A Experimental Results on Real-World Data

A.1 Categorical Data

We evaluate the performance of Algorithm 2 on two datasets: *Congressional Voting Records* and *Mushroom*. In the congressional voting records, the task is to group 435 Congressmen into two parties, Democrats and Republicans, based on whether they fovored or vetoed 16 different issues, or abstained their votes. The other categorical dataset consists of a total of 8124 mushroom species. Based on the 22 different physical or biological features, one needs to separate the edible from the poisonous apart. The authors in (Ghoshdastidar and Dukkipati, 2017) used the following way to embed the categorical data into a hypergraph. Take the political data for example. For each attribute (issue) and each possible value of that attribute (favor, veto or abstain), a hyperedge is formed among all instances (Congressmen) that share this particular value of the attribute. This generates a rather sparse non-uniform hypergraph. In order to have a fair comparison, though, we also adopt this non-uniform construction and implement our algorithm on top of these sparse hypergraphs. Table 1 compares the mismatch ratio of different algorithms with our spectral initialization.

Table 1: Experimental Results on Categorical Data

	ROCK	**COOLCAT**	**LIMBO**	**hMETIS**	**SnHP**	**GTS**	**Algo 2**
Voting	0.16	0.15	0.13	0.24	0.1263	0.3819	0.1129
Mushroom	0.43	0.27	0.11	0.48	0.1129	0.4979	0.1103

The results for ROCK, COOLCAT and LIMBO algorithms are taken from (Andritsos et al., 2004), while the multi-level approach (hMETIS) is taken from (Ghoshdastidar and Dukkipati, 2017). We can see that Algorithm 2 outperforms other algorithms on the voting databset and is on a par with the known best on the mushroom dataset.

A.2 Numerical Data

As for the numerical data, the authors in (Ghoshdastidar and Dukkipati, 2015) point out that a 3-way similarity measure can be utilized to construct a 3-uniform hypergraph, which gives an overall more robust performance empirically. In particular, for any three data points $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, the similarity among these three nodes is defined as

$$\exp\left(-\beta \max\left\{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2, \|\boldsymbol{x}-\boldsymbol{z}\|_2^2, \|\boldsymbol{y}-\boldsymbol{z}\|_2^2\right\}\right)$$

where β is a tuning parameter. It is inspired by the pairwise gaussian-density-like similarity $\exp(-\beta \|\boldsymbol{x}-\boldsymbol{y}\|_2^2)$ between data points $\boldsymbol{x}, \boldsymbol{y}$ often seen when performing spectral clustering (Ghoshdastidar and Dukkipati, 2015). In constrast to the affinity tensor where each entry is a real number in $[0, 1]$ considered in (Ghoshdastidar and Dukkipati, 2015), the adjacency tensor that we use is binary-valued. Instead of directly using these 3-way similarity values, we treat them as the connecting probability parameters, i.e. the success probabilities of the appearance of hyperedges. Table 2 compares our method with the generalized tensor spectral method (GTS) and conventional spectral clustering on similarity graph (SP) taken from (Ghoshdastidar and Dukkipati, 2015) in the first two columns. As above, our results are the average over 20 runs of realizations with regard to the randomness in the embedding process.

Table 2: Experimental Results on Numerical Data

Dataset	**SP**	**GTS**	**Algo 2**	**Algo 1**
Wine	0.342	0.331	0.105	0.087
Haberman's Survival	0.423	0.392	0.485	0.397
Vertebral Column	0.345	0.333	0.262	0.275
Ionosphere	0.325	0.316	0.319	0.315

First, we want to make clear that we does not optimize our results by choosing the β that leverages the best performance according to each specific ground truth. In our experiment, we choose the value of β to control the edge density in the hypergraph. From Table 2, we could see that our two-step algorithm does outperform the one proposed in (Ghoshdastidar and Dukkipati, 2015) and traditional graph-based method, endorsing the advantage of using a higher-order relational information to perform the task of clustering. Except for the *Vertebral Column* data where the second refinement step fails to have a gain from the first initialization step, we could see that, overally speaking, the refinement step helps reducing the number of misclassified nodes and achieves a better performance on the real-world data.

B Proof of Theorem 3.2

We first introduce the concept of local loss. The equivalence class of a community assignment σ is defined as $\Gamma(\sigma) \triangleq \{\sigma' \mid \exists\, \delta \in \mathcal{S}_K \text{ s.t. } \sigma' = \delta \circ \sigma\}$. Let $S_\sigma(\widehat{\sigma}) = \{\sigma' \in \Gamma(\widehat{\sigma}) \mid d_{\mathrm{H}}(\sigma', \sigma) = d_{\mathrm{H}}(\widehat{\sigma}, \sigma)\}$ be the set of all permutations in the equivalence class of $\widehat{\sigma}$ that achieve the minimum distance. For each $i \in [n]$, the local loss function is defined as the proportion of false labeling of node i in $S_\sigma(\widehat{\sigma})$.

$$\ell(\widehat{\sigma}(i), \sigma(i)) \triangleq \sum_{\sigma' \in S_\sigma(\widehat{\sigma})} \frac{d_{\mathrm{H}}(\widehat{\sigma}(i), \sigma(i))}{|S_\sigma(\widehat{\sigma})|}$$

It turns out that it is rather easy to study the local loss. Recall the sub-parameter space Θ_d^L of Θ_d^0 defined in (1) where the sizes of all communities are almost equal ($n_k \in \{n'-1, n', n'+1\}$ to be more specific). Since Θ_d^L is closed under permutation, we can apply the *global-to-local* lemma in Zhang and Zhou, 2016.

Lemma 2.1 (Lemma 2.1 in Zhang and Zhou, 2016)**:** *Let Θ be any homogeneous parameter space that is closed under permutation. Let* Unif *be the uniform prior over all the elements in Θ. Define the global Bayesian risk as* $\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}) = \frac{1}{|\Theta|}\sum_{\sigma\in\Theta} \mathbb{E}_\sigma \ell(\widehat{\sigma}, \sigma)$ *and the local Bayesian risk* $\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1)) = \frac{1}{|\Theta|}\sum_{\sigma\in\Theta} \mathbb{E}_\sigma \ell(\widehat{\sigma}(1), \sigma(1))$ *for the first node. Then*

$$\inf_{\widehat{\sigma}} \mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}) = \inf_{\widehat{\sigma}} \mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1))$$

Second, the local Bayesian risk can be transformed into the risk function of a hypothesis testing problem. We consider the most indistinguishable case where the potential candidate only disagrees with the ground truth on a single node. The key observation is that the situation is exactly the same as our testing one node at a time in the local version of the MLE method.

Lemma 2.2:

$$\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1)) \geq \mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_ir_j}} C_{r_ir_j}(X_u^{(r_j)} - X_u^{(r_i)}) \geq 0\right\}$$

where $f(t) \triangleq \frac{t}{1-t}$ *for* $C_{xy} \triangleq \log\frac{f(x)}{f(y)}$, *and for each* (r_i, r_j) *pair in* $\mathcal{N}_d$, $X_u^{(r_j)} \overset{i.i.d.}{\sim} \mathrm{Ber}(p_j), X_u^{(r_i)} \overset{i.i.d.}{\sim} \mathrm{Ber}(p_i)\ \forall u = 1, \ldots, m_{r_ir_j}$ *are all mutually indepedent random variables.*

With the aid of the Rozovsky lower bound Rozovsky, 2003, we are able to prove the following auxiliary result.

Lemma 2.3: *If* $\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \to \infty$, *then there exists a positive sequence* $\zeta_n \to 0$ *such that*

$$\mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_ir_j}} C_{r_ir_j}(X_u^{(r_j)} - X_u^{(r_i)}) \geq 0\right\} \geq \exp\Big(-(1+\zeta_n) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j}\Big) \tag{9}$$

Proof of Theorem 3.2. Finally, since the Bayesian risk always lower bound the minimax risk, we have

$$\begin{aligned}
\mathrm{R}_d^* &= \inf_{\widehat{\sigma}} \sup_{\sigma\in\Theta_d^0} \mathbb{E}_\sigma \ell(\widehat{\sigma},\sigma) \geq \inf_{\widehat{\sigma}} \sup_{\sigma\in\Theta_d^L} \mathbb{E}_\sigma \ell(\widehat{\sigma},\sigma) \\
&\geq \inf_{\widehat{\sigma}} \mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}) = \inf_{\widehat{\sigma}} \mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1)) \\
&\geq \mathbb{P}\left\{ \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_ir_j}} C_{r_ir_j}(X_u^{(r_j)} - X_u^{(r_i)}) \geq 0 \right\} \\
&\geq \exp\left(-(1+\zeta_n) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \right)
\end{aligned}$$

■

C Proof of Theorem 5.1

Fix any $(\mathbf{B},\sigma)$ in Θ_d^0. Let C_0, $\delta > 0$ and $\gamma = \gamma_n$ be constants in Condition 5.1. For each $u \in [n]$, there exists some $\pi_u \in \mathcal{S}_K$ so that

$$\mathbb{P}_\sigma \left\{ \ell_0\big((\widehat{\sigma}_u)_{\pi_u}, \sigma\big) \leq \gamma_n \right\} \geq 1 - C_0 n^{-(1+\delta)}$$

In consequence,

$$\begin{aligned}
\mathbb{E}_\sigma \ell_0(\widehat{\sigma},\sigma) &= \mathbb{E}_\sigma \left[\frac{1}{n} \sum_{u\in[n]} \mathbb{1}\left\{ \pi^{\mathrm{CSS}}(\widehat{\sigma}_u(u)) \neq \sigma(u) \right\} \right] \\
&= \frac{1}{n} \sum_{u\in[n]} \mathbb{P}_\sigma \left\{ \pi^{\mathrm{CSS}}(\widehat{\sigma}_u(u)) \neq \sigma(u) \right\} \\
&\leq \frac{1}{n} \sum_{u\in[n]} \mathbb{P}_\sigma \left\{ \big(\widehat{\sigma}_u(u)\big)_{\pi_u} \neq \sigma(u) \right\} + \mathbb{P}_\sigma \left\{ \pi^{\mathrm{CSS}} \neq \pi_u \right\}
\end{aligned}$$

where π^{CSS} is the consensus permutation (4) in Algorithm 1. By Lemma 5.2, for any $(\mathbf{B},\sigma) \in \Theta_d^0$ and each $u \in [n]$,

$$\mathbb{P}_\sigma \left\{ \big(\widehat{\sigma}_u(u)\big)_{\pi_u} \neq \sigma(u) \right\} \leq (K-1)\cdot \exp\left(-(1-\zeta_n^{''}) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \right) + Cn^{-(1+\delta)}$$

for some constants $C, \delta > 0$ and $\zeta_n^{''} \to 0$. Moreover, Lemma 5.3 implies that $\mathbb{P}\left\{\pi^{\mathrm{CSS}} \neq \pi_u\right\} \leq Cn^{-(1+\delta)}$. Together,

$$\sup_{(\mathbf{B},\sigma)\in\Theta_d^0} \mathbb{E}_\sigma \ell_0(\widehat{\sigma},\sigma) \leq (K-1)\cdot \exp\left(-(1-\zeta_n^{''}) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \right) + C' n^{-(1+\delta)}$$

Since we assume $\lim_{n\to\infty} \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \to \infty$, we further have

$$\begin{aligned}
\sup_{(\mathbf{B},\sigma)\in\Theta_d^0} \mathbb{E}_\sigma \ell_0(\widehat{\sigma},\sigma) &\leq \exp\left(-(1-\zeta_n^{'''}) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \right) + C' n^{-(1+\delta)} \\
&= ① + ②
\end{aligned}$$

for some $\zeta_n^{'''} \to 0$. If ① decays faster than ②, then $\mathrm{R}_d^* = o(\frac{1}{n^{1+\delta}}) < \frac{1}{n}$ for sufficiently large n. Therefore, $\mathrm{R}_d^* = 0$ and the corresponding parameters satisfy the criterion of exact recovery. On the other hand, if ① dominates ②, then there exists $\zeta_n \to 0$ such that

$$\mathrm{R}_d^* \leq \exp\left(-(1-\zeta_n) \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \right)$$

In either case, the claimed upper bound is achieved.

D Proof of Theorem 5.2

To prove Theorem 5.2, we first introduce some notations. For a matrix $\mathbf{M}$, we denote its operator norm by $\|\mathbf{M}\|_{\mathrm{op}}$ and its Frobenius norm by $\|\mathbf{M}\|_{\mathrm{F}}$. Also, let $\mathrm{O}(K_1,K_2) = \left\{\mathbf{V}\in\mathbb{R}^{K_1\times K_2} \mid \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_{K_2}\right\}$ for $K_1 \geq K_2$ be the set of all orthogonal $K_1\times K_2$ matrices. The proof requires the following two lemmas. First, we demonstrates that the hypergraph Laplacian after trimming does not deviate too much from its original expectation.

Lemma 4.1: *$\forall C' > 0$, $\exists\ C > 0$ such that*

$$\|\mathsf{T}_\tau(\mathcal{L}(\mathbf{A})) - \mathbb{E}\mathcal{L}(\mathbf{A})\|_{\mathrm{op}} \leq C\sqrt{n^{d-1}p_1 + 1}$$

with probability at least $1 - n^{-C'}$ uniformly over $\tau \in \left[C_1(n^{d-1}p_1+1), C_2(n^{d-1}p_1+1)\right]$ for some sufficiently large constants C_1 and C_2.

The next lemma analyzes the spectrum of $\mathbb{E}\mathcal{L}(\mathbf{A})$ and pinpoints a special structure.

Lemma 4.2 (Lemma 6 in Gao et al., 2017)**:** *We have*

$$\mathsf{SVD}_K\left(\mathbb{E}\mathcal{L}(\mathbf{A})\right) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$$

where $\mathbf{U} = \mathbf{Z}\mathbf{\Delta}^{-1}\mathbf{W}$ with $\mathbf{\Delta} = \mathsf{diag}(\sqrt{n_1},\ldots,\sqrt{n_K})$. $\mathbf{Z}\in\{0,1\}^{n\times K}$ is a matrix with exactly one nonzero entry in each row at $(i,\sigma(i))$ taking value 1 and $\mathbf{W}\in\mathrm{O}(K,K)$.

Proof of Theorem 5.2. Under the assumption $p_i \geq p_j\,\forall i < j$, we have $\mathbb{E}\tau \in [C_1' n^{d-1}p_1, C_2' n^{d-1}p_1]$ for some large constant C_1', C_2'. Thus by Bernstein's inequality, with probability at least $1 - e^{-C'n}$, we have $\tau \in [C_1 n^{d-1}p_1, C_2 n^{d-1}p_1]$. By Davis-Kahan theorem (Davis and Kahan, 1970), we have

$$||\widehat{\mathbf{U}} - \mathbf{U}\mathbf{W_1}||_{\mathrm{F}} \leq C\frac{\sqrt{K}}{\lambda_K}||\mathsf{T}_\tau(\mathcal{L}(\mathbf{A})) - \mathbb{E}\mathcal{L}(\mathbf{A})||_{\mathrm{op}}$$

for some $\mathbf{W_1}\in O(K,K)$ and some constant $C > 0$. Then applying Lemma 4.2, we have

$$||\widehat{\mathbf{U}} - \mathbf{V}||_{\mathrm{F}} \leq C\frac{\sqrt{K}}{\lambda_K}||\mathsf{T}_\tau(\mathcal{L}(\mathbf{A})) - \mathbb{E}\mathcal{L}(\mathbf{A})||_{\mathrm{op}} \tag{10}$$

where $\mathbf{V} = \mathbf{Z}\mathbf{\Delta}^{-1}\mathbf{W} = [\boldsymbol{v}_1^{\mathrm{T}}\ldots\boldsymbol{v}_n^{\mathrm{T}}]^{\mathrm{T}}$ as we state in Lemma 4.2. Combining (10), Lemma 4.1 and the conclusion $\tau\in[C_1n^{d-1}p_1, C_2n^{d-1}p_1]$ with probability at least $1 - e^{-C'n}$, we have

$$||\widehat{\mathbf{U}} - \mathbf{V}||_{\mathrm{F}} \leq \frac{C\sqrt{K}\sqrt{n^{d-1}p_1}}{\lambda_K} \tag{11}$$

with probability at least $1 - n^{-C'}$. The definition of $\mathbf{V}$ implies that

$$\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2 \begin{cases} \geq \sqrt{\frac{2K}{n}}, & \text{, if } \sigma(i)\neq\sigma(j) \\ = 0, & \text{, otherwise.}\end{cases}$$

Let $\mathbf{X} = \mathbf{\Delta}^{-1}\mathbf{W}$, which means $\boldsymbol{v}_i = \boldsymbol{x}_{\sigma(i)}$. Recall the definition of critical radius $r = \nu\sqrt{\frac{K}{n}}$ in Algorithm 2. Define the sets

$$T_i = \left\{s\in\sigma^{-1}(i) : \|\widehat{\boldsymbol{u}}_s - \boldsymbol{x}_i\|_2 < \frac{r}{2}\right\},\ i\in[K] \tag{12}$$

By definition, $T_i\cap T_j = \varnothing$ for $i\neq j$ and

$$\bigcup_{i\in[K]} T_i = \left\{s\in[n] : \|\widehat{\boldsymbol{u}}_s - \boldsymbol{v}_s\|_2 < \frac{r}{2}\right\}$$

Thus,

$$\left|\Big(\bigcup_{i\in[K]} T_i\Big)^c\right|\cdot\frac{r^2}{4} \leq \sum_{s\in[n]}\|\widehat{\boldsymbol{u}}_s - \boldsymbol{v}_s\|_2^2 \leq \frac{C^2Kn^{d-1}p_1}{\lambda_K^2}$$

where the last inequality is due to (11). The rest of the proof is identical to the proof of Theorem 3 in Gao et al., 2017. For completeness, we shall repeat it again here. After some rearrangements, we have

$$\left|\Big(\bigcup_{i\in[K]} T_i\Big)^c\right| \le \frac{4C^2n^dp_1}{\mu^2\lambda_K^2} \tag{13}$$

It means most nodes are close to the centers and are in the set we defined in (12). Also note that the sets $\{T_i\}_{i\in[K]}$ are disjoint. Suppose there is some $i\in[K]$ such that $|T_i|<|\sigma^{-1}(i)|-|(\cup_{i\in[K]}T_i)^c|$, we have $|\cup_{i\in[K]}T_i| = \sum_{i\in[K]}|T_i| < n-|(\cup_{i\in[K]}T_i)^c| = |\cup_{i\in[K]}T_i|$, which leads to a contradiction. Thus, we must have

$$|T_i| \ge |\sigma^{-1}(i)| - \left|\Big(\bigcup_{i\in[K]} T_i\Big)^c\right| \ge \frac{n}{K} - \frac{4C^2n^dp_1}{\mu^2\lambda_K^2} > \frac{n}{2K}$$

where the last inequality is from the assumption (8). Since the cluster centers are at least $\sqrt{\frac{2K}{n}}$ apart from each others and both $\{T_i\}_{i\in[K]}$, $\{\widehat{C}_i\}_{i\in[K]}$ (recall that $\widehat{C}_i$ are defined in Algorithm 2) are defined through the critical radius $r=\mu\sqrt{\frac{K}{n}}$, each $\widehat{C}_i$ should intersect with only one T_i. We claim that there is a permutation π of set $[K]$, such that

$$\widehat{C}_i \bigcap T_{\pi(i)} \ne \varnothing,\ \left|\widehat{C}_i\right| \ge \left|T_{\pi(i)}\right| \forall i\in[K] \tag{14}$$

We could now continue the proof with claim (14), where the proof of (14) can be found in Gao et al., 2017 (in their proof of Theorem 3). It is mainly established by an easy mathematical induction. From the definition of $\widehat{C}_i$ and (14), we have for any $i\ne j$, $T_{\pi(i)}\cap\widehat{C}_j=\varnothing$. This directly implies that for any $i\ne j$, $T_{\pi(i)}\subset\widehat{C}_j$. Thus, we know that $T_{\pi(i)}\cap\widehat{C}_i^c\subset\Big(\cup_{i\in[K]}\widehat{C}_i\Big)^c$. Therefore,

$$\bigcup_{i\in[K]}\Big(T_{\pi(i)}\bigcap\widehat{C}_i^c\Big)\subset\Big(\bigcup_{i\in[K]}\widehat{C}_i\Big)^c$$

Combining with the fact that $T_i\cap T_j=\varnothing\ \forall i\ne j$, we have

$$\sum_{i\in[K]}\left|T_{\pi(i)}\bigcap\widehat{C}_i^c\right| \le \left|\Big(\bigcup_{i\in[K]}\widehat{C}_i\Big)^c\right| \tag{15}$$

By definition, $\widehat{C}_i\bigcap\widehat{C}_j=\varnothing\ \forall i\ne j$. Along with (14), we have

$$\left|\Big(\bigcup_{i\in[K]}\widehat{C}_i\Big)^c\right| = n-\sum_{i\in[K]}\left|\widehat{C}_i\right| \le n-\sum_{i\in[K]}|T_i| = \left|\Big(\bigcup_{i\in[K]}T_i\Big)^c\right| \tag{16}$$

Together with (13), (15) and (16), we have

$$\sum_{i\in[K]}\left|T_{\pi(i)}\bigcap\widehat{C}_i^c\right| \le \frac{4C^2n^dp_1}{\mu^2\lambda_K^2} \tag{17}$$

Since for each $u\in\cup_{i\in[K]}\Big(T_{\pi(i)}\cap\widehat{C}_i\Big)$, we have $\widehat{\sigma}(u)=i$ when $\sigma(u)=\pi(i)$, the mis-classification ratio is bounded by

$$\begin{aligned}
\ell_0(\widehat{\sigma},\pi^{-1}(\sigma)) &\le \frac{1}{n}\left|\Big(\bigcup_{i\in[K]}(T_{\pi(i)}\bigcap\widehat{C}_i)\Big)^c\right| \\
&\le \frac{1}{n}\left[\left|\Big(\bigcup_{i\in[K]}(T_{\pi(i)}\bigcap\widehat{C}_i)\Big)^c\bigcap\Big(\bigcup_{i\in[K]}T_i\Big)\right| + \left|\Big(\bigcup_{i\in[K]}T_i\Big)^c\right|\right] \\
&\le \frac{1}{n}\left[\sum_{i\in[K]}\left|T_{\pi(i)}\bigcap\widehat{C}_i^c\right| + \left|\Big(\bigcup_{i\in[K]}T_i\Big)^c\right|\right] \\
&\le \frac{8C^2n^{d-1}p_1}{\mu^2\lambda_K^2}
\end{aligned}$$

where the last inequality is from (13) and (17). This proves the desired conclusion. ■

Remark. Essentially, Theorem 5.2 says that the performance of Algorithm 2 will be upper-bounded in regard to the K-th largest singular value. When λ_K is large, it means that the singular vectors are well separated, ensuring the algorithm to have a good performance. This is similar to classical spectral clustering methods.

E Proof of Theorem 5.3

Theorem 5.3 is established through controling the λ_K term. For convinience, we introduce the notions below regarding asymptotic relation. $f \approx g$ is to mean that f and g are in the same order if $f(n)/C \leq g(n) \leq Cf(n)$ for some constant $C \geq 1$ independent of n. $f \lesssim g$, defined by $\lim_{n\to\infty}(f(n) - g(n)) \leq 0$, means that $f(n)$ is asymtotically smaller than $g(n)$. $f \gtrsim g$ is equivalent to $g \lesssim f$. To take out the dependency on λ_K, we use the observation below.

Lemma 5.1: *For d-*`hSBM` *in $\Theta_d^0(n, K, \mathbf{p}, \eta)$, we have*

$$\lambda_K \gtrsim \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j}(p_i - p_j) \tag{18}$$

Proof. Let $\mathbf{P}_\mathrm{H} \triangleq \mathbb{E}\mathcal{L}(\mathbf{A})$ for an adjacency tensor $\mathbf{A}$. We start from analyzing the entries of $\mathbf{P}_\mathrm{H}$. Under the transformation from a d-dimensional tensor into a two-dimensional matrix, each entry in $\mathbf{P}_\mathrm{H}$ is a weighted combination of the probability parameters p_i's. To be specific, $(\mathrm{P}_\mathrm{H})_{ij}$ aggregates the contribution from other nodes $u \in [n] \setminus \{u, v\}$, and the value depends on the community relation induced by each hyperedge correspondingly. Depending on whether or not the two nodes u and v are in the same community, we have, $\forall i \neq j$

$$(\mathrm{P}_\mathrm{H})_{ij} \approx \begin{cases} u & \sigma(i) = \sigma(j) \\ v & \text{, otherwise.} \end{cases} \tag{19}$$

The explicit expression for $(\mathrm{P}_\mathrm{H})_{ij}$ changes for different values of d, the order of the underlying hypergraph. Observe that $u \geq v$ since we assume that p_i's are in decreasing order, i.e. $p_i \geq p_j$ for $i < j$. Below are u, v for the case $d = 3$ and 4.

$$\text{When } d = 3 \begin{cases} u \approx n'(p + (K-1)q) \\ v \approx n'(2q + (K-2)r) \end{cases} \tag{20}$$

$$\text{When } d = 4 \begin{cases} u \approx (n')^2\left(\frac{1}{2}p_1 + (K-1)p_2 + \frac{K-1}{2}p_3 + \binom{K-1}{2}p_4\right) \\ v \approx (n')^2\left(p_2 + p_3 + \frac{5(K-2)}{2}p_4 + \binom{K-2}{2}p_5\right) \end{cases} \tag{21}$$

Deducting v for each entry in $\mathbf{P}_\mathrm{H}$, we have

$$\mathbf{P}_\mathrm{H} - (1-\eta)v\mathbf{1}_n\mathbf{1}_n^T \approx (u - v)\sum_{t=1}^{K} \boldsymbol{v}_t\boldsymbol{v}_t^T \tag{22}$$

where $\boldsymbol{v}_t$ is defined as $\boldsymbol{v}_t = \left(\mathbf{0}_{n_1}^T, \ldots, \mathbf{1}_{n_t}^T, \ldots, \mathbf{0}_{n_K}^T\right)^T$ for each $t \in [K]$. Note that $\{\boldsymbol{v}_t\}_{t=1}^K$ are orthogonal to each other. Therefore,

$$\lambda_K\left(\sum_{t=1}^{K}\boldsymbol{v}_t\boldsymbol{v}_t^T\right) \geq \min_{t\in[K]} n_t \geq (1-\eta)n'$$

By Weyl's inequality,

$$\lambda_K(\mathbf{P}_\mathrm{H}) \geq (u-v)\,\lambda_K\left(\sum_{t=1}^{K}\boldsymbol{v}_t\boldsymbol{v}_t^T\right) + \lambda_n\left((1-\eta)v\mathbf{1}_n\mathbf{1}_n^T\right) \gtrsim (n')\,(u-v)$$

To further control $u - v$, let's first look at a few cases for lower-order d. For the case $d = 3$, we have

$$u - v \approx n'\big(p - q + (K-2)(q-r)\big)$$

while for the case $d = 4$,

$$u - v \approx (n')^2\left[\frac{1}{2}(p_1 - p_2) + \frac{1}{2}(p_2 - p_3) + (K-2)(p_3 - p_4) + \frac{K-2}{2}(p_2 - p_4) + \binom{K-1}{2}(p_4 - p_5)\right]$$

Note that $u-v$ could be represented as a weighted sum of pairwise comparisons, that is, $\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} M_{r_ir_j}(p_i-p_j)$ for some M_{r_i,r_j}'s. Recall that in our definition, $(i,j):(r_i,r_j)\in\mathcal{N}_d$ if the hyperedges of type r_i and r_j have community assignments that differ on only one node. The new coefficient $M_{r_ir_j}$ would be similar to $m_{r_ir_j}$. In fact, they will only differ up to a constant related only to d (in fact, up to $d-1$).

Moreover, $n'C_{r_ir_j}\geq m_{r_ir_j}$ for all possible (r_i,r_j). When counting, in m_{r_i,r_j} we fix one dimension (the first dimension to node 1), while in M_{r_i,r_j} two dimensions are fixated at u and v. Essentially, $u-v$ counts the difference of the number of random variables between two assignments, one being $\sigma(u)=\sigma(v)$ and the other being otherwise. Without loss of generality, we may think of the community labeled of u as a fixed number as in the operational definition of m_{r_i,r_j}, while the community label of v should be different from $\sigma(u)$. By multiplying back n' to get the expression $n'M_{r_i,r_j}$, we unshackle v and allow it to vary within the $\sigma(v)$-th community, the cardinality of which is approximately n'. Undoubtedly, there are double countings in both the number m_{r_i,r_j} and M_{r_i,r_j}. The value of m_{r_i,r_j} is normalized with respect to $d-1$ companions (only one dimension is fixed), while the value of M_{r_i,r_j} is normalized with respect to $d-2$ companions (two dimension are fixed). As a result, there are still some $\boldsymbol{l}=(u,v,l_3,\ldots,l_d)$'s being doubled counted in coefficient $n'M_{r_i,r_j}$ as opposed to coefficient m_{r_i,r_j}. This is the reason why the former is always larger than or equal to the latter.

Recall that the probability parameter $\boldsymbol{p}=\{p_1,\ldots,p_{\kappa_d}\}$ follows the majorization rule, which means that $p_i\leq p_j$ for all $i<j$. Combined with these fact, we have

$$\lambda_K(\mathbf{P}_{\mathrm{H}})\gtrsim(n')\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} C_{r_ir_j}(p_i-p_j)\gtrsim\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j}(p_i-p_j)$$

Hence we complete the proof. ∎

F Proof of Lemma 5.1

Fix any $(\mathbf{B},\sigma)\in\Theta_d^0$ and $u\in[n]$. Define the event

$$E_u\triangleq\left\{\ell_0\big((\widehat{\sigma}_u)_{\pi_u},\sigma\big)\leq\gamma\right\}\tag{23}$$

For simplicity, we assume that π_u is the identity permutation. Fix any $i\in[K]$. Then, on E_u we have

$$n_i\geq\left|\widetilde{C}_i^u\cap C_i\right|\geq n_i-\gamma_1 n,\ \left|\widetilde{C}_i^u\cap C_i\right|\leq\gamma_2 n\tag{24}$$

where $\gamma_1,\gamma_2\geq 0$ and $\gamma_1+\gamma_2\leq\gamma$. Let C_i' be any deterministic subset of $[n]$ such that (24) holds with $\widetilde{C}_i^u$ replaced by C_i'. By definition, there are at most

$$\sum_{l=0}^{\gamma n}\binom{n_i}{l}\sum_{m=0}^{\gamma n}\binom{n-n_i}{m}\leq\exp\left(C_1\gamma n\log\frac{1}{\gamma}\right)\tag{25}$$

different subsets with this property where $C_1>0$ is an absolute constant. We will only prove the case $\left|\widehat{\mathrm{B}}^u_{i\cdot\mathbf{1}}-\mathrm{B}_{i\cdot\mathbf{1}}\right|\leq o(\max_{(r_i,r_j)\in\mathcal{N}_d}p_i-p_j)$ where $\mathbf{1}$ is the d-dimensional all-one (row) vector. For the rest of the cases, we can easily follow an similar procedure to obtain the desired upper bound.

Let ξ_i' be the edges within C_i'. Note that ξ_i' consists of $\binom{n_i}{d}$ independent Bernoulli random variables. The number of truly $\mathrm{Ber}(\mathrm{B}_{i\cdot\mathbf{1}})$'s is at least $\binom{n_i-\gamma n}{d}$. By an simple combinatorial argument, we have

$$\mathbb{E}\left[\frac{|\xi_i'|}{\binom{|C_i'|}{d}}\right]\geq\min_{t\in[0,\gamma K]}\left\{p_i-(1-(1-t)^d)(p_i-p_{\mathcal{K}_d})\right\}\tag{26}$$

$$\mathbb{E}\left[\frac{|\xi_i'|}{\binom{|C_i'|}{d}}\right]\geq\max_{t\in[0,\gamma K]}\left\{p_i+(1-(1-t)^d)(p_1-p_i)\right\}\tag{27}$$

Note that p_i equals p_1 in this case. In general, though, the estimation of all the parameters have a similar formula, and therefore we use p_i still. Since K is constant, (26) becomes $p_i-o(\max_{(r_i,r_j)\in\mathcal{N}_d}p_i-p_j)$ by breaking

$p_i - p_{\mathcal{K}_d}$ into pairwise difference. Similarly, (27) would be $p_i + o(\max_{(r_i,r_j)\in\mathcal{N}_d} p_j - p_i)$ (since $K\gamma = o(1)$ is assumed). Together,

$$\left|\mathbb{E}\left[\frac{|\xi_i'|}{\binom{|C_i'|}{3}}\right] - \mathrm{B}_{i\cdot\mathbf{1}}\right| \le o\Big(\max_{(r_i,r_j)\in\mathcal{N}_d} p_i - p_j\Big) \tag{28}$$

On the other hand, by the Bernstein's inequality,

$$\mathbb{P}\left\{\big|\,|\xi_i'| - \mathbb{E}\,|\xi_i'|\,\big| > t\right\} \le 2\exp\left(-\frac{t^2}{2\left(\binom{n_i-\gamma n}{d}p_1 + \frac{2}{3}t\right)}\right)$$

Let

$$\begin{aligned} t^2 &= \binom{n_i - \gamma n}{d} p_1 \left(C_1 \gamma n \log\gamma^{-1} + (3+\delta)\log n\right) \vee \left(2C_1\gamma n\log\gamma^{-1} + 2(3+\delta)\log n\right)^2 \\ &\lesssim \left(\frac{n'}{K^{d-1}}\sqrt{n^{d-1}p_1\gamma\log\gamma^{-1}} + \gamma n\log\gamma^{-1}\right)^2 \end{aligned}$$

We have

$$\mathbb{P}\left\{\big|\,|\xi_i'| - \mathbb{E}\,|\xi_i'|\,\big| > C_\delta\Big(\frac{n'}{K^{d-1}}\sqrt{n^{d-1}p_1\gamma\log\gamma^{-1}} + \gamma n\log\gamma^{-1}\Big)\right\} \le \exp\left(-C_1\gamma n\log\gamma^{-1}\right)n^{-(3+\delta)}$$

Hence, with probability at least

$$1 - \exp\left(-C_1\gamma n\log\gamma^{-1}\right)n^{-(3+\delta)}$$

, we have

$$\left|\frac{|\xi_i'|}{\binom{|C_i'|}{d}} - \mathbb{E}\left[\frac{|\xi_i'|}{\binom{|C_i'|}{d}}\right]\right| \le C_\delta\left(\left(\frac{1}{n}\right)^{d-1}\sqrt{n^{d-1}p_1\gamma\log\gamma^{-1}} + \gamma\frac{K^3}{n^2}\log\gamma^{-1}\right)$$

Since $K\gamma\log\gamma^{-1} = O(1)$ and with the assumption $\max_{(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \to \infty$, $p_1 \asymp p_{\mathcal{K}_d}$, we further have

$$\left|\frac{|\xi_i'|}{\binom{|C_i'|}{d}} - \mathbb{E}\left[\frac{|\xi_i'|}{\binom{|C_i'|}{d}}\right]\right| \le o\Big(\max_{(r_i,r_j)\in\mathcal{N}_d} p_i - p_j\Big) \tag{29}$$

at least $1 - \exp\left(-C_1\gamma n\log\gamma^{-1}\right)n^{-(3+\delta)}$ in probability. Combining (29), (28) and apply the Union Bound over (25), we have

$$\left|\frac{|\xi_i'|}{\binom{|C_i'|}{d}} - \mathrm{B}_{i\cdot\mathbf{1}}\right| \le o\Big(\max_{(r_i,r_j)\in\mathcal{N}_d} p_i - p_j\Big)$$

with probability at least $1 - n^{-(3+\delta)}$.

The proof for the rest $\mathrm{B}_{\boldsymbol{s}}, \boldsymbol{s} \in [K]^d$ are all similar and thus omit. The key observation is that by the requirement on γ, we will only have $o(1)$ misclassification proportion. Which means for each sample mean, the proportion of "correct" random variables will dominate the "incorrect" ones. Thus we obtain the result of the expectation of sample mean will deviate from the true parameter no larger than $o(\max_{(r_i,r_j)\in\mathcal{N}_d} p_i - p_j)$. The second part is nothing but bounding the probability of sample mean deviate from its expectation. Note that we choose the proper t in Berstein's inequality to make sure the error probability will still be desirably small after the union bound. Hence we complete the proof.

G Proof of Lemma 5.2

Without loss of generality, we assume that π_u is the identity permutation and node u belongs to the first community. Also, the access index is denoted as $\boldsymbol{i}_u \triangleq (u, i_2, \ldots, i_d)$ and $M_p(t) \triangleq pe^t + 1 - p$ is the MGF of a $\mathrm{Ber}(p)$ random variable. We have

$$\mathbb{P}\left\{\widehat{\sigma}_u(u) \ne 1 \text{ and } E_u\right\} \le \sum_{l\ne 1} p_l$$

where E_u is the event (23) of a good initialization and, on E_u, p_l is defined as the probability of the following error event.

$$\left\{\widehat{L}_u(\widehat{\sigma}_u, l; \mathbf{A}) \geq \widehat{L}_u(\widehat{\sigma}_u, 1; \mathbf{A})\right\} \tag{30}$$

Recall that the initial method $\widehat{\sigma}_u$ determines all the assignments except for the u-th node before the refining process. We write $\boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(k_u)$ to indicate the fact that now the community relation r within nodes $u, i_2, \ldots, i_d$ depends on the label of node u, which is to be decided. Similarly, we denote the estimated connection probabilit parameter $\widehat{p}$ as $\widehat{p}(k_u)$. Then, (30) is equivalent to

$$\left\{\sum_{\boldsymbol{i}_u} \mathrm{A}_{\boldsymbol{i}_u} \log \frac{\widehat{p}(l)(1-\widehat{p}(1))}{\widehat{p}(1)(1-\widehat{p}(l))} + \log \frac{1-\widehat{p}(1)}{1-\widehat{p}(l)} \geq 0\right\}$$

Note that the summation is over all possible $i_2 < \cdots < i_d$. We can also write (30) in the form of pairwise comparison. Specifically, let $\widehat{\nu}_{ij} = \widehat{\nu}_{ij}(1,l) \triangleq \log \frac{\widehat{p}_i(l)(1-\widehat{p}_j(1))}{\widehat{p}_j(1)(1-\widehat{p}_i(l))}$ and $\widehat{\lambda}_{ij} = \widehat{\lambda}_{ij}(1,l) \triangleq \log \frac{1-\widehat{p}_i(1)}{1-\widehat{p}_j(l)}$. The error event is thus further equal to

$$\left\{\sum_{(r_i,r_j)\in\mathcal{N}_d}\Big(\sum_{\substack{\boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(1)=r_i \\ \boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(l)=r_j}} \widehat{\nu}_{ij}\mathrm{A}_{\boldsymbol{i}_u} + \widehat{\lambda}_{ij} + \sum_{\substack{\boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(1)=r_j \\ \boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(l)=r_i}} \widehat{\nu}_{ji}\mathrm{A}_{\boldsymbol{i}_u} + \widehat{\lambda}_{ji}\Big) \geq 0\right\} \tag{31}$$

Note that the inner two summations will contain $n_{ij}^{(1,l)}$ and $n_{ji}^{(1,l)}$ random variables, respectively, where

$$n_{ij}^{(1,l)} \triangleq \left|\left\{\boldsymbol{i}_u \mid \boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(1) = r_i \text{ and } \boldsymbol{i}_u \overset{\widehat{\sigma}_u}{\sim} r(l) = r_j\right\}\right|$$

Observe that not all $\mathrm{A}_{\boldsymbol{i}_u}$ in the summand associated with $n_{ij}^{(1,l)}$ would really be $\mathrm{Ber}(p_i)$. The reason is that there are still a few nodes misclassified by the initialization $\widehat{\sigma}_u$. Nevertheless, since we require that $\widehat{\sigma}_u$ satisfy Condition 5.1, it can be shown that there are only $o(1)n_{ij}^{(1,l)}$ of random variables in the summand associated with $n_{ij}^{(1,l)}$ are not $\mathrm{Ber}(p_i)$. Therefore, we apply the Chernoff bound on $\mathbb{P}\{(31)\}$ to obtain

$$\mathbb{P}\{(31)\} \leq \prod_{(r_i,r_j)\in\mathcal{N}_d} (\text{Part 1} \cdot \text{Part 2}) \tag{32}$$

where

$$\text{Part 1} = \exp\Big(-\frac{1}{2}\widehat{\lambda}_{ji}(n_{ji}^{(1,l)} - n_{ij}^{(1,l)})\Big) \cdot M_{p_j}\big(\frac{\widehat{\nu}_{ij}}{2}\big)^{n_{ji}^{(1,l)}} M_{p_i}\big(\frac{-\widehat{\nu}_{ij}}{2}\big)^{n_{ij}^{(1,l)}}$$

and

$$\text{Part 2} = \left[\sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{M_p(\frac{\widehat{\nu}_{ij}}{2})}{M_j(\frac{\widehat{\nu}_{ij}}{2})}\right]^{O(K\gamma)n_{ji}^{(1,l)}} \cdot \left[\sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{M_p(-\frac{\widehat{\nu}_{ij}}{2})}{M_i(\frac{-\widehat{\nu}_{ij}}{2})}\right]^{O(K\gamma)n_{ij}^{(1,l)}}$$

First, since the parameter space we consider is a approximately equal-size one, each community has a size $(1 \pm o(1))n'$. In addition, Condition 5.1 makes sure that the community size generated from $\widehat{\sigma}_u$ will still lie in $(1 \pm o(1))n'$. Thus, it is easy to find that

$$n_{ij}^{(1,l)} \asymp n_{ji}^{(1,l)} \asymp m_{r_i r_j} \ \forall l \neq 1$$

Moreover, by a similar combinatorial argument as in our proof of Lemma 5.1, we know that the proportion of wrongly added random variables is $O(K\gamma)$. That is the reason we use $O(K\gamma)n_{ij}^{(1,l)}$ for the number of wrongly added random variables.

In the following, we are going to show that Part 1 can be upper bounded by $\exp(-(1-o(1))m_{r_ir_j}I_{p_ip_j})$ and Part 2 can be upper bounded by a vanishing term with respect to Part 1. With a similar technique as in (Gao et al., 2017), we could immediately prove that

$$\text{Part 1} \leq \exp(-(1-o(1))m_{r_ir_j}I_{p_ip_j}) \tag{33}$$

For the second part, we have, for all $i < j$,

$$\left|e^{\frac{\widehat{\nu}_{ij}}{2}} - 1\right| \vee \left|e^{-\frac{\widehat{\nu}_{ij}}{2}} - 1\right| \le C_2 \frac{p_i - p_j}{p_i}$$

for some constant $C_2 > 0$. Thus,

$$\begin{aligned}\sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{M_p(\frac{\widehat{\nu}_{ij}}{2})}{M_j(\frac{\widehat{\nu}_{ij}}{2})} = 1 + \sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{(p-p_j)(e^{\frac{\widehat{\nu}_{ij}}{2}}-1)}{p_j e^{\frac{\widehat{\nu}_{ij}}{2}} + 1 - p_j} &\le 1 + O\Big(\sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{(p-p_j)(p_i-p_j)}{p_i}\Big) \\ &\le \exp\left(O\Big(\sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{(p-p_j)(p_i-p_j)}{p_i}\Big)\right)\end{aligned}$$

The second term of Part 2 can be bounded similarly. Together, Part 2 is upper bounded by

$$\exp\left(O(K\gamma) m_{r_ir_j} \sup_{p\in\{p_1,\ldots,p_{\mathcal{K}_d}\}} \frac{(p-p_j)(p_i-p_j)}{p_i}\right) = \exp\left(o(1) m_{r_ir_j} \max_{(r_i,r_j)\in\mathcal{N}_d} I_{r_ir_j}\right) \tag{34}$$

Note that this term will still be absorbed to the term corresponding to $\max_{(r_i,r_j)\in\mathcal{N}_d} I_{r_ir_j}$ since $K = O(1)$. Combining (33) and (34) into (32), we complete the proof.

H Proof of Lemma 4.1

In this section, we're going to proof Lemma 4.1. First we state the lemma we are going to use.

Lemma 8.1: *For independent Bernoulli random variables $X_u \sim Ber(p_u)$ and $p = \frac{1}{n}\sum_{u\in[n]} p_u$, we have*

$$\mathbb{P}\Big(\sum_{u\in[n]}(X_u - p_u) \ge t\Big) \le \exp\big(t - (np+t)\log(1+\frac{t}{np})\big), \forall t \ge 0$$

This lemma is Corollary A.1.10 in (Alon and Spencer, 2004).

Lemma 8.2: *Consider the matrix $\mathbf{A}_\text{H}$ derived from the unnormalized graph Laplacian for a hypergraph. Suppose $\max_{u\in[n]}\sum_{v\in[n]}(\text{A}_\text{H})_{uv} \le \tilde{d}$ and for any $S, T \subset [n]$, one of the following statements hold with some constant $C > 0$:*

1. $\frac{e(S,T)}{|S||T|\frac{\tilde{d}}{n}} \le C$
2. $e(S,T)\log(\frac{e(S,T)}{|S||T|\frac{\tilde{d}}{n}}) \le C|T|\log\frac{n}{|T|}$

where $e(S,T) = \sum_{u\in S}\sum_{v\in T}(\text{A}_\text{H})_{uv}$. Then, $\sum_{(u,v)\in U} x_u(\text{A}_\text{H})_{uv}y_v \le C'\sqrt{\tilde{d}}$ uniformly over all unit vectors x, y, where $U = \left\{(u,v) \mid |x_uy_v| \ge \frac{\sqrt{\tilde{d}}}{n}\right\}$ and $C' > 0$ is some constant.

Note that this is the direct result to the lemma 21 in (Chin et al., 2015).

Lemma 8.3: *For any $\tau > C\left(n^{d-1}p_1 + 1\right)$ with some sufficiently large $C > 0$, we have*

$$\left|\{u \in [n] \mid d_u \ge \tau\}\right| \le \frac{n}{\tau}$$

with probability at least $1 - \exp(-C'n)$ for some constant $C' > 0$.

Proof. Note that in this lemma, the edges $e(S)$ and $e(S, S^c)$ are counting the real hyperedges in $\mathbf{A}$. This is different with the definition in Lemma 8.2.

Let us consider a subset of nodes $S \subset [n]$ which contains all nodes with degree at least τ and $|S| = l$ for some $l \in [n]$. By the requirement on S, we have either $e(S) \ge C_1 l\tau$ or $e(S, S^c) \ge C_1 l\tau$ for some constant C_1. We want to show that both $\mathbb{P}\{e(S) \geq C_1 l\tau\}$ and $\mathbb{P}\{e(S,S^c) \geq C_1 l\tau\}$ are small. First, observe that $e(S)$ consists of $\binom{l}{d}$ Bernoulli random variables and $e(S,S^c)$ consists of $\sum_{s=1}^{d-1}\binom{n-l}{s}\binom{l}{d-s}$ Bernoulli random variables. Thus, $\mathbb{E}e(S) \leq C_2 l^d p_1$ and $\mathbb{E}e(S,S^c) \leq C_2 n^{d-1} l p_1$ for some constant C_2. Then, when $\tau > C\left(n^{d-1}p_1 + 1\right)$ for some sufficiently large $C > 0$, we have

$$\begin{aligned}
\mathbb{P}\{e(S) \geq C_1 l\tau\} &= \mathbb{P}\{e(S) - \mathbb{E}e(S) \geq C_1 l\tau - \mathbb{E}e(S)\} \\
&\leq \exp\left(C_1 l\tau - \mathbb{E}e(S) - C_1 l\tau \log\left(\frac{C_1 l\tau}{\mathbb{E}e(S)}\right)\right) && \text{by Lemma 8.1} \\
&\leq \exp\left(C_1 l\tau - C_1 l\tau \log\left(\frac{C_1\tau}{C_2 n^{d-1} p_1}\right)\right) \\
&\leq \exp\left(C_1 l\tau - C_1 l\tau \log(C_3)\right) && \text{where } C_3 = \frac{C_1 C}{C_2} \\
&\leq \exp\left(-C_4 l\tau\right) && \text{for some } C_4 > 0
\end{aligned}$$

where the last inequality holds since C_3 is sufficiently large. Similarly, the same bound applies for

$$\mathbb{P}\{e(S,S^c) \geq C_1 l\tau\}$$

Thus, we have, by Union Bound

$$\mathbb{P}\{|\{u \in [n] \mid d_u \geq \tau\}| > \xi n\} \leq \sum_{l > \xi n} 2\exp\left(l \log\left(\frac{en}{l}\right)\right) \cdot \exp\left(-C_4 l\tau\right) \leq \exp(-C_5 n)$$

where we choose $\xi = \frac{1}{\tau}$. We are done. ∎

Lemma 8.4: *Given $\tau > 0$, define the subset $J = \{u \in [n] \mid d_u \leq \tau\}$. Then for any $C' > 0$, there is some constant $C > 0$ such that*

$$\|(\mathrm{A_H})_{JJ} - \mathrm{P_H})_{JJ}\|_{\mathrm{op}} \leq C\left(\sqrt{n^{d-1}p_1} + \sqrt{\tau} + \frac{n^{d-1}p_1}{\sqrt{n^{d-1}p_1} + \sqrt{\tau}}\right)$$

with probability at least $1 - n^{-C'}$

Proof. By definition,

$$\|(\mathrm{A_H})_{JJ} - (\mathrm{P_H})_{JJ}\|_{\mathrm{op}} = \sup_{x,y \in S^{n-1}} \sum_{(u,v) \in J \times J} x_u\big((\mathrm{A_H})_{uv} - (\mathrm{P_H})_{uv}\big)y_v$$

where x, y are some unit vectors lying on the unit sphere S^{n-1} in $\mathbb{R}^{n-1}$. Define the following two sets

$$\begin{aligned}
L &= \left\{(u,v) : |x_u y_v| \leq \left(\sqrt{\tau} + \sqrt{n^{d-1}p_1}\right)/n\right\} \\
U &= \left\{(u,v) : |x_u y_v| \geq \left(\sqrt{\tau} + \sqrt{n^{d-1}p_1}\right)/n\right\}
\end{aligned}$$

Then we have

$$\begin{aligned}
\|(\mathrm{A_H})_{JJ} - (\mathrm{P_H})_{JJ}\|_{\mathrm{op}} \leq &\sup_{x,y \in S^{n-1}} \sum_{(u,v) \in L \cap J \times J} x_u\big((\mathrm{A_H})_{uv} - (\mathrm{P_H})_{uv}\big)y_v \\
&+ \sup_{x,y \in S^{n-1}} \sum_{(u,v) \in U \cap J \times J} x_u\big((\mathrm{A_H})_{uv} - (\mathrm{P_H})_{uv}\big)y_v
\end{aligned}$$

We will upper-bound these two parts separately. First we will bound the light pairs $\{(u,v) \in L\}$. A discretization argument in (Chin et al., 2015) implies that

$$\begin{aligned}
&\sup_{x,y \in S^{n-1}} \sum_{(u,v) \in L \cap J \times J} x_u\big((\mathrm{A_H})_{uv} - (\mathrm{P_H})_{uv}\big)y_v \\
\lesssim &\max_{x,y \in \mathcal{N}} \max_{S \subset [n]} \sum_{(u,v) \in L \cap J \times J} x_u\big((\mathrm{A_H})_{uv} - \mathbb{E}(\mathrm{A_H})_{uv}\big)y_v
\end{aligned}$$

where $\mathcal{N} \subset S^{n-1}$ and $|\mathcal{N}| \leq 5^n$. Let $r_{uv} = x_u y_v \mathbb{1}\left\{|x_u y_v| \leq \sqrt{\tilde{d}}/n\right\}$ and $\sqrt{\tilde{d}} = \sqrt{\tau} + \sqrt{n^{d-1}p_1}$. Then we have

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_{u<v} r_{uv}\big((\mathrm{A_H})_{uv} - \mathbb{E}(\mathrm{A_H})_{uv}\big)\right| \geq C\sqrt{\tilde{d}}\right\} \\
&= \mathbb{P}\left\{\left|\sum_{u<v}\sum_{\mathbf{i}_3^d\in[n]^{d-2}} r_{uv}\left(\mathrm{A}_{u,v,\mathbf{i}_3^d} - \mathbb{E}\mathrm{A}_{u,v,\mathbf{i}_3^d}\right)\right| \geq C\sqrt{\tilde{d}}\right\} && \text{by definition} \\
&= \mathbb{P}\left\{\left|\sum_{u<v}\sum_{i_3<\ldots<i_d} r_{uv}\left(\mathrm{A}_{u,v,\mathbf{i}_3^d} - \mathbb{E}\mathrm{A}_{u,v,\mathbf{i}_3^d}\right)\right| \geq C_1\sqrt{\tilde{d}}\right\} && \text{where } C_1 = C\times(d-2)! \\
&= \mathbb{P}\left\{\left|\sum_{i_1<\ldots<i_d}\Big(\sum_{1\leq a<b\leq d} r_{i_a i_b}\Big)\left(\mathrm{A}_{\mathbf{i}_1^d} - \mathbb{E}\mathrm{A}_{\mathbf{i}_1^d}\right)\right| \geq C_1\sqrt{\tilde{d}}\right\} && \text{Simple rearrangement according to independent terms} \\
&\leq \sum_{1\leq a<b\leq d}\mathbb{P}\left\{\left|\sum_{i_1<\ldots<i_d}(r_{i_a i_b})\left(\mathrm{A}_{\mathbf{i}_1^d} - \mathbb{E}\mathrm{A}_{\mathbf{i}_1^d}\right)\right| \geq C_2\sqrt{\tilde{d}}\right\} && (a)\text{ Union bound, } C_2 = C_1/\binom{d}{2} \\
&\leq 2\sum_{1\leq a<b\leq d}\exp\Big(-\frac{1/2C_2^2\tilde{d}}{p_1\sum_{i_1<\ldots<i_d} r_{i_a i_b}^2 + \frac{2}{3}\frac{\sqrt{\tilde{d}}}{n}C_2\sqrt{\tilde{d}}}\Big) && \text{Bernstein's inequality} \\
&\leq 2\binom{d}{2}\exp\Big(-\frac{1/2C_2^2\tilde{d}}{2p_1 n^{d-2} + \frac{2}{3}\frac{\sqrt{\tilde{d}}}{n}C_2\sqrt{\tilde{d}}}\Big) && (b) \\
&\leq 2\binom{d}{2}\exp\Big(-n\frac{C_2^2}{4+\frac{4C_2}{3}}\Big) && \text{Since } \tilde{d} > p_1 n^{d-1}
\end{aligned}$$

The inequality (b) holds since $\sum_{i_a<i_b} r_{i_a i_b}^2 \leq 2\sum_{1\leq i_a<i_b\leq n} x_{i_a}^2 y_{i_b}^2 \leq 2n^{d-2}$ (recall that $\boldsymbol{x}, \boldsymbol{y}$ are all unit vectors). Then, we apply the Union Bound over the space $\mathcal{N}$ and the other half of the Laplacian matrix $\mathbf{A}_{\mathrm{H}}$, we have

$$\max_{x,y\in\mathcal{N}}\max_{S\subset[n]}\sum_{(u,v)\in L\cap J\times J} x_u\big((\mathrm{A_H})_{uv} - \mathbb{E}(\mathrm{A_H})_{uv}\big)y_v \leq C\left(\sqrt{\tau} + \sqrt{n^{d-1}p_1}\right)$$

with probability at least $1-\exp(-C'n)$. Thus we complete the bound for light pairs. Here we want to highlight that the the above argument are all similar to (Chin et al., 2015), except the key step (a). Step (a) allows us to have a similar result under the d-`hSBM` setting.

Next we show how to bound the heavy pairs $\{(u,v)\in U\}$. Same as (Gao et al., 2017), we bound

$$\sup_{x,y\in S^{n-1}}\sum_{(u,v)\in U\cap J\times J} x_u(\mathrm{A_H})_{uv}y_v \tag{35}$$

and

$$\sup_{x,y\in S^{n-1}}\sum_{(u,v)\in U\cap J\times J} x_u(\mathrm{P_H})_{uv}y_v$$

separately. By the definition of U, we have

$$\sup_{x,y\in S^{n-1}}\sum_{(u,v)\in U\cap J\times J} x_u(\mathrm{P_H})_{uv}y_v \leq \sup_{x,y\in S^{n-1}}\sum_{(u,v)\in U\cap J\times J}\frac{x_u^2 y_v^2}{|x_u y_v|}(\mathrm{P_H})_{uv} \leq \frac{n^{d-1}p_1}{\sqrt{n^{d-1}p_1}+\sqrt{\tau}}$$

The last equation hold since $(\mathrm{P_H})_{max} \leq n^{d-2}p_1$. Then, we bound (35). Note that by the definition of the set J, the degree of the sub-graph $(\mathrm{A_H})_{JJ}$ is bounded above by τ. We need to prove that the condition (the *discrepancy* property) of Lemma 8.2 is satisfied with $\tilde{d} = \tau + n^{d-1}p_1$ with probability at least $1-n^{-C'}$. The proof mainly follows the arguments in (Lei and Rinaldo, 2015) and apply the Union Bound to make sure the independence (like what we have done in (a) above) . We have

$$\sup_{x,y\in S^{n-1}}\sum_{(u,v)\in U\cap J\times J} x_u(\mathrm{A_H})_{uv}y_v \leq C\left(\sqrt{\tau} + \sqrt{n^{d-1}p_1}\right)$$

with probability at least $1-n^{-C'}$. Together with all the results above, we are done. ■

Now we are going to prove Lemma 4.1.

Proof. By triangle inequality,

$$\|\mathsf{T}_\tau(\mathrm{A_H}) - \mathrm{P_H}\|_{\mathrm{op}} \le \|\mathsf{T}_\tau(\mathrm{A_H}) - \mathsf{T}_\tau(\mathrm{P_H})\|_{\mathrm{op}} + \|\mathsf{T}_\tau(\mathrm{P_H}) - \mathrm{P_H}\|_{\mathrm{op}}$$

Then we have $\|\mathsf{T}_\tau(\mathrm{A_H}) - \mathsf{T}_\tau(\mathrm{P_H})\|_{\mathrm{op}} = \|(\mathrm{A_H})_{JJ} - (\mathrm{P_H})_{JJ}\|_{\mathrm{op}}$, which is bounded by Lemma 8.4. By Lemma 8.3, we have $|J^c| \le \frac{n}{\tau}$ with probability at least $1 - \exp(-C'n)$. This implies

$$\|\mathsf{T}_\tau(\mathrm{P_H}) - \mathrm{P_H}\|_{\mathrm{op}} \le \|\mathsf{T}_\tau(\mathrm{P_H}) - \mathrm{P_H}\|_{\mathrm{F}} \le \sqrt{2n\,|J^c|\,(p_{\mathrm{H}})^2_{\max}} \le \frac{\sqrt{2}n(p_{\mathrm{H}})_{\max}}{\sqrt{\tau}} \le \frac{\sqrt{2}n^{d-1}p_1}{\sqrt{\tau}}$$

Taking $\tau \in \left[C_1(1+n^{d-1}p_1), C_2(1+n^{d-1}p_1)\right]$. Proof completed. ∎

I Proof of Lemma 2.2

First recall that

$$\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1)) = \frac{1}{|\Theta_d^L|}\sum_{\sigma\in\Theta_d^L} \mathbb{E}_\sigma \ell(\sigma(1), \widehat{\sigma}(1))$$

In order to connect $\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1))$ with the risk function of a hypothesis testing problem, we shall find an equivalent form of $\mathbb{E}_\sigma\ell(\sigma(1),\widehat{\sigma}(1))$. The idea is to find another assignment σ' such that $d(\sigma,\sigma') = d_{\mathrm{H}}(\sigma(1),\sigma'(1)) = 1$. σ' is the most indistinguishable opponent against σ in the sense that their assignments differ by only one node. Specifically, for each $\sigma_0 \in \Theta_d^L$, we construct a new assignment $\sigma[\sigma_0]$ based on σ_0:

$$\sigma[\sigma_0](1) = \operatorname*{arg\,min}_{2\le i\le n}\left\{n_{\sigma_0(i)} = n'\right\}$$

and $\sigma[\sigma_0](i) = \sigma_0(i)$ for $2 \le i \le n$. Note that $\{i \mid n_{\sigma_0(i)} = n'\} \neq \varnothing\, \forall \sigma_0 \in \Theta_d^L$ and $\sigma[\sigma_0] \in \Theta_d^L$. In addition, for any σ_1, $\sigma_2 \in \Theta_d^L$, we can see that $\sigma_1 \neq \sigma_2$ if and only if $\sigma[\sigma_1] \neq \sigma[\sigma_2]$. Therefore, $\{\sigma_0 \mid \sigma_0 \in \Theta_d^L\} = \{\sigma_0 \mid \sigma_0 \in \Theta_d^L\}$ and thus

$$\begin{aligned}
\mathrm{R}_{\sigma\sim\mathrm{Unif}}(\widehat{\sigma}(1)) &= \frac{1}{|\Theta_d^L|}\sum_{\sigma_0\in\Theta_d^L} \mathbb{E}_{\sigma_0}\ell(\sigma_0(1),\widehat{\sigma}(1))\\
&= \frac{1}{|\Theta_d^L|}\sum_{\sigma_0\in\Theta_d^L} \frac{1}{2}\left(\mathbb{E}_{\sigma_0}\ell(\sigma_0(1),\widehat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]}\ell(\sigma[\sigma_0](1),\widehat{\sigma}(1))\right)
\end{aligned}$$

In the testing problem, we can use the optimal Bayes risk as a lower bound. Let $\widehat{\sigma}_{\mathrm{Bayes}}$ be an assignment that achieves the minimum Bayes risk $\inf_{\widehat{\sigma}} \frac{1}{2}\left(\mathbb{E}_{\sigma_0}\ell(\sigma_0(1),\widehat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]}\ell(\sigma[\sigma_0](1),\widehat{\sigma}(1))\right)$. Notice that $\widehat{\sigma}_{\mathrm{Bayes}}(1)$ is a Bayes estimator concerning the 0-1 loss, indicating that $\widehat{\sigma}_{\mathrm{Bayes}}(1)$ must to be the mode of the posterior distribution. Roughly speaking, the team who has a larger value of sum of the supporting random variables wins the test.

Grouping terms together according to each community relation, the log-likelihood function under the true community assignment σ_0 given an observation $\mathbf{A}$ becomes

$$L(\sigma_0;\mathbf{A}) = \log\mathbb{P}\{\mathbf{A} \mid \sigma_0\} = \sum_{\boldsymbol{l}=(1,l_2,\ldots,l_d)}\sum_{i=1}^{\mathcal{K}_d} \mathrm{A}_{\boldsymbol{l}}\mathbb{1}\left\{\boldsymbol{l} \overset{\sigma_0}{\sim} r_i\right\}\left(\log\frac{p_i}{\bar{p}_i} + \log\bar{p}_i\right)$$

Similarly, we can obtain the expression $L(\sigma[\sigma_0];\mathbf{A})$ when the underlying community assignment changes to $\sigma[\sigma_0]$. Hence, the probability of error is

$$\begin{aligned}
\mathbb{E}_{\sigma_0}\ell(\sigma_0(1),\widehat{\sigma}_{\mathrm{Bayes}}(1)) &= \mathbb{P}_{\sigma_0}\{L(\sigma[\sigma_0];\mathbf{A}) \ge L(\sigma_0;\mathbf{A})\}\\
&= \mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d}\sum_{u=1}^{m_{r_ir_j}}\left(X_u^{(r_j)} - X_u^{(r_i)}\right) \ge 0\right\} \qquad (36)
\end{aligned}$$

where where $f(s) \triangleq \frac{s}{1-s}$ for $C_{xy} \triangleq \log\frac{f(x)}{f(y)}$, and for each (r_i, r_j) pair in $\mathcal{N}_d$, $X_u^{(r_j)} \overset{\text{i.i.d.}}{\sim} \text{Ber}(p_j), X_u^{(r_i)} \overset{\text{i.i.d.}}{\sim} \text{Ber}(p_i)$ $\forall u = 1, \ldots, m_{r_i r_j}$ are all mutually indepedent random variables. Note that when summing over all possible $\boldsymbol{l}$'s in the log-likelihood function, the indices can be partitioned into two kinds of set: one whose label changes from r_i to r_j for some $(r_i, r_j) \in \mathcal{N}_d$ when there is exactly one node disagreement and one whose label does not change whether the community assignment is σ_0 or $\sigma[\sigma_0]$. Specifically,

$$\{\boldsymbol{l} = (1, l_2, \ldots, l_d)\} = J \cup J^c$$

where

$$J = \bigcup_{i<j:(r_i,r_j)\in\mathcal{N}_d} \left\{\boldsymbol{l} \overset{\sigma_0}{\sim} r_i, \boldsymbol{l} \overset{\sigma[\sigma_0]}{\sim} r_j\right\} \quad \text{and} \quad J^c = \bigcup_{i=1}^{\mathcal{K}_d} \left\{\boldsymbol{l} \overset{\sigma_0}{\sim} r_i, \boldsymbol{l} \overset{\sigma[\sigma_0]}{\sim} r_i\right\}$$

The former contributes to the difference between two Bernoulli random variables with cardinality $m_{r_i r_j}$, while the latter is invariant to the hypothesis testing problem and its likelihood remains the same at both sides of the first inequality in (36). Note also that we rearrange terms on the specific side of the inequality to make $C_{r_i r_j} \geq 0$ $\forall (r_i, r_j) \in \mathcal{N}_d$ due to the non-decreasing property of the probability parameters p_i's.

By symmetry, the situation is exactly the same for $\mathbb{E}_{\sigma[\sigma_0]}\ell(\sigma[\sigma_0](1), \widehat{\sigma}_{\text{Bayes}}(1))$. Finally, since (36) holds for all $\sigma_0 \in \Theta_d^L$ and $\inf(\cdot)$ is a concave function, we have

$$\begin{aligned}
\mathrm{R}_{\sigma\sim\text{Unif}}(\widehat{\sigma}(1)) &\geq \inf_{\widehat{\sigma}} \mathrm{R}_{\sigma\sim\text{Unif}}(\widehat{\sigma}(1)) \\
&= \inf_{\widehat{\sigma}} \frac{1}{|\Theta_d^L|} \sum_{\sigma_0\in\Theta_d^L} \frac{1}{2}\left(\mathbb{E}_{\sigma_0}\ell(\sigma_0(1), \widehat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]}\ell(\sigma[\sigma_0](1), \widehat{\sigma}(1))\right) \\
&\geq \frac{1}{|\Theta_d^L|} \sum_{\sigma_0\in\Theta_d^L} \inf_{\widehat{\sigma}} \frac{1}{2}\left(\mathbb{E}_{\sigma_0}\ell(\sigma_0(1), \widehat{\sigma}(1)) + \mathbb{E}_{\sigma[\sigma_0]}\ell(\sigma[\sigma_0](1), \widehat{\sigma}(1))\right) \\
&= \frac{1}{|\Theta_d^L|} \sum_{\sigma_0\in\Theta_d^L} \mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\} \\
&= \mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\}
\end{aligned}$$

J Proof of Lemma 2.3

We can break the L.H.S. of (9) dirctly into

$$\mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\} \geq \prod_{i<j:(r_i,r_j)\in\mathcal{N}_d} \mathbb{P}\left\{\sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\}$$

Note that there are only finitely many terms involving in the product since we assume the order d is constant and so does the total number of community relations $\kappa_d = |\mathcal{K}_d|$ in d-`hSBM`. Though naïve, we could still arrive at the same order as the minimax rate. By symmetry, it suffices to focus on the first term in the above equation.

$$\mathbb{P}\left\{\sum_{u=1}^{m_{r_1 r_2}} C_{r_1 r_2}\left(X_u^{(r_2)} - X_u^{(r_1)}\right) \geq 0\right\}$$

Here, we utilize a result from large deviation.

Conseder i.i.d. random variables $\{X_i\}_{i=1}^n$ where each $X_i \sim X$. We assume X is nondegenerate and that

$$\mathbb{E}X^2 e^{\lambda X} < \infty \tag{37}$$

for some $\lambda > 0$. The former condition ensures, for $0 < u \leq \lambda$, the existence of the functions $m(u) \triangleq \left(\log L_X(u)\right)'$, $\sigma^2(u) \triangleq m'(u)$ and $Q(u) \triangleq u m(u) - \log L_X(u)$ where $L_X(u) \triangleq \mathbb{E}e^{uX}$ is the *Moment Generating Function* (MGF) of the random variable X. Recall some known results:

$$\lim_{u\downarrow 0} m(u) = m(0) = \mathbb{E}X < \infty$$

and

$$\sup_{0<u\leq\lambda} (ux - \log L_X(u)) = Q(u^*) \tag{38}$$

for $m(0) < x \leq m(\lambda)$, where u^* is the unique solution of the equation

$$m(u) = x \tag{39}$$

Note that it is the sup-achieving condition in (38). The main theorem goes as follows.

Theorem 10.1 (*Theorem 1* in Rozovsky, 2003)**:** *$\forall x$ such that $m(0) < x \leq m(\lambda)$ and $\forall n \geq 1$, the relation*

$$e^{-nQ(u^*)} \geq \mathbb{P}\left\{\sum_{i=1}^{n} X_i \geq nx\right\} \geq e^{-nQ(u^*)-c\left(1+\sqrt{nQ(u^*)}\right)}$$

holds, where the constant c does not depend on x and n.

The first inequality is essentially the Chernoff Bound, while here we use the second one, i.e. the lower bound result.

First, we identify that $X = C_{r_1 r_2}(X_u^{(r_2)} - X_u^{(r_1)})$ and $n = m_{r_1 r_2}$ for our problem. Besides, since $X < \infty$, we can take λ large enough so that (37) holds. The MGF now becomes

$$L_X(u) = \mathbb{E}e^{uX} = \mathbb{E}\left[e^{uC_{r_1 r_2} X_u^{(r_2)}}\right] \cdot \mathbb{E}\left[e^{-uC_{r_1 r_2} X_u^{(r_1)}}\right]$$

Also, since $m(0) = \mathbb{E}X < 0$, we make a trick here to take $x = 0$. The corresponding optimalilty condition (39) becomes

$$\begin{aligned} m(u) = x = 0 &\Leftrightarrow \frac{L'_X(u)}{L_X(u)} = 0 \\ &\Leftrightarrow L'_X(u) = 0 \end{aligned}$$

It can be shown that $u^* = \frac{1}{2}$ and the supremum achieved is

$$\begin{aligned} Q(u^*) &= \sup_{0<u\leq\lambda} (ux - \log L_X(u)) \\ &= -\log L_X(u^*) \\ &= I_{p_1 p_2} \end{aligned}$$

Combining the expressions for each $C_{r_i r_j}$ corresponding to a $(r_i, r_j) \in \mathcal{N}_d$, we can conclude that

$$\begin{aligned} &\mathbb{P}\left\{\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\} \\ \geq &\prod_{i<j:(r_i,r_j)\in\mathcal{N}_d} \mathbb{P}\left\{\sum_{u=1}^{m_{r_i r_j}} \left(X_u^{(r_j)} - X_u^{(r_i)}\right) \geq 0\right\} \\ \geq &\prod_{i<j:(r_i,r_j)\in\mathcal{N}_d} e^{-m_{r_i r_j} I_{p_i p_j} - c_{r_i r_j}\left(1+\sqrt{m_{r_i r_j} I_{p_i p_j}}\right)} \\ = &\exp\left(-\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \left(m_{r_i r_j} I_{p_i p_j} + c_{r_i r_j}\left(1+\sqrt{m_{r_i r_j} I_{p_i p_j}}\right)\right)\right) \\ \geq &\exp\left(-\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} \left(m_{r_i r_j} I_{p_i p_j} + c\left(1+\sqrt{m_{r_i r_j} I_{p_i p_j}}\right)\right)\right) \end{aligned}$$

where $c = \max_{i<j:(r_i,r_j)\in\mathcal{N}_d}\{c_{r_i r_j}\}$ is independent of n' and hence n. Finally, since we assume that $\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j}$ goes to infinity as n becomes large, the second term with the constant c in the above equation would be dominated by the first term. We have the desired asymtotic result consequently.