \leq \frac{2\|e'_i(\hat{Z} - Z\mathcal{P})\|_1}{\|e'_i Z\|_1} \\
 &= \frac{2\|e'_i(\hat{Z} - Z\mathcal{P})\|_1}{\|e'_i \Pi\|_1} = 2\|e'_i(\hat{Z} - Z\mathcal{P})\|_1 \leq 2\sqrt{K}\|e'_i(\hat{Z} - Z\mathcal{P})\|_F \\
 &= O(\omega K\kappa(\Pi'\Pi)\sqrt{\lambda_1(\Pi'\Pi)}).
 \end{aligned}$$

□

Appendix C.4. Proof of Corollary 1

Proof. Under the conditions of Corollary 1, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi} - \Pi\mathcal{P})\|_1 = O(\omega\sqrt{n}).$$

Under the conditions of Corollary 1, Lemma 2 gives $\omega = O(\frac{1}{\sigma_K(\tilde{P})} \frac{1}{\sqrt{n}} \sqrt{\frac{\log(n)}{\rho n}})$, which gives that

$$\max_{i \in [n]} \|e'_i(\hat{\Pi} - \Pi\mathcal{P})\|_1 = O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{\rho n}}).$$

□

Appendix D. Proof of Consistency under DCMM

Appendix D.1. Proof of Lemma 3

Proof. Since $\Omega = U\Lambda U'$, we have $U = \Omega U\Lambda^{-1}$ since $U'U = I_K$. Recall that $\Omega = \Theta\Pi\tilde{P}\Pi'\Theta$, we have $U = \Theta\Pi\tilde{P}\Pi'\Theta U\Lambda^{-1} = \Theta\Pi B$, where we set $B = \tilde{P}\Pi'\Theta U\Lambda^{-1}$ for convenience. Since $U(\mathcal{I}, :) = \Theta(\mathcal{I}, \mathcal{I})\Pi(\mathcal{I}, :)B = \Theta(\mathcal{I}, \mathcal{I})B$, we have $B = \Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :)$.

Set $M = \Pi B$. Then we have $U = \Theta M$, which gives that $U(i, :) = e'_i U = \Theta(i, i)M(i, :)$ for $i \in [n]$. Therefore, $U_*(i, :) = \frac{U(i, :)}{\|U(i, :)\|_F} = \frac{M(i, :)}{\|M(i, :)\|_F}$, and combined with the fact that $B = \Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :) \equiv \Theta^{-1}(\mathcal{I}, \mathcal{I})N_U^{-1}(\mathcal{I}, \mathcal{I})N_U(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :) \equiv \Theta^{-1}(\mathcal{I}, \mathcal{I})N_U^{-1}(\mathcal{I}, \mathcal{I})U_*(\mathcal{I}, :)$, we have

$$U_* = \begin{bmatrix} \Pi(1, :)/\|M(1, :)\|_F \\ \Pi(2, :)/\|M(2, :)\|_F \\ \vdots \\ \Pi(n, :)/\|M(n, :)\|_F \end{bmatrix} B = \begin{bmatrix} \Pi(1, :)/\|M(1, :)\|_F \\ \Pi(2, :)/\|M(2, :)\|_F \\ \vdots \\ \Pi(n, :)/\|M(n, :)\|_F \end{bmatrix} \Theta^{-1}(\mathcal{I}, \mathcal{I})N_U^{-1}(\mathcal{I}, \mathcal{I})U_*(\mathcal{I}, :).$$

Therefore, we have

$$Y = N_M \Pi \Theta^{-1}(\mathcal{I}, \mathcal{I}) N_U^{-1}(\mathcal{I}, \mathcal{I}),$$

where N_M is a diagonal matrix whose i -th diagonal entry is $\frac{1}{\|M(i, :)\|_F}$ for $i \in [n]$. □

Appendix D.2. Proof of Lemma 4

Proof. Similar to the proof of Lemma 1, set $W = A - \Omega$ and $W^{(i,j)} = W(i, j)(e_i e'_j + e_j e'_i)$, we have $W = \sum_{1 \leq i < j \leq n} W^{(i,j)}$, $\mathbb{E}[W^{(i,j)}] = 0$ and $\|W^{(i,j)}\| \leq 1$. Since

$$\begin{aligned} \mathbb{E}(W^2(i, j)) &= \mathbb{E}((A(i, j) - \Omega(i, j))^2) = \text{Var}(A(i, j)) = \Omega(i, j)(1 - \Omega(i, j)) \\ &\leq \Omega(i, j) = \theta(i)\theta(j)\Pi(i, :)\tilde{P}\Pi'(j, :) \leq \theta(i)\theta(j)\tilde{P}_{\max}, \end{aligned}$$

we have

$$\begin{aligned} \sigma^2 &= \left\| \sum_{1 \leq i < j \leq n} \mathbb{E}(W^2(i, j))(e_i e'_j + e_j e'_i)(e_i e'_j + e_j e'_i) \right\| = \left\| \sum_{1 \leq i < j \leq n} \mathbb{E}[W^2(i, j)(e_i e'_j + e_j e'_i)] \right\| \\ &\leq \max_{1 \leq i \leq n} \left| \sum_{j=1}^n \mathbb{E}(W^2(i, j)) \right| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n \theta(i)\theta(j)\tilde{P}_{\max} \leq \tilde{P}_{\max} \theta_{\max} \|\theta\|_1. \end{aligned}$$

Set $t = \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3} \sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)}$ for any $\alpha > 0$, combine Theorem A2 with $\sigma^2 \leq \tilde{P}_{\max}\theta_{\max}\|\theta\|_1, R = 1, d = n$, we have

$$\begin{aligned} \mathbb{P}(\|W\| \geq t) &= \mathbb{P}\left(\left\|\sum_{1 \leq i < j \leq n} W^{(i,j)}\right\| \geq t\right) \leq n \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right) \\ &\leq n \cdot \exp\left(\frac{-(\alpha + 1)\log(n)}{\frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} \sqrt{\frac{\log(n)}{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1}}}\right) \leq \frac{1}{n^\alpha}, \end{aligned}$$

where we use Assumption (A2) such that $\frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} \sqrt{\frac{\log(n)}{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1}} \leq \frac{18}{(\sqrt{\alpha+1}+\sqrt{\alpha+19})^2} + \frac{2\sqrt{\alpha+1}}{\sqrt{\alpha+1}+\sqrt{\alpha+19}} = 1. \quad \square$

Appendix D.3. Proof of Lemma 5

Proof. The proof is similar to that of Lemma 2, so we omit most details. Since $\mathbb{E}(A(i, j) - \Omega(i, j)) = 0, \mathbb{E}[(A(i, j) - \Omega(i, j))^2] \leq \theta(i)\theta(j)\tilde{P}_{\max} \leq \theta_{\max}^2\tilde{P}_{\max}, \frac{1}{\theta_{\max}\sqrt{\tilde{P}_{\max}n/(\mu\log(n))}} \leq O(1)$ holds by Assumption (A2) where $\mu = \frac{n\|U\|_{2 \rightarrow \infty}^2}{K}$. By Theorem 4.2 [64], with high probability, we have

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq C \frac{\theta_{\max}\sqrt{\tilde{P}_{\max}K}(\kappa(\Omega)\sqrt{\mu} + \sqrt{\log(n)})}{\sigma_K(\Omega)},$$

provided that $c_*\sigma_K(\Omega) \geq \theta_{\max}\sqrt{\tilde{P}_{\max}n\log(n)}$ for some sufficiently small constant c_* . By Lemma H.1 of [43], we know that $\|U\|_{2 \rightarrow \infty}^2 \leq \frac{\theta_{\max}^2}{\lambda_K(\Pi'\Theta^2\Pi)} \leq \frac{\theta_{\max}^2}{\theta_{\min}^2\lambda_K(\Pi'\Pi)}$ under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, which gives

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq C \frac{\theta_{\max}\sqrt{\tilde{P}_{\max}K} \left(\frac{\theta_{\max}\kappa(\Omega)}{\theta_{\min}} \sqrt{\frac{n}{K\lambda_K(\Pi'\Pi)}} + \sqrt{\log(n)}\right)}{\sigma_K(\Omega)},$$

Remark A3. Similar to the proof of Lemma 2, by Theorem 4.2 of [65], when $\sigma_K(\Omega) \geq 4\|A - \Omega\|_\infty$, we have

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq 14 \frac{\|A - \Omega\|_\infty}{\sigma_K(\Omega)} \|U\|_{2 \rightarrow \infty} \leq \frac{14\theta_{\max}\|A - \Omega\|_\infty}{\theta_{\min}\sigma_K(\Omega)\sqrt{\lambda_K(\Pi'\Pi)}}.$$

Let $y = (y_1, y_2, \dots, y_n)'$ be any $n \times 1$ vector, and by the Bernstein inequality, for any $t \geq 0$ and $i \in [n]$, we have

$$\mathbb{P}\left(\left|\sum_{j=1}^n (A(i, j) - \Omega(i, j))y(j)\right| > t\right) \leq 2\exp\left(-\frac{t^2/2}{\sum_{j=1}^n \Omega(i, j)y^2(j) + \frac{t\|y\|_\infty}{3}}\right).$$

By the proof of Lemma 4, we have $\Omega(i, j) \leq \theta(i)\theta(j)\tilde{P}_{\max}$, which gives $\sum_{j=1}^n \Omega(i, j) \leq \tilde{P}_{\max}\theta_{\max}\|\theta\|_1$. Set $y(j)$ as 1 or -1 such that $(A(i, j) - \Omega(i, j))y(j) = |A(i, j) - \Omega(i, j)|$, we have

$$\mathbb{P}(\|A - \Omega\|_\infty > t) \leq 2\exp\left(-\frac{t^2/2}{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 + \frac{t}{3}}\right).$$

Set $t = \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3} \sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)}$ for any $\alpha > 0$, by Assumption (A2), we have

$$\mathbb{P}(\|A - \Omega\|_{\infty} > t) \leq 2\exp\left(-\frac{t^2/2}{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 + \frac{t}{3}}\right) \leq n^{-\alpha}.$$

Hence, when $\sigma_K(\Omega) \geq C_0 \sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)}$ where $C_0 = 4 \frac{\alpha+1+\sqrt{(\alpha+1)(\alpha+19)}}{3}$, with probability at least $1 - o(n^{-\alpha})$,

$$\|\hat{U}\text{sgn}(H) - U\|_{2 \rightarrow \infty} \leq C \frac{\theta_{\max} \sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)}}{\theta_{\min}\sigma_K(\Omega) \sqrt{\lambda_K(\Pi'\Pi)}}.$$

Meanwhile, since $\sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)} \leq \theta_{\max} \sqrt{\tilde{P}_{\max}n \log(n)}$, for convenience, we let the lower bound requirement of $\sigma_K(\Omega)$ be $C\theta_{\max} \sqrt{\tilde{P}_{\max}n \log(n)}$.

Similar to the proof of Lemma 2, we have

$$\begin{aligned} \|\hat{U}\hat{U}' - UU'\|_{2 \rightarrow \infty} &= \max_{i \in [n]} \|e'_i(UU' - \hat{U}\hat{U}')\|_F \leq 2\|U - \hat{U}\text{sgn}(H)\|_{2 \rightarrow \infty} \\ &\leq C \frac{\theta_{\max} \sqrt{\tilde{P}_{\max}K} \left(\frac{\theta_{\max}K(\Omega)}{\theta_{\min}} \sqrt{\frac{n}{K\lambda_K(\Pi'\Pi)}} + \sqrt{\log(n)}\right)}{\sigma_K(\Omega)} \\ &\leq C \frac{\theta_{\max} \sqrt{\tilde{P}_{\max}K} \left(\frac{\theta_{\max}K(\Omega)}{\theta_{\min}} \sqrt{\frac{n}{K\lambda_K(\Pi'\Pi)}} + \sqrt{\log(n)}\right)}{\theta_{\min}^2 \sigma_K(\tilde{P}) \lambda_K(\Pi'\Pi)}, \end{aligned}$$

where the last inequality holds since $\sigma_K(\Omega) = \sigma_K(\Theta\Pi\tilde{P}\Pi'\Theta) \geq \theta_{\min}^2 \sigma_K(\Pi\Pi')$
 $= \theta_{\min}^2 \sigma_K(\Pi'\Pi\tilde{P}) \geq \theta_{\min}^2 \sigma_K(\tilde{P}) \sigma_K(\Pi'\Pi) = \theta_{\min}^2 \sigma_K(\tilde{P}) \lambda_K(\Pi'\Pi)$. And this bound is $C \frac{\theta_{\max} \sqrt{\tilde{P}_{\max}\theta_{\max}\|\theta\|_1 \log(n)}}{\theta_{\min}^3 \sigma_K(\tilde{P}) \lambda_K^{1.5}(\Pi'\Pi)}$ if we use Theorem 4.2 of [65]. \square

Appendix D.4. Proof of Theorem 2

Proof. To prove this theorem, we follow similar procedures as Theorem 3.2 of [43]. For $i \in [n]$, recall that $Z_* = Y_* J_* \equiv N_U^{-1} N_M \Pi$, $\hat{Z}_* = \hat{Y}_* \hat{J}_*$, $\Pi(i, :) = \frac{Z(i,:)}{\|Z(i,:)\|_1}$ and $\hat{\Pi}_*(i, :) = \frac{\hat{Z}_*(i,:)}{\|\hat{Z}_*(i,:)\|_1}$, where N_M and M are defined in the proof of Lemma 3 such that $U = \Theta M \equiv \Theta \Pi B_*$ and $N_M(i, i) = \frac{1}{\|M(i,:)\|_F}$, we have

$$\|e'_i(\hat{\Pi}_* - \Pi P_*)\|_1 \leq \frac{2\|e'_i(\hat{Z}_* - Z_* P_*)\|_1}{\|e'_i Z_*\|_1} \leq \frac{2\sqrt{K}\|e'_i(\hat{Z}_* - Z_* P_*)\|_F}{\|e'_i Z_*\|_1}.$$

Now, we provide a lower bound of $\|e'_i Z_*\|_1$ as below

$$\begin{aligned} \|e'_i Z_*\|_1 &= \|e'_i N_U^{-1} N_M \Pi\|_1 = \|N_U^{-1}(i, i) e'_i N_M \Pi\|_1 = N_U^{-1}(i, i) \|N_M(i, i) e'_i \Pi\|_1 = \frac{N_M(i, i)}{N_U(i, i)} \\ &= \|U(i, :)\|_F N_M(i, i) = \|U(i, :)\|_F \frac{1}{\|M(i, :)\|_F} = \|U(i, :)\|_F \frac{1}{\|e'_i M\|_F} \\ &= \|U(i, :)\|_F \frac{1}{\|e'_i \Theta^{-1} U\|_F} = \|U(i, :)\|_F \frac{1}{\|\Theta^{-1}(i, i) e'_i U\|_F} = \theta(i) \geq \theta_{\min}. \end{aligned}$$

Therefore, by Lemma A3, we have

$$\|e'_i(\hat{\Pi}_* - \Pi P_*)\|_1 \leq \frac{2\sqrt{K}\|e'_i(\hat{Z}_* - Z_* P_*)\|_F}{\|e'_i Z_*\|_1} \leq \frac{2\sqrt{K}\|e'_i(\hat{Z}_* - Z_* P_*)\|_F}{\theta_{\min}}$$

$$= O\left(\frac{\theta_{\max}^{15} K^5 \omega \kappa^{4.5} (\Pi' \Pi) \lambda_1^{1.5} (\Pi' \Pi)}{\theta_{\min}^{15} \pi_{\min}}\right).$$

□

Appendix D.5. Proof of Corollary 2

Proof. Under conditions of Corollary 2, we have

$$\max_{i \in [n]} \|e'_i(\hat{\Pi}_* - \Pi \mathcal{P}_*)\|_1 = O\left(\frac{\theta_{\max}^{15} K^5 \omega \kappa^{4.5} (\Pi' \Pi) \lambda_1^{1.5} (\Pi' \Pi)}{\theta_{\min}^{15} \pi_{\min}}\right) = O\left(\frac{\theta_{\max}^{15} \omega \sqrt{n}}{\theta_{\min}^{15}}\right).$$

Under the conditions of Corollary 2, Lemma 5 gives $\omega = O\left(\frac{\theta_{\max} \sqrt{\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \log(n)}}{\theta_{\min}^3 \sigma_K(\tilde{P}) n^{1.5}}\right)$, which gives that

$$\max_{i \in [n]} \|e'_i(\hat{\Pi}_* - \Pi \mathcal{P}_*)\|_1 = O\left(\frac{\theta_{\max}^{15} \omega \sqrt{n}}{\theta_{\min}^{15}}\right) = O\left(\frac{\theta_{\max}^{16} \sqrt{\tilde{P}_{\max} \theta_{\max} \|\theta\|_1 \log(n)}}{\theta_{\min}^{18} \sigma_K(\tilde{P}) n}\right).$$

By basic algebra, this corollary follows. □

Appendix D.6. Basic Properties of Ω under DCMM

Lemma A1. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, we have

$$\|U(i, :)\|_F \geq \frac{\theta_{\min}}{\theta_{\max} \sqrt{K \lambda_1(\Pi' \Pi)}} \text{ for } i \in [n], \text{ and } \eta \geq \frac{\theta_{\min}^4 \pi_{\min}}{\theta_{\max}^4 K \lambda_1(\Pi' \Pi)},$$

where $\eta = \min_{k \in [K]} ((U_*(\mathcal{I}, :)) U'_*(\mathcal{I}, :))^{-1} \mathbf{1}(k)$.

Proof. Since $I = U'U = U'(\mathcal{I}, :)\Theta^{-1}(\mathcal{I}, \mathcal{I})\Pi'\Theta^2\Pi\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :)$ by the proof of Lemma 3, we have $((\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))')^{-1} = \Pi'\Theta^2\Pi$, which gives that

$$\begin{aligned} \min_k \|e'_k(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))\|_F^2 &= \min_k e'_k(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))' e_k \\ &\geq \min_{\|x\|=1} x'(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))' x \\ &= \lambda_K((\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))') = \frac{1}{\lambda_1(\Pi'\Theta^2\Pi)}, \end{aligned}$$

where x is a $K \times 1$ vector whose l_2 norm is 1. Then, for $i \in [n]$, we have

$$\begin{aligned} \|U(i, :)\|_F &= \|\theta_i \Pi(i, :)\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :)\|_F = \theta_i \|\Pi(i, :)\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :)\|_F \\ &\geq \theta_i \min_i \|\Pi(i, :)\|_F \min_i \|e'_i(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))\|_F \\ &\geq \theta_i \min_i \|e'_i(\Theta^{-1}(\mathcal{I}, \mathcal{I})U(\mathcal{I}, :))\|_F / \sqrt{K} \\ &\geq \frac{\theta_i}{\sqrt{K \lambda_1(\Pi'\Theta^2\Pi)}} \geq \frac{\theta_{\min}}{\theta_{\max} \sqrt{K \lambda_1(\Pi' \Pi)}}, \end{aligned}$$

where we use the fact that $\min_i \|\Pi(i, :)\|_F \geq \frac{1}{\sqrt{K}}$ since $\sum_{k=1}^K \Pi(i, k) = 1$ and all entries of Π are non-negative.

Since $U_* = N_U U$, we have

$$\begin{aligned} (U_*(\mathcal{I}, :)) U'_*(\mathcal{I}, :))^{-1} &= N_U^{-1}(\mathcal{I}, \mathcal{I})\Theta^{-1}(\mathcal{I}, \mathcal{I})\Pi'\Theta^2\Pi\Theta^{-1}(\mathcal{I}, \mathcal{I})N_U^{-1}(\mathcal{I}, \mathcal{I}) \\ &\geq \frac{\theta_{\min}^2}{\theta_{\max}^2 N_{U, \max}^2} \Pi' \Pi \geq \frac{\theta_{\min}^4}{\theta_{\max}^4 K \lambda_1(\Pi' \Pi)} \Pi' \Pi, \end{aligned}$$

where we set $N_{U,\max} = \max_{i \in [n]} N_U(i, i)$ and we use the fact that N_U, Θ are diagonal matrices and $N_{U,\max} \leq \frac{\theta_{\max} \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}}$. Then we have

$$\begin{aligned} \eta &= \min_{k \in [K]} ((U_*(\mathcal{I}, :) U'_*(\mathcal{I}, :))^{-1} \mathbf{1})(k) \geq \frac{\theta_{\min}^4}{\theta_{\max}^4 K \lambda_1(\Pi' \Pi)} \min_{k \in [K]} e'_k \Pi' \Pi \mathbf{1} \\ &= \frac{\theta_{\min}^4}{\theta_{\max}^4 K \lambda_1(\Pi' \Pi)} \min_{k \in [K]} e'_k \Pi' \mathbf{1} = \frac{\theta_{\min}^4 \pi_{\min}}{\theta_{\max}^4 K \lambda_1(\Pi' \Pi)}. \end{aligned}$$

□

Appendix D.7. Bounds between Ideal SVM-cone-DCMMSB and SVM-cone-DCMMSB

The next lemma focuses on the 2nd step of SVM-cone-DCMMSB and is the cornerstone to characterize the behaviors of SVM-cone-DCMMSB.

Lemma A2. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, when conditions of Lemma 5 hold, there exists a permutation matrix $\mathcal{P}_* \in \mathbb{R}^{K \times K}$ such that with probability at least $1 - o(n^{-\alpha})$, we have

$$\max_{1 \leq k \leq K} \|e'_k(\hat{U}_{*,2}(\hat{\mathcal{I}}_*, :) - \mathcal{P}'_* U_{*,2}(\mathcal{I}, :))\|_F = O\left(\frac{K^3 \theta_{\max}^{11} \omega \kappa^3 (\Pi' \Pi) \lambda_1^{1.5}(\Pi' \Pi)}{\theta_{\min}^{11} \pi_{\min}}\right),$$

where $U_{*,2} = U_* U'$, $\hat{U}_{*,2} = \hat{U}_* \hat{U}'$, i.e., $U_{*,2}, \hat{U}_{*,2}$ are the row-normalized versions of $U U'$ and $\hat{U} \hat{U}'$, respectively.

Proof. Lemma G.1. of [43] says that using $\hat{U}_{*,2}$ as input of the SVM-cone algorithm returns the same result as using \hat{U}_* as the input. By Lemma F.1 of [43], there exists a permutation matrix $\mathcal{P}_* \in \mathbb{R}^{K \times K}$ such that

$$\max_{k \in [K]} \|e'_k(\hat{U}_{*,2}(\hat{\mathcal{I}}_*, :) - \mathcal{P}'_* U_{*,2}(\mathcal{I}, :))\|_F = O\left(\frac{\sqrt{K} \zeta \epsilon_*}{\lambda_K^{1.5}(U_{*,2}(\mathcal{I}, :)) U'_{*,2}(\mathcal{I}, :)}\right),$$

where $\zeta \leq \frac{4K}{\eta \lambda_K^{1.5}(U_{*,2}(\mathcal{I}, :)) U'_{*,2}(\mathcal{I}, :)} = O\left(\frac{K}{\eta \lambda_K^{1.5}(U_*(\mathcal{I}, :)) U'_*(\mathcal{I}, :)}\right)$, $\epsilon_* = \max_{i \in [n]} \|\hat{U}_{*,2}(i, :) - U_{*,2}(i, :)\|_F$ and $\eta = \min_{1 \leq k \leq K} ((U_*(\mathcal{I}, :) U'_*(\mathcal{I}, :))^{-1} \mathbf{1})(k)$. Next we give upper bound of ϵ_* .

$$\begin{aligned} \|\hat{U}_{*,2}(i, :) - U_{*,2}(i, :)\|_F &= \left\| \frac{\hat{U}_2(i, :)\|U_2(i, :)\|_F - U_2(i, :)\|\hat{U}_2(i, :)\|_F}{\|\hat{U}_2(i, :)\|_F \|U_2(i, :)\|_F} \right\|_F \leq \frac{2\|\hat{U}_2(i, :) - U_2(i, :)\|_F}{\|U_2(i, :)\|_F} \\ &\leq \frac{2\|\hat{U}_2 - U_2\|_{2 \rightarrow \infty}}{\|U_2(i, :)\|_F} \leq \frac{2\omega}{\|U_2(i, :)\|_F} = \frac{2\omega}{\|(U U')(i, :)\|_F} = \frac{2\omega}{\|U(i, :)\|_F} \\ &\leq \frac{2\theta_{\max} \omega \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}}, \end{aligned}$$

where the last inequality holds by Lemma A1. Then, we have $\epsilon_* = O\left(\frac{\theta_{\max} \omega \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}}\right)$. By Lemma H.2. of [43], $\lambda_K(U_*(\mathcal{I}, :)) U'_*(\mathcal{I}, :)) \geq \frac{\theta_{\min}^2 \kappa^{-1} (\Pi' \Pi)}{\theta_{\max}^2}$. By the lower bound of η given in Lemma A1, we have

$$\max_{k \in [K]} \|e'_k(\hat{U}_{*,2}(\hat{\mathcal{I}}_*, :) - \mathcal{P}'_* U_{*,2}(\mathcal{I}, :))\|_F = O\left(\frac{K^3 \theta_{\max}^{11} \omega \kappa^3 (\Pi' \Pi) \lambda_1^{1.5}(\Pi' \Pi)}{\theta_{\min}^{11} \pi_{\min}}\right).$$

□

Next the lemma focuses on the 3rd step of SVM-cone-DCMMSB and bounds $\max_{i \in [n]} \|e'_i(\hat{Z}_* - Z_* \mathcal{P}_*)\|_F$.

Lemma A3. Under $DCMM_n(K, \tilde{P}, \Pi, \Theta)$, when the conditions of Lemma 5 hold, with a probability of at least $1 - o(n^{-\alpha})$, we have

$$\max_{i \in [n]} \|e'_i(\hat{Z}_* - Z_* \mathcal{P}_*)\|_F = O\left(\frac{\theta_{\max}^{15} K^{4.5} \omega \kappa^{4.5} (\Pi' \Pi) \lambda_1^{1.5} (\Pi' \Pi)}{\theta_{\min}^{14} \tau_{\min}}\right).$$

Proof. For $i \in [n]$, since $Z_* = Y_* J_*$, $\hat{Z}_* = \hat{Y}_* \hat{J}_*$ and J_*, \hat{J}_* are diagonal matrices, we have

$$\begin{aligned} \|e'_i(\hat{Z}_* - Z_* \mathcal{P}_*)\|_F &= \|e'_i(\max(0, \hat{Y}_* \hat{J}_*) - Y_* J_* \mathcal{P}_*)\|_F \leq \|e'_i(\hat{Y}_* \hat{J}_* - Y_* J_* \mathcal{P}_*)\|_F \\ &= \|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*) \hat{J}_* + e'_i Y_* \mathcal{P}_* (\hat{J}_* - J_* \mathcal{P}_*)\|_F \\ &\leq \|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F \|\hat{J}_*\| + \|e'_i Y_* \mathcal{P}_*\|_F \|\hat{J}_* - J_* \mathcal{P}_*\| \\ &= \|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F \|\hat{J}_*\| + \|e'_i Y_*\|_F \|\hat{J}_* - J_* \mathcal{P}_*\| \\ &= \|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F \|\hat{J}_*\| + \|e'_i Y_*\|_F \|J_* - \mathcal{P}_* \hat{J}_* \mathcal{P}_*\|. \end{aligned}$$

Therefore, the bound of $\|e'_i(\hat{Z}_* - Z_* \mathcal{P}_*)\|_F$ can be obtained as long as we bound $\|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F$, $\|\hat{J}_*\|$, $\|e'_i Y_*\|_F$ and $\|J_* - \mathcal{P}_* \hat{J}_* \mathcal{P}_*\|$. We bound the four terms as below:

- We bound $\|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F$ first. Set $U_*(\mathcal{I}, :) = B_*$, $\hat{U}_*(\hat{\mathcal{I}}_*, :) = \hat{B}_*$, $U_{*,2}(\mathcal{I}, :) = B_{2*}$, $\hat{U}_{*,2}(\hat{\mathcal{I}}_*, :) = \hat{B}_{2*}$ for convenience. For $i \in [n]$, we have

$$\begin{aligned} \|e'_i(\hat{Y}_* - Y_* \mathcal{P}_*)\|_F &= \|e'_i(\hat{U} \hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1} - U B'_*(B_* B'_*)^{-1} \mathcal{P}_*)\|_F \\ &= \|e'_i(\hat{U} - U(U' \hat{U})) \hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1} + e'_i(U(U' \hat{U}) \hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1} \\ &\quad - U(U' \hat{U})(\mathcal{P}'_*(B_* B'_*)(B'_*)^{-1}(U' \hat{U}))^{-1})\|_F \\ &\leq \|e'_i(\hat{U} - U(U' \hat{U})) \hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1}\|_F + \|e'_i U(U' \hat{U})(\hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1} \\ &\quad - (\mathcal{P}'_*(B_* B'_*)(B'_*)^{-1}(U' \hat{U}))^{-1})\|_F \\ &\leq \|e'_i(\hat{U} - U(U' \hat{U}))\|_F \|\hat{B}_*^{-1}\|_F + \|e'_i U(U' \hat{U})(\hat{B}'_*(\hat{B}_* \hat{B}'_*)^{-1} - (\mathcal{P}'_*(B_* B'_*)(B'_*)^{-1}(U' \hat{U}))^{-1})\|_F \\ &\leq \sqrt{K} \|e'_i(\hat{U} - U(U' \hat{U}))\|_F / \sqrt{\lambda_K(\hat{B}_* \hat{B}'_*)} + \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F \\ &\stackrel{(i)}{=} \sqrt{K} \|e'_i(\hat{U} \hat{U}' - U U') \hat{U}\|_F O\left(\frac{\theta_{\max} \sqrt{\kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F \\ &\leq \sqrt{K} \|e'_i(\hat{U} \hat{U}' - U U')\|_F O\left(\frac{\theta_{\max} \sqrt{\kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F \\ &\leq \sqrt{K} \omega O\left(\frac{\theta_{\max} \sqrt{\kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F \\ &= O\left(\omega \frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F, \end{aligned}$$

where we have used similar idea in the proof of Lemma VII.3 in [44] such that we apply $O\left(\frac{1}{\lambda_K(\hat{B}_* \hat{B}'_*)}\right)$ to estimate $\frac{1}{\lambda_K(\hat{B}_* \hat{B}'_*)}$.

Now we aim to bound $\|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F$. For convenience, set $T = U' \hat{U}$, $S = \mathcal{P}'_* B_* T$. We have

$$\begin{aligned} \|e'_i U(U' \hat{U})(\hat{B}_*^{-1} - (\mathcal{P}'_* B_*(U' \hat{U}))^{-1})\|_F &= \|e'_i U T S^{-1} (S - \hat{B}_*) \hat{B}_*^{-1}\|_F \\ &\leq \|e'_i U T S^{-1} (S - \hat{B}_*)\|_F \|\hat{B}_*^{-1}\|_F \leq \|e'_i U T S^{-1} (S - \hat{B}_*)\|_F \frac{\sqrt{K}}{|\lambda_K(\hat{B}_*)|} \\ &= \|e'_i U T S^{-1} (S - \hat{B}_*)\|_F \frac{\sqrt{K}}{\sqrt{\lambda_K(\hat{B}_* \hat{B}'_*)}} \leq \|e'_i U T S^{-1} (S - \hat{B}_*)\|_F O\left(\frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) \\ &= \|e'_i U T T^{-1} B'_*(B_* B'_*)^{-1} \mathcal{P}_* (S - \hat{B}_*)\|_F O\left(\frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) \end{aligned}$$

$$\begin{aligned}
 &= \|e'_i U B'_*(B_* B'_*)^{-1} \mathcal{P}_*(S - \hat{B}_*)\|_F O\left(\frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) \\
 &= \|e'_i Y_* \mathcal{P}_*(S - \hat{B}_*)\|_F O\left(\frac{\theta_{\max} \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}}\right) \leq \|e'_i Y_*\|_F \|S - \hat{B}_*\|_F O\left(\frac{\theta_{\max} \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}}\right) \\
 &\stackrel{\text{By Eq (A3)}}{\leq} \frac{\theta_{\max}^2 \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}^2 \lambda_K(\Pi' \Pi)} \max_{1 \leq k \leq K} \|e'_k (S - \hat{B}_*)\|_F O\left(\frac{\theta_{\max} K \sqrt{\kappa(\Pi' \Pi)}}{\theta_{\min}}\right) \\
 &= \max_{1 \leq k \leq K} \|e'_k (\hat{B}_* - \mathcal{P}'_* B_* U' \hat{U})\|_F O\left(\frac{\theta_{\max}^3 K^{1.5} \kappa(\Pi' \Pi)}{\theta_{\min}^3 \sqrt{\lambda_K(\Pi' \Pi)}}\right) \\
 &= \max_{1 \leq k \leq K} \|e'_k (\hat{B}_* \hat{U}' - \mathcal{P}'_* B_* U') \hat{U}\|_F O\left(\frac{\theta_{\max}^3 K^{1.5} \kappa(\Pi' \Pi)}{\theta_{\min}^3 \sqrt{\lambda_K(\Pi' \Pi)}}\right) \\
 &\leq \max_{1 \leq k \leq K} \|e'_k (\hat{B}_* \hat{U}' - \mathcal{P}'_* B_* U')\|_F O\left(\frac{\theta_{\max}^3 K^{1.5} \kappa(\Pi' \Pi)}{\theta_{\min}^3 \sqrt{\lambda_K(\Pi' \Pi)}}\right) \\
 &= \max_{1 \leq k \leq K} \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*})\|_F O\left(\frac{\theta_{\max}^3 K^{1.5} \kappa(\Pi' \Pi)}{\theta_{\min}^3 \sqrt{\lambda_K(\Pi' \Pi)}}\right) \\
 &\stackrel{\text{By Lemma A2}}{=} O\left(\frac{K^{4.5} \theta_{\max}^{14} \omega \kappa^{4.5}(\Pi' \Pi) \lambda_1(\Pi' \Pi)}{\theta_{\min}^{14} \pi_{\min}}\right).
 \end{aligned}$$

Then, we have

$$\begin{aligned}
 \|e'_i (\hat{Y}_* - Y_* \mathcal{P}_*)\|_F &\leq O\left(\omega \frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + \|e'_i U (U' \hat{U}) (\hat{B}_*^{-1} - (\mathcal{P}'_* B_* U' \hat{U})^{-1})\|_F \\
 &\leq O\left(\omega \frac{\theta_{\max} \sqrt{K \kappa(\Pi' \Pi)}}{\theta_{\min}}\right) + O\left(\frac{K^{4.5} \theta_{\max}^{14} \omega \kappa^{4.5}(\Pi' \Pi) \lambda_1(\Pi' \Pi)}{\theta_{\min}^{14} \pi_{\min}}\right) \\
 &= O\left(\frac{K^{4.5} \theta_{\max}^{14} \omega \kappa^{4.5}(\Pi' \Pi) \lambda_1(\Pi' \Pi)}{\theta_{\min}^{14} \pi_{\min}}\right).
 \end{aligned}$$

- for $\|e'_i Y_*\|_F$, since $Y_* = U U_*^{-1}(\mathcal{I}, :)$, we have

$$\|e'_i Y_*\|_F \leq \|U(i, :)\|_F \|U_*^{-1}(\mathcal{I}, :)\|_F \leq \frac{\sqrt{K} \|U(i, :)\|_F}{\sqrt{\lambda_K(U_*(\mathcal{I}, :), U_*(\mathcal{I}, :))}} \leq \frac{\theta_{\max}^2 \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}^2 \lambda_K(\Pi' \Pi)}. \tag{A3}$$

- for $\|\hat{J}_*\|$, recall that $\hat{J}_* = \sqrt{\text{diag}(\hat{U}_*(\hat{I}_*, :)\hat{\Lambda}\hat{U}'_*(\hat{\mathcal{I}}_*, :))}$, we have

$$\begin{aligned}
 \|\hat{J}_*\|^2 &= \max_{k \in [K]} \hat{J}_*^2(k, k) = \max_{k \in [K]} e'_k \hat{U}_*(\hat{I}_*, :)\hat{\Lambda}\hat{U}'_*(\hat{\mathcal{I}}_*, :)e_k \\
 &= \max_{k \in [K]} \|e'_k \hat{U}_*(\hat{I}_*, :)\hat{\Lambda}\hat{U}'_*(\hat{\mathcal{I}}_*, :)e_k\| \\
 &\leq \max_{k \in [K]} \|e'_k \hat{U}_*(\hat{I}_*, :)\|^2 \|\hat{\Lambda}\| \leq \max_{k \in [K]} \|e'_k \hat{U}_*(\hat{I}_*, :)\|_F^2 \|\hat{\Lambda}\| = \|\hat{\Lambda}\|,
 \end{aligned}$$

where we have used the fact that $\|\hat{U}_*(i, :)\|_F = 1$ for $i \in [n]$ in the last equality. Since we need $\sigma_K(\Omega) \geq C \theta_{\max} \sqrt{\tilde{P}_{\max} n \log(n)} \geq C \|A - \Omega\|$ in the proof of Lemma 5, we have $\|\hat{\Lambda}\| = \|A\| = \|A - \Omega + \Omega\| \leq \|A - \Omega\| + \|\Omega\| \leq \sigma_K(\Omega) + \|\Omega\| \leq 2\|\Omega\| = 2\|\Theta \Pi \tilde{P} \Pi' \Theta\| \leq 2C \theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi' \Pi) = O(\theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi' \Pi))$. Then we have

$$\|\hat{J}_*\| = O(\theta_{\max} \sqrt{\tilde{P}_{\max} \lambda_1(\Pi' \Pi)}).$$

- for $\|J_* - \mathcal{P}_* \hat{J}_* \mathcal{P}'_*\|$, we provide some simple facts first: $\|\hat{\Lambda}\| = \|A\|, \|\Lambda\| = \|\Omega\|, \Omega = U \Lambda U', \tilde{A} = \hat{U} \hat{\Lambda} \hat{U}', \|\hat{U}\| = 1, \|U\| = 1, \|e'_k \mathcal{P}_* \hat{B}_{2*}\| = \|\hat{B}_{2*} e_k\| = \|e'_k \hat{B}_{2*}\| \leq \|e'_k \hat{B}_{2*}\|_F = 1$. Since \tilde{A} is the best rank K approximation to A in the spectral norm, and therefore $\|\tilde{A} - A\| \leq \|\Omega - A\|$ since $\Omega = U \Lambda U'$ with rank K and Ω can also be viewed as a

rank K approximation to A . This leads to $\|\Omega - \tilde{A}\| = \|\Omega - A + A - \tilde{A}\| \leq 2\|A - \Omega\|$. By Lemma H.2 [43], $\|B_*\| = \|U_*(\mathcal{I}, :)\| = \sqrt{\lambda_1(U_*(\mathcal{I}, :)\tilde{U}'_*(\mathcal{I}, :))} \leq \sqrt{\kappa(\Pi'\Pi)} \leq \frac{\theta_{\max}\kappa^{0.5}(\Pi'\Pi)}{\theta_{\min}}$. $\|A\| = \|A - \Omega + \Omega\| \leq \|A - \Omega\| + \|\Omega\| \leq \sigma_K(\Omega) + \|\Omega\| \leq 2\|\Omega\|$ by the lower bound requirement of $\sigma_K(\Omega)$ in Lemma 5, and we also have $\|A - \Omega\| \leq \sigma_K(\Omega) \leq \|\Omega\|$. For $k \in [K]$, let $\tau_k = J_*(k, k)$, $\hat{\tau}_k = (\mathcal{P}_*\hat{J}_*\mathcal{P}'_*)(k, k)$ for convenience. Based on the above facts and Lemma A2, we have

$$\begin{aligned} \max_{k \in [K]} |\tau_k^2 - \hat{\tau}_k^2| &= \max_{k \in [K]} \|e'_k U_*(\mathcal{I}, :)\Lambda U'_*(\mathcal{I}, :)\tilde{e}_k - e'_k \mathcal{P}_* \tilde{U}_*(\hat{\mathcal{I}}_*, :)\hat{\Lambda} \hat{U}'_*(\hat{\mathcal{I}}_*, :)\mathcal{P}'_* \tilde{e}_k\| \\ &= \max_{k \in [K]} \|e'_k B_{2*} U \Lambda U' B_{2*}' \tilde{e}_k - e'_k \mathcal{P}_* \hat{B}_{2*} \hat{U} \hat{\Lambda} \hat{U}' \hat{B}_{2*}' \mathcal{P}'_* \tilde{e}_k\| \\ &\leq \|e'_k (B_{2*} - \mathcal{P}_* \hat{B}_{2*}) U \Lambda U' B_{2*}' \tilde{e}_k\| + \|e'_k \mathcal{P}_* \hat{B}_{2*} (U \Lambda U' - \hat{U} \hat{\Lambda} \hat{U}') B_{2*}' \tilde{e}_k\| \\ &\quad + \|e'_k \mathcal{P}_* \hat{B}_{2*} \hat{U} \hat{\Lambda} \hat{U}' (B_{2*}' - \hat{B}_{2*}' \mathcal{P}'_*) \tilde{e}_k\| \\ &\leq \|e'_k (B_{2*} - \mathcal{P}_* \hat{B}_{2*})\| \|U\| \|\Lambda\| \|U'\| \|B_{2*}' \tilde{e}_k\| + \|e'_k \mathcal{P}_* \hat{B}_{2*}\| \|U \Lambda U' - \hat{U} \hat{\Lambda} \hat{U}'\| \|B_{2*}' \tilde{e}_k\| \\ &\quad + \|e'_k \mathcal{P}_* \hat{B}_{2*}\| \|\hat{U}\| \|\hat{\Lambda}\| \|\hat{U}'\| \|B_{2*}' - \hat{B}_{2*}' \mathcal{P}'_*\| \| \tilde{e}_k \| \\ &\leq \|e'_k (B_{2*} - \mathcal{P}_* \hat{B}_{2*})\| \|\Lambda\| \|B_{2*}' \tilde{e}_k\| + \|U \Lambda U' - \hat{U} \hat{\Lambda} \hat{U}'\| \|B_{2*}' \tilde{e}_k\| + \|\hat{\Lambda}\| \|B_{2*}' - \hat{B}_{2*}' \mathcal{P}'_*\| \| \tilde{e}_k \| \\ &= \|e'_k (B_{2*} - \mathcal{P}_* \hat{B}_{2*})\| (\|\Omega\| \|B_{2*}' \tilde{e}_k\| + \|A\|) + \|\Omega - \tilde{A}\| \|B_{2*}' \tilde{e}_k\| \\ &= \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|\Omega\| \|B_{2*}' \tilde{e}_k\| + \|A\|) + \|\Omega - \tilde{A}\| \|B_{2*}' \tilde{e}_k\| \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|\Omega\| \|B_{2*}' \tilde{e}_k\| + \|A\|) + 2\|A - \Omega\| \|B_{2*}' \tilde{e}_k\| \\ &= \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|\Omega\| \|U B_{2*}' \tilde{e}_k\| + \|A\|) + 2\|A - \Omega\| \|U B_{2*}' \tilde{e}_k\| \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|\Omega\| \|B_{2*}' \tilde{e}_k\| + \|A\|) + 2\|A - \Omega\| \|B_{2*}' \tilde{e}_k\| \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|\Omega\| \|B_{2*}' \tilde{e}_k\| + 2\|\Omega\|) + 2\|\Omega\| \|B_{2*}' \tilde{e}_k\| \\ &= \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|B_{2*}' \tilde{e}_k\| + 1) O(\theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi'\Pi)) + \|B_{2*}' \tilde{e}_k\| O(\theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi'\Pi)) \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| (\|B_*\| + 1) O(\theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi'\Pi)) + \|B_*\| O(\theta_{\max}^2 \tilde{P}_{\max} \lambda_1(\Pi'\Pi)) \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\| O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) \\ &\quad + O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) \\ &\leq \|e'_k (\hat{B}_{2*} - \mathcal{P}'_* B_{2*}')\|_F O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) \\ &\quad + O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) \\ &= O\left(\frac{K^3 \theta_{\max}^{11} \omega \kappa^3(\Pi'\Pi) \lambda_1^{1.5}(\Pi'\Pi)}{\theta_{\min}^{11} \pi_{\min}}\right) O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) \\ &\quad + O(\theta_{\max}^3 \tilde{P}_{\max} \kappa^{0.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi) / \theta_{\min}) = O\left(\frac{K^3 \theta_{\max}^{14} \tilde{P}_{\max} \omega \kappa^{3.5}(\Pi'\Pi) \lambda_1^{2.5}(\Pi'\Pi)}{\theta_{\min}^{12} \pi_{\min}}\right). \end{aligned}$$

Recall that $J_* = N_U(\mathcal{I}, \mathcal{I}) \Theta(\mathcal{I}, \mathcal{I})$, we have $\|J_*\| \leq N_{U, \max} \theta_{\max} \leq \frac{\theta_{\max}^2 \sqrt{K \lambda_1(\Pi'\Pi)}}{\theta_{\min}}$ where the last inequality holds by Lemma A1. Similarly, we have

$J_*(k, k) \geq \theta_{\min} \min_{i \in [n]} \frac{1}{\|U(i, :)\|_F} \geq \frac{\theta_{\min}^2 \sqrt{\lambda_K(\Pi'\Pi)}}{\theta_{\max}}$ where the last inequality holds by the proof of Lemma 5. Then we have

$$\begin{aligned} \|J_* - \mathcal{P}_* \hat{J}_* \mathcal{P}'_*\| &= \max_{k \in [K]} |\hat{\tau}_k - \tau_k| = \max_{k \in [K]} \frac{|\hat{\tau}_k^2 - \tau_k^2|}{\hat{\tau}_k + \tau_k} \leq \max_{k \in [K]} \frac{|\hat{\tau}_k^2 - \tau_k^2|}{\tau_k} \\ &\leq \frac{\theta_{\max}}{\theta_{\min}^2 \sqrt{\lambda_K(\Pi'\Pi)}} \max_{k \in [K]} |\hat{\tau}_k^2 - \tau_k^2| = O\left(\frac{K^3 \theta_{\max}^{15} \tilde{P}_{\max} \omega \kappa^{3.5}(\Pi'\Pi) \lambda_1^{2.5}(\Pi'\Pi)}{\theta_{\min}^{14} \pi_{\min} \sqrt{\lambda_K(\Pi'\Pi)}}\right). \end{aligned}$$

Combining the above results, we have

$$\begin{aligned} \|e'_i (\hat{Z}_* - Z_* \mathcal{P}_*)\|_F &\leq \|e'_i (\hat{Y}_* - Y_* \mathcal{P}_*)\|_F \|\hat{J}_*\| + \|e'_i Y_*\|_F \|J_* - \mathcal{P}_* \hat{J}_* \mathcal{P}'_*\| \\ &\leq O\left(\frac{K^{4.5} \theta_{\max}^{14} \omega \kappa^{4.5}(\Pi'\Pi) \lambda_1(\Pi'\Pi)}{\theta_{\min}^{14} \pi_{\min}}\right) O(\theta_{\max} \sqrt{\tilde{P}_{\max} \lambda_1(\Pi'\Pi)}) \end{aligned}$$

$$\begin{aligned}
& + \frac{\theta_{\max}^2 \sqrt{K \lambda_1(\Pi' \Pi)}}{\theta_{\min}^2 \lambda_K(\Pi' \Pi)} O\left(\frac{K^3 \theta_{\max}^{15} \tilde{P}_{\max} \omega \kappa^{3.5} (\Pi' \Pi) \lambda_1^{2.5} (\Pi' \Pi)}{\theta_{\min}^{14} \pi_{\min} \sqrt{\lambda_K(\Pi' \Pi)}}\right) \\
& = O\left(\frac{\theta_{\max}^{15} K^{4.5} \omega \kappa^{4.5} (\Pi' \Pi) \lambda_1^{1.5} (\Pi' \Pi)}{\theta_{\min}^{14} \pi_{\min}}\right).
\end{aligned}$$

□

References

1. Watts, D.J.; Strogatz, S.H. Collective dynamics of ‘small-world’ networks. *Nature* **1998**, *393*, 440–442.
2. Newman, M.E. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **2001**, *64*, 016132.
3. Dunne, J.A.; Williams, R.J.; Martinez, N.D. Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12917–12922.
4. Newman, M.E.J. Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5200–5205.
5. Notebaart, R.A.; van Enckevort, F.H.; Francke, C.; Siezen, R.J.; Teusink, B. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinform.* **2006**, *7*, 296.
6. Pizzuti, C. Ga-net: A genetic algorithm for community detection in social networks. In *International Conference on Parallel Problem Solving from Nature*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1081–1090.
7. Jackson, M.O. *Social and Economic Networks*; Princeton University Press: Princeton, NJ, USA, 2010.
8. Gao, J.; Liang, F.; Fan, W.; Wang, C.; Sun, Y.; Han, J. On community outliers and their efficient detection in information networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 24–28 July 2010; pp. 813–822.
9. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **2010**, *52*, 1059–1069.
10. Su, G.; Kuchinsky, A.; Morris, J.H.; States, D.J.; Meng, F. GLay: Community structure analysis of biological networks. *Bioinformatics* **2010**, *26*, 3135–3137.
11. Lin, W.; Kong, X.; Yu, P.S.; Wu, Q.; Jia, Y.; Li, C. Community detection in incomplete information networks. In *Proceedings of the 21st International Conference on World Wide Web*, Lyon, France, 16–20 April 2012; pp. 341–350.
12. Scott, J.; Carrington, P.J. *The SAGE Handbook of Social Network Analysis*; SAGE Publications: London, UK, 2014.
13. Bedi, P.; Sharma, C. Community detection in social networks. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2016**, *6*, 115–135.
14. Ji, P.; Jin, J. Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* **2016**, *10*, 1779–1812.
15. Ji, P.; Jin, J.; Ke, Z.T.; Li, W. Co-citation and Co-authorship Networks of Statisticians. *J. Bus. Econ. Stat.* **2022**, *40*, 469–485.
16. Newman, M.E. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256.
17. Newman, M.E.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113.
18. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308.
19. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174.
20. Fortunato, S.; Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **2016**, *659*, 1–44.
21. Abbe, E.; Bandeira, A.S.; Hall, G. Exact Recovery in the Stochastic Block Model. *IEEE Trans. Inf. Theory* **2016**, *62*, 471–487.
22. Fortunato, S.; Newman, M.E. 20 years of network community detection. *Nat. Phys.* **2022**, 1–3. <https://doi.org/10.1038/s41567-022-01716-7>.
23. Goldenberg, A.; Zheng, A.X.; Fienberg, S.E.; Airoldi, E.M. A survey of statistical network models. *Found. Trends Mach. Learn.* **2010**, *2*, 129–233.
24. Holland, P.W.; Laskey, K.B.; Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **1983**, *5*, 109–137.
25. Snijders, T.A.; Nowicki, K. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classif.* **1997**, *14*, 75–100.
26. Rohe, K.; Chatterjee, S.; Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Stat.* **2011**, *39*, 1878–1915.
27. Choi, D.S.; Wolfe, P.J.; Airoldi, E.M. Stochastic blockmodels with a growing number of classes. *Biometrika* **2012**, *99*, 273–284.
28. Sussman, D.L.; Tang, M.; Fishkind, D.E.; Priebe, C.E. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Am. Stat. Assoc.* **2012**, *107*, 1119–1128.
29. Latouche, P.; Birmelé, E.; Ambroise, C. Model selection in overlapping stochastic block models. *Electron. J. Stat.* **2014**, *8*, 762–794.
30. Lei, J.; Rinaldo, A. Consistency of spectral clustering in stochastic block models. *Ann. Stat.* **2015**, *43*, 215–237.
31. Sarkar, P.; Bickel, P.J. Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Stat.* **2015**, *43*, 962–990.
32. Lyzinski, V.; Tang, M.; Athreya, A.; Park, Y.; Priebe, C.E. Community detection and classification in hierarchical stochastic blockmodels. *IEEE Trans. Netw. Sci. Eng.* **2016**, *4*, 13–26.
33. Valles-Catala, T.; Massucci, F.A.; Guimera, R.; Sales-Pardo, M. Multilayer stochastic block models reveal the multilayer structure of complex networks. *Phys. Rev. X* **2016**, *6*, 011036.
34. Lei, J. A goodness-of-fit test for stochastic block models. *Ann. Stat.* **2016**, *44*, 401–424.

35. Tabouy, T.; Barbillon, P.; Chiquet, J. Variational inference for stochastic block models from sampled data. *J. Am. Stat. Assoc.* **2020**, *115*, 455–466.
36. Airoldi, E.M.; Blei, D.M.; Fienberg, S.E.; Xing, E.P. Mixed Membership Stochastic Blockmodels. *J. Mach. Learn. Res.* **2008**, *9*, 1981–2014.
37. Wang, F.; Li, T.; Wang, X.; Zhu, S.; Ding, C. Community discovery using nonnegative matrix factorization. *Data Min. Knowl. Discov.* **2011**, *22*, 493–521.
38. Airoldi, E.M.; Wang, X.; Lin, X. Multi-way blockmodels for analyzing coordinated high-dimensional responses. *Ann. Appl. Stat.* **2013**, *7*, 2431–2457.
39. Panov, M.; Slavnov, K.; Ushakov, R. Consistent Estimation of Mixed Memberships with Successive Projections. In *International Conference on Complex Networks and Their Applications*; Springer: Cham, Switzerland, 2017; pp. 53–64.
40. Zhang, Y.; Levina, E.; Zhu, J. Detecting overlapping communities in networks using spectral methods. *SIAM J. Math. Data Sci.* **2020**, *2*, 265–283.
41. Jin, J.; Ke, Z.T.; Luo, S. Estimating network memberships by simplex vertex hunting. *arXiv* **2017**, arXiv:1708.07852.
42. Mao, X.; Sarkar, P.; Chakrabarti, D. On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations. In Proceedings of the 34th International Conference of Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 2324–2333.
43. Mao, X.; Sarkar, P.; Chakrabarti, D. Overlapping Clustering Models, and One (class) SVM to Bind Them All. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; Volume 31, pp. 2126–2136.
44. Mao, X.; Sarkar, P.; Chakrabarti, D. Estimating Mixed Memberships With Sharp Eigenvector Deviations. *J. Am. Stat. Assoc.* **2020**, *116*, 1928–1940.
45. Karrer, B.; Newman, M.E.J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **2011**, *83*, 16107.
46. Kaufmann, E.; Bonald, T.; Lelarge, M. A spectral algorithm with additive clustering for the recovery of overlapping communities in networks. *Theor. Comput. Sci.* **2017**, *742*, 3–26.
47. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416.
48. Qin, T.; Rohe, K. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3120–3128.
49. Joseph, A.; Yu, B. Impact of regularization on spectral clustering. *Ann. Stat.* **2016**, *44*, 1765–1791.
50. Jin, J. Fast community detection by SCORE. *Ann. Stat.* **2015**, *43*, 57–89.
51. Gillis, N.; Vavasis, S.A. Semidefinite Programming Based Preconditioning for More Robust Near-Separable Nonnegative Matrix Factorization. *SIAM J. Optim.* **2015**, *25*, 677–698.
52. Mossel, E.; Neeman, J.; Sly, A. Consistency thresholds for binary symmetric block models. *arXiv* **2014**, arXiv:1407.1591.
53. Abbe, E. Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **2017**, *18*, 6446–6531.
54. Hajek, B.; Wu, Y.; Xu, J. Achieving Exact Cluster Recovery Threshold via Semidefinite Programming: Extensions. *IEEE Trans. Inf. Theory* **2016**, *62*, 5918–5937.
55. Agarwal, N.; Bandeira, A.S.; Koiliaris, K.; Kolla, A. Multisection in the Stochastic Block Model using Semidefinite Programming. *arXiv* **2017**, arXiv:1507.02323.
56. Bandeira, A.S. Random Laplacian Matrices and Convex Relaxations. *Found. Comput. Math.* **2018**, *18*, 345–379.
57. Abbe, E.; Sandon, C. Community Detection in General Stochastic Block models: Fundamental Limits and Efficient Algorithms for Recovery. In Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, Berkeley, CA, USA, 17–20 October 2015; pp. 670–688.
58. Gao, C.; Ma, Z.; Zhang, A.Y.; Zhou, H.H. Achieving Optimal Misclassification Proportion in Stochastic Block Models. *J. Mach. Learn. Res.* **2017**, *18*, 1–45.
59. McSherry, F. Spectral partitioning of random graphs. In *Proceedings of the 2001 IEEE International Conference on Cluster Computing*, Newport Beach, CA, USA, 8–11 October 2001; pp. 529–537.
60. Newman, M.E. Assortative mixing in networks. *Phys. Rev. Lett.* **2002**, *89*, 208701.
61. Erdős, P.; Rényi, A. On the evolution of random graphs. In *The Structure and Dynamics of Networks*; Princeton University Press: Princeton, NJ, USA, 2011; pp. 38–82. <https://doi.org/doi:10.1515/9781400841356.38>.
62. Blum, A.; Hopcroft, J.; Kannan, R. *Foundations of Data Science*; Number 1; Cambridge University Press: Cambridge, UK, 2020; pp. 1–465.
63. Lei, L. Unified $\ell_{2 \rightarrow \infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv* **2019**, arXiv:1909.04798.
64. Chen, Y.; Chi, Y.; Fan, J.; Ma, C. Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **2021**, *14*, 566–806.
65. Cape, J.; Tang, M.; Priebe, C.E. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Stat.* **2019**, *47*, 2405–2439.
66. Abbe, E.; Fan, J.; Wang, K.; Zhong, Y. Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank. *Ann. Stat.* **2020**, *48*, 1452–1474.
67. Rohe, K.; Qin, T.; Yu, B. Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12679–12684.
68. Wang, Z.; Liang, Y.; Ji, P. Spectral Algorithms for Community Detection in Directed Networks. *J. Mach. Learn. Res.* **2020**, *21*, 1–45.

69. Cai, T.T.; Li, X. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Stat.* **2015**, *43*, 1027–1059.
70. Tropp, J.A. User-Friendly Tail Bounds for Sums of Random Matrices. *Found. Comput. Math.* **2012**, *12*, 389–434.
71. Zhou, Z.; Amini, A. Analysis of spectral clustering algorithms for community detection: the general bipartite setting. *J. Mach. Learn. Res.* **2019**, *20*, 1–47.
72. Zhao, Y.; Levina, E.; Zhu, J. Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Stat.* **2012**, *40*, 2266–2292.
73. Ghoshdastidar, D.; Dukkipati, A. Consistency of Spectral Partitioning of Uniform Hypergraphs under Planted Partition Model. In Proceedings of the Advances in Neural Information Processing Systems 27, Montreal, QC, Canada, 8–13 December 2014; Volume 27, pp. 397–405.
74. Ke, Z.T.; Shi, F.; Xia, D. Community Detection for Hypergraph Networks via Regularized Tensor Power Iteration. *arXiv* **2019**, arXiv:1909.06503.
75. Cole, S.; Zhu, Y. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra Its Appl.* **2020**, *593*, 45–73.
76. Bandeira, A.S.; van Handel, R. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **2016**, *44*, 2479–2506.
77. Cape, J. Orthogonal Procrustes and norm-dependent optimality. *Electron. J. Linear Algebra* **2020**, *36*, 158–168.