

CONSISTENCY OF SPECTRAL HYPERGRAPH PARTITIONING UNDER PLANTED PARTITION MODEL

DEBARGHYA GHOSHDASTIDAR, AMBEDKAR DUKKIPATI

ABSTRACT. Hypergraph partitioning lies at the heart of a number of problems in machine learning and network sciences. Many algorithms for hypergraph partitioning have been proposed that extend standard approaches for graph partitioning to the case of hypergraphs. However, theoretical aspects of such methods have seldom received attention in the literature as compared to the extensive studies on the guarantees of graph partitioning. For instance, consistency results of spectral graph partitioning under the stochastic block model are well known. In this paper, we present a planted partition model for sparse random non-uniform hypergraphs that generalizes the stochastic block model. We derive an error bound for a spectral hypergraph partitioning algorithm under this model using matrix concentration inequalities. To the best of our knowledge, this is the first consistency result related to partitioning non-uniform hypergraphs.

1. INTRODUCTION

A wide variety of complex real-world systems can be understood by analyzing the interactions among various entities or components of the system. This has made network analysis a subject of both theoretical and practical interest. A plethora of challenging problems related to social, biological, communication networks have intrigued researchers over the past decades, and has led to the development of some sophisticated techniques for network analysis. This is clearly witnessed in the problems related to network or graph partitioning, where the task is to find strongly connected groups of nodes with sparse connections across groups. The problem appears in several engineering applications such as circuit or program segmentation (Kernighan and Lin, 1970), community detection in social or biological networks (Wasserman, 1994; Guimera and Amaral, 2005), data analysis and clustering (Ng et al., 2002) among others.

Graph partitioning and the stochastic block model. The problem of finding a balanced partition of a graph is known to be computationally hard. However, a number of approximate methods have been studied in the literature. These include spectral algorithms (Fiedler, 1973; Ng et al., 2002; Krzakala et al., 2013), modularity and likelihood based methods (Girvan and Newman, 2002; Bickel and Chen, 2009; Choi et al., 2012), convex optimization (Amini and Levina, 2014; Chen et al., 2014), belief propagation (Decelle et al., 2011) among others. The empirical success of such methods is not a mere coincidence, and theoretical guarantees for most of these methods have been extensively studied. In this respect, it is quite common to study partitioning algorithms under statistical models for random networks, such

as the stochastic block model or planted partition model (Holland et al., 1983) or its variants. In this model, one considers a random graph on n nodes with a well defined k -way partition $\psi : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$. The edges are randomly added with probabilities depending on the class labels of the participating nodes. Thus, the following interesting question arises:

Question: Let ψ' be the partition obtained from an algorithm, then what is the number of mismatches between ψ and ψ' ?

One typically asks for a high probability bound on the above error in terms of n . Such error bounds have been established for a variety of partitioning algorithms including aforementioned approaches. Chen et al. (2014) compare the theoretical guarantees for different approaches.

In the case of random graphs under stochastic block model, analysis of a spectral algorithm was first considered by McSherry (2001). However, the popular variant of spectral graph partitioning, commonly known as *spectral clustering*, was studied only in recent times (Rohe et al., 2011; Lei and Rinaldo, 2015). It is now known that for a planted graph with $\Omega(\ln n)$ minimum node degree, spectral clustering achieves an $o(n)$ error rate. One commonly refers to this property as the *weak consistency* of spectral clustering. This not the best known error rate as exact recovery of the partitions are known to be possible using other approaches (Amini and Levina, 2014). Recent results (Vu, 2014; Lei and Zhu, 2014; Gao et al., 2015) show that an additional refinement process can improve the partitioning of spectral clustering to exactly recover the partitions, thereby achieving *strong consistency*. The condition on minimum node degree can be also relaxed by considering alternative spectral techniques (Krzakala et al., 2013; Le et al., 2015), and these algorithms can detect partitions in sparse random graphs that are close to the algorithmic barrier for community detection (Decelle et al., 2011).

Hypergraph partitioning. In spite of the vast applicability of network modeling and analysis, there exists more complex scenarios, where pairwise interactions cannot accurately model the system of interest. A common example is folksonomy, where individuals annotate online resources, such as images or research papers. Such problems appear to have a tri-partite structure in form of “user–resource–annotation”, and is naturally represented as a 3-uniform hypergraph (Ghoshal et al., 2009), where each edge connects three nodes. Earlier works in data mining (Gibson et al., 2000) as well as in computer vision (Govindu, 2005) have also demonstrated the necessity of uniform hypergraphs. Moreover, in large scale circuit design (Karypis and Kumar, 2000) and molecular interaction networks (Michoel and Nachtergaele, 2012), one needs to consider group interactions that is appropriately modeled by a non-uniform hypergraph.

The current work focuses on hypergraph partitioning that appears in various applications such as circuit partitioning (Schweikert and Kernighan, 1979), categorical data clustering (Gibson et al., 2000), geometric grouping (Govindu, 2005) and others. Various partitioning techniques are used in practice including move-based algorithms (Schweikert and Kernighan, 1979; Karypis and Kumar, 2000), spectral

algorithms (Rodríguez, 2002; Zhou et al., 2007), tensor based methods (Govindu, 2005) etc.

Hypergraph partitioning and related problems have been of theoretical interest for quite some time (Berge, 1984). While early works on hypergraph partitioning studied various properties of hypergraph cuts (Bolla, 1993; Chung, 1992), more recent results provide insights into the algebraic connectivity and chromatic numbers of hypergraphs (Hu and Qi, 2012; Cooper and Dutle, 2012). However, till date, little is known about the theoretical guarantees of hypergraph partitioning methods that are popular amongst practitioners. The primary reason for this lack of results, at least in a stochastic framework, is due to the absence of random models of non-uniform hypergraphs that can incorporate a planted structure. Planted structures in uniform hypergraphs have been studied in context of hypergraph coloring (Chen and Frieze, 1996), and in a more general setting in (Ghoshdastidar and Dukkipati, 2014), where almost sure error bounds are derived for partitioning planted uniform hypergraphs via tensor decomposition. But, extension of such models or analysis to non-uniform hypergraphs does not follow directly.

Contributions. The primary focus of this paper is to derive an error bound for a hypergraph partitioning algorithm that solves a spectral relaxation of the normalized hypergraph cut problem. This is achieved in the form of a two-fold contribution.

We present a model for generating random hypergraphs with a planted solution. Extensions of the Erdős-Rényi model to non-uniform hypergraphs have been studied in the literature (Schmidt-Pruzan and Shamir, 1985; Darling and Norris, 2005), where, for each m , the probability of generating edges of size m is controlled by a parameter p_m . The recent work of Stasi et al. (2014) present a similar model, but with a specified degree sequence. Such models implicitly suggest that one can consider a non-uniform hypergraph as a collection of m -uniform hypergraphs for varying m . Thus, it is possible to construct planted models for non-uniform hypergraphs from a collection of uniform hypergraph models. Based on this idea, we present a planted hypergraph model that naturally extends the sparse stochastic block model for graphs (Lei and Rinaldo, 2015), and also encompasses previously studied models for uniform hypergraphs (Ghoshdastidar and Dukkipati, 2014).

We consider a popular spectral algorithm for hypergraph partitioning, and derive a bound on the number of nodes incorrectly assigned by the algorithm under the above model. We prove that for random planted hypergraphs with minimum node degree above a certain threshold, the algorithm is weakly consistent in general. However, the algorithm can also exactly recover the partitions from dense hypergraphs without any subsequent refinement procedure. Our analysis relies on an alternative characterization of the incidence matrix of the random hypergraph, and the use of matrix concentration inequalities (Chung and Radcliffe, 2011; Tropp, 2012).

Typically, spectral partitioning algorithms involve a post-processing stage of distance based clustering. Though the k -means algorithm (Lloyd, 1982) or its approximate variants (Kumar et al., 2004; Ostrovsky et al., 2012) are the popular choice in practice, such algorithms are not always guaranteed to provide good clustering. Gao

et al. (2015) discusses the implication of this drawback on the consistency results for spectral clustering (Lei and Rinaldo, 2015). On the other hand, we establish that under certain conditions, the approximate k -means algorithm of (Ostrovsky et al., 2012) indeed provides a good clustering with very high probability.

Finally, we consider special cases of the planted model. We comment on the allowable model parameters, and illustrate their effect on the derived error bound. Numerical studies reveal the practical significance of spectral hypergraph partitioning as well as the applicability of our analysis.

Organization. We first describe the spectral hypergraph partitioning algorithm under consideration in Section 2. Section 3 describes the model for random hypergraphs with a planted partition. We provide the main consistency result in Section 4, followed by a series of examples of planted models studied in Section 5. Section 6 contains experimental results that validate our model and analysis, and Section 7 presents the concluding remarks. The proofs of the technical lemmas can be found in the appendices.

Notations. Some of the notations that is often used in this paper are mentioned here. $\mathbb{1}\{\cdot\}$ is the indicator function and $\ln(\cdot)$ refers to the natural logarithm. $\mathbf{E}[\cdot]$ denotes expectation with respect to the distribution of the planted model. For a matrix A , we use A_i to refer to the i^{th} row of A and $A_{\cdot i}$ refers to its i^{th} column. $\|\cdot\|_2$ denotes the Euclidean norm for vectors and the spectral norm for matrices, while $\|\cdot\|_F$ denotes the Frobenius norm. We sometimes compute standard matrix functions like $\text{Trace}(\cdot)$ and $\det(\cdot)$. In addition, we also use asymptotic notations $O(\cdot), o(\cdot), \Omega(\cdot)$ etc., where we view these quantities as functions of the number of nodes n .

2. SPECTRAL HYPERGRAPH PARTITIONING

A hypergraph is defined as a tuple $(\mathcal{V}, \mathcal{E})$, where \mathcal{V} is a set of objects and \mathcal{E} is a collection of subsets of \mathcal{V} . Though early works in combinatorics viewed this structure purely as a set system, it was soon realized that one may view \mathcal{V} as a set of nodes and every element of \mathcal{E} as an edge (or connection) among a subset of nodes. As noted in (Berge, 1984), such a generalization of graphs helps to simplify several combinatorial results in the graph literature. A hypergraph is said to be r -uniform if every edge $e \in \mathcal{E}$ contains exactly r nodes.

In this paper, we assume that there are no edges of size 0 or 1 as they do not convey any information in a partitioning framework. We also assume that the hypergraph is undirected, *i.e.*, there is no ordering of nodes in any edge. Under this setting, the most simple representation of a hypergraph is in terms of its incidence matrix $H \in \{0, 1\}^{|\mathcal{V}| \times |\mathcal{E}|}$, where $H_{ve} = 1$ if the node v is contained in the edge e , and 0 otherwise. One can note that the degree of any node v can be written as $\deg(v) = \sum_{e \in \mathcal{E}} H_{ve}$, which is simply the sum of the v^{th} row of H . Similarly, the cardinality of any edge e is $|e| = \sum_{v \in \mathcal{V}} H_{ve}$.

Several notions of hypergraph cut and hypergraph Laplacian have been proposed in the literature (Bolla, 1993; Chung, 1992; Rodríguez, 2002) that generalize the standard notions well studied in the graph literature. In this work, we consider the generalization studied in (Zhou et al., 2007). Let $\mathcal{V}_1 \subset \mathcal{V}$, then $\text{vol}(\mathcal{V}_1) = \sum_{v \in \mathcal{V}_1} \deg(v)$ is called the volume of \mathcal{V}_1 , while the boundary of \mathcal{V}_1 , defined as $\partial\mathcal{V}_1 = \{e \in \mathcal{E} : e \cap \mathcal{V}_1 \neq \emptyset, e \cap \mathcal{V}_1^c \neq \emptyset\}$, denotes the set of edges that are cut when the nodes are divided into \mathcal{V}_1 and $\mathcal{V} \setminus \mathcal{V}_1$. The volume of $\partial\mathcal{V}_1$ is defined as

$$\text{vol}(\partial\mathcal{V}_1) = \sum_{e \in \partial\mathcal{V}_1} \frac{|e \cap \mathcal{V}_1| |e \cap \mathcal{V}_1^c|}{|e|}.$$

We consider the problem of partitioning the vertex set \mathcal{V} into k disjoint sets, $\mathcal{V}_1, \dots, \mathcal{V}_k$, that minimizes the normalized hypergraph cut

$$\text{NH-cut}(\mathcal{V}_1, \dots, \mathcal{V}_k) = \sum_{j=1}^k \frac{\text{vol}(\partial\mathcal{V}_j)}{\text{vol}(\mathcal{V}_j)}. \quad (1)$$

One can observe that for graphs, the above definition (1) retrieves the standard notion of a normalized cut (von Luxburg, 2007). Zhou et al. (2007) also define the notion of a normalized hypergraph Laplacian matrix $L \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ given by

$$L = I - D^{-1/2} H \Delta^{-1} H^T D^{-1/2}, \quad (2)$$

where the matrices $D \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$, $\Delta \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{E}|}$ are diagonal with $D_{vv} = \deg(v)$ and $\Delta_{ee} = |e|$. A simple calculation shows that the problem of minimizing the quantity in (1) is equivalent to the problem:

$$\underset{\mathcal{V}_1, \dots, \mathcal{V}_k}{\text{minimize}} \quad \text{Trace} \left(\widehat{X}^T L \widehat{X} \right), \quad (3)$$

where $\widehat{X} \in \mathbb{R}^{|\mathcal{V}| \times k}$ is such that $\widehat{X}_{vj} = \sqrt{\frac{\deg(v)}{\text{vol}(\mathcal{V}_j)}} \mathbb{1}\{v \in \mathcal{V}_j\}$, and satisfies $\widehat{X}^T \widehat{X} = I$. Since the optimization in (3) is NP-hard, one relaxes the problem by minimizing over all $X \in \mathbb{R}^{|\mathcal{V}| \times k}$ with orthonormal columns. It is well known that the solution to this relaxed problem is the matrix of k leading orthonormal eigenvectors of L . Note that L is a positive semi-definite matrix with at least one eigenvalue equal to zero. The term ‘‘leading eigenvectors’’ refers to the eigenvectors that correspond to the k smallest eigenvalues of L .

The above discussion motivates a spectral k -way partitioning approach based on minimizing NH-cut. The method is listed in Algorithm 1. The form of Laplacian matrix in (2) also suggests that the problem of minimizing NH-cut may be alternatively expressed as the problem of partitioning a graph with weighted adjacency matrix

$$A = H \Delta^{-1} H^T. \quad (4)$$

Such a graph is related to the star expansion of the hypergraph (Agarwal et al., 2006).

The intuition behind the k -means step in Algorithm 1 is as follows. If the solution of the spectral relaxation results in $X = \widehat{X}$, where \widehat{X} is defined as in (3), then after row normalization, \overline{X} corresponds to a binary matrix with exactly one non-zero term

in each row. Hence, one obtains the partitions desired in (3) by performing k -means on the rows of \bar{X} . In this paper, we assume that the approximate k -means method of (Ostrovsky et al., 2012) is used, that provides a near optimal solution in a single iteration.

Algorithm 1 Spectral Hypergraph Partitioning Algorithm

input Incidence matrix H of the hypergraph.

- 1: Compute the hypergraph Laplacian L as in (2).
- 2: Compute the leading eigenvector matrix $X \in \mathbb{R}^{|\mathcal{V}| \times k}$.
- 3: Normalize rows of X to have unit norm. Call this matrix \bar{X} .
- 4: Run k -means on the rows of \bar{X} .

output Partition of \mathcal{V} that corresponds to the clusters obtained from k -means.

3. PLANTED PARTITION IN RANDOM HYPERGRAPHS

In the rest of the paper, we study the error incurred by Algorithm 1. For this, we consider a model for generating random hypergraphs with a planted solution.

3.1. The model. Let $\mathcal{V} = \{1, 2, \dots, n\}$ be a set of nodes, and let $\psi : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, k\}$ be a partition of the nodes into k classes. Here, ψ is the unknown planted partition that one needs to extract from a hypergraph generated on \mathcal{V} . For a node i , we denote its class by ψ_i . Let $M \geq 2$ be an integer, representing the range or the maximum edge cardinality in the hypergraph. In view of practical situations, we allow both k and M to vary with n though this dependence is not made explicit in the notation. One may set $M = n$ to allow occurrence of all possible edges, but in practice, one can assume that $M = O(\ln n)$. Also, for each n and for each $m = 2, \dots, M$, let $\alpha_{m,n} \in [0, 1]$, and $B^{(m)} \in [0, 1]^{k \times k \times \dots \times k}$ be a symmetric k -dimensional tensor of order m .

A random hypergraph on \mathcal{V} is generated as follows. For each $m = 2, \dots, M$, and for every set $\{i_1, i_2, \dots, i_m\} \subset \mathcal{V}$, an edge is included independently with probability $\alpha_{m,n} B_{\psi_{i_1} \psi_{i_2} \dots \psi_{i_m}}^{(m)}$. This process generates a random hypergraph of maximum edge cardinality M . The tensor $B^{(m)}$ contains the probabilities of forming m -way edges among the different classes if $\alpha_{m,n} = 1$. On the other hand, $\alpha_{m,n}$ allows for a sparsity scaling that does not depend on the partitions. In the case of sparse graphs, $\alpha_{2,n}$ regulates the edge density. However, in real-world non-uniform hypergraphs, one often finds that the density of 2 or 3-way edges is much more than edges of larger size (say, 10). To account for this generality, we allow $\alpha_{m,n}$ to vary both with m and n . For instance, if $\alpha_{2,n} = 1$ and $\alpha_{m,n} = \frac{1}{n^{m-1}}$ for all $m > 2$, then the generated hypergraph contains $O(n^2)$ number of 2-way edges, but only $O(n)$ number of m -way edges for every $m > 2$.

As a special case, note that for graphs, $M = 2$ for all n , and the model corresponds to the sparse stochastic block model, where an edge (u, v) is formed with probability $\alpha_{2,n} B_{\psi_u \psi_v}^{(2)}$. In other words, if $Z \in \{0, 1\}^{n \times k}$ denotes the assignment matrix, then the

probability of edge (u, v) is same as the corresponding entry of $\alpha_{2,n} Z B^{(2)} Z^T$. For r -uniform uniform hypergraphs, one has $\alpha_{m,n} = 0$ for all $m \neq r$. Ghoshdastidar and Dukkipati (2014) considered a dense uniform hypergraph, *i.e.*, $\alpha_{r,n} = 1$, and edge probabilities specified by a r^{th} -order k -dimensional tensor $B^{(r)}$. It was shown that the population adjacency tensor can be expressed in terms of $B^{(r)}$ and Z .

The intuition behind the above described model is that one may view a hypergraph of range M as a collection of uniform hypergraphs of orders $m = 2, \dots, M$. In the random setting, each m -uniform hypergraph is specified in terms of $\alpha_{m,n}$ and $B^{(m)}$. The above model can be easily extended to directed hypergraphs, and also to the case of weighted hypergraphs.

3.2. The random hypergraph Laplacian. For stochastic block model, a random instance of a graph is specified by its $n \times n$ adjacency matrix. However, for random hypergraphs, the size of the incidence matrix H is a random quantity as it depends on the number of generated edges. Since this poses difficulties in working with the form of hypergraph Laplacian in (2), we present an alternative representation. The Laplacian can be written as

$$L = I - \sum_{e \in \mathcal{E}} \frac{1}{|e|} D^{-1/2} a_e a_e^T D^{-1/2}, \quad (5)$$

where for $e \subset \mathcal{E}$, $a_e \in \{0, 1\}^n$ with $(a_e)_i = 1$, if node $i \in e$, and 0 otherwise.

Let $\beta_M = \sum_{m=2}^M \binom{n}{m}$. Note that β_M is the maximum number of edges the hypergraph can contain given the fact that its range is M . For convenience, we define a bijective map $\xi : \{1, 2, \dots, \beta_M\} \rightarrow \{e \subset \mathcal{V} : 2 \leq |e| \leq M\}$, where each ξ_j refers to a subset of nodes, *i.e.*, a possible edge in the given hypergraph. Then the Laplacian can be expressed as

$$L = I - \sum_{j=1}^{\beta_M} \frac{\mathbb{1}\{\xi_j \in \mathcal{E}\}}{|\xi_j|} D^{-1/2} a_{\xi_j} a_{\xi_j}^T D^{-1/2}, \quad (6)$$

where the summation is over all possible edges of size at most M , but the missing edges do not contribute to the sum. Similarly, one can express the degree matrix D as

$$D_{ii} = \deg(i) = \sum_{e \in \mathcal{E}} (a_e)_i = \sum_{j=1}^{\beta_M} \mathbb{1}\{\xi_j \in \mathcal{E}\} (a_{\xi_j})_i. \quad (7)$$

The above representation corresponds to an ‘extended’ version of the incidence matrix as $\bar{H} \in \{0, 1\}^{n \times \beta_M}$, whose j^{th} column is $\mathbb{1}\{\xi_j \in \mathcal{E}\} a_{\xi_j}$, *i.e.*, \bar{H} contains the columns of H with additional zero columns inserted to account for missing edges. This holds for any hypergraph of range M defined on the set \mathcal{V} . We use this representation to keep the number of columns as a deterministic quantity. We now discuss how the described planted partition model for hypergraphs, with maximum edge size M , can be expressed in terms of the extended incidence matrix $\bar{H} \in \{0, 1\}^{n \times \beta_M}$. Let h_j , $j = 1, 2, \dots, \beta_M$ be independent Bernoulli random variables that indicate the

presence of the edge $\xi_j \subset \mathcal{V}$. By description of the model, if $\xi_j = \{i_1, i_2, \dots, i_{m_j}\}$ for some $m_j \in \{2, \dots, M\}$, then the random variable $h_j \sim \text{Bernoulli}\left(\alpha_{m_j, n} B_{\psi_{i_1} \psi_{i_2} \dots \psi_{i_{m_j}}^{(m_j)}}\right)$. The j^{th} column of \overline{H} is $h_j a_{\xi_j}$, and hence, the Laplacian matrix for the random hypergraph is

$$L = I - \sum_{j=1}^{\beta_M} \frac{h_j}{|\xi_j|} D^{-1/2} a_{\xi_j} a_{\xi_j}^T D^{-1/2}, \quad \text{where } D_{ii} = \sum_{j=1}^{\beta_M} h_j (a_{\xi_j})_i. \quad (8)$$

At this stage, we note that the above matrices depend on the number of nodes n . For ease of notation, we do not explicitly mention this dependence.

4. CONSISTENCY OF SPECTRAL HYPERGRAPH PARTITIONING

This section presents the main result of this paper that gives a bound on the error incurred by the spectral hypergraph partitioning algorithm described in Algorithm 1. Let $\psi' : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$ denote the labels obtained from the algorithm. The partitioning error is given by the number of nodes incorrectly assigned by Algorithm 1, *i.e.*,

$$\text{Err}(\psi, \psi') = \min_{\sigma} \sum_{i=1}^n \mathbf{1}\{\psi_i \neq \sigma(\psi'_i)\}, \quad (9)$$

where the minimum is taken over all permutation σ of labels. We show that if **(i)** the partitions are *identifiable*, and **(ii)** the hypergraph is not too *sparse*, then indeed $\text{Err}(\psi, \psi')$ is bounded by a quantity that is at most sub-linear in n . Furthermore, the bound holds with probability $(1 - o(1))$. This immediately implies that Algorithm 1 is *weakly consistent*. However, we show later that for particular model parameters, Algorithm 1 can even recover the partitions exactly, *i.e.*, $\text{Err}(\psi, \psi') = o(1)$.

4.1. The main result. The consistency result studied in this paper is quite similar, in spirit, to those studied in the case of stochastic block model for graphs. In such a case, one typically analyzes the population version of a spectral algorithm, and then uses the fact that the spectral properties of the Laplacian eventually concentrates around those of the population Laplacian.

From this point of view, we consider the population version of the hypergraph Laplacian (8) defined as

$$\mathcal{L} = I - \sum_{j=1}^{\beta_M} \frac{\mathbb{E}[h_j]}{|\xi_j|} \mathcal{D}^{-1/2} a_{\xi_j} a_{\xi_j}^T \mathcal{D}^{-1/2}, \quad (10)$$

where \mathcal{D} is the expected degree matrix, *i.e.*, $\mathcal{D}_{ii} = \sum_{j=1}^{\beta_M} \mathbb{E}[h_j] (a_{\xi_j})_i$. We also define the quantity $d = \min_{i \in \{1, \dots, n\}} \mathcal{D}_{ii}$. Without loss of generality, we may also assume that for a given n , the community sizes are $n_1 \geq n_2 \geq \dots \geq n_k$

Before stating the main result, it is useful to elaborate on the aforementioned conditions under which the derived error bound holds. A lower bound on the sparsity of the hypergraph is a standard requirement to ensure that the concentration of the spectral properties eventually hold, and has been often used in the graph literature (Lei and Rinaldo, 2015; Le et al., 2015). In our setting, this can be stated in terms of the sparsity factors $\alpha_{m,n}$, or more simply, in terms of the minimum expected degree d , that grows with n but at a rate controlled by the sparsity factors.

A more critical condition is the identifiability of the partitions. Note that the definition of the hypergraph Laplacian essentially implies that the hypergraph is reduced to a graph with self loops. Hence, the performance of Algorithm 1 crucially depends on the identifiability of the partitions from \mathcal{L} , or rather from this reduced graph with the population adjacency matrix

$$\mathcal{A} = \mathbb{E}[A] = \sum_{j=1}^{\beta_M} \frac{\mathbb{E}[h_j]}{|\xi_j|} a_{\xi_j} a_{\xi_j}^T.$$

The following result provides a characterization of \mathcal{L} and \mathcal{A} , which in turn helps to quantify the condition for identifiability of the partitions from \mathcal{L} .

Lemma 4.1. *Let $Z \in \{0, 1\}^{n \times k}$ denote the assignment matrix corresponding to the partition ψ . Then the population hypergraph Laplacian is given by*

$$\mathcal{L} = I - \mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}, \quad (11)$$

where \mathcal{A} can be expressed as

$$\mathcal{A} = ZGZ^T - J. \quad (12)$$

Here, $J \in \mathbb{R}^{n \times n}$ is diagonal with $J_{ii} = J_{jj}$ whenever $\psi_i = \psi_j$, and $G \in \mathbb{R}^{k \times k}$.

Furthermore, \mathcal{L} contains k eigenvalues for which the corresponding orthonormal eigenvectors are the columns of the matrix $\mathcal{X} = Z(Z^T Z)^{-1/2} U$, where $U \in \mathbb{R}^{k \times k}$ is orthonormal.

The representation in (12) shows that \mathcal{A} is essentially of rank k , except for the diagonal entries. Owing to the first term in (12), one does expect \mathcal{L} to have k eigenvectors whose entries are constant in each community. As discussed later, a close inspection of \mathcal{X} reveals that indeed the columns of \mathcal{X} satisfy this property. Thus, if the spectral stage of Algorithm 1 can extract \mathcal{X} , then zero error can be achieved from the k -means step.

In general, \mathcal{X} need not correspond to leading eigenvectors \mathcal{L} (as computed in Algorithm 1). This is true even for certain types of graphs, for instance k -colorable graphs (Alon and Kahale, 1997). This effect is more pronounced in non-uniform hypergraphs due to the presence of a large number of model parameters. To account for this factor, we define the following quantity

$$\delta = \left(\lambda_{\min}(G) \min_{1 \leq i \leq n} \frac{n_{\psi_i}}{\mathcal{D}_{ii}} \right) - \max_{1 \leq i, j \leq n} \left| \frac{J_{ii}}{\mathcal{D}_{ii}} - \frac{J_{jj}}{\mathcal{D}_{jj}} \right|, \quad (13)$$

where n_{ψ_i} is the size of the community in which node i belongs. We show that if $\delta > 0$, then the columns of \mathcal{X} are the k leading eigenvectors of \mathcal{L} . Here, $\lambda_{\min}(G)$

refers to the smallest eigenvalue of G . Thus, we can state the consistency result for Algorithm 1 as below.

Theorem 4.2. *Consider a random hypergraph on n nodes generated according to the planted partition model described in Section 3. Assume that n is sufficiently large, and the size of the k partitions are $n_1 \geq n_2 \geq \dots \geq n_k$. Let d be the minimum expected degree, and δ be the quantity defined in (13).*

There exists an absolute constant $C > 0$, such that, if $\delta > 0$ and

$$d > C \frac{kn_1(\ln n)^2}{\delta^2 n_k} \quad (14)$$

then with probability at least $1 - O((\ln n)^{-1/4})$,

$$\text{Err}(\psi, \psi') = O\left(\frac{kn_1 \ln n}{\delta^2 d}\right). \quad (15)$$

Note here that the quantities δ, d and k can vary with n . On substituting the condition on d into (14), one can see that $\text{Err}(\psi, \psi') = o(n)$ with probability $(1 - o(1))$. Hence, Algorithm 1 is weakly consistent if the conditions of the theorem are satisfied. However, we show later that in certain dense hypergraphs, the bound in (15) may eventually decay to zero. Thus, Algorithm 1 is guaranteed to exactly recover the communities in such cases.

In Section 5, we consider particular instances of the planted model, and illustrate the dependence of the above result on the model parameters. For instance, (14) implies that the result holds if the sparsity factor ($\alpha_{m,n}$) is above a certain threshold (see Corollaries 5.1 and 5.2). Even when (14) holds, higher error is incurred for a sparse hypergraph (small d) or when the number of communities k is large.

One may note that $\delta > 0$ is the condition for identifiability of the partitions, and is essential for success of the algorithm. Typically, one does find that $\delta \downarrow 0$ as $n \rightarrow \infty$. To this end, the condition (14) implies that δ cannot decay rapidly as $\delta^2 d$ needs to maintain a minimum growth rate. We also note that δ quantifies identifiability of the partitions and $\text{Err}(\psi, \psi')$ varies as $\frac{1}{\delta^2}$. Hence, if the model parameters are such that δ is small, for instance if the probability of inter-community edges is very close to that of within community edges, then $\text{Err}(\psi, \psi')$ is larger.

Before presenting the proof of Theorem 4.2, we comment on the assumption of sufficiently large n . Note that the sole purpose of this assumption is to ensure the success of the k -means algorithm. Later, in the proof, we establish that if n is large enough, the condition (14) ensures that the approximate k -means method of Ostrovsky et al. (2012) provides a near optimal solution, which is worse by only a constant factor. Earlier works on spectral graph partitioning (Rohe et al., 2011; Lei and Rinaldo, 2015) assumed the existence of such a near optimal solution with probability 1. To demonstrate the effect of such an assumption, we state the following result, which is a modification of Theorem 4.2 under the above assumption.

Corollary 4.3. *Consider a random hypergraph on n nodes generated according to the planted partition model, and let the other quantities be as defined in Theorem 4.2.*

Assume that for a constant $\gamma > 1$, there is a γ -approximate¹ k -means algorithm that succeeds with probability 1.

There exists an absolute constant $C > 0$, such that, if $\delta > 0$ and

$$d > C \frac{\ln n}{\delta^2} \tag{16}$$

then with probability at least $1 - \frac{4}{n^2}$,

$$\text{Err}(\psi, \psi') = O\left(\frac{kn_1 \ln n}{\delta^2 d}\right). \tag{17}$$

The result reveals that if a good k -means algorithm is available, then the success probability of Algorithm 1 increases, and the result is also applicable for more sparse hypergraph since the condition (16) is weaker than (14). However, based on the existing results in the k -means literature, one should consider the following remark.

Remark. If the data satisfies certain *clusterability* criterion², then the efficient variants of k -means (Kumar et al., 2004; Ostrovsky et al., 2012) provide a γ -approximate solution with a constant probability $\rho < 1$. Both γ and ρ depend on various factors including k , clusterability criterion etc.

In view of the above remark, Corollary 4.3 is too optimistic. Recently, Gao et al. (2015) pointed that if one uses the method of Kumar et al. (2004), then γ grows with k . In addition, one should also note that the success probability of this method is $\rho = c^k$ for an absolute constant $c \in (0, 1)$. Hence, a spectral partitioning algorithm using this method cannot succeed with probability $(1 - o(1))$. Instead, we use the method of Ostrovsky et al. (2012) to achieve a higher success rate as stated in Theorem 4.2. The only additional assumption is that of sufficiently large n . We note that this requirement, along with condition (14), can be relaxed if one only aims for a constant success probability. This is shown in the following modification of Theorem 4.2, where we assume that the k -means algorithm of Ostrovsky et al. (2012) is used.

Corollary 4.4. *Consider a random hypergraph on n nodes generated according to the planted partition model, and let the other quantities be as defined in Theorem 4.2.*

There exist absolute constants $C > 0$ and $\epsilon \in (0, 0.015)$, such that, if $\delta > 0$ and

$$d > \frac{C kn_1 \ln n}{\epsilon^2 \delta^2 n_k} \tag{18}$$

then with probability at least $1 - O(\sqrt{\epsilon})$,

$$\text{Err}(\psi, \psi') = O\left(\frac{kn_1 \ln n}{\delta^2 d}\right). \tag{19}$$

¹Informally, a γ -approximate k -means methods returns a solution for which the objective of the k -means problem is at most γ times the global minimum, where $\gamma > 1$. A formal definition is postponed to (22) and subsequent discussions.

²Various clusterability criteria have been studied in the literature. In this work, we consider the notion of ϵ -separability proposed by Ostrovsky et al. (2012).

4.2. Proof of Theorem 4.2. We now present an outline of the proof of Theorem 4.2 using a series of lemmas. The proofs for these lemmas are given in the appendix. The result is obtained by proving the following facts:

- (1) If Algorithm 1 is performed on the population Laplacian \mathcal{L} , then under the condition of $\delta > 0$, the obtained partitions are correct.
- (2) The deviation of L from \mathcal{L} is bounded above, and the bound holds with probability at least $(1 - \frac{4}{n^2})$.
- (3) As a consequence of above facts, the standard matrix perturbation bounds (Stewart and Sun, 1990) imply that the eigenvalues and the corresponding eigenspaces of L concentrate about those of \mathcal{L} .
- (4) If (14) holds, then k -means stage of Algorithm 1 succeeds in obtaining a near optimal solution with probability at least $1 - O((\ln n)^{-1/4})$.
- (5) The partitioning error can be expressed in terms of the above bounds, which leads to (15).

Corollaries 4.3 and 4.4 can be proved in similar manner. This is discussed in the appendix. We now prove the above facts. The following result extends Lemma 4.1.

Lemma 4.5. *If $\delta > 0$, then the k leading orthonormal eigenvectors of \mathcal{L} correspond to the columns of the matrix $\mathcal{X} = Z(Z^T Z)^{-1/2}U$.*

In the above result, $Z^T Z$ is a diagonal matrix with entries being the sizes of the k partitions. Hence, both $Z^T Z$ and U are of the rank k . Due to this, one can observe that the matrix \mathcal{X} contains exactly k distinct rows, each corresponding to a particular partition, *i.e.*, if A_i denotes i^{th} row of a matrix A , then for any two nodes $i, j \in \mathcal{V}$,

$$\mathcal{X}_i = \mathcal{X}_j \iff Z_i = Z_j \iff \psi_i = \psi_j .$$

Moreover, since U is orthonormal, the distinct rows of \mathcal{X} are orthogonal. Hence, after row normalization, the distinct rows correspond to k orthonormal vectors in \mathbb{R}^k , which can be easily clustered by k -means algorithm to obtain the true communities. Technically, δ is a lower bound on the eigen-gap between the k^{th} and $(k+1)^{\text{th}}$ smallest eigenvalues of \mathcal{L} . Since, it is difficult to obtain a simple characterization of the eigen gap, we resort to the use of δ as defined in (13).

Next, we bound the deviation of a random instance of L from the population Laplacian \mathcal{L} . This bound relies on the use of matrix Bernstein inequality (Chung and Radcliffe, 2011; Tropp, 2012). We note that for graphs, sharp deviation bounds have been used (Lei and Rinaldo, 2015), but such techniques cannot be directly extended to the case of hypergraphs.

Lemma 4.6. *If $d > 9 \ln n$, then with probability at least $(1 - \frac{4}{n^2})$,*

$$\|L - \mathcal{L}\|_2 \leq 12\sqrt{\frac{\ln n}{d}} . \tag{20}$$

We now use the principle subspace perturbation result due to (Lei and Rinaldo, 2015) to comment on the deviation of the leading eigenvectors of L from those of \mathcal{L} . A modified version of their result is proved that incorporates the row normalization

of the eigenvector matrix. Let X be the matrix of the k leading eigenvectors of L , and \bar{X} be its row normalized version. We have the following result. Note that since $\delta < 1$, the condition of Lemma 4.6 is subsumed by the condition stated below.

Lemma 4.7. *If $\delta > 0$ and $d > \frac{576 \ln n}{\delta^2}$, then there is an orthonormal matrix $Q \in \mathbb{R}^{k \times k}$ such that*

$$\|\bar{X} - ZQ\|_F \leq \frac{24}{\delta} \sqrt{\frac{2kn_1 \ln n}{d}} \quad (21)$$

with probability at least $(1 - \frac{4}{n^2})$.

We now derive a bound on the error incurred by the k -means step in the algorithm. Formally, k -means minimizes $\|\bar{X} - S\|_F$ over all $S \in \mathcal{M}_{n \times k}(k)$, where $\mathcal{M}_{n \times k}(r)$ is the set of all $n \times k$ matrices with at most r distinct rows. In practice, the rows of S correspond to the centers of obtained clusters. Achieving a global optimum for this problem is NP-hard. However, there are algorithms (Kumar et al., 2004; Ostrovsky et al., 2012) that can provide a solution S^* from the above class of matrices such that

$$\|\bar{X} - S^*\|_F \leq \gamma \min_S \|\bar{X} - S\|_F \quad (22)$$

for some $\gamma > 1$. The factor γ depends on the algorithm under consideration. For instance, γ grows with k in the case of (Kumar et al., 2004). On the other hand, Ostrovsky et al. (2012) showed that a constant factor approximation is possible if the data (rows of \bar{X} in our case) is *well-separated*.

To be precise, define $\eta_r(\bar{X})$ to be the minimum of the objective function when r clusters are found, *i.e.*,

$$\eta_r(\bar{X}) = \min_{S \in \mathcal{M}_{n \times k}(r)} \|\bar{X} - S\|_F. \quad (23)$$

The rows of \bar{X} is said to be ϵ -separated if $\eta_k(\bar{X}) \leq \epsilon \eta_{k-1}(\bar{X})$. Theorem 4.15 in (Ostrovsky et al., 2012) claims that if this condition holds for small enough ϵ , then the solution $S^* \in \mathcal{M}_{n \times k}(k)$ obtained from their approximate k -means algorithm satisfies (22) with probability $(1 - O(\sqrt{\epsilon}))$, where γ is given as $\gamma = \sqrt{\frac{1 - \epsilon^2}{1 - 37\epsilon^2}}$.

The following result shows that in our case, the rows of \bar{X} are indeed well-separated.

Lemma 4.8. *If the condition in (14) holds, then the rows of \bar{X} are ϵ -separated with $\epsilon = (\ln n)^{-1/2}$.*

As a consequence of Lemma 4.8, it follows that if n is sufficiently large, then the result of (Ostrovsky et al., 2012) holds. Moreover, one can also observe that for large n , we have $\gamma = O(1)$.

Finally, one needs to combine the above results in order to prove Theorem 4.2. For this, define the set $\mathcal{V}_{err} \subset \mathcal{V}$ as

$$\mathcal{V}_{err} = \left\{ i \in \mathcal{V} : \|S_i^* - Z_i \cdot Q\|_2 \geq \frac{1}{\sqrt{2}} \right\}. \quad (24)$$

Rohe et al. (2011) used a similar definition for the number of incorrectly assigned nodes, and discussed the intuition behind this definition. In the following result, we formally prove that the nodes that are not in \mathcal{V}_{err} are correctly assigned. We also provide an upper bound on the size of \mathcal{V}_{err} .

Lemma 4.9. *Let $i, j \notin \mathcal{V}_{err}$ and $S_i^* = S_j^*$, then $\psi_i = \psi_j$. As a consequence, $\text{Err}(\psi, \psi') \leq |\mathcal{V}_{err}|$. In addition,*

$$|\mathcal{V}_{err}| \leq 4(1 + \gamma^2) \|\bar{X} - ZQ\|_F^2.$$

Theorem 4.2 follows by combining the above bound with (21).

5. CONSISTENCY FOR SPECIAL CASES

We now study the implications of Theorem 4.2 for partitioning particular models of uniform and non-uniform hypergraphs. We also discuss the conditions for identifiability in special cases.

5.1. Balanced partitions in uniform hypergraph. Let the n nodes be divided into k groups such that each group contains $\frac{n}{k}$ nodes. We now consider a random r -uniform hypergraph on the nodes generated as follows. Let $p, q \in [0, 1]$ be constants with $(p + q) \leq 1$, and $\alpha_{r,n} \in (0, 1]$ be the sparsity factor dependent on n . For any r nodes from the same group, there is an edge among them with probability $\alpha_{r,n}(p + q)$. If all the r nodes do not belong to same group, then there is an edge with probability $\alpha_{r,n}q$.

In terms of the model in Section 3, one can see that $M = r$, and for all $m < r$, $\alpha_{m,n} = 0$. The r^{th} order k -dimensional tensor $B^{(r)}$ is given by

$$B_{j_1 j_2 \dots j_r}^{(r)} = \begin{cases} p + q & \text{if } j_1 = j_2 = \dots = j_r, \\ q & \text{otherwise.} \end{cases}$$

One can see that for $r = 2$, this model corresponds to the sparse stochastic block model considered in (Lei and Rinaldo, 2015) with balanced community sizes, and if $\alpha_{2,n} = 1$, one has the standard four parameter stochastic block model (Rohe et al., 2011). The following corollary to Theorem 4.2 shows the consistency of Algorithm 1.

Corollary 5.1. *In the above model,*

$$\delta = \frac{p\alpha_{r,n}n}{rkd} \binom{\frac{n}{k} - 2}{r - 2}, \quad (25)$$

and hence, the partitions are identifiable for all $p > 0$. Moreover, if

$$\alpha_{r,n} \geq C \frac{k^{2r-1} n (\ln n)^2}{\binom{n}{r}} \quad (26)$$

for some absolute constant $C > 0$, then the conditions in Theorem 4.2 are satisfied, and hence, we have

$$\text{Err}(\psi, \psi') = O\left(\frac{k^{2r-2} n^2 \ln n}{p^2 \alpha_{r,n} \binom{n}{r}}\right) = o(n) \quad (27)$$

with probability $(1 - o(1))$,

The lower bound on $\alpha_{r,n}$ mentioned in Corollary 5.1 needs some discussion. One can verify that in the above model, the expected number of edges lie in the range $[q\alpha_{r,n}\binom{n}{r}, (p+q)\alpha_{r,n}\binom{n}{r}]$, *i.e.*, it is about $\alpha_{r,n}\binom{n}{r}$ up to a constant scaling. The lower bound on $\alpha_{r,n}$ specifies that the number of edges must be at least $\Omega(k^{2r-1}n(\ln n)^2)$. This also indicates that for a larger r , more edges are required to ensure the error bound of Corollary 5.1. Since, $\alpha_{r,n} \leq 1$, one can see that the result is applicable for $k = O(n^{0.5-\epsilon})$ for all $\epsilon > \frac{1}{2(2r-1)}$. Even consistency results for graph partitioning require similar condition (Rohe et al., 2011; Choi et al., 2012).

A closer look at the condition (26) shows that if k is constant or increases slowly, $k = O(\ln n)$, then a sufficient condition for weak consistency of Algorithm 1 is $\alpha_{r,n} \geq C_r \frac{(\ln n)^{2r+1}}{n^{r-1}}$, where the constant C_r depends only on r . In case of graph partitioning, this level of sparsity is needed when one relies on matrix Bernstein inequality. However, recent results (Lei and Rinaldo, 2015) reduced the lower bound by using sharp concentration bounds for the binary adjacency matrix. Corollary 5.1 also indicates that if k increases at a higher rate, for example $k = n^a$, then consistency can be guaranteed only when hypergraph is more dense.

On the other extreme are uniform hypergraphs encountered in computer vision (Ghoshdastidar and Dukkipati, 2014, 2015) that are usually dense, *i.e.*, $\alpha_{r,n} = 1$. In this case, if $k = O(\ln n)$ then $\text{Err}(\psi, \psi') = O\left(\frac{(\ln n)^{2r-1}}{n^{r-2}}\right)$. Thus, the error decreases at a faster rate for r -uniform hypergraphs with larger r . In fact, for $r \geq 3$, above bound indicates that $\text{Err}(\psi, \psi') = o(1)$, *i.e.*, Algorithm 1 guarantees exact recovery of the partitions for large n .

Lastly, we discuss the effect of δ and the parameters p, q in this setting. Note that the case $q = 0$ is not interesting as there are no edges among different groups, and hence, the partition can be identified by a simple breadth-first search. On the other hand, $p = 0$ generates a random uniform hypergraph with all identical edges. Hence, the partitions cannot be identified in this case. This can also be seen from (25), where $\delta = 0$. In general, p denotes the gap between the probability of edge occurrence among nodes from same community and the probability with which nodes from different communities form an edge. Since δ is linear in p , one can observe from Theorem 4.2 that $\text{Err}(\psi, \psi')$ varies as $\frac{1}{p^2}$ with p . However, note that the model assumes that p does not vary with n , and may be treated as a constant in the asymptotic case.

5.2. Balanced partitions in non-uniform hypergraph. We now consider the case of non-uniform hypergraph of range M , where M may vary with n . As in Section 5.1, assume that n nodes are equally split into k groups. Also let $p, q \in (0, 1)$ such that $(p+q) \leq 1$, and for $m = 2, \dots, M$, let $B^{(m)}$ be the m^{th} -order symmetric k -dimensional tensor with

$$B_{j_1 j_2 \dots j_m}^{(m)} = \begin{cases} p+q & \text{if } j_1 = j_2 = \dots = j_m, \\ q & \text{otherwise.} \end{cases}$$

Setting $\alpha_{m,n} \in (0, 1]$ as the sparsity factors, we obtain a model, where the edges appear independently, and for each m , an edge on m nodes from the same group appears with probability $\alpha_{m,n}(p + q)$. For any set of m nodes from different groups, there is an edge among them with probability $\alpha_{m,n}q$.

Since, the non-uniform hypergraph is a superposition of the m -uniform hypergraphs for $m = 2, \dots, M$, one can easily derive a consistency result in the non-uniform case by applying Corollary 5.1 for each of the uniform components. However, observe that the number of edges of size m is $\Theta(\alpha_{m,n} \binom{n}{m})$, and hence, the requirement $\alpha_{m,n} \binom{n}{m} \geq C_m k^{2m-1} n (\ln n)^2$ for each m implies that the number of m -size edges should increase with m . This contradicts the natural intuition in existing random models (Darling and Norris, 2005), where the hypergraph contains less edges of higher cardinality. The same phenomenon is also observed in practice (see Section 6.1). The following consistency result takes this fact into account.

Corollary 5.2. *The partitions in the above model is identifiable for all $p > 0$. In addition, let $(\theta_m)_{m=2}^\infty$ be a non-negative sequence independent of n , and assume that for any $n \in \mathbb{N}$ and $m = 2, \dots, M$, the sparsity factor*

$$\alpha_{m,n} = \frac{\theta_m n^a (\ln n)^b}{\binom{n}{m}}$$

for some $a \geq 1$ and $b \geq 2$. There exists an absolute constant C , such that, if

$$\sum_{m=r}^M m \theta_m \leq C \left(\frac{n^{a-1} (\ln n)^{b-2}}{k^{2r-1}} \right) \quad (28)$$

for $r = \min\{m : \theta_m > 0\}$, then $\text{Err}(\psi, \psi') = o(n)$ with probability $(1 - o(1))$.

In the above result, r denotes the smallest size of an edge in the hypergraph. In practice, $(\theta_m)_{m=2}^\infty$ is a decreasing sequence, and hence, the number of m -size edges also decreases with m . In particular, if $\theta_2 > 0$, $\sum_{m=2}^\infty m \theta_m < \infty$, and $k = O(n^{(a-1)/3} (\ln n)^{(b-2)/3})$, then Algorithm 1 is weakly consistent. Thus, if the hypergraph is sparse, *i.e.*, $a = 1$, consistency is guaranteed only for logarithmic growth in k , whereas larger number of partitions can be consistently detected only in dense hypergraphs. Observe that the problem gets harder if $r > 2$.

5.3. Identifiability of the partitions. In the previous two sections, we considered problems where the partitions are identifiable from \mathcal{L} . This need not hold for arbitrary model parameters. We now briefly discuss few cases, which show that the partitions are typically identifiable under reasonable choice of model parameters.

Example 1. Consider a 3-uniform hypergraph on n nodes. For simplicity assume there are $k \geq 3$ partitions of equal size. We define $B^{(3)}$ as follows

$$B_{j_1 j_2 j_3}^{(3)} = \begin{cases} p_1 & \text{if } j_1 = j_2 = j_3, \\ p_2 & \text{if exactly two of them are identical,} \\ p_3 & \text{if } j_1, j_2, j_3 \text{ are all different.} \end{cases}$$

for some constants $p_1, p_2, p_3 \in [0, 1]$. Observe that the above situation is the most general case provided that the partitions are statistically identical. In this setting, it is easy to see that the following statement holds.

Lemma 5.3. *Assume that n is a multiple of k . Then $\delta > 0$ if and only if*

$$(p_2 - p_3) + \frac{1}{k}(p_1 - 3p_2 + 2p_3) - \frac{2}{n}(p_1 - p_2) > 0. \quad (29)$$

In particular, $\delta > 0$ when $p_1 > p_2 > p_3$, or at most one inequality is replaced by equality.

Note that the setting of Section 5.1 follows when $p_1 > p_2 = p_3$, while the case $p_1 = p_2 = p_3$ corresponds to a random hypergraph with all edges following the same law. Obviously, the partitions are not identifiable in the latter case. More generally, the order of probabilities $p_1 > p_2 > p_3$ is intuitive as it implies that an edge has a larger probability of occurrence if it has more nodes from the same community. One may compare this observation with the case of graphs, where partitioning based on the leading eigenvectors of Laplacian works only when edges within each community occur more frequently than edges across communities. The opposite scenario, found in colorable graphs, requires one to consider eigenvectors corresponding to the other end of the spectrum (Alon and Kahale, 1997). Moreover, if $p_2 > p_3$ and k grows with n , one can observe that δ mostly depends on the gap $(p_2 - p_3)$, and hence, the error $\text{Err}(\psi, \psi')$ is proportional to $\frac{1}{(p_2 - p_3)^2}$.

Example 2. We now modify the above model by allowing edges of size 2 to be present. In particular, assume $\alpha_{2,n} = 1$ and $B^{(2)} = I$, which means all pairwise edges within each community are present, and no two nodes from different communities form a pairwise edge. In addition, let $\alpha_{3,n} \in [0, 1]$ be arbitrary. Then, one can observe the following.

Lemma 5.4. *Assume that n is a multiple of k . Then $\delta > 0$ if and only if*

$$\frac{1}{2} + \frac{n\alpha_{3,n}}{3} \left((p_2 - p_3) + \frac{(p_1 - 3p_2 + 2p_3)}{k} - \frac{2(p_1 - p_2)}{n} \right) > 0. \quad (30)$$

It is easy to see that if $\alpha_{3,n} = 0$, then the hypergraph is a graph with k disconnected components, and hence, the partitions are identifiable. However, even when $\alpha_{3,n} = o(\frac{1}{n})$, the pairwise edges eventually dominate and the partitions can be identified for arbitrary values of p_1, p_2, p_3 . On the other hand, if $\alpha_{3,n}$ grows faster than $\frac{1}{n}$ (for instance $\alpha_{3,n} = 1$), then the situation is eventually similar to that of Lemma 5.3. The critical case is $\alpha_{3,n} = \Theta(\frac{1}{n})$, where the expected number of 2-way and 3-way edges are of similar order. In this case, (30) suggests that the partitions can be identified ($\delta > 0$) even when $p_2 < p_3$ provided p_3 is sufficiently small.

Example 3. In the above cases, we restricted ourselves to communities of equal size. The arguments also hold for $\frac{n_1}{n_k} = O(1)$. However, if $n_k \ll n_1$ or the probability of edges vary across different communities, then the second term in (13) can lead to $\delta \leq 0$, or equivalently, may affect the identifiability of the partitions. To study this effect, we consider the following model for r -uniform hypergraphs.

Let $\alpha_{r,n} = 1$, and there are $k = 2$ partitions of size s and $(n - s)$. We assume $s = o(n)$, and define $B^{(r)} \in \mathbb{R}^{2 \times 2 \times \dots \times 2}$ as

$$B_{j_1 j_2 \dots j_r}^{(r)} = \begin{cases} 1 & \text{if } j_1 = j_2 = \dots = j_r = 1, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

For $r = 2$, the model is same as that of a s clique planted in a Erdős-Rényi graph. This model presents a high disparity in both community sizes and degree distributions. We make the following comment on the identifiability of the partitions under this model.

Lemma 5.5. *For a given $r \geq 2$, there exists a finite constant s_r such that $\delta > 0$ for the above model for all $s \geq s_r$.*

Thus, when s grows with n , the partitions can be eventually identified from \mathcal{L} . The proof of the above result shows that both the terms in (13) decay with n , but the ratio of the first term to the second grows as $\Omega(s)$. We believe that a similar observation can be made in more general situations, where this growth rate depends on the size of the smallest community.

In view of the above lemma, it is interesting to know whether Algorithm 1 is able to detect small cliques in uniform hypergraphs. This is indeed true, but due to the generality of the approach, as listed in this paper, the minimal growth rate for s needed to accurately find the partitions from L is not optimal. More precisely, it is worse by a logarithmic factor in the case of graphs. However, Lemma 5.5 shows that one can use spectral techniques similar to (Alon et al., 1998) for finding planted cliques in hypergraphs.

6. EXPERIMENTAL RESULTS

In this section, we empirically demonstrate that the conditions in Corollaries 5.1 and 5.2 are reasonable. For this, we consider a number of hypergraphs that have been studied in practical problems (Ghoshal et al., 2009; Alpert, 1998). We also study the performance of Algorithm 1 in some benchmark clustering problems.

6.1. Sparsity of real-world hypergraphs. The consistency results in this paper are applicable only under certain restrictions on the hypergraph to be partitioned. To be precise, Corollaries 5.1 and 5.2 hold when the sparsity of the hypergraph is above a certain threshold. We study the practicability of the conditions in the case of real-world hypergraphs. We consider two types of applications – *folksonomy*, where the underlying model is a 3-uniform hypergraph, and *circuit design*, which involves non-uniform hypergraph partitioning.

To study the nature of hypergraphs in folksonomy, we consider 11 networks from KONECT, HetRec’2011 and MovieLens³. Each network is a tri-partite 3-uniform hypergraph containing three types of nodes – user, resource and annotation. Each

³The HetRec’2011 and MovieLens datasets are maintained by the GroupLens research group, and are available at: <http://grouplens.org/>

KONECT refers to the Koblenz network collection: <http://konect.uni-koblenz.de/>

edge is an entry in the database that occurs when an user describes a certain resource by a particular tag or rating. The number of nodes vary between 2630 to 9.8×10^5 . Assuming that $k = O(1)$, the sufficient condition in Corollary 5.1 requires that the number of edges in a 3-uniform hypergraph grows as $\Omega(n(\ln n)^2)$. In Figure 1, we compare the number of edges $|\mathcal{E}|$ with $n(\ln n)^2$ for above networks. We observe that in few cases (last four in Figure 1), these quantities are similar, whereas for the remaining networks, $|\mathcal{E}|$ is smaller by a nearly constant factor.

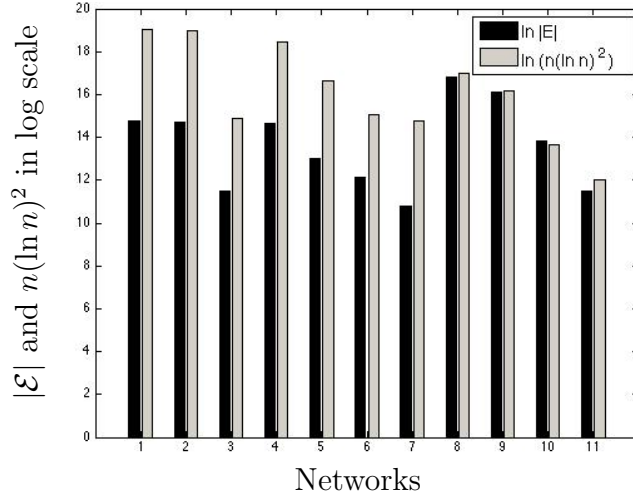


FIGURE 1. Bar plot for $|\mathcal{E}|$ and $n(\ln n)^2$ in logarithmic scale for 11 folksonomy networks.

The next study is related to non-uniform hypergraphs that are encountered in circuit partitioning. We consider 18 circuits from the ISPD98 circuit benchmark suite (Alpert, 1998). From a hypergraph view, the components of the circuit are the nodes of the hypergraph, while the multi-way connections among them are the edges. These networks are also sparse as the number of nodes vary from 1.27×10^4 to 2.1×10^5 , while the number of edges range between 1.4×10^4 to 2×10^5 . Moreover, these networks contain relatively large number of edges of sizes 2 or 3, and the number of edges of size m gradually decreases with m . We assume $a = 1$, $b = 2$, and ignoring constant factors, we estimate θ_m as $\theta_m = \frac{|\mathcal{E}_m|}{n(\ln n)^2}$, where \mathcal{E}_m is the set of edges of size m in the network. Figure 2 shows a plot of this quantity as a function of m for different networks. We find that the estimate of θ_m is bounded by exponentially decaying functions, and hence, one can argue that $\sum_m m\theta_m < \infty$.

6.2. Experiments on benchmark problems. Partitioning the networks discussed in the previous section is an interesting problem. However, for such networks, the underlying partitions are not known, and hence, for these networks, the performance of Algorithm 1 cannot be measured in terms of the number of incorrectly assigned nodes. So, we consider benchmark problems for categorical data clustering, where the true partitions are known. Here, one needs to group instances of a database, each described by a number of categorical attributes. Two such benchmark databases include the *1984 US Congressional Voting Records* and the *Mushroom Database* available at

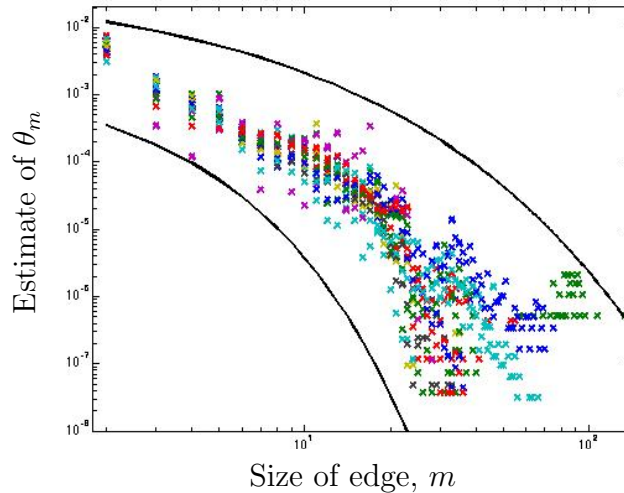


FIGURE 2. Scatter plot for estimated $\theta_m = \frac{|\mathcal{E}_m|}{n(\ln n)^2}$ versus m for the 18 circuits. Plot for each circuit is shown in a different color. The bounding curves correspond to the functions $0.05 \exp(-m^{0.5})$ from above and $0.002 \exp(-m^{0.8})$ from below.

the UCI repository (Lichman, 2013). The first set contains votes of 435 Congress men on 16 issues. The task is to group the Congress men into Democrats and Republicans based on whether they voted for or against the issue, or abstained their votes. The mushroom database contains information about 22 features of 8124 varieties of mushrooms. Based on the categorical features, one needs to separate the edible varieties from the poisonous ones. Thus, both databases have two well-defined partitions.

One may consider the instances of the database as the nodes of the hypergraph. For each possible value of each attribute, an edge is considered among all instances that take the particular value of the attribute. This generates a sparse non-uniform hypergraph that can be partitioned to obtain the clusters. Table 1 compares the performance of Algorithm 1 with some popular categorical clustering algorithms. We also study the performance when the hypergraph partitioning is done using a multi-level approach (hMETIS) (Karypis and Kumar, 2000) or eigen-decomposition of the normalized Laplacian obtained from clique expansion (Rodríguez, 2002). The error is measured as $\frac{1}{n} \text{Err}(\psi, \psi')$. Table 1 shows that Algorithm 1 performs quite well compared to other methods.

TABLE 1. Fraction of nodes incorrectly assigned by different clustering algorithms. The results for ROCK, COOLCAT and LIMBO are taken from (Andritsos et al., 2004).

Database	ROCK	COOLCAT	LIMBO	hMETIS	Clique	Algorithm 1
Voting	0.16	0.15	0.13	0.24	0.12	0.12
Mushroom	0.43	0.27	0.11	0.48	0.11	0.11

7. CONCLUSION

The primary focus of this work was to study the consistency of hypergraph partitioning in the presence of a planted structure in the hypergraph. This is achieved by considering a model for random hypergraphs that extends the stochastic block model in a natural way. The algorithm studied in this work is quite simple, where one essentially reduces a given hypergraph to a graph with weighted adjacency matrix given by (4), and then performs spectral clustering on this graph. Our analysis mainly relies on a matrix concentration inequality that was previously used to derive concentration bounds for the Laplacian matrix of sparse random graphs (Chung and Radcliffe, 2011). We also establish that the k -means step indeed achieves a constant factor approximation with probability $(1 - o(1))$. This question had remained unanswered for a long time in the spectral clustering literature.

Note on the optimality of our result. Theorem 4.2 is quite similar, in spirit, to the existing results in the block model literature, for instance (Lei and Rinaldo, 2015), where it is shown that spectral clustering is weakly consistent when the expected degree of any node is $\Omega(\ln n)$, or equivalently the edge density $\alpha_{2,n} = \Omega(\frac{\ln n}{n})$. One can easily see from Corollary 5.1 that our result is not optimal, at least in the case of graphs. The primary factor contributing to this difference is a sharp concentration result (Friedman et al., 1989) that holds for the sparse binary adjacency matrix. While such a sharp bound may not hold for a weighted adjacency matrix, we do wonder whether one can consider the following approach.

Question: Viewing a hypergraph as a collection of m -uniform hypergraphs, one can represent the adjacencies as a collection of m -way binary tensors for varying m . Does a generalization of (Friedman et al., 1989) hold for sparse binary tensors? If so, then what is its implication on the allowable sparsity for community detection in hypergraphs?

One can show that for dense tensors, the operator norm (equivalently, largest eigenvalue for matrices) does concentrate similar to the matrix case (Ghoshdastidar and Dukkipati, 2015), but the sparse case has not been studied yet.

Considering the most sparse regime for community detection (Decelle et al., 2011), it is now known that spectral techniques based on eigenvectors of suitably defined matrices work even for graphs with density $\alpha_{2,n} = \Theta(\frac{1}{n})$ (Le et al., 2015; Krzakala et al., 2013). To this end, the following problem is quite interesting.

Question: What is an appropriate extension of the regularized adjacency matrix (Le et al., 2015) and the non-backtracking matrix (Krzakala et al., 2013) in the case of hypergraphs? More generally, what is the algorithmic barrier for community detection in hypergraphs?

Phase transitions in uniform hypergraphs have been studied in the literature (Achlioptas and Coja-Oghlan, 2008; Panagiotou and Coja-Oghlan, 2012), and thresholds for 2-colorability and boolean satisfiability are known up to constant factors. However, the case of non-uniform hypergraphs still remains unexplored.

Extensions of our results. One can observe that both the model and the analysis can be further extended to more general situations. For instance, one often encounters weighted hypergraphs in practical applications (Ghoshdastidar and Dukkipati, 2014), where every edge e has a weight $w(e)$ associated with it. In our random model, we assumed $w(e)$ to be a Bernoulli random variable. A direct extension to the weighted hypergraphs is obtained by allowing $w(e)$ to take real values. To this end, we note that our results are only based on the first two moments of $w(e)$. Hence, if we restrict $w(e) \in [0, 1]$ and assume that first moment is same as that of the Bernoulli variables in our model, then Theorem 4.2 holds even in this setting.

In the case of planted graphs, the stochastic block model has been extended to account for factors such as degree heterogeneity or overlapping communities (Lei and Rinaldo, 2015; Zhang et al., 2014). Similar modifications of the hypergraph model is an interesting extension. However, it also seems possible that some information, such as community overlap, may be lost when the edge information is ‘compressed’ into the hypergraph Laplacian. Hence, one may have to consider spectral properties of the incidence matrix H or other alternatives for the Laplacian in (2).

While we restricted our discussions to a popular hypergraph partitioning approach, the results can be extended to variants of Algorithm 1. For instance, one may use the eigenvectors of the weighted adjacency matrix A instead of the Laplacian L . Minor modifications to Theorem 4.2 can guarantee the consistency of such an approach. Moreover, Theorem 4.2 is based on a theoretical result of approximate k -means (Ostrovsky et al., 2012). As mentioned in Section 4, we need to assume n to be sufficiently large in order to ensure that the k -means step provides a near optimal solution. Alternatively, one could also use the greedy clustering algorithm of Gao et al. (2015) that may alleviate the condition on n .

It also is known that one can iteratively refine the solution of a spectral algorithm to exactly recover the partitions (Vu, 2014; Lei and Zhu, 2014). Such an approach usually constructs an embedding of the nodes based on the adjacency matrix. We believe that similar results will hold for hypergraphs if one constructs the embedding using the weighted adjacency matrix A defined in (4).

We noted in Lemma 5.5 that Algorithm 1 can be used to find a planted clique in a random hypergraph. For a more optimal result, one could possibly extend the approach of (Alon et al., 1998) to the case of hypergraphs. To this end, it is interesting to note that the hypergraph clique problem is often encountered in computer vision applications (Ghoshdastidar and Dukkipati, 2014). Thus, several variations of the hypergraph partitioning problem surface in engineering applications.

This work explores into the theoretical analysis of hypergraph partitioning, and provides the first step for expanding the extensive studies on planted graphs to the case of hypergraphs.

ACKNOWLEDGMENT

The authors thank the reviewers for pointing out key references and for helpful suggestions that led to inclusion of the result for k -means algorithm, and discussions

on identifiability of the partitions. This work is supported by Department of Science and Technology (DST:SB/S3/EECE/093/2014)

APPENDIX A. PROOFS FOR COROLLARIES IN SECTION 4

The proofs are similar to that of Theorem 4.2. To prove Corollary 4.3, one needs to use the lemmas in Section 4, except Lemma 4.8. On the other hand, Corollary 4.4 follows when we use Lemma 4.8 with constant $\epsilon > 0$. The condition $\epsilon < 0.015$ follows immediately from the requirement of Lemma 4.11 in (Ostrovsky et al., 2012).

APPENDIX B. PROOFS FOR LEMMAS IN SECTION 4

B.1. Proof of Lemma 4.1. We observe that for $i \neq j$,

$$\begin{aligned} \mathcal{A}_{ij} &= \sum_{\ell=1}^{\beta_M} \frac{\mathbb{E}[h_\ell]}{|\xi_\ell|} (a_{\xi_\ell})_i (a_{\xi_\ell})_j = \sum_{m=2}^M \sum_{\substack{\ell: |\xi_\ell|=m, \\ i, j \in \xi_\ell}} \frac{\mathbb{E}[h_\ell]}{m} \\ &= \sum_{m=2}^M \sum_{\substack{i_3 < i_4 < \dots < i_m, \\ i, j \notin \{i_3, \dots, i_m\}}} \frac{1}{m} \alpha_{m,n} B_{\psi_i \psi_j \psi_{i_3} \dots \psi_{i_m}}^{(m)}. \end{aligned} \quad (31)$$

The last equality follows by noting that for every ξ_ℓ such that $|\xi_\ell| = m$ and $\xi_\ell \ni i, j$, we can write ξ_ℓ as $\xi_\ell = \{i, j, i_3, \dots, i_m\}$, where the nodes i, j, i_3, \dots, i_m are distinct. It is interesting to note that the above sum remains same if i, j are replaced by some i', j' such that $\psi_i = \psi_{i'}$ and $\psi_j = \psi_{j'}$. This is true since the terms in (31) depend on ψ_i, ψ_j instead of i, j . This observation motivates us to define the matrix $G \in \mathbb{R}^{k \times k}$ such that for any $i, j \in \mathcal{V}$, $i \neq j$,

$$G_{\psi_i \psi_j} = \sum_{m=2}^M \sum_{\substack{i_3 < i_4 < \dots < i_m, \\ i', j' \notin \{i_3, \dots, i_m\}}} \frac{1}{m} \alpha_{m,n} B_{\psi_{i'} \psi_{j'} \psi_{i_3} \dots \psi_{i_m}}^{(m)}, \quad (32)$$

where i', j' are arbitrary nodes satisfying $\psi_i = \psi_{i'}$ and $\psi_j = \psi_{j'}$. Hence, one can write $\mathcal{A}_{ij} = (ZGZ^T)_{ij}$ for all $i \neq j$, where Z is the assignment matrix. However,

$$\mathcal{A}_{ii} = \sum_{\ell: i \in \xi_\ell} \frac{\mathbb{E}[h_\ell]}{|\xi_\ell|} \neq G_{\psi_i \psi_i}.$$

So, one can write the matrix \mathcal{A} as $\mathcal{A} = ZGZ^T - J$, where $J \in \mathbb{R}^{n \times n}$ is a diagonal matrix defined as $J_{ii} = G_{\psi_i \psi_i} - \mathcal{A}_{ii}$. We also note that for i, i' in the same group, *i.e.*, $\psi_i = \psi_{i'}$, we have $\mathcal{D}_{ii} = \mathcal{D}_{i'i'}$ and $J_{ii} = J_{i'i'}$. So we can define matrices $\tilde{\mathcal{D}}, \tilde{J} \in \mathbb{R}^{k \times k}$ diagonal such that $\mathcal{D}_{ii} = \tilde{\mathcal{D}}_{\psi_i \psi_i}$ and $J_{ii} = \tilde{J}_{\psi_i \psi_i}$ for all $i \in \mathcal{V}$. It is easy to see that $\mathcal{D}Z = Z\tilde{\mathcal{D}}$ and $JZ = Z\tilde{J}$.

Using above definitions, we now characterize the eigenpairs of the matrix $\mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}$. To this end, note that (λ, v) is an eigenpair of \mathcal{L} if and only if $((1 - \lambda), v)$ is an eigenpair of $\mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}$. Hence, it suffices to consider the eigenvalues of $\mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}$, and their corresponding eigen spaces.

First, observe that since $G \in \mathbb{R}^{k \times k}$, \mathcal{A} is composed of a matrix of rank at most k that is perturbed by the diagonal matrix J . We show that the orthonormal basis for the range space of ZGZ^T are the eigenvectors that are of interest to us. For this, consider $\mathcal{G} = (\tilde{\mathcal{D}}^{-1}Z^T Z)^{1/2}G(Z^T Z\tilde{\mathcal{D}}^{-1})^{1/2} - \tilde{J}\tilde{\mathcal{D}}^{-1} \in \mathbb{R}^{k \times k}$, and suppose its eigen-decomposition is given by $\mathcal{G} = U\Lambda_1 U^T$, where $U \in \mathbb{R}^{k \times k}$ contains the orthonormal eigenvectors and $\Lambda_1 \in \mathbb{R}^{k \times k}$ is a diagonal matrix of eigenvalues of \mathcal{G} . Defining $\mathcal{X} = Z(Z^T Z)^{-1/2}U \in \mathbb{R}^{n \times k}$, we can write that

$$\begin{aligned} \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}\mathcal{X} &= \mathcal{D}^{-1/2}(ZGZ^T - J)\mathcal{D}^{-1/2}Z(Z^T Z)^{-1/2}U \\ &= \mathcal{D}^{-1/2}(ZG(Z^T Z)^{1/2} - Z(Z^T Z)^{-1/2}\tilde{J})\tilde{\mathcal{D}}^{-1/2}U \\ &= Z(Z^T Z)^{-1/2}\mathcal{G}U \\ &= Z(Z^T Z)^{-1/2}U\Lambda_1 = \mathcal{X}\Lambda_1, \end{aligned}$$

which implies that the columns of \mathcal{X} are the eigenvectors of $\mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$ corresponding to the k eigenvalues in Λ_1 . Alternatively, the columns of \mathcal{X} are the eigenvectors of \mathcal{L} corresponding to the k eigenvalues in $(I - \Lambda_1)$. Note that the above equalities are derived by repeated use of the facts that diagonal matrices commute and $\mathcal{D}Z = Z\tilde{\mathcal{D}}$, $JZ = Z\tilde{J}$. Also, since U is orthonormal, it is easy to verify that the columns of \mathcal{X} are orthonormal.

B.2. Proof of Lemma 4.5. We continue from the proof of Lemma 4.1. Note that we need to derive conditions under which \mathcal{X} contain the leading eigenvectors of \mathcal{L} . Equivalently, we need to show that the eigenvalues in Λ_1 are strictly larger than other eigenvalues of $\mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$.

Since, $\mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$ is symmetric and hence, diagonalizable, we can conclude that remaining eigenvectors of the matrix are orthogonal to columns of \mathcal{X} . Let the columns of $Y \in \mathbb{R}^{n \times (n-k)}$ be the matrix of the remaining orthonormal eigenvectors of $\mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$, with corresponding eigenvalues given by the diagonal matrix $\Lambda_2 \in \mathbb{R}^{(n-k) \times (n-k)}$. So $Y^T Z(Z^T Z)^{-1/2}U = 0$. Due to the non-singularity of $Z^T Z$ or U , it follows that $Z^T Y = 0$, and

$$Y\Lambda_2 = \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}Y = -\mathcal{D}^{-1}JY,$$

that is, the columns of Y are eigenvectors of $(-\mathcal{D}^{-1}J)$. Further, since $\mathcal{D}^{-1}J$ is diagonal, the eigenvalues in Λ_2 are a subset of the entries of $(-\mathcal{D}^{-1}J)$. Thus, to ensure that \mathcal{X} are the leading eigenvectors, one needs to ensure $\min_i(\Lambda_1)_{ii} > \max_i(\Lambda_2)_{ii}$, and hence, one may define $\tilde{\delta}$ as the eigen-gap,

$$\tilde{\delta} = \min_{1 \leq i \leq k} (\Lambda_1)_{ii} - \max_{1 \leq i \leq (n-k)} (\Lambda_2)_{ii}. \quad (33)$$

Hence, the condition $\tilde{\delta} > 0$ ensures that columns of \mathcal{X} are leading eigenvectors of \mathcal{L} . Though the above definition of $\tilde{\delta}$ suffices, it cannot be easily verified for a given model. Below, we show that $\tilde{\delta} \geq \delta$, where the latter is as defined in (13). Note that

$$\max_{1 \leq i \leq (n-k)} (\Lambda_2)_{ii} \leq \max_{1 \leq i \leq n} \left(-\frac{J_{ii}}{\mathcal{D}_{ii}} \right) = \min_{1 \leq i \leq n} \frac{J_{ii}}{\mathcal{D}_{ii}}$$

On the other hand, using Weyl's inequality, we have

$$\min_{1 \leq i \leq k} (\Lambda_1)_{ii} = \lambda_{\min}(\mathcal{G}) \geq \lambda_{\min}((\tilde{\mathcal{D}}^{-1} Z^T Z)^{1/2} G (Z^T Z \tilde{\mathcal{D}}^{-1})^{1/2}) - \|\tilde{J} \tilde{\mathcal{D}}^{-1}\|_2,$$

where $\lambda_{\min}(\mathcal{G})$ denotes the minimum eigenvalue of \mathcal{G} . The inequality follows by viewing \mathcal{G} as the matrix $(\tilde{\mathcal{D}}^{-1} Z^T Z)^{1/2} G (Z^T Z \tilde{\mathcal{D}}^{-1})^{1/2}$ perturbed by $-\tilde{\mathcal{D}}^{-1} \tilde{J}$. To simplify further, we note

$$\|\tilde{J} \tilde{\mathcal{D}}^{-1}\|_2 = \max_{1 \leq i \leq k} \frac{\tilde{J}_{ii}}{\tilde{\mathcal{D}}_{ii}} = \max_{1 \leq i \leq n} \frac{J_{ii}}{\mathcal{D}_{ii}},$$

and using Rayleigh's principle, one can show that

$$\lambda_{\min}((\tilde{\mathcal{D}}^{-1} Z^T Z)^{1/2} G (Z^T Z \tilde{\mathcal{D}}^{-1})^{1/2}) \geq \lambda_{\min}(G) \min_{1 \leq i \leq k} \frac{(Z^T Z)_{ii}}{\tilde{\mathcal{D}}_{ii}}.$$

Combining the above bounds, we conclude that $\tilde{\delta} \geq \delta$. Here, we use the observation that $(Z^T Z)_{jj}$ equals the size of the j^{th} community. Thus, $\delta > 0$ is a sufficient condition for the claim of the lemma.

B.3. Proof of Lemma 4.6. Define $\hat{L} = I - \mathcal{D}^{-1/2} A \mathcal{D}^{-1/2}$. Note that

$$\|L - \mathcal{L}\|_2 \leq \|L - \hat{L}\|_2 + \|\mathcal{L} - \hat{L}\|_2. \quad (34)$$

We deal with the two terms separately. First, we show that if $d > 9 \ln n$, then

$$\mathbb{P} \left(\|\mathcal{L} - \hat{L}\|_2 \geq 3 \sqrt{\frac{\ln n}{d}} \right) \leq \frac{2}{n^2}. \quad (35)$$

To prove (35), we note that

$$\begin{aligned} \mathcal{L} - \hat{L} &= \mathcal{D}^{-1/2} (A - \mathcal{A}) \mathcal{D}^{-1/2} \\ &= \sum_{\ell} (h_{\ell} - \mathbb{E}[h_{\ell}]) \frac{1}{|\xi_{\ell}|} \mathcal{D}^{-1/2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \mathcal{D}^{-1/2}. \end{aligned}$$

Denoting, each matrix in the sum as Y_{ℓ} , it is easy to see that $\{Y_{\ell}\}_{\ell}$ are independent with $\mathbb{E}[Y_{\ell}] = 0$. Hence, we can apply matrix Bernstein inequality (Chung and Radcliffe, 2011) to obtain

$$\begin{aligned} \mathbb{P} \left(\|\mathcal{L} - \hat{L}\|_2 \geq 3 \sqrt{\frac{\ln n}{d}} \right) &= \mathbb{P} \left(\left\| \sum_{\ell} Y_{\ell} \right\|_2 \geq 3 \sqrt{\frac{\ln n}{d}} \right) \\ &\leq 2n \exp \left(\frac{-\frac{9 \ln n}{d}}{2 \left\| \sum_{\ell} \text{Var}(Y_{\ell}) \right\|_2 + \frac{2}{3} \sqrt{\frac{9 \ln n}{d}} \max_{\ell} \|Y_{\ell}\|_2} \right), \end{aligned} \quad (36)$$

where

$$\text{Var}(Y_{\ell}) = \mathbb{E}[Y_{\ell}^2] = \text{Var}(h_{\ell}) \frac{(a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}})}{|\xi_{\ell}|^2} \mathcal{D}^{-1/2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \mathcal{D}^{-1/2}.$$

Thus,

$$\sum_{\ell} \text{Var}(Y_{\ell}) = \mathcal{D}^{-1/2} \left(\sum_{\ell} \text{Var}(h_{\ell}) \frac{(a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}})}{|\xi_{\ell}|^2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \right) \mathcal{D}^{-1/2}.$$

Note that for any matrix B , $\mathcal{D}^{-1/2} B \mathcal{D}^{-1/2}$ and $\mathcal{D}^{-1} B$ have same eigenvalues, and hence, using Gerschgorin's theorem (Stewart and Sun, 1990), one has

$$\begin{aligned} \left\| \sum_{\ell} \text{Var}(Y_{\ell}) \right\|_2 &\leq \max_{1 \leq i \leq n} \frac{1}{\mathcal{D}_{ii}} \sum_{j=1}^n \left(\sum_{\ell} \text{Var}(h_{\ell}) \frac{(a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}})}{|\xi_{\ell}|^2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \right)_{ij} \\ &= \max_{1 \leq i \leq n} \frac{1}{\mathcal{D}_{ii}} \sum_{\ell} \text{Var}(h_{\ell}) \frac{(a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}})}{|\xi_{\ell}|^2} (a_{\xi_{\ell}})_i \sum_{j=1}^n (a_{\xi_{\ell}})_j. \end{aligned}$$

Observing that $a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}} \leq \frac{a_{\xi_{\ell}}^T a_{\xi_{\ell}}}{d}$ and $|\xi_{\ell}| = \sum_j (a_{\xi_{\ell}})_j = a_{\xi_{\ell}}^T a_{\xi_{\ell}}$, we have

$$\left\| \sum_{\ell} \text{Var}(Y_{\ell}) \right\|_2 \leq \frac{1}{d} \max_{1 \leq i \leq n} \frac{1}{\mathcal{D}_{ii}} \sum_{\ell} \text{Var}(h_{\ell}) (a_{\xi_{\ell}})_i \leq \frac{1}{d}$$

since $\text{Var}(h_{\ell}) = \mathbf{E}[h_{\ell}](1 - \mathbf{E}[h_{\ell}]) \leq \mathbf{E}[h_{\ell}]$. Similarly, one can also compute

$$\|Y_{\ell}\|_2 \leq |h_{\ell} - \mathbf{E}[h_{\ell}]| \frac{1}{|\xi_{\ell}|} \|\mathcal{D}^{-1/2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \mathcal{D}^{-1/2}\|_2 \leq \frac{|a_{\xi_{\ell}}^T \mathcal{D}^{-1} a_{\xi_{\ell}}|}{|\xi_{\ell}|} \leq \frac{1}{d},$$

where second inequality holds since $h_{\ell} \in \{0, 1\}$ and $\mathcal{D}^{-1/2} a_{\xi_{\ell}} a_{\xi_{\ell}}^T \mathcal{D}^{-1/2}$ is a rank-1 matrix. Substituting above bounds in (36) and noting that $\frac{9 \ln n}{d} < 1$, we have

$$\mathbf{P} \left(\|\mathcal{L} - \widehat{L}\|_2 \geq 3\sqrt{\frac{\ln n}{d}} \right) \leq 2n \exp \left(-\frac{9 \ln n}{\frac{d}{2} + \frac{1}{d}} \right) = \frac{2}{n^2},$$

which proves (35). To bound the other term in (34), we note that

$$\begin{aligned} \|L - \widehat{L}\|_2 &\leq \|\mathcal{D}^{-1/2} A \mathcal{D}^{-1/2} - D^{-1/2} A D^{-1/2}\|_2 \\ &\leq \|(\mathcal{D}^{-1/2} - D^{-1/2}) A \mathcal{D}^{-1/2} + D^{-1/2} A (\mathcal{D}^{-1/2} - D^{-1/2})\|_2 \\ &\leq \|(\mathcal{D}^{-1} D)^{1/2} - I\|_2 \| (D \mathcal{D}^{-1})^{1/2} \|_2 + \|(\mathcal{D}^{-1} D)^{1/2} - I\|_2. \end{aligned}$$

In above, we use the fact that $D_{ii} = \sum_j A_{ij}$ to conclude that $\|D^{-1/2} A \mathcal{D}^{-1/2}\|_2 = 1$. Note that $\mathcal{D}^{-1} D$ is a diagonal matrix with non-negative diagonal entries, and hence,

$$\|(\mathcal{D}^{-1} D)^{1/2} - I\|_2 = \max_{1 \leq i \leq n} \left| \sqrt{\frac{D_{ii}}{\mathcal{D}_{ii}}} - 1 \right| \leq \max_{1 \leq i \leq n} \left| \frac{D_{ii}}{\mathcal{D}_{ii}} - 1 \right|,$$

where the inequality follows from the fact that $|\sqrt{x} - 1| \leq |x - 1|$ for all $x \geq 0$. We now claim that for all $i = 1, \dots, n$,

$$\mathbf{P} \left(|D_{ii} - \mathcal{D}_{ii}| > 3\mathcal{D}_{ii} \sqrt{\frac{\ln n}{d}} \right) \leq \frac{2}{n^3}. \quad (37)$$

Hence, with probability at least $(1 - \frac{2}{n^2})$,

$$\max_{1 \leq i \leq n} \left| \frac{D_{ii}}{\mathcal{D}_{ii}} - 1 \right| \leq 3\sqrt{\frac{\ln n}{d}}.$$

From above and the relation $\|(DD^{-1})^{1/2}\|_2 \leq 1 + \|(DD^{-1})^{1/2} - I\|_2$, we have

$$\|L - \widehat{L}\|_2 \leq \frac{9 \ln n}{d} + 6\sqrt{\frac{\ln n}{d}} \leq 9\sqrt{\frac{\ln n}{d}}$$

where the last inequality holds since $3\sqrt{\frac{\ln n}{d}} < 1$. The lemma follows by combining above bound with (35).

Finally, we prove (37). Since, $D_{ii} = \sum_{\ell} h_{\ell}(a_{\xi_{\ell}})_i = \sum_{\ell: i \in \xi_{\ell}} h_{\ell}$, we use Bernstein inequality to write

$$\begin{aligned} \mathbb{P} \left(|D_{ii} - \mathcal{D}_{ii}| > 3\mathcal{D}_{ii}\sqrt{\frac{\ln n}{d}} \right) &= \mathbb{P} \left(\left| \sum_{\ell: i \in \xi_{\ell}} (h_{\ell} - \mathbb{E}[h_{\ell}]) \right| > 3\mathcal{D}_{ii}\sqrt{\frac{\ln n}{d}} \right) \\ &\leq 2 \exp \left(\frac{-\frac{9\mathcal{D}_{ii}^2 \ln n}{d}}{2 \sum_{\ell: i \in \xi_{\ell}} \text{Var}(h_{\ell}) + 2\mathcal{D}_{ii}\sqrt{\frac{\ln n}{d}}} \right) \\ &\leq 2 \exp \left(-\frac{3\mathcal{D}_{ii} \ln n}{d} \right) \end{aligned}$$

for $d > 9 \ln n$. Since, $\mathcal{D}_{ii} \geq d$, we obtain (37).

B.4. Proof of Lemma 4.7. To perform a valid row normalization of X , first we need to ensure that the rows of X are non-zero with high probability. From (37), we see that for all $i \in \mathcal{V}$,

$$\mathbb{P} \left(D_{ii} \geq \mathcal{D}_{ii} \left(1 - 3\sqrt{\frac{\ln n}{d}} \right) \right) \geq 1 - \frac{2}{n^3},$$

where the second term in lower bound of D_{ii} is smaller than 1 for $d > 9 \ln n$. In addition, we know $\mathcal{D}_{ii} \geq d > 9 \ln n$. Combining above with union bound, we can say that $\mathbb{P}(\min_i D_{ii} > 0) \geq 1 - \frac{2}{n^2}$, *i.e.*, the hypergraph contains no disconnected node with probability at least $(1 - \frac{2}{n^2})$.

One can verify the following properties for L in (2), which can be derived using arguments similar to those for graph Laplacian (von Luxburg, 2007).

- L is positive semi-definite with eigenvalues in $[0, 2]$.
- The multiplicity of 0 eigenvalue of L is equal to the number of connected components of the hypergraph.
- Provided that $D_{ii} > 0$ for all i , for each zero eigenvalue of L , there is an eigenvector with non-zero coordinates for one connected component.

From the condition $\delta > 0$ in Lemma 4.5, there is a strictly positive eigen-gap between the k^{th} and $(k+1)^{\text{th}}$ eigenvalues of \mathcal{L} , and hence, \mathcal{L} has at most k zero eigenvalues. Denoting $\lambda_i(L), \lambda_i(\mathcal{L})$ as the i^{th} smallest eigenvalues of L, \mathcal{L} respectively, we have $\delta \leq \tilde{\delta} = \lambda_{k+1}(\mathcal{L}) - \lambda_k(\mathcal{L})$, where $\tilde{\delta}$ is defined in (33). Also, we can use Weyl's inequality to claim that for all $i = 1, \dots, n$,

$$|\lambda_i(L) - \lambda_i(\mathcal{L})| \leq \|L - \mathcal{L}\|_2 \leq 12\sqrt{\frac{\ln n}{d}} < \frac{\delta}{2}$$

if δ and d satisfy the prescribed condition. Thus

$$\lambda_{k+1}(L) \geq \lambda_{k+1}(\mathcal{L}) - \frac{\delta}{2} = \lambda_k(\mathcal{L}) + \frac{\delta}{2} > 0,$$

which means L has at most k zero eigenvalues, *i.e.*, at most k connected components. Since, all nodes have positive degrees almost surely, hence, every node corresponds to a connected component. Due to third property of L , for every node, at least one of the k leading eigenvectors has a non-zero component, and hence, every row of X is non-zero. Thus, \bar{X} is well-defined.

We now derive a perturbation bound involving X and \mathcal{X} . By Davis-Kahan $\sin \Theta$ theorem (Stewart and Sun, 1990), we have if $2\|L - \mathcal{L}\|_2 < \delta$, then

$$\|\sin \Theta(X, \mathcal{X})\|_2 \leq \frac{\|L - \mathcal{L}\|_2}{\delta}, \quad (38)$$

where $\sin \Theta(X, \mathcal{X}) \in \mathbb{R}^{k \times k}$ is diagonal with entries same as the sine of the canonical angles between the subspaces X and \mathcal{X} . Let these angles be denoted as $\theta_1, \dots, \theta_k \in [0, \frac{\pi}{2}]$ such that $\theta_1 \geq \dots \geq \theta_k$. Then $\|\sin \Theta(X, \mathcal{X})\|_2 = \sin \theta_1$. On the other hand, one can see that the singular values for the matrix $X^T \mathcal{X}$ are given by $\cos \theta_1, \dots, \cos \theta_k$. Thus, if $X^T \mathcal{X} = U_1 \Sigma U_2^T$ is the singular value decomposition of $X^T \mathcal{X}$, then

$$\begin{aligned} \|X - \mathcal{X}U_2U_1^T\|_F^2 &= \text{Trace} \left((X - \mathcal{X}U_2U_1^T)^T (X - \mathcal{X}U_2U_1^T) \right) \\ &= 2\text{Trace}(I - U_1 \Sigma U_1^T) \\ &= 2 \sum_{i=1}^k (1 - \cos \theta_i) \leq 2 \sum_{i=1}^k (1 - \cos^2 \theta_i) \leq 2k \sin^2 \theta_1. \end{aligned} \quad (39)$$

From (38) and (39), we can conclude that for $\delta > 24\sqrt{\frac{\ln n}{d}} \geq 2\|L - \mathcal{L}\|_2$,

$$\|X - Z(Z^T Z)^{-1/2} Q\|_F = \|X - \mathcal{X}U_2U_1^T\|_F \leq \sqrt{2k} \frac{\|L - \mathcal{L}\|_2}{\delta}, \quad (40)$$

where $Q = UU_2U_1^T \in \mathbb{R}^{k \times k}$ is orthonormal. One can see that i^{th} row of $Z(Z^T Z)^{-1/2} Q$ is $Z_{i \cdot} (Z^T Z)^{-1/2} Q$, and the norm of i^{th} row is $(Z^T Z)_{\psi_i \psi_i}^{-1/2}$. Thus, on row normalization of this matrix, one obtains ZQ . Hence,

$$\begin{aligned} \|\bar{X} - ZQ\|_F^2 &= \sum_{i=1}^n \left\| \frac{1}{\|X_{i \cdot}\|_2} X_{i \cdot} - Z_{i \cdot} Q \right\|_2^2 \\ &= \sum_{i=1}^n \left\| \left(\frac{1}{\|X_{i \cdot}\|_2} - (Z^T Z)_{\psi_i \psi_i}^{1/2} \right) X_{i \cdot} + (Z^T Z)_{\psi_i \psi_i}^{1/2} (X_{i \cdot} - Z_{i \cdot} (Z^T Z)^{-1/2} Q) \right\|_2^2 \end{aligned}$$

Now,

$$\begin{aligned}
 & \left\| \left(\frac{1}{\|X_i\|_2} - (Z^T Z)_{\psi_i \psi_i}^{1/2} \right) X_i + (Z^T Z)_{\psi_i \psi_i}^{1/2} (X_i - Z_i (Z^T Z)^{-1/2} Q) \right\|_2^2 \\
 & \leq \sqrt{(Z^T Z)_{\psi_i \psi_i}} \left(\|Z_i (Z^T Z)^{-1/2} Q\|_2 - \|X_i\|_2 \right) + \|X_i - Z_i (Z^T Z)^{-1/2} Q\|_2 \\
 & \leq 2 \sqrt{(Z^T Z)_{\psi_i \psi_i}} \|X_i - Z_i (Z^T Z)^{-1/2} Q\|_2 .
 \end{aligned}$$

Substituting this bound above, we get

$$\begin{aligned}
 \|\bar{X} - ZQ\|_F^2 & \leq 4 \sum_{i=1}^n (Z^T Z)_{\psi_i \psi_i} \|X_i - Z_i (Z^T Z)^{-1/2} Q\|_2^2 \\
 & \leq 4n_1 \|X - Z(Z^T Z)^{-1/2} Q\|_F^2 ,
 \end{aligned}$$

where $n_1 = \max_j (Z^T Z)_{jj}$ is the size of the largest partition since we assumed that $n_1 \geq \dots \geq n_k$. The bound in (21) follows by combining above bound with (40) and Lemma 4.6.

B.5. Proof of Lemma 4.8. Let $\epsilon = (\ln n)^{-1/2}$. From (21), we have an upper bound on $\|\bar{X} - ZQ\|_F$ with probability $(1 - O(n^{-2}))$. For convenience, let us denote this upper bound by β . The condition in (14) implies that

$$\beta \leq 24 \sqrt{\frac{2n_k}{C \ln n}} \leq \frac{\epsilon \sqrt{n_k}}{2}$$

if C is chosen sufficiently large. For large enough n , above inequality implies

$$\beta \leq \epsilon (\sqrt{n_k} - \beta) , \tag{41}$$

which will be used to prove ϵ -separability, *i.e.*, $\eta_k(\bar{X}) \leq \epsilon \eta_{k-1}(\bar{X})$. Since ZQ has exactly k distinct rows, we have $\eta_k(\bar{X}) \leq \|X - ZQ\|_F \leq \beta$. On the other hand, observe that all matrices in $\mathcal{M}_{n \times k}(r)$ have rank at most r . Hence,

$$\eta_{k-1}(\bar{X}) = \min_{S \in \mathcal{M}_{n \times k}(k-1)} \|\bar{X} - S\|_F \geq \min_{\text{rank}(S) \leq (k-1)} \|\bar{X} - S\|_F .$$

It is well known that the minimum of the last quantity is $\lambda_k(\bar{X})$, which is the smallest singular value of \bar{X} . Also Mirsky's theorem (Stewart and Sun, 1990) gives a bound on the perturbation of singular values, and hence, we have

$$|\lambda_i(\bar{X}) - \lambda_i(ZQ)| \leq \|\bar{X} - ZQ\|_2 \leq \beta$$

for $i = 1, \dots, k$. Note here that the singular values of ZQ are $\lambda_i(ZQ) = \sqrt{n_i}$, where the ordering of the values is due to our assumption $n_1 \geq \dots \geq n_k$. From above arguments

$$\eta_{k-1}(\bar{X}) \geq \lambda_k(\bar{X}) \geq (\lambda_k(ZQ) - \beta) = (\sqrt{n_k} - \beta) .$$

Hence, it follows that ϵ -separability holds when (41) is satisfied, which is true under the assumption in (14).

B.6. Proof of Lemma 4.9. We observe from the proof of Lemma 4.7 that the k distinct rows of ZQ , form an orthonormal set of vectors in \mathbb{R}^k . Hence, for any $i, j \in \mathcal{V}$, $\|Z_i.Q - Z_j.Q\|_2 = 0$ or $\sqrt{2}$, where the former occurs if $Z_i = Z_j$, *i.e.*, $\psi_i = \psi_j$, and the latter occurs if $\psi_i \neq \psi_j$. Now, consider $i, j \notin \mathcal{V}_{err}$. We have $\|S_i^* - Z_i.Q\|_2 < \frac{1}{\sqrt{2}}$, $\|S_j^* - Z_j.Q\|_2 < \frac{1}{\sqrt{2}}$, and hence,

$$\|Z_i.Q - Z_j.Q\|_2 \leq \|S_i^* - Z_i.Q\|_2 + \|S_j^* - Z_j.Q\|_2 + \|S_i^* - S_j^*\|_2 < \sqrt{2}$$

whenever $S_i^* = S_j^*$. So, from the previous observation, $\|Z_i.Q - Z_j.Q\|_2 = 0$, *i.e.*, $\psi_i = \psi_j$, which proves the first claim.

The second claim follows by using arguments similar to (Rohe et al., 2011). Note that for all $i \in \mathcal{V}_{err}$, we have $2\|S_i^* - Z_i.Q\|_2^2 \geq 1$. Therefore,

$$|\mathcal{V}_{err}| = \sum_{i \in \mathcal{V}_{err}} 1 \leq 2 \sum_{i \in \mathcal{V}_{err}} \|S_i^* - Z_i.Q\|_2^2 \leq 2\|S^* - ZQ\|_F^2. \quad (42)$$

Since, S^* is a sub-optimal solution satisfying (22), we can write $\|\bar{X} - S^*\|_F \leq \gamma\|\bar{X} - S\|_F$, for all $S \in \mathcal{M}_{n \times k}(k)$. In particular, $ZQ \in \mathcal{M}_{n \times k}(k)$ and so $\|\bar{X} - S^*\|_F \leq \gamma\|\bar{X} - ZQ\|_F$. Hence,

$$\|S^* - ZQ\|_F \leq \|\bar{X} - ZQ\|_F + \|\bar{X} - S^*\|_F \leq (1 + \gamma)\|\bar{X} - ZQ\|_F.$$

The lemma follows by combining above inequality with (42), and using the relation $(1 + \gamma)^2 \leq 2(1 + \gamma^2)$.

APPENDIX C. PROOFS OF COROLLARIES IN SECTION 5

C.1. Proof of Corollary 5.1. We observe that for the specified model, the matrix $G \in \mathbb{R}^{k \times k}$, as defined in Lemma 4.1, is given by

$$G_{ij} = \begin{cases} \frac{p\alpha_{r,n}}{r} \binom{\frac{n}{k} - 2}{r-2} + \frac{q\alpha_{r,n}}{r} \binom{n-2}{r-2} & \text{if } i = j, \\ \frac{q\alpha_{r,n}}{r} \binom{n-2}{r-2} & \text{if } i \neq j. \end{cases}$$

Thus, G is of the form $G = aI + b\mathbf{1}$, where $\mathbf{1}$ is constant matrix of ones. It is easy to verify that for such a matrix, the minimum eigenvalue is a . Hence, we have $\lambda_{\min}(G) = \frac{p\alpha_{r,n}}{r} \binom{\frac{n}{k} - 2}{r-2}$. Also, we can compute

$$d = \left(p\alpha_{r,n} \binom{\frac{n}{k} - 1}{r-1} + q\alpha_{r,n} \binom{n-1}{r-1} \right) > q\alpha_{r,n} \binom{n-1}{r-1}. \quad (43)$$

Note that since the partitions are balanced and node degrees behave identically, the second term in (13) is zero, and we can compute δ as

$$\delta = \frac{np\alpha_{r,n}}{kdr} \binom{\frac{n}{k} - 2}{r-2} \geq \frac{n}{kr} \frac{p\alpha_{r,n} \binom{\frac{n}{k} - 2}{r-2}}{(p+q)\alpha_{r,n} \binom{n-1}{r-1}} \geq \frac{p(r-1)}{r(p+q)} \frac{k \binom{n/k}{r}}{\binom{n}{r}}, \quad (44)$$

where the first inequality follows by observing that $d \leq (p+q)\alpha_{r,n} \binom{n-1}{r-1}$. Now, observe that under the given condition on $\alpha_{r,n}$, we have

$$\delta^2 d \geq C' \frac{\alpha_{r,n}}{n} \binom{n}{r} \left(\frac{k \binom{n/k}{r}}{\binom{n}{r}} \right)^2 \geq C'' k^{2r-1} (\ln n)^2 \left(\frac{k \left(\frac{n}{k} \right)^r}{n^r} \right)^2 \geq C'' k (\ln n)^2,$$

where C'' is a constant depending only on C, p, q and r , that is obtained from (43) and (44). The last inequality uses the relation $\frac{a^b}{4(b!)} \leq \binom{a}{b} \leq \frac{a^b}{b!}$. Choosing C sufficiently large, the condition of Theorem 4.2 is satisfied, and hence, we can conclude from Theorem 4.2 that $\text{Err}(\psi, \psi') = O\left(\frac{n \ln n}{\delta^2 d}\right)$, which simplifies to the stated claim.

C.2. Proof of Corollary 5.2. The proof is quite similar to that of Corollary 5.1 since the matrices G and \mathcal{D} are linear combinations of the corresponding matrices for the m -uniform hypergraphs. We compute

$$\lambda_{\min}(G) = \sum_{m=2}^M \frac{p\alpha_{m,n}}{m} \binom{\frac{n}{k} - 2}{m-2} \geq \frac{p\alpha_{r,n}}{r} \binom{\frac{n}{k} - 2}{r-2},$$

where $\theta_r > 0$ and we ignore all terms for $m > r$. Substituting the relation for $\alpha_{r,n}$, we can write

$$\lambda_{\min}(G) \geq C_1 \frac{\theta_r n^{a-2} (\ln n)^b}{k^{r-2}}$$

for some constant C_1 depending on p, q and r . Also,

$$\begin{aligned} d &= \sum_{m=2}^M \left(p\alpha_{m,n} \binom{\frac{n}{k} - 1}{m-1} + q\alpha_{m,n} \binom{n-1}{m-1} \right) \\ &> q \sum_{m=2}^M \frac{m}{n} \alpha_{m,n} \binom{n}{m} \geq n^{a-1} (\ln n)^b q \sum_{m=2}^M m\theta_m. \end{aligned}$$

Similarly, one can verify that $d \leq (p+q)n^{a-1} (\ln n)^b \sum_{m=2}^M m\theta_m$. Thus,

$$\delta^2 d \geq \frac{n^2 (\lambda_{\min}(G))^2}{k^2 d} \geq \frac{C_1^2 \theta_r^2}{(p+q)} \frac{n^{a-1} (\ln n)^b}{k^{2r-2} \sum_{m=2}^M m\theta_m}, \quad (45)$$

where the inequality follows from above bounds on $\lambda_{\min}(G)$ and d . Under the condition (28) with large enough C , the condition of Theorem 4.2 holds and the claim follows.

C.3. Proofs for Lemmas 5.3 and 5.4. As in the previous corollaries, we can say that, for both cases, the second term in (13) is zero. Hence, $\delta > 0$ if and only if

$\lambda_{\min}(G) > 0$. For the setting of Lemma 5.3, one can compute

$$G_{ij} = \begin{cases} \frac{\alpha_{3,n}}{3} \left[p_1 \left(\frac{n}{k} - 2 \right) + p_2 \left(n - \frac{n}{k} \right) \right] & \text{if } i = j, \\ \frac{\alpha_{3,n}}{3} \left[p_2 \left(\frac{2n}{k} - 2 \right) + p_3 \left(n - \frac{2n}{k} \right) \right] & \text{if } i \neq j. \end{cases}$$

Using an observation made in the proof of Corollary 5.1, we have $\lambda_{\min}(G) = G_{11} - G_{12}$, and (29) is equivalent to stating $G_{11} > G_{12}$.

The same arguments are valid for Lemma 5.4, where G is of the form

$$G_{ij} = \begin{cases} \frac{1}{2} + \frac{\alpha_{3,n}}{3} \left[p_1 \left(\frac{n}{k} - 2 \right) + p_2 \left(n - \frac{n}{k} \right) \right] & \text{if } i = j, \\ \frac{\alpha_{3,n}}{3} \left[p_2 \left(\frac{2n}{k} - 2 \right) + p_3 \left(n - \frac{2n}{k} \right) \right] & \text{if } i \neq j. \end{cases}$$

C.4. Proof of Lemma 5.5. As mentioned in Lemma 4.1, one can write \mathcal{A} as $\mathcal{A} = ZGZ^T - J$. In the present case, $G \in \mathbb{R}^{2 \times 2}$ is given by

$$G_{ij} = \begin{cases} \frac{1}{2r} \binom{s-2}{r-2} + \frac{1}{2r} \binom{n-2}{r-2} & \text{if } i = j = 1, \\ \frac{1}{2r} \binom{n-2}{r-2} & \text{otherwise.} \end{cases}$$

One can also verify that the diagonal matrices \mathcal{D} and J are given by

$$\mathcal{D}_{ii} = \begin{cases} \frac{1}{2} \binom{s-1}{r-1} + \frac{1}{2} \binom{n-1}{r-1} & \text{if } i \in s\text{-clique,} \\ \frac{1}{2} \binom{n-1}{r-1} & \text{otherwise.} \end{cases}$$

and

$$J_{ii} = \begin{cases} -\frac{1}{2r} \binom{s-2}{r-1} - \frac{1}{2r} \binom{n-2}{r-1} & \text{if } i \in s\text{-clique,} \\ -\frac{1}{2r} \binom{n-2}{r-1} & \text{otherwise.} \end{cases}$$

Hence, the claim follows by substituting above relations in (13). It is more convenient to write (13) using the notations $\tilde{\mathcal{D}}$ and \tilde{J} defined in the proof of Lemma 4.5, and it can be written as

$$\delta = \frac{s\lambda_{\min}(G)}{\tilde{\mathcal{D}}_{11}} - \left| \frac{\tilde{J}_{11}}{\tilde{\mathcal{D}}_{11}} - \frac{\tilde{J}_{22}}{\tilde{\mathcal{D}}_{22}} \right|, \quad (46)$$

where we use the fact $s < (n - s)$. Note that G has non-negative eigenvalues, and we can use the following relation for 2×2 matrices

$$\lambda_{\min}(G) \geq \frac{\det(G)}{\text{Trace}(G)} = \frac{\frac{1}{2r} \binom{s-2}{r-2} \binom{n-2}{r-2}}{\binom{s-2}{r-2} + 2 \binom{n-2}{r-2}} \geq \frac{1}{6r} \binom{s-2}{r-2}.$$

Combining this with the observation $\tilde{\mathcal{D}}_{11} \leq \binom{n-1}{r-1}$, we can argue that the first term in (46) is at least

$$\frac{s \binom{s-2}{r-2}}{6r \binom{n-1}{r-1}} \geq C_1 \left(\frac{s}{n}\right)^{r-1}$$

for some $C_1 > 0$. On the hand, the second term in (46) can be computed as

$$\begin{aligned} \left| \frac{\tilde{J}_{11}}{\tilde{\mathcal{D}}_{11}} - \frac{\tilde{J}_{22}}{\tilde{\mathcal{D}}_{22}} \right| &= \left| \frac{\binom{s-1}{r-1} \binom{n-2}{r-1} - \binom{n-1}{r-1} \binom{s-2}{r-1}}{\binom{n-1}{r-1} \left[\binom{n-1}{r-1} + \binom{s-1}{r-1} \right]} \right| \\ &\leq \frac{\binom{s-1}{r-1}}{\binom{n-1}{r-1}} \left(\frac{n-r}{n-1} - \frac{s-r}{s-1} \right) \\ &\leq \frac{\binom{s-1}{r-1}}{\binom{n-1}{r-1}} \left(\frac{r-1}{s-1} \right) \leq \frac{C_2}{s} \left(\frac{s}{n}\right)^{r-1} \end{aligned}$$

for some $C_2 > 0$. From above, one can conclude that $\delta > 0$ when $C_1 > \frac{C_2}{s}$, where both C_1 and C_2 depend only on r . Hence, the claim.

REFERENCES

Achlioptas, D. and A. Coja-Oghlan (2008). Algorithmic barriers from phase transitions. In *Proceedings of 49th Annual Symposium on Foundations of Computer Science*.

Agarwal, S., K. Branson, and S. Belongie (2006). Higher order learning with graphs. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 17–24.

Alon, N. and N. Kahale (1997). A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing* 26(6), 1733–1748.

Alon, N., M. Krivlevich, and B. Sudakov (1998). Finding a large hidden clique in a random graph. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 594–598.

Alpert, C. J. (1998). The ISPD98 circuit benchmark suite. In *ISPD '98 Proceedings of the 1998 International Symposium on Physical Design*, pp. 80–85.

Amini, A. A. and E. Levina (2014). On semi-definite relaxations for the block model. *arXiv preprint arXiv:1406.5647*.

Andritsos, P., P. Tsaparas, R. J. Miller, and K. C. Sevcik (2004). LIMBO: scalable clustering of categorical data. In *International Conference on Extending Database Technology*, pp. 123–146.

Berge, C. (1984). *Hypergraphs: combinatorics of finite sets*, Volume 45. Elsevier.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences* 106, 21068–21073.

Bolla, M. (1993). Spectra, euclidean representations and clusterings of hypergraphs. *Discrete Mathematics* 117(1), 19–39.

Chen, H. and A. Frieze (1996). Coloring bipartite hypergraphs. *Integer Programming and Combinatorial Optimization*, 345–358.

Chen, Y., S. Sanghavi, and H. Xu (2014). Improved graph clustering. *IEEE Transactions on Information Theory* 60(10), 6440–6455.

- Choi, D. S., P. J. Wolfe, and E. M. Airoidi (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* 99(2), 273–284.
- Chung, F. (1992). The laplacian of a hypergraph. In *Expanding Graphs: Proceedings of a DIMACS Workshop*, pp. 21–36.
- Chung, F. and M. Radcliffe (2011). On the spectra of general random graphs. *Electronic Journal of Combinatorics* 18(1), 215–229.
- Cooper, J. and A. Dutle (2012). Spectra of uniform hypergraphs. *Linear Algebra and its Applications* 436(9), 3268–3292.
- Darling, R. W. R. and J. R. Norris (2005). Structure of large random hypergraphs. *Annals of Applied Probability* 15(1A), 125–152.
- Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E* 84(066106).
- Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal* 23(2), 298–305.
- Friedman, J., J. Kahn, and E. Szemerédi (1989). On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM Symposium on Theory of Computing*.
- Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2015). Achieving optimal misclassification proportion in stochastic block model. *arXiv preprint arXiv:1505.03772*.
- Ghoshal, G., V. Zlatić, G. Caldarelli, and M. E. J. Newman (2009). Random hypergraphs and their applications. *Physical Review E* 79(066118).
- Ghoshdastidar, D. and A. Dukkipati (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 397–405.
- Ghoshdastidar, D. and A. Dukkipati (2015). A provable generalized tensor spectral method for uniform hypergraph partitioning. In *Proceedings of the International Conference on Machine Learning (ICML)*.
- Gibson, D., J. Kleinberg, and P. Raghavan (2000). Clustering categorical data: An approach based on dynamical systems. *VLDB Journal* 8(3-4), 222–236.
- Girvan, M. and M. E. J. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99(12), 7821–7826.
- Govindu, V. M. (2005). A tensor decomposition for geometric grouping and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1150–1157.
- Guimera, R. and L. A. N. Amaral (2005). Functional cartography of complex metabolic networks. *Nature* 433(7028), 895–900.
- Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks* 5(2), 109–137.
- Hu, S. and L. Qi (2012). Algebraic connectivity of even uniform hypergraph. *Journal of Combinatorial Optimization* 24, 564–579.
- Karypis, G. and V. Kumar (2000). Multilevel k-way hypergraph partitioning. *VLSI Design* 11(3), 285–300.
- Kernighan, B. W. and S. Lin (1970). An efficient heuristic procedure for partitioning graphs. *Bell system technical journal* 49(2), 291–307.
- Krzakala, F., C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences* 110(52), 20935–20940.
- Kumar, A., Y. Sabharwal, and S. Sen (2004). A simple linear time $(1 + \epsilon)$ -approximation algorithm for geometric k-means clustering in any dimensions. In *Proceedings-Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 454–462.
- Le, C. M., E. Levina, and R. Vershynin (2015). Sparse random graphs: Regularization and concentration of the Laplacian. *arXiv preprint arXiv:1502.03049*.
- Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics* 43(1), 215–237.

- Lei, J. and L. Zhu (2014). A generic sample splitting approach for refined community recovery in stochastic block models. *arXiv preprint arXiv:1411.1469*.
- Lichman, M. (2013). UCI machine learning repository. <http://archive.ics.uci.edu/ml>.
- Llyod, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28(2), 129–137.
- McSherry, F. (2001). Spectral partitioning of random graphs. In *Proceedings-Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 529–537.
- Michoel, T. and B. Nachtergaele (2012). Alignment and integration of complex networks by hypergraph-based spectral clustering. *Physical Review E* 86(056111).
- Ng, A., M. Jordan, and Y. Weiss (2002). On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pp. 849–856.
- Ostrovsky, R., Y. Rabani, L. J. Schulman, and C. Swamy (2012). The effectiveness of lloyd-type methods for the k-means problem. *Journal of the ACM* 59(6), 28:1–28:22.
- Panagiotou, K. and A. Coja-Oghlan (2012). Catching the k-NAESAT threshold. In *ACM Symposium on Theory of Computing*.
- Rodríguez, J. A. (2002). On the laplacian eigenvalues and metric parameters of hypergraphs. *Linear and Multilinear Algebra* 50, 1–14.
- Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics* 39(4), 1878–1915.
- Schmidt-Pruzan, J. and E. Shamir (1985). Component structure in the evolution of random hypergraphs. *Combinatorica* 5(1), 81–94.
- Schweikert, G. and B. W. Kernighan (1979). A proper model for the partitioning of electrical circuits. In *Proceedings of 9th Design Automation Workshop*, Dallas, pp. 57–62.
- Stasi, D., K. Sadeghi, A. Rinaldo, S. Petrović, and S. E. Fienberg (2014). β models for random hypergraphs with a given degree sequence. *arXiv preprint arXiv:1407.1004*.
- Stewart, G. W. and J. Sun (1990). *Matrix Perturbation Theory*. Academic Press.
- Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* 12(4), 389–434.
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing* 17(4), 395–416.
- Vu, V. (2014). A simple SVD algorithm for finding hidden partitions. *arXiv preprint arXiv:1404.3918*.
- Wasserman, S. (1994). *Social network analysis: Methods and applications*. Cambridge university press.
- Zhang, Y., E. Levina, and J. Zhu (2014). Detecting overlapping communities in networks with spectral methods. *arXiv preprint (arXiv:1412.3432v2)*.
- Zhou, D., J. Huang, and B. Schölkopf (2007). Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1601–1608.

DEPARTMENT OF COMPUTER SCIENCE & AUTOMATION,
INDIAN INSTITUTE OF SCIENCE,
BANGALORE - 560012, INDIA.

E-mail address: {debarghya.g,ad}@csa.iisc.ernet.in