PARTIAL RECOVERY AND WEAK CONSISTENCY IN THE NON-UNIFORM HYPERGRAPH STOCHASTIC BLOCK MODEL

IOANA DUMITRIU, HAIXIAO WANG, AND YIZHE ZHU

ABSTRACT. We consider the community detection problem in sparse random hypergraphs under the non-uniform hypergraph stochastic block model (HSBM), a general model of random networks with community structure and higher-order interactions. When the random hypergraph has bounded expected degrees, we provide a spectral algorithm that outputs a partition with at least a γ fraction of the vertices classified correctly, where $\gamma \in (0.5, 1)$ depends on the signal-to-noise ratio (SNR) of the model. When the SNR grows slowly as the number of vertices goes to infinity, our algorithm achieves weak consistency, which improves the previous results in [24] for non-uniform HSBMs.

Our spectral algorithm consists of three major steps: (1) Hyperedge selection: select hyperedges of certain sizes to provide the maximal signal-to-noise ratio for the induced sub-hypergraph; (2) Spectral partition: construct a regularized adjacency matrix and obtain an approximate partition based on singular vectors; (3) Correction and merging: incorporate the hyperedge information from adjacency tensors to upgrade the error rate guarantee. The theoretical analysis of our algorithm relies on the concentration and regularization of the adjacency matrix for sparse non-uniform random hypergraphs, which can be of independent interest.

1. INTRODUCTION

Clustering is one of the central problems in network analysis and machine learning [46, 51, 47]. Many clustering algorithms make use of graph models, which represent pairwise relationships among data. A well-studied probabilistic model is the stochastic block model (SBM), which was first introduced in [28] as a random graph model that generates community structure with given ground truth for clusters so that one can study algorithm accuracy. The past decades have brought many notable results in the analysis of different algorithms and fundamental limits for community detection in SBMs in different settings. A major breakthrough was the proof of phase transition behaviors of community detection algorithms in various connectivity regimes [41, 10, 43, 45, 44, 2, 4]. See the survey [1] for more references.

Hypergraphs can represent more complex relationships among data [9, 8], including recommendation systems [11, 38], computer vision [26, 55], and biological networks [42, 53], and they have been shown empirically to have advantages over graphs [61]. Besides community detection problems, sparse hypergraphs and their spectral theory have also found applications in data science [29, 62, 27], combinatorics [20, 22, 52], and statistical physics [12, 50].

With the motivation given by this broad set of applications, many efforts have been made in recent years to study community detection on random hypergraphs. The hypergraph stochastic block model (HSBM), as a generalization of graph SBM, was first introduced and studied in [23]. In this model, we observe a random uniform hypergraph where each hyperedge appears independently with some given probability depending on the community structure of the vertices in the hyperedge.

Succinctly put, the HSBM recovery problem is to find the ground truth clusters either approximately or exactly, given a sample hypergraph and estimates of model parameters. We may ask the following questions about the quality of the solutions (see [1] for further details in the graph case).

(1) **Strong consistency (a.k.a. exact recovery):** With high probability, find all clusters exactly (up to permutation).
(2) **Weak consistency:** With high probability, find a partition of the vertex set such that at most $o(n)$ vertices are misclassified.
(3) **Partial recovery:** Given a fixed $\gamma \in (0.5, 1)$, with high probability, find a partition of the vertex set such that at least a fraction γ of the vertices are clustered correctly.

Date: December 23, 2021.

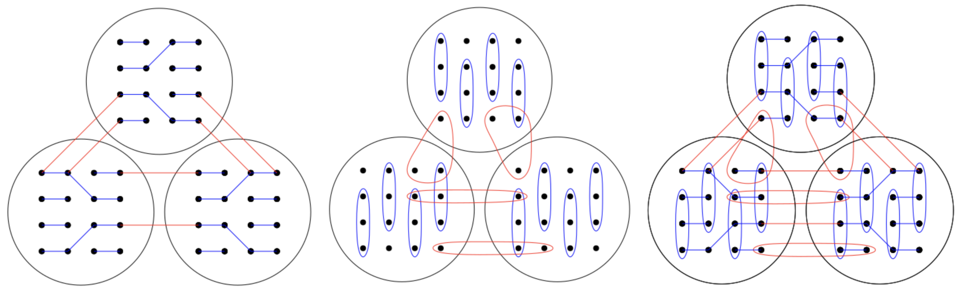

FIGURE 1. Uniform HSBM and Non-uniform HSBM.

(4) **Detection:** With high probability, find a partition correlated with the true partition.

For exact recovery of uniform HSBMs, it was shown that the phase transition occurs in the regime of logarithmic expected degrees in [39, 15, 14] and the exact threshold was given in [32, 59], by a generalization of the techniques in [2, 3]. Spectral methods were considered in [14, 5, 17, 59, 57], while semidefinite programming methods were analyzed in [32, 35]. Weak consistency for HSBMs was studied in [14, 15, 24, 25, 31]. For detection of the HSBM, the authors of [7] proposed a conjecture that the phase transition occurs in the regime of constant expected degrees, and the positive part of the conjecture for the binary case was solved in [49].

1.1. Non-uniform hypergraph stochastic block model.

The non-uniform HSBM was first studied in [24], which removed the uniform hypergraph assumption in previous works, and it is a more realistic model to study higher-order interaction on networks [40, 55]. It can be seen as a superposition of several uniform HSBMs with different model parameters. We first define the uniform HSBM in our setting and extend it to non-uniform hypergraphs.

Definition 1.1 (Uniform HSBM)**.** *Let $V = \{V_1, \dots V_k\}$ be a partition of the set $[n]$ into k blocks of size n/k (assuming n is divisible by k). For any set of m distinct vertices $i_1, \dots i_m$, a hyperedge $\{i_1, \dots i_m\}$ is generated with probability $a_m/\binom{n}{m-1}$ if the vertices $i_1, \dots i_m$ are in the same block; otherwise with probability $b_m/\binom{n}{m-1}$. We denote this distribution on the set of m-uniform hypergraphs as*

$$H_m \sim \mathrm{HSBM}_m\left(\frac{n}{k}, \frac{a_m}{\binom{n}{m-1}}, \frac{b_m}{\binom{n}{m-1}}\right), \quad 2 \le m \le M\,, \tag{1.1}$$

where $a_m/\binom{n}{m-1}$ and $b_m/\binom{n}{m-1}$ denote the connecting probabilities within and across blocks, respectively.

Definition 1.2 (Non-uniform HSBM)**.** *Let $H = (V, E)$ be a non-uniform random hypergraph, which can be considered as a collection of m-uniform hypergraphs, i.e., $H = \bigcup_{m=2}^{M} H_m$ with each H_m from Equation (1.1).*

Examples of 2-uniform and 3-uniform HSBM, and an example of non-uniform HSBM with $M = 3$ and $k = 3$ can be seen in Figure 1.

1.2. Main results.

In this paper, we consider *partial recovery* and *weak consistency* in the non-uniform HSBM given in Definition 1.2. The accuracy of the recovery is measured as follows.

Definition 1.3 (γ-correctness)**.** *Suppose we have k disjoint blocks $V_1, \dots, V_k$. A collection of subsets $\widehat{V}_1, \dots, \widehat{V}_k$ of V is γ-correct if $|V_i \cap \widehat{V}_i| \ge \gamma |V_i|$ for all $i \in [k]$. Note that we do not require $\widehat{V}_1, \dots, \widehat{V}_k$ to be a partition of V.*

The following theorem provides an algorithm that outputs a γ-correct partition of a non-uniform HSBM with high probability, when $k \ge 3$ is fixed.

Theorem 1.4 ($k \geq 3$). *Given any $\nu \in (0.5, 1)$, assume $a_m \geq b_m$ are constants independent of n for all $m \in \mathcal{M}$, where $\mathcal{M} \subset \{2, \dots, M\}$ is obtained from Algorithm 4.1, and*

$$d := \sum_{m\in\mathcal{M}} (m-1)a_m \geq C,$$

$$\sum_{m\in\mathcal{M}} (m-1)(a_m - b_m) \geq C_\nu \cdot k^{\mathcal{M}_{\max}-1}\sqrt{\log\left(\frac{k}{1-\nu}\right)d} \tag{1.2}$$

for some constants C, C_ν. Then with probability $1 - O(n^{-2})$, for sufficiently large n, Algorithm 1.1 outputs a γ-correct partition with $\gamma = \max\{\nu, 1-k\rho\}$ and $\rho := \exp(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}})$, where

$$\mathrm{SNR}_{\mathcal{M}} := \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}+b_m\right)}, \quad C_{\mathcal{M}} := \frac{[\nu^{\mathcal{M}_{\max}-1}-(1-\nu)^{\mathcal{M}_{\max}-1}]^2}{(\mathcal{M}_{\max}-1)^2\cdot 2^{2\mathcal{M}_{\max}+3}}. \tag{1.3}$$

Remark 1.5. *Taking $M = 2$, Theorem 1.4 can be reduced to [16, Lemma 9] for the graph case. The failure probability $O(n^{-2})$ can be decreased to $O(n^{-p})$ for any $p > 0$, as long as one is willing to pay a price in making the graph denser (increasing d).*

Our algorithm for Theorem 1.4 requires the input of model parameters $a_m, b_m, m \in \mathcal{M}$, and they can be estimated by counting cycles in hypergraphs as shown in [43, 56]. Estimation of the number of blocks can be done by counting the outliers in the spectrum of the non-backtracking operator [33, 7].

Our Algorithm 1.1 can be summarized in 3 steps:

(1) **Hyperedge selection**: select hyperedges of certain sizes to provide the maximal signal-to-noise ratio (SNR) for the induced sub-hypergraph.
(2) **Spectral partition**: construct a regularized adjacency matrix and obtain an approximate partition based on singular vectors (first approximation).
(3) **Correction and merging**: incorporate the hyperedge information from adjacency tensors to upgrade the error rate guarantee (second, better approximation).

Algorithm 1.1 Partition

Input: The adjacency tensors $\boldsymbol{\mathcal{T}}_m$, k, a_m, b_m for $m \in \{2, \cdots, M\}$.
Output: The corrected sets $\widehat{V}_1, \cdots, \widehat{V}_k$.

1: **function Partition**($\mathbf{A}$, a_m, b_m for $m \in \mathcal{M}$)
2: Run **Algorithm 4.1 Pre-processing** to obtain the subset $\mathcal{M}$ which achieves maximal SNR.
3: Randomly color the hyperedges red or blue with equal probability.
4: Randomly partition V into 2 disjoint subsets Y and Z by assigning $+1$ or -1 to each vertex with equal probability.
5: Let $\mathbf{B}$ denote the adjacency matrix of the bipartite hypergraph between Y and Z consisting only of the red hyperedges, with rows indexed by Z and columns indexed by Y.
6: Run **Algorithm 4.2 Spectral Partition** on the red hypergraph and output $U'_1, \cdots, U'_k$.
This step only uses the red hyperedges between vertices in Y and Z and outputs approximate clusters for $U_i := V_i \cap Z$, with $i = 1, \dots, k$
7: Run **Algorithm 4.3 Correction** on the red hypergraph and output $\widehat{U}_1, \cdots, \widehat{U}_k$.
8: Run **Algorithm 4.4 Merging** on the blue hypergraph and output $\widehat{V}_1, \cdots, \widehat{V}_k$.
This step only uses the blue hyperedges between vertices in Y and Z and assigns the vertices in Y to an appropriate approximate cluster.
9: **end function**

We generalize the innovative graph algorithm in [16] to uniform hypergraphs with a more detailed analysis. In particular, the random partitioning step in Algorithm 1.2 was explained only in the case when the output is an equi-partition; in general, the output sets $\widehat{V}_1, \dots, \widehat{V}_k$ are only approximately equal in size, and we give complete explanations of the general case here.

A non-uniform HSBM can be seen as a collection of noisy observations for the same underlying community structure through several uniform HSBMs of different orders. A possible issue is that some uniform hypergraphs with small SNR might not be informative (if we observe an m-uniform hypergraph with parameters $a_m = b_m$, including hyperedge information from it ultimately increases the noise). To improve our error rate guarantees, we start by adding a pre-processing step for hyperedge selection according to SNR, and then apply the algorithm on the sub-hypergraph with maximal SNR.

When the number of blocks is 2, we give a simpler algorithm (Algorithm 1.2) for partial recovery with a better error rate guarantee, as shown in the following Theorem 1.6. Algorithm 1.2 does not need the merging routine in Algorithm 1.1, since once we cluster one block, the other one is clustered automatically.

Theorem 1.6 ($k = 2$)**.** *Given $\nu \in (0.5, 1)$, assume $a_m \geq b_m$ are constants independent of n for all $m \in \mathcal{M}$,*

$$d := \sum_{m\in\mathcal{M}} (m-1)a_m \geq C\,,$$
$$\sum_{m\in\mathcal{M}} (m-1)(a_m - b_m) \geq C_\nu \cdot 2^{\mathcal{M}_{\max}+2}\sqrt{d}\,,$$

for some constants C, C_ν. Then with probability $1 - O(n^{-2})$, for sufficiently large n, Algorithm 1.2 outputs a γ-correct partition with $\gamma = \max\{\nu,\, 1-2\rho\}$ and $\rho = 2\exp(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}})$, where

$$\mathrm{SNR}_{\mathcal{M}} := \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{2^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{2^{m-1}}+b_m\right)}\,, \quad C_{\mathcal{M}} := \frac{[(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}]^2}{8(\mathcal{M}_{\max}-1)^2}\,. \tag{1.5}$$

Algorithm 1.2 Binary Partition

Input: The adjacency tensors $\boldsymbol{\mathcal{T}}^{(m)}$, a_m, b_m for $m \in \{2, \cdots, M\}$.
Output: The corrected sets $\widehat{V}_1, \widehat{V}_2$.

function Partition($\boldsymbol{\mathcal{T}}^{(m)}$, a_m, b_m for $m \in \{2, \cdots, M\}$)
  Run **Algorithm 4.1 Pre-processing** to obtain the subset $\mathcal{M}$ which achieves maximal SNR.
  Randomly color the hyperedges red or blue with equal probability.
  Run **Algorithm 4.5 Spectral Partition** on the red hypergraph and output V_1', V_2'.
  Run **Algorithm 4.6 Correction** on the blue hypergraph and output $\widehat{V}_1, \widehat{V}_2$.
  return the corrected sets $\widehat{V}_1, \widehat{V}_2$.
end function

Throughout the proofs for Theorems 1.4 and 1.6, we make only one assumption on the growth or finiteness of d and $\mathrm{SNR}_{\mathcal{M}}$, and that happens in estimating the failure probability in the proof of Lemma 5.9 as noted in Remark 5.13. As a consequence, we can also give the following corollary, which covers the case when d and $\mathrm{SNR}_{\mathcal{M}}$ grow with n.

Corollary 1.7 (Weak consistency)**.** *For fixed M and k, if*

$$\mathrm{SNR}_{\mathcal{M}} := \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}+b_m\right)} \to \infty\,,$$

and $\mathrm{SNR}_{\mathcal{M}} = o(\log n)$, then with probability $1 - O(n^{-2})$, Algorithm 1.1 outputs a partition with only $o(n)$ misclassified vertices.

Here, the failure probability is of order $O(\exp[-n\exp(-\mathrm{SNR}_{\mathcal{M}})])$, and it fails to go to zero if $\mathrm{SNR}_{\mathcal{M}}$ is not smaller than $o(\log(n))$.

1.3. Comparison with existing results.

Although many algorithms and theoretical results have been developed for hypergraph community detection, most of them are restricted to uniform hypergraphs, and few results are known for non-uniform ones. We will discuss the most relevant ones here and compare our results to theirs.

In [31], the authors studied the degree-corrected HSBM with general connection probability parameters by using a tensor power iteration algorithm and Tucker decomposition. Their algorithm achieves weak consistency for uniform hypergraphs when the average degree is $\omega(\log^2 n)$, which is the regime complementary to the regime we studied here. They discussed a way to generalize the algorithm to non-uniform hypergraphs, but the theoretical analysis remains open. The recent paper [60] analyzed non-uniform hypergraph community detection by using hypergraph embedding and optimization algorithms and obtained weak consistency when the expected degrees are of $\omega(\log n)$, again a complementary regime to ours. Results on spectral norm concentration of sparse random tensors were obtained in [19, 48, 29, 36, 62], but no provable tensor algorithm in the bounded expected degree is known. Testing for the community structure for non-uniform hypergraphs was studied in [56, 30], which is a problem different from community detection.

In our approach, we relied on knowing the tensors for each uniform hypergraph. However, in computations, we only ran the spectral algorithm on the adjacency matrix of the entire hypergraph, since the stability of tensor algorithms does not yet come with guarantees due to the lack of concentration, and for non-uniform hypergraphs, $M-1$ adjacency tensors would be needed. This approach presented the challenge that, unlike for graphs, the adjacency matrix of a random non-uniform hypergraph has dependent entries, and the concentration properties of such a random matrix were previously unknown. We were able to overcome this issue and prove concentration bounds from scratch, down to the bounded degree regime. Similar to [21, 34], we provided here a regularization analysis by removing rows in the adjacency matrix with large row sums (suggestive of large degree vertices) and proving a concentration result for the regularized matrix down to the bounded expected degree regime (see Theorem 3.3).

In terms of partial recovery for hypergraphs, our results are new, even in the uniform case. In [6, Theorem 1], for uniform hypergraphs, the authors showed detection (not partial recovery) is possible when the average degree is $\Omega(\frac{1}{n})$; in addition, the error rate is not exponential in the model parameters. In the graph case, it was shown in [58] that the error rate in Equation (1.3) is optimal up to a constant in the exponent. It's an interesting open problem to extend the analysis in [58] to obtain a minimax error rate for hypergraphs.

In [24], the authors considered weak consistency in a non-uniform HSBM model with a spectral algorithm based on the hypergraph Laplacian matrix, and showed that weak consistency is achievable if the expected degree is of $\Omega(\log^2 n)$ with high probability [23, Theorem 4.2]. Their algorithm can't be applied to sparse regime straightforwardly since the normalized Laplacian is not well-defined due to the existence of isolated vertices in the bounded degree case. In addition, our weak consistency results obtained here are valid as long as the expected degree is $\omega(1)$ and $o(\log n)$, which is the entire set of problems on which weak consistency is expected. By contrast, in [24], weak consistency is shown only when the expected degree is $\Omega(\log^2(n))$, which is a regime complementary to ours and where exact recovery should (in principle) be possible: for example, this is known to be an exact recovery regime in the uniform case [15, 32, 35, 59].

Finally, although our analysis does not provide a way to identify the exact threshold for the phase transition of detection in the non-uniform hypergraph, we conjecture that (similar to [7] in the uniform case) $\mathrm{SNR}_{\mathcal{M}} = 1$ in Equation (2.6) is the exact threshold for detection.

1.4. **Organization of the paper.** In Section 2, we include the definitions of adjacency matrices of hypergraphs. The concentration results for the adjacency matrices are provided in Section 3. The algorithms for partial recovery are presented in Section 4. The proof for correctness of our algorithms for Theorem 1.4 and Corollary 1.7 are given in Section 5. The proof of Theorem 1.6 as well as the proofs of many auxiliary lemmas, and useful lemmas in the literature are provided in the Appendices.

2. Preliminaries

Definition 2.1 (Adjacency tensor). *Given an m-uniform hypergraph $H_m = ([n], E_m)$, we can associate to it an order-m adjacency tensor $\boldsymbol{\mathcal{T}}^{(m)}$. For any m-hyperedge $e = \{i_1, \dots, i_m\}$, let $\boldsymbol{\mathcal{T}}_e^{(m)}$ denote the corresponding entry $\boldsymbol{\mathcal{T}}_{[i_1,\dots,i_m]}^{(m)}$, such that*

$$\boldsymbol{\mathcal{T}}_e^{(m)} := \boldsymbol{\mathcal{T}}_{[i_1,\dots,i_m]}^{(m)} = \mathbb{1}_{\{e \in E_m\}} \,. \tag{2.1}$$

Definition 2.2 (Adjacency matrix). *For the non-uniform hypergraph H (Definition 1.2), let $\boldsymbol{\mathcal{T}}^{(m)}$ be the order-m adjacency tensor of the underlying m-uniform hypergraph for each $m \in \{2, \dots, M\}$. The adjacency* *matrix* $\mathbf{A} := [\mathbf{A}_{ij}]_{n\times n}$ *of the non-uniform hypergraph* H *is defined by*

$$\mathbf{A}_{ij} = \mathbb{1}_{\{i\neq j\}} \cdot \sum_{m=2}^{M} \sum_{\substack{e\in E_m\\ \{i,j\}\subset e}} \boldsymbol{\mathcal{T}}_e^{(m)} \,. \tag{2.2}$$

We compute the expectation of $\mathbf{A}$ first. In each H_m, we pick two vertices $i, j \in V$ arbitrarily, then the expected number of m-hyperedges containing i and j can be computed as follows. Since our model does not allow for loops, in the below, we only consider the case $i \neq j$. Recall that $n/k \in \mathbb{N}$.

- If i and j are from the same block, the m-hyperedge can either be sampled with probability $a_m/\binom{n}{m-1}$ when the other $m-2$ vertices are from the same block as i and j, otherwise they are sampled with probability $b_m/\binom{n}{m-1}$. Then
$$\alpha_m := \mathbb{E}\mathbf{A}_{ij} = \binom{\frac{n}{k}-2}{m-2}\frac{a_m}{\binom{n}{m-1}} + \left[\binom{n-2}{m-2} - \binom{\frac{n}{k}-2}{m-2}\right]\frac{b_m}{\binom{n}{m-1}} \,.$$
- If i and j are not from the same block, we sample the m-hyperedge with probability $b_m/\binom{n}{m-1}$, and
$$\beta_m := \mathbb{E}\mathbf{A}_{ij} = \binom{n-2}{m-2}\frac{b_m}{\binom{n}{m-1}} \,.$$

By assumption $a_m \geq b_m$, then $\alpha_m \geq \beta_m$ for each $m \in \{2, \dots, M\}$. Summing over m, the *expected adjacency* matrix under the k-block non-uniform HSBM can written as

$$\mathbb{E}\mathbf{A} = \begin{bmatrix} \alpha\mathbf{J}_{\frac{n}{k}} & \beta\mathbf{J}_{\frac{n}{k}} & \cdots & \beta\mathbf{J}_{\frac{n}{k}} \\ \beta\mathbf{J}_{\frac{n}{k}} & \alpha\mathbf{J}_{\frac{n}{k}} & \cdots & \beta\mathbf{J}_{\frac{n}{k}} \\ \vdots & \vdots & \ddots & \vdots \\ \beta\mathbf{J}_{\frac{n}{k}} & \beta\mathbf{J}_{\frac{n}{k}} & \cdots & \alpha\mathbf{J}_{\frac{n}{k}} \end{bmatrix} - \alpha\mathbf{I}_n \,, \tag{2.3}$$

with

$$\alpha := \sum_{m=2}^{M} \alpha_m \,, \quad \beta := \sum_{m=2}^{M} \beta_m \,, \tag{2.4}$$

where $\mathbf{J}_{\frac{n}{k}} \in \mathbb{R}^{\frac{n}{k}\times\frac{n}{k}}$ denotes the all-one matrix. The Lemma 2.3 can be verified through direct computation.

Lemma 2.3. *The eigenvalues of* $\mathbb{E}\mathbf{A}$ *are given below:*

$$\begin{aligned} \lambda_1(\mathbb{E}\mathbf{A}) &= \frac{n}{k}(\alpha + (k-1)\beta) - \alpha \,, & \\ \lambda_i(\mathbb{E}\mathbf{A}) &= \frac{n}{k}(\alpha - \beta) - \alpha \,, & 2 \leq i \leq k \,, \\ \lambda_i(\mathbb{E}\mathbf{A}) &= -\alpha \,, & k+1 \leq i \leq n \,. \end{aligned}$$

In our proof, Lemma 2.4 is used for more approximately equal partitions, meaning that we can approximate the eigenvalues of $\widetilde{\mathbb{E}\mathbf{A}}$ by eigenvalues $\mathbb{E}\mathbf{A}$ when n is sufficiently large.

Lemma 2.4. *For any partition* $(V_1, \dots, V_k)$ *of* V, *consider the following matrix*

$$\widetilde{\mathbb{E}\mathbf{A}} = \begin{bmatrix} \alpha\mathbf{J}_{n_1} & \beta\mathbf{J}_{n_1\times n_2} & \cdots & \beta\mathbf{J}_{n_1\times n_{k-1}} & \beta\mathbf{J}_{n_1\times n_k} \\ \beta\mathbf{J}_{n_2\times n_1} & \alpha\mathbf{J}_{n_2} & \cdots & \beta\mathbf{J}_{n_2\times n_{k-1}} & \beta\mathbf{J}_{n_2\times n_k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta\mathbf{J}_{n_{k-1}\times n_1} & \beta\mathbf{J}_{n_{k-1}\times n_2} & \cdots & \alpha\mathbf{J}_{n_{k-1}} & \beta\mathbf{J}_{n_{k-1}\times n_k} \\ \beta\mathbf{J}_{n_k\times n_1} & \beta\mathbf{J}_{n_k\times n_2} & \cdots & \beta\mathbf{J}_{n_k\times n_{k-1}} & \alpha\mathbf{J}_{n_k} \end{bmatrix} - \alpha\mathbf{I}_n \,,$$

where $n_i := |V_i|$. *Assume that* $n_i = (n/k) + O(\sqrt{n}\log n)$ *for all* $i \in [k]$. *Then, for all* $1 \leq i \leq k$,

$$\frac{|\lambda_i(\widetilde{\mathbb{E}\mathbf{A}}) - \lambda_i(\mathbb{E}\mathbf{A})|}{|\lambda_i(\mathbb{E}\mathbf{A})|} = O\Big(n^{-1/4}\log^{1/2}(n)\Big) \,.$$

Note that both $(\widetilde{\mathbb{E}\mathbf{A}} + \alpha\mathbf{I}_n)$ *and* $(\mathbb{E}\mathbf{A} + \alpha\mathbf{I}_n)$ *are rank* k *matrices, then* $\lambda_i(\widetilde{\mathbb{E}\mathbf{A}}) = \lambda_i(\mathbb{E}\mathbf{A}) = -\alpha$ *for all* $(k+1) \leq i \leq n$.

Definition 2.5 (signal-to-noise ratio, SNR). *We define the signal-to-noise ratio as*

$$\mathrm{SNR} := \frac{\left[\sum_{m=2}^{M}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}\right)\right]^2}{\sum_{m=2}^{M}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}+b_m\right)}. \tag{2.6}$$

SNR is related to the following quantity

$$\frac{[\lambda_2(\mathbb{E}\mathbf{A})]^2}{\lambda_1(\mathbb{E}\mathbf{A})} = \frac{[(n-k)\alpha - n\beta]^2}{k[(n-k)\alpha + n(k-1)\beta]} = \frac{\left[\sum_{m=2}^{M}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}\right)\right]^2}{\sum_{m=2}^{M}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}+b_m\right)}(1+o(1)).$$

When $M=2$ and k is fixed, (2.6) is equal to $\frac{(a-b)^2}{k[a+(k-1)b]}$, which corresponds to the SNR for the undirected graph in [16], see also [1, Section 6].

3. Spectral norm concentration

The correctness of Algorithm 1.1 and Algorithm 1.2 relies on the concentration of the adjacency matrix of H. We include the following two concentration results for general random hypergraphs. The proofs are given in Appendix A.

Theorem 3.1. *Let $H = \bigcup_{m=2}^{M} H_m$, where $H_m = ([n], E_m)$ is an Erdős-Rényi inhomogeneous hypergraph of order m with $2 \leq m \leq M$ with a probability tensor $\boldsymbol{\mathcal{T}}^{(m)}$ such that $\boldsymbol{\mathcal{T}}^{(m)}_{[i_1,\dots,i_m]} = d_{[i_1,\dots,i_m]}/\binom{n}{m-1}$ and $d_m = \max d_{[i_1,\dots,i_m]}$. Suppose for some constant $c>0$,*

$$d := \sum_{m=2}^{M}(m-1)\cdot d_m \geq c\log n\,. \tag{3.1}$$

Then for any $K>0$, there exists a constant $C = 512M(M-1)(K+6)\left[2+(M-1)(1+K)/c\right]$ such that with probability at least $1-2n^{-K}-2e^{-n}$, the adjacency matrix $\mathbf{A}$ of H satisfies

$$\|\mathbf{A}-\mathbb{E}\mathbf{A}\| \leq C\sqrt{d}\,. \tag{3.2}$$

Taking $M=2$, Equation (3.2) is the result for graph case obtained in [21, 37]. Result for uniform hypergraph is obtained in [35]. Note that d is a fixed constant in our community detection problem, thus Equation (3.1) does not hold and Theorem 3.1 cannot be directly applied. However, we can still prove a concentration bound for a regularized version of $\mathbf{A}$, following the strategy of the proof for Theorem 3.1.

Definition 3.2 (Regularized matrix). *Given any $n\times n$ matrix $\mathbf{A}$ and an index set $\mathcal{I}$, let $\mathbf{A}_{\mathcal{I}}$ be the $n\times n$ matrix obtained from $\mathbf{A}$ by zeroing out the rows and columns not indexed by $\mathcal{I}$. Namely,*

$$(\mathbf{A}_{\mathcal{I}})_{ij} = \mathbb{1}_{\{i,j\in\mathcal{I}\}}\cdot \mathbf{A}_{ij}\,. \tag{3.3}$$

Since every hyperedge of size m containing i is counted $(m-1)$ times in the i-th row sum of $\mathbf{A}$, the i-th row sum of $\mathbf{A}$ is given by

$$\mathrm{row}(i) := \sum_j \mathbf{A}_{ij} := \sum_j \mathbb{1}_{\{i\neq j\}}\sum_{m=2}^{M}\sum_{\substack{e\in E_m\\ \{i,j\}\subset e}}\boldsymbol{\mathcal{T}}^{(m)}_e = \sum_{m=2}^{M}(m-1)\sum_{e\in E_m:\, i\in e}\boldsymbol{\mathcal{T}}^{(m)}_e.$$

Theorem 3.3 is the concentration result for the regularized $\mathbf{A}_{\mathcal{I}}$, by zeroing out rows and columns corresponding to vertices with high row sums.

Theorem 3.3. *Following all the notations in Theorem 3.1, for any constant $\tau>1$, define*

$$\mathcal{I} = \{i\in[n] : \mathrm{row}(i)\leq \tau d\}.$$

Let $\mathbf{A}_{\mathcal{I}}$ be the regularized version of $\mathbf{A}$, as in Definition 3.2. Then for any $K>0$, there exists a constant C_τ depending on M,τ,K, such that $\|(\mathbf{A}-\mathbb{E}\mathbf{A})_{\mathcal{I}}\| \leq C_\tau\sqrt{d}$ with probability at least $1-2(e/2)^{-n}-n^{-K}$.

4. Spectral algorithms

In this section, we present the algorithmic blocks that we use to construct our main partition algorithm (Algorithm 1.1): pre-processing (Algorithm 4.1), first attempt at partition (Algorithm 4.2), correction of blemishes via majority rule (Algorithm 4.3), and merging (Algorithm 4.4), together with their proofs of correctness. We start by giving the algorithms.

Algorithm 4.1 Pre-processing

function Pre-processing(Parameters a_m, b_m for all $m \in \{2, \cdots, M\}$)

For each subset $\mathcal{S} \subset \{2, \cdots, M\}$, let $H_{\mathcal{S}} = \bigcup_{m \in \mathcal{S}} H_m$ denote the restriction of the non-uniform hypergraph H on $\mathcal{S}$. Compute SNR of $H_{\mathcal{S}}$, denoted by

$$\mathrm{SNR}_{\mathcal{S}} := \frac{\left[\sum_{m\in\mathcal{S}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{S}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}} + b_m\right)} .$$

Among all the $\mathcal{S}$, find the subset $\mathcal{M}$ such that

$$\mathcal{M} := \arg \max_{\mathcal{S} \subset \{2,\cdots,M\}} \mathrm{SNR}_{\mathcal{S}} ,$$

with $\mathcal{M}_{\max}$ denoting its maximal element.

return $\mathcal{M}$.

end function

Algorithm 4.2 Spectral Partition

function Spectral Partition($\mathbf{B}$, k, a_m, b_m for $m \in \mathcal{M}$)

Randomly label vertices in Y with $+1$ and -1 sign with equal probability, and partition Y into 2 disjoint subsets Y_1 and Y_2.

Let $\mathbf{B}_1$ (resp. $\mathbf{B}_2$) denote the adjacency matrices with all vertices in $Z \cup Y_1$, with rows indexed by Z and columns indexed by Y_1 (resp. Y_2). Pad $\mathbf{B}_1$, $\mathbf{B}_2$ with zeros to obtain the $n \times n$ matrices $\mathbf{A}_1$ and $\mathbf{A}_2$.

Let $d := \sum_{m\in\mathcal{M}}(m-1)a_m$. Zero out all the rows and columns of $\mathbf{A}_1$ corresponding to vertices whose row sum is bigger than $20\mathcal{M}_{\max}d$, to obtain the matrix $(\mathbf{A}_1)_{\mathcal{I}_1}$.

Find the space $\mathbf{U}$, spanned by the first k left singular vectors of $(\mathbf{A}_1)_{\mathcal{I}_1}$.

Randomly sample $s = 2k\log^2 n$ vertices from Y_2 without replacement. Denote the corresponding columns in $\mathbf{A}_2$ by $\boldsymbol{a}_{i_1}, \dots, \boldsymbol{a}_{i_s}$. For each $i \in \{i_1, \cdots, i_s\}$, project $\boldsymbol{a}_i - \overline{\boldsymbol{a}}$ onto $\mathbf{U}$, where the elements in vector $\overline{\boldsymbol{a}} \in \mathbb{R}^n$ is defined by $\overline{\boldsymbol{a}}(j) = \mathbb{1}_{j\in Z} \cdot (\overline{\alpha} + \overline{\beta})/2$

For each projected vector $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$, identify the top $n/(2k)$ coordinates in value as a set U_i. Discard half of the s sets U_i, those with the lowest blue hyperedge density in them.

Sort the remaining sets according to blue hyperedge density and identify first k distinct subsets $U'_1, \cdots, U'_k$ such that $|U'_i \cap U'_j| < \lceil (1-\nu)n/k \rceil$ if $i \neq j$.

return $U'_1, \cdots, U'_k$.

end function

Algorithm 4.3 Correction

function Correction(A collection of subsets $U'_1, \cdots, U'_k \subset Z$ and a graph on Z)

For every $u \in Z$, if $i \in \{1, 2, \cdots, k\}$ is such that u has more neighbors in U'_i than in any other U'_j, for $j \neq i$, then add u to $\widehat{U}_i$. Break ties arbitrarily.

return $\widehat{U}_1, \cdots, \widehat{U}_k$.

end function

4.1. **Three or more blocks ($k \geq 3$).** The proof of Theorem 1.4 is structured as follows.

Algorithm 4.4 Merging

function Merging(A partition $\widehat{U}_1, \cdots, \widehat{U}_k$ of $(Z \cap V_1) \cup (Z \cap V_2) \cup \cdots \cup (Z \cap V_k)$ and a graph between vertices Y and Z.)
  For all $u \in Y$, add u to $\widehat{U_i}$ if the number of hyperedges, which contains u with the remaining vertices located in vertex set $\widehat{U}_i$, is at least μ_{M}. Label the conflicts arbitrarily.
  return the label classes as the clusters $\widehat{V}_1, \cdots, \widehat{V}_k$.
end function

Lemma 4.1. *Under the assumptions of Theorem 1.4, Algorithm 4.2 outputs a ν-correct partition $U_1', \cdots, U_k'$ of $Z = (Z \cap V_1) \cup \cdots \cup (Z \cap V_k)$ with probability at least $1 - O(n^{-2})$.*

Lemma 4.2. *Under the assumptions of Theorem 1.4, given any ν-correct partition $U_1', \cdots, U_k'$ of $Z = (Z \cap V_1) \cup \cdots \cup (Z \cap V_k)$ and the red hypergraph over Z, Algorithm 4.3 computes a γ_{C}-correct partition $\widehat{U}_1, \cdots, \widehat{U}_k$ with probability $1 - O(e^{-n\rho})$, while $\gamma_{\mathrm{C}} = \max\{\nu, 1 - k\rho\}$ with $\rho := k \exp\left(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}}\right)$ where $\mathcal{M}$ is obtained from Algorithm 4.1, and $\mathrm{SNR}_{\mathcal{M}}$ and $C_{\mathcal{M}}$ are defined in Equation (1.3).*

Lemma 4.3. *Given any ν-correct partition $\widehat{U}_1, \cdots, \widehat{U}_k$ of $Z = (Z \cap V_1) \cup \cdots \cup (Z \cap V_k)$ and the blue hypergraph between Y and Z, with probability $1 - O(e^{-n\rho})$, Algorithm 4.4 outputs a γ-correct partition $\widehat{V}_1, \cdots, \widehat{V}_k$ of $V_1 \cup V_2 \cup \cdots \cup V_k$, while $\gamma = \max\{\nu, 1 - k\rho\}$.*

4.2. **The binary case ($k = 2$).** The spectral partition step is given in Algorithm 4.5, and the correction step is given in Algorithm 4.6.

Algorithm 4.5 Spectral Partition

function Spectral Partition($\mathbf{A}$, $\boldsymbol{\mathcal{T}}$)
  Zero out all the rows and the columns of $\mathbf{A}$ corresponding to vertices with row sums greater than $20\mathcal{M}_{\max} d$, to obtain the regularized matrix $\mathbf{A}_{\mathcal{I}}$.
  Find the subspace $\mathbf{U}$, which is spanned by the eigenvectors corresponding to the largest two eigenvalues of $\mathbf{A}_{\mathcal{I}}$.
  Compute $P_{\mathbf{U}} \mathbf{1}_n$, the projection of all-ones vector onto $\mathbf{U}$.
  Let $\boldsymbol{v}$ be the unit vector in $\mathbf{U}$ perpendicular to $P_{\mathbf{U}} \mathbf{1}_n$.
  Sort the vertices according to their values in $\boldsymbol{v}$. Let $V_1' \subset V$ be the corresponding top $n/2$ vertices, and $V_2' \subset V$ be the remaining $n/2$ vertices.
  return V_1', V_2'.
end function

Lemma 4.4. *Under the conditions of Theorem 1.6, the Algorithm 4.5 outputs a ν-correct partition V_1', V_2' of $V = V_1 \cup V_2$ with probability at least $1 - O(n^{-2})$.*

Algorithm 4.6 Correction

function Correction($\boldsymbol{\mathcal{T}}$, partition V_1', V_2', and a blue graph on $V_1' \cup V_2'$)
  For any $v \in V_1'$, label v "bad" if the number of blue neighbors of v in V_2' is at least $\frac{a+b}{4}$ and "good" otherwise.
  Do the same for $v \in V_2'$.
  If $v \in V_i'$ is good, assign it to $\widehat{V}_i$, otherwise $\widehat{V}_{3-i}$.
  return $\widehat{V}_1$, $\widehat{V}_2$.
end function

Lemma 4.5. *Given any ν-correct partition V_1', V_2' of $V = V_1 \cup V_2$, with probability at least $1 - O(e^{-n\rho})$, the Algorithm 4.6 computes a γ-correct partition $\widehat{V}_1, \widehat{V}_2$ with $\gamma = \{\nu, 1 - 2\rho\}$ and $\rho = 2\exp(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}})$, where $\mathrm{SNR}_{\mathcal{M}}$ and $C_{\mathcal{M}}$ are defined in Equation (1.5).*

5. Algorithm's correctness

In this section, we will show the correctness of Algorithm 1.1. We first introduce some definitions.

Vertex set splitting and adjacency matrix. In Algorithm 1.1, we first randomly partition the vertex set V into two disjoint subsets Z and Y by assigning $+1$ and -1 to each vertex independently with equal probability. Let $\mathbf{B} \in \mathbb{R}^{|Z| \times |Y|}$ denote the submatrix of $\mathbf{A}$, while $\mathbf{A}$ was defined in Equation (2.2), where rows and columns of $\mathbf{B}$ correspond to vertices in Z and Y respectively. Let n_i denote the number of vertices in $Z \cap V_i$, where V_i denotes the true partition with $|V_i| = n/k$ for all $i \in [k]$, then n_i can be written as a sum of independent Bernoulli random variables, i.e.,

$$n_i = |Z \cap V_i| = \sum_{v \in V_i} \mathbb{1}_{\{v \in Z\}}\,, \tag{5.1}$$

and $|Y \cap V_i| = |V_i| - |Z \cap V_i| = n/k - n_i$ for each $i \in [k]$.

Definition 5.1. *The splitting $V = Z \cup Y$ is perfect if $|Z \cap V_i| = |Y \cap V_i| = n/(2k)$ for all $i \in [k]$. And the splitting $Y = Y_1 \cup Y_2$ is perfect if $|Y_1 \cap V_i| = |Y_2 \cap V_i| = n/(4k)$ for all $i \in [k]$.*

However, the splitting will actually be *imperfect* in most cases, since the size of Z and Y would not be exactly the same under the independence assumption. The random matrix $\mathbf{B}$ is parameterized by $\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m \in \mathcal{M}}$ and $\{n_i\}_{i=1}^k$. If we take expectation over $\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m \in \mathcal{M}}$ given the block size information $\{n_i\}_{i=1}^k$, then it gives rise to the expectation of the *imperfect* splitting, denoted by $\widetilde{\mathbf{B}}$,

$$\widetilde{\mathbf{B}} := \begin{bmatrix} \alpha \mathbf{J}_{n_1 \times (\frac{n}{k} - n_1)} & \beta \mathbf{J}_{n_1 \times (\frac{n}{k} - n_2)} & \cdots & \beta \mathbf{J}_{n_1 \times (\frac{n}{k} - n_k)} \\ \beta \mathbf{J}_{n_2 \times (\frac{n}{k} - n_1)} & \alpha \mathbf{J}_{n_2 \times (\frac{n}{k} - n_2)} & \cdots & \beta \mathbf{J}_{n_2 \times (\frac{n}{k} - n_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta \mathbf{J}_{n_k \times (\frac{n}{k} - n_1)} & \beta \mathbf{J}_{n_k \times (\frac{n}{k} - n_2)} & \cdots & \alpha \mathbf{J}_{n_k \times (\frac{n}{k} - n_k)} \end{bmatrix},$$

where α, β are defined in Equation (2.4). In the *perfect* splitting case, the dimension of each block is $n/(2k) \times n/(2k)$ since $\mathbb{E} n_i = n/(2k)$ for all $i \in [k]$, and the expectation matrix $\overline{\mathbf{B}}$ can be written as

$$\overline{\mathbf{B}} := \begin{bmatrix} \alpha \mathbf{J}_{\frac{n}{2k}} & \beta \mathbf{J}_{\frac{n}{2k}} & \cdots & \beta \mathbf{J}_{\frac{n}{2k}} \\ \beta \mathbf{J}_{\frac{n}{2k}} & \alpha \mathbf{J}_{\frac{n}{2k}} & \cdots & \beta \mathbf{J}_{\frac{n}{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \beta \mathbf{J}_{\frac{n}{2k}} & \beta \mathbf{J}_{\frac{n}{2k}} & \cdots & \alpha \mathbf{J}_{\frac{n}{2k}} \end{bmatrix}.$$

In Algorithm 4.2, Y_1 is a random subset of Y obtained by selecting each element with probability $1/2$ independently, and $Y_2 = Y \setminus Y_1$. Let n'_i denote the number of vertices in $Y_1 \cap V_i$, then n'_i can be written as a sum of independent Bernoulli random variables,

$$n'_i = |Y_1 \cap V_i| = \sum_{v \in V_i} \mathbf{1}_{\{v \in Y_1\}}\,, \tag{5.2}$$

and $|Y_2 \cap V_i| = |V_i| - |Z \cap V_i| - |Y_1 \cap V_i| = n/k - n_i - n'_i$ for all $i \in [k]$.

Induced sub-hypergraph.

Definition 5.2 (Induced Sub-hypergraph)**.** *Let $H = (V, E)$ be a non-uniform random hypergraph and $S \subset V$ be any subset of the vertices of H. Then the induced sub-hypergraph $H[S]$ is the hypergraph whose vertex set is S and whose hyperedge set $E[S]$ consists of all of the edges in E that have all endpoints located in S.*

Let $H[Y_1 \cup Z]$(resp. $H[Y_2 \cup Z]$) denote the induced sub-hypergraph on vertex set $Y_1 \cup Z$ (resp. $Y_2 \cup Z$), and $\mathbf{B}_1 \in \mathbb{R}^{|Z| \times |Y_1|}$ (resp. $\mathbf{B}_2 \in \mathbb{R}^{|Z| \times |Y_2|}$) denote the adjacency matrices corresponding to the sub-hypergraphs, where rows and columns of $\mathbf{B}_1$ (resp. $\mathbf{B}_2$) are corresponding to elements in Z and Y_1 (resp., Z and Y_2). Therefore, $\mathbf{B}_1$ and $\mathbf{B}_2$ are parameterized by $\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m \in \mathcal{M}}$, $\{n_i\}_{i=1}^k$ and $\{n'_i\}_{i=1}^k$, and the entries in $\mathbf{B}_1$ are independent of the entries in $\mathbf{B}_2$, due to the independence of hyperedges. If we take expectation over

$\{\mathcal{T}^{(m)}\}_{m\in\mathcal{M}}$ conditioning on $\{n_i\}_{i=1}^k$ and $\{n_i'\}_{i=1}^k$, then it gives rise to the expectation of the *imperfect* splitting, denoted by $\widetilde{\mathbf{B}}_1$,

$$\widetilde{\mathbf{B}}_1 := \begin{bmatrix} \widetilde{\alpha}_{11}\mathbf{J}_{n_1\times n_1'} & \dots & \widetilde{\beta}_{1k}\mathbf{J}_{n_1\times n_k'} \\ \vdots & \ddots & \vdots \\ \widetilde{\beta}_{k1}\mathbf{J}_{n_k\times n_1'} & \dots & \widetilde{\alpha}_{kk}\mathbf{J}_{n_k\times n_k'} \end{bmatrix}, \tag{5.3}$$

where

$$\widetilde{\alpha}_{ii} := \sum_{m\in\mathcal{M}} \left\{ \binom{n_i + n_i' - 2}{m-2} \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\sum_{l=1}^k (n_l + n_l') - 2}{m-2} \frac{b_m}{\binom{n}{m-1}} \right\}, \tag{5.4a}$$

$$\widetilde{\beta}_{ij} := \sum_{m\in\mathcal{M}} \binom{\sum_{l=1}^k (n_l + n_l') - 2}{m-2} \frac{b_m}{\binom{n}{m-1}}, \quad i \neq j,\ i,j \in [k]\,. \tag{5.4b}$$

The expectation of the *perfect* splitting, denoted by $\overline{\mathbf{B}}_1$, can be written as

$$\overline{\mathbf{B}}_1 := \begin{bmatrix} \overline{\alpha}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \dots & \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} \\ \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \overline{\alpha}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \dots & \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \overline{\beta}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} & \dots & \overline{\alpha}\mathbf{J}_{\frac{n}{2k}\times\frac{n}{4k}} \end{bmatrix}, \tag{5.5}$$

where

$$\overline{\alpha} := \sum_{m\in\mathcal{M}} \left\{ \binom{\frac{3n}{4k} - 2}{m-2} \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\frac{3n}{4} - 2}{m-2} \frac{b_m}{\binom{n}{m-1}} \right\}, \quad \overline{\beta} := \sum_{m\in\mathcal{M}} \binom{\frac{3n}{4} - 2}{m-2} \frac{b_m}{\binom{n}{m-1}}\,. \tag{5.6}$$

The matrices $\widetilde{\mathbf{B}}_2, \overline{\mathbf{B}}_2$ can be defined similarly, since the dimensions of $|Y_2 \cap V_i|$ are also determined by n_i and n_i'. Obviously, $\overline{\mathbf{B}}_2 = \overline{\mathbf{B}}_1$ since $\mathbb{E}n_i' = \mathbb{E}(n/k - n_i - n_i') = n/(4k)$ for all $i \in [k]$.

Fixing Dimensions. The dimensions of $\widetilde{\mathbf{B}}_1$ and $\widetilde{\mathbf{B}}_2$, as well as blocks they consist of, are not deterministic—since n_i and n_i', defined in Equation (5.1) and Equation (5.2) respectively, are sums of independent random variables. As such we cannot directly compare them. In order to overcome this difficulty, we embed $\mathbf{B}_1$ and $\mathbf{B}_2$ into the following $n \times n$ matrices:

$$\mathbf{A}_1 := \begin{bmatrix} \mathbf{0}_{|Z|\times|Z|} & \mathbf{B}_1 & \mathbf{0}_{|Z|\times|Y_2|} \\ \mathbf{0}_{|Y|\times|Z|} & \mathbf{0}_{|Y|\times|Y_1|} & \mathbf{0}_{|Y|\times|Y_2|} \end{bmatrix}, \quad \mathbf{A}_2 := \begin{bmatrix} \mathbf{0}_{|Z|\times|Z|} & \mathbf{0}_{|Z|\times|Y_1|} & \mathbf{B}_2 \\ \mathbf{0}_{|Y|\times|Z|} & \mathbf{0}_{|Y|\times|Y_1|} & \mathbf{0}_{|Y|\times|Y_2|} \end{bmatrix}. \tag{5.7}$$

Note that $\mathbf{A}_1$ and $\mathbf{A}_2$ have the same size. Also by definition, the entries in $\mathbf{A}_1$ are independent of the entries in $\mathbf{A}_2$. If we take expectation over $\{\mathcal{T}^{(m)}\}_{m\in\mathcal{M}}$ conditioning on $\{n_i\}_{i=1}^k$ and $\{n_i'\}_{i=1}^k$, then we obtain the expectation matrices of the *imperfect* splitting, denoted by $\widetilde{\mathbf{A}}_1$(resp. $\widetilde{\mathbf{A}}_2$), written as

$$\widetilde{\mathbf{A}}_1 := \begin{bmatrix} \mathbf{0}_{|Z|\times|Z|} & \widetilde{\mathbf{B}}_1 & \mathbf{0}_{|Z|\times|Y_2|} \\ \mathbf{0}_{|Y|\times|Z|} & \mathbf{0}_{|Y|\times|Y_1|} & \mathbf{0}_{|Y|\times|Y_2|} \end{bmatrix}, \quad \widetilde{\mathbf{A}}_2 := \begin{bmatrix} \mathbf{0}_{|Z|\times|Z|} & \mathbf{0}_{|Z|\times|Y_1|} & \widetilde{\mathbf{B}}_2 \\ \mathbf{0}_{|Y|\times|Z|} & \mathbf{0}_{|Y|\times|Y_1|} & \mathbf{0}_{|Y|\times|Y_2|} \end{bmatrix}. \tag{5.8}$$

The expectation matrix of the *perfect* splitting, denoted by $\overline{\mathbf{A}}_1$(resp. $\overline{\mathbf{A}}_2$), can be written as

$$\overline{\mathbf{A}}_1 := \begin{bmatrix} \mathbf{0}_{\frac{n}{2}\times\frac{n}{2}} & \overline{\mathbf{B}}_1 & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} \\ \mathbf{0}_{\frac{n}{2}\times\frac{n}{2}} & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} \end{bmatrix}, \quad \overline{\mathbf{A}}_2 := \begin{bmatrix} \mathbf{0}_{\frac{n}{2}\times\frac{n}{2}} & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} & \overline{\mathbf{B}}_2 \\ \mathbf{0}_{\frac{n}{2}\times\frac{n}{2}} & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} & \mathbf{0}_{\frac{n}{2}\times\frac{n}{4}} \end{bmatrix}. \tag{5.9}$$

Obviously, $\widetilde{\mathbf{A}}_i$ and $\widetilde{\mathbf{B}}_i$(resp. $\overline{\mathbf{A}}_i$ and $\overline{\mathbf{B}}_i$) have the same non-zero singular values for $i = 1, 2$. In the remaining of this section, we will deal with $\widetilde{\mathbf{A}}_i$ and $\overline{\mathbf{A}}_i$ instead of $\widetilde{\mathbf{B}}_i$ and $\overline{\mathbf{B}}_i$ for $i = 1, 2$.

5.1. Spectral Partition: Proof of Lemma 4.1.

5.1.1. *Proof Outline.* Recall that $\mathbf{A}_1$ is defined as the adjacency matrix of the induced sub-hypergraph $H[Y_1 \cup Z]$ in Section 5. Consequently, the index set should contain information only from $H[Y_1 \cup Z]$. Define the index sets

$$\mathcal{I} = \{i \in [n] : \mathrm{row}(i) \leq 20\mathcal{M}_{\max} d\}, \quad \mathcal{I}_1 = \left\{ i \in [n] : \mathrm{row}(i)\Big|_{Y_1 \cup Z} \leq 20\mathcal{M}_{\max} d \right\},$$

where $d = \sum_{m \in \mathcal{M}} (m-1) a_m$, and $\mathrm{row}(i)\big|_{Y_1 \cup Z}$ is the row sum of i on $H[Y_1 \cup Z]$. We say $\mathrm{row}(i)\big|_{Y_1 \cup Z} = 0$ if $i \notin Y_1 \cup Z$, and for vertex $i \in Y_1 \cup Z$,

$$\mathrm{row}(i)\Big|_{Y_1 \cup Z} := \sum_{j=1}^{n} \sum_{m \in \mathcal{M}} \sum_{\substack{e \in E_m[Y_1 \cup Z] \\ \{i,j\} \subset e}} \boldsymbol{\mathcal{T}}_e^{(m)} = \sum_{m \in \mathcal{M}} (m-1) \sum_{\substack{e \in E_m[Y_1 \cup Z] \\ \{i,j\} \subset e}} \boldsymbol{\mathcal{T}}_e^{(m)}.$$

As a result, the matrix $(\mathbf{A}_1)_{\mathcal{I}_1}$ is obtained by restricting $\mathbf{A}_1$ on index set $\mathcal{I}_1$. The next 4 steps guarantee that Algorithm 4.2 outputs a ν-correct partition.

(i) Find the singular subspace $\mathbf{U}$, which is spanned by the first k left singular vectors of $(\mathbf{A}_1)_{\mathcal{I}_1}$.
(ii) Randomly pick $s = 2k \log^2 n$ vertices from Y_2 and denote the corresponding columns in $\mathbf{A}_2$ by $\boldsymbol{a}_{i_1}, \dots, \boldsymbol{a}_{i_s}$. Project each vector $\boldsymbol{a}_i - \overline{\boldsymbol{a}}$ onto the singular subspace $\mathbf{U}$, with $\overline{\boldsymbol{a}} \in \mathbb{R}^n$ defined by $\overline{\boldsymbol{a}}(j) = \mathbb{1}_{j \in Z} \cdot (\overline{\alpha} + \overline{\beta})/2$, where $\overline{\alpha}$, $\overline{\beta}$ were defined in Equation (5.6).
(iii) For each projected vector $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$, identify the top $n/(2k)$ coordinates in value and place the corresponding vertices into a set U_i'. Discard half of the obtained s subsets, those with the lowest *blue* edge densities.
(iv) Sort the remaining sets according to blue hyperedge density and identify k distinct subsets $U_1', \cdots, U_k'$ such that $|U_i' \cap U_j'| < \lceil (1-\nu) n/k \rceil$ if $i \neq j$.

Based on the 4 steps above in Algorithm 4.2, the proof of Lemma 4.1 is structured in 4 parts.

(i) Let $\widetilde{\mathbf{U}}$ denote the subspace spanned by the first k left singular vectors of $\widetilde{\mathbf{A}}_1$, as defined in Equation (5.8). Section 5.1.2 shows that the subspace angle between $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ is of order $o(1)$.
(ii) The vector $\widetilde{\boldsymbol{\delta}}_i$, defined in Equation (5.14), reflects the underlying true partition $Z \cap V_{k(i)}$ for each $i \in [s]$, where $k(i)$ denotes the membership of vertex i. Section 5.1.3 shows that $\widetilde{\boldsymbol{\delta}}_i$ can be approximately recovered by the projected vector $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$, since the projection error $\|P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}}) - \widetilde{\boldsymbol{\delta}}_i\|_2$ is of order $O(1/\sqrt{n})$.
(iii) Section 5.1.4 indicates that the coincidence ratio between the remaining sets and the true partition is at least ν, after discarding half of the sets with lowest blue hyperedge densities.
(iv) Lemma 5.11 proves that we can find k distinct subsets U_i' within $k \log^2 n$ trials *w.h.p.*

5.1.2. *Bounding the angle between* $\mathbf{U}$ *and* $\widetilde{\mathbf{U}}$. Recall that the angle between subspaces $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ is defined as $\sin \angle(\mathbf{U}, \widetilde{\mathbf{U}}) := \|P_{\mathbf{U}} - P_{\widetilde{\mathbf{U}}}\|$. A natural idea is to apply Wedin's $\sin\Theta$ Theorem (Lemma D.7).

Lemma 5.3 indicates that the difference of $\sigma_i(\widetilde{\mathbf{A}}_1)$ and $\sigma_i(\overline{\mathbf{A}}_1)$ is relatively small, compared to $\sigma_i(\overline{\mathbf{A}}_1)$.

Lemma 5.3. *Let $\sigma_i(\overline{\mathbf{A}}_1)$(resp. $\sigma_i(\widetilde{\mathbf{A}}_1)$) denote the singular values of $\overline{\mathbf{A}}_1$ (resp. $\widetilde{\mathbf{A}}_1$) for all $i \in [k]$, where the matrices $\overline{\mathbf{A}}_1$ and $\widetilde{\mathbf{A}}_1$ are defined in Equation (5.9) and Equation (5.8) respectively. Then*

$$\sigma_1(\overline{\mathbf{A}}_1) = \frac{n\left[\overline{\alpha} + (k-1)\overline{\beta}\right]}{2\sqrt{2}k} = \frac{n}{2\sqrt{2}k} \sum_{m \in \mathcal{M}} \left[\binom{\frac{3n}{4k} - 2}{m-2} \frac{a_m - b_m}{\binom{n}{m-1}} + k \binom{\frac{3n}{4} - 2}{m-2} \frac{b_m}{\binom{n}{m-1}} \right],$$

$$\sigma_i(\overline{\mathbf{A}}_1) = \frac{n(\overline{\alpha} - \overline{\beta})}{2\sqrt{2}k} = \frac{n}{2\sqrt{2}k} \sum_{m \in \mathcal{M}} \binom{\frac{3n}{4k} - 2}{m-2} \frac{a_m - b_m}{\binom{n}{m-1}}, 2 \leq i \leq k,$$

$$\sigma_i(\overline{\mathbf{A}}_1) = 0, k+1 \leq i \leq n.$$

with $\overline{\alpha}$, $\overline{\beta}$ defined in Equation (5.6). Moreover, with probability at least $1 - 2k\exp(-k\log^2(n))$,

$$\frac{|\sigma_i(\overline{\mathbf{A}}_1) - \sigma_i(\widetilde{\mathbf{A}}_1)|}{\sigma_i(\overline{\mathbf{A}}_1)} = O\left(n^{-1/4} \log^{1/2}(n)\right).$$

Therefore, with Lemma 5.3, we can write $\sigma_i(\widetilde{\mathbf{A}}_1) = \sigma_i(\overline{\mathbf{A}}_1)(1+o(1))$. Define $\mathbf{E}_1 := \mathbf{A}_1 - \widetilde{\mathbf{A}}_1$ and its restriction on $\mathcal{I}_1$ as

$$(\mathbf{E}_1)_{\mathcal{I}_1} := (\mathbf{A}_1 - \widetilde{\mathbf{A}}_1)_{\mathcal{I}_1} = (\mathbf{A}_1)_{\mathcal{I}_1} - (\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1}\,, \tag{5.10}$$

as well as $\mathbf{\Delta}_1 := (\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1$. Then $(\mathbf{A}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1$ is decomposed as

$$(\mathbf{A}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1 = [(\mathbf{A}_1)_{\mathcal{I}_1} - (\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1}] + [(\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1] = (\mathbf{E}_1)_{\mathcal{I}_1} + \mathbf{\Delta}_1\,.$$

Lemma 5.4 shows that the number of high degree vertices is relatively small with high probability.

Lemma 5.4. *Let $d = \sum_{m\in\mathcal{M}}(m-1)a_m$, where $\mathcal{M}$ is obtained from Algorithm 4.1. There exists a constant C_1 such that if $d \geq C_1$, then with probability at least $1-\exp\left(-d^{-2}n/\mathcal{M}_{\max}\right)$, no more than $d^{-3}n$ vertices have row sums greater than $20\mathcal{M}_{\max}d$.*

Consequently, Corollary 5.5 indicates $\|\mathbf{\Delta}_1\| \leq \sqrt{d}$ with high probability.

Corollary 5.5. *Assume $d \geq \max\{C_1, \sqrt{2}\}$, where C_1 is the constant in Lemma 5.4, then $\|\mathbf{\Delta}_1\| \leq \sqrt{d}$ w.h.p.*

Proof. Note that $n - |\mathcal{I}| \leq d^{-3}n$ and $\mathcal{I} \subset \mathcal{I}_1$, then $n - |\mathcal{I}_1| \leq d^{-3}n$. From Lemma 5.4, there are at most $d^{-3}n$ vertices with row sum greater than $20\mathcal{M}_{\max}d$ in the adjacency matrix $\mathbf{A}_1$, then the matrix $\mathbf{\Delta}_1 = (\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1$ has at most $2d^{-3}n^2$ non-zero entries. As defined in Equation (5.8), every entry in $\widetilde{\mathbf{A}}_1$ is bounded by α, then,

$$\begin{aligned}
\|\mathbf{\Delta}_1\| &\leq \|\mathbf{\Delta}_1\|_{\mathrm{F}} = \|(\widetilde{\mathbf{A}}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1\|_{\mathrm{F}} \\
&\leq \sqrt{2d^{-3}n^2}\,\alpha = \sqrt{2d^{-3}}n\sum_{m\in\mathcal{M}}\left[\binom{\frac{n}{k}-2}{m-2}\frac{a_m-b_m}{\binom{n}{m-1}} + \binom{n-2}{m-2}\frac{b_m}{\binom{n}{m-1}}\right] \\
&\leq \sqrt{2d^{-3}}\sum_{m\in\mathcal{M}}(m-1)a_m \leq \sqrt{2d^{-1}} \leq \sqrt{d}\,.
\end{aligned}$$

∎

Moreover, taking $\tau = 20\mathcal{M}_{\max}, K = 3$ in Theorem 3.3, with probability at least $1-n^{-2}$, we have

$$\|(\mathbf{E}_1)_{\mathcal{I}_1}\| \leq C_3\sqrt{d} \tag{5.12}$$

where C_3 is a constant depending only on $\mathcal{M}_{\max}$. Together with upper bounds for $\|(\mathbf{E}_1)_{\mathcal{I}_1}\|$ and $\|\mathbf{\Delta}_1\|$, Lemma 5.6 shows that the angle between $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ is relatively small with high probability.

Lemma 5.6. *For any $c \in (0,1)$, there exists some constant C_2 such that, if $\sum_{m\in\mathcal{M}}(m-1)(a_m - b_m) \geq C_2 k^{\mathcal{M}_{\max}-1}\sqrt{d}$, then $\sin\angle(\mathbf{U}, \widetilde{\mathbf{U}}) \leq c$ with probability $1-n^{-2}$. Here $\angle(\mathbf{U}, \widetilde{\mathbf{U}})$ is the angle between $\mathbf{U}$ and $\widetilde{\mathbf{U}}$.*

Proof. From Equation (5.12) and Corollary 5.5, with high probability,

$$\|(\mathbf{A}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1\| \leq \|(\mathbf{E}_1)_{\mathcal{I}_1}\| + \|\mathbf{\Delta}_1\| \leq (C_3+1)\sqrt{d}.$$

Since $\sigma_{k+1}(\widetilde{\mathbf{A}}_1) = 0$, using Lemma 5.3 to approximate $\sigma_k(\widetilde{\mathbf{A}}_1)$, we obtain

$$\begin{aligned}
&\sigma_k(\widetilde{\mathbf{A}}_1) - \sigma_{k+1}(\widetilde{\mathbf{A}}_1) = \sigma_k(\widetilde{\mathbf{A}}_1) = (1+o(1))\sigma_k(\overline{\mathbf{A}}_1) \geq \frac{1}{2}\sigma_k(\overline{\mathbf{A}}_1) \\
\geq{}& \frac{n}{4\sqrt{2}k}\sum_{m\in\mathcal{M}}\binom{\frac{3n}{4k}-2}{m-2}\frac{a_m-b_m}{\binom{n}{m-1}} \geq \frac{1}{8k}\sum_{m\in\mathcal{M}}\left(\frac{3}{4k}\right)^{m-2}(m-1)(a_m-b_m) \\
\geq{}& \frac{1}{8k}\left(\frac{1}{2k}\right)^{\mathcal{M}_{\max}-2}\sum_{m\in\mathcal{M}}(m-1)(a_m-b_m) \geq \frac{C_2\sqrt{d}}{2^{\mathcal{M}_{\max}+1}}\,.
\end{aligned}$$

Then for any $c \in (0,1)$, we can find $C_2 = [2^{\mathcal{M}_{\max}+2}(C_3+1)/c]$ such that $\|(\mathbf{A}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1\| \leq (1-1/\sqrt{2})\sigma_k(\widetilde{\mathbf{A}}_1)$, and by Wedin's Theorem (Lemma D.7), the angle between $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ is bounded by

$$\|P_{\mathbf{U}} - P_{\widetilde{\mathbf{U}}}\| \leq \frac{\sqrt{2}\|(\mathbf{A}_1)_{\mathcal{I}_1} - \widetilde{\mathbf{A}}_1\|}{\sigma_k(\widetilde{\mathbf{A}}_1)} \leq \frac{\sqrt{2}(C_3+1)}{C_2/2^{\mathcal{M}_{\max}+1}} = \frac{\sqrt{2}}{2}c < c\,.$$

∎

5.1.3. *Bound the projection error.* Randomly pick $s = 2k\log^2 n$ vertices from Y_2. Let $\boldsymbol{a}_{i_1}, \dots, \boldsymbol{a}_{i_s}, \widetilde{\boldsymbol{a}}_{i_1}, \dots, \widetilde{\boldsymbol{a}}_{i_s}$, $\overline{\boldsymbol{a}}_{i_1}, \dots, \overline{\boldsymbol{a}}_{i_s}$ and $\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_s}$ be the corresponding columns of $\mathbf{A}_2$, $\widetilde{\mathbf{A}}_2$, $\overline{\mathbf{A}}_2$ and $\mathbf{E}_2 := \mathbf{A}_2 - \widetilde{\mathbf{A}}_2$ respectively, where $\mathbf{A}_2$, $\widetilde{\mathbf{A}}_2$ and $\overline{\mathbf{A}}_2$ were defined in Equation (5.7), Equation (5.8) and Equation (5.9). Let $k(i)$ denote the membership of vertex i, hence $i \in V_{k(i)} \cap Y_2$ for any $i \in \{i_1, \cdots, i_s\}$. Define the corresponding vector $\widetilde{\boldsymbol{\delta}}_i \in \mathbb{R}^n$ with the entries of $\widetilde{\boldsymbol{a}}_i$ and $\widetilde{\boldsymbol{\delta}}_i$ given by

$$\widetilde{\boldsymbol{a}}_i(j) = \begin{cases} \widetilde{\alpha}_{ii}, & \text{if } j \in Z \cap V_{k(i)} \\ \widetilde{\beta}_{ij}, & \text{if } j \in Z \setminus V_{k(i)} \\ 0\,, & \text{if } j \in Y \end{cases}, \quad \widetilde{\boldsymbol{\delta}}_i(j) = \begin{cases} (\widetilde{\alpha}_{ii} - \widetilde{\beta}_{ij})/2 > 0, & \text{if } j \in Z \cap V_{k(i)} \\ (\widetilde{\beta}_{ij} - \widetilde{\alpha}_{ii})/2 < 0, & \text{if } j \in Z \setminus V_{k(i)} \\ 0, & \text{if } j \in Y \end{cases}. \tag{5.14}$$

where $\widetilde{\alpha}_{ii}$, $\widetilde{\beta}_{ij}$ were defined in Equation (5.4a), Equation (5.4b) respectively. Consequently, we can identify the block $Z \cap V_{k(i)}$ by recovering $\widetilde{\boldsymbol{\delta}}_i$ according to the sign difference between elements in $Z \cap V_{k(i)}$ and $Z \setminus V_{k(i)}$. However, it is hard to handle with $\widetilde{\boldsymbol{\delta}}_i$ due to the randomness of $\widetilde{\alpha}_{ii}$, $\widetilde{\beta}_{ij}$. A good approximation of $\widetilde{\boldsymbol{\delta}}_i$ was given by $\overline{\boldsymbol{\delta}}_i := \overline{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}$, where $\overline{\boldsymbol{a}}(j) := \mathbb{1}_{j\in Z} \cdot (\overline{\alpha} + \overline{\beta})/2$ and entries of $\overline{\boldsymbol{\delta}}_i$ given by

$$\overline{\boldsymbol{a}}_i(j) = \begin{cases} \overline{\alpha}, & \text{if } j \in Z \cap V_{k(i)} \\ \overline{\beta}, & \text{if } j \in Z \setminus V_{k(i)} \\ 0, & \text{if } j \in Y \end{cases}, \quad \overline{\boldsymbol{\delta}}_i(j) = \begin{cases} (\overline{\alpha} - \overline{\beta})/2 > 0, & \text{if } j \in Z \cap V_{k(i)} \\ (\overline{\beta} - \overline{\alpha})/2 < 0, & \text{if } j \in Z \setminus V_{k(i)} \\ 0, & \text{if } j \in Y \end{cases}.$$

By construction, $\overline{\boldsymbol{\delta}}_i$ identifies $Z \cap V_{k(i)}$ in the case of *perfect splitting* for any $i \in \{i_1, \cdots, i_s\}$. However, the fluctuations of n_i and n_i' are up to $O(\sqrt{n}\log n)$ *w.h.p.*, as proved in Lemma 5.3. The error in this approximation can be then corrected in Algorithm 4.3. Note that

$$\boldsymbol{a}_i - \overline{\boldsymbol{a}} = (\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i) + (\widetilde{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}_i) + (\overline{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}) = \boldsymbol{e}_i + (\widetilde{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}_i) + \overline{\boldsymbol{\delta}}_i\,, \quad \forall i \in V_{k(i)} \cap Y_2 \cap \{i_1, \cdots, i_s\}\,.$$

Since we do not have access to $\widetilde{\boldsymbol{\delta}}_i$ and $\overline{\boldsymbol{\delta}}_i$, we use $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$ as a good approximation, and

$$P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}}) - \widetilde{\boldsymbol{\delta}}_i = P_{\mathbf{U}}\boldsymbol{e}_i + P_{\mathbf{U}}(\widetilde{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}_i) + P_{\mathbf{U}}\overline{\boldsymbol{\delta}}_i - \widetilde{\boldsymbol{\delta}}_i\,.$$

By definitions of $\overline{\alpha}$ and $\overline{\beta}$ in Equation (5.6), we have

$$\|P_{\mathbf{U}}(\widetilde{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}_i)\|_2 \le \|\widetilde{\boldsymbol{a}}_i - \overline{\boldsymbol{a}}_i\|_2 \le O\big[k \cdot \sqrt{n}\log n \cdot (\overline{\alpha} - \overline{\beta})^2\big]^{1/2} = O[n^{-3/4} \cdot \log^{1/2}(n)]$$
$$\|P_{\mathbf{U}}\overline{\boldsymbol{\delta}}_i - \widetilde{\boldsymbol{\delta}}_i\|_2 \le \|\overline{\boldsymbol{\delta}}_i\|_2 + \|\widetilde{\boldsymbol{\delta}}_i\|_2 = 2\left[\left(\sum_{i=1}^{k} n_i\right) \cdot (\overline{\alpha} - \overline{\beta})^2\right]^{1/2} = O(n^{-1/2})\,.$$

Lemma 5.7 shows that with high probability, at least half of the vectors in $\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_s}$ has small l_2-norm of order $O(n^{-1/2})$, after projection onto $\mathbf{U}$.

Lemma 5.7. *With probability* $1 - O(n^{-k\log n})$*, at least* $s/2$ *of the vectors* $\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_s}$ *satisfy*

$$\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2 \le 2\sqrt{k(\alpha + d\mathcal{M}_{\max}/n)} = O(n^{-1/2})\,, \quad \forall i_1, \cdots, i_s \in [n]\,. \tag{5.16}$$

To avoid introducing extra notations, let $\boldsymbol{a}_{i_1}, \dots, \boldsymbol{a}_{i_{s_1}}$ denote those vectors satisfying

$$\|P_{\mathbf{U}}\boldsymbol{e}_{i_j}\|_2 \le 2\sqrt{k(\alpha + d\mathcal{M}^2_{\max}/n)}$$

for $j \in [s_1]$, hence referred as *good* vectors. Lemma 5.7 indicates that the number of good vectors satisfies $s_1 \ge s/2 = k\log^2 n$. Together with Lemma 5.6 and discussion in Section 5.1.3, Lemma 5.8 proves that *good* vectors have vanishing projection errors.

Lemma 5.8 (Projection error)**.** *If conditions in Lemma 5.6 are satisfied, then for all good vectors* $\boldsymbol{a}_{i_j}$ *with* $j \in [s_1]$*,* $\|P_{\mathbf{U}}(\boldsymbol{a}_{i_j} - \overline{\boldsymbol{a}}) - \widetilde{\boldsymbol{\delta}}_{i_j}\|_2 = O(n^{-1/2})$ *holds with probability* $1 - O(n^{-k})$*.*

5.1.4. *Accuracy.* For each projected vector $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$, let U_i' denote the set of its top $n/(2k)$ coordinates in value, where $i \in \{i_1, \cdots, i_s\}$ and $s = 2k\log^2 n$. In Algorithm 4.2, we discard half of the obtained subsets U_i' with lowest blue hyperedge densities. Lemma 5.9 shows that for every good index i_j with vanishing projection errors as referred in Lemma 5.8, the vertex set U_{i_j}' contains at least ν fraction of the vertices in $V_{k(i_j)} \cap Z$, where $k(i_j)$ denotes the membership of vertex i_j.

Lemma 5.9. *Suppose that we are given a set $X \subset Z$ with size $|X| = n/(2k)$. Define*

$$\mu_1 := \frac{1}{2}\sum_{m\in\mathcal{M}} m(m-1)\left\{\left[\binom{\frac{\nu n}{2k}}{m} + \binom{\frac{(1-\nu)n}{2k}}{m}\right]\frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m}\frac{b_m}{\binom{n}{m-1}}\right\},$$

$$\mu_2 := \frac{1}{2}\sum_{m\in\mathcal{M}} m(m-1)\left\{\left[\binom{\frac{(1+\nu)n}{4k}}{m} + (k-1)\binom{\frac{(1-\nu)n}{4k(k-1)}}{m}\right]\frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m}\frac{b_m}{\binom{n}{m-1}}\right\},$$

and $\mu_{\mathrm{T}} := (\mu_1 + \mu_2)/2 \in [\mu_1, \mu_2]$. There is a constant $c > 0$ depending on k, a_m, ν such that for sufficiently large n,

- *(i) If $|X \cap V_i| \le \nu|X|$ for all $i \in [k]$, then with probability at least $1 - e^{-cn}$, the number of blue hyperedges in the hypergraph induced by X is at most μ_{T}.*
- *(ii) Conversely, if $|X \cap V_i| \ge (1/2)\cdot(1+\nu)|X|$ for some $i \in \{1, \ldots, k\}$, then with probability at least $1 - e^{-cn}$, the number of blue hyperedges in the hypergraph induced by X is at least μ_{T}.*

Remark 5.10. *When taking $M = 2$, Lemma 5.9 reduces to [16, Lemma 31]. The original proof of [16, Lemma 31] contains an error and we are able to fix it in our proof.*

Let U_i' denote the set of $n/(2k)$ largest coordinates of the projected vector $P_{\mathbf{U}}(\boldsymbol{a}_i - \overline{\boldsymbol{a}})$ with $i \in V_{k(i)} \cap Y_2$. According to Lemma 5.9, if the blue hyperedge density inside U_i' is at least μ_{T}, then it guarantees that $|U_i' \cap V_{k(i)}| \ge \nu|U_i'|$. Therefore, it is enough to consider half of the sets with highest blue edge densities in Algorithm 4.2.

Lemma 5.11. *Through random sampling without replacement in Step 6 of Algorithm 4.2, we can find at least k vertices in Y_2 among $k\log^2 n$ samples which belongs to distinct V_i, $i \in [k]$ with probability $1 - n^{-\Omega(\log n)}$.*

5.2. **Local Correction: Proof of Lemma 4.2.** For notional convenience, let $U_i := Z \cap V_i$ denote the intersection of Z and true partition V_i for all $i \in [k]$. In Algorithm 1.1, we first color the hyperedges by red and blue with equal probability. By running Algorithm 4.2 on the red hypergraph, we obtain a ν-correct partition $U_1', \ldots, U_k'$, i.e.,

$$|U_i \setminus U_i'| \le (1-\nu)\cdot|U_i'| = (1-\nu)\cdot\frac{n}{2k}, \quad \forall i \in [k]. \tag{5.18}$$

In the rest of this subsection, we condition on the event (5.18). Consider a hyperedge $e = \{i_1, \cdots, i_m\}$ in the underlying m-uniform hypergraph. If vertices $i_1, \cdots, i_m$ are from the same block, then e is a red hyperedge with probability $a_m/2\binom{n}{m-1}$; if vertices $i_1, \cdots, i_m$ are not from the same block, then e is a red hyperedge with probability $b_m/2\binom{n}{m-1}$. The presence of those two types of hyperedges can be denoted by

$$T_e^{(a_m)} \sim \mathrm{Bernoulli}\left(\frac{a_m}{2\binom{n}{m-1}}\right), \quad T_e^{(b_m)} \sim \mathrm{Bernoulli}\left(\frac{b_m}{2\binom{n}{m-1}}\right),$$

respectively. For any finite set S, let $[S]^l$ denote the family of l-subsets of S, i.e., $[S]^l = \{Z | Z \subseteq S, |Z| = l\}$. Consider a vertex $u \in U_1 := Z \cap V_1$. The weighted number of red hyperedges, which contains $u \in U_1$ with the remaining vertices in U_j', can be written as

$$S_{1j}'(u) := \sum_{m\in\mathcal{M}} (m-1)\cdot\left\{\sum_{e\in\mathcal{E}_{1,j}^{(a_m)}} T_e^{(a_m)} + \sum_{e\in\mathcal{E}_{1,j}^{(b_m)}} T_e^{(b_m)}\right\}, \quad u \in U_1, \tag{5.19}$$

where $\mathcal{E}_{1,j}^{(a_m)} := E_m([U_1]^1, [U_1 \cap U_j']^{m-1})$ denotes the set of m-hyperedges with one vertex from $[U_1]^1$ and the other $m-1$ from $[U_1 \cap U_j']^{m-1}$, while $\mathcal{E}_{1,j}^{(b_m)} := E_m\Big([U_1]^1,\ [U_j']^{m-1} \setminus [U_1 \cap U_j']^{m-1}\Big)$ denotes the set of

m-hyperedges with one vertex in $[U_1]^1$ while the remaining $m-1$ vertices in $[U'_j]^{m-1} \setminus [U_1 \cap U'_j]^{m-1}$(not all m vertices are from V_1) with their cardinalities

$$|\mathcal{E}_{1,j}^{(a_m)}| = \binom{|U_1 \cap U'_j|}{m-1}, \quad |\mathcal{E}_{1,j}^{(b_m)}| = \left[\binom{|U'_j|}{m-1} - \binom{|U_1 \cap U'_j|}{m-1}\right].$$

We multiply $(m-1)$ in Equation (5.19) as weight since the rest $m-1$ vertices are all located in U'_j, which can be regarded as u's neighbors in U'_j. According to the fact $|U'_j \cap U_j| \geq (\nu n/2k)$ in Equation (5.18) and $|U'_j| = n/(2k)$ for $j \in [k]$,

$$|\mathcal{E}_{1,1}^{(a_m)}| \geq \binom{\frac{\nu n}{2k}}{m-1}, \quad |\mathcal{E}_{1,j}^{(a_m)}| \leq \binom{\frac{(1-\nu)n}{2k}}{m-1}, \; j \neq 1.$$

To simplify the calculation, we take the lower and upper bound of $|\mathcal{E}_{1,1}^{(a_m)}|$ and $|\mathcal{E}_{1,j}^{(a_m)}|\,(j \neq 1)$ respectively. By taking expectation with respect to $T_e^{(a_m)}$ and $T_e^{(b_m)}$, then for any $u \in U_1$, we have

$$\mathbb{E}S'_{11}(u) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{\nu n}{2k}}{m-1} \frac{a_m - b_m}{2\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m-1} \frac{b_m}{2\binom{n}{m-1}}\right],$$

$$\mathbb{E}S'_{1j}(u) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{(1-\nu)n}{2k}}{m-1} \frac{a_m - b_m}{2\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m-1} \frac{b_m}{2\binom{n}{m-1}}\right], \quad j \neq 1.$$

By assumptions in Theorem 1.4, $\mathbb{E}S'_{11}(u) - \mathbb{E}S'_{1j}(u) = \Omega(1)$. Define

$$\mu_{\mathrm{C}} := \frac{1}{2} \sum_{m \in \mathcal{M}} (m-1) \cdot \left\{\left[\binom{\frac{\nu n}{2k}}{m-1} + \binom{\frac{(1-\nu)n}{2k}}{m-1}\right] \frac{a_m - b_m}{2\binom{n}{m-1}} + 2 \cdot \binom{\frac{n}{2k}}{m-1} \frac{b_m}{2\binom{n}{m-1}}\right\}. \tag{5.21}$$

In Algorithm 4.3, vertex u is assigned to $\widehat{U}_i$ if it has the maximal number of neighbors in U'_i. If $u \in U_1$ is mislabeled, then one of the following events must happen:

- $S'_{11}(u) \leq \mu_{\mathrm{C}}$, meaning that u was mislabeled by Algorithm 4.3.
- $S'_{1j}(u) \geq \mu_{\mathrm{C}}$, for some $j \neq 1$, meaning that u survived Algorithm 4.3 without being corrected.

Lemma 5.12 shows that the probabilities of those two events are exponentially small.

Lemma 5.12. *For sufficiently large n and any $u \in U_1 = Z \cap V_1$, we have*

$$\rho'_1 := \mathbb{P}\left(S'_{11}(u) \leq \mu_{\mathrm{C}}\right) \leq \rho, \quad \rho'_j := \mathbb{P}\left(S'_{1j}(u) \geq \mu_{\mathrm{C}}\right) \leq \rho, \; (j \neq 1), \tag{5.22}$$

where $\rho := \exp\left(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}}\right)$ with $\mathrm{SNR}_{\mathcal{M}}$ and $C_{\mathcal{M}}$ defined in Equation (1.3).

As a result, the probability of that either of those events happened is bounded by ρ. The number of mislabeled vertices in U_1 after Algorithm 4.3 is at most

$$R_1 = \sum_{t=1}^{|U_1|} \Gamma_t + \sum_{j=2}^{k} \sum_{t=1}^{|U_1 \cap U'_j|} \Lambda_t,$$

where Γ_t (resp. Λ_t) are i.i.d indicator random variables with mean ρ'_1 (resp. ρ'_j, $j \neq 1$). Then

$$\mathbb{E}R_1 \leq \frac{n}{2k}\rho'_1 + \sum_{j=2}^{k} \frac{(1-\nu)n}{2k}\rho'_j \leq \frac{n}{2k} \cdot k\rho = \frac{n\rho}{2}.$$

Let $t_1 := n\rho/2$, where ν is the correctness after Algorithm 4.2, then by Chernoff bound (Lemma D.1),

$$\mathbb{P}\left(R_1 \geq n\rho\right) = \mathbb{P}\left(R_1 - n\rho/2 \geq t_1\right) \leq \mathbb{P}\left(R_1 - \mathbb{E}R_1 \geq t_1\right) \leq e^{-ct_1} = O(e^{-n\rho}). \tag{5.23}$$

Then with probability $1 - O(e^{-n\rho})$, the fraction of mislabeled vertices in U_1 is smaller than $k\rho$, i.e., the correctness of U_1 is at least $\gamma_{\mathrm{C}} := \max\{\nu, 1 - k\rho\}$. Therefore, Algorithm 4.3 outputs a γ_{C}-correct partition $\widehat{U}_1, \cdots, \widehat{U}_k$ with probability $1 - O(e^{-n\rho})$.

Remark 5.13. *When $\mathrm{SNR}_{\mathcal{M}} = \Omega(\log n)$, we have $\rho = \Omega(n^{-1})$ and $e^{-n\rho} = \Omega(1)$ in Equation (5.23), which may not decrease to 0 as $n \to \infty$. As a result, we assume $\mathrm{SNR}_{\mathcal{M}} = o(\log n)$ in the statement of Corollary 1.7.*

5.3. **Merging: Proof of Lemma 4.3.** By running Algorithm 4.3 on the red graph, we obtain a γ_{C}-correct partition $\widehat{U}_1, \cdots, \widehat{U}_k$ where $\gamma_{\mathrm{C}} := \max\{\nu\,, 1-k\rho\} \geq \nu$, i.e.,

$$|U_j \cap \widehat{U}_j| \geq \nu \cdot |\widehat{U}_j| = \frac{\nu n}{2k}\,, \quad \forall j \in [k]\,. \tag{5.24}$$

In the rest of this subsection, we shall condition on this event and abbreviate $Y \cap V_l$ by $W_l := Y \cap V_l$. The failure probability of Algorithm 4.4 is estimated by presence of hyperedges between vertex sets Y and Z.

Consider a hyperedge $e = \{i_1, \cdots, i_m\}$ in the underlying m-uniform hypergraph. If vertices $i_1, \cdots, i_m$ are all from the same cluster V_l, then the probability that e is an existing blue edge conditioning on the event that e is not a red edge is

$$\psi_m := \mathbb{P}\left[e \text{ is a blue edge } \Big| e \text{ is not a red edge } \right] = \frac{\frac{a_m}{2\binom{n}{m-1}}}{1 - \frac{a_m}{2\binom{n}{m-1}}} = \frac{a_m}{2\binom{n}{m-1}}(1+o(1))\,, \tag{5.25}$$

and the presence of e can be represented by an indicator random variable $\zeta_e^{(a_m)} \sim \text{Bernoulli}\,(\psi_m)$. Similarly, if vertices $i_1, \cdots, i_m$ are not all from the same cluster V_l, the probability that e is an existing blue edge conditioning on the event that e is not a red edge is

$$\phi_m := \mathbb{P}\left[e \text{ is a blue edge } \Big| e \text{ is not a red edge } \right] = \frac{\frac{b_m}{2\binom{n}{m-1}}}{1 - \frac{b_m}{2\binom{n}{m-1}}} = \frac{b_m}{2\binom{n}{m-1}}(1+o(1))\,, \tag{5.26}$$

and the presence of this hyperedge can be represented by an indicator random variable $\xi_e^{(b_m)} \sim \text{Bernoulli}\,(\phi_m)$.

For any vertex $w \in W_l := Y \cap V_l$ with fixed $l \in [k]$, we want to compute the number of hyperedges containing w with all remaining vertices located in vertex set $\widehat{U}_j$ for some fixed $j \in [k]$. Following a similar argument given in Section 5.2, this number can be written as

$$\widehat{S}_{lj}(w) := \sum_{m \in \mathcal{M}} (m-1) \cdot \left\{ \sum_{e \in \widehat{\mathcal{E}}_{l,j}^{(a_m)}} \zeta_e^{(a_m)} + \sum_{e \in \widehat{\mathcal{E}}_{l,j}^{(b_m)}} \xi_e^{(b_m)} \right\}\,, \quad w \in W_l\,, \tag{5.27}$$

where $\widehat{\mathcal{E}}_{l,j}^{(a_m)} := E_m([W_l]^1, [U_l \cap \widehat{U}_j]^{m-1})$ denotes the set of m-hyperedges with 1 vertex from $[W_l]^1$ and the other $m-1$ vertices from $[U_l \cap \widehat{U}_j]^{m-1}$, while $\widehat{\mathcal{E}}_{l,j}^{(b_m)} := E_m([W_l]^1,\ [\widehat{U}_j]^{m-1} \setminus [U_l \cap \widehat{U}_j]^{m-1})$ denotes the set of m-hyperedges with 1 vertex in $[W_l]^1$ while the remaining $m-1$ vertices are in $[\widehat{U}_j]^{m-1} \setminus [U_l \cap \widehat{U}_j]^{m-1}$, with their cardinalities

$$|\widehat{\mathcal{E}}_{l,j}^{(a_m)}| = \binom{|U_l \cap \widehat{U}_j|}{m-1}\,, \quad |\widehat{\mathcal{E}}_{l,j}^{(b_m)}| = \left[\binom{|\widehat{U}_j|}{m-1} - \binom{|U_l \cap \widehat{U}_j|}{m-1}\right]\,.$$

Similarly, we multiply $(m-1)$ in Equation (5.27) as weight since the rest $m-1$ vertices can be regarded as u's neighbors in $\widehat{U}_j$. By accuracy of Algorithm 4.3 in Equation (5.24), $|\widehat{U}_j \cap U_j| \geq \nu n/(2k)$, then

$$|\widehat{\mathcal{E}}_{l,l}^{(a_m)}| \geq \binom{\frac{\nu n}{2k}}{m-1}\,, \quad |\widehat{\mathcal{E}}_{l,j}^{(a_m)}| \leq \binom{\frac{(1-\nu)n}{2k}}{m-1}\,,\ j \neq l\,.$$

Taking expectation with respect to $\zeta_e^{(a_m)}$ and $\xi_e^{(b_m)}$, for any $w \in W_l$, we have

$$\mathbb{E}\widehat{S}_{ll}(w) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{\nu n}{2k}}{m-1}(\psi_m - \phi_m) + \binom{\frac{n}{2k}}{m-1}\phi_m\right]\,,$$

$$\mathbb{E}\widehat{S}_{lj}(w) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{(1-\nu)n}{2k}}{m-1}(\psi_m - \phi_m) + \binom{\frac{n}{2k}}{m-1}\phi_m\right]\,,\ j \neq l\,.$$

By assumptions in Theorem 1.4, $\mathbb{E}\widehat{S}_{ll}(w) - \mathbb{E}\widehat{S}_{lj}(w) = \Omega(1)$. We define

$$\mu_{\mathrm{M}} := \frac{1}{2}\sum_{m \in \mathcal{M}} (m-1) \cdot \left\{\left[\binom{\frac{\nu n}{2k}}{m-1} + \binom{\frac{(1-\nu)n}{2k}}{m-1}\right](\psi_m - \phi_m) + 2\binom{\frac{n}{2k}}{m-1}\phi_m\right\}\,.$$

After Algorithm 4.4, if a vertex $w \in W_l$ is mislabelled, one of the following events must happen

- $\widehat{S}_{ll}(w) \leq \mu_{\rm M}$, which implies that u was mislabelled by Algorithm 4.4.
- $\widehat{S}_{lj}(w) \geq \mu_{\rm M}$ if $j \neq l$, which implies that u survived Algorithm 4.4 without being corrected.

By an argument similar to Lemma 5.12, we can prove that for any $w \in W_l$,

$$\hat{\rho}_l := \mathbb{P}(\widehat{S}_{ll}(w) \leq \mu_{\rm M}) \leq \rho\,, \quad \hat{\rho}_j := \mathbb{P}(\widehat{S}_{lj}(w) \geq \mu_{\rm M}) \leq \rho\,,\, (j \neq l),$$

where $\rho := \exp\left(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}}\right)$. The mis-classified probability for $w \in W_l$ is upper bounded by $\sum_{j=1}^{k} \hat{\rho}_j \leq k\rho$. The number of mislabelled vertices in W_l is at most $R_l = \sum_{t=1}^{|W_l|} \Gamma_t$, where Γ_t are i.i.d indicator random variables with mean $k\rho$ and $\mathbb{E}R_l \leq n/(2k) \cdot k\rho = n\rho/2$. Let $t_l := n\rho/2$, by Chernoff bound (Lemma D.1),

$$\mathbb{P}\left(R_l \geq n\rho\right) = \mathbb{P}\left(R_l - n\rho/2 \geq t_l\right) \leq \mathbb{P}\left(R_l - \mathbb{E}R_l \geq t_l\right) \leq e^{-ct_l} = O(e^{-n\rho})\,.$$

Hence with probability $1 - O(e^{-n\rho})$, the fraction of mislabeled vertices in W_l is smaller than $k\rho$, i.e., the correctness in W_l is at least $\gamma_{\rm M} := \max\{\nu,\, 1 - k\rho\}$.

5.4. **Proof of Theorem 1.4.** Now we are ready to prove our main result (Theorem 1.4). Note that the correctness of Algorithm 4.3 and Algorithm 4.4 are $\gamma_{\rm C}$ and $\gamma_{\rm M}$ respectively, then with probability at least $1 - o(1)$, the correctness γ of Algorithm 1.1 is $\gamma := \min\{\gamma_{\rm C}, \gamma_{\rm M}\} = \max\{\nu,\, 1 - k\rho\}$. We will have $\gamma = 1 - k\rho$ if $\nu \leq 1 - k\rho$, equivalently,

$$\mathrm{SNR}_{\mathcal{M}} \geq \frac{1}{C_{\mathcal{M}}} \log \frac{k}{1 - \nu}\,, \tag{5.29}$$

otherwise $\gamma = \nu$. The inequality (5.29) holds since

$$\begin{aligned}\mathrm{SNR}_{\mathcal{M}} =& \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}} + b_m\right)}\\ \geq& \frac{\left[\sum_{m\in\mathcal{M}}(a_m - b_m)\right]^2}{k^{2\mathcal{M}_{\max}-2}(\mathcal{M}_{\max} - 1)d} \geq \frac{(C_\nu)^2}{\mathcal{M}_{\max} - 1} \log\frac{k}{1-\nu} \geq \frac{1}{C_{\mathcal{M}}}\log\frac{k}{1-\nu}\,.\end{aligned}$$

where the first two inequalities holds according to $d := \sum_{m\in\mathcal{M}}(m-1)a_m$ and Condition (1.2), while the last inequality holds when taking C_ν in (1.2) satisfying $C_\nu \geq \sqrt{(\mathcal{M}_{\max} - 1)/C_{\mathcal{M}}}$, which finishes the proof.

Remark 5.14. *Condition (5.29) indicates that the improvement of accuracy from local refinement (Algorithm 4.3 and Algorithm 4.4) will be guaranteed when $\mathrm{SNR}_{\mathcal{M}}$ is large enough. If $\mathrm{SNR}_{\mathcal{M}}$ is small, we use correctness of Algorithm 4.2 instead, i.e., $\gamma = \nu$, to represent the correctness of Algorithm 1.1.*

5.5. **Proof of Corollary 1.7.** For any fixed $\nu \in (0.5, 1)$, $\mathrm{SNR}_{\mathcal{M}} \to \infty$ implies $\rho \to 0$ and $d = \sum_{m\in\mathcal{M}}(m-1)a_m \to \infty$. Since

$$\frac{\sum_{m\in\mathcal{M}}(m-1)(a_m - b_m)^2}{\sum_m (m-1)a_m} \geq \frac{\sum_{m\in\mathcal{M}}(m-1)(a_m - b_m)^2}{\sum_m (m-1)k^{m-1}(a_m + k^{m-1}b_m)} = \mathrm{SNR}\,,$$

Condition (1.2) is satisfied. Applying Theorem 1.4, we find $\gamma = 1 - o(1)$, which implies weak consistency. The constraint of $\mathrm{SNR}_{\mathcal{M}} = o(\log n)$ is used in the proof of Lemma 4.2, see Remark 5.13.

Acknowledgements. This work is partially supported by NSF DMS-1949617. I.D. and Y.Z. acknowledge support from NSF DMS-1928930 during their participation in the program "Universality and Integrability in Random Matrix Theory and Interacting Particle Systems" hosted by the Mathematical Sciences Research Institute in Berkeley, California during the Fall semester of 2021. Y.Z. thanks Zhixin Zhou for helpful comments.

References

[1] Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[2] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016.

[3] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015.

[4] Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

[5] Kwangjun Ahn, Kangwook Lee, and Changho Suh. Community recovery in hypergraphs. In *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*, pages 657–663. IEEE, 2016.

[6] Kwangjun Ahn, Kangwook Lee, and Changho Suh. Hypergraph spectral clustering in the weighted stochastic block model. *IEEE Journal of Selected Topics in Signal Processing*, 12(5):959–974, 2018.

[7] Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*, pages 66–73. IEEE, 2015.

[8] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92, 2020.

[9] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[10] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Annals of probability*, 46(1):1–71, 2018.

[11] Jiajun Bu, Shulong Tan, Chun Chen, Can Wang, Hao Wu, Lijun Zhang, and Xiaofei He. Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 391–400, 2010.

[12] Elena Cáceres, Anderson Misobuchi, and Rafael Pimentel. Sparse SYK and traversable wormholes. *Journal of High Energy Physics*, 2021(11):1–32, 2021.

[13] Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021.

[14] I Chien, Chung-Yi Lin, and I-Hsiang Wang. Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 871–879, 2018.

[15] I Eli Chien, Chung-Yi Lin, and I-Hsiang Wang. On the minimax misclassification ratio of hypergraph community detection. *IEEE Transactions on Information Theory*, 65(12):8095–8118, 2019.

[16] Peter Chin, Anup Rao, and Van Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory*, pages 391–423, 2015.

[17] Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593:45–73, 2020.

[18] Nicholas Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. *The Annals of Probability*, 46(1):72–125, 2018.

[19] Joshua Cooper. Adjacency spectra of random and complete hypergraphs. *Linear Algebra and its Applications*, 596:184–202, 2020.

[20] Ioana Dumitriu and Yizhe Zhu. Spectra of random regular hypergraphs. *Electronic Journal of Combinatorics*, 28(3):P3.36, 2021.

[21] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.

[22] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[23] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *Advances in Neural Information Processing Systems*, pages 397–405, 2014.

[24] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, 2017.

[25] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Uniform hypergraph partitioning: Provable tensor methods and sampling techniques. *The Journal of Machine Learning Research*, 18(1):1638–1678, 2017.
[26] Venu Madhav Govindu. A tensor decomposition for geometric grouping and segmentation. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 1150–1157. IEEE, 2005.
[27] Kameron Decker Harris and Yizhe Zhu. Deterministic tensor completion with hypergraph expanders. *SIAM Journal on Mathematics of Data Science*, 3(4):1117–1140, 2021.
[28] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
[29] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems*, pages 1431–1439, 2014.
[30] Jiashun Jin, Tracy Ke, and Jiajun Liang. Sharp impossibility results for hyper-graph testing. *Advances in Neural Information Processing Systems*, 34, 2021.
[31] Zheng Tracy Ke, Feng Shi, and Dong Xia. Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*, 2019.
[32] Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach. *arXiv preprint arXiv:1807.02884*, 2018.
[33] Can M Le and Elizaveta Levina. Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*, 2015.
[34] Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.
[35] Jeonghwan Lee, Daesung Kim, and Hye Won Chung. Robust hypergraph clustering via convex relaxation of truncated MLE. *IEEE Journal on Selected Areas in Information Theory*, 2020.
[36] Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73, 2020.
[37] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.
[38] Lei Li and Tao Li. News recommendation via hypergraph learning: encapsulation of user behavior and news content. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 305–314, 2013.
[39] Chung-Yi Lin, I Eli Chien, and I-Hsiang Wang. On the fundamental statistical limit of community detection in random hypergraphs. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pages 2178–2182. IEEE, 2017.
[40] Rodica Ioana Lung, Noémi Gaskó, and Mihai Alexandru Suciu. A hypergraph model for representing scientific output. *Scientometrics*, 117(3):1361–1379, 2018.
[41] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703. ACM, 2014.
[42] Tom Michoel and Bruno Nachtergaele. Alignment and integration of complex networks by hypergraph-based spectral clustering. *Physical Review E*, 86(5):056111, 2012.
[43] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015.
[44] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. *Electron. J. Probab*, 21(21):1–24, 2016.
[45] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.
[46] Mark EJ Newman, Duncan J Watts, and Steven H Strogatz. Random graph models of social networks. *Proceedings of the national academy of sciences*, 99(suppl 1):2566–2572, 2002.
[47] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.
[48] Nam H Nguyen, Petros Drineas, and Trac D Tran. Tensor sparsification via a bound on the spectral norm of random tensors. *Information and Inference: A Journal of the IMA*, 4(3):195–229, 2015.
[49] Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, 59(3):407–463, 2021.

[50] Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures & Algorithms*, 53(3):504–536, 2018.
[51] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
[52] Tasuku Soma and Yuichi Yoshida. Spectral sparsification of hypergraphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2570–2581. SIAM, 2019.
[53] Ze Tian, TaeHyun Hwang, and Rui Kuang. A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge. *Bioinformatics*, 25(21):2831–2838, 2009.
[54] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
[55] Longyin Wen, Dawei Du, Shengkun Li, Xiao Bian, and Siwei Lyu. Learning non-uniform hypergraph for multi-object tracking. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8981–8988, 2019.
[56] Mingao Yuan, Ruiqi Liu, Yang Feng, and Zuofeng Shang. Testing community structures for hypergraphs. *arXiv preprint arXiv:1810.04617*, 2018.
[57] Mingao Yuan, Bin Zhao, and Xiaofeng Zhao. Community detection in censored hypergraph. *arXiv preprint arXiv:2111.03179*, 2021.
[58] Anderson Y Zhang and Harrison H Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.
[59] Qiaosheng Zhang and Vincent YF Tan. Exact recovery in the general hypergraph stochastic block model. *arXiv preprint arXiv:2105.04770*, 2021.
[60] Yaoming Zhen and Junhui Wang. Community detection in general hypergraph via graph embedding. *arXiv preprint arXiv:2103.15035*, 2021.
[61] Denny Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems*, pages 1601–1608, 2007.
[62] Zhixin Zhou and Yizhe Zhu. Sparse random tensors: Concentration, regularization and applications. *Electronic Journal of Statistics*, 15(1):2483–2516, 2021.

Appendix A. Proof of Theorem 3.1 and Theorem 3.3

A.1. **Discretization.** To prove Theorem 3.1, we start with a standard ε-net argument.

Lemma A.1 (Lemma 4.4.1 in [54])**.** *Let* $\mathbf{W}$ *be any Hermitian* $n \times n$ *matrix and let* $\mathcal{N}_\varepsilon$ *be an* ε*-net on the unit sphere* $\mathbb{S}^{n-1}$ *with* $\varepsilon \in (0,1)$*, then* $\|\mathbf{W}\| \leq \frac{1}{1-\varepsilon} \sup_{\boldsymbol{x} \in \mathcal{N}_\varepsilon} |\langle \mathbf{W}\boldsymbol{x}, \boldsymbol{x}\rangle|$.

By [54, Corollary 4.2.13], the size of $\mathcal{N}_\varepsilon$ is bounded by $|\mathcal{N}_\varepsilon| \leq (1+2/\varepsilon)^n$. We would have $\log|\mathcal{N}| \leq n\log(5)$ when $\mathcal{N}$ is taken as an $(1/2)$-net of $\mathbb{S}^n$. Define $\mathbf{W} := \mathbf{A} - \mathbb{E}\mathbf{A}$, then $\mathbf{W}_{ii} = 0$ for each $i \in [n]$ by the definition of adjacency matrix in Equation (2.3), and we obtain

$$\|\mathbf{A} - \mathbb{E}\mathbf{A}\| := \|\mathbf{W}\| \leq 2 \sup_{\boldsymbol{x} \in \mathcal{N}} |\langle \mathbf{W}\boldsymbol{x}, \boldsymbol{x}\rangle|\,. \tag{A.1}$$

For any fixed $\boldsymbol{x} \in \mathbb{S}^{n-1}$, consider the *light* and *heavy* pairs as follows.

$$\mathcal{L}(\boldsymbol{x}) = \left\{(i,j) : |\boldsymbol{x}_i \boldsymbol{x}_j| \leq \frac{\sqrt{d}}{n}\right\}, \quad \mathcal{H}(\boldsymbol{x}) = \left\{(i,j) : |\boldsymbol{x}_i \boldsymbol{x}_j| > \frac{\sqrt{d}}{n}\right\}, \tag{A.2}$$

where $d = \sum_{m=2}^{M}(m-1)d_m$. Thus by the triangle inequality,

$$|\langle \boldsymbol{x}, \mathbf{W}\boldsymbol{x}\rangle| \leq \left|\sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right| + \left|\sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right|,$$

and by Equation (A.1),

$$\|\mathbf{A} - \mathbb{E}\mathbf{A}\| \leq 2\sup_{\boldsymbol{x}\in\mathcal{N}} \left|\sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right| + 2\sup_{\boldsymbol{x}\in\mathcal{N}} \left|\sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right|. \tag{A.3}$$

A.2. **Contribution from light pairs.** For each m-hyperedge $e \in E_m$, we define $\boldsymbol{\mathcal{W}}_e^{(m)} := \boldsymbol{\mathcal{T}}_e^{(m)} - \mathbb{E}\boldsymbol{\mathcal{T}}_e^{(m)}$. Then for any fixed $\boldsymbol{x} \in \mathbb{S}^{n-1}$, the contribution from light couples can be written as

$$
\begin{aligned}
\sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j &= \sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \left(\sum_{m=2}^{M} \sum_{\substack{e\in E_m \\ \{i,j\}\subset e}} \boldsymbol{\mathcal{W}}_e^{(m)} \right) \boldsymbol{x}_i\boldsymbol{x}_j \\
&= \sum_{m=2}^{M} \sum_{e\in E_m} \boldsymbol{\mathcal{W}}_e^{(m)} \left(\sum_{\substack{(i,j)\in\mathcal{L}(\boldsymbol{x}) \\ i\neq j,\, \{i,j\}\subset e}} \boldsymbol{x}_i\boldsymbol{x}_j \right) = \sum_{m=2}^{M} \sum_{e\in E_m} \boldsymbol{\mathcal{Y}}_e^{(m)} , \qquad \text{(A.4)}
\end{aligned}
$$

where the constraint $i \neq j$ comes from the fact $\mathbf{W}_{ii} = 0$ and we denote

$$
\boldsymbol{\mathcal{Y}}_e^{(m)} := \boldsymbol{\mathcal{W}}_e^{(m)} \left(\sum_{\substack{(i,j)\in\mathcal{L}(\boldsymbol{x}) \\ i\neq j,\, \{i,j\}\subset e}} \boldsymbol{x}_i\boldsymbol{x}_j \right) .
$$

Note that $\mathbb{E}\boldsymbol{\mathcal{Y}}_e^{(m)} = 0$, and by the definition of light pair (A.2),

$$
|\boldsymbol{\mathcal{Y}}_e^{(m)}| \leq m(m-1)\sqrt{d}/n \leq M(M-1)\sqrt{d}/n , \quad \forall m \in \{2, \cdots, M\} .
$$

Moreover, Equation (A.4) is a sum of independent, mean-zero random variables, and

$$
\begin{aligned}
&\sum_{m=2}^{M} \sum_{e\in E_m} \mathbb{E}[(\boldsymbol{\mathcal{Y}}_e^{(m)})^2] := \sum_{m=2}^{M} \sum_{e\in E_m} \left[\mathbb{E}[(\boldsymbol{\mathcal{W}}_e^{(m)})^2] \left(\sum_{\substack{(i,j)\in\mathcal{L}(\boldsymbol{x}) \\ i\neq j, \{i,j\}\subset e}} \boldsymbol{x}_i\boldsymbol{x}_j \right)^2 \right] \\
\leq& \sum_{m=2}^{M} \sum_{e\in E_m} \left[\mathbb{E}[\boldsymbol{\mathcal{T}}_e^{(m)}] \cdot m(m-1) \left(\sum_{\substack{(i,j)\in\mathcal{L}(x) \\ i\neq j, \{i,j\}\subset e}} \boldsymbol{x}_i^2\boldsymbol{x}_j^2 \right) \right] \leq \sum_{m=2}^{M} \frac{d_m \cdot m(m-1)}{\binom{n}{m-1}} \binom{n}{m-2} \sum_{(i,j)\in[n]^2} \boldsymbol{x}_i^2\boldsymbol{x}_j^2 \\
\leq& \sum_{m=2}^{M} \frac{d_m m(m-1)^2}{n-m+2} \leq \frac{2}{n} \sum_{m=2}^{M} d_m (m-1)^3 \leq \frac{2d(M-1)^2}{n} ,
\end{aligned}
$$

when $n \geq 2m-2$, where $d_m = \max d_{i_1,\dots,i_m}$ and $d = \sum_{m=2}^{M}(m-1)d_m$. Then Bernstein's inequality (Lemma D.3) implies that for any $\alpha > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\left| \sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \geq \alpha\sqrt{d} \right) = \mathbb{P}\left(\left| \sum_{m=2}^{M} \sum_{e\in E_m} \boldsymbol{\mathcal{Y}}_e^{(m)} \right| \geq \alpha\sqrt{d} \right) \\
\leq& 2\exp\left(- \frac{\frac{1}{2}\alpha^2 d}{\frac{2d}{n}(M-1)^2 + \frac{1}{3}(M-1)M\frac{\sqrt{d}}{n}\alpha\sqrt{d}} \right) \leq 2\exp\left(- \frac{\alpha^2 n}{4(M-1)^2 + \frac{2\alpha(M-1)M}{3}} \right).
\end{aligned}
$$

Therefore by taking a union bound,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{x}\in\mathcal{N}} \left| \sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \geq \alpha\sqrt{d} \right) \leq |\mathcal{N}| \cdot \mathbb{P}\left(\left| \sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})} \mathbf{W}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \geq \alpha\sqrt{d} \right) \\
\leq& 2\exp\left(\log(5)\cdot n - \frac{\alpha^2 n}{4(M-1)^2 + \frac{2\alpha(M-1)M}{3}} \right) \leq 2e^{-n} , \qquad \text{(A.7)}
\end{aligned}
$$

where we choose $\alpha = 5M(M-1)$ in the last line.

A.3. **Contribution from heavy pairs.** Note that for any $i \neq j$,

$$\mathbb{E}\mathbf{A}_{ij} \leq \sum_{m=2}^{M} \binom{n-2}{m-2} \frac{d_m}{\binom{n}{m-1}} = \sum_{m=2}^{M} \frac{(m-1)d_m}{n} = \frac{d}{n}. \tag{A.8}$$

and

$$\left| \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \mathbb{E}\mathbf{A}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| = \left| \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \mathbb{E}\mathbf{A}_{ij} \frac{\boldsymbol{x}_i^2\boldsymbol{x}_j^2}{\boldsymbol{x}_i\boldsymbol{x}_j} \right| \leq \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \frac{d}{n} \frac{\boldsymbol{x}_i^2\boldsymbol{x}_j^2}{|\boldsymbol{x}_i\boldsymbol{x}_j|} \leq \sqrt{d} \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \boldsymbol{x}_i^2\boldsymbol{x}_j^2 \leq \sqrt{d}. \tag{A.9}$$

Therefore it suffices to show that, with high probability,

$$\sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} \mathbf{A}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j = O\Big(\sqrt{d}\Big). \tag{A.10}$$

Here we use the discrepancy analysis from [21, 18]. We consider the weighted graph associated with the adjacency matrix $\mathbf{A}$.

Definition A.2 (Uniform upper tail property, **UUTP**). *Let $\mathbf{M}$ be an $n \times n$ random symmetric matrix with non-negative entries and define*

$$\mu := \sum_{i,j=1}^{n} \mathbf{Q}_{ij}\mathbb{E}\mathbf{M}_{ij}, \quad \tilde{\sigma}^2 := \sum_{i,j=1}^{n} \mathbf{Q}_{ij}^2\mathbb{E}\mathbf{M}_{ij}.$$

We say that $\mathbf{M}$ satisfies the uniform upper tail property $\mathbf{UUTP}(c_0, \gamma_0)$ with $c_0 > 0, \gamma_0 \geq 0$, if for any $a, t > 0$ and any $n \times n$ symmetric matrix $\mathbf{Q}$ with entries $\mathbf{Q}_{ij} \in [0, a]$ for all $i, j \in [n]$,

$$\mathbb{P}\left(f_{\mathbf{Q}}(\mathbf{M}) \geq (1+\gamma_0)\mu + t \right) \leq \exp\left(-c_0 \frac{\tilde{\sigma}^2}{a^2} h\left(\frac{at}{\tilde{\sigma}^2} \right) \right).$$

where function $f_{\mathbf{Q}}(\mathbf{M}) : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$ is defined by $f_{\mathbf{Q}}(\mathbf{M}) := \sum_{i,j=1}^{n} \mathbf{Q}_{ij}\mathbf{M}_{ij}$ for $\mathbf{M} \in \mathbb{R}^{n\times n}$, and function $h(x) := (1+x)\log(1+x) - x$ for all $x > -1$.

Lemma A.3. *Let $\mathbf{A}$ be the adjacency matrix of the non-uniform hypergraph $H = \cup_{m=2}^{M} H_m$, then $\mathbf{A}$ satisfies $\mathbf{UUTP}(c_0, \gamma_0)$ with $c_0 = [M(M-1)]^{-1}, \gamma_0 = 0$.*

Proof. Note that

$$\begin{aligned} f_{\mathbf{Q}}(\mathbf{A}) - \mu &= \sum_{i,j=1}^{n} \mathbf{Q}_{ij}(\mathbf{A}_{ij} - \mathbb{E}\mathbf{A}_{ij}) = \sum_{i,j=1}^{n} \mathbf{Q}_{ij}\mathbf{W}_{ij} \\ &= \sum_{i,j=1}^{n} \mathbf{Q}_{ij} \Bigg(\sum_{m=2}^{M} \sum_{\substack{e\in E_m \\ i\neq j,\, \{i,j\}\subset e}} \boldsymbol{\mathcal{W}}_e^{(m)} \Bigg) = \sum_{m=2}^{M} \sum_{e\in E_m} \boldsymbol{\mathcal{W}}_e^{(m)} \Bigg(\sum_{\{i,j\}\subset e,\, i\neq j} \mathbf{Q}_{ij} \Bigg) = \sum_{m=2}^{M} \sum_{e\in E_m} \boldsymbol{\mathcal{Z}}_e^{(m)}, \end{aligned}$$

where $\boldsymbol{\mathcal{Z}}_e^{(m)} = \boldsymbol{\mathcal{W}}_e^{(m)} \big(\sum_{\{i,j\}\subset e, i\neq j} \mathbf{Q}_{ij} \big)$ are independent centered random variables upper bounded by $|\boldsymbol{\mathcal{Z}}_e^{(m)}| \leq \sum_{\{i,j\}\subset e,\, i\neq j} \mathbf{Q}_{ij} \leq M(M-1)a$ for each $m \in \{2, \dots, M\}$ since $\mathbf{Q}_{ij} \in [0, a]$. Moreover, the variance of the sum can be written as

$$\begin{aligned} &\sum_{m=2}^{M} \sum_{e\in E_m} \mathbb{E}(\boldsymbol{\mathcal{Z}}_e^{(m)})^2 := \sum_{m=2}^{M} \sum_{e\in E_m} \mathbb{E}(\boldsymbol{\mathcal{W}}_e^{(m)})^2 \Bigg(\sum_{\{i,j\}\subset e,\, i\neq j} \mathbf{Q}_{ij} \Bigg)^2 \\ \leq &\sum_{m=2}^{M} \sum_{e\in E_m} \mathbb{E}[\boldsymbol{\mathcal{T}}_e^{(m)}] \cdot m(m-1) \sum_{\{i,j\}\subset e, i\neq j} \mathbf{Q}_{ij}^2 \leq M(M-1) \sum_{i,j=1}^{n} \mathbf{Q}_{ij}^2 \mathbb{E}\mathbf{A}_{ij} = M(M-1)\tilde{\sigma}^2. \end{aligned}$$

where the last inequality holds since by definition $\mathbb{E}\mathbf{A}_{ij} = \sum_{m=2}^{M} \sum_{\substack{e\in E_m \\ \{i,j\}\subset e}} \mathbb{E}[\boldsymbol{\mathcal{T}}_e^{(m)}]$. Then by Bennett's inequality D.4, we obtain

$$\mathbb{P}(f_{\mathbf{Q}}(\mathbf{A}) - \mu \geq t) \leq \exp\left(- \frac{\tilde{\sigma}^2}{M(M-1)a^2} h\left(\frac{at}{\tilde{\sigma}^2} \right) \right)$$

where the inequality holds since the function $x \cdot h(1/x) = (1+x)\log(1+1/x) - 1$ is decreasing with respect to x. □

Definition A.4 (Discrepancy property, **DP**)**.** *Let $\mathbf{M}$ be an $n \times n$ matrix with non-negative entries. For $S, T \subset [n]$, define $e_{\mathbf{M}}(S,T) = \sum_{i\in S, j\in T} \mathbf{M}_{ij}$. We say $\mathbf{M}$ has the discrepancy property with parameter $\delta > 0$, $\kappa_1 > 1, \kappa_2 \geq 0$, denoted by $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$, if for all non-empty $S, T \subset [n]$, at least one of the following hold:*

(1) $e_{\mathbf{M}}(S,T) \leq \kappa_1 \delta |S||T|$;

(2) $e_{\mathbf{M}}(S,T) \cdot \log\left(\frac{e_{\mathbf{M}}(S,T)}{\delta |S| \cdot |T|} \right) \leq \kappa_2 (|S| \vee |T|) \cdot \log\left(\frac{en}{|S| \vee |T|} \right)$.

The following shows that if a symmetric random matrix $\mathbf{A}$ satisfies the upper tail property $\mathbf{UUTP}(c_0, \gamma_0)$ with parameter $c_0 > 0, \gamma_0 \geq 0$, then the discrepancy property holds with high probability.

Lemma A.5 (Lemma 6.4 in [18])**.** *Let $\mathbf{M}$ be an $n \times n$ symmetric random matrix with non-negative entries. Assume that for some $\delta > 0$, $\mathbb{E}\mathbf{M}_{ij} \leq \delta$ for all $i, j \in [n]$ and $\mathbf{M}$ has $\mathbf{UUTP}(c_0, \gamma_0)$ with parameter $c_0, \gamma_0 > 0$. Then for any $K > 0$, the discrepancy property $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ holds for $\mathbf{M}$ with probability at least $1 - n^{-K}$ with $\kappa_1 = e^2(1+\gamma_0)^2, \kappa_2 = \frac{2}{c_0}(1+\gamma_0)(K+4)$.*

When the discrepancy property holds, then deterministically the contribution from heavy pairs is $O(\sqrt{d})$, as shown in the following lemma.

Lemma A.6 (Lemma 6.6 in [18])**.** *Let $\mathbf{M}$ be a non-negative symmetric $n \times n$ matrix with all row sums bounded by d. Suppose $\mathbf{M}$ has $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ with $\delta = Cd/n$ for some $C > 0, \kappa_1 > 1, \kappa_2 \geq 0$. Then for any $x \in \mathbb{S}^{n-1}$,*

$$\left| \sum_{(i,j)\in\mathcal{H}(x)} \mathbf{M}_{ij} \boldsymbol{x}_i \boldsymbol{x}_j \right| \leq \alpha_0 \sqrt{d},$$

where $\alpha_0 = 16 + 32C(1+\kappa_1) + 64\kappa_2(1 + \frac{2}{\kappa_1 \log \kappa_1})$.

The next lemma shows $\mathbf{A}$ has bounded row and column sums with high probability.

Lemma A.7. *For any $K > 0$, there is a constant $\alpha_1 > 0$ such that with probability at least $1 - n^{-K}$,*

$$\max_{1\leq i\leq n} \sum_{j=1}^{n} \mathbf{A}_{ij} \leq \alpha_1 d \tag{A.13}$$

with $\alpha_1 = 4 + \frac{2(M-1)(1+K)}{3c}$ and $d \geq c \log n$.

Proof. For a fixed $i \in [n]$,

$$\sum_{j=1}^{n} \mathbf{A}_{ij} = \sum_{m=2}^{M} \sum_{e\in E_m : i\in e} (m-1)\boldsymbol{\mathcal{T}}_e^{(m)}, \quad \sum_{j=1}^{n} (\mathbf{A}_{ij} - \mathbb{E}\mathbf{A}_{ij}) = \sum_{m=2}^{M} \sum_{e\in E_m : i\in e} (m-1)\boldsymbol{\mathcal{W}}_e^{(m)},$$

$$\sum_{j=1}^{n} \mathbb{E}\mathbf{A}_{ij} \leq \sum_{m=2}^{M} \binom{n}{m-1} \frac{(m-1)d_m}{\binom{n}{m-1}} = d,$$

$$\sum_{m=2}^{M} (m-1)^2 \sum_{e\in E_m : i\in e} \mathbb{E}[(\boldsymbol{\mathcal{W}}_e^{(m)})^2] \leq \sum_{m=2}^{M} (m-1)^2 \sum_{e\in E_m : i\in e} \mathbb{E}[\boldsymbol{\mathcal{T}}_e^{(m)}] \leq (M-1)d\,.$$

Then for $\alpha_1 = 4 + \frac{2(M-1)(1+K)}{3c}$, by Bernstein's inequality, with the assumption that $d \geq c \log n$,

$$\mathbb{P}\left(\sum_{j=1}^{n} \mathbf{A}_{ij} \geq \alpha_1 d\right) \leq \mathbb{P}\left(\sum_{j=1}^{n} \mathbf{A}_{ij} - \mathbb{E}\mathbf{A}_{ij} \geq (\alpha_1 - 1)d\right)$$

$$\leq \exp\left(-\frac{\frac{1}{2}(\alpha_1-1)^2 d^2}{(M-1)d + \frac{1}{3}(M-1)(\alpha_1-1)d}\right) \leq n^{-\frac{3c(\alpha_1-1)^2}{(M-1)(2\alpha_1+4)}} \leq n^{-1-K}. \tag{A.15}$$

Taking a union bound over $i \in [n]$, then Equation (A.13) holds with probability $1 - n^{-K}$. $\square$

Now we are ready to obtain Equation (A.10).

Lemma A.8. *For any $K > 0$, there is a constant β depending on K, c, M such that with probability at least $1 - 2n^{-K}$,*

$$\left|\sum_{(i,j)\in\mathcal{H}(x)} \mathbf{A}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right| \leq \beta\sqrt{d}. \tag{A.16}$$

Proof. By Lemma A.3, $\mathbf{A}$ satisfies $\mathbf{UUTP}(\frac{1}{M(M-1)}, 0)$. From Equation (A.8) and Lemma A.5, the property $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ holds for $\mathbf{A}$ with probability at least $1 - n^{-K}$ with

$$\delta = \frac{d}{n}, \quad \kappa_1 = e^2, \quad \kappa_2 = 2M(M-1)(K+4).$$

Let $\mathcal{E}_1$ be the event that $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ holds for $\mathbf{A}$. Let $\mathcal{E}_2$ be the event that all row sums of $\mathbf{A}$ are bounded by $\alpha_1 d$. Then $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - 2n^{-K}$. On the event $\mathcal{E}_1 \cap \mathcal{E}_2$, by Lemma A.6, Equation (A.16) holds with $\beta = \alpha_0\alpha_1$, where

$$\alpha_0 = 16 + 32(1+e^2) + 128M(M-1)(K+4)(1+e^{-2}), \quad \alpha_1 = 4 + \frac{2(M-1)(1+K)}{3c}.$$

$\square$

A.4. **Proof of Theorem 3.1.**

Proof. From (A.7), with probability at least $1 - 2e^{-n}$, the contribution from light pairs in (A.3) is bounded by $2\alpha\sqrt{d}$ with $\alpha = 5M(M-1)$. From (A.9) and (A.16), with probability at least $1-2n^{-K}$, the contribution from heavy pairs in (A.3) is bounded by $2\sqrt{d} + 2\beta\sqrt{d}$. Therefore with probability at least $1 - 2e^{-n} - 2n^{-K}$,

$$\|\mathbf{A} - \mathbb{E}\mathbf{A}\| \leq C_M\sqrt{d},$$

where C_M is a constant depending only on c, K, M such that $C_M = 2(\alpha+1+\beta)$. In particular, we can take $\alpha = 5M(M-1)$, $\beta = 512M(M-1)(K+5)\left(2 + \frac{(M-1)(1+K)}{c}\right)$, and $C_M = 512M(M-1)(K+6)\left(2 + \frac{(M-1)(1+K)}{c}\right)$. This finishes the proof of Theorem 3.1. $\square$

A.5. **Proof of Theorem 3.3.** Let $\mathcal{S} \subset [n]$ be any given subset. From (A.7), with probability at least $1 - 2e^{-n}$,

$$\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})}\left(\mathbf{A}_{\mathcal{S}} - \mathbb{E}\mathbf{A}_{\mathcal{S}}\right)_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right| \leq 5M(M-1)\sqrt{d}. \tag{A.17}$$

Since there are at most 2^n many choices for $\mathcal{S}$, by taking a union bound, with probability at least $1-2(e/2)^{-n}$, we have for all $\mathcal{S} \subset [n]$, Equation (A.17) holds. In particular, by taking $\mathcal{S} = \mathcal{I} = \{i \in [n] : \mathrm{row}(i) \leq \tau d\}$, with probability at least $1 - 2(e/2)^{-n}$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{N}}\left|\sum_{(i,j)\in\mathcal{L}(\boldsymbol{x})}[(\mathbf{A} - \mathbb{E}\mathbf{A})_{\mathcal{I}}]_{ij}\boldsymbol{x}_i\boldsymbol{x}_j\right| \leq 5M(M-1)\sqrt{d}. \tag{A.18}$$

Similar to Equation (A.9), deterministically,

$$\left| \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} [(\mathbb{E}\mathbf{A})_{\mathcal{I}}]_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \le (M-1)\sqrt{d}. \tag{A.19}$$

Next we show the contribution from heavy pairs for $\mathbf{A}_{\mathcal{I}}$ is bounded.

Lemma A.9. *For any $K > 0$, there is a constant β_τ depending on K, c, M, τ such that with probability at least $1 - n^{-K}$,*

$$\left| \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} [(\mathbf{A})_{\mathcal{I}}]_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \le \beta_\tau\sqrt{d}. \tag{A.20}$$

Proof. Note that $\mathbf{A}$ satisfies $\mathbf{UUTP}\left(\frac{1}{M(M-1)}, 0\right)$ from Lemma A.3. Then from Lemma A.5, with probability at least $1 - n^{-K}$, $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ holds for $\mathbf{A}$ with

$$\delta = \frac{d}{n}, \quad \kappa_1 = e^2, \quad \kappa_2 = 2M(M-1)(K+4).$$

The $\mathbf{DP}(\delta, \kappa_1, \kappa_2)$ property holds for $\mathbf{A}_{\mathcal{I}}$ as well, since $\mathbf{A}_{\mathcal{I}}$ is obtained from $\mathbf{A}$ by restricting to $\mathcal{I}$. Note that all row sums in $\mathbf{A}_{\mathcal{I}}$ are bounded by τd. By Lemma A.6,

$$\left| \sum_{(i,j)\in\mathcal{H}(\boldsymbol{x})} [\mathbf{A}_{\mathcal{I}}]_{ij}\boldsymbol{x}_i\boldsymbol{x}_j \right| \le \alpha_0\sqrt{\tau d}, \tag{A.21}$$

where we can take $\alpha_0 = 16 + \frac{32}{\tau}(1 + e^2) + 128M(M-1)(K+4)\left(1 + \frac{1}{e^2}\right)$. □

We can then take $\beta_\tau = \alpha_0\sqrt{\tau}$ in Equation (A.20). Therefore, combining Equation (A.18), Equation (A.19), Equation (A.21), with probability at least $1 - 2(e/2)^{-n} - n^{-K}$, there exists a constant C_τ depending only on τ, M, K such that $\|(\mathbf{A} - \mathbb{E}\mathbf{A})_{\mathcal{I}}\| \le C_\tau\sqrt{d}$, where $C_\tau = 2((5M+1)(M-1) + \alpha_0\sqrt{\tau})$. This finishes the proof of Theorem 3.3.

Appendix B. Technical Lemmas

B.1. Proof of Lemma 2.4.

Proof. By Weyl's inequality D.5, the difference between eigenvalues of $\widetilde{\mathbb{E}\mathbf{A}}$ and $\mathbb{E}\mathbf{A}$ can be upper bounded by

$$\begin{aligned} |\lambda_i(\widetilde{\mathbb{E}\mathbf{A}}) - \lambda_i(\mathbb{E}\mathbf{A})| &\le \|\widetilde{\mathbb{E}\mathbf{A}} - \mathbb{E}\mathbf{A}\|_2 \le \|\widetilde{\mathbb{E}\mathbf{A}} - \mathbb{E}\mathbf{A}\|_{\mathrm{F}} \\ &\le \left[2k \cdot \frac{n}{k} \cdot \sqrt{n}\log(n) \cdot (\alpha-\beta)^2\right]^{1/2} = O\Big(n^{3/4}\log^{1/2}(n)(\alpha-\beta)\Big). \end{aligned}$$

The lemma follows, as $\lambda_i(\mathbb{E}\mathbf{A}) = \Omega\left(n(\alpha-\beta)\right)$ for all $1 \le i \le k$. □

B.2. Proof of Lemma 5.3.

Proof. We first compute the singular values of $\overline{\mathbf{B}}_1$. From Equation (5.5), the rank of matrix $\overline{\mathbf{B}}_1$ is k, and the least non-trivial singular value of $\overline{\mathbf{B}}_1$ is

$$\sigma_k(\overline{\mathbf{B}}_1) = \frac{n}{2\sqrt{2}k}(\overline{\alpha} - \overline{\beta}) = \frac{n}{2\sqrt{2}k}\sum_{m\in\mathcal{M}} \binom{\frac{3n}{4k} - 2}{m-2}\frac{a_m - b_m}{\binom{n}{m-1}},$$

where $\mathcal{M}$ is obtained from Algorithm 4.1. By the definition of $\overline{\mathbf{A}}_1$ in Equation (5.9), the least non-trivial singular value of $\overline{\mathbf{A}}_1$ is

$$\sigma_k(\overline{\mathbf{A}}_1) = \sigma_k(\overline{\mathbf{B}}_1) = \frac{n}{2\sqrt{2}k}(\overline{\alpha} - \overline{\beta}) = \frac{n}{2\sqrt{2}k}\sum_{m\in\mathcal{M}} \binom{\frac{3n}{4k} - 2}{m-2}\frac{a_m - b_m}{\binom{n}{m-1}}.$$

Recall that n_i, defined in Equation (5.1), denotes the number of vertices in $Z \cap V_i$, which can be written as $n_i = \sum_{v \in V_i} \mathbf{1}_{\{v \in Z\}}$. By Hoeffding's Lemma D.2,

$$\mathbb{P}\left(\left| n_i - \frac{n}{2k} \right| \geq \sqrt{n} \log(n) \right) \leq 2 \exp\left(-k \log^2(n)\right) .$$

Similarly, n_i', defined in Equation (5.2), satisfies

$$\mathbb{P}\left(\left| n_i' - \frac{n}{4k} \right| \geq \sqrt{n} \log(n) \right) \leq 2 \exp\left(-k \log^2(n)\right) .$$

As defined in Equation (5.3) and Equation (5.5), both $\widetilde{\mathbf{B}}_1$ and $\overline{\mathbf{B}}_1$ are deterministic block matrices. Then with probability at least $1 - 2k \exp\left(-k \log^2(n)\right)$, the dimensions of each block inside $\widetilde{\mathbf{B}}_1$ and $\overline{\mathbf{B}}_1$ are approximately the same, with deviations up to $\sqrt{n} \log(n)$. By Weyl's Lemma D.5, for any $i \in [k]$,

$$\begin{aligned} |\sigma_i(\overline{\mathbf{B}}_1) - \sigma_i(\widetilde{\mathbf{B}}_1)| &= |\sigma_i(\overline{\mathbf{A}}_1) - \sigma_i(\widetilde{\mathbf{A}}_1)| \leq \|\overline{\mathbf{A}}_1 - \widetilde{\mathbf{A}}_1\|_2 \leq \|\overline{\mathbf{A}}_1 - \widetilde{\mathbf{A}}_1\|_{\mathrm{F}} \\ &\leq \left[2k \cdot \frac{n}{k} \cdot \sqrt{n} \log(n) \cdot (\overline{\alpha} - \overline{\beta})^2 \right]^{1/2} = O\left(n^{3/4} \log^{1/2}(n) \cdot (\overline{\alpha} - \overline{\beta}) \right) . \end{aligned}$$

As a result, with probability at least $1 - 2k \exp(-k \log^2(n))$, we have

$$\frac{|\sigma_k(\overline{\mathbf{A}}_1) - \sigma_k(\widetilde{\mathbf{A}}_1)|}{\sigma_k(\overline{\mathbf{A}}_1)} = \frac{|\sigma_k(\overline{\mathbf{B}}_1) - \sigma_k(\widetilde{\mathbf{B}}_1)|}{\sigma_k(\overline{\mathbf{B}}_1)} = O\left(n^{-1/4} \log^{1/2}(n) \right) .$$

□

B.3. **Proof of Lemma 5.4.**

Proof. Without loss of generality, we can assume $\mathcal{M} = \{2, \ldots, M\}$. If $\mathcal{M}$ is a subset of $\{2, \ldots, M\}$, we can take $a_m = b_m = 0$ for $m \notin \mathcal{M}$. Note that in fact, if the best SNR is obtained when $\mathcal{M}$ is a strict subset, we can substitute $\mathcal{M}_{\max}$ for M.

Let $X \subset V$ be a subset of vertices in hypergraph $H = (V, E)$ with size $|X| = cn$ for some $c \in (0, 1)$ to be decided later. Suppose X is a set of vertices with high degrees that we want to zero out. We first count the m-uniform hyperedges on X separately, then weight them by $(m-1)$, and finally sum over m to compute the row sums in $\mathbf{A}$ corresponding to each vertex in X. Let $E_m(X)$ denote the set of m-uniform hyperedges with all vertices located in X, and $E_m(X^c)$ denote the set of m-uniform hyperedges with all vertices in $X^c = V \setminus X$, respectively. Let $E_m(X, X^c)$ denote the set of m-uniform hyperedges with at least 1 endpoint in X and 1 endpoint in X^c. The relationship between total row sums and the number of non-uniform hyperedges in vertex set X can be expressed as

$$\sum_{v \in X} \mathrm{row}(v) \leq \sum_{m=2}^{M} (m-1) \Big(m|E_m(X)| + (m-1)|E_m(X, X^c)| \Big) \tag{B.3}$$

If the row sum of each vertex $v \in X$ is at least $20Md$, where $d = \sum_{m=2}^{M} (m-1) a_m$, it follows that

$$\sum_{m=2}^{M} (m-1) \Big(m|E_m(X)| + (m-1)|E_m(X, X^c)| \Big) \geq cn \cdot (20Md) . \tag{B.4}$$

Then either

$$\sum_{m=2}^{M} m(m-1)|E_m(X)| \geq 4Mcnd, \quad \text{or} \quad \sum_{m=2}^{M} (m-1)^2 |E_m(X, X^c)| \geq 16Mcnd.$$

B.3.1. *Concentration of* $\sum_{m=2}^{M} m(m-1)|E_m(X)|$. Recall that $\left|E_m(X)\right|$ denotes the number of m-uniform hyperedges with all vertices located in X, which can be viewed as the sum of independent Bernoulli random variables $T_e^{(a_m)}$ and $T_e^{(b_m)}$ given by

$$T_e^{(a_m)} \sim \mathrm{Bernoulli}\left(\frac{a_m}{\binom{n}{m-1}} \right), \quad T_e^{(b_m)} \sim \mathrm{Bernoulli}\left(\frac{b_m}{\binom{n}{m-1}} \right) . \tag{B.5}$$

Let $\{V_1, \dots, V_k\}$ be the true partition of V. Suppose that there are $\eta_i cn$ vertices in block $V_i \cap X$ for each $i \in [k]$ with restriction $\sum_{i=1}^k \eta_i = 1$, then $\big|E_m(X)\big|$ can be written as

$$\big|E_m(X)\big| = \sum_{e \in E_m(X, a_m)} T_e^{(a_m)} + \sum_{e \in E_m(X, b_m)} T_e^{(b_m)},$$

where $E_m(X, a_m) := \cup_{i=1}^k E_m(V_i \cap X)$ denotes the union for sets of hyperedges with all vertices in the same block $V_i \cap X$ for some $i \in [k]$, and

$$E_m(X, b_m) := E_m(X) \setminus E_m(X, a_m) = E_m(X) \setminus \Big(\cup_i^k E_m(V_i \cap X) \Big)$$

denotes the set of hyperedges with vertices crossing different $V_i \cap X$. We can compute the expectation of $\big|E_m(X)\big|$ as

$$\mathbb{E}|E_m(X)| = \sum_{i=1}^k \binom{\eta_i cn}{m} \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{cn}{m} \frac{b_m}{\binom{n}{m-1}}. \tag{B.6}$$

Then

$$\sum_{m=2}^M m(m-1) \cdot \mathbb{E}|E_m(X)| = \sum_{m=2}^M m(m-1) \left[\sum_{i=1}^k \binom{\eta_i cn}{m} \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{cn}{m} \frac{b_m}{\binom{n}{m-1}} \right]. \tag{B.7}$$

As $\sum\limits_{i=1}^k \eta_i = 1$, it follows that $\sum_{i=1}^k \binom{\eta_i cn}{m} \leq \binom{cn}{m}$ by induction, thus

$$\frac{a_m - b_m}{\binom{n}{m-1}} \sum_{i=1}^k \binom{\eta_i cn}{m} + \frac{b_m}{\binom{n}{m-1}} \binom{cn}{m} = \frac{a_m}{\binom{n}{m-1}} \sum_{i=1}^k \binom{\eta_i cn}{m} + \frac{b_m}{\binom{n}{m-1}} \left(\binom{cn}{m} - \sum_{i=1}^k \binom{\eta_i cn}{m} \right)$$

where both terms on the right are positive numbers. Using this and taking $b_m = a_m$, we obtain the following upper bound for all n,

$$\sum_{m=2}^M m(m-1)\mathbb{E}|E_m(X)| \leq \sum_{m=2}^M m(m-1) \binom{cn}{m} \frac{a_m}{\binom{n}{m-1}} \leq cn \sum_{m=2}^M (m-1) a_m = cnd\,.$$

Note that $\sum_{m=2}^M m(m-1)|E_m(X)|$ is a weighted sum of independent Bernoulli random variables (corresponding to hyperedges), each upper bounded by M^2. Also, its variance is bounded by

$$\begin{aligned} \sigma^2 := \mathbb{V}\mathrm{ar} \left(\sum_{m=2}^M m(m-1)|E_m(X)| \right) &= \sum_{m=2}^M m^2(m-1)^2 \mathbb{V}\mathrm{ar}\left(|E_m(X)|\right) \\ &\leq \sum_{m=2}^M m^2(m-1)^2 \mathbb{E}|E_m(X)| \leq M^2 cnd. \end{aligned}$$

We can apply Bernstein's Lemma D.3 and obtain

$$\begin{aligned} \mathbb{P}\left(\sum_{m=2}^M m(m-1)|E_m(X)| \geq 4Mcnd \right) &\leq \mathbb{P}\left(\sum_{m=2}^M m(m-1)(|E_m(X)| - \mathbb{E}|E_m(X)|) \geq 3Mcnd \right) \\ &\leq \exp\left(- \frac{(3Mcnd)^2}{M^2 cnd + M^2 cnd/3} \right) \leq \exp(-6cnd)\,. \end{aligned} \tag{B.8}$$

B.3.2. *Concentration of* $\sum_{m=2}^M (m-1)^2 |E_m(X, X^c)|$. For any finite set S, let $[S]^j$ denote the family of j-subsets of S, i.e., $[S]^j = \{Z | Z \subseteq S, |Z| = j\}$. Let $E_m([Y]^j, [Z]^{m-j})$ denote the set of m-hyperedges, where j vertices are from Y and $m - j$ vertices are from Z within each m-hyperedge. We want to count the number of m-hyperedges between X and X^c, according to the number of vertices located in X^c within each m-hyperedge. Suppose that there are j vertices from X^c within each m-hyperedge for some $1 \leq j \leq m-1$.

(i) Assume that all those j vertices are in the same $[V_i \setminus X]^j$. If the remaining $m-j$ vertices are from $[V_i \cap X]^{m-j}$, then this m-hyperedge is connected with probability $a_m/\binom{n}{m-1}$, otherwise $b_m/\binom{n}{m-1}$. The number of this type m-hyperedges can be written as

$$\sum_{i=1}^{k}\left[\sum_{e\in\mathcal{E}_{j,i}^{(a_m)}} T_e^{(a_m)} + \sum_{e\in\mathcal{E}_{j,i}^{(b_m)}} T_e^{(b_m)}\right],$$

where $\mathcal{E}_{j,i}^{(a_m)} := E_m([V_i \cap X^c]^j, [V_i \cap X]^{m-j})$, and

$$\mathcal{E}_{j,i}^{(b_m)} := E_m\Big([V_i \cap X^c]^j,\ [X]^{m-j} \setminus [V_i \cap X]^{m-j}\Big)$$

denotes the set m-hyperedges with j vertices in $[V_i \cap X^c]^j$ and the remaining $m-j$ vertices in $[X]^j \setminus [V_i \cap X]^j$. We compute all possible choices and upper bound the cardinality of $\mathcal{E}_{j,i}^{(a_m)}$ and $\mathcal{E}_{j,i}^{(b_m)}$ by

$$|\mathcal{E}_{j,i}^{(a_m)}| \leq \binom{(\frac{1}{k}-\eta_i c)n}{j}\binom{\eta_i cn}{m-j}, \quad |\mathcal{E}_{j,i}^{(b_m)}| \leq \binom{(\frac{1}{k}-\eta_i c)n}{j}\left[\binom{cn}{m-j} - \binom{\eta_i cn}{m-j}\right].$$

(ii) If those j vertices in $[V \setminus X]^j$ are not in the same $[V_i \cap X]^j$ (which only happens $j \geq 2$), then the number of this type hyperedges can be written as $\sum_{e\in\mathcal{E}_j^{(b_m)}} T_e^{(b_m)}$, where

$$\mathcal{E}_j^{(b_m)} := E_m\Big([V \setminus X]^j \setminus \big(\cup_{i=1}^k [V_i \setminus X]^j\big),\ [X]^{m-j}\Big),$$

$$|\mathcal{E}_j^{(b_m)}| \leq \left[\binom{(1-c)n}{j} - \sum_{i=1}^{k}\binom{(\frac{1}{k}-\eta_i c)n}{j}\right]\binom{cn}{m-j}.$$

Therefore, $|E_m(X, X^c)|$ can be written as a sum of independent Bernoulli random variables,

$$|E_m(X, X^c)| = \sum_{j=1}^{m-1}\sum_{i=1}^{k}\left[\sum_{e\in\mathcal{E}_{j,i}^{(a_m)}} T_e^{(a_m)} + \sum_{e\in\mathcal{E}_{j,i}^{(b_m)}} T_e^{(b_m)}\right] + \sum_{j=2}^{m-1}\sum_{e\in\mathcal{E}_j^{(b_m)}} T_e^{(b_m)}. \tag{B.10}$$

Then the expectation can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left(|E_m(X, X^c)|\right) \\
&= \sum_{j=1}^{m-1}\sum_{i=1}^{k}\binom{(\frac{1}{k}-\eta_i c)n}{j}\left\{\binom{\eta_i cn}{m-j}\frac{a_m}{\binom{n}{m-1}} + \left[\binom{cn}{m-j} - \binom{\eta_i cn}{m-j}\right]\frac{b_m}{\binom{n}{m-1}}\right\} \\
&\quad + \sum_{j=1}^{m-1}\left[\binom{(1-c)n}{j} - \sum_{i=1}^{k}\binom{(\frac{1}{k}-\eta_i c)n}{j}\right]\binom{cn}{m-j}\frac{b_m}{\binom{n}{m-1}} \\
&= \sum_{j=1}^{m-1}\sum_{i=1}^{k}\binom{(\frac{1}{k}-\eta_i c)n}{j}\binom{\eta_i cn}{m-j}\frac{a_m-b_m}{\binom{n}{m-1}} + \sum_{j=1}^{m-1}\binom{(1-c)n}{j}\binom{cn}{m-j}\frac{b_m}{\binom{n}{m-1}} \\
&= \sum_{i=1}^{k}\left[\binom{\frac{n}{k}}{m} - \binom{\eta_i cn}{m} - \binom{(\frac{1}{k}-\eta_i c)n}{m}\right]\frac{a_m-b_m}{\binom{n}{m-1}} + \left[\binom{n}{m} - \binom{cn}{m} - \binom{(1-c)n}{m}\right]\frac{b_m}{\binom{n}{m-1}},
\end{aligned} \tag{B.11}$$

where we used the fact $\binom{(1-c)n}{1} = \sum_{i=1}^{k}\binom{(1/k-\eta_i c)n}{1}$ in the first equality and Vandermonde's identity $\binom{n_1+n_2}{m} = \sum_{j=0}^{m}\binom{n_1}{j}\binom{n_2}{m-j}$ in last equality. Note that

$$f_c := \binom{n}{m} - \binom{cn}{m} - \binom{(1-c)n}{m}$$

counts the number of subsets of V with m elements such that at least one element belongs to X and at least one element belongs to X^c. On the other hand,

$$g_c = \sum_{i=1}^{k} \left[\binom{\frac{n}{k}}{m} - \binom{\eta_i cn}{m} - \binom{(\frac{1}{k} - \eta_i c)n}{m} \right] .$$

counts the number of subsets of V with m elements such that all elements belong to a single V_i, and given such an i, that at least one element belongs to $X \cap V_i$ and at least one belongs to $X^c \cap V_i$. As Figure 2

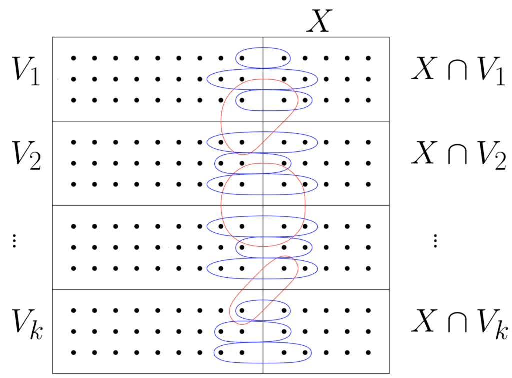

FIGURE 2. Comparison of f_c and g_c

shows, g_c only counts the blue pairs while f_c counts red pairs in addition. By virtue of the fact that there are fewer conditions imposed on the sets included in the count for f_c, we must have $f_c \geq g_c$. Thus, rewriting Equation (B.11), we obtain

$$\mathbb{E}(|E_m(X, X^c)|) = g_c \frac{a_m}{\binom{n}{m-1}} + (f_c - g_c) \frac{b_m}{\binom{n}{m-1}} .$$

Since both terms in the above sum are positive, we can upper bound by taking $a_m = b_m$ to obtain

$$\mathbb{E}\left(|E_m(X, X^c)|\right) \leq f_c \frac{a_m}{\binom{n}{m-1}} = \left[\binom{n}{m} - \binom{cn}{m} - \binom{(1-c)n}{m} \right] \frac{a_m}{\binom{n}{m-1}} .$$

By summing over m, the expectation of $\sum_{m=2}^{M} (m-1)^2 |E_m(X, X^c)|$ satisfies

$$\begin{aligned} \sum_{m=2}^{M} (m-1)^2 \cdot \mathbb{E}\left(|E_m(X, X^c)|\right) &\leq \sum_{m=2}^{M} (m-1)^2 \left[\binom{n}{m} - \binom{cn}{m} - \binom{(1-c)n}{m} \right] \frac{a_m}{\binom{n}{m-1}} , \\ &\leq 2n \sum_{m=2}^{M} (1 - c^m - (1-c)^m)(m-1) a_m \leq 8Mcnd , \end{aligned}$$

where the last upper inequality holds when $c \leq c_0 = c_0(m)$ is sufficiently small, since

$$\begin{aligned} &[(1-c)+c]^m - c^m - (1-c)^m \\ &= \binom{m}{1}(1-c)^{m-1}c + \binom{m}{2}(1-c)^{m-2}c^2 + \cdots + \binom{m}{m-1}(1-c)^1 c^{m-1} \\ &\leq \binom{m}{1}c + \binom{m}{2}c^2 + \cdots + \binom{m}{m-1}c^{m-1} \leq (1+c)^m - 1 \leq 2(1+mc-1) \leq 2mc . \end{aligned} \tag{B.13}$$

Similarly, we apply Bernstein Lemma D.3 again with $K = M^2$, $\sigma^2 \le 8M^3cnd$ and obtain

$$\mathbb{P}\left(\sum_{m=2}^{M}(m-1)^2|E_m(X,X^c)| \ge 16Mcnd\right)$$

$$\le \mathbb{P}\left(\sum_{m=2}^{M}(m-1)^2(|E_m(X,X^c)| - \mathbb{E}|E_m(X,X^c)|) \ge 8Mcnd\right) \le \exp(-6cnd/M)\,. \tag{B.14}$$

By the binomial coefficient upper bound $\binom{n}{k} \le (\frac{en}{k})^k$ for $1 \le k \le n$, there are at most

$$\binom{n}{cn} \le \left(\frac{e}{c}\right)^{cn} = \exp(-c(\log c - 1)n) \tag{B.15}$$

many subsets X of size $|X| = cn$. Let d be sufficiently large so that $d^{-3} \le c_0$. Substituting $c = d^{-3}$ in (B.15), we have

$$\binom{n}{d^{-3}n} \le \exp\left[3d^{-3}\log(d)n\right].$$

Taking $c = d^{-3}$ in (B.8) and (B.14), we obtain

$$\mathbb{P}\left(\sum_{m=2}^{M}(m-1)(m|E_m(X)| + (m-1)|E_m(X,X^c)|) \ge 20Md^{-2}n\right) \le 2\exp(-2d^{-2}n/M)\,.$$

Taking a union bound over all possible X with $|X| = d^{-3}n$, we obtain with probability at least $1 - 2\exp(3d^{-3}\log dn - 2d^{-2}n/M) \le 1 - 2\exp(-d^{-2}n/M)$, no more than $d^{-3}n$ many vertices have total row sum greater than $20Md$. Note that we have imposed the condition that $d^{-3} \le c_0$ (as in B.13), so d needs to be sufficiently large. □

B.4. **Proof of Lemma 5.7.**

Proof. Note that $\mathbf{U}$ is spanned by the first k singular vectors of $(\mathbf{A}_1)_{\mathcal{I}_1}$. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ be an orthonormal basis of the subspace $\mathbf{U}$, then the projection $P_{\mathbf{U}}$ is given by $P_{\mathbf{U}} := \sum_{l=1}^{k}\langle \boldsymbol{u}_l, \cdot\rangle \boldsymbol{u}_l$. For some fixed vertex i with $k(i)$ indexing its membership, we have for $i \in V_{k(i)} \cap Y_2 \cap \{i_1, \cdots, i_s\}$,

$$P_{\mathbf{U}}\boldsymbol{e}_i = \sum_{l=1}^{k}\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle \boldsymbol{u}_l, \quad \|P_{\mathbf{U}}\boldsymbol{e}_i\|_2^2 = \sum_{l=1}^{k}\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle^2.$$

By definition of $\mathbf{A}_1$ and $\mathbf{A}_2$ in Equation (5.7), and the independence between entries in $\mathbf{A}_1$ and entries in $\mathbf{A}_2$, we know that $\{\boldsymbol{u}_l\}_{l=1}^k$ and e_i are independent of each other, since $\{\boldsymbol{u}_l\}_{l=1}^k$ is constructed from $\mathbf{A}_1$ and e_i is chosen from columns of $\mathbf{E}_2 := \mathbf{A}_2 - \widetilde{\mathbf{A}}_2$. If we take expectation over $\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}$ conditioning on $\{\boldsymbol{u}_l\}_{l=1}^k$, then by direct calculation,

$$\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle \Big| \{\boldsymbol{u}_l\}_{l=1}^k\right] = \sum_{j=1}^{n}\boldsymbol{u}_l(j)\cdot\mathbb{E}\left(\Big[(\mathbf{A}_2)_{ji} - (\mathbb{E}\mathbf{A}_2)_{ji}\Big]\right) = 0\,,$$

$$\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2^2 \Big| \{\boldsymbol{u}_l\}_{l=1}^k\right] = \sum_{l=1}^{k}\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle^2 \Big| \{\boldsymbol{u}_l\}_{l=1}^k\right],$$

where $\mathcal{M}$ is obtained from Algorithm 4.1. We expand each $\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle^2$ and rewrite it into 2 parts,

$$\begin{aligned}\langle \boldsymbol{u}_l, \boldsymbol{e}_i\rangle^2 &= \sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\boldsymbol{u}_l(j_1)\boldsymbol{e}_i(j_1)\boldsymbol{u}_l(j_2)\boldsymbol{e}_i(j_2)\\ &= \underbrace{\sum_{j=1}^{k}[\boldsymbol{u}_l(j)]^2[\boldsymbol{e}_i(j)]^2}_{(a)} + \underbrace{\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{e}_i(j_1)\boldsymbol{u}_l(j_2)\boldsymbol{e}_i(j_2)}_{(b)}\,, \quad l \in [k]\,.\end{aligned} \tag{B.16}$$

The expectation of part (a) in Equation (B.16) is upper bounded by α as defined in Equation (2.4), since

$$
\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\sum_{j=1}^{k}[\boldsymbol{u}_l(j)]^2[\boldsymbol{e}_i(j)]^2\middle|\{\boldsymbol{u}_l\}_{l=1}^k\right]\\
&=\sum_{j=1}^{n}[\boldsymbol{u}_l(j)]^2\cdot\mathbb{V}\mathrm{ar}\Big((\mathbf{A}_2)_{ji}\Big)\le\sum_{j=1}^{n}[\boldsymbol{u}_l(j)]^2\cdot(\mathbb{E}\mathbf{A}_2)_{ji}\le\alpha\,,\quad\forall l\in[k]\,,
\end{aligned}
$$

where $\|\boldsymbol{u}_l\|_2=1$ and $\mathbf{A}_2$ is defined in Equation (5.7). For part (b),

$$
\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{e}_i(j_1)\boldsymbol{u}_l(j_2)\boldsymbol{e}_i(j_2)\middle|\{\boldsymbol{u}_l\}_{l=1}^k\right]\\
&=\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\mathbb{E}\left[\Big((\mathbf{A}_2)_{j_1i}-(\mathbb{E}\mathbf{A}_2)_{j_1i}\Big)\Big((\mathbf{A}_2)_{j_2i}-(\mathbb{E}\mathbf{A}_2)_{j_2i}\Big)\right]\\
&=\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\mathbb{E}\left(\sum_{m\in\mathcal{M}}\sum_{\substack{e\in E_m[Y_2\cup Z]\\ \{i,j_1\}\subset e}}(\boldsymbol{\mathcal{T}}_e^{(m)}-\mathbb{E}\boldsymbol{\mathcal{T}}_e^{(m)})\right)\left(\sum_{m\in\mathcal{M}}\sum_{\substack{e\in E_m[Y_2\cup Z]\\ \{i,j_2\}\subset e}}(\boldsymbol{\mathcal{T}}_e^{(m)}-\mathbb{E}\boldsymbol{\mathcal{T}}_e^{(m)})\right).
\end{aligned}
$$

According to Definition 2.1 of the adjacency tensor, $\boldsymbol{\mathcal{T}}_{e_1}^{(m)}$ and $\boldsymbol{\mathcal{T}}_{e_2}^{(m)}$ are independent if hyperedge $e_1\neq e_2$, then only the terms with hyperedge $e\supset\{i,j_1,j_2\}$ have nonzero expectation. Then the expectation of part (b) can be rewritten as

$$
\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{e}_i(j_1)\boldsymbol{u}_l(j_2)\boldsymbol{e}_i(j_2)\middle|\{\boldsymbol{u}_l\}_{l=1}^k\right]\\
&=\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\sum_{m\in\mathcal{M}}\sum_{\substack{e\in E_m[Y_2\cup Z]\\ \{i,j_1,j_2\}\subset e}}\mathbb{E}\big(\boldsymbol{\mathcal{T}}_e^{(m)}-\mathbb{E}\boldsymbol{\mathcal{T}}_e^{(m)}\big)^2\\
&\le\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\sum_{m\in\mathcal{M}}\sum_{\substack{e\in E_m[Y_2\cup Z]\\ \{i,j_1,j_2\}\subset e}}\mathbb{E}\boldsymbol{\mathcal{T}}_e^{(m)}\\
&=\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\sum_{m\in\mathcal{M}}\sum_{\substack{e\in E_m[Y_2\cup Z]\\ \{i,j_1,j_2\}\subset e}}\frac{a_m}{\binom{n}{m-1}}\,.
\end{aligned}
\tag{B.19}
$$

Note that $|Y_2\cup Z|\le n$, then the number of possible hyperedges e, while $e\in E_m[Y_2\cup Z]$ and $e\supset\{i,j_1,j_2\}$, is at most $\binom{n}{m-3}$. Thus Equation (B.19) is upper bounded by

$$
\begin{aligned}
&\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\sum_{m\in\mathcal{M}}\binom{n}{m-3}\frac{a_m}{\binom{n}{m-1}}\\
&\le\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\sum_{m\in\mathcal{M}}\frac{(m-1)(m-2)}{(n-m)^2}a_m\le\frac{d\mathcal{M}_{\max}}{n^2}\sum_{j_1\neq j_2}\boldsymbol{u}_l(j_1)\boldsymbol{u}_l(j_2)\\
&\le\frac{d\mathcal{M}_{\max}}{2n^2}\sum_{j_1\neq j_2}\Big([\boldsymbol{u}_l(j_1)]^2+[\boldsymbol{u}_l(j_2)]^2\Big)\le\frac{d\mathcal{M}_{\max}(n-1)}{2n^2}\left(\sum_{j_1=1}^{n}[\boldsymbol{u}_l(j_1)]^2+\sum_{j_2=1}^{n}[\boldsymbol{u}_l(j_2)]^2\right)\le\frac{d\mathcal{M}_{\max}}{n}\,,
\end{aligned}
$$

where $\|\boldsymbol{u}_l\|_2=1$, $d=\sum_{m\in\mathcal{M}}(m-1)a_m$. With the upper bounds for part (a) and (b) in Equation (B.16), the conditional expectation of $\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2$ is bounded by

$$
\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2^2\middle|\{\boldsymbol{u}_l\}_{l=1}^k\right]=\sum_{l=1}^{k}\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\langle\boldsymbol{u}_l,\boldsymbol{e}_i\rangle^2\middle|\{\boldsymbol{u}_l\}_{l=1}^k\right]\le k\Big(\alpha+\frac{d\mathcal{M}_{\max}}{n}\Big)\,,
$$

By Markov's inequality,

$$\mathbb{P}\left(\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2 > 2\sqrt{k\Big(\alpha + \frac{d\mathcal{M}_{\max}}{n}\Big)}\right) \le \frac{\mathbb{E}_{\{\boldsymbol{\mathcal{T}}^{(m)}\}_{m\in\mathcal{M}}}\left[\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2^2 \Big| \{\boldsymbol{u}_l\}_{l=1}^k\right]}{4k(\alpha + d\mathcal{M}_{\max}/n)} \le \frac{1}{4}\,.$$

Let X_i be the Bernoulli random variable defined by

$$X_i = \mathbb{1}\{\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2 > 2\sqrt{k(\alpha + d\mathcal{M}_{\max}/n)}\}\,, \quad i \in \{i_1, \cdots, i_s\}\,.$$

Obviously, $\mathbb{E}X_i = \mathbb{P}(X_i = 1) \le 1/4$. Define

$$\delta := \frac{s}{2\sum_{j=1}^s \mathbb{E}X_{i_j}} - 1 \ge 1\,.$$

Since $s = 2k\log^2 n$, then by Chernoff's Lemma D.1,

$$\mathbb{P}\left(\sum_{j=1}^s X_{i_j} \ge \frac{s}{2}\right) = \mathbb{P}\left(\sum_{j=1}^s X_{i_j} \ge (1+\delta)\sum_{j=1}^s \mathbb{E}X_{i_j}\right) \le \exp\left(-c\delta^2\sum_{j=1}^s \mathbb{E}X_{i_j}\right) = O(e^{-k\log^2 n})\,,$$

for some constant c depending on k. Therefore, with probability $1 - O(n^{-k\log n})$, at least $s/2$ of the vectors $\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_s}$ satisfy $\|P_{\mathbf{U}}\boldsymbol{e}_i\|_2 \le 2\sqrt{k(\alpha + d\mathcal{M}_{\max}/n)}$ for each $i \in \{i_1, i_2, \cdots, i_s\}$. □

B.5. **Proof of Lemma 5.9.**

Proof. We start with the following simple claim: for any $m \ge 2$ and any $\nu \in [1/2, 1)$,

$$\nu^m + (1-\nu)^m < \left(\frac{1+\nu}{2}\right)^m\,. \tag{B.21}$$

Indeed, one quick way to see this is by induction on m; we will induct from m to $m+2$. Assume the inequality is true for m; then

$$\begin{aligned}
\nu^{m+2} + (1-\nu)^{m+2} &= \nu^2\nu^m + (1-\nu)^2(1-\nu)^m \\
&= \nu^2\nu^m + (1-2\nu+\nu^2)(1-\nu)^m \;\le\; \nu^2\nu^m + \nu^2(1-\nu)^m \\
&= \nu^2(\nu^m + (1-\nu)^m) \;<\; (\nu^2 + (1-\nu)^2)(\nu^m + (1-\nu)^m) \\
&< \left(\frac{1+\nu}{2}\right)^2\left(\frac{1+\nu}{2}\right)^m = \left(\frac{1+\nu}{2}\right)^{m+2}\,,
\end{aligned}$$

where we have used the induction hypothesis together with $1-2\nu \le 0$ and $(1-\nu)^2 > 0$. After easily checking that the inequality works for $m = 2, 3$, the induction is complete. We shall now check that the quantities defined in Lemma 5.9 obey the relationship $\mu_2 \ge \mu_1$ and $\mu_2 - \mu_1 = \Omega(n)$, for n large enough. First, note that the only thing we need to check is that, for sufficiently large n,

$$\binom{\frac{\nu n}{2k}}{m} + \binom{\frac{(1-\nu)n}{2k}}{m} \le \binom{\frac{(1+\nu)n}{4k}}{m} + (k-1)\binom{\frac{(1-\nu)n}{4k(k-1)}}{m}\,;$$

in fact, we will show the stronger statement that for any $m \ge 2$ and n large enough,

$$\binom{\frac{\nu n}{2k}}{m} + \binom{\frac{(1-\nu)n}{2k}}{m} < \binom{\frac{(1+\nu)n}{4k}}{m}\,, \tag{B.22}$$

and this will suffice to see that the second part of the assertion, $\mu_2 - \mu_1 = \Omega(n)$, is also true. Asymptotically, $\binom{\frac{\nu n}{2k}}{m} \sim \frac{\nu^m}{m!}\left(\frac{n}{2k}\right)^m$, $\binom{\frac{(1-\nu)n}{2k}}{m} \sim \frac{(1-\nu)^m}{m!}\left(\frac{n}{2k}\right)^m$, and $\binom{\frac{(1+\nu)n}{4k}}{m} \sim \frac{\left(\frac{1+\nu}{2}\right)^m}{m!}\left(\frac{n}{2k}\right)^m$. Note then that (B.22) follows from (B.21).

Let $\{V_1, \ldots, V_k\}$ be the true partition of V. Recall that hyperedges in $H = \cup_{m\in\mathcal{M}} H_m$ are colored red and blue with equal probability in Algorithm 1.1. Let $E_m(X)$ denote the set of blue m-uniform hyperedges with all vertices located in vertex set X. Assume $|X \cap V_i| = \eta_i|X|$ with $\sum_{i=1}^k \eta_i = 1$. For each $m \in \mathcal{M}$, the presence of hyperedge $e \in E_m(X)$ can be represented by independent Bernoulli random variables

$$T_e^{(a_m)} \sim \text{Bernoulli}\left(\frac{a_m}{2\binom{n}{m-1}}\right)\,, \quad T_e^{(b_m)} \sim \text{Bernoulli}\left(\frac{b_m}{2\binom{n}{m-1}}\right)\,,$$

depending on whether e is a hyperedge with all vertices in the same block. Denote by

$$E_m(X, a_m) := \cup_{i=1}^k E_m(V_i \cap X)$$

the union of all m-uniform sets of hyperedges with all vertices in the same $V_i \cap X$ for some $i \in [k]$, and by

$$E_m(X, b_m) := E_m(X) \setminus E_m(X, a_m) = E_m(X) \setminus \Big(\cup_i^k E_m(V_i \cap X) \Big)$$

the set of m-uniform hyperedges with vertices across different blocks $V_i \cap X$. Then the cardinality $|E_m(X)|$ can be written as the

$$|E_m(X)| = \sum_{e \in E_m(X, a_m)} T_e^{(a_m)} + \sum_{e \in E_m(X, b_m)} T_e^{(b_m)},$$

and by summing over m, the weighted cardinality $|E(X)|$ is written as

$$|E(X)| := \sum_{m \in \mathcal{M}} m(m-1)|E_m(X)| = \sum_{m \in \mathcal{M}} m(m-1) \left\{ \sum_{e \in E_m(X, a_m)} T_e^{(a_m)} + \sum_{e \in E_m(X, b_m)} T_e^{(b_m)} \right\},$$

with its expectation

$$\mathbb{E}|E(X)| = \sum_{m \in \mathcal{M}} m(m-1) \left\{ \sum_{i=1}^k \binom{\eta_i \frac{n}{2k}}{m} \frac{a_m - b_m}{2\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m} \frac{b_m}{2\binom{n}{m-1}} \right\}, \tag{B.23}$$

since

$$|E_m(X, a_m)| = \sum_{i=1}^k |E_m(V_i \cap X)| = \sum_{i=1}^k \binom{\eta_i \frac{n}{2k}}{m}, \quad |E_m(X, b_m)| = \binom{\frac{n}{2k}}{m} - \sum_{i=1}^k \binom{\eta_i \frac{n}{2k}}{m},$$

Next, we prove the two statements in Lemma 5.9 separately. First, assume that $|X \cap V_i| \leq \nu|X|$ (i.e., $\eta_i \leq \nu$) for all $i \in [k]$. Then

$$\mathbb{E}|E(X)| \leq \frac{1}{2} \sum_{m \in \mathcal{M}} m(m-1) \left\{ \left[\binom{\frac{\nu n}{2k}}{m} + \binom{\frac{(1-\nu)n}{2k}}{m} \right] \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m} \frac{b_m}{\binom{n}{m-1}} \right\} =: \mu_1 .$$

To justify the above inequality, note that since $\sum\limits_{i=1}^{k} \eta_i = 1$, the sum $\sum_{i=1}^k \binom{\eta_i \frac{n}{2k}}{m}$ is maximized when all but 2 of the η_i are 0, and since all $\eta_i \leq \nu$, this means that

$$\sum_{i=1}^k \binom{\eta_i \frac{n}{2k}}{m} \leq \binom{\frac{\nu n}{2k}}{m} + \binom{\frac{(1-\nu)n}{2k}}{m}.$$

Note that $m(m-1)(T_e^{(a_m)} - \mathbb{E}T_e^{(a_m)})$ and $m(m-1)(T_e^{(b_m)} - \mathbb{E}T_e^{(b_m)})$ are independent mean-zero random variables bounded by $M(M-1)$ for all $m \in \mathcal{M}$, and $\mathbb{V}\mathrm{ar}(|E(X)|) \leq M^2(M-1)^2\mathbb{E}|E(X)| = \Omega(n)$. Recall that $\mu_{\mathrm{T}} := (\mu_1 + \mu_2)/2$. Define $t = \mu_{\mathrm{T}} - \mathbb{E}|E(X)|$, then $0 < (\mu_2 - \mu_1)/2 \leq t \leq \mu_{\mathrm{T}}$, hence $t = \Omega(n)$. By Bernstein's Lemma D.3, we have

$$\mathbb{P}\Big(|E(X)| \geq \mu_{\mathrm{T}}\Big) = \mathbb{P}\Big(|E(X)| - \mathbb{E}|E(X)| \geq t\Big) \leq \exp\left(-\tfrac{t^2/2}{\mathbb{V}\mathrm{ar}(|E(X)|) + M(M-1)t/3}\right) = O(e^{-cn}),$$

where $c > 0$ is some constant. On the other hand, if $|X \cap V_i| \geq \frac{1+\nu}{2}|X|$ for some $i \in [k]$, then

$$\mathbb{E}|E(X)| \geq \frac{1}{2} \sum_{m \in \mathcal{M}} m(m-1) \left\{ \left[\binom{\frac{(1+\nu)n}{4k}}{m} + (k-1)\binom{\frac{(1-\nu)n}{4k(k-1)}}{m} \right] \frac{a_m - b_m}{\binom{n}{m-1}} + \binom{\frac{n}{2k}}{m} \frac{b_m}{\binom{n}{m-1}} \right\} =: \mu_2 .$$

The above can be justified by noting that at least one $|X \cap V_i| \geq \frac{1+\nu}{2}|X|$, and that the rest of the vertices will yield a minimal binomial sum when they are evenly split between the remaining V_j. Similarly, define $t = \mu_{\mathrm{T}} - \mathbb{E}|E(X)|$, then $0 < (\mu_2 - \mu_1)/2 \leq -t = \Omega(n)$, and Bernstein's Lemma D.3 gives

$$\mathbb{P}\Big(|E(X)| \leq \mu_{\mathrm{T}}\Big) = \mathbb{P}\Big(|E(X)| - \mathbb{E}|E(X)| \leq -t\Big) \leq \exp\left(-\frac{t^2/2}{\mathbb{V}\mathrm{ar}(|E(X)|) + M(M-1)(-t)/3}\right) = O(e^{-c'n}),$$

where $c' > 0$ is some other constant. □

B.6. **Proof of Lemma 5.11.**

Proof. If vertex i is uniformly chosen from Y_2, the probability that $i \notin V_l$ for some $l \in [k]$ is

$$\mathbb{P}(i \notin V_l | i \in Y_2) = \frac{\mathbb{P}(i \notin V_l, i \in Y_2)}{\mathbb{P}(i \in Y_2)} = 1 - \frac{|V_l \cap Y_2|}{|Y_2|} = 1 - \frac{\frac{n}{k} - n_l - n_l'}{n - \sum_{t=1}^{k}(n_t + n_t')}\,, \quad l \in [k]\,,$$

where n_t and n_t', defined in Equation (5.1) and Equation (5.2), denote the cardinality of $Z \cap V_t$ and $Y_1 \cap V_t$ respectively. As proved in Appendix B.2, with probability at least $1 - 2\exp(-k\log^2(n))$, we have $|n_t - n/(2k)| \le \sqrt{n}\log(n)$ and $|n_t' - n/(4k)| \le \sqrt{n}\log(n)$, then $\mathbb{P}(i \notin V_l | i \in Y_2) = 1 - \frac{1}{k}\Big(1 + o(1)\Big)$. After $k\log^2 n$ samples from Y_2, the probability that there exists at least one node which belongs to V_l is at least

$$1 - \left(1 - \frac{1 + o(1)}{k}\right)^{k\log^2 n} = 1 - n^{-(1+o(1))k\log(\frac{k}{k-1})\log n}.$$

The proof is completed by a union bound over $l \in [k]$.

□

B.7. **Proof of Lemma 5.12.**

Proof. We calculate $\mathbb{P}(S_{11}'(u) \le \mu_{\mathrm{C}})$ first. Define $t_{1\mathrm{C}} := \mu_{\mathrm{C}} - \mathbb{E}S_{11}'(u)$, then by Bernstein's inequality (Lemma D.3) and taking $K = \mathcal{M}_{\max} - 1$,

$$\begin{aligned}
&\mathbb{P}\left(S_{11}'(u) \le \mu_{\mathrm{C}}\right) = \mathbb{P}\left(S_{11}'(u) - \mathbb{E}S_{11}'(u) \le t_{1\mathrm{C}}\right)\\
&\le \exp\left(-\frac{t_{1\mathrm{C}}^2/2}{\mathbb{V}\mathrm{ar}[S_{11}'(u)] + (\mathcal{M}_{\max} - 1)\cdot t_{1\mathrm{C}}/3}\right) \le \exp\left(-\frac{3t_{1\mathrm{C}}^2/(\mathcal{M}_{\max} - 1)}{6(\mathcal{M}_{\max} - 1)\cdot \mathbb{E}S_{11}'(u) + 2t_{1\mathrm{C}}}\right)\\
&\le \exp\left(-\frac{[(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}]^2}{(\mathcal{M}_{\max} - 1)^2 \cdot 2^{2\mathcal{M}_{\max}+3}} \cdot \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}} + b_m\right)}\right),
\end{aligned}$$

where $\mathcal{M}$ is obtained from Algorithm 4.1 with $\mathcal{M}_{\max}$ denoting the maximum value in $\mathcal{M}$, and the last two inequalities hold since $\mathbb{V}\mathrm{ar}[S_{11}'(u)] \le (\mathcal{M}_{\max} - 1)^2\mathbb{E}S_{11}'(u)$, and for sufficiently large n,

$$\begin{aligned}
t_{1\mathrm{C}} :=& \mu_{\mathrm{C}} - \mathbb{E}S_{11}'(u) = -\frac{1}{2}\sum_{m\in\mathcal{M}}(m-1)\cdot\left[\binom{\frac{\nu n}{2k}}{m-1} - \binom{\frac{(1-\nu)n}{2k}}{m-1}\right]\frac{a_m - b_m}{2\binom{n}{m-1}}\\
&\le -\frac{1}{2}\sum_{m\in\mathcal{M}}\frac{(\nu)^{m-1} - (1-\nu)^{m-1}}{2^m}\cdot(m-1)\frac{a_m - b_m}{k^{m-1}}(1 + o(1))\\
&\le -\frac{(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}}{2^{\mathcal{M}_{\max}+2}}\sum_{m\in\mathcal{M}}(m-1)\cdot\frac{a_m - b_m}{k^{m-1}}\,,
\end{aligned}$$

$$\begin{aligned}
&6(\mathcal{M}_{\max} - 1)\mathbb{E}S_{11}'(u) + 2t_{1\mathrm{C}} = 2\mu_{\mathrm{C}} + (6\mathcal{M}_{\max} - 8)\mathbb{E}S_{11}'(u)\\
=& \sum_{m\in\mathcal{M}}(m-1)\left\{\left[(6\mathcal{M}_{\max} - 7)\binom{\frac{\nu n}{2k}}{m-1} + \binom{\frac{(1-\nu)n}{2k}}{m-1}\right]\frac{a_m - b_m}{2\binom{n}{m-1}} + 6(\mathcal{M}_{\max} - 1)\binom{\frac{n}{2k}}{m-1}\frac{b_m}{2\binom{n}{m-1}}\right\}\\
=& \sum_{m\in\mathcal{M}}(m-1)\left[\frac{(6\mathcal{M}_{\max} - 7)\cdot(\nu)^{m-1} + (1-\nu)^{m-1}}{2^m}\cdot\frac{a_m - b_m}{k^{m-1}} + \frac{(6\mathcal{M}_{\max} - 6)b_m}{2^m k^{m-1}}\right](1 + o(1))\\
\le& \sum_{m\in\mathcal{M}}\frac{(6\mathcal{M}_{\max} - 7)\cdot(\nu)^{m-1} + (1-\nu)^{m-1}}{2^m}\cdot(m-1)\left(\frac{a_m - b_m}{k^{m-1}} + b_m\right)(1 + o(1))\\
\le& \frac{3(\mathcal{M}_{\max} - 1)}{2}\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m - b_m}{k^{m-1}} + b_m\right).
\end{aligned}$$

Similarly, for $\mathbb{P}(S'_{1j}(u) \geq \mu_{\rm C})$, define $t_{\rm jC} := \mu_{\rm C} - \mathbb{E}S'_{1j}(u)$ for $j \neq 1$, by Bernstein's Lemma D.3,

$$
\begin{aligned}
&\mathbb{P}\left(S'_{1j}(u) \geq \mu_{\rm C}\right) = \mathbb{P}\left(S'_{1j}(u) - \mathbb{E}S'_{1j}(u) \geq t_{\rm jC}\right)\\
\leq{}& \exp\left(-\frac{t_{\rm jC}^2/2}{\mathbb{V}\mathrm{ar}[S'_{1j}(u)] + (\mathcal{M}_{\max}-1)\cdot t_{\rm jC}/3}\right) \leq \exp\left(-\frac{3t_{\rm jC}^2/(\mathcal{M}_{\max}-1)}{6(\mathcal{M}_{\max}-1)\cdot \mathbb{E}S'_{1j}(u) + 2t_{\rm jC}}\right)\\
\leq{}& \exp\left(-\frac{[(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}]^2}{(\mathcal{M}_{\max}-1)^2\cdot 2^{2\mathcal{M}_{\max}+3}} \cdot \frac{\left[\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}}\right)\right]^2}{\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}} + b_m\right)}\right).
\end{aligned}
$$

The last two inequalities holds since $\mathbb{V}\mathrm{ar}[S'_{1j}(u)] \leq (\mathcal{M}_{\max}-1)^2\mathbb{E}S'_{1j}(u)$, and for sufficiently large n,

$$
\begin{aligned}
t_{\rm jC} := \mu_{\rm C} - \mathbb{E}S'_{1j}(u) &= \frac{1}{2}\sum_{m\in\mathcal{M}}(m-1)\cdot\left[\binom{\frac{\nu n}{2k}}{m-1} - \binom{\frac{(1-\nu)n}{2k}}{m-1}\right]\frac{a_m-b_m}{2\binom{n}{m-1}}\\
&\geq \frac{(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}}{2^{\mathcal{M}_{\max}+2}}\sum_{m\in\mathcal{M}}(m-1)\cdot\frac{a_m-b_m}{k^{m-1}}\,,
\end{aligned}
$$

$$
\begin{aligned}
&6\mathbb{E}S'_{1j}(u) + 2t_{\rm jC} = 2\mu_{\rm C} + (6\mathcal{M}_{\max}-8)\mathbb{E}S'_{1j}(u)\\
={}& \sum_{m\in\mathcal{M}}(m-1)\left\{\left[\binom{\frac{\nu n}{2k}}{m-1} + (6\mathcal{M}_{\max}-7)\binom{\frac{(1-\nu)n}{2k}}{m-1}\right]\frac{a_m-b_m}{2\binom{n}{m-1}} + 6(\mathcal{M}_{\max}-1)\binom{\frac{n}{2k}}{m-1}\frac{b_m}{2\binom{n}{m-1}}\right\}\\
={}& \sum_{m\in\mathcal{M}}(m-1)\cdot\left(\frac{(\nu)^{m-1} + (6\mathcal{M}_{\max}-7)\cdot(1-\nu)^{m-1}}{2^m}\cdot\frac{a_m-b_m}{k^{m-1}} + \frac{(6\mathcal{M}_{\max}-6)\cdot b_m}{2^m k^{m-1}}\right)(1+o(1))\\
\leq{}& \sum_{m\in\mathcal{M}}\frac{(\nu)^{m-1} + (6\mathcal{M}_{\max}-7)\cdot(1-\nu)^{m-1}}{2^m}\cdot(m-1)\left(\frac{a_m-b_m}{k^{m-1}} + b_m\right)(1+o(1))\\
\leq{}& \frac{3(\mathcal{M}_{\max}-1)}{2}\sum_{m\in\mathcal{M}}(m-1)\left(\frac{a_m-b_m}{k^{m-1}} + b_m\right).
\end{aligned}
$$

□

Appendix C. Algorithm's correctness for the binary case

We will show the correctness of Algorithm 1.2 and prove Theorem 1.6 in this section. The analysis will mainly follow from the analysis in Section 5. We only detail the differences.

Without loss of generality, we assume n is even to guarantee the existence of a binary partition of size $n/2$. The method to deal with the odd n case was discussed in Lemma 2.4. Then, let the index set be $\mathcal{I} = \{i \in [n] : \mathrm{row}(i) \leq 20\mathcal{M}_{\max}d\}$, as shown in Equation (3.3). Let $\boldsymbol{u}_i$ (resp. $\bar{\boldsymbol{u}}_i$) denote the eigenvector associated to $\lambda_i(\mathbf{A}_{\mathcal{I}})$ (resp. $\lambda_i(\overline{\mathbf{A}})$) for $i = 1, 2$. Define two linear subspaces $\mathbf{U} := \mathrm{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ and $\overline{\mathbf{U}} := \mathrm{Span}\{\bar{\boldsymbol{u}}_1, \bar{\boldsymbol{u}}_2\}$, then the angle between $\mathbf{U}$ and $\overline{\mathbf{U}}$ is defined as $\sin\angle(\mathbf{U}, \overline{\mathbf{U}}) := \|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\|$, where $P_{\mathbf{U}}$ and $P_{\overline{\mathbf{U}}}$ are the orthogonal projections onto $\mathbf{U}$ and $\overline{\mathbf{U}}$, respectively.

C.1. **Proof of Lemma 4.4.**

C.1.1. *Bound the angle between* $\mathbf{U}$ *and* $\overline{\mathbf{U}}$. The strategy to bound the angle is similar to Section 5.1.2, except that we apply Davis-Kahan Theorem (Lemma D.6) here.

Define $\mathbf{E} := \mathbf{A} - \overline{\mathbf{A}}$ and its restriction on $\mathcal{I}$, namely $\mathbf{E}_{\mathcal{I}} := (\mathbf{A} - \overline{\mathbf{A}})_{\mathcal{I}} = \mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}_{\mathcal{I}}$, as well as $\boldsymbol{\Delta} := \overline{\mathbf{A}}_{\mathcal{I}} - \overline{\mathbf{A}}$. Then the deviation $\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}$ is decomposed as

$$
\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}} = (\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}_{\mathcal{I}}) + (\overline{\mathbf{A}}_{\mathcal{I}} - \overline{\mathbf{A}}) = \mathbf{E}_{\mathcal{I}} + \boldsymbol{\Delta}\,.
$$

Theorem 3.3 indicates $\|\mathbf{E}_{\mathcal{I}}\| \leq C_3\sqrt{d}$ with probability at least $1 - n^{-2}$ when taking $\tau = 20\mathcal{M}_{\max}, K = 3$, where C_3 is a constant depending only on $\mathcal{M}_{\max}$. Moreover, Lemma 5.4 shows that the number of vertices with high degrees is relatively small. Consequently, an argument similar to Corollary 5.5 leads to the conclusion $\|\boldsymbol{\Delta}\| \leq \sqrt{d}$ *w.h.p.* Together with upper bounds for $\|\mathbf{E}_{\mathcal{I}}\|$ and $\|\boldsymbol{\Delta}\|$, Lemma C.1 shows that the angle between $\mathbf{U}$ and $\overline{\mathbf{U}}$ is relatively small *w.h.p.*

Lemma C.1. *For any $c \in (0,1)$, there exists a constant C_2 depending on $\mathcal{M}_{\max}$ and c such that if*

$$\sum_{m\in\mathcal{M}} (m-1)(a_m - b_m) \geq C_2 \cdot 2^{\mathcal{M}_{\max}+2}\sqrt{d},$$

then $\sin\angle(\mathbf{U}, \overline{\mathbf{U}}) \leq c$ with probability $1 - n^{-2}$.

Proof. First, with probability $1 - n^{-2}$, we have

$$\|\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}\| \leq \|\mathbf{E}_{\mathcal{I}}\| + \|\mathbf{\Delta}\| \leq (C_3+1)\sqrt{d}.$$

According to the definitions in Equation (2.4), $\alpha \geq \beta$ and $\alpha = O(1/n)$, $\beta = O(1/n)$. Meanwhile, Lemma 2.3 shows that $|\lambda_2(\overline{\mathbf{A}})| = [-\alpha + (\alpha-\beta)n/2]$ and $|\lambda_3(\overline{\mathbf{A}})| = \alpha$. Then

$$|\lambda_2(\overline{\mathbf{A}})| - |\lambda_3(\overline{\mathbf{A}})| = \frac{n}{2}(\alpha-\beta) - 2\alpha \geq \frac{3}{4}\cdot\frac{n}{2}(\alpha-\beta) = \frac{3n}{8}\sum_{m\in\mathcal{M}} \binom{\frac{n}{2}-2}{m-2}\frac{(a_m-b_m)}{\binom{n}{m-1}}$$
$$\geq \frac{1}{4}\sum_{m\in\mathcal{M}} \frac{(m-1)(a_m-b_m)}{2^{m-2}} \geq \frac{1}{2^{\mathcal{M}_{\max}}}\sum_{m\in\mathcal{M}}(m-1)(a_m-b_m) \geq 4C_2\sqrt{d}\,.$$

Then for some large enough C_2, the following condition for Davis-Kahan Theorem (Lemma D.6) is satisfied

$$\|\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}\| \leq (1-1/\sqrt{2})\left(|\lambda_2(\overline{\mathbf{A}})| - |\lambda_3(\overline{\mathbf{A}})|\right).$$

Then for any $c \in (0,1)$, we can choose $C_2 = (C_3+1)/c$ such that

$$\|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\| \leq \frac{2\|\mathbf{A}_{\mathcal{I}} - \overline{\mathbf{A}}\|}{|\lambda_2(\overline{\mathbf{A}})| - |\lambda_3(\overline{\mathbf{A}})|} \leq \frac{2(C_3+1)\sqrt{d}}{4C_2\sqrt{d}} = \frac{c}{2} \leq c\,.$$

□

Now, we focus on the accuracy of Algorithm 4.5, once the conditions in Lemma C.1 are satisfied.

Lemma C.2 (Lemma 23 in [16]). *If $\sin\angle(\overline{\mathbf{U}}, \mathbf{U}) \leq c \leq \frac{1}{4}$, there exists a unit vector $\boldsymbol{v} \in \mathbf{U}$ such that $\sin\angle(\bar{\boldsymbol{u}}_2, \boldsymbol{v}) \leq 2\sqrt{c}$.*

The desired vector $\boldsymbol{v}$, as constructed in Algorithm 4.5, is the unit vector perpendicular to $P_{\mathbf{U}}\mathbf{1}_n$, where $P_{\mathbf{U}}\mathbf{1}_n$ is the projection of all-ones vector onto $\mathbf{U}$. Lemma C.1 and Lemma C.2 together give the following corollary.

Corollary C.3. *For any $c \in (0,1)$, there exists a unit vector $\boldsymbol{v} \in \mathbf{U}$ such that $\sin\angle(\bar{\boldsymbol{u}}_2, \boldsymbol{v}) \leq c < 1$ with probability $1 - O(e^{-n})$.*

Lemma C.4 (Lemma 23 in [16]). *If $\sin\angle(\bar{\boldsymbol{u}}_2, \boldsymbol{v}) < c \leq 0.5$, then we can identify at least $(1 - 4c^2/3)n$ vertices from each block correctly.*

The proof of Lemma 4.4 is completed when choosing C_2, C_3 in Lemma C.1 such that $c \leq 1/4$.

C.2. **Proof of Lemma 4.5.** The proof strategy is similar to Section 5.2 and Section 5.3. In Algorithm 1.2, we first color the hyperedges with red and blue with equal probability. By running Algorithm 4.6 on the red graph, we obtain a ν-correct partition V_1', V_2' of $V = V_1 \cup V_2$, i.e., $|V_l \cap V_l'| \geq \nu n/2$ for $l = 1, 2$. In the rest of the proof, we condition on this event and the event that the maximum red degree of a vertex is at most $\log^2(n)$ with probability at least $1 - o(1)$. This can be proved by Bernstein's inequality (Lemma D.3).

Similarly, we consider the probability of a hyperedge $e = \{i_1, \cdots, i_m\}$ being blue conditioning on the event that e is not a red hyperedge, in each underlying m-uniform hypergraph separately. If vertices $i_1, \cdots, i_m$ are all from the same true cluster, then the probability is ψ_m, otherwise ϕ_m, where ψ_m and ϕ_m are defined in Equation (5.25) and Equation (5.26), and the presence of those hyperedges are represented by random variables $\zeta_e^{(a_m)} \sim \text{Bernoulli}(\psi_m)$, $\xi_e^{(b_m)} \sim \text{Bernoulli}(\phi_m)$, respectively.

Following a similar argument in Section 5.2, the row sum of u can be written as

$$S_{lj}'(u) := \sum_{m\in\mathcal{M}} (m-1)\cdot\left\{\sum_{e\in\,\mathcal{E}_{l,j}^{(a_m)}} \zeta_e^{(a_m)} + \sum_{e\in\mathcal{E}_{l,j}^{(b_m)}} \xi_e^{(b_m)}\right\}, \quad u \in V_l\,,$$

where $\mathcal{E}_{l,j}^{(a_m)} := E_m([V_l]^1, [V_l \cap V_j']^{m-1})$ denotes the set of m-hyperedges with 1 vertex from $[V_l]^1$ and the other $m-1$ vertices from $[V_l \cap V_j']^{m-1}$, while

$$\mathcal{E}_{l,j}^{(b_m)} := E_m\Big([V_l]^1,\ [V_j']^{m-1} \setminus [V_l \cap V_j']^{m-1}\Big)$$

denotes the set of m-hyperedges with 1 vertex in $[V_l]^1$ while the remaining $m-1$ vertices in $[V_j']^{m-1} \setminus [V_l \cap V_j']^{m-1}$, with their cardinalities

$$|\mathcal{E}_{l,j}^{(a_m)}| \leq \binom{|V_l \cap V_j'|}{m-1}, \quad |\mathcal{E}_{l,j}^{(b_m)}| \leq \left[\binom{|V_j'|}{m-1} - \binom{|V_l \cap V_j'|}{m-1}\right].$$

According to the fact $|V_l \cap V_l'| \geq \nu n/2$, $|V_l| = n/2$, $|V_l'| = n/2$ for $l = 1, 2$, we have

$$|\mathcal{E}_{l,l}^{(a_m)}| \geq \binom{\frac{\nu n}{2}}{m-1}, \quad |\mathcal{E}_{l,j}^{(a_m)}| \leq \binom{\frac{(1-\nu)n}{2}}{m-1},\ j \neq l\,.$$

To simplify the calculation, we take the lower and upper bound of $|\mathcal{E}_{l,l}^{(a_m)}|$ and $|\mathcal{E}_{l,j}^{(a_m)}|(j \neq l)$ respectively. Taking expectation with respect to $\zeta_e^{(a_m)}$ and $\xi_e^{(b_m)}$, for any $u \in V_l$, we have

$$\mathbb{E}S_{ll}'(u) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{\nu n}{2}}{m-1}(\psi_m - \phi_m) + \binom{\frac{n}{2}}{m-1}\phi_m\right],$$

$$\mathbb{E}S_{lj}'(u) = \sum_{m \in \mathcal{M}} (m-1) \cdot \left[\binom{\frac{(1-\nu)n}{2}}{m-1}(\psi_m - \phi_m) + \binom{\frac{n}{2}}{m-1}\phi_m\right],\ j \neq l\,.$$

By assumptions in Theorem 1.4, $\mathbb{E}S_{ll}'(u) - \mathbb{E}S_{lj}'(u) = \Omega(1)$. We define

$$\mu_{\mathrm{C}} := \frac{1}{2}\sum_{m \in \mathcal{M}} (m-1) \cdot \left\{\left[\binom{\frac{\nu n}{2}}{m-1} + \binom{\frac{(1-\nu)n}{2}}{m-1}\right](\psi_m - \phi_m) + 2\binom{\frac{n}{2}}{m-1}\phi_m\right\}.$$

After Algorithm 4.4, if a vertex $u \in V_l$ is mislabelled, one of the following event must happen

- $S_{ll}'(u) \leq \mu_{\mathrm{C}}$,
- $S_{lj}'(u) \geq \mu_{\mathrm{C}}$, for some $j \neq l$.

By an argument similar to Lemma 5.12, we can prove that

$$\rho_1' = \mathbb{P}\left(S_{ll}'(u) \leq \mu_{\mathrm{C}}\right) \leq \rho\,, \quad \rho_2' = \mathbb{P}\left(S_{lj}'(u) \geq \mu_{\mathrm{C}}\right) \leq \rho\,,$$

where $\rho = \exp\left(-C_{\mathcal{M}} \cdot \mathrm{SNR}_{\mathcal{M}}\right)$ and

$$C_{\mathcal{M}} := \frac{[(\nu)^{\mathcal{M}_{\max}-1} - (1-\nu)^{\mathcal{M}_{\max}-1}]^2}{8(\mathcal{M}_{\max}-1)^2}, \quad \mathrm{SNR}_{\mathcal{M}} := \frac{\left[\sum_{m \in \mathcal{M}}(m-1)\left(\frac{a_m - b_m}{2^{m-1}}\right)\right]^2}{\sum_{m \in \mathcal{M}}(m-1)\left(\frac{a_m - b_m}{2^{m-1}} + b_m\right)}.$$

As a result, the probability that either of those events happened is bounded by ρ. The number of mislabeled vertices in V_1 after Algorithm 4.3 is at most

$$R_l = \sum_{i=1}^{|V_l \setminus V_l'|} \Gamma_i + \sum_{i=1}^{|V_l \cap V_l'|} \Lambda_i\,,$$

where Γ_i (resp. Λ_i) are i.i.d indicator random variables with mean ρ_1' (resp. ρ_2'). Then

$$\mathbb{E}R_l \leq \frac{n}{2}\rho_1' + \frac{(1-\nu)n}{2}\rho_2' = (1 - \nu/2)n\rho\,.$$

where ν is the correctness after Algorithm 4.2. Let $t_l := (1 + \nu/2)n\rho$, then by Chernoff Lemma D.1,

$$\mathbb{P}\left(R_l \geq n\rho\right) = \mathbb{P}\left[R_l - (1-\nu/2)n\rho \geq t_l\right] \leq \mathbb{P}\left(R_l - \mathbb{E}R_l \geq t_l\right) \leq e^{-ct_l} = O(e^{-n\rho})\,,$$

which means that with probability $1 - O(e^{-n\rho})$, the fraction of mislabeled vertices in V_l is smaller than 2ρ, i.e., the correctness of V_l is at least $\gamma := \max\{\nu, 1 - 2\rho\}$.

Appendix D. Useful Lemmas

Lemma D.1 (Chernoff's inequality, Theorem 2.3.6 in [54])**.** *Let X_i be independent Bernoulli random variables with parameters p_i. Consider their sum $S_N = \sum_{i=1}^N X_i$ and denote its mean by $\mu = \mathbb{E}S_N$. Then for any $\delta \in (0,1]$,*

$$\mathbb{P}\big(|S_N - \mu| \geq \delta\mu\big) \leq 2\exp(-c\delta^2\mu)\,.$$

Lemma D.2 (Hoeffding's inequality, Theorem 2.2.6 in [54])**.** *Let $X_1, \dots, X_N$ be independent random variables with $X_i \in [a_i, b_i]$ for each $i \in \{1, \dots, N\}$. Then for any, $t \geq 0$, we have*

$$\mathbb{P}\left(\Big|\sum_{i=1}^N (X_i - \mathbb{E}X_i)\Big| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\right).$$

Lemma D.3 (Bernstein's inequality, Theorem 2.8.4 in [54])**.** *Let $X_1, \dots, X_N$ be independent mean-zero random variables such that $|X_i| \leq K$ for all i. Let $\sigma^2 = \sum_{i=1}^N \mathbb{E}X_i^2$. Then for every $t \geq 0$, we have*

$$\mathbb{P}\left(\Big|\sum_{i=1}^N X_i\Big| \geq t\right) \leq 2\exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right).$$

Lemma D.4 (Bennett's inequality, Theorem 2.9.2 in [54])**.** *Let $X_1, \dots, X_N$ be independent random variables. Assume that $|X_i - \mathbb{E}X_i| \leq K$ almost surely for every i. Then for any $t > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^N (X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{\sigma^2}{K^2}\cdot h\left(\frac{Kt}{\sigma^2}\right)\right),$$

where $\sigma^2 = \sum_{i=1}^N \mathrm{Var}(X_i)$, and $h(u) := (1+u)\log(1+u) - u$.

Lemma D.5 (Weyl's inequality)**.** *Let $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{m\times n}$ be two real $m \times n$ matrices, then $|\sigma_i(\mathbf{A}+\mathbf{E}) - \sigma_i(\mathbf{A})| \leq \|\mathbf{E}\|$ for every $1 \leq i \leq \min\{m,n\}$. Furthermore, if $m = n$ and $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n\times n}$ are real symmetric, then $|\lambda_i(\mathbf{A}+\mathbf{E}) - \lambda_i(\mathbf{A})| \leq \|\mathbf{E}\|$ for all $1 \leq i \leq n$.*

Lemma D.6 (Davis-Kahan's $\sin\boldsymbol{\Theta}$ Theorem, Theorem 2.2.1 in [13])**.** *Let $\overline{\mathbf{M}}$ and $\mathbf{M} = \overline{\mathbf{M}} + \mathbf{E}$ be two real symmetric $n \times n$ matrices, with eigenvalue decompositions given respectively by*

$$\overline{\mathbf{M}} = \sum_{i=1}^n \overline{\lambda}_i \overline{\boldsymbol{u}}_i \overline{\boldsymbol{u}}_i^{\mathsf{T}} = \begin{bmatrix}\overline{\mathbf{U}} & \overline{\mathbf{U}}_\perp\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{\Lambda}} & \mathbf{0}\\ \mathbf{0} & \overline{\boldsymbol{\Lambda}}_\perp\end{bmatrix}\begin{bmatrix}\overline{\mathbf{U}}^{\mathsf{T}}\\ \overline{\mathbf{U}}_\perp^{\mathsf{T}}\end{bmatrix},$$
$$\mathbf{M} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathsf{T}} = \begin{bmatrix}\mathbf{U} & \mathbf{U}_\perp\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Lambda}_\perp\end{bmatrix}\begin{bmatrix}\mathbf{U}^{\mathsf{T}}\\ \mathbf{U}_\perp^{\mathsf{T}}\end{bmatrix}.$$

Here, $\{\overline{\lambda}_i\}_{i=1}^n$(resp. $\{\lambda_i\}_{i=1}^n$) stand for the eigenvalues of $\overline{\mathbf{M}}$(resp. $\mathbf{M}$), and $\overline{\boldsymbol{u}}_i$(resp. $\boldsymbol{u}_i$) denotes the eigenvector associated $\overline{\lambda}_i$(resp. λ_i). Additionally, for some fixed integer $r \in [n]$, we denote

$$\overline{\boldsymbol{\Lambda}} := \mathrm{diag}\{\overline{\lambda}_1, \dots, \overline{\lambda}_r\}, \quad \overline{\boldsymbol{\Lambda}}_\perp := \mathrm{diag}\{\overline{\lambda}_{r+1}, \dots, \overline{\lambda}_n\},$$
$$\overline{\mathbf{U}} := [\overline{\boldsymbol{u}}_1, \dots, \overline{\boldsymbol{u}}_r] \in \mathbb{R}^{n\times r}, \quad \overline{\mathbf{U}}_\perp := [\overline{\boldsymbol{u}}_{r+1}, \dots, \overline{\boldsymbol{u}}_n] \in \mathbb{R}^{n\times(n-r)}.$$

The matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}_\perp$, $\mathbf{U}$, $\mathbf{U}_\perp$ are defined analogously. Assume that

$$\text{eigenvalues}(\overline{\boldsymbol{\Lambda}}) \subseteq [\alpha, \beta]\,, \quad \text{eigenvalues}(\boldsymbol{\Lambda}_\perp) \subseteq (-\infty, \alpha - \Delta] \cup [\beta + \Delta, \infty), \quad \alpha, \beta \in \mathbb{R}\,, \Delta > 0\,,$$

and the projection matrices are given by $P_{\mathbf{U}} := \mathbf{U}\mathbf{U}^{\mathsf{T}}$, $P_{\overline{\mathbf{U}}} := \overline{\mathbf{U}}\,\overline{\mathbf{U}}^{\mathsf{T}}$, then one has $\|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\| \leq (2\|\mathbf{E}\|/\Delta)$. In particular, suppose that $|\overline{\lambda}_1| \geq |\overline{\lambda}_2| \geq \cdots \geq |\overline{\lambda}_r| \geq |\overline{\lambda}_{r+1}| \geq \cdots |\overline{\lambda}_n|$ (resp. $|\lambda_1| \geq \cdots \geq |\lambda_n|$). If $\|\mathbf{E}\| \leq (1 - 1/\sqrt{2})(|\overline{\lambda}|_r - |\overline{\lambda}|_{r+1})$, then one has

$$\|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\| \leq \frac{2\|\mathbf{E}\|}{|\overline{\lambda}_r| - |\overline{\lambda}_{r+1}|}\,.$$

Lemma D.7 (Wedin's $\sin\mathbf{\Theta}$ Theorem, Theorem 2.3.1 in [13]). *Let $\overline{\mathbf{M}}$ and $\mathbf{M} = \overline{\mathbf{M}} + \mathbf{E}$ be two $n_1 \times n_2$ real matrices and $n_1 \geq n_2$, with SVDs given respectively by*

$$\overline{\mathbf{M}} = \sum_{i=1}^{n_1} \overline{\sigma}_i \overline{\boldsymbol{u}}_i \overline{\boldsymbol{v}}_i^{\mathsf{T}} = \begin{bmatrix} \overline{\mathbf{U}} & \overline{\mathbf{U}}_{\perp} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{\Sigma}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{\Sigma}}_{\perp} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{V}}^{\mathsf{T}} \\ \overline{\mathbf{V}}_{\perp}^{\mathsf{T}} \end{bmatrix}$$

$$\mathbf{M} = \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{T}} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_{\perp} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\perp} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}^{\mathsf{T}} \\ \mathbf{V}_{\perp}^{\mathsf{T}} \end{bmatrix}.$$

Here, $\overline{\sigma}_1 \geq \cdots \geq \overline{\sigma}_{n_1}$ (resp. $\sigma_1 \geq \cdots \geq \sigma_{n_1}$) stand for the singular values of $\overline{\mathbf{M}}$(resp. $\mathbf{M}$), $\overline{\boldsymbol{u}}_i$(resp. $\boldsymbol{u}_i$) denotes the left singular vector associated with the singular value $\overline{\sigma}_i$(resp. σ_i), and $\overline{\boldsymbol{v}}_i$(resp. $\boldsymbol{v}_i$) denotes the right singular vector associated with the singular value $\overline{\sigma}_i$(resp. σ_i). In addition, for any fixed integer $r \in [n]$, we denote

$$\mathbf{\Sigma} := \operatorname{diag}\{\sigma_1, \ldots, \sigma_r\}, \quad \mathbf{\Sigma}_{\perp} := \operatorname{diag}\{\sigma_{r+1}, \ldots, \sigma_{n_1}\},$$
$$\mathbf{U} := [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r] \in \mathbb{R}^{n_1 \times r}, \quad \mathbf{U}_{\perp} := [\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_{n_1}] \in \mathbb{R}^{n_1 \times (n_1 - r)},$$
$$\mathbf{V} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r] \in \mathbb{R}^{n_2 \times r}, \quad \mathbf{V}_{\perp} := [\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_{n_2}] \in \mathbb{R}^{n_2 \times (n_2 - r)}.$$

The matrices $\overline{\mathbf{\Sigma}}$, $\overline{\mathbf{\Sigma}}_{\perp}$, $\overline{\mathbf{U}}$, $\overline{\mathbf{U}}_{\perp}$, $\overline{\mathbf{V}}$, $\overline{\mathbf{V}}_{\perp}$ are defined analogously. If $\mathbf{E} = \mathbf{M} - \overline{\mathbf{M}}$ satisfies $\|\mathbf{E}\| \leq \overline{\sigma}_r - \overline{\sigma}_{r+1}$, then with the projection matrices $P_{\mathbf{U}} := \mathbf{U}\mathbf{U}^{\mathsf{T}}$, one has

$$\max\left\{\|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\|, \|P_{\mathbf{V}} - P_{\overline{\mathbf{V}}}\|\right\} \leq \frac{\sqrt{2}\max\left\{\|\mathbf{E}^{\mathsf{T}}\overline{\mathbf{U}}\|, \|\mathbf{E}\overline{\mathbf{V}}\|\right\}}{\overline{\sigma}_r - \overline{\sigma}_{r+1} - \|\mathbf{E}\|}.$$

In particular, if $\|\mathbf{E}\| \leq (1 - 1/\sqrt{2})(\overline{\sigma}_r - \overline{\sigma}_{r+1})$, then one has

$$\max\left\{\|P_{\mathbf{U}} - P_{\overline{\mathbf{U}}}\|, \|P_{\mathbf{V}} - P_{\overline{\mathbf{V}}}\|\right\} \leq \frac{\sqrt{2}\|\mathbf{E}\|}{\overline{\sigma}_r - \overline{\sigma}_{r+1}}.$$

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093
Email address: idumitriu@ucsd.edu

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093
Email address: h9wang@ucsd.edu

Department of Mathematics, University of California, San Diego, La Jolla, CA 92093
Email address: yiz084@ucsd.edu