

Strong Consistency of Spectral Clustering for the Sparse Degree-Corrected Hypergraph Stochastic Block Model

Chong Deng, Xin-Jian Xu, and Shihui Ying, *Member, IEEE*

Abstract—We prove strong consistency of spectral clustering under the degree-corrected hypergraph stochastic block model in the sparse regime where the maximum expected hyperdegree is as small as $\Omega(\log n)$ with n denoting the number of nodes. We show that the basic spectral clustering without preprocessing or postprocessing is strongly consistent in an even wider range of the model parameters, in contrast to previous studies that either trim high-degree nodes or perform local refinement. At the heart of our analysis is the entry-wise eigenvector perturbation bound derived by the “leave-one-out” technique. To the best of our knowledge, this is the first entry-wise error bound for degree-corrected hypergraph models, resulting in the strong consistency for clustering non-uniform hypergraphs with heterogeneous hyperdegrees.

Index Terms—Hypergraph, stochastic block model, spectral clustering, consistency.

I. INTRODUCTION

COMMUNITY detection, also known as graph clustering, occupies a central position in modern network science. The goal is to partition nodes in a network into communities such that the nodes in the same community are more similar than those belonging to different communities. Among numerous algorithms designed for this purpose, the model-based algorithms stand out for their statistical guarantees. The stochastic block model (SBM) [1] is one of the well known generative models for random graphs with community structure. In this model, edges are independently generated with probabilities depending only on the community memberships of nodes, and thus the nodes in the same community have the same expected degrees. As a consequence, the SBM delivers a poor description of real networks that exhibit degree heterogeneity even within communities. To overcome this limitation, the degree-corrected stochastic block model (DCSBM) [2] has been proposed by introducing a set of node-specific parameters to allow any possible degree distribution. It shows that the consideration of the degree heterogeneity significantly improves the ability of the model to fit real networks.

For community detection in a generative model, i.e., recovering the hidden community assignments from a single

instance generated by the model, it is usually concerned with consistency of an algorithm:

- *strong consistency* (or exact recovery). Finding the true partition of all nodes (up to a permutation) with high probability;
- *weak consistency* (or almost exact recovery). Finding the true partition of all but a vanishing fraction of nodes (up to a permutation) with high probability.

The fundamental limits of community detection under the (DC)SBM have been widely studied [3], [4], [5], [6], and the performance of many clustering algorithms, such as spectral algorithms [7], [8], has also been investigated both theoretically and numerically. However, as graphs model pairwise relationships by edges connecting pairs of nodes, they fail to capture higher-order interactions, and this is where hypergraphs come into play. Different from graphs, edges in a hypergraph, called hyperedges, can connect any number of nodes. Due to this merit, community detection in hypergraphs, or hypergraph clustering, has received increasing attention recently. In particular, the SBM has been generalized to the hypergraph SBM (HSBM) [9], in which the probability of occurrence of a hyperedge depends on the community memberships of all nodes in it. Considering the degree heterogeneity, the DCSBM has also been generalized to the degree-corrected HSBM (DCHSBM). Yet, there are very few studies on the theoretical limits for community detection under the DCHSBM and the algorithms achieving the detectability thresholds.

As one of the most popular algorithms, spectral clustering has been adopted for graph clustering. In its basic form, nodes are first represented as points in a low-dimensional space based on the leading eigenvectors, and these points are then clustered using standard clustering algorithms such as k -means. To develop spectral algorithms for hypergraph clustering, several matrices have been considered including adjacency matrix [10], hypergraph Laplacian [11] and adjacency tensor [12]. All these algorithms are weakly consistent in the HSBM under certain conditions on the model parameters. As for strong consistency, Cole and Zhu [13] achieved exact recovery for dense hypergraphs but it is suboptimal in the sparse case. On the other hand, two-stage algorithms have been proposed [10], [14], [15]: first obtain a weakly consistent community assignment from the trimmed adjacency matrix or hypergraph Laplacian and then perform local refinement to guarantee strong consistency.

C. Deng is with the Department of Mathematics, Shanghai University, Shanghai 200444, China (e-mail: dengchong@shu.edu.cn).

X.-J. Xu is with Qianweichang College, Shanghai University, Shanghai 200444, China (e-mail: xinjxu@shu.edu.cn).

S. Ying is with the Department of Mathematics, Shanghai University, Shanghai 200444, China (e-mail: shying@shu.edu.cn).

For graph clustering, it has been proved that the basic spectral clustering without trimming or refinement is strongly consistent [16], [17]. Therefore, it is natural to ask a question: whether the basic hypergraph spectral clustering could achieve strong consistency or not? Answering this question requires a more refined entry-wise analysis of eigenvector perturbations. The goal of this paper is to prove strong consistency of the basic spectral clustering under the DCHSBM in the sparse regime where the maximum expected hyperdegree might be of order $\Omega(\log n)$.

A. Main contributions

Let A be the weighted adjacency matrix (Equation (1)) of a hypergraph generated by a DCHSBM (Definition 1) and $P = \mathbb{E}[A]$ be the population counterpart of A . Let \hat{U} and U be the matrices respectively formed by stacking the K leading eigenvectors of A and P , where K is the number of communities. As the matrix U contains all information about the true community assignment, the task becomes bounding the deviation between \hat{U} and U . While the well known Davis-Kahan theorem specifies an upper bound on $\|\hat{U}\hat{O} - U\|_F$ for some orthogonal matrix \hat{O} (this matrix is used to align \hat{U} and U ; see Equation (8) for its specific expression), it gives a loose bound on $\|\hat{U}\hat{O} - U\|_{2,\infty}$, hence no guarantee for node-wise behavior. We show that simply applying k -means on the row-normalized leading eigenvector matrix of the weighted adjacency matrix achieves strong consistency in an even wider range of parameters. We develop a sharp deviation bound on $\|A - P\|$ for non-uniform hypergraphs with general edge probabilities. As a byproduct, we also derive an upper bound on mis-clustered nodes incurred by an approximate k -means algorithm which leads to weak consistency of Algorithm 1 (see Section II-C). Via the leave-one-out analysis, we obtain an eigenvector perturbation bound in two-to-infinity norm which is the first result on node-wise error bounds for hypergraph models. With this bound, we also study strong consistency of Algorithm 2 (see Section II-C), which is simple in nature but will be extremely hard to analyze if one only has a bound on the deviation between \hat{U} and U in the Frobenius norm. Finally, we consider a special case of the model where the conditions for exact recovery will be expressed more clearly. To the best of our knowledge, our study gives the first strong consistency result for clustering non-uniform hypergraphs with heterogeneous hyperdegrees.

B. Related Work

To date, there are few studies on strong consistency for hypergraph stochastic block models. Kim et al [18] demonstrated that exact recovery shows a sharp phase-transition behavior for the uniform HSBM with two equal-sized and symmetric communities. They proposed a semidefinite programming algorithm which is strongly consistent in an order-wise optimal parameter regime. Ahn et al [10] investigated consistency of spectral clustering in weighted uniform HSBMs where the number of clusters K is constant. Cole and Zhu [13] proposed a spectral algorithm based on the hypergraph adjacency matrix and proves that the algorithm achieves exact recovery in the

dense uniform HSBM where $K = \Theta(\sqrt{n})$. Chien et al [14] showed that spectral algorithms with local refinement achieve the exact recovery criteria in the sparse HSBM. Zhang and Tan [15] studied fundamental limits of exact recovery in the general uniform HSBM and develop a two-stage algorithm that meets the achievability threshold.

In [17], the authors performed an entry-wise eigenvector analysis and proved that spectral clustering achieves the threshold of exact recovery in a graph SBM with two blocks. In our study, we adopt the idea of the leave-one-out technique. Compared with [17], our main contributions are as follows. First, we generalize the entry-wise eigenvector perturbation analysis for graphs to hypergraphs. This is non-trivial because the weighted adjacency matrix A violates the ‘‘row- and column-wise independence’’ assumption made in [17]. Based on the idea of the leave-one-out method, we introduce a set of suitably defined surrogate matrices $A^{(l)}$ where randomness contributed by hyperedges containing l is eliminated, and then resort to the matrix Bernstein inequality where sequential transformations and inequalities are performed to address the dependency across entries of A . Second, we develop a sharp deviation bound on $\|A - P\|$ for non-uniform hypergraphs. Finally, we derive the strong consistency result for spectral clustering in the non-uniform DCHSBM with multiple communities.

Gaudio and Joshi [19] derived an entry-wise bound on the second eigenvector of the adjacency matrix and prove strong consistency of spectral clustering for the uniform HSBM with two communities. Compared with [19], we consider different algorithms and have a different goal: we try to find under what conditions on degree heterogeneity, the number of communities, sparsity and the minimal non-zero eigenvalue of the matrix (See Lemma 1), spectral clustering without pre-processing and post-processing could achieve strong consistency. Gaudio and Joshi [19] instead aimed at proving their algorithm can achieve the theoretical threshold. While our settings are more general, our result does not achieve the theoretical limit in this special case. Both our work and Ref. [19] perform the entry-wise eigenvector analysis. The main difference is as follows. Ref. [19] obtains a more refined l_∞ eigenvector perturbation bound by generalizing Theorem 1.1 in Ref. [17] to the two-block HSBM case. We derive an entry-wise bound for deviation between \hat{U} and U in $l_{2,\infty}$ norm by applying the matrix Bernstein inequality together with the Davis-Kahan theorem and the triangle inequality, where we carefully handle the dependence across the entries of A . Our entry-wise bound makes it possible to guarantee strong consistency of the considered algorithms in the general K -block case, where K is even allowed to diverge to infinity at a slow rate.

Wang [20] established the information-theoretical threshold for strong consistency in non-uniform HSBM with two equal-sized communities. However, the algorithms considered in Ref. [20] are not applicable to the case of multiple blocks. Dumitriu and Wang [21] derived sharp threshold for exact recovery in non-uniform HSBM with multiple communities and provided multi-stage algorithms that successfully achieve exact recovery above the threshold. Compared with Ref. [21],

we consider strong consistency of spectral clustering without preprocessing or postprocessing and derive the first entry-wise eigenvector perturbation bound for non-uniform hypergraphs with heterogeneous hyperdegrees.

C. Paper Organization

We start in Section II with an introduction of the DCHSBM, the weighted adjacency matrix and two hypergraph clustering algorithms. We describe our main results as well as its proof in Section III-A. The consistency results in a special case is given in Section IV. Section VI contains concluding remarks.

D. Notations

Given a matrix X , we use X_i to refer the i -th row of X . Let $\|\cdot\|$ denote the spectral norm of a matrix or the l_2 norm of a vector and $\|\cdot\|_F$ denote the matrix Frobenius norm. Let $\|X\|_{2,\infty} = \max_i \|X_i\|$ be the two-to-infinity norm of X . For any vector $v \in \mathbb{R}^n$, let $v_{max} = \max_i v_i$, $v_{min} = \min_i v_i$ and $\text{diag}(v)$ be an $n \times n$ matrix with zero off-diagonal whose ii -th element is v_i . For any positive integer n , we use $[n] = \{1, 2, \dots, n\}$ to denote the set of positive integers not greater than n . Denote δ as the Kronecker function: for $i, j \in \mathbb{Z}$, $\delta_{i,j} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise. Throughout the paper, we use the standard asymptotic notations: $o(\cdot)$, $O(\cdot)$, $\Theta(\cdot)$, $\Omega(\cdot)$ and $\omega(\cdot)$.

II. PRELIMINARIES

A. The Model

Let n be the number of nodes in a hypergraph. The nodes are divided into K communities and each node i is assigned a label $g_i \in [K]$ representing the community to which it belongs. Additionally, each node i is assigned a parameter $\theta_i > 0$ to control its expected degree. Let $M \geq 2$ be the maximum hyperedge cardinality in the hypergraph. Let E be the hyperedge set of a complete hypergraph on $[n]$ with edge size between 2 and M . Each possible hyperedge $e \in E$ (hyperedge that contains duplicate nodes is allowed) is associated with an indicator variable h_e such that $h_e = 1$ if e is present and $h_e = 0$ otherwise.

Definition 1 (Degree-corrected hypergraph stochastic block model): In a DCHSBM, $\{h_e\}_{e \in E}$ are independent Bernoulli random variables satisfying

$$\mathbb{P}(h_e = 1) = b_e \pi(\theta_e) \Phi(g_e) \in [0, 1],$$

where $\pi(\theta_e) = \prod_{i \in e} \theta_i$ is the product of the hyperdegree parameters of the nodes in e , Φ is an affinity function that maps the group assignment to a non-negative real number and the coefficient b_e denotes the number of distinct ways to order the nodes of e .

For example, suppose $n = 8, K = 2, M = 5$ and $g = (1, 1, 1, 1, 1, 2, 2, 2)$. For three possible hyperedges $e_1 = (2, 6), e_2 = (1, 3, 8), e_3 = (1, 4, 4, 7)$, we have $b_{e_1} = 2, b_{e_2} = 6, b_{e_3} = 12$ and thus $\mathbb{P}(h_{e_1} = 1) = 2\theta_2\theta_6\Phi(1, 2), \mathbb{P}(h_{e_2} = 1) = 6\theta_1\theta_3\theta_8\Phi(1, 1, 2)$ and $\mathbb{P}(h_{e_3} = 1) = 12\theta_1\theta_4^2\theta_7\Phi(1, 1, 1, 2)$.

Let n_k denote the size of the k -th community. Without loss of generality, we assume $n_1 \geq n_2 \geq \dots \geq n_K$. The community memberships can also be represented by an assignment matrix $Z \in \{0, 1\}^{n \times K}$ where $Z_{ik} = 1$ if $g_i = k$ and $Z_{ik} = 0$ otherwise. For each $e \in E$, we define g_e as a vector of the cluster labels for nodes in e .

To ensure model identifiability, we impose the constraint that $\sum_i \theta_i \delta_{g_i, k} = n_k$ for each $k \in [K]$. Thus, the DCHSBM nests the HSBM as a special case by setting $\theta_i = 1$ for all $i \in [n]$. Similarly, we define θ_e as a vector of the hyperdegree parameters of the nodes in e .

Furthermore, we introduce vectors $a_e \in \mathbb{R}^n$, where a_{ei} is the number of occurrences of node i in e . The hyperdegree of node i is defined by $d_i = \sum_{e \in E} a_{ei} h_e$ and the cardinality of each possible hyperedge $e \in E$ is given by $|e| = \sum_{i=1}^n a_{ei}$.

A basic but important choice of Φ is the so-called ‘‘all-or-nothing’’ affinity function, where the hyperedge probability depends on whether all nodes in the hyperedge are in the same community or not.

Definition 2 (Degree-corrected hypergraph planted partition model): For each $e = \{i_1, \dots, i_m\} \in E$, it is generated with probability

$$\mathbb{P}(h_e = 1) = \alpha_m b_e \pi(\theta_e) \cdot ((p - q) \mathbb{1}_{\{g_{i_1} = \dots = g_{i_m}\}} + q),$$

where $p > q > 0$ are constants independent of n and $\alpha_m \geq 0$ is a scaling factor that varies with n and controls the number of the hyperedge cardinality m .

We investigate consistency of spectral algorithms for this model in Section IV as a case study and test the performance of the considered algorithms in this model in Section V.

B. The Weight Adjacency Matrix

Spectral algorithms for community detection in the model depend on the weighted adjacency matrix $A \in \mathbb{R}^{n \times n}$ which is non-negative symmetric with entries:

$$A_{ij} = \begin{cases} \sum_{e \in E} \frac{a_{ei} a_{ej}}{|e|-1} h_e, & \text{if } i \neq j, \\ \sum_{e \in E} \frac{a_{ei}(a_{ei}-1)}{|e|-1} h_e, & \text{if } i = j. \end{cases} \quad (1)$$

When there is no hyperedge containing repeated nodes, the diagonal entries of A are all zero. The so-defined weighted adjacency matrix corresponds to the adjacency matrix of a weighted undirected graph projected by a hypergraph [22]. In this work, we investigate consistency of the hypergraph spectral clustering based on the weighted adjacency matrix in the DCHSBM.

To understand why the spectrum of A contains information about the hidden community structure, we first take a close look at its expectation $P = \mathbb{E}[A]$. Denote the eigenvalues of P by $\{\lambda_i\}_{i=1}^n$ such that $|\lambda_1| \geq \dots \geq |\lambda_n|$ and let $\{u_i\}_{i=1}^n$ be the corresponding unit-norm eigenvectors. Define $U = (u_1, \dots, u_K) \in \mathbb{R}^{n \times K}$ and U^* the normalized matrix with $U_i^* = U_i / \|U_i\|$. The following lemma characterizes the eigen-structure of P .

Lemma 1 (The eigen-structure of P):

- (i). There exists a symmetric matrix $B \in \mathbb{R}^{K \times K}$ such that

$$P = \text{diag}(\theta) Z B Z^T \text{diag}(\theta). \quad (2)$$

- (ii). If B is full rank, then there exists an orthogonal matrix $Q \in \mathbb{R}^{K \times K}$ such that for any $i \in [n]$,

$$U_i = \tilde{\theta}_i Q_{g_i}, \quad (3)$$

where $\tilde{\theta}_i = \theta_i / \phi_{g_i}$ and $\phi_k = \sqrt{\sum_{j=1}^n \theta_j^2 \delta_{g_j, k}}$ for $k \in [K]$.

As an immediate consequence of (2), we have $\text{rank}(P) \leq K$, which means $\lambda_{K+1} = \dots = \lambda_n = 0$. Readers familiar with spectral clustering in graphs will find that P could be regarded as the population adjacency matrix of an ordinary DCSBM parameterized by g, θ and B , which could be viewed as a ‘‘projected model’’. It is known that graph projection may cause the community structure unidentifiable under some parameter space [11], [15]. Therefore, the condition that matrix B is full rank, i.e., $|\lambda_K| > 0$, is crucial for the success of the graph-projection-based clustering algorithms.

According to (3), for any two nodes i and j , if they belong to the same community, i.e., $g_i = g_j$, then U_i and U_j point to the same direction in \mathbb{R}^K ; otherwise, U_i and U_j are orthogonal to each other. By normalizing the rows of U to have unit length, we have

$$\|U_i^* - U_j^*\| = \sqrt{2} \cdot \mathbb{1}_{\{g_i \neq g_j\}}.$$

Thus, the community memberships of all nodes will be exactly recovered from U^* without difficulty.

In practice, P is not observed and spectral clustering is applied to the noisy observation A . Thus, if \hat{U}^* is close enough to U^* , we can expect that the spectral algorithms could still successfully recover the true community memberships. However, bounding the deviation of \hat{U}^* from U^* (especially entry-wise) turns out to be a non-trivial task once the entries of A are no longer independent. More care must be taken when dealing with the complex dependency across entries.

C. Spectral Algorithms

Denote the K leading eigenvalues of A by $\{\hat{\lambda}_k\}_{k=1}^K$ such that $|\hat{\lambda}_1| \geq \dots \geq |\hat{\lambda}_K|$ and let $\{\hat{u}_k\}_{k=1}^K$ be the corresponding unit-norm eigenvectors. Let $\hat{U} \in \mathbb{R}^{n \times K}$ be the matrix that contains the K leading eigenvectors of A as columns and $\hat{U}^* \in \mathbb{R}^{n \times K}$ be the row-normalized version of \hat{U} , i.e., $\hat{U}_i^* = \hat{U}_i / \|\hat{U}_i\|$ for all $i \in [n]$.

The first algorithm we consider is the classical spectral clustering summarized in Algorithm 1. For strong consistency of the k -means step, we utilize the result of Theorem 2.3 in Ref. [16]. The second one is a simple thresholding algorithm listed in Algorithm 2. Two nodes will be assigned to the same community when the distance between the corresponding rows of \hat{U}^* is small enough. Concretely, we try to find a threshold τ such that, starting from an empty graph $G = ([n], \emptyset)$, after connecting all node pairs (i, j) satisfying $\|\hat{U}_i^* - \hat{U}_j^*\| < \tau$, G has exactly K connected components, which corresponds to the true K communities. Initially there are n connected components, each of which contains only one node. After adding an edge, the number of components either decreases by one or remains the same. If we connect all node pairs, there will be only one connected component. Therefore, Algorithm

2 can successfully output a partition of $[n]$ into K communities and the output is unique.

Algorithm 1 Hypergraph spectral clustering with k -means

Input: The hypergraph \mathcal{H} and the number of communities K .

- 1: Construct the weighted adjacency matrix A and compute \hat{U}^* .
- 2: Run k -means algorithm with $k = K$ on the rows of \hat{U}^* .

Output: A community assignment \hat{g} where \hat{g}_i is the cluster index of i -th row.

Algorithm 2 Hypergraph spectral clustering via thresholding

Input: The hypergraph \mathcal{H} and the number of communities K .

- 1: Construct the weighted adjacency matrix A and compute \hat{U}^* .
- 2: Let G be a simple graph with node set $[n]$ and an empty edge set.
- 3: Sort node pairs $\{(i, j) | i, j \in [n]\}$ in ascending order by $\|\hat{U}_i^* - \hat{U}_j^*\|$.
- 4: Add edges to node pairs in G in the above order until G has exactly K connected components.

Output: A community assignment \tilde{g} where \tilde{g}_i is the index of the component to which i belongs in G .

For both algorithms, the successful recovery of community labels of all nodes depends strongly on \hat{U}^* and U^* being sufficiently close entry-wise. When proving strong consistency of Algorithm 2, we show that with high probability: if $|\hat{U}_i^* - \hat{U}_j^*| < 1/\sqrt{2}$ then $g_i = g_j$, otherwise $g_i \neq g_j$. Here we comment on the threshold $1/\sqrt{2}$, which could be replaced by any constant $\tau \in (0, \sqrt{2})$. For any nodes i and j , by the triangle inequality, we have $\|\hat{U}_i^* - \hat{U}_j^*\| \leq 2\|\hat{U}^* \hat{O} - U^*\|_{2, \infty}$ if $g_i = g_j$ and $\|\hat{U}_i^* - \hat{U}_j^*\| \geq \sqrt{2} - 2\|\hat{U}^* \hat{O} - U^*\|_{2, \infty}$ otherwise. Since both bounds are tight in the worst case, a sufficient and necessary condition to ensure strong consistency of Algorithm 2 is

$$\|\hat{U}^* \hat{O} - U^*\|_{2, \infty} < \frac{1}{2} \min\{\tau, \sqrt{2} - \tau\}. \quad (4)$$

Because $1/\sqrt{2}$ maximizes the right-hand side of (4), it is optimal in the sense that it imposes the mildest requirement on $\|\hat{U}^* \hat{O} - U^*\|_{2, \infty}$.

III. CONSISTENCY OF HYPERGRAPH SPECTRAL CLUSTERING

A. The Main Result

We show that under mild conditions on the model parameters, both algorithms mentioned above are strongly consistent. Let g' be an estimator of g that partitions nodes into K communities. The number of mis-clustered nodes of g' is defined by

$$l(g, g') = \min_{\sigma \in \mathcal{S}_K} \sum_{i=1}^n \mathbb{1}_{\{g_i \neq \sigma(g'_i)\}},$$

where \mathcal{S}_K is the K -th order symmetric group and the minimum is taken over all possible permutation of $[K]$. We say

g' is weakly consistent if $l(g, g') = o(n)$ with probability $1 - o(1)$, which only needs the fraction of mis-clustered nodes to vanish in the large n limit. In contrast, strong consistency requires $l(g, g') = 0$ with probability $1 - o(1)$, which means that the community membership of all nodes should be exactly identified in large samples. For notational simplicity, we write “ $g' = g$ with high probability” in short.

To ensure strong consistency of spectral algorithms, we make the following assumptions about the model parameters.

Assumption 1: $M = O(1)$.

That is, M is not allowed to vary with n . A constant maximum hyperedge cardinality is necessary for both the sharp upper bound on $\|A - P\|$ and the bound on the node-wise deviation $\|\hat{U}\hat{O} - U\|_{2,\infty}$ to hold. This is a mild condition for M since in practice M is usually much smaller than n .

Assumption 2: $n_1/n_K = O(1)$.

In other words, the nodes form communities of fairly balanced size, which is a standard assumption for strong consistency [10], [15], [16].

Assumption 3: $\kappa = |\lambda_1|/|\lambda_K| = O(\sqrt{\log n})$.

This is a necessary condition for Theorem 3 in Section III-B. In the literature, κ is usually assumed to be bounded from above by a constant [23]. A weaker requirement is made here thanks to the sharp deviation bound on $\|A - P\|$.

Assumption 4: B is full rank, or equivalently, $\lambda_K \neq 0$.

This assumption is in fact a critical condition for the identifiability of the partitions. Since the hypergraph is projected to a weighted graph, a large eigen-gap, i.e., a large value of $|\lambda_K|$, ensures that the partition is identifiable and thus could be successfully extracted. It is unclear whether Algorithms 1 and 2 are still strongly consistent under milder conditions on model parameters (e.g., B contains distinct rows).

Let $d = \max\{n \max_{ij} P_{ij}, c_0 \log n\}$ be an upper bound of the expected node hyperdegrees for some constant $c_0 > 0$. Define

$$\gamma = \frac{\max_{i \in [n]} \|U_i\|}{\min_{i \in [n]} \|U_i\|} = \frac{\tilde{\theta}_{max}}{\tilde{\theta}_{min}}. \quad (5)$$

γ depends only on the model parameter θ and specifies an upper bound on $\|U_i\|/\|U_j\|$, which is useful when characterizing the upper bound for clustering error rate. For a HSBM, we have $\gamma = \sqrt{n_1/n_K} = O(1)$ due to Assumption 2, which corresponds to the absence of degree heterogeneity in the HSBM. In general, the stronger the degree heterogeneity is, the larger γ is and the harder for clustering algorithms to achieve exact recovery.

The main result is given by the following theorem.

Theorem 1: Let A be a weighted adjacency matrix of a hypergraph generated by a DCHSBM that Assumptions 1-4 hold.

- (i). There exists a constant $C_1 = C_1(M, c_0)$ such that if n is sufficiently large¹ and

$$\frac{\gamma K^{3/2} \sqrt{d \log n}}{|\lambda_K|} \leq C_1, \quad (6)$$

then $\hat{g} = g$ with high probability.

¹Note that the number of nodes n being sufficiently large is a necessary condition for the success of the k -means algorithm [11], [16].

- (ii). There exists a constant $C_2 = C_2(M, c_0)$ such that if

$$\frac{\gamma \sqrt{d \log n}}{|\lambda_K|} \leq C_2, \quad (7)$$

then $\tilde{g} = g$ with high probability.

Though the condition (7) is milder than (6), Algorithm 1 may have better theoretical and/or practical performance than Algorithm 2 since Theorem 1 only provides sufficient conditions for the considered algorithms to achieve exact recovery. Note that conditions (6) and (7) do not directly reveal the fact that strong consistency should be more achievable when a hypergraph gets denser since both d and λ_K are affected by the sparsity of the hypergraph. To better demonstrate this, we consider the consistency of \hat{g} and \tilde{g} in the m -uniform HPPM, which has been widely studied in the literature [10], [13], [18]. In the m -uniform HPPM, all hyperedges are of size m and the generation probability of $e = \{i_1, \dots, i_m\}$ is $\mathbb{P}(e) = \alpha_m b_e ((p - q) \cdot \mathbb{1}_{\{g_{i_1} = \dots = g_{i_m}\}} + q)$.

Corollary 1: Let A be a weighted adjacency matrix of a hypergraph generated by the m -uniform HPPM, then Assumption 4 holds. If

$$K = O((\log n)^{\frac{1}{2m-2}}),$$

then Assumption 3 holds. When Assumptions 1-4 all hold, we have the following conclusions:

- (i). There exists a constant $C_1 > 0$ such that for sufficiently large n , if

$$\alpha_m \geq C_1 \frac{K^{2m+1} \log n}{n^{m-1}},$$

then $\hat{g} = g$ with high probability.

- (ii). There exists a constant $C_2 > 0$ such that if

$$\alpha_m \geq C_2 \frac{K^{2m-2} \log n}{n^{m-1}},$$

then $\tilde{g} = g$ with high probability.

This corollary is a special case of Corollary 2 in Section IV-A. The proof could be found in Appendix A-C. We compare Algorithms 1 and 2 with existing strongly consistent algorithms in Table I. While [10], [14], [15] require the number of communities K to be a constant, [13] allows K to grow like $O(\sqrt{\log n})$ which is much faster than ours, but that algorithm only works for dense hypergraphs. In contrast, we allow K to diverge to infinity at a slow rate in the sparse regime. When K is a constant, the sparsity required by Algorithms 1 and 2 meets the theoretical limit $\Omega(\log n/n^{m-1})$.

B. Proof of the main result

The outline of the proof is as follows. We first derive a sharp bound on the deviation $\|A - P\|$ by the combinatorial technique [8]. Then, we obtain an upper bound on $\|\hat{U}\hat{O} - U\|_{2,\infty}$ using the leave-one-out technique [17]. Finally, we bound $\|\hat{U}^*\hat{O} - U^*\|_{2,\infty}$ and analyze the performance of Algorithms 1 and 2.

To derive a tight bound on $\|A - P\|$, we adopt the Kahn-Szemerédi argument [24], which has been used to bound the second largest eigenvalue of ER graphs [25] and the spectral norm of general binary symmetric random matrices [8].

TABLE I
COMPARISON OF STRONGLY CONSISTENT ALGORITHMS

Paper	Algorithm type	m	K	α_m	Sizes of blocks
[18]	SDP	$O(1)$	2	$\Omega(\log n / \binom{n-1}{m-1})$	Equal
[10]	Spectral+LR	$O(1)$	$O(1)$	$\Omega(n \log n / \binom{n}{m})$	Almost equal
[14]	Spectral+LR	$O(1)$	$O(1)$	$\Omega(\log n / n^{m-1})$	Almost equal
[13]	Spectral	$O(1)$	$O(\sqrt{n})$	$\Theta(1)$	Equal
[15]	Spectral+LR	$O(1)$	$O(1)$	$\Theta(\log n / n^{m-1})$	Almost equal
Algorithm 1	Spectral	$O(1)$	$O((\log n)^{\frac{1}{2m-2}})$	$\Omega(\frac{K^{2m+1} \log n}{n^{m-1}})$	Almost equal
Algorithm 2	Spectral	$O(1)$	$O((\log n)^{\frac{1}{2m-2}})$	$\Omega(\frac{K^{2m-2} \log n}{n^{m-1}})$	Almost equal

Theorem 2: Let Assumption 1 hold. For any $r > 0$, there exists a constant $C = C(M, c_0, r)$ such that

$$\|A - P\| \leq C\sqrt{d}$$

with probability at least $1 - O(n^{-r})$.

It should be mentioned that [10] first extended the above techniques to hypergraphs and obtained a concentration bound for the weighted uniform hypergraph stochastic block model in which the edge weight has binary expected value. On the contrary, the model we consider here is non-uniform and the edge probabilities are more general. In this scenario, the DCHSBM allows variation of density of edges of different cardinality. Since we focus on sparse hypergraphs, the hyperdegrees of nodes, which could be as small as $\Omega(\log n)$, will be regarded as a summary of the sparsity of edges of different cardinalities.

As a byproduct, we establish the weak consistency of Algorithm 1 based on Theorem 2. The classical way to derive the error bound of Algorithm 1 consists of three steps: (i) Bound $\|A - P\|$; (ii) Bound $\|\hat{U}\hat{Q} - U\|_F$ and $\|\hat{U}^*\hat{Q} - U^*\|_F$ for some orthogonal matrix $\hat{Q} \in \mathbb{R}^{K \times K}$; and (iii) Bound the error incurred by an $(1 + \epsilon)$ -approximate k -means algorithm [26]. Combining Theorem 2, Lemmas 5.1 and 5.3 in [8], we have the following lemma.

Lemma 2: Let Assumption 1 hold. Suppose an $(1 + \epsilon)$ -approximate k -means algorithm is used in Algorithm 1 for a constant $\epsilon > 0$. There exists a constant $c > 0$ such that if

$$\frac{Kd}{n_K \lambda_K^2 \tilde{\theta}_{min}^2} < c,$$

then

$$l(g, \hat{g}) = O\left(\frac{Kd}{\lambda_K^2 \tilde{\theta}_{min}^2}\right)$$

with high probability.

Again we consider the m -uniform HPPM with Assumptions 1 and 2 being held. Algorithm 1 is weakly consistent when $K = O(n^{\frac{m-1}{2m-1}})$ and $\alpha_m = \omega(\frac{K^{2m-1}}{n^{m-1}})$, which is a weaker condition than that of [11].

Remark 1: As reported in [11], for dense enough hypergraphs, one has $l(g, \hat{g}) = O(1)$, which also implies strong consistency. Taking the m -uniform HPPM where Assumptions 1 and 2 hold for an example, one has $l(g, \hat{g}) = O(1)$ whenever $m \geq 3$, $K = o(n^{\frac{m-2}{2m-2}})$ and $\alpha_m = \omega(\frac{K^{2m-2}}{n^{m-2}})$. In this case, it follows that $d = \omega(K^{2m-2}n)$, which means a much denser hypergraph than that implied by Theorem 1.

By the Davis-Kahan theorem, one can bound $\|\hat{U}^*\hat{O} - U^*\|_F$. However, it will lead to a trivial bound on $\|\hat{U}^*\hat{O} - U^*\|_{2,\infty}$, and thus could not guarantee strong consistency. As an alternative, we turn to more refined analysis of the perturbation of eigenvectors. For this purpose, we study the entry-wise eigenvector deviation between \hat{U} and U . Due to the possibility of the presence of identical eigenvalues of A (P), \hat{U} (U) may not be uniquely determined. Therefore, a $K \times K$ orthogonal matrix is involved to align \hat{U} and U .

Let $H = \hat{U}^T U$ and its singular value decomposition be $H = \bar{U} \bar{\Sigma} \bar{V}^T$, then the matrix sign function [27] is a $K \times K$ orthogonal matrix given by

$$\text{sgn}(H) = \bar{U} \bar{V}^T. \quad (8)$$

We have the following theorem specifying an upper bound on $\|\hat{U} \text{sgn}(H) - U\|_{2,\infty}$.

Theorem 3: Let Assumptions 1 and 3 hold. For any $r > 0$, there exist some constants $C_3 = C_3(M, c_0, r)$ and $C_4 = C_4(M, c_0, r)$ such that if $\sqrt{d} \log n \leq C_3 |\lambda_K|$, then

$$\|\hat{U} \text{sgn}(H) - U\|_{2,\infty} \leq C_4 \frac{\sqrt{d} \log n}{|\lambda_K|} \|U\|_{2,\infty}$$

with probability at least $1 - O(n^{-r})$.

One can find the proof in Appendix C, which is based on the leave-one-out technique used to study the entry-wise eigenvector deviation for graphs where the adjacency matrix has independent entries [17]. The main challenge here stems from the dependency across entries of A .

To derive the bound on $\|\hat{U} \text{sgn}(H) - U\|_{2,\infty}$, a key step is to bound

$$\|(A - P)(\hat{U}H - U)\|_{2,\infty} = \max_{l \in [n]} \|(A - P)_l \cdot (\hat{U}H - U)\|, \quad (9)$$

which is challenging because of the dependence between A_l and $\hat{U}H$. We employ the leave-one-out method to tackle this problem.

For each $l \in [n]$, we define a matrix $A^{(l)} \in \mathbb{R}^{n \times n}$ as

$$A_{ij}^{(l)} = \sum_{e \in E: a_{el}=0} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} h_e + \sum_{e \in E: a_{el}>0} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} \mathbb{E}[h_e], \quad (10)$$

where all the randomness contributed by the possible hyperedges containing node l is eliminated. Denote the eigenvalues of $A^{(l)}$ by $\{\hat{\lambda}_i^{(l)}\}_{i=1}^n$, which are arranged in decreasing order of absolute value. Let $\hat{U}^{(l)}$ be a matrix that contains the K leading eigenvectors of $A^{(l)}$ as columns and let $H^{(l)} = (\hat{U}^{(l)})^T U$. We have the following observations:

- $\mathbb{E}[A^{(l)}] = P$ and $A^{(l)}$ concentrates around P . As a consequence, $\hat{U}^{(l)}$ and $H^{(l)}$ should be close to \hat{U} and H , respectively;
- A_l and $\hat{U}^{(l)}$ are independent, which implies the independence between A_l and $\hat{U}^{(l)}H^{(l)}$;
- $A - A^{(l)}$ and $\hat{U}^{(l)}$ are independent.

Now we can bound (9) as

$$\begin{aligned} & \|(A - P)_l(\hat{U}H - U)\| \\ & \leq \|(A - P)_l(\hat{U}H - \hat{U}^{(l)}H^{(l)})\| + \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\| \\ & \leq \|A - P\| \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| + \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\|. \end{aligned}$$

Applying the Davis-Kahan theorem yields

$$\|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| \leq \frac{2\|(A - A^{(l)})\hat{U}^{(l)}\|}{|\lambda_K|}.$$

Therefore, we are able to apply the matrix Bernstein inequality [28, Theorem 1.6.2] to bound $\|(A - A^{(l)})\hat{U}^{(l)}\|$ and $\|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\|$. See the proof for more details.

Proof [Proof of Theorem 1]: By Theorem 3, there are constants $c_1, c_2 > 0$ such that if $\sqrt{d \log n}/|\lambda_K| \leq c_1$, then

$$\|\hat{U} \text{sgn}(H) - U\|_{2,\infty} \leq c_2 \frac{\sqrt{d \log n}}{|\lambda_K|} \|U\|_{2,\infty}$$

with probability at least $1 - O(n^{-3})$. We first let $C_1, C_2 < 1/c_2$ and have

$$\begin{aligned} \|\hat{U}_i\| & \geq \|U_i\| - \|\hat{U}_i \text{sgn}(H) - U_i\| \\ & \geq \min_{i \in [n]} \|U_i\| - \|\hat{U} \text{sgn}(H) - U\|_{2,\infty} \\ & \geq \min_{i \in [n]} \|U_i\| \left(1 - c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}\right) > 0, \end{aligned}$$

with probability at least $1 - O(n^{-3})$. For any two non-zero vectors $x, y \in \mathbb{R}^K$, we have

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq 2 \frac{\|x - y\|}{\|y\|},$$

which yields

$$\begin{aligned} \|\hat{U}_i^* \text{sgn}(H) - U_i^*\| & = \left\| \frac{\hat{U}_i}{\|\hat{U}_i\|} \text{sgn}(H) - \frac{U_i}{\|U_i\|} \right\| \\ & \leq 2 \frac{\|\hat{U}_i \text{sgn}(H) - U_i\|}{\|U_i\|} \\ & \leq 2 \frac{\|\hat{U} \text{sgn}(H) - U\|_{2,\infty}}{\min_{i \in [n]} \|U_i\|} \\ & \leq 2c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}. \end{aligned}$$

Thus, we obtain

$$\|\hat{U}^* \text{sgn}(H) - U^*\|_{2,\infty} \leq 2c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}$$

with probability exceeding $1 - O(n^{-2})$.

For strong consistency of the k -means step of Algorithm 1, we use the result of Theorem 2.3 in [16]. Notice that the K distinct rows of U^* are orthogonal unit-length vectors, one

can choose the two deterministic sequences as $c_{1n} = 1$ and $c_{2n} = 2c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}$. Since

$$\sum_{n=1}^{\infty} \mathbb{P}(\|\hat{U}^* \text{sgn}(H) - U^*\|_{2,\infty} \geq c_{2n}) < \infty,$$

then $\|\hat{U}^* \text{sgn}(H) - U^*\|_{2,\infty} \leq c_{2n}$ almost surely. Under the current settings, a sufficient condition for Assumption 4.3 in [16] is $c_{2n} \leq c_3/(260K^{3/2})$, where the constant $c_3 \in (0, 1)$ is a lower bound on Kn_K/n that does exist due to Assumption 2. The result follows by choosing $C_1 = \min\{c_1, c_3/(520c_2)\}$.

Next, we consider the consistency of \hat{g} . When nodes i and j belong to the same community, we have

$$\begin{aligned} \|\hat{U}_i^* - \hat{U}_j^*\| & = \|(\hat{U}_i^* - \hat{U}_j^*) \text{sgn}(H)\| \\ & \leq \|\hat{U}_i^* \text{sgn}(H) - U_i^*\| + \|\hat{U}_j^* \text{sgn}(H) - U_j^*\| \\ & \leq 4c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}. \end{aligned}$$

While they belong to different communities, it follows that

$$\begin{aligned} \|\hat{U}_i^* - \hat{U}_j^*\| & = \|(\hat{U}_i^* - \hat{U}_j^*) \text{sgn}(H)\| \\ & \geq \|U_i^* - U_j^*\| - \|\hat{U}_i^* \text{sgn}(H) - U_i^*\| - \|\hat{U}_j^* \text{sgn}(H) - U_j^*\| \\ & \geq \sqrt{2} - 4c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|}. \end{aligned}$$

Therefore, Algorithm 2 exactly recovers the true community structure whenever

$$4c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|} < 1/\sqrt{2} \text{ and } \sqrt{2} - 4c_2 \frac{\gamma \sqrt{d \log n}}{|\lambda_K|} \geq 1/\sqrt{2},$$

which are satisfied if $C_2 < \min\{c_1, 1/(4\sqrt{2}c_2)\}$. ■

IV. CONSISTENCY FOR THE DCHPPM

A. The Uniform Case

Corollary 2: Let A be a weighted adjacency matrix of a hypergraph generated by an m -uniform DCHPPM, then Assumption 4 holds. If

$$K^{m-1} \phi_{max}^2 / \phi_{min}^2 = O(\sqrt{\log n}),$$

then Assumption 3 holds. Suppose Assumptions 1-4 hold.

- (i). There exists a constant $C_1 > 0$ such that for sufficiently large n , if

$$\alpha_m \geq C_1 \frac{\gamma^2 \theta_{max}^2 K^{2m+1} \log n}{n^{m-1}}, \quad (11)$$

then $\hat{g} = g$ with high probability.

- (ii). There exists a constant $C_2 > 0$ such that if

$$\alpha_m \geq C_2 \frac{\gamma^2 \theta_{max}^2 K^{2m-2} \log n}{n^{m-1}}, \quad (12)$$

then $\tilde{g} = g$ with high probability.

B. The Non-uniform Case

Corollary 3: Let A be a weighted adjacency matrix of a hypergraph generated by the above model, then Assumption 4 holds. Let $m_0 = \max\{m \in \mathbb{Z} | 2 \leq m \leq M, \alpha_m > 0\}$. If

$$K^{m_0-1} \phi_{max}^2 / \phi_{min}^2 = O(\sqrt{\log n}),$$

then Assumption 3 holds. Suppose Assumptions 1-4 hold.

- (i). There exists a constant $C_1 > 0$ such that for sufficiently large n , if

$$\sum_{m=2}^M m \alpha_m n^{m-1} \geq C_1 \gamma^2 \theta_{max}^2 K^{2m_0+1} \log n, \quad (13)$$

then $\hat{g} = g$ with high probability.

- (ii). There exists a constant $C_2 > 0$ such that if

$$\sum_{m=2}^M m \alpha_m n^{m-1} \geq C_2 \gamma^2 \theta_{max}^2 K^{2m_0-2} \log n, \quad (14)$$

then $\tilde{g} = g$ with high probability.

Remark 2: As the number of hyperedges of size m is of order $\Theta(\alpha_m n^m)$, (13) essentially requires that the total number of hyperedges is $\Omega(\gamma^2 \theta_{max}^2 K^{2m_0+1} n \log n)$, which reduces to $\Omega(n \log n)$ when $\theta_i = 1$ for all $i \in [n]$ and K is a constant.

V. NUMERICAL EXPERIMENT

In this section, we test the practical performance of Algorithms 1 and 2 in DCHPPMs. We set $n = 3000, M = 4, p = 10, q = 1$ and vary K, θ , and the averaged expected degree. We choose K from $\{2, 3\}$ and let the communities have the same size in both cases. We consider two choices of θ : (i) $\theta_i^{(1)} = 1$ for all $i \in [n]$; and (ii) first draw ψ_i uniformly and independently from $[1, 2]$, and then set $\theta_i^{(2)} = n \psi_i / (K \sum_{j \in [n]} \psi_j \delta_{g_i, g_j})$. We choose α_m such that the number of hyperedges of different sizes is the same. By multiplying all α_m by the same number, we are able to vary sparsity of the hypergraph. In implementation of Algorithm 1, we use MATLAB “kmeans” algorithm.

The clustering accuracy measured by an error rate $l(g, g')/n$ is shown in Figure 1. For both Algorithms 1 and 2, larger number of communities or stronger degree heterogeneity will make it harder to achieve exact recovery. As the hypergraph gets denser, both algorithms are able to fully recover the true community assignment. In the experiment, Algorithm 1 consistently outperforms Algorithms 2, which demonstrates the superiority of the k -means algorithm over the simple thresholding method.

VI. CONCLUSION

This paper characterizes the performance of spectral clustering based on the weighted adjacency matrix in a sparse DCHSBM. We consider both heterogeneous hyperdegrees and general edge-connecting probability. By establishing an entry-wise eigenvector perturbation bound, we show that even the simplest spectral clustering algorithm could exactly recover the true community structure under mild conditions on the model parameters, hence an important advancement over the

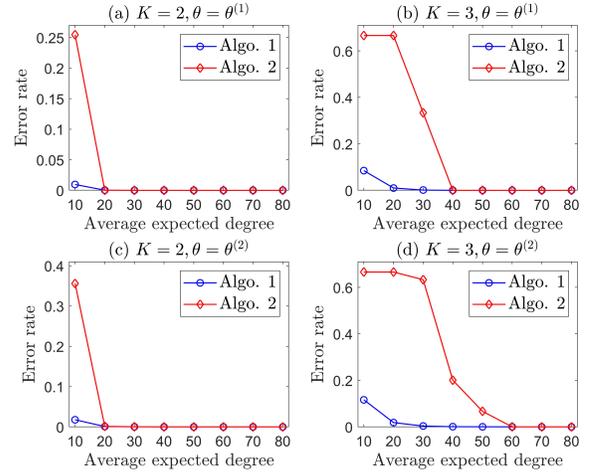


Fig. 1. Clustering result for Algorithms 1 and 2 in the DCHPPM. Each value is the average over 10 independent trials. The first and second rows show the results corresponding to $\theta = \theta^{(1)}$ and $\theta^{(2)}$, respectively. The first and second columns are the results corresponding to $K = 2$ and 3 , respectively.

literature. We suggest two further extensions. As spectral clustering assumes that the number of communities K is known, an important question would be the estimation of K . In this paper, we allow K to grow with n , but at a very slow rate to ensure that Assumption 3 holds. One possible way to improve this is to use the method developed in Ref. [29] to establish the entry-wise bound. However, this may fail when the maximum expected degree grows as $\Omega(\log n)$, the most challenging regime. On the other hand, a recent study [15] characterized the theoretical limit for exact recovery in a general uniform HSBM by the GCH-divergence. While the success of hypergraph spectral clustering depends on a large eigen-gap, it is important to explore and interpret the relation between λ_K and the GCH-divergence. The proposed spectral algorithm, relying on local refinement to be strongly consistent, is shown to meet the exact recovery threshold. It is worth investigating whether the spectral algorithms considered in this work achieve the theoretical limit. If not, can the threshold be achieved by adding a local refinement step [15]? Since hypergraph projection itself may incur information loss which makes it hard to achieve the limit, a possible approach is to consider tensor-based methods.

APPENDIX A

PROOF OF LEMMAS AND COROLLARIES

A. Proof of Lemma 1

- (i) For $i \neq j$, consider a hyperedge e such that $a_{ei} \geq 1$ and $a_{ej} \geq 1$, one has $b_e = \frac{|e|!}{\prod_{k=1}^n a_{ek}!}$ and $b_{e \setminus \{i,j\}} = \frac{(|e|-2)!}{(\prod_{k \neq i,j} a_{ek}!) \cdot (a_{ei}-1)! \cdot (a_{ej}-1)!} = \frac{a_{ei} a_{ej}}{|e|(|e|-1)} b_e$, and thus

$$A_{ij} = \sum_{e \in E: a_{ei} \geq 1, a_{ej} \geq 1} |e| \frac{b_{e \setminus \{i,j\}}}{b_e} h_e.$$

Then

$$\begin{aligned}
 P_{ij} &= \sum_{e \in E: a_{ei} \geq 1, a_{ej} \geq 1} |e| \frac{b_{e \setminus \{i,j\}}}{b_e} \mathbb{E}[h_e] \\
 &= \sum_{m=2}^M m \sum_{i_3=1}^n \cdots \sum_{i_m=1}^n \theta_i \theta_j \prod_{l=3}^m \theta_{i_l} \cdot \Phi(g_i, g_j, g_{i_3}, \dots, g_{i_m}) \\
 &= \theta_i \theta_j \sum_{m=2}^M m \sum_{k_3=1}^K \cdots \sum_{k_m=1}^K \prod_{l=3}^m n_{k_l} \cdot \Phi(g_i, g_j, k_3, \dots, k_m).
 \end{aligned}$$

For each $e \in E$ such that $a_{ei} \geq 2$, one obtains $b_{e \setminus \{i,i\}} = \frac{(|e|-2)!}{(\prod_{j \neq i} a_{ej}) \cdot (a_{ei}-2)!} = \frac{a_{ei}(a_{ei}-1)}{|e|(|e|-1)} b_e$. Therefore

$$A_{ii} = \sum_{e \in E: a_{ei} \geq 2} |e| \frac{b_{e \setminus \{i,i\}}}{b_e} h_e,$$

and

$$\begin{aligned}
 P_{ii} &= \sum_{e \in E: a_{ei} \geq 2} |e| \frac{b_{e \setminus \{i,i\}}}{b_e} \mathbb{E}[h_e] \\
 &= \sum_{m=2}^M m \sum_{i_3=1}^n \cdots \sum_{i_m=1}^n \theta_i^2 \prod_{l=3}^m \theta_{i_l} \cdot \Phi(g_i, g_i, g_{i_3}, \dots, g_{i_m}) \\
 &= \theta_i^2 \sum_{m=2}^M m \sum_{k_3=1}^K \cdots \sum_{k_m=1}^K \prod_{l=3}^m n_{k_l} \cdot \Phi(g_i, g_i, k_3, \dots, k_m).
 \end{aligned}$$

The result follows by letting B be a $K \times K$ matrix with

$$B_{rs} = \sum_{m=2}^M m \sum_{k_3=1}^K \cdots \sum_{k_m=1}^K \prod_{l=3}^m n_{k_l} \cdot \Phi(r, s, k_3, \dots, k_m),$$

for $1 \leq r, s \leq K$.

(ii) See, for example, Lemma 4.1 in [8].

B. Proof of Lemma 2

By Lemma 5.1 in [8], there exists an orthogonal matrix $\hat{Q} \in \mathbb{R}^{K \times K}$ such that

$$\|\hat{U}\hat{Q} - U\|_F \leq \frac{2\sqrt{2K}}{|\lambda_K|} \|A - P\| \leq C \frac{2\sqrt{2Kd}}{|\lambda_K|}$$

with high probability.

Note that one cannot guarantee $\|\hat{U}_i\| > 0$ for all $i \in [n]$ with high probability. Let $S_0 = \{i \in [n] \mid \hat{U}_i = 0\}$ be the set of nodes that correspond to zero rows of \hat{U} . The nodes in S_0 are regarded as mis-clustered and are not involved in the k -means step. Since $\|\hat{U}\hat{Q} - U\|_F^2 \geq \sum_{i=1}^n \mathbb{1}_{\{\hat{U}_i=0\}} \|U_i\|^2 \geq \hat{\theta}_{min}^2 |S_0|$, we have $|S_0| \leq \|\hat{U}\hat{Q} - U\|_F^2 / \hat{\theta}_{min}^2$.

Let \hat{U}' and U' be the matrices whose rows are given by the normalized non-zero rows of \hat{U} and U , respectively, it follows that

$$\begin{aligned}
 \|\hat{U}'\hat{Q} - U'\|_F &= \sqrt{\sum_{i \in S_0^c} \left\| \frac{\hat{U}_i}{\|\hat{U}_i\|} \hat{Q} - \frac{U_i}{\|U_i\|} \right\|^2} \\
 &\leq \sqrt{\sum_{i=1}^n \frac{\|\hat{U}_i \hat{Q} - U_i\|^2}{\|U_i\|^2}} \\
 &\leq \frac{\|\hat{U}\hat{Q} - U\|_F}{\hat{\theta}_{min}}.
 \end{aligned}$$

Define $S_1 = \{i \in S_0^c \mid \|\hat{U}'_i \hat{Q} - U'_i\| \geq 1/\sqrt{2}\}$, then

$$|S_1| \leq 4(2 + \epsilon) \|\hat{U}'\hat{Q} - U'\|_F^2$$

By Lemma 5.3 in [8]. When

$$n_K > 4(2 + \epsilon) \|\hat{U}'\hat{Q} - U'\|_F^2,$$

the nodes outside $S_0 \cup S_1$ are all correctly assigned.

Thus, one can choose $c = \frac{1}{32(2+\epsilon)C}$ so that when

$$\frac{Kd}{n_K \lambda_K^2 \tilde{\theta}_{min}^2} < c,$$

the nodes outside $S_0 \cup S_1$ are all correctly clustered, and the number of incorrectly clustered nodes is bounded by

$$l(g, \hat{g}) \leq |S_0| + |S_1| \leq 8(9 + 4\epsilon)C \frac{Kd}{\lambda_K^2 \tilde{\theta}_{min}^2}.$$

C. Proof of Corollary 2

Under the current settings, it is easy to obtain

$$B_{rs} = m\alpha_m ((p - q)n_r^{m-2} \delta_{rs} + qn^{m-2}).$$

Notice that the eigenvalues of P are identical to those of $\text{diag}(\phi)B\text{diag}(\phi)$, we have

$$\begin{aligned}
 \lambda_{max}(B) &\leq m\alpha_m ((p - q)n_1^{m-2} + Kqn^{m-2}) \\
 &\leq Kpm\alpha_m n^{m-2}, \\
 \lambda_1 &\leq \lambda_{max}(B) \cdot \max_k \phi_k^2 \leq Kpm\alpha_m \phi_{max}^2 n^{m-2}, \\
 \lambda_{min}(B) &\geq (p - q)m\alpha_m n_K^{m-2}, \\
 \lambda_K &\geq \lambda_{min}(B) \cdot \min_k \phi_k^2 \geq (p - q)m\alpha_m \phi_{min}^2 n_K^{m-2},
 \end{aligned}$$

and then $\kappa = \lambda_1/\lambda_K \leq c_1 K^{m-1} \phi_{max}^2 / \phi_{min}^2$ for some constant $c_1 > 0$. Since $\phi_k^2 \geq (\sum_i \theta_i \delta_{g_i, k})^2 / n_k = n_k$, we have $\lambda_K \geq (p - q)m\alpha_m n_K^{m-1}$. When $\max_{ij} P_{ij} \leq \theta_{max}^2 \max_{rs} B_{rs} \leq pm\alpha_m \theta_{max}^2 n^{m-2}$, we have $d \leq \max\{pm\alpha_m \theta_{max}^2 n^{m-1}, c_0 \log n\}$.

Consider the consistency of Algorithm 1, a sufficient condition for $\gamma K^{3/2} \frac{\sqrt{d \log n}}{|\lambda_K|}$ being sufficiently small is

$$\alpha_m \geq C_1 \frac{\gamma^2 \theta_{max}^2 K^{2m+1} \log n}{n^{m-1}}$$

for some constant $C_1 > 0$.

Analogously, there exists a constant $C_2 > 0$ such that if

$$\alpha_m \geq C_2 \frac{\gamma^2 \theta_{max}^2 K^{2m-2} \log n}{n^{m-1}},$$

then Algorithm 2 is strongly consistent.

D. Proof of Corollary 3

It is easy to obtain the following conclusions:

$$B_{rs} = \sum_{m=2}^M m\alpha_m ((p-q)n_r^{m-2}\delta_{rs} + qn^{m-2}),$$

$$\lambda_{\max}(B) \leq Kp \sum_{m=2}^M m\alpha_m n^{m-2},$$

$$\lambda_{\min}(B) \geq (p-q) \sum_{m=2}^M m\alpha_m n_K^{m-2} \geq c_1 \frac{p-q}{K^{m_0-2}} \sum_{m=2}^M m\alpha_m n^{m-2},$$

$$\lambda_1 \leq \phi_{\max}^2 \lambda_{\max}(B) \leq Kp\phi_{\max}^2 \sum_{m=2}^M m\alpha_m n^{m-2},$$

$$\lambda_K \geq \phi_{\min}^2 \lambda_{\min}(B) \geq \frac{c_1(p-q)\phi_{\min}^2}{K^{m_0-2}} \sum_{m=2}^M m\alpha_m n^{m-2},$$

$$\kappa \leq \frac{p}{c_1(p-q)} K^{m_0-1} \phi_{\max}^2 / \phi_{\min}^2,$$

where $c_1 > 0$ is a constant.

From the fact that $\phi_{\min} \geq \sqrt{n_K}$, we have $\lambda_K \geq \frac{c_2(p-q)}{K^{m_0-1}} \sum_{m=2}^M m\alpha_m n^{m-1}$ for some constant $c_2 > 0$. Since $\max_{ij} P_{ij} \leq p\theta_{\max}^2 \sum_{m=2}^M m\alpha_m n^{m-2}$, we have $d \leq \max\{p\theta_{\max}^2 \sum_{m=2}^M m\alpha_m n^{m-1}, c_0 \log n\}$. Then, the sufficient conditions for $\gamma K^{3/2} \sqrt{d} \log n / |\lambda_K|$ and $\gamma \sqrt{d} \log n / |\lambda_K|$ being sufficiently small are

$$\sum_{m=2}^M m\alpha_m n^{m-1} \geq C_1 \gamma^2 \theta_{\max}^2 K^{2m_0+1} \log n$$

for some constant $C_1 > 0$ and

$$\sum_{m=2}^M m\alpha_m n^{m-1} \geq C_2 \gamma^2 \theta_{\max}^2 K^{2m_0-2} \log n$$

for some constant $C_2 > 0$, respectively.

APPENDIX B
PROOF OF THEOREM 2

The overall structure of the proof is similar to that in [8]. Let $W = A - P$ and $S = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ be the unit ball in \mathbb{R}^n , the goal is to bound the spectral norm of W :

$$\|W\| = \sup_{x \in S} |x^T W x|.$$

We first outline the three major steps of the proof.

(1). Discretization. Fix a constant $\delta \in (0, 1/3)$ and define a set of grid points in S :

$$T = \{x = (x_1, \dots, x_n)^T \in S \mid \sqrt{n}x_i/\delta \in \mathbb{Z}, \text{ for } i \in [n]\},$$

we will prove the following results:

- (i) T is a δ -net of S . That is, T is a finite subset of S such that for any $x \in S$, there exists a point $y \in T$ satisfying $\|x - y\| \leq \delta$.
- (ii) $\|W\| \leq (1 - 3\delta)^{-1} \sup_{x, y \in T} |x^T W y|$.

Bounding $\|W\|$ is then reduced to bound the supremum of $|x^T W y|$ over all $x, y \in T$. Next, we split the point pairs in the grid into two parts, called the *light pairs*

and *heavy pairs*, and then bound each of the two parts separately.

- (2). Bounding the light pairs. We use Bernstein inequality to bound the contribution of the light pairs.
- (3). The contribution of heavy pairs, however, cannot be simply bounded by the standard Bernstein inequality. We show that the two key properties of the random hypergraph, i.e., *bounded degree property* and *bounded discrepancy property*, still hold and thus the contribution of heavy pairs is bounded.

Lemma 3: T is a δ -net of S .

Proof: It is straightforward to see that T is a finite subset of S . For any $x = (x_1, \dots, x_n)^T \in S$, without loss of generality, assume $x_i \geq 0$ for $i = 1, \dots, n$. For each $1 \leq i \leq n$, there exists $k_i \in \mathbb{Z}$ such that

$$0 \leq k_i \delta / \sqrt{n} \leq x_i \leq (k_i + 1) \delta / \sqrt{n}.$$

Let $y = (k_1 \delta / \sqrt{n}, \dots, k_n \delta / \sqrt{n})^T$, we have $\|y\| \leq \|x\| \leq 1$, which means $y \in T$. Moreover, we have $\|x - y\| \leq \sqrt{n(\delta/\sqrt{n})^2} = \delta$.

Lemma 4: $\|W\| \leq (1 - 3\delta)^{-1} \sup_{x, y \in T} |x^T W y|$.

Proof: For any $x_0 \in S$, according to Lemma 3, there exist $x_1, x_2 \in T$ such that $\|x_0 - x_1\| \leq \delta$ and $\|x_0 - x_2\| \leq \delta$. We have

$$\begin{aligned} |x_0^T W x_0| &= |(x_0 - x_1 + x_1)^T W (x_0 - x_2 + x_2)| \\ &\leq |(x_0 - x_1)^T W (x_0 - x_2)| + |(x_0 - x_1)^T W x_2| \\ &\quad + |x_1^T W (x_0 - x_2)| + |x_1^T W x_2| \\ &\leq (\delta^2 + 2\delta) \|W\| + |x_1^T W x_2| \\ &\leq 3\delta \|W\| + \sup_{x, y \in T} |x^T W y|. \end{aligned}$$

Then $\|W\| = \sup_{x \in S} |x^T W x| \leq 3\delta \|W\| + \sup_{x, y \in T} |x^T W y|$.

Split the pairs (x_i, y_j) into light pairs: $L_{xy} = \{(i, j) \mid |x_i y_j| \leq \sqrt{d}/n\}$ and heavy pairs $\bar{L}_{xy} = \{(i, j) \mid |x_i y_j| > \sqrt{d}/n\}$. The corresponding contributions are $I_{xy} = \sum_{(i, j) \in L_{xy}} x_i y_j w_{ij}$ and $\bar{I}_{xy} = \sum_{(i, j) \in \bar{L}_{xy}} x_i y_j w_{ij}$.

Lemma 5: For any constant $c > 0$,

$$\begin{aligned} &\mathbb{P} \left(\sup_{x, y \in T} |I_{xy}| \geq c\sqrt{d} \right) \\ &\leq 2 \exp \left[- \left(\frac{c^2/2}{M(M^2 + c/3)} - 2 \log \left(\frac{2}{\delta} + 1 \right) \right) n \right]. \end{aligned}$$

Proof: For any $x, y \in T$, we have

$$\begin{aligned} I_{xy} &= \sum_{(i, j) \in L_{xy}} x_i y_j \sum_{e \in E} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} (h_e - \mathbb{E}[h_e]) \\ &= \sum_{e \in E} \left(\sum_{(i, j) \in L_{xy}} \frac{a_{ei}(a_{ej} - \delta_{ij}) x_i y_j}{|e| - 1} \right) (h_e - \mathbb{E}[h_e]). \end{aligned}$$

Let $S_e = \left(\sum_{(i, j) \in L_{xy}} \frac{a_{ei}(a_{ej} - \delta_{ij}) x_i y_j}{|e| - 1} \right) (h_e - \mathbb{E}[h_e])$, then $I_{xy} = \sum_{e \in E} S_e$. Note that $\{S_e\}_{e \in E}$ are independent random variables, and

$$\mathbb{E}[S_e] = 0,$$

$$|S_e| \leq \frac{1}{|e|-1} \cdot \frac{\sqrt{d}}{n} \sum_{(i,j) \in L_{xy}} a_{ei}(a_{ej} - \delta_{ij}) \leq M \frac{\sqrt{d}}{n},$$

$$\begin{aligned} \sum_{e \in E} \mathbb{E}[S_e^2] &= \sum_{e \in E} \frac{\text{var}(h_e)}{(|e|-1)^2} \left(\sum_{(i,j) \in L_{xy}} a_{ei}(a_{ej} - \delta_{ij})x_i y_j \right)^2 \\ &\leq \sum_{e \in E} \mathbb{E}[h_e] \frac{|e|}{|e|-1} \sum_{(i,j) \in L_{xy}} a_{ei}^2 (a_{ej} - \delta_{ij})^2 x_i^2 y_j^2 \\ &\leq \sum_{e \in E} \mathbb{E}[h_e] \frac{|e|}{|e|-1} \sum_{1 \leq i,j \leq n} a_{ei}^2 (a_{ej} - \delta_{ij})^2 x_i^2 y_j^2 \\ &\leq \sum_{e \in E} \mathbb{E}[h_e] |e|^2 \cdot \sum_{1 \leq i,j \leq n} a_{ei}(a_{ej} - \delta_{ij}) x_i^2 y_j^2 \\ &\leq M^3 \sum_{e \in E} \mathbb{E}[h_e] \cdot \sum_{1 \leq i,j \leq n} \frac{a_{ei}(a_{ej} - \delta_{ij}) x_i^2 y_j^2}{|e|-1} \\ &= M^3 \sum_{1 \leq i,j \leq n} \left(\sum_{e \in E} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} \mathbb{E}[h_e] \right) x_i^2 y_j^2 \\ &= M^3 \sum_{1 \leq i,j \leq n} P_{ij} x_i^2 y_j^2 \\ &\leq M^3 d/n. \end{aligned}$$

Applying the Bernstein inequality yields

$$\begin{aligned} \mathbb{P}(|I_{xy}| \geq c\sqrt{d}) &= \mathbb{P}\left(\left| \sum_{e \in E} S_e \right| \geq c\sqrt{d} \right) \\ &\leq 2 \exp\left(\frac{-c^2 d/2}{\frac{M^3 d}{n} + M \frac{\sqrt{d}}{n} \cdot c\sqrt{d}/3} \right) \\ &= 2 \exp\left(-\frac{c^2/2}{M(M^2 + c/3)} n \right). \end{aligned}$$

Since $|T| \leq (\frac{2}{\delta} + 1)^n$, we have

$$\begin{aligned} \mathbb{P}\left(\sup_{x,y \in T} |I_{xy}| \geq c\sqrt{d} \right) &\leq |T|^2 \cdot 2 \exp\left(-\frac{c^2/2}{M(M^2 + c/3)} n \right) \\ &\leq 2 \exp\left[-\left(\frac{c^2/2}{M(M^2 + c/3)} - 2 \log\left(\frac{2}{\delta} + 1 \right) \right) n \right]. \end{aligned}$$

Next, we show that $\sup_{x,y \in T} |\bar{I}_{xy}|$ is bounded by $c\sqrt{d}$. Since for any $x, y \in T$,

$$\begin{aligned} \left| \sum_{(i,j) \in \bar{L}_{xy}} x_i y_j P_{ij} \right| &\leq \sum_{(i,j) \in \bar{L}_{xy}} \frac{x_i^2 y_j^2}{|x_i y_j|} P_{ij} \\ &\leq \frac{n}{\sqrt{d}} p_{max} \sum_{1 \leq i,j \leq n} x_i^2 y_j^2 \\ &\leq \sqrt{d}, \end{aligned}$$

we need only to show that

$$\left| \sum_{(i,j) \in \bar{L}_{xy}} x_i y_j A_{ij} \right| = O(\sqrt{d})$$

for any $x, y \in T$ (not with high probability but definitely). In what follows, we prove that both the bounded degree property and the bounded discrepancy property hold.

Lemma 6 (Bounded degree property): For any constant $c > 0$, there exists a constant $c_1 = c_1(c)$ such that $\mathbb{P}(d_i \geq c_1 d) \leq n^{-c}$ for any node $i \in [n]$.

Proof: Recall that $d_i = \sum_{j=1}^n A_{ij} = \sum_{e \in E} a_{ei} h_e$. Notice that $\mathbb{E}[d_i] = \sum_{j=1}^n \mathbb{E}[A_{ij}] = \sum_{j=1}^n P_{ij} \leq d$, we have

$$\begin{aligned} \mathbb{P}(d_i \geq c_1 d) &= \mathbb{P}(d_i - \mathbb{E}[d_i] \geq c_1 d - \mathbb{E}[d_i]) \\ &\leq \mathbb{P}(d_i - \mathbb{E}[d_i] \geq (c_1 - 1)d) \\ &= \mathbb{P}\left(\sum_{e \in E} a_{ei}(h_e - \mathbb{E}[h_e]) \geq (c_1 - 1)d \right). \end{aligned}$$

Let $S_e = a_{ei}(h_e - \mathbb{E}[h_e])$. Since $\mathbb{E}[S_e] = 0$, $|S_e| \leq M$, and

$$\begin{aligned} \sum_{e \in E} \mathbb{E}[S_e^2] &= \sum_{e \in E} a_{ei}^2 \text{var}(h_e) \\ &\leq M \sum_{e \in E} a_{ei} \mathbb{E}[h_e] \\ &= M \cdot \mathbb{E}[d_i] \\ &\leq Md, \end{aligned}$$

we have

$$\begin{aligned} \mathbb{P}(d_i \geq c_1 d) &\leq \exp\left(\frac{-(c_1 - 1)^2 d^2 / 2}{Md + M \cdot (c_1 - 1)d/3} \right) \\ &\leq \exp\left(-\frac{c_0(c_1 - 1)^2 / 2}{M(2 + c_1)/3} \log n \right) \\ &= n^{-\frac{3c_0(c_1 - 1)^2}{2M(2 + c_1)}} \end{aligned}$$

by the Bernstein inequality.

For any $I, J \subset [n]$, let $e(I, J) = \sum_{i \in I} \sum_{j \in J} A_{ij}$ and define $\mu(I, J) = |I||J| \frac{d}{n}$ which is an upper bound on the expectation of $e(I, J)$.

Lemma 7 (Bounded discrepancy property): For any constant $c > 0$, there exist constants $c_2 = c_2(c)$ and $c_3 = c_3(c)$, such that, with probability least $1 - 2n^{-c}$, for any $I, J \subset [n]$ with $|I| \leq |J|$, either of the following holds:

- $e(I, J)/\mu(I, J) \leq ec_2$ (here e denotes Euler's number).
- $e(I, J) \log \frac{e(I, J)}{\mu(I, J)} \leq c_3 |J| \log \frac{n}{|J|}$.

Proof: If $|J| \geq n/e$, according to Lemma 6, there exists $c_1 = c_1(c)$ such that $e(I, J) \leq \sum_{i \in I} d_i \leq c_1 d |I|$ with probability at least $1 - n^{-c}$. Since $\mu(I, J) \geq \frac{d}{e} |I|$, we have $e(I, J)/\mu(I, J) \leq ec_1$ with high probability.

Otherwise, suppose $|J| < n/e$. Note that

$$\begin{aligned} e(I, J) &= \sum_{i \in I} \sum_{j \in J} \sum_{e \in E} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} h_e \\ &= \sum_{e \in E} \left(\sum_{i \in I, j \in J} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} \right) h_e \end{aligned}$$

is a sum over independent random variables and

$$0 \leq \left(\sum_{i \in I, j \in J} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} \right) h_e \leq |e| \leq M.$$

According to Lemma 5 in [10], there exists a constant $c' = c'(M) > 1$ such that

$$\mathbb{P}(e(I, J) > k\mu(I, J)) \leq \exp\left(-\frac{1}{2M} k \log k \cdot \mu(I, J)\right)$$

for any $k \geq c'$. Given $c_3 > 0$, we choose $k' = \max\{c', t(I, J)\}$ where $t(I, J) \geq 1$ is the unique solution of

$$t \log t = c_3 \frac{|J|}{\mu(I, J)} \log \frac{n}{|J|},$$

then

$$\begin{aligned} \mathbb{P}(e(I, J) \geq k' \mu(I, J)) &\leq \exp\left(-\frac{1}{2M} k' \log k' \cdot \mu(I, J)\right) \\ &\leq \exp\left(-\frac{c_3}{2M} |J| \log \frac{n}{|J|}\right). \end{aligned}$$

According to the proof of Lemma 4.2 in [8], we have

$$\mathbb{P}(\exists(I, J) : |I| \leq |J| \leq n/e, e(I, J) \geq k' \mu(I, J)) \leq n^{-(c_3/M-12)/3}. \quad (15)$$

Therefore, when $c_3 > 12M$, for any $I, J \subset [n]$ with $|I| \leq |J| \leq n/e$, with high probability at least one of the following holds:

- $k' = c' \geq t(I, J)$, which implies $e(I, J) \leq c' \mu(I, J)$.
- $k' = t(I, J) \geq c'$, which means $k' \log k' = c_3 \frac{|J|}{\mu(I, J)} \log \frac{n}{|J|}$.

In the second case, according to (15), we have $e(I, J) \leq k' \mu(I, J)$ with high probability, and then

$$\frac{e(I, J)}{\mu(I, J)} \log \frac{e(I, J)}{\mu(I, J)} \leq k' \log k' = c_3 \frac{|J|}{\mu(I, J)} \log \frac{n}{|J|},$$

which yields

$$e(I, J) \log \frac{e(I, J)}{\mu(I, J)} \leq c_3 |J| \log \frac{n}{|J|}$$

with high probability. Note that both c_2 and c_3 depend only on c and M . The result follows by setting $c_2 = \max\{c_1, c'\}$ and $c_3 = (2c + 12)M$.

Lemma 8: If both the bounded degree property and the bounded discrepancy property hold with some constant c_1, c_2, c_3 , then

$$\sup_{x, y \in T} \left| \sum_{(i, j) \in \bar{L}_{xy}} x_i y_j A_{ij} \right| = O(\sqrt{d}).$$

We refer to [8], [25] for a proof.

APPENDIX C PROOF OF THEOREM 3

Let $r > 0$ be a fixed constant. We first provide several useful lemmas.

Lemma 9: The following two conclusions hold.

- (i). There exists a constant $c_1 = c_1(M, c_0, r)$ such that

$$\max_{l \in [n]} \|A^{(l)} - P\| \leq c_1 \sqrt{d}, \|A - P\| \leq c_1 \sqrt{d}$$

with probability at least $1 - O(n^{-r})$.

- (ii). $\|H\| \leq 1, \|H^{(l)}\| \leq 1$. If $|\lambda_K| \geq 4 \max\{\|A - P\|, \|A^{(l)} - P\|\}$, then $\|H - \text{sgn}(H)\|^{1/2} \leq 2\|A - P\|/|\lambda_K|, \|H^{-1}\| \leq 2, \|(H^{(l)})^{-1}\| \leq 2$, and

$$\|\hat{U} \hat{U}^T - \hat{U}^{(l)} (\hat{U}^{(l)})^T\| \leq \frac{2\|(A - A^{(l)}) \hat{U}^{(l)}\|}{|\lambda_K|}. \quad (16)$$

Proof:

- (i). For any $l \in [n]$, it follows that

$$(A^{(l)} - P)_{ij} = \sum_{e \in E: a_{ei}=0} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} (h_e - \mathbb{E}[h_e]).$$

According to the proof of Theorem 2, there exists a constant $c_1 > 0$ such that $\mathbb{P}(\|A - P\| \leq c_1 \sqrt{d}) \geq 1 - n^{-r}$ and $\mathbb{P}(\|A^{(l)} - P\| \leq c_1 \sqrt{d}) \geq 1 - n^{-r-1}$. Then

$$\begin{aligned} &\mathbb{P}\left(\max_{l \in [n]} \|A^{(l)} - P\| \leq c_1 \sqrt{d}, \|A - P\| \leq c_1 \sqrt{d}\right) \\ &\geq 1 - n \cdot n^{-r-1} - n^{-r} = 1 - O(n^{-r}). \end{aligned}$$

We only prove (16). See Lemma 2 in [17] and Lemma 4.14 in [30] for the proof of the other conclusions.

According to Lemma 2.5 in [30] and the Davis-Kahan theorem, we have

$$\|\hat{U} \hat{U}^T - \hat{U}^{(l)} (\hat{U}^{(l)})^T\| \leq \frac{\|(A - A^{(l)}) \hat{U}^{(l)}\|}{\delta}$$

with $\delta = \min_{1 \leq i \leq K < j \leq n} |\hat{\lambda}_i - \hat{\lambda}_j^{(l)}|$. By Weyl's inequality, it follows that $|\hat{\lambda}_K - \lambda_K| \leq \|A - P\| \leq |\lambda_K|/4$. Then $|\lambda_K| \geq 3/(4|\lambda_K|)$. On the other hand, Weyl's inequality forces that

$$|\hat{\lambda}_{K+1}^{(l)}| = |\hat{\lambda}_{K+1}^{(l)} - \lambda_{K+1}| \leq \|A^{(l)} - P\| \leq |\lambda_K|/4.$$

Then $\delta = |\hat{\lambda}_K| - |\hat{\lambda}_{K+1}^{(l)}| \geq |\lambda_K|/2$, and the result follows.

Lemma 10: For any fixed matrix $X \in \mathbb{R}^{n \times K}$, there exists a constant $C = C(M, c_0, r)$ such that

$$\|(A - P)X\|_{2, \infty} \leq C \sqrt{d \log n} \|X\|_{2, \infty}$$

with probability at least $1 - O(n^{-r-1})$.

Proof: We use the matrix Bernstein inequality to derive the bound. For each $i \in [n]$, we have

$$\begin{aligned} \|(A - P)_i X\| &= \left\| \sum_{j=1}^n (A_{ij} - P_{ij}) X_j \right\| \\ &= \left\| \sum_{j=1}^n \left(\sum_{e \in E} \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} (h_e - \mathbb{E}[h_e]) \right) X_j \right\| \\ &= \left\| \sum_{e \in E} \left(\sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} X_j \right) (h_e - \mathbb{E}[h_e]) \right\|. \end{aligned}$$

Let $S_e = \left(\sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} X_j \right) (h_e - \mathbb{E}[h_e])$.

- (i). $\mathbb{E}[S_e] = 0$, and

$$\begin{aligned} \|S_e\| &\leq \left\| \sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} X_j \right\| \\ &\leq \sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} \|X_j\| \\ &\leq \sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e| - 1} \|X\|_{2, \infty} \\ &= a_{ei} \|X\|_{2, \infty} \leq M \|X\|_{2, \infty}. \end{aligned}$$

(ii). Define $W_e = \sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} X_j$. Then

$$\begin{aligned} v &= \max\left\{\left\|\sum_{e \in E} \mathbb{E}[S_e S_e^T]\right\|, \left\|\sum_{e \in E} \mathbb{E}[S_e^T S_e]\right\|\right\} \\ &= \max\left\{\left\|\sum_{e \in E} \text{var}(h_e) W_e W_e^T\right\|, \left\|\sum_{e \in E} \text{var}(h_e) W_e^T W_e\right\|\right\} \\ &\leq \max\left\{\sum_{e \in E} \text{var}(h_e) \|W_e W_e^T\|, \sum_{e \in E} \text{var}(h_e) \|W_e^T W_e\|\right\} \\ &= \sum_{e \in E} \text{var}(h_e) \|W_e\|^2. \end{aligned}$$

Since W_e is an $1 \times K$ row vector, we have

$$\begin{aligned} v &\leq \sum_{e \in E} \text{var}(h_e) \left\|\sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} X_j\right\|^2 \\ &\leq \sum_{e \in E} \mathbb{E}[h_e] \left(\sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} \|X_j\|\right)^2 \\ &\leq \sum_{e \in E} \mathbb{E}[h_e] (a_{ei} \|X\|_{2,\infty})^2 \\ &= M \|X\|_{2,\infty}^2 \sum_{e \in E} a_{ei} \mathbb{E}[h_e] \\ &= M \|X\|_{2,\infty}^2 \mathbb{E}[d_i] \leq Md \|X\|_{2,\infty}^2. \end{aligned}$$

Let $C > 0$ be a constant such that $\frac{3C^2}{2M(3+C/\sqrt{c_0})} \geq r + 3$. By the matrix Bernstein inequality, it follows that

$$\begin{aligned} \mathbb{P}(\|(A - P)_i X\| \geq C\sqrt{d \log n} \|X\|_{2,\infty}) &\leq (1 + K) \exp\left(\frac{-3C^2 d \log n}{2M(3d + C\sqrt{d \log n})}\right) \\ &\leq (1 + K) \exp\left(\frac{-3C^2 d \log n}{2M(3d + Cd/\sqrt{c_0})}\right) \\ &= (1 + K) n^{-\frac{3C^2}{2M(3+C/\sqrt{c_0})}} \\ &\leq (1 + K) n^{-r-3} \leq n^{-r-2}. \end{aligned}$$

Then

$$\mathbb{P}\left(\|(A - P)X\|_{2,\infty} \geq C\sqrt{d \log n} \|X\|_{2,\infty}\right) \leq n^{-r-1}.$$

As a consequence, there exists a constant $c_2 = c_2(M, c_0, r)$ such that

$$\mathbb{P}(\|(A - P)U\|_{2,\infty} \leq c_2 \sqrt{d \log n} \|U\|_{2,\infty}) \geq 1 - O(n^{-r-1}).$$

Furthermore, due to the independence between $(A - P)_l$ and $\hat{U}^{(l)} H^{(l)} - U$, by the proof of Lemma 10, there is a constant $c_3 = c_3(M, c_0, r)$ such that for each $l \in [n]$

$$\|(A - P)_l (\hat{U}^{(l)} H^{(l)} - U)\| \leq c_3 \sqrt{d \log n} \|\hat{U}^{(l)} H^{(l)} - U\|_{2,\infty}$$

with probability at least $1 - O(n^{-r-1})$. Then

$$\|(A - P)_l (\hat{U}^{(l)} H^{(l)} - U)\| \leq c_3 \sqrt{d \log n} \|\hat{U}^{(l)} H^{(l)} - U\|_{2,\infty},$$

for any $l \in [n]$, with probability exceeding $1 - O(n^{-r})$.

Lemma 11: For any fixed matrix $X \in \mathbb{R}^{n \times K}$, there exists a constant $C = C(M, c_0, r)$ such that

$$\|(A - A^{(l)})X\| \leq C\sqrt{d \log n} \|X\|_{2,\infty}, \forall l \in [n],$$

with probability at least $1 - O(n^{-r})$.

Proof: We use the matrix Bernstein inequality to derive the bound. Define $W_e = \frac{a_e a_e^T - \text{diag}(a_e)}{|e|-1}$, then $A - A^{(l)} = \sum_{e \in E: a_{el} > 0} (h_e - \mathbb{E}[h_e]) W_e$ and

$$\|(A - A^{(l)})X\| = \left\| \sum_{e \in E: a_{el} > 0} (h_e - \mathbb{E}[h_e]) W_e X \right\|.$$

Define $S_e = (h_e - \mathbb{E}[h_e]) W_e X$, then

(i). $\mathbb{E}[S_e] = 0$, and

$$\begin{aligned} \|S_e\| &\leq \|W_e X\| \\ &\leq \|W_e X\|_F \\ &= \sqrt{\sum_{i=1}^n \sum_{j=1}^n (W_e)_{ij} X_j} \\ &= \sqrt{\sum_{i=1}^n \sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} X_j} \\ &\leq \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^n \frac{a_{ei}(a_{ej} - \delta_{ij})}{|e|-1} \|X_j\|\right)^2} \\ &\leq \sqrt{\sum_{i=1}^n (a_{ei} \|X\|_{2,\infty})^2} \\ &\leq \|X\|_{2,\infty} \sqrt{M \sum_{i=1}^n a_{ei}} \\ &\leq M \|X\|_{2,\infty}. \end{aligned}$$

(ii).

$$\begin{aligned} v &= \max\left\{\left\|\sum_{e \in E: a_{el} > 0} \mathbb{E}[S_e S_e^T]\right\|, \left\|\sum_{e \in E: a_{el} > 0} \mathbb{E}[S_e^T S_e]\right\|\right\} \\ &\leq \sum_{e \in E: a_{el} > 0} \text{var}(h_e) \|W_e X\|^2 \\ &\leq \sum_{e \in E: a_{el} > 0} \mathbb{E}[h_e] \|W_e X\|_F^2 \\ &\leq \sum_{e \in E: a_{el} > 0} \mathbb{E}[h_e] \cdot M^2 \|X\|_{2,\infty}^2 \\ &\leq M^2 \|X\|_{2,\infty}^2 \sum_{e \in E} a_{el} \mathbb{E}[h_e] \\ &= M^2 \|X\|_{2,\infty}^2 \mathbb{E}[d_l] \\ &\leq M^2 d \|X\|_{2,\infty}^2. \end{aligned}$$

Let $C > 0$ be a constant such that $\frac{3C^2}{2M(3M+C/\sqrt{c_0})} \geq r + 2$. By the matrix Bernstein inequality, it follows that

$$\begin{aligned} \mathbb{P}(\|(A - A^{(l)})X\| \geq C\sqrt{d \log n} \|X\|_{2,\infty}) &\leq (1 + K) \exp\left(\frac{-3C^2 d \log n}{2M(3Md + C\sqrt{d \log n})}\right) \\ &\leq (1 + K) \exp\left(\frac{-3C^2 d \log n}{2M(3Md + Cd/\sqrt{c_0})}\right) \\ &= (1 + K) n^{-\frac{3C^2}{2M(3M+C/\sqrt{c_0})}} \\ &\leq (1 + K) n^{-r-2} \leq n^{-r-1}. \end{aligned}$$

Then

$$\|(A - A^{(l)})X\| \leq C\sqrt{d\log n}\|X\|_{2,\infty}, \forall l \in [n]$$

with probability at least $1 - n^{-r}$.

Since $A - A^{(l)}$ is independent of $\hat{U}^{(l)}$, according to Lemma 11, there exists a constant $c_4 = c_4(M, c_0, r)$ such that

$$\|(A - A^{(l)})\hat{U}^{(l)}\| \leq c_4\sqrt{d\log n}\|\hat{U}^{(l)}\|_{2,\infty}, \forall l \in [n]$$

with probability exceeding $1 - O(n^{-r})$.

Define the following events:

$$\mathcal{E}_1 = \{\|A^{(l)} - P\| \leq c_1\sqrt{d}, \forall l \in [n], \text{ and } \|A - P\| \leq c_1\sqrt{d}\},$$

$$\mathcal{E}_2 = \{\|(A - P)U\|_{2,\infty} \leq c_2\sqrt{d\log n}\|U\|_{2,\infty}\},$$

$$\mathcal{E}_3 = \{\|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\| \leq c_3\sqrt{d\log n}\|\hat{U}^{(l)}H^{(l)} - U\|_{2,\infty}, \forall l \in [n]\},$$

$$\mathcal{E}_4 = \{\|(A - A^{(l)})\hat{U}^{(l)}\| \leq c_4\sqrt{d\log n}\|\hat{U}^{(l)}\|_{2,\infty}, \forall l \in [n]\}.$$

Lemma 12: There exist some constants $c_5, c_6 > 0$ (depending only on M, c_0 and r) such that if $\sqrt{d\log n} \leq c_5|\lambda_K|$, then

$$\|A(\hat{U}H - U)\|_{2,\infty} \leq c_6 \left(\sqrt{d\log n}\|\hat{U}H - U\|_{2,\infty} + (\sqrt{d\log n} + \frac{\kappa d}{|\lambda_K|})\|U\|_{2,\infty} \right)$$

with probability at least $1 - O(n^{-r})$.

Proof: By the triangle inequality, it follows that

$$\|A(\hat{U}H - U)\|_{2,\infty} \leq \|(A - P)(\hat{U}H - U)\|_{2,\infty} + \|P(\hat{U}H - U)\|_{2,\infty}$$

We first use the leave-one-out technique to develop an upper bound on $\|(A - P)(\hat{U}H - U)\|_{2,\infty}$. Applying the triangle inequality again yields

$$\begin{aligned} \|(A - P)(\hat{U}H - U)\|_{2,\infty} &= \max_{l \in [n]} \|(A - P)_l(\hat{U}H - U)\|_{2,\infty} \\ &\leq \max_{l \in [n]} (\|(A - P)_l(\hat{U}H - \hat{U}^{(l)}H^{(l)})\| \\ &\quad + \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\|) \\ &\leq \max_{l \in [n]} (\|A - P\| \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| \\ &\quad + \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\|). \end{aligned}$$

That is, $\hat{U}^{(l)}H^{(l)}$ is employed here as a surrogate of $\hat{U}H$.

Set $c_5 \leq 1/(4c_1)$ such that $|\lambda_K| \geq 4c_1\sqrt{d}$. When $\mathcal{E}_1 \cap \mathcal{E}_4$ happens, which has probability at least $1 - O(n^{-r})$, we have

$$\begin{aligned} \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| &= \|[\hat{U}\hat{U}^T - \hat{U}^{(l)}(\hat{U}^{(l)})^T]U\| \\ &\leq \|\hat{U}\hat{U}^T - \hat{U}^{(l)}(\hat{U}^{(l)})^T\| \\ &\leq \frac{2\|(A - A^{(l)})\hat{U}^{(l)}\|}{|\lambda_K|} \\ &\leq \frac{2c_4\sqrt{d\log n}\|\hat{U}^{(l)}\|_{2,\infty}}{|\lambda_K|} \\ &\leq 2c_4c_5\|\hat{U}^{(l)}\|_{2,\infty} \\ &\leq 4c_4c_5\|\hat{U}^{(l)}H^{(l)}\|_{2,\infty} \\ &\leq 4c_4c_5(\|\hat{U}H\|_{2,\infty} + \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\|_{2,\infty}) \\ &\leq 4c_4c_5(\|\hat{U}H\|_{2,\infty} + \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\|), \end{aligned}$$

for all $l \in [n]$. So long as $4c_4c_5 \leq \frac{1}{2}$, we have

$$\begin{aligned} \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| &\leq 8c_4c_5\|\hat{U}H\|_{2,\infty} \\ &\leq \|\hat{U}H\|_{2,\infty} \\ &\leq \|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}, \end{aligned}$$

for all $l \in [n]$. When $\mathcal{E}_1 \cap \mathcal{E}_3 \cap \mathcal{E}_4$ happens, which has probability at least $1 - O(n^{-r})$, we can deduce that for all $l \in [n]$,

$$\begin{aligned} \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\| &\leq c_3\sqrt{d\log n}\|\hat{U}^{(l)}H^{(l)} - U\|_{2,\infty} \\ &\leq c_3\sqrt{d\log n}(\|\hat{U}H - \hat{U}^{(l)}H^{(l)}\|_{2,\infty} + \|\hat{U}H - U\|_{2,\infty}) \\ &\leq c_3\sqrt{d\log n}(\|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| + \|\hat{U}H - U\|_{2,\infty}) \\ &\leq c_3\sqrt{d\log n}(2\|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}). \end{aligned}$$

Combining these two results, we have

$$\begin{aligned} \|(A - P)(\hat{U}H - U)\|_{2,\infty} &\leq \max_{l \in [n]} (\|A - P\| \|\hat{U}H - \hat{U}^{(l)}H^{(l)}\| + \|(A - P)_l(\hat{U}^{(l)}H^{(l)} - U)\|) \\ &\leq c_1\sqrt{d}(\|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}) \\ &\quad + c_3\sqrt{d\log n}(2\|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}) \\ &\leq (c_1 + c_3)\sqrt{d\log n}(3\|\hat{U}H - U\|_{2,\infty} + 2\|U\|_{2,\infty}) \end{aligned}$$

with probability exceeding $1 - O(n^{-r})$.

Next, we bound $\|P(\hat{U}H - U)\|_{2,\infty}$. According to Eq. (4.114) in [30], when \mathcal{E}_1 happens and $|\lambda_K| \geq 4c_1\sqrt{d}$, we have

$$\begin{aligned} \|P(\hat{U}H - U)\|_{2,\infty} &\leq \|U\|_{2,\infty} \cdot \|\Lambda\| \cdot \|\hat{U}\hat{U}^T - UU^T\|^2 \\ &\leq \|U\|_{2,\infty} \cdot \|\Lambda\| \cdot \frac{2\|A - P\|^2}{\lambda_K^2} \\ &\leq |\lambda_1| \cdot \frac{2c_1^2d}{\lambda_K^2}\|U\|_{2,\infty} \\ &= \frac{2c_1^2\kappa d}{|\lambda_K|}\|U\|_{2,\infty}. \end{aligned}$$

Combining these two bounds, the result follows by letting $c_5 = \min\{1/(4c_1), 1/(8c_4)\}$ and $c_6 = \max\{3(c_1 + c_3), 2(c_1 + c_3) + 2c_1^2\}$.

Define $\mathcal{E}_5 = \{\|A(\hat{U}H - U)\|_{2,\infty} \leq c_6(\sqrt{d\log n}\|\hat{U}H - U\|_{2,\infty} + (\sqrt{d\log n} + \kappa d/|\lambda_K|)\|U\|_{2,\infty})\}$. We are now ready to prove Theorem 3.

Proof: Suppose $\sqrt{d\log n} \leq c_5|\lambda_K|$ and $\bigcap_{i=1}^5 \mathcal{E}_i$ happens, which has probability exceeding $1 - O(n^{-r})$. According to Lemma 4.16 in [30], we have

$$\|\hat{U}H - U\|_{2,\infty} \leq \gamma_1 + \gamma_2 + \gamma_3$$

with $\gamma_1 = \frac{2\|A(\hat{U}H - U)\|_{2,\infty}}{|\lambda_K|}$, $\gamma_2 = \frac{4c_1\sqrt{d}\|AU\|_{2,\infty}}{\lambda_K^2}$, and $\gamma_3 = \frac{\|(A - P)U\|_{2,\infty}}{|\lambda_K|}$. The last two terms could be directly bounded.

First, when \mathcal{E}_2 happens, we have

$$\gamma_3 \leq c_2 \frac{\sqrt{d\log n}}{|\lambda_K|}\|U\|_{2,\infty}.$$

Next, since

$$\begin{aligned} \|AU\|_{2,\infty} &\leq \|(A-P)U\|_{2,\infty} + \|PU\|_{2,\infty} \\ &= \|(A-P)U\|_{2,\infty} + \|U\Lambda\|_{2,\infty} \\ &\leq c_2\sqrt{d\log n}\|U\|_{2,\infty} + \|U\|_{2,\infty}\|\Lambda\| \\ &= (c_2\sqrt{d\log n} + |\lambda_1|)\|U\|_{2,\infty}, \end{aligned}$$

we have

$$\begin{aligned} \gamma_2 &\leq 4c_1(c_2\frac{\sqrt{d\log n}}{|\lambda_K|} + \kappa)\frac{\sqrt{d}}{|\lambda_K|}\|U\|_{2,\infty} \\ &\leq 4c_1(c_2c_5 + \kappa)\frac{\sqrt{d}}{|\lambda_K|}\|U\|_{2,\infty}. \end{aligned}$$

Now we turn to bound γ_1 . For large n such that $c_5\kappa \leq \log n$, we have

$$\begin{aligned} \gamma_1 &= \frac{2}{|\lambda_K|}\|A(\hat{U}H - U)\|_{2,\infty} \\ &\leq 2c_6\frac{\sqrt{d\log n}}{|\lambda_K|}\left(\|\hat{U}H - U\|_{2,\infty} + \left(1 + \frac{\kappa\sqrt{d}}{\sqrt{\log n}|\lambda_K|}\right)\|U\|_{2,\infty}\right) \\ &\leq 2c_6\frac{\sqrt{d\log n}}{|\lambda_K|}\left(\|\hat{U}H - U\|_{2,\infty} + \left(1 + \frac{c_5\kappa}{\log n}\right)\|U\|_{2,\infty}\right) \\ &\leq 2c_6\frac{\sqrt{d\log n}}{|\lambda_K|}\left(\|\hat{U}H - U\|_{2,\infty} + 2\|U\|_{2,\infty}\right). \end{aligned}$$

The upper bound on $\|\hat{U}H - U\|_{2,\infty}$ gives that

$$\begin{aligned} \|\hat{U}H - U\|_{2,\infty} &\leq 2c_6\frac{\sqrt{d\log n}}{|\lambda_K|}\left(\|\hat{U}H - U\|_{2,\infty} + 2\|U\|_{2,\infty}\right) \\ &\quad + \gamma_2 + \gamma_3. \end{aligned}$$

As long as $c_5 \cdot 2c_6 \leq 1/2$, we have

$$\begin{aligned} \|\hat{U}H - U\|_{2,\infty} &\leq 8c_6\frac{\sqrt{d\log n}}{|\lambda_K|}\|U\|_{2,\infty} + 2\gamma_2 + 2\gamma_3 \\ &\leq \left(8c_6 + \frac{8c_1(c_2c_5 + \kappa)}{\sqrt{\log n}} + 2c_2\right)\frac{\sqrt{d\log n}}{|\lambda_K|}\|U\|_{2,\infty}. \end{aligned}$$

Recall that $\kappa = O(\log n)$ by Assumption 3. Let the constant $c_7 = c_7(M, c_0, r)$ be an upper bound on $8c_6 + \frac{8c_1(c_2c_5 + \kappa)}{\sqrt{\log n}} + 2c_2$. When $\log n > 1$, we have

$$\begin{aligned} \|\hat{U}\text{sgn}(H) - \hat{U}H\|_{2,\infty} &\leq \|\hat{U}\|_{2,\infty}\|H - \text{sgn}(H)\| \\ &\leq 2\|\hat{U}H\|_{2,\infty}\|H - \text{sgn}(H)\| \\ &\leq 2(\|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}) \cdot \frac{4\|A - P\|^2}{\lambda_K^2} \\ &\leq \frac{8c_1^2d}{\lambda_K^2}(\|\hat{U}H - U\|_{2,\infty} + \|U\|_{2,\infty}) \\ &\leq 8c_1^2c_5(1 + c_5c_7)\frac{\sqrt{d\log n}}{|\lambda_K|}\|U\|_{2,\infty}. \end{aligned}$$

Applying the triangle inequality yields

$$\|\hat{U}\text{sgn}(H) - U\|_{2,\infty} \leq \|\hat{U}\text{sgn}(H) - \hat{U}H\|_{2,\infty} + \|\hat{U}H - U\|_{2,\infty}.$$

Thus, the result follows by choosing $C_3 = \min\{c_5, 1/(4c_6)\}$ and $C_4 = c_7 + 8c_1^2c_5(1 + c_5c_7)$. ■

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of China (No. 12071281) and the National Key R & D Program of China (No. 2021YFA1003004).

REFERENCES

- [1] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: first steps," *Social Netw.*, vol. 5, no. 2, pp. 109–137, 1983.
- [2] B. Karrer, and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," *Phys. Rev. E*, vol. 83, no. 1, 2011, Art. no. 016107.
- [3] Y. Zhao, E. Levina, and J. Zhu, "Consistency of community detection in networks under degree-corrected stochastic block models," *Ann. Statist.*, vol. 40, no. 4, pp. 2266–2292, 2012.
- [4] F. Krzakala *et al.*, "Spectral redemption in clustering sparse networks," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 52, pp. 20935–20940, 2013.
- [5] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou, "Community detection in degree-corrected block models," *Ann. Statist.*, vol. 46, no. 5, pp. 2153–2185, 2018.
- [6] L. Gulikers, M. Lelarge, and L. Massoulié, "An impossibility result for reconstruction in the degree-corrected stochastic block model," *Ann. Appl. Probab.*, vol. 28, no. 5, pp. 3002–3027, 2018.
- [7] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic block model," *Ann. Statist.*, vol. 39, no. 4, pp. 1878–1915, 2011.
- [8] J. Lei, and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," *Ann. Statist.*, vol. 43, no. 1, pp. 215–237, 2015.
- [9] P. S. Chodrow, N. Veldt, and A. R. Benson, "Generative hypergraph clustering: from blockmodels to modularity," *Sci. Adv.*, vol. 7, 2021, Art. no. eabh1303.
- [10] K. Ahn, K. Lee, and C. Suh, "Hypergraph spectral clustering in the weighted stochastic block model," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 959–974, 2018.
- [11] D. Ghoshdastidar, and A. Dukkipati, "Consistency of spectral hypergraph partitioning under planted partition model," *Ann. Statist.*, vol. 45, no. 1, pp. 289–315, 2017.
- [12] Z. T. Ke, F. Shi, and D. Xia, "Community detection for hypergraph networks via regularized tensor power iteration," arXiv:1909.06503, 2019.
- [13] S. Cole, and Y. Zhu, "Exact recovery in the hypergraph stochastic block model: a spectral algorithm," *Linear Algebra Appl.*, vol. 593, pp. 45–73, 2020.
- [14] I. E. Chien, C.-Y. Lin, and I.-H. Wang, "On the minimax misclassification ratio of hypergraph community detection," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8095–8118, 2019.
- [15] Q. Zhang, and V. Y. Tan, "Exact recovery in the general hypergraph stochastic block model," *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 453–471, 2022.
- [16] L. Su, W. Wang, and Y. Zhang, "Strong consistency of spectral clustering for stochastic block models," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 324–338, 2019.
- [17] E. Abbe, J. Fan, K. Wang, and Y. Zhong, "Entrywise eigenvector analysis of random matrices with low expected rank," *Ann. Statist.*, vol. 48, no. 3, pp. 1452–1474, 2020.
- [18] C. Kim, A. S. Bandeira, and M. X. Goemans, "Stochastic block model for hypergraphs: statistical limits and a semidefinite programming approach," arXiv:1807.02884, 2018.
- [19] J. Gaudio, and N. Joshi, "Community detection in the hypergraph SBM: optimal recovery given the similarity matrix," arXiv:2208.12227, 2022.
- [20] H. Wang, "Strong consistency and optimality of spectral clustering in symmetric binary non-uniform hypergraph stochastic block model," arXiv:2306.06845, 2023.
- [21] I. Dumitriu and H. Wang, "Exact recovery for the non-uniform hypergraph stochastic block model," arXiv:2304.13139, 2023.
- [22] T. Carletti, D. Fanelli, and R. Lambiotte, "Random walks and community detection in hypergraphs," *J. Phys. Complex*, vol. 2, no. 1, 2021, Art. no. 015011.
- [23] E. Abbe, "Community detection and stochastic block models: recent developments," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6446–6531, 2017.
- [24] J. Friedman, J. Kahn, and E. Szemerédi, "On the second eigenvalue of random regular graphs," in *Proc. 21st Annu. ACM Symp. Theory Comput.*, 1989, pp. 587–598.

- [25] U. Feige, and E. Ofek, “Spectral techniques applied to sparse random graphs,” *Random Struct. Algorithms*, vol. 27, no. 2, pp. 251–275, 2005.
- [26] A. Kumar, Y. Sabharwal, and S. Sen, “A simple linear time $(1+\epsilon)$ -approximation algorithm for k -means clustering in any dimensions,” *45th Annual IEEE Symp. Found. Comput. Sci.*, 2004, pp. 454–462.
- [27] D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [28] J. A. Tropp, “An introduction to matrix concentration inequalities,” *Found. Trends Mach. Learn.*, vol. 8, no. 1, pp. 1–230, 2015.
- [29] X. Mao, P. Sarkar, and D. Chakrabati, “Estimating mixed memberships with sharp eigenvector deviations,” *J. Amer. Stat. Assoc.*, vol. 116, no. 536, pp. 1928–1940, 2021.
- [30] Y. Chen, Y. Chi, J. Ffan, and C. Ma, “Spectral methods for data science: a statistical perspective,” *Found. Trends Mach. Learn.*, vol. 14, no. 5, pp. 566–806, 2021.

Chong Deng received the B.Sc. degree in applied mathematics from Shanghai University in 2020. He is currently working toward the M.S. degree in applied mathematics at Shanghai University. His research interests are the theory and applications of stochastic block models.

Xin-Jian Xu received the B.Sc. and Ph.D. degrees in theoretical physics from Lanzhou University in 2001 and 2005 respectively. He is currently a professor in Qianweichang College at Shanghai University. His research interests include statistical analysis of complex networks and its applications in society and biology.

Shihui Ying (Member, IEEE) received the B.Eng. degree in mechanical engineering and the Ph.D. degree in applied mathematics from Xi’an Jiaotong University in 2001 and 2008 respectively. He is currently a professor in the Department of Mathematics at Shanghai University. He was a postdoctor in Biomedical Research Imaging Center (BRIC) at University of North Carolina at Chapel Hill from 2012 to 2013. His research interests cover geometric theory and methods for machine learning with applications in medical image analysis.