Community Detection in Hypergraphs: Optimal Statistical Limit and Efficient Algorithms

I (Eli) Chien
University of Illinois, Urbana Champaign

Chung-Yi Lin
National Taiwan University

I-Hsiang Wang
National Taiwan University

Abstract

In this paper, community detection in hypergraphs is explored. Under a generative hypergraph model called "d-wise hypergraph stochastic block model" (d-hSBM) which naturally extends the Stochastic Block Model (SBM) from graphs to d-uniform hypergraphs, the fundamental limit on the asymptotic minimax misclassified ratio is characterized. For proving the achievability, we propose a two-step polynomial time algorithm that provably achieves the fundamental limit in the sparse hypergraph regime. For proving the optimality, the lower bound of the minimax risk is set by finding a smaller parameter space which contains the most dominant error events, inspired by the analysis in the achievability part. It turns out that the minimax risk decays exponentially fast to zero as the number of nodes tends to infinity, and the rate function is a weighted combination of several divergence terms, each of which is the Rényi divergence of order 1/2 between two Bernoulli distributions. The Bernoulli distributions involved in the characterization of the rate function are those governing the random instantiation of hyperedges in d-hSBM. Experimental results on both synthetic and real-world data validate our theoretical finding.

1 INTRODUCTION

Community detection (clustering) has received great attention recently across many applications, including social science, biology, computer science, and machine learning, while it is usually an ill-posed problem due to the lack of ground truth. A prevalent way to circumvent the difficulty is to formulate it as an inverse problem on a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where each node $i \in \mathcal{V} = [n] \triangleq \{1, \ldots, n\}$ is assigned a community (label) $\sigma(i) \in [K] \triangleq \{1, \ldots, K\}$ that serves as the ground truth. The ground-truth *community assignment* $\sigma : [n] \to [K]$ is hidden while the graph $\mathcal{G}$ is revealed. Each edge in the graph models a certain kind of *pairwise* interaction between the two nodes. The goal of community detection is to determine σ from $\mathcal{G}$, by leveraging the fact that different combination of community relations leads to different likeliness of edge connectivity. A canonical statistical model is the *stochastic block model* (SBM) (Holland et al., 1983) (also known as planted partition model (Condon and Karp, 2001)) which generates randomly connected edges from a set of labeled nodes. The presence of the $\binom{n}{2}$ edges is governed by $\binom{n}{2}$ independent Bernoulli random variables, and the parameter of each of them depends on the community assignments of the two nodes in the corresponding edge.

Through the lens of statistical decision theory, the fundamental statistical limits of community detection provides a way to benchmark various community detection algorithms. Under SBM, the fundamental statistical limits have been characterized recently. One line of work takes a Bayesian perspective, where the unknown labeling σ of nodes in $\mathcal{V}$ is assumed to be distributed according to certain prior, and one of the most common assumption is i.i.d. over nodes. Along this line, the fundamental limit for exact recovery is characterized (Abbe et al., 2016) in the full generality, while partial recovery remains open in general. See the survey (Abbe, 2017) for more details and references therein. A second line of work takes a minimax perspective, and the goal is to characterize the minimax risk, which is typically the *mismatch ratio* between the true community assignment and the recovered one. In (Zhang and Zhou, 2016), a tight asymptotic characterization of the minimax risk for community detection in SBM is found. Along with these theoretical results, several algorithms have been proposed

Proceedings of the 21^{st} International Conference on Artificial Intelligence and Statistics (AISTATS) 2018, Lanzarote, Spain. PMLR: Volume 84.

to achieve these limits, including degree-profiling comparison (Abbe and Sandon, 2015) for exact recovery, spectral MLE (Yun and Proutiere, 2015) for almost-exact recovery, and a two-step mechanism (Gao et al., 2017) under the minimax framework.

However, graphs can only capture pairwise relational information, while such dyadic measure may be inadequate in many applications, such as the task of 3-D subspace clustering (Agarwal et al., 2005) and the higher-order graph matching problem in computer vision (Duchenne et al., 2011). Therefore, it is natural to model such *beyond-pairwise* interaction by a hyperedge in a hypergraph and study the clustering problem in a hypergraph setting. *Hypergraph partitioning* has been investigated in computer science, and several algorithms have been proposed, including spectral methods based on clique expansion (Agarwal et al., 2006), hypergraph Laplacian (Zhou et al., 2006), tensor method (Ghoshdastidar and Dukkipati, 2015), linear programming (Li et al., 2016), to name a few. Existing approaches, though, mainly focus on optimizing a certain score function entirely based on the connectivity of the observed hypergraph and do not view it as a statistical estimation problem.

In this paper, we investigate the community detection problem in hypergraphs through the lens of statistical decision theory. Our goal is to characterize the fundamental statistical limit and develop computationally feasible algorithms to achieve it. As for the generative model for hypergraphs, one natural extension of the SBM model to a hypergraph setting is the *hypergraph stochastic block model* (hSBM), where the presence of an order-h hyperedge $\mathsf{e} \subset \mathcal{V}$ (i.e. $|\mathsf{e}| = h \leq M$, the maximum edge cardinality) is governed by a Bernoulli random variable with parameter θ_{e} and the presence of different hyperedges are mutually independent. Despite the success of the aforementioned algorithms applied on many practical datasets, it remains open how they perform in hSBM since the the fundamental limits have not been characterized and the probabilistic nature of hSBM has not been fully utilized.

The hypergraph stochastic block model is first introduced in (Ghoshdastidar and Dukkipati, 2014) as the planted partition model in random uniform hypergraphs where each hyperedge has the same cardinality. The uniform assumption is later relaxed in a follow-up work (Ghoshdastidar and Dukkipati, 2017) and a more general hSBM with mixing edge orders is considered. In (Angelini et al., 2015), the authors consider the sparse regime and propose a spectral method based on a generalization of non-backtracking operator. Besides, a weak consistency condition is derived in (Ghoshdastidar and Dukkipati, 2017) for hSBM by using the hypergraph Laplacian. Departing from SBM, an extension to the *censored block model* to the hypergraph setting is considered in (Ahn et al., 2016), where an information theoretic limit on the sample complexity for exact recovery is characterized.

As a first step towards characterizing the fundamental limit of community detection in hypergraphs, in this work we focus on the "d-wise hypergraph stochastic block model" (d-hSBM), where all hyperedges generated in the hypergraph stochastic block model are of order d. Our main contributions are two-fold.First, we give a tight asymptotic characterization of the optimal minimax risk in d-hSBM for any d. Second, we propose a polynomial time algorithm which provably achieves the minimax risk under mild regularity conditions. Throughout the paper, the order d and the number of communities K are both treated as constants, while other parameters (hyperedge connection probability) may be coupled with n. The proposed algorithm consists of two steps. The first step is a global estimator that roughly recovers the hidden community assignment to a certain precision level, and the second step refines the estimated assignment based on the underlying probabilistic model. This *refine-after-initialize* concept has also been used in graph clustering (Abbe and Sandon, 2015; Yun and Proutiere, 2015; Gao et al., 2017) and ranking (Chen and Suh, 2015). The proposed algorithm performs well on both synthetic data and real-world data. The experimental results validate the theoretical finding that not only is the refinement step critical in achieving the optimal statistical limit, but it is significantly better to use hypergraphs for community detection problem rather than graphs.

The characterized minimax risk in d-hSBM is an exponential rate, and the error exponent turns out to be a linear combination of Rényi divergences of order $1/2$. Each divergence term in the sum corresponds to a pair of community relations that would be confused with one another when there is only one misclassification, and the weighted coefficient associated with it indicates the total number of such confusing patterns. Probabilistically, there may well be two or more misclassifications, with each confusing relation pair pertaining to a Rényi divergence when analyzing the error probability. However, we demonstrate technically that these situations are all dominated by the error event with a single misclassified node, which leaves out only the "neighboring" divergence terms in the asymptotic expression. The main technical challenge resolved in this work is attributed to the fact that the community relations become much more complicated as the order d increases, meaning that more error events may arise compared to the much simpler homogeneous graph SBM case. In the proof of achievability, we show that the re-

finement step is able to achieve the fundamental limit provided that the initilization step satisfies a certain weak consistency condition. The converse part of the minimax risk follows a standard approach in statistics by finding a smaller parameter space where we can analyze the risk.

Finally, we would like to note that an extended version (especially when K is allowed to scale with n) of this paper can be found online (Chien et al., 2018).

2 PROBLEM FORMULATION

2.1 Community Relations

Let $\mathcal{K}_d \triangleq \{r_i\}$ be the set of all possible community relations under d-hSBM and $\kappa_d \triangleq |\mathcal{K}_d|$. Contrary to the dichotomy situation (same-community or not) concerning the appearance of an edge between two nodes in a symmetric graph SBM, there is a multitude of community relations in d-hSBM. In order not to mess up with them as d increases, we use the idea of majorization to organize $\mathcal{K}_d$ with each r_i in the form of a histogram. Specifically, the histogram operator $\mathsf{hist}(\cdot)$ is used to transform a vector $\boldsymbol{r} \in [K]^d$ into its histogram vector $\mathsf{hist}(\boldsymbol{r})$. For convinience, we sort the histogram vector in descending order and append zero's to make $\mathsf{hist}(\boldsymbol{r})$ remain d-dimensional. The notion of majorization is introduced as follows. For any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, we say that $\boldsymbol{a}$ majorizes $\boldsymbol{b}$, written as $\boldsymbol{a} \succ \boldsymbol{b}$, if $\sum_{i=1}^k a_i^{\downarrow} \geq \sum_{i=1}^k b_i^{\downarrow}$ for $k = 1, \ldots, d$ and $\sum_{i=1}^d a_i = \sum_{i=1}^d b_i$, where $x_i^{\downarrow}$'s are elements of $\boldsymbol{x}$ sorted in descending order. Observe that each relation $r_i \in \mathcal{K}_d$ can be uniquely represented by it histogram counterpart $\boldsymbol{h}_i$. We arrange $\mathcal{K}_d$ in majorization (pre)order such that $\boldsymbol{h}_i \succ \boldsymbol{h}_j \Leftrightarrow i < j$.

Example 2.1 ($\mathcal{K}_4$ in 4-hSBM)**:** $|\mathcal{K}_4| = \kappa_4 = 5$ *with histogram vectors being*

	Relation	Histogram	Connecting Probability
r_1	(all-same)	$h_1 = (4,0,0,0)$	p_1
r_2	(only-1-diff)	$h_2 = (3,1,0,0)$	p_2
r_3		$h_3 = (2,2,0,0)$	p_3
r_4	(only-2-same)	$h_4 = (2,1,1,0)$	p_4
r_5	(all-diff)	$h_5 = (1,1,1,1)$	p_5

2.2 Probabilistic Model: d-hSBM

In a d-uniform hypergraph, the adjacency relation is indicated by a d-dimensional $n \times \cdots \times n$ random tensor $\mathbf{A} \triangleq [\mathrm{A}_{\boldsymbol{l}}]$ (the size of each dimension being n), where $\boldsymbol{l} = (l_1, \ldots, l_d) \in [n]^d$ is the access index of an element in the tensor. Let $\boldsymbol{x}_\pi \triangleq (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for a permutation $\pi \in \mathcal{S}_n$ denote the permuted version of a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Also, $\mathcal{S}_n$ is the symmetric group of degree n which contains all the permutations from $[n]$ to itself. The following two natural conditions on this adjacency tensor come from hypergraph:

$$\begin{aligned} &\text{No self-loop:} \quad \mathrm{A}_{\boldsymbol{l}} \neq 0 \iff |\{l_1, \ldots, l_d\}| = d. \\ &\text{Symmetry:} \quad \mathrm{A}_{\boldsymbol{l}} = \mathrm{A}_{\boldsymbol{l}_\pi} \quad \forall \pi \in \mathcal{S}_d. \end{aligned}$$

In d-SBM, $\mathrm{A}_{\boldsymbol{l}}$ is a Bernoulli random variable with success probability $\mathrm{Q}_{\boldsymbol{l}}$ for each $\boldsymbol{l}$. The parameter tensor $\mathbf{Q} \triangleq [\mathrm{Q}_{\boldsymbol{l}}]$ depends on the community assignment and forms a block structure. The block structure is characterized by a symmetric community connectivity d-dimensional tensor $\mathbf{B} \in [0,1]^{K \times \cdots \times K}$ where $\mathrm{Q}_{\boldsymbol{l}} = \mathrm{B}_{\sigma(\boldsymbol{l})}$. Here the function $\sigma(\boldsymbol{x}) \triangleq (\sigma(x_1), \ldots, \sigma(x_n))$ is the community label vector assigned by σ for a node vector $\boldsymbol{x} = (x_1, \ldots, x_n)$. Let $n_k = \big|\{i | \sigma(i) = k\}\big|$ be the size of the k-th community for $k \in [K]$. In addition, let $\boldsymbol{p} = (p_1, \ldots, p_{\kappa_d}) \in (0,1)^{\kappa_d}$ where p_i is to denote the success probability of the Bernoulli random variable that corresponds to the appearance of a hyperedge with relation $r_i \in \mathcal{K}_d$. We assume, without loss of generality, that $p_i \geq p_j \ \forall i < j$. The more concentrated a group is, the higher the chances that the members will be connected by a hyperedge. To guarantee the solvability of weak recovery in d-hSBM, we set the probability parameter $\boldsymbol{p}$ at least in the order of $\Omega(1/n^{d-1})$. Therefore, we would write $\boldsymbol{p} = \frac{1}{n^{d-1}}(a_1, \ldots, a_{\kappa_d})$ where $a_i = \Omega(1) \, \forall i = 1, \ldots, \kappa_d$. The sparse regime $\Omega(1/n^{d-1})$ considered here is first motivated in (Lin et al., 2017). Under 3-hSBM, the authors in (Lin et al., 2017) consider $\boldsymbol{p} = \Theta(1/n^2)$, which is orderwise-lower than the one (i.e. $\Theta(1/n)$) required for partial recovery (Abbe and Sandon, 2015) and the minimax risk (Zhang and Zhou, 2016) under SBM.

The parameter space that we consider is a *homogeneous* and *approximately equal-sized* case where each n_k is roughly $\lfloor \frac{n}{K} \rfloor$. Formally speaking (let $n' \triangleq \lfloor \frac{n}{K} \rfloor$),

$$\Theta_d^0(n, K, \boldsymbol{p}, \eta) \triangleq \Big\{ (\mathbf{B}, \sigma) \mid \sigma : [n] \to [K], \\ n_k \in [(1-\eta)n', (1+\eta)n'] \ \forall k \in [K] \Big\}$$

where $\mathbf{B}$ has the property that $\mathrm{B}_{\sigma(\boldsymbol{l})} = p_i$ if and only if $\mathsf{hist}(\sigma(\boldsymbol{l})) = \boldsymbol{h}_i$. In other words, only the histogram of the community labels within a group matters when it comes to connectivity. η is a parameter that controls how much n_k could vary. We assume the more interesting case that $\eta \geq \frac{1}{n'}$ where the community sizes are not restricted to be exactly equal. Interchangeably, we would write $\boldsymbol{l} \overset{\sigma}{\sim} r_i$ to indicate the community relation within nodes $l_1, \ldots, l_d$ under the assignment σ.

2.3 Performance Measure

To gauge how good an estimator $\widehat{\sigma} : \mathcal{G} \to [K]^n$ is, we use the *mismatch ratio* as the performance measure to the community detection problem. The un-permuted loss function is defined as $\ell_0(\sigma_1, \sigma_2) \triangleq \frac{1}{n} d_{\mathrm{H}}(\sigma_1, \sigma_2)$ where d_{H} is the Hamming distance. It directly counts the proportion of misclassified nodes between an estimator and the ground truth. Concerning the issue of possible re-labeling, the mismatch ratio is defined as the loss function which maximizes the agreements between an estimator and the ground truth after an alignment by label permutation.

$$\ell(\widehat{\sigma}, \sigma) \triangleq \min_{\pi \in \mathcal{S}_K} \ell_0(\widehat{\sigma}_\pi, \sigma)$$

As convention, we use $\mathrm{R}_\sigma(\widehat{\sigma}) \triangleq \mathbb{E}_\sigma \ell(\widehat{\sigma}, \sigma)$ to denote the corresponding risk function. Finally, the minimax risk for the parameter space $\Theta_d^0(n, K, \boldsymbol{p}, \eta)$ under d-hSBM is denoted as

$$\mathrm{R}_d^* \triangleq \inf_{\widehat{\sigma}} \sup_{(\mathbf{B},\sigma) \in \Theta_d^0} \mathrm{R}_\sigma(\widehat{\sigma})$$

3 MAIN CONTRIBUTIONS

For the case $d = 2$, the asymptotic minimax risk R_2^* is characterized in (Zhang and Zhou, 2016), which decays to zero exponentially fast as $n \to \infty$. In addition, the (negative) exponent of R_2^* is determined by n' and the Rényi divergence of order $1/2$ between two Bernoulli distributions $\mathrm{Ber}(p)$ and $\mathrm{Ber}(q)$

$$I_{pq} \triangleq -2 \log \left(\sqrt{pq} + \sqrt{1-p}\sqrt{1-q} \right).$$

It turns out the worst-case risk of our proposed algorithm also decays to zero exponentially fast, given that the outcome of the initialization algorithm satisfies certain conditions. The exponent is a weighted combination of divergence terms. To specify the weight, we introduce further notations below. We use $\mathcal{N}_d \triangleq \left\{ (r_i, r_j) \mid i < j,\ \|h_i - h_j\|_1 = 2 \right\}$ to denote the collection of ordered pairs of relations in $\mathcal{K}_d$ that are at a one-hop distance to each other where $\|\cdot\|_1$ stands for the ℓ_1 norm. There is a weighted coefficient associated with every pairwise divergence term. Appearing in the hypothesis testing problem when deriving the minimax lower bound, it represents the number of error events arising from confusing relation r_i with r_j for each pair $(r_i, r_j) \in \mathcal{N}_d$. It turns out that this situation also happens for any order d. Precisely, let's consider a least favorable sub-parameter space of Θ_d^0:

$$\Theta_d^L(n, K, \boldsymbol{p}, \eta) \triangleq \Big\{ (\mathbf{B}, \sigma) \in \Theta_d^0 \mid \forall k \in [K] \quad n_k \in \{n' - 1, n', n' + 1\},\ n_{\sigma(1)} = n' + 1 \Big\} \tag{1}$$

In Θ_d^L, each community takes on only 3 possible sizes and there are exactly $n' + 1$ members in the community where the first node belongs. We pick a σ_0 in Θ_d^L and construct a new assignment $\sigma[\sigma_0]$ based on σ_0: $\sigma[\sigma_0](i) = \sigma_0(i)$ for $2 \le i \le n$ and $\sigma[\sigma_0](1) = \arg\min_{2 \le k \le K} \{n_k = n'\}$. In other words, $\sigma[\sigma_0]$ and σ_0 only disagree on the label of the first node. For each pair $(r_i, r_j) \in \mathcal{N}_d$, we define the weighted coefficient

$$m_{r_i r_j} \triangleq \left| \left\{ \boldsymbol{l} = (1, l_2, \ldots, l_d) \mid \boldsymbol{l} \overset{\sigma_0}{\sim} r_i, \boldsymbol{l} \overset{\sigma[\sigma_0]}{\sim} r_j \right\} \right|$$

as the number of how many r_i-edges do we mistake as r_j-edges. Note that the above definition is independent of the choice of $\sigma_0 \in \Theta_d^L$.

Example 3.1 ($\mathcal{N}_4$ in 4-hSBM)**:** $|\mathcal{N}_4| = 5$ *with elements*

Relation Pair	Combinatorial Number
(r_1, r_2)	$m_{r_1 r_2} \asymp \binom{n'}{3}$
(r_2, r_3)	$m_{r_2 r_3} \asymp \binom{n'}{2} n'$
(r_3, r_4)	$m_{r_3 r_4} \asymp n'(K-2)\binom{n'}{2}$
(r_2, r_4)	$m_{r_2 r_4} \asymp \binom{n'}{2}(K-2)n'$
(r_4, r_5)	$m_{r_4 r_5} \asymp n'\binom{K-2}{2}(n')^2$

Note: $m_{r_1 r_2}$ is the smallest while $m_{r_4 r_5}$ is the largest.

Here the asymptotic equality between two functions $f(n)$ and $g(n)$, denoted as $f \asymp g$ (as $n \to \infty$), holds if $\lim_{n\to\infty} f(n)/g(n) = 1$. The asymptotic optimal minimax risk for the parameter space $\Theta_d^0(n, K, \boldsymbol{p}, \eta)$ under d-hSBM is characterized by the following theorem.

Theorem 3.1 (Main Theorem)**:** *Suppose as $n \to \infty$,*

$$\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j} \to \infty \tag{2}$$

Then

$$\log \mathrm{R}_d^* \asymp - \sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j}$$

The converse part of Theorem 3.1 is established via the following lower bound on the minimax risk.

Theorem 3.2: *If $\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j} \to \infty$ as $n \to \infty$, then there exists a positive sequence $\zeta_n \to 0$ as $n \to \infty$ such that*

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

We would like to note that the minimax result obtained in (Gao et al., 2017) can be recovered from our main theorem by specializing $d = 2$. Their condition can be identified with (2) by using the approximation $I_{p_1 p_2} \asymp \frac{(a_1 - a_2)^2}{n a_1}$. Moreover, the only weighted coefficient associated with the community relation pair $(r_1, r_2) \in \mathcal{N}_2$ under $\Theta_2^0(n, K, \boldsymbol{p}, \eta)$ is $m_{r_1 r_2} = n' \asymp \frac{n}{k}$.

4 TWO-STEP ALGORITHM

4.1 Refinement Step

The refinement step (Algorithm 1) comprises two major parts. First, for each node $u \in [n]$, we generate an estimated assignment $\widehat{\sigma}_u$ of all nodes except u by applying an initialization algorithm Alg_{init} on the sub-hypergraph without the vertex u. The sub-hypergraph is represented by $\mathbf{A}_{-u}$, which is the $(n-1)\times\cdots\times(n-1)$ sub-tensor of $\mathbf{A}$ when the u-th coordinate is removed in each dimension. Then, the label of u under $\widehat{\sigma}_u$ is determined by maximizing a local likelihood function. Specifically, let us start with the *global* likelihood function defined as follows. Let

$$L(\sigma;\mathbf{A}) \triangleq \sum_{\{i|r_i\in\mathcal{K}_d\}} \sum_{\{\boldsymbol{l}|\boldsymbol{l}\overset{\sigma}{\sim}r_i\}} \Big(\log p_i \mathrm{A}_{\boldsymbol{l}} + \log(1-p_i)(1-\mathrm{A}_{\boldsymbol{l}})\Big)$$

denote the log-likelihood of an adjacency tensor $\mathbf{A}$ when the hidden community structure is determined by σ. For each $u \in [n]$, we use

$$L_u(\sigma,k;\mathbf{A}) \triangleq \sum_{\{i|r_i\in\mathcal{K}_d\}} \sum_{\{\boldsymbol{l}|l_1=u,\boldsymbol{l}\overset{\sigma}{\sim}r_i\}} \Big(\log p_i \mathrm{A}_{\boldsymbol{l}} + \log(1-p_i)(1-\mathrm{A}_{\boldsymbol{l}})\Big)$$

to denote those likelihood terms pertaining to the u-th node when its label is k. Since $\mathbf{A}$ is symmetric, we can assume without loss of generality that the first index in $\boldsymbol{l}$ is u. Based on the estimated assignment of the other $n-1$ nodes, we make use of the following *local Maximum Likelihood Estimation* method

$$\widehat{\sigma}(u) \triangleq \arg\max_{k\in[K]} L(\sigma,k;\mathbf{A})$$

to predict the label of u. While the parameter $\mathbf{B}$ that governs the underlying random hypergraph model is unknown when evaluating the likelihood, we will use $\widehat{L}(\sigma;\mathbf{A})$ and $\widehat{L}_u(\sigma_u,k_u;\mathbf{A})$ to denote the global and local likelihood function with the true $\mathbf{B}$ replaced by its estimated counterpart $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{B}}^u$, respectively. Note that the superscript u is to indicate the fact that the estimation $\widehat{\mathbf{B}}^u$ is calculated with node u taken out.

The final step of the refinement algorithm is to form a consensus through a majority neighbor voting. The consensus step seeks for a jointly agreed community assignment among n different estimated assignments $\{\widehat{\sigma}_u : u \in [n]\}$ derived since they would all be close to the ground truth up to some permutation.

4.2 Spectral Clustering

To devise a good initialization algorithm Alg_{init}, we develop a hypergraph version of the unnormalized spectral clustering (Von Luxburg, 2007) with regularization (Chin et al., 2015). In particular, a modified version of the *hypergraph Laplacian* described below is employed. Let $\mathbf{H} = [\mathrm{H}_{ve}]$ be the $|\mathcal{V}| \times |\mathcal{E}|$ incidence matrix, where each entry H_{ve} is the indicator function whether or not node v belongs to hyperedge e. Note that the incidence matrix $\mathbf{H}$ contains the same amouont of information as the the adjacency tensor $\mathbf{A}$. Let d_u denote the degree of the u-th node, and $\bar{d}$ be the average degree across the hypergraph. The *unnormalized* hypergraph Laplacian is defined as

$$\mathcal{L}(\mathbf{A}) \triangleq \mathbf{H}\mathbf{H}^{\mathrm{T}} - \mathbf{D} \tag{5}$$

where $\mathbf{D} = \mathsf{diag}(d_1,\ldots,d_n)$ is a diagonal matrix representing the degree distribution in that hypergraph with adjacency tensor $\mathbf{A}$ and $(\cdot)^{\mathrm{T}}$ is the usual matrix transpose. Note that $\mathcal{L}$ can be thought of as an encoding of the higher-dimensional connectivity relationship into a two-dimensional matrix.

Before we directly apply the spectral method, high-degree nodes in the hypergraph is first trimmed to ensure the performance of the clustering algorithm. Specifically, we use $\mathbf{A}_\tau$ to denote the modification of $\mathbf{A}$ where all coordinates pertinent to the set $\{u \in [n] \mid d_u \geq \tau\}$ are replaced with all-zero vectors. Let $\mathbf{H}_\tau$ and $\mathbf{D}_\tau$ be the corresponding incidence matrix and degree matrix of $\mathbf{A}_\tau$. The spectrum we are looking for is the trimmed version of $\mathcal{L}$, denoted as $\mathsf{T}_\tau(\mathcal{L}(\mathbf{A})) \triangleq \mathbf{H}_\tau\mathbf{H}_\tau^{\mathrm{T}} - \mathbf{D}_\tau$ where the operator $\mathsf{T}_\tau(\cdot)$ is the trimming process with a degree threshold τ. Let

$$\mathsf{SVD}_K\left(\mathsf{T}_\tau(\mathcal{L}(\mathbf{A}))\right) \triangleq \widehat{\mathbf{U}} = \left[\boldsymbol{u}_1^{\mathrm{T}} \ \cdots \ \boldsymbol{u}_n^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{n\times K}$$

Algorithm 1: Refinement Scheme

Input: Adjacency tensor $\mathbf{A} \in \{0,1\}^{n\times\cdots\times n}$,
number of communities K,
initialization algorithm Alg_{init}.

Local MLE:

for $u = 1$ **to** n **do**

Apply Alg_{init} on $\mathbf{A}_{-u}$ to obtain $\widehat{\sigma}_u(v)\ \forall v \neq u$.

Estimate entries of $\mathbf{B}$ using the sample mean $\widehat{\mathbf{B}}^u$.

Assign the label of node u based on

$$\widehat{\sigma}_u(u) = \arg\max_{k\in[K]} \widehat{L}_u(\widehat{\sigma}_u,k;\mathbf{A}) \tag{3}$$

end

Consensus:

Define $\widehat{\sigma}(1) = \widehat{\sigma}_1(1)$. For $u = 2,\ldots,n$, define

$$\widehat{\sigma}(u) = \arg\max_{k\in[K]} \left| \{v|\widehat{\sigma}_1(v) = k\} \cap \{v|\widehat{\sigma}_u(v) = \widehat{\sigma}_u(u)\} \right| \tag{4}$$

Output: Community assignment $\widehat{\sigma}$.

denote the K leading singular vectors generated from the singular value decomposition of the trimmed matrix $\mathsf{T}_\tau(\mathcal{L}(\mathbf{A}))$. Also, Let λ_K be the K-th largest singular value of it. Note that in a conventional spetral clustering method, each node is represented by a reduced K-dimensional row vector. The hypergraph spectral clustering algorithm is described in Algorithm 2.

Algorithm 2: Spectral Initialization

Input: Vectors $\mathsf{SVD}_K\left(\mathsf{T}_\tau(\mathcal{L}(\mathbf{A}))\right) = \left[\boldsymbol{u}_1^{\mathrm{T}} \ \cdots \ \boldsymbol{u}_n^{\mathrm{T}}\right]^{\mathrm{T}}$, number of communities K, critical radius $r = \mu\sqrt{\frac{K}{n}}$ with some $\mu > 0$.

Set $S = [n]$.
for $k = 1$ **to** K **do**
  Let $t_k = \arg\max_{i\in S} \left|\left\{j \in S : \|\boldsymbol{u}_j - \boldsymbol{u}_i\|_2 < r\right\}\right|$.
  Set $\widehat{C}_k = \left\{j \in S : \|\boldsymbol{u}_j - \boldsymbol{u}_{t_k}\|_2 < r\right\}$.
  Label $\widehat{\sigma}(i) = k \ \forall i \in \widehat{C}_k$.
  Update $S \leftarrow S \setminus \widehat{C}_k$.
end
If $S \neq \varnothing$, then for any $i \in S$, set $\widehat{\sigma}(i) = \arg\min_{k\in[K]} \frac{1}{|\widehat{C}_k|}\sum_{j\in\widehat{C}_k} \|\boldsymbol{u}_j - \boldsymbol{u}_i\|_2$.

Output: Community assignment $\widehat{\sigma}$.

4.3 Time Complexity

Algorithm 2 has a time complexity of $O(n^3)$, the bottleneck of which being the SVD_K step. Still, the computation of SVD could be done approximately in $O(n^2 \log n)$ time with high probability (Yun and Proutiere, 2015) if we are only interested in the first k spectrums. As for the refinement scheme, the sparsity of the underlying hypergraph can be utilized to reduce the complextiy since the whole network structure could be stored in the incidence matrix $\mathbf{H}$ equivalently as in the d-dimensional adjacency tensor $\mathbf{A}$. As a result, the parameter estimation stage only requires $O(dm)$ where $m = |\mathcal{E}|$ is the total number of hyperedges realized. Similary, the time complexity would be $O(Kdm)$ and $O(Kn^2)$ for the calculation of likelihood function and the consensus step, respectively. Hence, the overall complexity for Algorithm 1 and Algorithm 2 combined are $O(n^3 \log n + nKm + Kn^2)$ for a constant order d. It further reduces to $O(n^3 \log n)$ in the sparse regime $\boldsymbol{p} = O(\log n/n^{d-1})$ where $m = O(n \log n)$ with high probability.

5 THEORETICAL GUARANTEES

We first consider the theoretical guarantee for Algorithm 1, which requires that the first-step algorithm satisfy the following condition.

Condition 5.1: *There exists constants $C_0, \delta > 0$ and a positive sequence $\gamma = \gamma_n$ such that*

$$\inf_{(\mathbf{B},\sigma)\in\Theta_d^0} \min_{u\in[n]} \mathbb{P}_\sigma\left\{\ell(\widehat{\sigma}_u, \sigma) \le \gamma_n\right\} \ge 1 - C_0 n^{-(1+\delta)}.$$

5.1 Refinement Step

We have the following upper bound for the risk obtained by the refinement scheme, which serves as the achievability part to the minimax risk.

Theorem 5.1: *As $n \to \infty$, if*

$$\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j} \to \infty \tag{6}$$

and Condition 5.1 is satisfied for

$$\gamma = o(1). \tag{7}$$

Then, with Algorithm 1, there exists a positive sequence $\zeta_n \to 0$ as $n \to \infty$ such that

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

To prove Theorem 5.1, we need a couple of technical lemmas. First, the accuracy of the parameter estimation step can be ensured with a qualified initialization.

Lemma 5.1: *Suppose as $n \to \infty$, $m_{r_i r_j} I_{p_i p_j} \to \infty$ for each $(r_i, r_j) \in \mathcal{N}_d$, and Condition 5.1 holds with γ satisfying (7) for some $\delta > 0$. Then there exists a sequence $\zeta_n' \to 0$ as $n \to \infty$ and a constant $C > 0$ such that*

$$\inf_{(\mathbf{B},\sigma)\in\Theta_d^0} \min_{u\in[n]} \mathbb{P}_\sigma\Big\{\min_{\pi\in\mathcal{S}_K}\max_{\boldsymbol{s}\in[K]^d}\left|\widehat{\mathrm{B}}^u_{\boldsymbol{s}} - \mathrm{B}_{\boldsymbol{s}_\pi}\right| \le \zeta_n' \max_{(r_i,r_j)\in\mathcal{N}_d}(p_i - p_j)\Big\} \ge 1 - Cn^{-(1+\delta)}.$$

Based on Lemma 5.1, the next lemma shows that the local MLE method is able to achieve a risk that decays exponentially fast.

Lemma 5.2: *Suppose as $n \to \infty$, $m_{r_i r_j} I_{p_i p_j} \to \infty$ for each $(r_i, r_j) \in \mathcal{N}_d$. If there are two sequences $\gamma = o(1)$ and $\zeta_n' = o(1)$, constants $C, \delta > 0$ and permutations $\{\pi_u\}_{u=1}^n \subset \mathcal{S}_K$ such that*

$$\inf_{(\mathbf{B},\sigma)\in\Theta_d^0} \min_{u\in[n]} \mathbb{P}_\sigma\Big\{\ell_0((\widehat{\sigma}_u)_{\pi_u}, \sigma) \le \gamma,\ \left|\widehat{\mathrm{B}}^u_{\boldsymbol{s}} - \mathrm{B}_{\boldsymbol{s}_\pi}\right| \le \zeta_n' \max_{(r_i,r_j)\in\mathcal{N}_d}(p_i - p_j)\Big\} \ge 1 - Cn^{-(1+\delta)}.$$

Then for the local estimator $\widehat{\sigma}_u(u)$ (3), there exists a sequence $\zeta_n'' = o(1)$ such that

$$\sup_{(\mathbf{B},\sigma)\in\Theta_d^0} \max_{u\in[n]} \mathbb{P}_\sigma\Big\{(\widehat{\sigma}_u(u))_{\pi_u} \neq \sigma(u)\Big\} \le (K-1)\cdot \exp\left(-(1-\zeta_n'')\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_i r_j} I_{p_i p_j}\right) + Cn^{-(1+\delta)}.$$

Finally, we justify the usage of (4) as a consensus majority voting.

Lemma 5.3 (Lemma 4 in (Gao et al., 2017))**:** *For any community assignments σ and σ': $[n] \to [K]$, such that for some constant $C \geq 1$*

$$\min_{k\in[K]} \big|\{u|\sigma(u)=k\}\big|,\ \min_{k\in[K]} \big|\{u|\sigma'(u)=k\}\big| \geq \frac{n}{CK}$$

and

$$\min_{\pi\in\mathcal{S}_K} \ell_0(\sigma'_\pi) < \frac{1}{CK}.$$

Define map $\xi : [K] \to [K]$ as

$$\xi(i) = \arg\max_{k\in[K]} \big|\{u \mid \sigma(u)=k\} \cap \{u \mid \sigma'(u)=k\}\big|$$

for each $i \in [K]$. Then $\xi \in \mathcal{S}_K$ and $\ell_0(\sigma'_\xi, \sigma)$ is equal to $\min_{\pi\in\mathcal{S}_K} \ell_0(\sigma'_\pi, \sigma)$

We are now ready to sketch the proof of Theorem 5.1.

Sketch Proof of Theorem 5.1. First, we use the union bound to upper bound the risk as follows.

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

where π^{CSS} is the consensus permutation (4) in Algorithm 1. The first part is the risk with correct label permutation, and the second part is the probability that the consensus step fails. The former could be further controled by Lemma 5.2, while the latter could be further upper-bounded by an exponentially decaying term using Lemma 5.3 together with Condition 5.1. Finally, we discuss two cases, depending on whether the exponential term is larger or smaller than $n^{-(1+\delta)}$. In either case, the claimed upper bound is achieved. ■

Theorem 5.1 implies that as long as there is a good initialization achieving Condition 5.1, the refinement scheme could be applied to further reduce the risk. This is because once Condition 5.1 is satisfied, we could estimate the parameters accurately and find the correct permutation by the consensus step. Compared to analysis in graph `SBM` (Gao et al., 2017), we are dealing with more kinds of community relations and thus more kinds of random variables. Such perplexity makes the generalization from a homogeneous `SBM` with only two possible community relations more difficult to analyze.

5.2 Spectral Clustering

Next, we show that our proposed spectral clustering algorithm achieves Condition 5.1. We have the following performance guarantee for Algorithm 2.

Theorem 5.2: *If*

$$\frac{Ka_1}{\lambda_K^2} \leq C_1 \tag{8}$$

for some sufficiently small $C_1 \in (0,1)$ where $p_1 = \frac{a_1}{n^{d-1}}$. Apply Algorithm 2 with a sufficiently small constant $\mu > 0$ and $\tau = C_2\bar{d}$ for some sufficiently large constant C_2. For any constant $C' > 0$, there exists some $C > 0$ depending only on C', C_2 and μ so that

$$\ell(\widehat{\sigma}, \sigma) \leq C\frac{a_1}{\lambda_K^2}$$

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

Our technical contribution here is to generalize the analysis on adjacency matrix for a graph to the hypergraph Laplacian (5) associated with a hypergraph. This is not a trivial work because now the entries in a hypergraph Laplacian are not independent any more. Still, we successfully arrive at a similar expression as the lower bound under the d-`hSBM` model.

Combining Algorithm 1 and Algorithm 2, we have the following achievability part to Theorem 3.1. The key step is to further lower bound λ_K in Theorem 5.2 and demonstrate that it would satisfy Condition 5.1.

Theorem 5.3: *Suppose $\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j} \to \infty$ as $n \to \infty$ Then there exists a positive sequence $\zeta'_n \to 0$ such that*

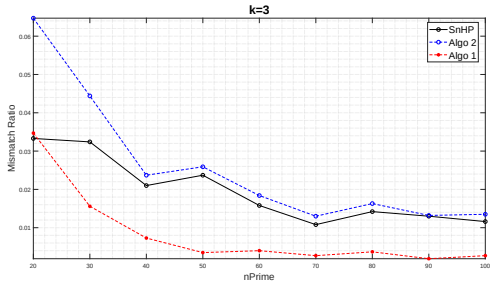
$$\mathrm{R}_d^* \leq \exp\Big(-(1-\zeta'_n)\sum_{i<j:(r_i,r_j)\in\mathcal{N}_d} m_{r_ir_j} I_{p_ip_j}\Big)$$

5.3 Discussion

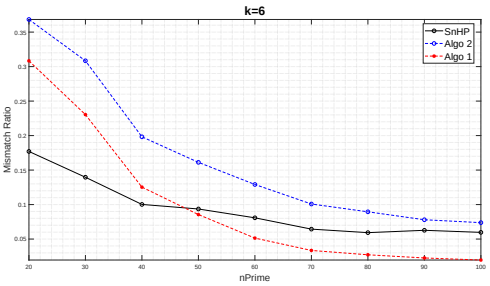
Parallel to our work, (Ghoshdastidar and Dukkipati, 2017) also proposed a similar d-`hSBM` model and analyze the performance of the normalized hypergraph Laplacian. The parameter space they consider is more general. They prove that their risk could be $o(1)$ with probability at least $1 - O((\log n)^{-1/4})$ if the minimum expected degree satisfying a certain condition. This is different from ours since we guarantee that our risk will be $o(1)$ with probability at least $1 - O(n^{-(1+\delta)})$ for some $\delta > 0$ with Algorithm 2. It turns out that we allow the observed hypergraph to be sparser (a lower connecting probability) yet acquire higher mismatch ratio. However, if we raise the probability $\boldsymbol{p}$ to the same order as considered in (Ghoshdastidar and Dukkipati, 2017), Algorithm 1 is guaranteed to have a $(n')^2$-times lower risk. In either case, we always have a success probability converging faster to 1.

6 EXPERIMENTAL RESULTS

The advantage of clustering with a hypergraph representation over traditional graph-based approaches has been reported in the literature (Agarwal et al., 2005; Zhou et al., 2006). Here, we present a comparative study of our two-step algorithm with existing clustering methods on hypergraphs, especially the spectral non-uniform hypergraph partitioning algorithm (SnHP) in (Ghoshdastidar and Dukkipati, 2017) using the hypergraph Laplacian proposed in (Zhou et al., 2006) and the generalized tensor spectral method (GTS) in (Ghoshdastidar and Dukkipati, 2015). In order to have a fair comparison, we run different algorithms on the same hypergraphs generated and calculate the corresponding mismatch ratio as the performance measure. We will not elaborate on how to choose a best way or define a proper way of embedding, as the topic itself would require a whole line of research and is beyond the scope of this paper. In what follows, Algo 2 refers to our first-step spectral clustering algorithm and Algo 1 refers to the combined two-step workflow (i.e. Algorithm 1 on top of Algorithm 2).



(a) $K = 3$



(b) $K = 6$

Figure 1: Simulation Results on 3-`hSBM`.

6.1 Synthetic Data

We implement different algorithms on generative 3-`hSBM` data. The parameter spaces considered are homogeneous and exactly equal-sized, which means that each community has the same number of members. This nodes-per-community parameter n' scales from $20, 30, \cdots$ to 100, while the number of communities K varies from $2, 3, \cdots$ to 10. We set the connecting probability parameter $\boldsymbol{p}$ to be $(60, 30, 10) \cdot \log n/n^2$ for each possible value of $n = K \cdot n'$. Note that the order of $\boldsymbol{p}$ is as prescribed for the sparse regime in Section 5. The choice of this particular triplet is to ensure that the generated hypergraphs are not too sparse. Empirically, the total number of realized hyperedges is roughly $4n \log n$ to $5n \log n$. The performance under each scenario, i.e. each pair of (K, n'), is averaged over 25 realizations of the random hypergraph model, which is large enough for the mismatch ratio to converge for all algorithms implemented. Figure 1 summarizes our simulation results.

Except for the first few scenarios where the total number of nodes n are quite small, we can see that Algo 2 performs roughly as well as the SnHP algorithm. This somewhat indicates that the weak consistency condition Condition 5.1 can also be satisfied with the hypergraph Laplacian proposed by (Zhou et al., 2006) as the first step. Furthermore, the refinement scheme indeed has a better performance over the spectral clustering methods. Observe that the improvement due to the second step becomes larger as K (and hence n) increases. The performance gain should be more evident for a larger network.

6.2 Real-World Data

The data analyzed in this work are obtained from the UCI repository (Lichman, 2013), which is widely used as a benchmark database that admits ground-truth community labels. To perform clustering on a hypergraph, we first embed the entities with various attributes into a hypergraph. One caveat is that the embedded hypergraphs are no longer homogeneous nor approximately equal-sized as assumed when deriving the theoretical guarantees. Nevertheless, the experimental results show that our two-step algorithm does have a performance that is comparable to or even better than existing methods on either graph or hypergraph models. More thorough experimental results are given in the supplementary matrial.

References

Abbe, E. and C. Sandon (2015). "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*, pp. 670–688.

Abbe, Emmanuel (2017). "Community detection and stochastic block models: recent developments". In: *CoRR* abs/1703.10146.

Abbe, Emmanuel, Afonso S Bandeira, and Georgina Hall (2016). "Exact recovery in the stochastic block model". In: *IEEE Transactions on Information Theory* 62.1, pp. 471–487.

Agarwal, S. et al. (2005). "Beyond pairwise clustering". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2, 838–845 vol. 2.

Agarwal, Sameer, Kristin Branson, and Serge Belongie (2006). "Higher Order Learning with Graphs". In: *Proceedings of International Conference on Machine Learning (ICML)*. ICML '06, pp. 17–24.

Ahn, Kwangjun, Kangwook Lee, and Changho Suh (2016). "Community Recovery in Hypergraphs". In: *Allerton Conference on Communication, Control and Computing*. UIUC.

Alon, Noga and Joel H Spencer (2004). *The probabilistic method*. John Wiley & Sons.

Andritsos, Periklis et al. (2004). "LIMBO: Scalable clustering of categorical data". In: *International Conference on Extending Database Technology*. Springer, pp. 123–146.

Angelini, Maria Chiara et al. (2015). "Spectral detection on sparse hypergraphs". In: *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, pp. 66–73.

Chen, Yuxin and Changho Suh (2015). "Spectral MLE: Top-K Rank Aggregation from Pairwise Comparisons". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pp. 371–380.

Chien, I (Eli), Chung-Yi Lin, and I-Hsiang Wang (2018). "On the Minimax Misclassification Ratio of Hypergraph Community Detection". In: *ArXiv e-prints*. arXiv: 1802.00926.

Chin, Peter, Anup Rao, and Van Vu (2015). "Stochastic Block Model and Community Detection in Sparse Graphs: A spectral algorithm with optimal rate of recovery." In: *COLT*, pp. 391–423.

Condon, Anne and Richard M. Karp (2001). "Algorithms for Graph Partitioning on the Planted Partition Model". In: *Random Structures and Algorithms* 18.2, pp. 116–140.

Davis, Chandler and William Morton Kahan (1970). "The rotation of eigenvectors by a perturbation. III". In: *SIAM Journal on Numerical Analysis* 7.1, pp. 1–46.

Duchenne, O. et al. (2011). "A Tensor-Based Algorithm for High-Order Graph Matching". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.12, pp. 2383–2395.

Gao, Chao et al. (2017). "Achieving Optimal Misclassification Proportion in Stochastic Block Models". In: *Journal of Machine Learning Research* 18.60, pp. 1–45.

Ghoshdastidar, Debarghya and Ambedkar Dukkipati (2014). "Consistency of Spectral Partitioning of Uniform Hypergraphs under Planted Partition Model". In: *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, pp. 397–405.

– (2015). "A Provable Generalized Tensor Spectral Method for Uniform Hypergraph Partitioning". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pp. 400–409.

– (2017). "Consistency of spectral hypergraph partitioning under planted partition model". In: *Ann. Statist.* 45.1, pp. 289–315.

Holland, Paul W., Kathryn Blackmond Laskey, and Samuel Leinhardt (1983). "Stochastic Blockmodels: First Steps". In: *Social Networks* 5.2, pp. 109–137.

Lei, Jing, Alessandro Rinaldo, et al. (2015). "Consistency of spectral clustering in stochastic block models". In: *The Annals of Statistics* 43.1, pp. 215–237.

Li, Pan et al. (2016). "Motif Clustering and Overlapping Clustering for Social Network Analysis". In: *arXiv preprint arXiv:1612.00895*.

Lichman, M. (2013). *UCI Machine Learning Repository*. URL: http://archive.ics.uci.edu/ml.

Lin, C. Y., I. E. Chien, and I. H. Wang (2017). "On the fundamental statistical limit of community detection in random hypergraphs". In: *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 2178–2182.

Rozovsky, LV (2003). "A lower bound of large-deviation probabilities for the sample mean under the Cramér condition". In: *Journal of Mathematical Sciences* 118.6, pp. 5624–5634.

Von Luxburg, Ulrike (2007). "A tutorial on spectral clustering". In: *Statistics and computing* 17.4, pp. 395–416.

Yun, Se-Young and Alexandre Proutiere (2015). "Optimal Cluster Recovery in the Labeled Stochastic Block Model". In: *ArXiv e-prints*. arXiv: 1510.05956.

Zhang, Anderson Y. and Harrison H. Zhou (2016). "Minimax rates of community detection in stochastic block models". In: *Ann. Statist.* 44.5, pp. 2252–2280.

Zhou, Dengyong, Jiayuan Huang, and Bernhard Schölkopf (2006). "Learning with hypergraphs: Clustering, classification, and embedding". In: *NIPS*. Vol. 19, pp. 1633–1640.