

PAPER

Exact recovery of community detection in k -community Gaussian mixture models

Zhongyang Li

Department of Mathematics, University of Connecticut, Storrs, Connecticut, USA
Email: zhongyang.li@uconn.edu

Received: 02 May 2023; **Revised:** 22 April 2024; **Accepted:** 07 May 2024

Keywords: Gaussian mixture model; community detection; maximum likelihood estimation; exact recovery

2020 Mathematics Subject Classification: 62H30 (Primary); 62F12, 68W40, 68Q25 (Secondary)

Abstract

We study the community detection problem on a Gaussian mixture model, in which vertices are divided into $k \geq 2$ distinct communities. The major difference in our model is that the intensities for Gaussian perturbations are different for different entries in the observation matrix, and we do not assume that every community has the same number of vertices. We explicitly find the necessary and sufficient conditions for the exact recovery of the maximum likelihood estimation, which can give a sharp phase transition for the exact recovery even though the Gaussian perturbations are not identically distributed; see Section 7. Applications include the community detection on hypergraphs.

1. Introduction

Community structures are ubiquitous in graphs modelling natural and social phenomena. In natural sciences, atoms form molecules so that atoms in the same molecule have stronger connections compared to those in different molecules. In social sciences, individuals form groups in such a way that individuals in the same group have more communications compared to individuals in different groups. The main aim for community detection is to determine the specific groups that specific individuals belong to based on observations of (random) connections between individuals. Identifying different communities in the stochastic block model is a central topic in many fields of science and technology; see [1] for a summary.

In this paper, we study the community detection problem for the Gaussian mixture model, in which there are n vertices belonging to k ($k \geq 2$) different communities. We observe a $p \times 1$ vector for each one of the n vertices, perturbed by a $p \times 1$ Gaussian vector with independent (but not necessarily identically distributed), mean-0 entries. More precisely, each entry of the $p \times n$ perturbation matrix is obtained by a multiple of a standard Gaussian random variable, while the intensities of different entries are different. Given such an observation, we find the maximum likelihood estimation (MLE) for the community assignment and study the probability that the MLE equals the true community assignment as the number of vertices $n \rightarrow \infty$. If this probability tends to 1 as $n \rightarrow \infty$, we say exact recovery occurs. Heuristically, it is natural to conjecture that exact recovery may occur when the intensities of the perturbations are small but does not occur when these intensities are large. The major theme of the paper is to investigate how small the intensities of the perturbations are needed in order to ensure the exact recovery and how large the intensities are required to stop the occurrences of the exact recovery.

Clustering problems in the Gaussian mixture model have been studied extensively; see [5, 7, 15, 18] for an incomplete list. We mention some recent related work here.



The Gaussian mixture model when all the entries of the perturbation matrix are i.i.d was studied in [4], in which a condition for the exact recovery of the semi-definite programming (SDP) is proved. When all the communities have the same size, a condition that exact recovery does not occur was also proved in [4] when the number of communities $k \leq \log n$. The case of unbalanced communities was investigated in [8]. In this paper, we obtain conditions when the exact recovery happens and does not happen for the more general Gaussian mixture model when the entries of the perturbation matrix are not necessarily identically distributed. Our result can be applied to the special case when intensities of the Gaussian perturbations are all equal, and in particular, we obtain a condition that the exact recovery of MLE does not occur when the number of communities k is $e^{o(\log n)}$ in the hypergraph model; see the explanations at the end of Example 4.5. We can also see from the sufficient condition (17) for the exact recovery and the sufficient condition (22) that the exact recovery does not occur that when k is $e^{o(\log n)}$, these two conditions match and there is a sharp phase transition. In Section 7, we see an example in which these necessary and sufficient conditions of the exact recovery give a sharp phase transition when the Gaussian perturbations are not i.i.d.

When $p = n$ in our model, we may consider the rows and columns of the observation matrix are indexed by the n vertices, and each entry represents an edge. In this case, we obtain the community detection problem on a graph. When $p = n^s$ with $s \geq 2$, we may consider the rows of the observation matrix are indexed by ordered s -tuples of vertices, and each entry of the observation matrix represents a $(s + 1)$ -hyperedge. In this case, we obtain the community detection problem on a hypergraph. Community detections on hypergraphs with Gaussian perturbations were studied in [11], where the vertices are divided into two equal-sized communities, and a weight-1 $(d + 1)$ -hyperedge exists if and only if all the vertices are in the same group. The results proved in this paper can be applied to the community detection problems on hypergraphs with Gaussian perturbation to obtain necessary and sufficient conditions for the exact recovery, in which the number of communities is arbitrary and communities are not necessarily equal-sized; moreover, the hyperedges have general weights as represented in the (unperturbed) observation matrix. Community detection problems on random graphs were also studied in [2, 3, 6, 9, 14, 16]. Algorithms to efficiently implement the community recovery in mixture models include the Expectation-Maximization (EM) algorithm ([5]) and the spectral method ([10, 12, 19]). The EM algorithm is a local search heuristic that can fail. The spectral method is based on the singular value decomposition of the dataset and then only use data related to a fixed finite number of largest singular values to achieve dimension reduction. The performance of the spectral method when the Gaussian perturbations consist of i.i.d. Gaussian random variables was discussed in [12]. The major differences between results in this paper and those in [12] are (1) this paper focuses on exact recovery, that is, the probability that the estimation is exactly equal to the true community assignment and no mislabel is allowed, while [12] allows a small number of mislabelled vertices (almost exact recovery) and (2) [12] focuses on Gaussian mixture model with isotropic covariance matrix; while in this paper, the covariances matrices for the model can be more general. Partial recovery was also discussed in [8].

The implement of the MLE is usually very slow; in some cases, relaxing some constraints in MLE leads to the more efficient SDP. However, most SDPs require the partition to be balanced or that, at least, the size of each group is known in advance. The MLE optimisation discussed here has the advantage to attack the case of unbalanced sizes of groups, and the case when the size of each group is unknown. When the perturbations for the entries of the observation matrix are assumed to be i.i.d. Gaussian, and using the average value of the observations associated with each group to approximated the expectation of the observation to the group, in this case the MLE becomes the K-means estimation. The SDP relaxation for K-means have been studied extensively, see for example [17]. We expect a similar SDP relaxation for K-means in the dependent case in which the inner product in the objective function should be defined with respect to the inverse of the covariance matrix; yet in this paper, we shall focus on the more fundamental MLE and try to investigate its statistical limit.

The organisation of the paper is as follows. In Section 2, we review the definition of the Gaussian mixture models and hypergraphs and state the main results proved in this paper. In Section 3, we prove conditions for the exact recovery of the Gaussian mixture model when the number of vertices in each

community is unknown. In Section 4, we apply the results proved in Section 3 to the exact recovery of the community detection in hypergraphs and also prove conditions when exact recovery does not occur in hypergraphs under the assumption that the number of vertices in each community is unknown. In Section 5, we prove conditions for the exact recovery of the Gaussian mixture model when the number of vertices in each community is known and fixed. In Section 6, we prove conditions when exact recovery does not occur in hypergraphs under the assumption that the number of vertices in each community is known and fixed. In Section 7, we give an example in which these necessary and sufficient conditions of the exact recovery give a sharp phase transition when the Gaussian perturbations are not i.i.d.; results of numerical experiments are also included to show the performance of MLE with difference parameters to illustrate the sharp phase transition. In Section A, we prove a lemma used to obtain the main results of the paper.

2. Backgrounds and main results

In this section, we review the definition of the Gaussian mixture models and hypergraphs and state the main results proved in this paper.

2.1. Gaussian mixture model

Let $n \geq k \geq 2$ be positive integers. Let

$$[n] = \{1, 2, \dots, n\}$$

be a set of n vertices divided into k different communities. Let

$$[k] := \{1, \dots, k\}$$

be the set of communities. A mapping $x : [n] \rightarrow [k]$ which assigns a unique community represented by an integer in $[k]$ to each one of the n vertices in $[n]$ is called a community assignment mapping. Let Ω be the set consisting of all the possible mappings from $[n]$ to $[k]$, that is,

$$\Omega := \{x : [n] \rightarrow [k]\}.$$

Each mapping in Ω is a community assignment mapping.

We shall define a function θ , which assigns to each vertex a $p \times 1$ vector, depending on the community of this vertex. Indeed, we require that all the vertices in the same community correspond to the same $p \times 1$ vector under θ . Observing θ perturbed by some Gaussian noise, our goal is to identify the community assignment and determine when exactly recovery occurs. It is natural to believe that those vertices whose corresponding observations are close to each other are in the same community, while those vertices whose corresponding observations are far away from each other are in different communities, but this intuition may be complicated by the existence of the Gaussian noise. See Sections 4.1, 4.2 and 4.3 for examples with specific θ .

More precisely, let $p \geq 1$ be a positive integer. Let

$$\theta : \Omega \times [p] \times [k] \rightarrow \mathbb{R}$$

be a function on the set $\Omega \times [p] \times [k]$ taking real values.

For a community assignment mapping $x \in \Omega$, let \mathbf{A}_x be a $p \times n$ matrix whose entries are given by:

$$(\mathbf{A}_x)_{i,j} = \theta(x, i, x(j)), \quad \forall i \in [p], j \in [n]. \tag{1}$$

Let Σ be a $p \times n$ matrix with positive real entries defined by:

$$\Sigma := (\sigma_{i,j})_{i \in [p], j \in [n]} \in (\mathbb{R}^+)^{p \times n}$$

Let P, Q be two $p \times n$ matrices. Define the inner product of P, Q by

$$\langle P, Q \rangle = \sum_{i \in [p]} \sum_{j \in [n]} P_{ij} Q_{ij}.$$

The norm $\|P\|$ for a matrix P is defined by:

$$\|P\| = \sqrt{\langle P, P \rangle}.$$

Let $P * Q$ be a $p \times n$ matrix defined by:

$$P * Q := (P_{i,j} Q_{i,j})_{i \in [p], j \in [n]}$$

Define a random observation matrix \mathbf{K}_x by:

$$\mathbf{K}_x = \mathbf{A}_x + \Sigma * \mathbf{W}; \tag{2}$$

where \mathbf{W} is a random $p \times n$ matrix with i.i.d. standard Gaussian entries. Note that if the entries of Σ are not all equal, the perturbation matrix $\Sigma * W$ has independent but not identically distributed entries.

Let $y \in \Omega$ be the true community assignment mapping. Given the observation \mathbf{K}_y , the goal is to determine the true community assignment mapping y . We shall apply the MLE.

Let n_1, \dots, n_k be positive integers satisfying

$$\sum_{i=1}^k n_i = n.$$

and

$$|y^{-1}(i)| = n_i, \forall i \in [k];$$

that is, n_i is the number of vertices in community i for each $i \in [k]$ under the mapping y .

Let

$$\Omega_{n_1, \dots, n_k} := \{x \in \Omega : |x^{-1}(i)| = n_i, \forall i \in [k]\}$$

be the set of all the community assignment mappings such that there are exactly n_i vertices in the community i , for each $i \in [k]$.

For each real number $c \in (0, 1)$, let

$$\Omega_c := \left\{ x \in \Omega : \frac{|x^{-1}(i)|}{\sum_{j \in [k]} |x^{-1}(j)|} \geq c, \forall i \in [k] \right\},$$

that is, Ω_c consists of all the community assignment mappings such that the percentage of the numbers of vertices in each community is at least c .

Assume the true community assignment mapping $y \in \Omega_c$ for some $c \in (0, 1)$. Let Φ be an $p \times n$ matrix whose entries are given by:

$$(\Phi)_{ij} = \frac{1}{\sigma_{ij}}, \forall i \in [p], j \in [n];$$

in other words, the (i, j) -entry of Φ is the reciprocal of the (i, j) -entry of Σ . Define

$$\hat{y} := \operatorname{argmin}_{x \in \Omega_{\frac{2c}{3}}} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 \tag{3}$$

and

$$\check{y} := \operatorname{argmin}_{x \in \Omega_{n_1, \dots, n_k}} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 \tag{4}$$

Then we have the following lemma

Lemma 2.1. \hat{y} is the MLE with respect to the observation \mathbf{K}_y in $\Omega_{\frac{2c}{3}}$. \check{y} is the MLE with respect to the observation \mathbf{K}_y in Ω_{n_1, \dots, n_k} .

Proof. By definition, the MLE with respect to the observation \mathbf{K}_y in $\Omega_{\frac{2c}{3}}$ (resp. Ω_{n_1, \dots, n_k}) should maximise the probability density of the observation \mathbf{K}_y among all $x \in \Omega_{\frac{2c}{3}}$ (resp. $x \in \Omega_{n_1, \dots, n_k}$). If the true community assignment mapping $y = x$, we may consider \mathbf{K}_y as a random matrix with mean value \mathbf{A}_x and independent entries, such that variance of its (i, j) -entry is σ_{ij}^2 . Therefore, the probability density of \mathbf{K}_y is given by:

$$\left(\prod_{i \in [p], j \in [n]} \frac{1}{\sqrt{2\pi} \sigma_{ij}} \right) e^{-\sum_{i \in [p], j \in [n]} \frac{(\mathbf{K}_y - \mathbf{A}_x)_{ij}^2}{2\sigma_{ij}^2}},$$

where the exponent is exactly

$$-\frac{1}{2} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2.$$

It is straightforward to check that the minimiser of $\|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2$ is exactly the maximiser of the probability density. Then the lemma follows. \square

We shall investigate under which conditions we have $\check{y} = y$ and $\hat{y} = y$ with high probability.

To state the main theorems proved in this paper, we first introduce a few assumptions.

For $x, y \in \Omega$, let

$$L_\Phi(x, y) := \|\Phi * (\mathbf{A}_x - \mathbf{A}_y)\|^2. \tag{5}$$

For $x \in \Omega$, let

$$n_i(x) = |x^{-1}(i)|, \forall i \in [k];$$

then $n_i(x)$ is the number of vertices in community i under the community assignment mapping x . It is straightforward to check that

$$\sum_{i=1}^k n_i(x) = n.$$

For $i, j \in [k]$ and $x, z \in \Omega$, let $t_{ij}(x, z)$ be a non-negative integer given by:

$$t_{ij}(x, z) = |x^{-1}(i) \cap z^{-1}(j)|.$$

That is, $t_{ij}(x, z)$ is the number of vertices in $[n]$ which are in community i under the mapping x and in community j under the mapping z . Then

$$\sum_{j \in [k]} t_{ij}(x, z) = n_i(x); \quad \sum_{i \in [k]} t_{ij}(x, z) = n_j(z); \tag{6}$$

We now introduce an equivalence condition on Ω .

Definition 1. For $x \in \Omega$, let $C(x)$ consist of all the $x' \in \Omega$ such that x' can be obtained from x by a θ -preserving bijection of communities. More precisely, $x' \in C(x) \subset \Omega$ if and only if the following condition holds

- for $i \in [p]$ and $j \in [n]$, $\theta(x, i, x(j)) = \theta(x', i, x'(j))$.

We define an equivalence relation on Ω as follows: we say $x, z \in \Omega$ are equivalent if and only if $x \in C(z)$. Let $\bar{\Omega}$ be the set of all the equivalence classes in Ω .

We see that if x and z are equivalent in the sense of Definition 1, the non-random parts of the observation matrices corresponding to x and z satisfy $\mathbf{A}_x = \mathbf{A}_z$. Hence, any algorithm based on this observation will not distinguish equivalent community assignments; and the best we expect is to exactly recover the equivalent class of the community assignment. We want to assume this θ function to be able to distinguish community assignments up to the composition with a bijection of communities more precisely.

Assumption 2.2. Let $x, z \in \Omega$. If for any $i \in [p]$ and $j \in [n]$,

$$\theta(x, i, x(j)) = \theta(z, i, z(j)); \tag{7}$$

then there is a bijection $\eta : [k] \rightarrow [k]$, such that

$$x = \eta \circ z \tag{8}$$

where \circ denotes the composition of two mappings.

Note that (8) is equivalent of saying that for $i, j \in [n]$, $x(i) = x(j)$ if and only if $x'(i) = x'(j)$. See Section 4.1 for examples.

Define a set

$$\mathcal{B} := \left\{ (t_{1,1}, t_{1,2}, \dots, t_{k,k}) \in \{0, 1, 2, \dots, n\}^{k^2} : \sum_{i=1}^k t_{ij} = n_j \right\}. \tag{9}$$

For $\varepsilon > 0$, define a set \mathcal{B}_ε consisting of all the $(t_{1,1}, t_{1,2}, \dots, t_{k,k}) \in \mathcal{B}$ satisfying all the following conditions:

1. $\forall i \in [k], \max_{j \in [k]} t_{j,i} \geq n_i - n\varepsilon$.
2. For $i \in [k]$, let $t_{w(i),i} = \max_{j \in [k]} t_{j,i}$. Then w is a bijection from $[k]$ to $[k]$.
3. w is θ -preserving, that is, for any $x \in \Omega, i \in [p]$ and $a \in [k]$, we have

$$\theta(x, i, a) = \theta(w \circ x, i, w(a)).$$

We may assume θ and Σ satisfy the following assumptions.

Assumption 2.3. Suppose $\varepsilon \in (0, \frac{2c}{3k}), x \in \Omega_{\frac{2c}{3k}}$ and $y \in \Omega_c$. Then for all $x, y \in \Omega$, and

$$(t_{1,1}(x, y), t_{1,2}(x, y), \dots, t_{k,k}(x, y)) \in \mathcal{B} \setminus \mathcal{B}_\varepsilon, \tag{10}$$

we have

$$L_\Phi(x, y) \geq R(n) > 0. \tag{11}$$

Here $R(n)$ is some function of n , such that (11) combined with (15) leads to the desired exact recovery result in Theorem 2.6.

Assumption 2.4. Assume $\varepsilon \in (0, \frac{2c}{3k}), x \in \Omega_{\frac{2c}{3k}}$ and $y \in \Omega_c$. Assume there exists $\Delta > 0$ such that the following holds:

Let $y_1, y_2 \in \Omega_{\frac{2c}{3k}}$ and $a, b \in [k]$ and $a \neq b$. Let $i, j \in [n]$ such that $i \in y_1^{-1}(a) \cap x^{-1}(b)$. Let $y_2 : [n] \rightarrow [k]$ be defined as follows:

$$y_2(j) := \begin{cases} b & \text{if } j = i \\ y_1(j) & \text{if } j \in [n] \setminus \{i\} \end{cases}.$$

When

$$(t_{1,1}(x, y_1), t_{1,2}(x, y_1), \dots, t_{k,k}(x, y_1)) \in \mathcal{B}_\varepsilon \tag{12}$$

such that for all $i \in [k]$

$$t_{i,i} = \max_{j \in [k]} t_{j,i}(x, y_1);$$

$\varepsilon \in (0, \frac{2c}{3k})$, and

$$y_1 \notin C(x);$$

we have

$$L_\Phi(x, y_1) - L_\Phi(x, y_2) \geq \Delta(1 + o(1)).$$

where $o(1) \rightarrow 0$, as $n \rightarrow \infty$.

Assumptions 2.3 and 2.4 are technical. One may interpret Assumption 2.4 as follows: when a community assignment mapping $y_1 \in \Omega$ is sufficiently close to the community assignment mapping x in the sense of (12), change the community of exactly one vertex j from $y_1(j)$ to $x(j)$ and obtain a new community assignment mapping $y_2(j)$, then the distance between y_2 and x , in the sense of L_ϕ , approximately decreases by Δ from the distance between y_1 and x . In other words, as y_1 is close to x and approaching x , $L_\phi(y_1, x)$ decreases linearly with rate Δ . As we see later, this will create a convergent geometric series when computing the lower bound for the probability of exact recovery.

Note that Assumption 2.3 can be guaranteed by the following stronger assumption with $R(n) = \frac{T(n)}{B_1^2}$; see Lemma 3.5.

Assumption 2.5.

1. There exists $B_1 > 0$, such that for all $i, j \in [p] \times n$, we have

$$|\sigma_{i,j}| \leq B_1.$$

2. Assume $\varepsilon \in (0, \frac{2c}{3k})$, $x \in \Omega_{\frac{2c}{3k}}$ and $y \in \Omega_c$. Then for all $x, y \in \Omega$, and

$$(t_{1,1}(x, y), t_{1,2}(x, y), \dots, t_{k,k}(x, y)) \in \mathcal{B} \setminus \mathcal{B}_\varepsilon, \tag{13}$$

we have

$$\sum_{i \in [p], j \in [n]} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2 \geq T(n) > 0. \tag{14}$$

Here $T(n)$ is some function of n , such that (14) combined with (16) leads to the desired exact recovery result in Theorem 2.6.

Assumption 2.5(1) gives an upper bound on the standard deviation of the Gaussian perturbations (which may depend on n); Assumption 2.5(2) may be interpreted as when $x, y \in \Omega$ are “far away” from each other in the sense of (13), the corresponding θ functions are “far away” from each other in the sense of (11).

Theorem 2.6. Assume $y \in \Omega_c$ is the true community assignment mapping. Suppose that Assumptions 2.5 and 2.4 hold. Let $\varepsilon \in (0, \frac{2c}{3k})$. Suppose one of the following cases occurs

1. Assumption 2.3 holds and

$$\lim_{n \rightarrow \infty} n \log k - \frac{R(n)}{8} = -\infty, \tag{15}$$

2. Assumption 2.5 holds and

$$\lim_{n \rightarrow \infty} n \log k - \frac{T(n)}{8B_1^2} = -\infty. \tag{16}$$

Moreover, suppose that for any constant $\delta > 0$ independent of n ,

$$\lim_{n \rightarrow \infty} \log k + \log n - \frac{\Delta(1 - \delta)}{8} = -\infty, \tag{17}$$

then $\lim_{n \rightarrow \infty} \Pr(\hat{y} \in C(y)) = 1$.

Theorem 2.6 gives a sufficient condition for the exact recovery of MLE in the Gaussian mixture model. It is proved in Section 3. An application of Theorem 2.6 on the exact recovery of community detection on hypergraphs is discussed in Section 4.3.

We also obtain a condition for the exact recovery when the sample space of the MLE is restricted to Ω_{n_1, \dots, n_k} , that is, the number of vertices in each community is known and fixed.

Assumption 2.7. Assume $x, y_m, y_h \in \Omega$ such that

1. $D_\Omega(y_m, y_h) = j$, where $j \geq 2$ is a positive integer; and

2. There exist $u_1, \dots, u_j \in [n]$, such that

- (a) $y_m(v) = y_h(v)$, for all $v \in [n] \setminus \{u_1, \dots, u_j\}$; and
- (b) $y_m(u_i) \neq y_h(u_i) = x(u_i) = y_m(u_{i-1})$ for all $i \in [j]$.
- (c) $(t_{1,1}(x, y_m), t_{1,2}(x, y_m), \dots, t_{k,k}(x, y_m)) \in \mathcal{B}_\varepsilon$ with $\varepsilon \in (0, \frac{2c}{3k})$ and $w(i) = i$.

Then

$$L_\Phi(x, y_m) - L_\Phi(x, y_h) \geq j\Delta(1 + o(1)) \tag{18}$$

Theorem 2.8. Suppose that Assumptions 2.5, 2.7, (15) and (17) hold. Then $\lim_{n \rightarrow \infty} \Pr(\check{y} \in C(y)) = 1$.

Indeed, Assumption 2.4 implies Assumption 2.7; see Lemma 5.5. Theorem 2.8 is proved in Section (5).

2.2. Hypergraphs

A special case for the Gaussian mixture model is the hypergraph model. Let s, s_1, s_2 be positive integers satisfying

$$2 \leq s_1 \leq s \leq s_2.$$

A hypergraph $H = (V, E)$ has vertex set $V := [n]$ and hyperedge set E defined as follows:

$$E := \{(a_1, \dots, a_s) : a_1, \dots, a_s \in [n], s \in \{s_1, s_1 + 1, \dots, s_2\}\}$$

Let $\phi : \cup_{s=s_1}^{s_2} [k]^s \rightarrow [0, \infty)$ be a function which assigns a unique real number $\phi(c_1, \dots, c_s)$ to each s -tuple of communities $(c_1, \dots, c_s) \in [k]^s$, and $s \in [s_1, s_2]$.

For a community assignment mapping x , the weighted adjacency tensor \mathbf{A}_x is defined by:

$$(\mathbf{A}_x)_{a_1 \dots a_s} = \begin{cases} \phi(x(a_1), \dots, x(a_s)), & \text{if } (a_1, \dots, a_s) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

and

$$\Sigma_{(a_1, \dots, a_s)} := \sigma_{(a_1, \dots, a_s)}$$

Define a random tensor \mathbf{K}_x as in (2). Recall that $y \in \Omega_c$ is the true community assignment mapping. Define \hat{y} and \check{y} as in (3) and (4).

Recall that $y \in \Omega$ is the true community assignment mapping satisfying $|y^{-1}(i)| = n_i$, for all $i \in [k]$. Let $a \in [n]$. Let $y^{(a)} \in \Omega$ be defined by:

$$y^{(a)}(i) = \begin{cases} y(i) & \text{if } i \in [n], \text{ and } i \neq a \\ y^{(a)}(a) & \text{if } i = a. \end{cases} \tag{20}$$

such that

$$y(a) \neq y^{(a)}(a) \in [k].$$

Theorem 2.9. Assume

$$\lim_{n \rightarrow \infty} \min_{i \in [k]} n_i = \infty. \tag{21}$$

Suppose that there exists a subset $H \subset [n]$ satisfying all the following conditions:

- 1. $|H| = h = o(n)$;
- 2. $\lim_{n \rightarrow \infty} \frac{\log h}{\log n} = 1$;

3. For each $g \in H$,

$$\sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i_j}, \dots, i_s) \in ([n] \setminus H)^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, g, i_{j+1}, \dots, i_s)}^2} \times (\phi(y(i_1), \dots, y^{(g)}(g), \dots, y(i_s)) - \phi(y(i_1), \dots, y(g), \dots, y(i_s)))^2 = (1 + o(1))L_{\Phi}(y^{(g)}, y)$$

4. there exists a constant $\beta > 0$ independent of n , such that

$$\frac{\max_{a \in H} L_{\Phi}(y^{(a)}, y)}{\min_{a \in H} L_{\Phi}(y^{(a)}, y)} \leq \beta^2, \forall n.$$

If there exists a constant $\delta > 0$ independent of n , such that

$$\max_{a \in H} L_{\Phi}(y^{(a)}, y) < 8(1 - \delta) \log n \tag{22}$$

Then $\lim_{n \rightarrow \infty} \Pr(\hat{y} \in C(y)) = 0$.

Theorem 2.9 is proved in Section 4. An example is given in Section 4.2.

Let $a, b \in [n]$ such that $y(a) \neq y(b)$. Let $y^{(ab)} \in \Omega_{n_1, \dots, n_k}$ be the community assignment mapping defined by:

$$y^{(ab)}(i) = \begin{cases} y(i) & \text{if } i \in [n] \setminus \{a, b\} \\ y(b) & \text{if } i = a \\ y(a) & \text{if } i = b \end{cases} \tag{23}$$

In other words, $y^{(ab)}$ is obtained from y by exchanging $y(a)$ and $y(b)$.

We also prove a condition when the exact recovery does not occur if the sample space of the MLE is restricted in Ω_{n_1, \dots, n_k} .

Theorem 2.10. Assume

$$\lim_{n \rightarrow \infty} \min_{i \in [k]} n_i = \infty. \tag{24}$$

Suppose that there exist two subsets $H_1, H_2 \subset [n]$ satisfying all the following conditions:

1. $|H_1| = |H_2| = h = o(n)$;
2. $\lim_{n \rightarrow \infty} \frac{\log h}{\log n} = 1$;
3. For any $u_1, u_2 \in H_1$ and $v_1, v_2 \in H_2$,

$$y(u_1) = y(u_2) \neq y(v_1) = y(v_2);$$

4. For any $u \in H_1$ and $v \in H_2$

$$\sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i_j}, \dots, i_s) \in ([n] \setminus (H_1 \cup H_2))^{s-1}} \left(\frac{1}{\sigma_{(i_1, \dots, i_{j-1}, u, i_{j+1}, \dots, i_s)}^2} + \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, v, i_{j+1}, \dots, i_s)}^2} \right) (\phi(y(i_1), \dots, y(v), \dots, y(i_s)) - \phi(y(i_1), \dots, y(u), \dots, y(i_s)))^2 = (1 + o(1))L_{\Phi}(y^{(uv)}, y)$$

5. For any $g \in H_1 \cup H_2$, the quantity

$$\sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i_j}, \dots, i_s) \in ([n] \setminus (H_1 \cup H_2))^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, g, i_{j+1}, \dots, i_s)}^2} (\phi(y(i_1), \dots, y(b), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s)))^2$$

is a constant and is independent of g .

If there exists a constant $\delta > 0$ independent of n , such that

$$\max_{u \in H_1, v \in H_2} L_\Phi(y^{(uv)}, y) < 16(1 - \delta) \log n, \tag{25}$$

$\lim_{n \rightarrow \infty} \Pr(\check{y} \in C(y)) = 0$.

Theorem 2.10 is proved in Section 6.

3. Community detection on K-community Gaussian mixture models

In this section, we consider the MLE when the number of vertices in each community is unknown. We shall obtain a sufficient condition for the occurrence of the exact recovery. The main goal is to prove Theorem 2.6.

Recall that we defined an equivalence relation on Ω in Definition 1. It is straightforward to check that

$$\mathbf{K}_y = \mathbf{K}_{y'}, \text{ and } \mathbf{A}_y = \mathbf{A}_{y'}, \quad \text{if } y' \in C(y).$$

Therefore, the MLE based on the observation \mathbf{K}_y can only recover the community assignment mapping up to equivalence.

The main idea to prove Theorem 2.6 is as follows:

- We find an upper bound for the probability that the MLE is not in the equivalence class of the true community assignment; see (32).
- We show this upper bound converges to zero as $n \rightarrow \infty$ under the assumptions of Theorem 2.6. This is achieved by splitting the upper bounds into two parts: one part I_2 is the sum over all the community assignment functions that differs from the true community assignment functions by at most $n\epsilon$ and other part I_1 is the sum over all the community assignment functions that differs from the true community assignment functions by at least $n\epsilon$. The fact that I_2 converges to 0 as $n \rightarrow \infty$ follows directly from Assumption 2.5. To prove that I_2 converges to 0 as $n \rightarrow \infty$, we bound I_1 by a geometric series that converges to 0 as $n \rightarrow \infty$.

Note that

$$\begin{aligned} \langle \Phi * \mathbf{A}_x, \Phi * \mathbf{A}_z \rangle &= \sum_{i \in [p], j \in [n]} \frac{(\mathbf{A}_x)_{ij}(\mathbf{A}_z)_{ij}}{\sigma_{ij}^2} \\ &= \sum_{i \in [p], j \in [n]} \frac{\theta(x, i, x(j))\theta(z, i, z(j))}{\sigma_{ij}^2} \end{aligned} \tag{26}$$

In particular for each $x \in \Omega$, we have

$$\|\Phi * \mathbf{A}_x\|^2 = \sum_{i \in [p], j \in [n]} \frac{(\theta(x, i, x(j)))^2}{\sigma_{ij}^2} \tag{27}$$

Recall that $y \in \Omega_{n_1, \dots, n_k}$ is the true community assignment mapping. Note that

$$\|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 = \|\Phi * \mathbf{K}_y\|^2 - 2\langle \Phi * \mathbf{K}_y, \Phi * \mathbf{A}_x \rangle + \|\Phi * \mathbf{A}_x\|^2 \tag{28}$$

For each fixed observation \mathbf{K}_y , $\|\Phi * \mathbf{K}_y\|^2$ is fixed and independent of $x \in \Omega$. Therefore,

$$\begin{aligned} \hat{y} &:= \operatorname{argmin}_{x \in \Omega_{\frac{2c}{3}}} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 \\ &= \operatorname{argmin}_{x \in \Omega_{\frac{2c}{3}}} (-2\langle \Phi * \mathbf{K}_y, \Phi * \mathbf{A}_x \rangle + \|\Phi * \mathbf{A}_x\|^2) \end{aligned}$$

For $x \in \Omega$, define

$$f(x) := -2\langle \Phi * \mathbf{K}_y, \Phi * \mathbf{A}_x \rangle + \|\Phi * \mathbf{A}_x\|^2 \tag{29}$$

Then

$$\begin{aligned}
 f(x) - f(y) &= \|\Phi * \mathbf{A}_x\|^2 - \|\Phi * \mathbf{A}_y\|^2 \\
 &\quad - 2\langle \Phi * \mathbf{A}_y, \Phi * (\mathbf{A}_x - \mathbf{A}_y) \rangle - 2\langle \mathbf{W}, \Phi * (\mathbf{A}_x - \mathbf{A}_y) \rangle \\
 &= \|\Phi * (\mathbf{A}_x - \mathbf{A}_y)\|^2 - 2\langle \mathbf{W}, \Phi * (\mathbf{A}_x - \mathbf{A}_y) \rangle,
 \end{aligned}
 \tag{30}$$

where we use the identity

$$\Phi * (\Sigma * \mathbf{W}) = \mathbf{W}.$$

Then $f(x) - f(y)$ is a Gaussian random variable with mean value

$$\mathbb{E}(f(x) - f(y)) = L_\Phi(x, y);$$

and variance

$$\text{Var}(f(x) - f(y)) = 4L_\Phi(x, y).$$

Lemma 3.1. For $x, z \in \Omega$. If $x \in C(z)$, then

$$f(x) = f(z).$$

Proof. By Definition 1, if $x \in C(z)$, then for any $i \in [p]$ and $j, h \in [n]$, $x(j) = x(h)$ if and only if $z(j) = z(h)$ and $\theta(x, i, x(j)) = \theta(z, i, z(j))$, then $\mathbf{A}_x = \mathbf{A}_z$ by (19). Moreover, since for any $i \in [p]$ and $j \in [n]$,

$$(\mathbf{A}_x)_{ij} = (\mathbf{A}_z)_{ij};$$

we have

$$\frac{(\mathbf{A}_x)_{ij}}{\sigma_{ij}} = \frac{(\mathbf{A}_z)_{ij}}{\sigma_{ij}};$$

this implies

$$\Phi * \mathbf{A}_x = \Phi * \mathbf{A}_z. \tag{31}$$

Then the lemma follows from (29). □

Define

$$p(\hat{y}; \sigma) := \Pr(\hat{y} \in C(y)) = \Pr\left(f(y) < \min_{C(x) \in \bar{\Omega}_{\frac{2\sigma}{3}}, C(x) \neq C(y)} f(x)\right)$$

Then

$$\begin{aligned}
 1 - p(\hat{y}; \sigma) &\leq \sum_{C(x) \in \bar{\Omega}_{\frac{2\sigma}{3}} : C(x) \neq C(y)} \Pr(f(x) - f(y) \leq 0) \\
 &= \sum_{C(x) \in \bar{\Omega}_{\frac{2\sigma}{3}} : C(x) \neq C(y)} \Pr_{\xi \in \mathcal{N}(0,1)}\left(\xi \leq \frac{-\sqrt{L_\Phi(x, y)}}{2}\right) \\
 &\leq \sum_{C(x) \in \bar{\Omega}_{\frac{2\sigma}{3}} : C(x) \neq C(y)} e^{-\frac{(L_\Phi(x, y))^2}{8}}.
 \end{aligned}
 \tag{32}$$

Lemma 3.2. Let $x, y, x', y' \in \Omega$, such that $x' \in C(x)$ and $y' \in C(y)$, then

$$L_\Phi(x, y) = L_\Phi(x', y').$$

Proof. By (31), we obtain that when $x' \in C(x)$ and $y' \in C(y)$,

$$\Phi * \mathbf{A}_x = \Phi * \mathbf{A}_{x'}; \quad \Phi * \mathbf{A}_y = \Phi * \mathbf{A}_{y'}.$$

Then the lemma follows from (5). □

Lemma 3.3. For $x, y \in \Omega$, $L_\Phi(x, y) \geq 0$. Moreover,

1. If $x \in C(y)$, then $L_\Phi(x, y) = 0$.
2. If θ satisfies Assumption 2.2 and $L_\Phi(x, y) = 0$, then $x \in C(y)$.

Proof. From (5), it is straightforward to check that $L_\Phi(x, y) \geq 0$ for any $x, y \in \Omega$. Moreover, from (5), we obtain

$$L_\Phi(x, y) = \sum_{i \in [p], j \in [n]} \sigma_{ij}^2 (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2$$

By the fact that $\sigma_{ij} > 0$ for all $i \in [p], j \in [n]$, we obtain that $L_\Phi(x, y) = 0$ if and only if

$$\theta(x, i, x(j)) = \theta(y, i, y(j)), \quad \forall i \in [p], j \in [n]. \tag{33}$$

If $x \in C(y)$, then there exists a θ -preserving bijection $\eta : [k] \rightarrow [k]$, such that $x = \eta \circ y$. Then (33) holds by the θ -preserving property of η , then we obtain Part(1).

On the other hand, if $L_\Phi(x, y) = 0$, we have (33) holds. Then $x \in C(y)$ follows from Assumption 2.2. □

Lemma 3.4. Assume that $y \in \Omega_c$ and $x \in \Omega_{\frac{2\varepsilon}{3k}}$. For $i \in [k]$, let

$$t_{w(i),i}(x, y) = \max_{j \in [k]} t_{ji}(x, y), \tag{34}$$

where $w(i) \in [k]$. When $\varepsilon \in (0, \frac{2\varepsilon}{3k})$ and $(t_{1,1}(x, y), t_{1,2}(x, y), \dots, t_{k,k}(x, y)) \in \mathbb{R}^{k^2}$ satisfies

$$\max_{j \in [k]} t_{ji}(x, y) \geq n_i - n\varepsilon, \quad \forall i \in [k]$$

w is a bijection from $[k]$ to $[k]$.

Proof. See Lemma 5.6 of [13]. □

Definition 2. Define the distance function $D_\Omega : \Omega \times \Omega \rightarrow [n]$ as follows:

$$D_\Omega(x, y) = \sum_{i,j \in [k], i \neq j} t_{ij}(x, y).$$

for $x, y \in \Omega$.

From Definition 2, it is straightforward to check that

$$D_\Omega(x, y) = n - \sum_{i \in [k]} t_{ii}(x, y)$$

Lemma 3.5. Assume that θ, Σ satisfies Assumptions 2.5. Then for all the $x, y \in \Omega$ such that (10) holds, we have

$$L_\Phi(x, y) \geq \frac{T(n)}{B_1^2}.$$

Proof. Note that

$$L_\Phi(x, y) = \left(\sum_{i \in [p], j \in [n]} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2 \right) \left(\frac{\sum_{i \in [p], j \in [n]} \frac{1}{\sigma_{ij}^2} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2}{\sum_{i \in [p], j \in [n]} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2} \right)$$

By Assumption 2.5(1), we have

$$\frac{\sum_{i \in [p], j \in [n]} \frac{1}{\sigma_{ij}^2} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2}{\sum_{i \in [p], j \in [n]} (\theta(x, i, x(j)) - \theta(y, i, y(j)))^2} \geq \frac{1}{B_1^2}$$

Then the lemma follows from Assumption 2.5(2). □

Proof of Theorem 2.6. Note that

$$\sum_{C(x) \in \overline{\Omega} \setminus C(y)} e^{-\frac{L_{\Phi}(x,y)}{8}} \leq I_1 + I_2$$

where

$$I_1 = \sum_{C(x) \in \overline{\Omega}_{\frac{2\varepsilon}{3}} : (t_{1,1}(x,y), \dots, t_{k,k}(x,y)) \in [\mathcal{B} \setminus \mathcal{B}_{\varepsilon}], C(x) \neq C(y)} e^{-\frac{(L_{\Phi}(x,y))^2}{8}}$$

and

$$I_2 = \sum_{C(x) \in \overline{\Omega}_{\frac{2\varepsilon}{3}} : (t_{1,1}(x,y), \dots, t_{k,k}(x,y)) \in \mathcal{B}_{\varepsilon}, C(x) \neq C(y)} e^{-\frac{(L_{\Phi}(x,y))^2}{8}}.$$

and $\varepsilon \in (0, \frac{2\varepsilon}{3k})$.

By Lemma 3.5, when Assumption 2.5 holds, we have

$$I_1 \leq k^n e^{-\frac{T(n)}{8B_1^2}}$$

When (15) holds, we obtain

$$\lim_{n \rightarrow \infty} I_1 = 0. \tag{35}$$

Now let us consider I_2 . Let w be the bijection from $[k]$ to $[k]$ as defined in (34). Let $y^* \in \Omega$ be defined by:

$$y^*(z) = w(y(z)), \quad \forall z \in [n].$$

Then $y^* \in C(y)$ since w is θ -preserving by the definition of $\mathcal{B}_{\varepsilon}$. Moreover, x and y^* satisfies

$$t_{i,i}(x, y^*) \geq n_i(y^*) - n\varepsilon, \quad \forall i \in [k]. \tag{36}$$

We consider the following community changing process to obtain x from y^* :

1. If for all $(j, i) \in [k]^2$, and $j \neq i$, $t_{j,i}(x, y^*) = 0$, then $x = y^*$.
2. If (1) does not hold, find the least $(j, i) \in [k]^2$ in lexicographic order such that $j \neq i$ and $t_{j,i}(x, y^*) > 0$. Choose an arbitrary vertex $u \in \{x^{-1}(j) \cap (y^*)^{-1}(i)\}$. Define $y_1 \in \Omega$ as follows:

$$y_1(z) = \begin{cases} j & \text{if } z = u \\ y^*(z) & \text{if } z \in [n] \setminus \{u\} \end{cases}$$

Then we have

$$t_{j,i}(x, y_1) = t_{j,i}(x, y^*) - 1 \tag{37}$$

$$t_{j,j}(x, y_1) = t_{j,j}(x, y^*) + 1 \tag{38}$$

$$t_{a,b}(x, y_1) = t_{a,b}(x, y^*) \quad \forall (a, b) \in ([k]^2 \setminus \{(j, i), (j, j)\}).$$

$$n_i(y_1) = n_i(y^*) - 1$$

$$n_j(y_1) = n_j(y^*) + 1$$

$$n_l(y_1) = n_l(y^*) \quad \forall l \in [k] \setminus \{i, j\}.$$

Therefore, x, y_1 and y^* satisfy

$$t_{l,l}(x, y_1) \geq n_l(y_1) - n\varepsilon;$$

$$t_{l,l}(x, y_1) \geq n_l(y^*) - n\varepsilon;$$

$$n_l(y_1) \geq n_l(y^*) - n\varepsilon;$$

for all $l \in [k]$.

From Assumption 2.4 and Lemma 3.2, we obtain

$$L_\Phi(x, y_1) - L_\Phi(x, y) = L_\Phi(x, y_1) - L_\Phi(x, y^*) \leq -\Delta(1 + o(1)).$$

Therefore,

$$e^{-\frac{L_\Phi(x,y)}{8}} \leq e^{-\frac{L_\Phi(x,y_1)}{8}} e^{-\frac{\Delta(1+o(1))}{8}} \tag{39}$$

In general, if we have constructed $y_l \in \Omega$ ($l \geq 1$) satisfying all the following conditions:

$$\begin{aligned} t_{l,i}(x, y_r) &\geq n_l(y_r) - n\varepsilon; \\ t_{l,i}(x, y_r) &\geq n_l(y^*) - n\varepsilon; \\ n_l(y_r) &\geq n_l(y^*) - n\varepsilon; \end{aligned} \tag{40}$$

for all $l \in [k]$. We now construct $y_{r+1} \in \Omega$ as follows:

- (1) If for all $(j, i) \in [k]^2$, and $j \neq i$, $t_{j,i}(x, y_r) = 0$, then $x = y_r$; then the construction process stops at this step.
- (2) If (a) does not hold, find the least $(j, i) \in [k]^2$ in lexicographic order such that $j \neq i$ and $t_{j,i}(x, y_r) > 0$. Choose an arbitrary vertex $u \in \{x^{-1}(j) \cap y_r^{-1}(i)\}$. Define $y_{r+1} \in \Omega$ as follows:

$$y_{r+1}(z) = \begin{cases} j & \text{if } z = u \\ y_r(z) & \text{if } z \in [n] \setminus \{u\} \end{cases}$$

Then it is straightforward to check that

$$\begin{aligned} t_{l,i}(x, y_{r+1}) &\geq n_l(y_{r+1}) - n\varepsilon; \\ t_{l,i}(x, y_{r+1}) &\geq n_l(y^*) - n\varepsilon; \\ n_l(y_{r+1}) &\geq n_l(y^*) - n\varepsilon; \end{aligned}$$

for all $l \in [k]$.

Then if (36) holds with y^* replaced by y_r , then (36) holds with y^* replaced by y_{r+1} . By Assumption 2.4, we obtain

$$e^{-\frac{L_\Phi(x,y_r)}{8}} \leq e^{-\frac{L_\Phi(x,y_{r+1})}{8}} e^{-\frac{\Delta(1+o(1))}{8}}.$$

Recall that the distance D_Ω in Ω is defined in Definition 2. From the constructions of y_{r+1} , we have

$$D_\Omega(x, y_{r+1}) = D_\Omega(x, y_r) - 1.$$

Therefore, there exists $h \in [n]$, such that $y_h = x$. By (39) and Assumption 2.4, we obtain

$$e^{-\frac{L_\Phi(x,y)}{8}} \leq e^{-\frac{h\Delta(1+o(1))}{8}}.$$

Since any x in \mathcal{B}_ε can be obtained from y by the community changing process described above, we have

$$I_2 \leq \sum_{l=1}^{\infty} (nk)^l e^{-\frac{l\Delta(1+o(1))}{8}}; \tag{41}$$

The right-hand side of (41) is the sum of geometric series with both initial term and common ratio equal to

$$V := e^{\log k + \log n - \frac{\Delta(1+o(1))}{8}} \tag{42}$$

When (17) holds, we obtain

$$\lim_{n \rightarrow \infty} I_2 = 0 \tag{43}$$

Then the proposition follows from (35) and (43).

4. Community detection on k-community hypergraphs

In this section, we apply the results proved in Section 3 to the exact recovery of the community detection in hypergraphs and also prove conditions when exact recovery does not occur in hypergraphs under the assumption that the number of vertices in each community is unknown.

The main idea to prove Theorem 2.9 is as follows:

- We obtain an lower bound for the probability that the MLE is not in the equivalent class of the true community assignment function; see (50).
- We show this lower bound converges to 1 as $n \rightarrow \infty$. The proof utilises the inequalities regarding the distribution of the maximum of a number of Gaussian random variables as discussed in Section A.

In the case of a hypergraph, from (26), when

$$i = (i_1, i_2, \dots, i_{s-1}) \in [n]^{s-1};$$

$$\theta(x, i, a) = \phi(x(i_1), \dots, x(i_{s-1}), a);$$

we obtain for $x, z \in \Omega$

$$\langle \Phi * \mathbf{A}_x, \Phi * \mathbf{A}_z \rangle \tag{44}$$

$$= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{(\mathbf{A}_x)_{(i_1, \dots, i_s)} (\mathbf{A}_z)_{(i_1, \dots, i_s)}}{\sigma_{(i_1, \dots, i_s)}^2}$$

$$= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{\phi(x(i_1), \dots, x(i_s)) \phi(z(i_1), \dots, z(i_s))}{\sigma_{(i_1, \dots, i_s)}^2}$$

In particular,

$$\|\Phi * \mathbf{A}_x\|^2 = \langle \Phi * \mathbf{A}_x, \Phi * \mathbf{A}_x \rangle$$

$$= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{(\phi(x(i_1), \dots, x(i_s)))^2}{\sigma_{(i_1, \dots, i_s)}^2}$$

Recall that $y \in \Omega_c$ is the true community assignment mapping. Then

$$\hat{y} = \operatorname{argmin}_{x \in \Omega_{\frac{2c}{3}}} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 = \operatorname{argmin}_{x \in \Omega_{\frac{2c}{3}}} f(x)$$

where $f(x)$ is given by (29).

By (30), we obtain that in the hypergraph case

$$f(x) - f(y) \tag{45}$$

$$= \|\Phi * (\mathbf{A}_x - \mathbf{A}_y)\|^2 - 2\langle \mathbf{W}, \Phi * (\mathbf{A}_x - \mathbf{A}_y) \rangle$$

$$= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{(\phi(x(i_1), \dots, x(i_s)) - \phi(y(i_1), \dots, y(i_s)))^2}{\sigma_{(i_1, \dots, i_s)}^2}$$

$$- 2\langle \mathbf{W}, \Phi * (\mathbf{A}_x - \mathbf{A}_y) \rangle$$

Then $f(x) - f(y)$ is a Gaussian random variable with mean value $L_\Phi(x, y)$ and variance $4L_\Phi(x, y)$, where $L_\Phi(x, y)$ is defined by (5).

Proof of Theorem 2.9. When $y^{(a)} \in \Omega$ is defined by (20),

$$t_{y^{(a)}(a), y^{(a)}}(y^{(a)}, y) = 1; \tag{46}$$

$$t_{y^{(a)}, y^{(a)}}(y^{(a)}, y) = n_{y^{(a)}} - 1; \tag{47}$$

$$t_{i, i}(y^{(a)}, y) = n_i; \forall i \in [k] \setminus \{y^{(a)}\}; \tag{48}$$

$$t_{i, j}(y^{(a)}, y) = 0; \forall (i, j) \in [k]^2 \setminus \{(y^{(a)}(a), y^{(a)})\}, \text{ and } i \neq j. \tag{49}$$

and

$$\begin{aligned} n_{y^{(a)}(a)}(y^{(a)}) &= n_{y^{(a)}(a)} + 1; \\ n_{y^{(a)}}(y^{(a)}) &= n_{y^{(a)}} - 1; \\ n_i(y^{(a)}) &= n_i; \forall i \in [k] \setminus \{y^{(a)}(a), y^{(a)}\}. \end{aligned}$$

Moreover,

$$1 - p(\hat{y}; \sigma) \geq \Pr \left(\cup_{a \in [n]} \{f(y^{(a)}) - f(y) < 0\} \right) \tag{50}$$

Since any of the event $\{f(y^{(a)}) - f(y) < 0\}$ implies $\hat{y} \neq y$.

Let $H \subset [n]$ be given as in the assumptions of the proposition. Under Assumption (3) of the proposition when $a \in H$, we have

$$\begin{aligned} & \|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2 \\ &= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{(\phi(y^{(a)}(i_1)), \dots, y^{(a)}(i_s)) - \phi(y(i_1), \dots, y(i_s))}{\sigma_{(i_1, \dots, i_s)}^2} \\ &= L_\Phi(y^{(a)}, y) \\ &= (1 + o(1)) \left\{ \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \hat{i}_j, \dots, i_s) \in ([n] \setminus H)^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, a, i_{j+1}, \dots, i_s)}^2} \right. \\ & \quad \left. \times (\phi(y(i_1), \dots, y^{(a)}(a), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s)))^2 \right\} \end{aligned}$$

Then from (45), we have

$$\begin{aligned} & f(y^{(a)}) - f(y) \\ &= -2\langle \mathbf{W}, \Phi * \mathbf{A}_{y^{(a)}} - \mathbf{A}_y \rangle + (1 + o(1)) \left\{ \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \hat{i}_j, \dots, i_s) \in ([n] \setminus H)^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, a, i_{j+1}, \dots, i_s)}^2} \right. \\ & \quad \left. \times (\phi(y(i_1), \dots, y^{(a)}(a), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s)))^2 \right\}. \end{aligned}$$

Then $1 - p(\hat{y}; \sigma)$ is at least

$$\begin{aligned} & \Pr \left(\cup_{a \in [n]} \{f(y^{(a)}) - f(y) < 0\} \right) \\ & \geq \Pr \left(\max_{a \in [n]} \frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} > 1 \right) \\ & \geq \Pr \left(\max_{a \in H} \frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} > 1 \right) \end{aligned}$$

Let $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ be a partition of $\cup_{s=s_1}^{s_2} [n]^s$ defined by:

$$\begin{aligned} \mathcal{X} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \cup_{s=s_1}^{s_2} [n]^s, \{\alpha_1, \dots, \alpha_s\} \cap H = \emptyset\} \\ \mathcal{Y} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \cup_{s=s_1}^{s_2} [n]^s, |\{i \in [s] : \alpha_i \in H\}| = 1\} \\ \mathcal{Z} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \cup_{s=s_1}^{s_2} [n]^s, |\{i \in [s] : \alpha_i \in H\}| \geq 2\} \end{aligned}$$

For $\eta \in \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, define the random tensor \mathbf{W}_η from the entries of \mathbf{W} as follows:

$$(\mathbf{W}_\eta)_{(i_1, i_2, \dots, i_s)} = \begin{cases} 0 & \text{if } (i_1, \dots, i_s) \notin \eta \\ (\mathbf{W})_{(i_1, \dots, i_s)}, & \text{if } (i_1, \dots, i_s) \in \eta \end{cases}$$

For each $a \in H$, let

$$\begin{aligned} \mathcal{X}_a &= \langle \mathbf{W}_{\mathcal{X}}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle \\ \mathcal{Y}_a &= \langle \mathbf{W}_{\mathcal{Y}}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle \\ \mathcal{Z}_a &= \langle \mathbf{W}_{\mathcal{Z}}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle \end{aligned}$$

For $s \in \{s_1, s_1 + 1, \dots, s_2\}$, let

$$J_s := (j_1, \dots, j_s) \subset [n]^s$$

Explicit computations show that

$$\begin{aligned} & (\mathbf{A}_{y^{(a)}})_{J_s} - (\mathbf{A}_y)_{J_s} \\ &= \begin{cases} \phi(y^{(a)}(j_1), \dots, y^{(a)}(j_s)) - \phi(y(j_1), \dots, y(j_s)) & \text{if } a \in \{j_1, \dots, j_s\} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{51}$$

Claim 4.1. *The followings are true:*

1. $\mathcal{X}_a = 0$ for $a \in H$.
2. For each $a \in H$, the variables \mathcal{Y}_a and \mathcal{Z}_a are independent.

Proof. It is straightforward to check (1). (2) holds because $\mathcal{Y} \cap \mathcal{Z} = \emptyset$. □

For $g \in H$, let $\mathcal{Y}^g \subseteq \mathcal{Y}$ be defined by

$$\mathcal{Y}^g = \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \mathcal{Y} : s \in \{s_1, s_1 + 1, \dots, s_2\}, \exists l \in [s], \text{ s.t. } \alpha_l = g\}.$$

Note that for $g_1, g_2 \in H$ and $g_1 \neq g_2$, $\mathcal{Y}^{g_1} \cap \mathcal{Y}^{g_2} = \emptyset$. Moreover, $\mathcal{Y} = \cup_{g \in H} \mathcal{Y}^g$. Therefore,

$$\mathcal{Y}_a = \sum_{g \in H} \langle \mathbf{W}_{\mathcal{Y}^g}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle$$

Note also that $\langle \mathbf{W}_{\mathcal{Y}^g}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle = 0$, if $g \neq a$. Hence,

$$\mathcal{Y}_a = \sum_{\alpha \in \mathcal{Y}^a} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha}$$

So by (51), we obtain

$$\begin{aligned} \sum_{\alpha \in \mathcal{Y}^a} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha} &= \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{j}, \dots, i_s) \in ([n] \setminus H)^{s-1}} \frac{1}{\sigma_{(i_1, \dots, j-1, a, j+1, \dots, i_s)}} \\ &\quad \{(\phi(y(i_1), \dots, y^{(a)}(a), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s))) (\mathbf{W})_{(i_1, \dots, j-1, a, j+1, \dots, i_s)}\} \end{aligned}$$

Then $\{\mathcal{Y}_g\}_{g \in H}$ is a collection of independent centred Gaussian random variables. Moreover, the variance of \mathcal{Y}_g is equal to

$$\begin{aligned} & \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{j}, \dots, i_s) \in ([n] \setminus H)^{s-1}} \frac{1}{\sigma_{(i_1, \dots, j-1, g, j+1, \dots, i_s)}^2} \\ & \quad (\phi(y(i_1), \dots, y^{(g)}(g), \dots, y(i_s)) - \phi(y(i_1), \dots, y(g), \dots, y(i_s)))^2 \\ & = (1 + o(1)) L_\Phi(y^{(g)}, y) \end{aligned} \tag{52}$$

by Assumption (3) of the proposition.

By Claim 4.1, we obtain

$$\frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} = \frac{2\mathcal{Y}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} + \frac{2\mathcal{Z}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2}$$

Moreover,

$$\max_{a \in H} \frac{2(\mathcal{Y}_a + \mathcal{Z}_a)}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} \geq \max_{a \in H} \frac{2\mathcal{Y}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} - \max_{a \in H} \frac{-2\mathcal{Z}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2}$$

Recall that

$$\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2 = L_\Phi(y^{(a)}, y)$$

By Lemma A.1 about the tail bound result of the maximum of Gaussian random variables, if (A1) holds with N replaced by h , the event

$$E_1 := \left\{ \max_{a \in H} \frac{2\mathcal{Y}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} \geq (1 - \varepsilon) \sqrt{2 \min_{a \in H} \frac{4}{L_\Phi(y^{(a)}, y)} \log h} \right\}$$

has probability at least $1 - e^{-h^\varepsilon}$, and the event

$$E_2 := \left\{ \max_{a \in H} \frac{2\mathcal{Z}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} \leq (1 + \varepsilon) \sqrt{2 \log h \cdot \max_{a \in H} \frac{4\text{Var}(\mathcal{Z}_a)}{(L_\Phi(y^{(a)}, y))^2}} \right\}$$

has probability $1 - h^{-\varepsilon}$.

Moreover, by Assumption (3) of the Proposition and (52),

$$\text{Var} \mathcal{Z}_a = \|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2 - \text{Var}(\mathcal{Y}_a) = o(1)L_\Phi(y^{(a)}, y)$$

Define an event E by

$$E := \left\{ \max_{a \in H} \frac{2\mathcal{Y}_a + 2\mathcal{Z}_a}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} \geq \left(1 - \varepsilon - (1 + \varepsilon)o(1) \sqrt{\frac{\max_{a \in H} L_\Phi(y^{(a)}, y)}{\min_{a \in H} L_\Phi(y^{(a)}, y)}} \right) \times \sqrt{8 \log h \min_{a \in H} \frac{1}{L_\Phi(y^{(a)}, y)}} \right\}$$

Then $E_1 \cap E_2 \subseteq E$.

When n is large, and (22) holds

$$\begin{aligned} & \Pr \left(\max_{a \in H} \frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2} > 1 \right) \\ & \geq \Pr(E) \geq \Pr(E_1 \cap E_2) \geq 1 - \Pr(E_1^c) - \Pr(E_2^c) \rightarrow 1, \end{aligned}$$

as $n \rightarrow \infty$. Then the proposition follows.

4.1. Examples of Assumption 2.2

We shall see some examples of the function $\theta : \Omega \times [k] \times [k] \rightarrow \mathbb{R}$ satisfying Assumption 2.2. We first see an example when θ can uniquely determine the community assignment mapping in Ω .

Example 4.2. Assume $p = k$. For $a, b \in [k]$, $x \in \Omega$

$$\theta(x, a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise.} \end{cases}$$

Then if for all $a \in [p]$ and $j \in [n]$, (7) holds, we have $x(j) = a$ if and only if $z(j) = a$, then $x = z$.

We now see an example when θ cannot uniquely determine the community assignment mapping but determines the community assignment mappings up to the equivalent class as defined in Definition 1.

Example 4.3. Assume $p = n$.

$$\theta(x, i, a) = \begin{cases} 1 & \text{if } x(i) = a; \\ 0 & \text{otherwise.} \end{cases}$$

Then if for all $i, j \in [n]$, (7) holds, we have $x(j) = x(i)$ if and only if $z(j) = z(i)$, then $x \in C(z)$.

Example 4.4. Assume $p = n$, $a \in [k]$ and $i \in [n]$.

$$\theta(x, i, a) = x(i) - a;$$

Then if for all $i, j \in [n]$, (7) holds, we have $x(i) - x(j) = z(i) - z(j)$. This implies that $x(i) = x(j)$ if and only if $z(i) = z(j)$, therefore $x \in C(z)$. If both x and z are surjective onto $[k]$, then $x = z$.

4.2. Example of Theorem 2.9

Example 4.5. Here we see an example about how to apply Theorem 2.9 to the exact recovery of community detection on hypergraphs. Let $y \in \Omega_{n_1, \dots, n_k}$ be the true community assignment mapping. Assume that for any $s \in \{s_1, \dots, s_2\}$, $(i_1, i_2, \dots, i_s), (j_1, j_2, \dots, j_s) \in [n]^s$, we have

$$\sigma_{(i_1, i_2, \dots, i_s)} = \sigma_{(j_1, j_2, \dots, j_s)}$$

whenever

$$y(i_r) = y(j_r), \forall r \in [s];$$

that is, $\sigma_{(i_1, \dots, i_s)}$ depends only on the communities of (i_1, \dots, i_s) under the mapping y . In this case, we can define $\bar{\sigma}: \cup_{s=s_1}^{s_2} [k]^s \rightarrow (0, \infty)$, such that

$$\sigma_{(i_1, \dots, i_s)} = \bar{\sigma}(y(i_1), \dots, y(i_s)), \forall (i_1, \dots, i_s) \in [n]^s \tag{53}$$

Then for any $a \in [n]$,

$$\begin{aligned} L_\Phi(y^{(a)}, y) &= \|\Phi * (\mathbf{A}_{y^{(a)}} - \mathbf{A}_y)\|^2 \\ &= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{(\phi(y^{(a)}(i_1), \dots, y^{(a)}(i_s)) - \phi(y(i_1), \dots, y(i_s)))^2}{(\bar{\sigma}(y(i_1), \dots, y(i_s)))^2} \end{aligned} \tag{54}$$

Moreover, for any $a, b \in [n]$ such that

$$y(a) = y(b); y^{(a)}(a) = y^{(b)}(b)$$

we have

$$L_\Phi(y^{(a)}, y) = L_\Phi(y^{(b)}, y).$$

We consider

$$\min_{(y^{(a)}(a), y^{(a)}(a)) \in [k]^2, y^{(a)}(a) \neq y(a)} L_\Phi(y^{(a)}, y) \tag{55}$$

Assume that when $y(a) = r_0, y^{(a)}(a) = r_1, L_\Phi(y^{(a)}, y)$ achieves its minimum. Let $H \subset y^{-1}(r_0)$, then $h = |H| \leq n_{r_0}$. Assume

$$\lim_{n \rightarrow \infty} \frac{\log n_{r_0}}{\log n} = 1.$$

Then we may choose $h = \frac{n_{r_0}}{\log n}$ such that Assumptions (1)(2) in Theorem 2.9 hold. Moreover, Assumption (4) in Theorem 2.9 holds because if for all $a \in H$, let $y^{(a)}(a) = r_1$, then $L_\Phi(y^{(a)}, y)$ takes the same value for all $a \in H$. There are many mappings $\phi: \cup_{s=s_1}^{s_2} [k]^s \rightarrow \mathbb{R}$ to guarantee Assumption (3) in Theorem 2.9. For example, one may choose

$$\phi(b_1, \dots, b_s) = \begin{cases} 2^s & \text{if } b_1 = \dots = b_s \\ 0 & \text{otherwise.} \end{cases} \tag{56}$$

for $s \in \{s_1, s_1 + 1, \dots, s_2\}$ and $b_1, \dots, b_s \in [k]$. Then from (54), we obtain

$$\begin{aligned} L_\Phi(y^{(a)}, y) &= \sum_{s=s_1}^{s_2} \sum_{(b_1, \dots, b_s) \in [k]^s} \sum_{(d_1, \dots, d_s) \in [k]^s} \\ &\frac{(\phi(d_1, \dots, d_s) - \phi(b_1, \dots, b_s))^2}{(\bar{\sigma}(b_1, \dots, b_s))^2} \left(\prod_{j=1}^s t_{d_j, b_j}(y^{(a)}, y) \right) \end{aligned} \tag{57}$$

From (46)–(49) and (56), we obtain that the terms actually contributing to the sum must satisfy

$$\{(d_1, b_1), \dots, (d_s, b_s)\} \subseteq \{(r_1, r_1), (r_1, r_0)\}$$

or

$$\{(d_1, b_1), \dots, (d_s, b_s)\} \subseteq \{(r_0, r_0), (r_1, r_0)\}$$

Then we obtain

$$L_{\Phi}(y^{(a)}, y) = \sum_{s=s_1}^{s_2} 2^{2s} (L_{0,s} + L_{1,s})$$

where

$$L_{0,s} = \sum_{(b_1, \dots, b_s) \in [k]^s, (d_1, \dots, d_s) \in [k]^s, (d_1, b_1), \dots, (d_s, b_s) \subseteq \{(r_0, r_0), (r_1, r_0)\}} \frac{(\prod_{j=1}^s t_{d_j, b_j}(y^{(a)}, y))}{(\bar{\sigma}(r_0, \dots, r_0))^2}$$

$$L_{1,s} = \sum_{(b_1, \dots, b_s) \in [k]^s, (d_1, \dots, d_s) \in [k]^s, (d_1, b_1), \dots, (d_s, b_s) \subseteq \{(r_1, r_1), (r_1, r_0)\}} \frac{(\prod_{j=1}^s t_{d_j, b_j}(y^{(a)}, y))}{(\bar{\sigma}(b_1, \dots, b_s))^2}$$

Assume

$$\lim_{n \rightarrow \infty} \min\{n_{r_0}, n_{r_1}\} = \infty.$$

If in $\{(d_1, b_1), \dots, (d_s, b_s)\}$, there exist more than one $g \in [s]$, such that $(d_g, b_g) = (r_1, r_0)$, then the sum of such terms will be of order $o((n_{r_0})^{s-1})$ (resp. $o((n_{r_1})^{s-1})$) in $L_{0,s}$ (resp. $L_{1,s}$). Therefore, we obtain

$$L_{0,s} = \frac{s (n_{r_0})^{s-1} (1 + o(1))}{(\bar{\sigma}(r_0, \dots, r_0))^2}$$

To analyse $L_{1,s}$, assume that there exists a positive constant $C > 0$ independent of n , such that

$$0 < C < \frac{\min_{b_1, \dots, b_s \in \{r_0, r_1\}} \bar{\sigma}(b_1, \dots, b_s)}{\max_{b_1, \dots, b_s \in \{r_0, r_1\}} \bar{\sigma}(b_1, \dots, b_s)}, \forall n \in \mathbb{N}, s \in \{s_1, s_1 + 1, \dots, s_2\}. \tag{58}$$

Then we obtain

$$L_{1,s} = \sum_{j=1}^s \frac{(n_{r_1})^{s-1} (1 + o(1))}{(\bar{\sigma}(r_1, \dots, r_1, r_0, r_1, \dots, r_1))^2}$$

Then Assumption (3) of Theorem 2.9 follows from the fact that $|H| = \frac{n_{r_0}}{\log n} = o(n_{r_0})$.

In the special case when all the communities have equal size, we may obtain a sufficient condition that the exact recovery of MLE does not occur in the hypergraph case when the number of communities $k = e^{o(\log n)}$. Since in this case we have $n_1 = n_2 = \dots = n_k \geq e^{\log n - o(\log n)}$, then (21) holds. Choose $h = \frac{n_1}{\log n}$, then Assumptions (1) and (2) of Theorem 2.9 hold.

4.3. Example of Theorem 2.6

Example 4.6. We can also apply Theorem 2.6 to the case of exact recovery of community detection on hypergraphs. Again we consider the case when $\sigma_{(i_1, \dots, i_s)}$ depends only on $(y(i_1), \dots, y(i_s))$. Hence, we may define $\bar{\sigma}$ as in (53).

To check Assumption 2.4, let $y_m, y_{m+1}, x \in \Omega$ be given as in the proof of Proposition 2.6. For the simplicity of notation, we use y instead of y^* . By (57), we obtain

$$\begin{aligned} & L_\Phi(x, y_m) - L_\Phi(x, y_{m+1}) \\ &= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{1}{(\bar{\sigma}(y(i_1), \dots, y(i_s))))^2} \\ & \quad [(\phi(x(i_1), \dots, x(i_s)) - \phi(y_m(i_1), \dots, y_m(i_s)))^2 - (\phi(x(i_1), \dots, x(i_s)) - \phi(y_{m+1}(i_1), \dots, y_{m+1}(i_s))))^2] \\ &= \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [n]^s} \frac{1}{(\bar{\sigma}(y(i_1), \dots, y(i_s))))^2} \{(\phi(y_m(i_1), \dots, y_m(i_s)))^2 - (\phi(y_{m+1}(i_1), \dots, y_{m+1}(i_s)))^2 \\ & \quad - 2\phi(x(i_1), \dots, x(i_s)) [\phi(y_m(i_1), \dots, y_m(i_s)) - \phi(y_{m+1}(i_1), \dots, y_{m+1}(i_s))]\} \end{aligned}$$

For $j, p, q \in [k]$ and $x, y, z \in \Omega$, we define

$$t_{j,p,q}(x, y, z) = |\{i \in [n] : x(i) = j, y(i) = p, z(i) = q\}| = |x^{-1}(j) \cap y^{-1}(p) \cap z^{-1}(q)|.$$

Then

$$\begin{aligned} L_\Phi(x, y_m) - L_\Phi(x, y_{m+1}) &= \sum_{s=s_1}^{s_2} \sum_{(b_1, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, b_s))^2} \sum_{(d_1, \dots, d_s) \in [k]^s} \tag{59} \\ & \left\{ (\phi(d_1, \dots, d_s))^2 \left(\prod_{r=1}^s t_{b_r, d_r}(y, y_m) - \prod_{r=1}^s t_{b_r, d_r}(y, y_{m+1}) \right) - 2 \sum_{(l_1, \dots, l_s) \in [k]^s} \right. \\ & \left. \phi(l_1, \dots, l_s) \phi(d_1, \dots, d_s) \left(\prod_{r=1}^s t_{b_r, d_r, l_r}(y, y_m, x) - \prod_{r=1}^s t_{b_r, d_r, l_r}(y, y_{m+1}, x) \right) \right\} \end{aligned}$$

Recall that $D_\Omega(y_m, y_{m+1}) = 1$, and there exists $u \in [n]$ such that

$$x(u) = j = y_{m+1}(u) \neq y_m(u) = i = y(u).$$

where $i, j \in [k]$ and $i \neq j$, while $y_m(v) = y_{m+1}(v)$ for all the $v \in [n] \setminus \{u\}$. This implies that if $\{d_1, \dots, d_s\} \cap \{i, j\} = \emptyset$, then the corresponding summand in (59) is 0 and does not contribute to the sum. Under the assumption that

1. $(t_{1,1}(x, y), t_{1,2}(x, y), \dots, t_{k,k}(x, y)) \in \mathcal{B}_\varepsilon$ with $w : [k] \rightarrow [k]$ the identity map; and
2. $n_1 \geq n_2 \geq \dots \geq n_k$; and
3. $\min_{(b_1, \dots, b_s) \in [k]^s} |\bar{\sigma}(b_1, \dots, b_s)| \geq B_3 > 0$; and
4. $\lim_{n \rightarrow \infty} \frac{n\varepsilon}{n_1} = 0$.

we obtain

$$\begin{aligned} L_\Phi(x, y_m) - L_\Phi(x, y_{m+1}) &= \sum_{s=s_1}^{s_2} \sum_{g=1}^s \sum_{(b_1, \dots, \hat{b}_g, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, i, \dots, b_s))^2} \sum_{(d_1, \dots, \hat{d}_g, \dots, d_s) \in [k]^s} \\ & \left\{ (\phi(d_1, \dots, i, \dots, d_s))^2 - (\phi(d_1, \dots, j, \dots, d_s))^2 \prod_{r \in [s] \setminus \{g\}} t_{b_r, d_r}(y, y_m) - 2 \sum_{(l_1, \dots, \hat{l}_g, \dots, l_s) \in [k]^s} \right. \\ & \left. \phi(l_1, \dots, j, \dots, l_s) (\phi(d_1, \dots, i, \dots, d_s) - \phi(d_1, \dots, j, \dots, d_s)) \right. \\ & \left. \left(\prod_{r \in [s] \setminus \{g\}} t_{b_r, d_r, l_r}(y, y_m, x) \right) \right\} + O\left(\frac{n_1^{k-2}}{B_3^2}\right) \end{aligned}$$

The identity above can be interpreted as follows. We can classify the terms satisfying $\{d_1, \dots, d_s\} \cap \{i, j\} \neq \emptyset$ by the number

$$N_{ij} = \{l \in [s] : d_l \in \{i, j\}\},$$

and obtain that the leading term of $L_\Phi(x, y_m) - L_\Phi(x, y_{m+1})$ is given by the terms when $N_{i,j} = 1$. Moreover, by Assumption (1) we have

$$\begin{aligned} L_\Phi(x, y_m) - L_\Phi(x, y_{m+1}) &= \sum_{s=s_1}^{s_2} \sum_{g=1}^s \sum_{(b_1, \dots, \hat{b}_g, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, i, \dots, b_s))^2} \\ &\quad \left\{ (\phi(b_1, \dots, i, \dots, b_s))^2 - (\phi(b_1, \dots, j, \dots, b_s))^2 \prod_{r \in [s] \setminus \{g\}} n_r - 2 \sum_{(i_1, \dots, \hat{i}_g, \dots, i_s) \in [k]^s} \right. \\ &\quad \left. \phi(b_1, \dots, j, \dots, b_s) (\phi(b_1, \dots, i, \dots, b_s) - \phi(b_1, \dots, j, \dots, b_s)) \right. \\ &\quad \left. \left(\prod_{r \in [s] \setminus \{g\}} n_r \right) \right\} + o\left(\frac{n_1^{s-2}}{B_3^2}\right) + o\left(\frac{\varepsilon n n_1^{s-2}}{B_3^2}\right) \\ &= \sum_{s=s_1}^{s_2} \sum_{g=1}^s \sum_{(b_1, \dots, \hat{b}_g, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, i, \dots, b_s))^2} \\ &\quad (\phi(b_1, \dots, i, \dots, b_s) - \phi(b_1, \dots, j, \dots, b_s))^2 \prod_{r \in [s] \setminus \{g\}} n_{b_r} \\ &\quad + o\left(\frac{n_1^{s-2}}{B_3^2}\right) + o\left(\frac{\varepsilon n n_1^{s-2}}{B_3^2}\right) \end{aligned}$$

Define

$$\begin{aligned} \Delta := & \min_{i,j \in [k], i \neq j} \sum_{s=s_1}^{s_2} \sum_{g=1}^s \sum_{(b_1, \dots, \hat{b}_g, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, i, \dots, b_s))^2} \\ & \times (\phi(b_1, \dots, i, \dots, b_s) - \phi(b_1, \dots, j, \dots, b_s))^2 \prod_{r \in [s] \setminus \{g\}} n_{b_r} \end{aligned}$$

We further make the assumptions below:

$$\lim_{n \rightarrow \infty} \frac{n_1^{s-2} + \varepsilon n n_1^{s-2}}{B_3^2 \Delta} = 0. \tag{60}$$

Then

$$L_\Phi(x, y_m) - L_\Phi(x, y_{m+1}) \geq \Delta(1 + o(1))$$

Then by Assumption 2.5(1), the exact recovery occurs with probability 1 when $n \rightarrow \infty$ if (15) and (17) hold, and

$$\begin{aligned} & \sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [k]^s} \sum_{(j_1, \dots, j_s) \in [k]^s} (\phi(i_1, \dots, i_s) - \phi(j_1, \dots, j_s))^2 \\ & \times \left(\prod_{r=1}^s t_{i_r, j_r}(x, y) \right) \geq T(n) \end{aligned} \tag{61}$$

when (10) holds.

There are a lot of functions $\phi: \cup_{s=s_1}^{s_2} [k]^s \rightarrow \mathbb{R}$ satisfying (60) and (61) and Assumption 2.2. For example, we may consider the function ϕ as defined in (56). Assume

$$\begin{aligned} \Delta = & \sum_{s=s_1}^{s_2} \sum_{g=1}^s \sum_{(b_1, \dots, \hat{b}_g, \dots, b_s) \in [k]^s} \frac{1}{(\bar{\sigma}(b_1, \dots, r_0, \dots, b_s))^2} \\ & \times (\phi(b_1, \dots, r_0, \dots, b_s) - \phi(b_1, \dots, r_1, \dots, b_s))^2 \prod_{r \in [s] \setminus \{g\}} n_{b_r} \end{aligned}$$

where $r_0, r_1 \in [k]$ and $r_0 \neq r_1$. As in Example 4.5, we obtain that

$$\Delta = \sum_{s=s_1}^{s_2} 2^{2s} (\Delta_{0,s} + \Delta_{1,s}),$$

where

$$\begin{aligned} \Delta_{0,s} &= \frac{s(n_{r_0})^{s-1}}{(\bar{\sigma}(r_0, \dots, r_0))^2} \geq \frac{s((n_{r_0}))^{s-1}}{B_1^2} \\ \Delta_{1,s} &= \sum_{j=1}^s \frac{(n_{r_1})^{s-1}}{(\bar{\sigma}(r_1, \dots, r_1, r_0, r_1, \dots, r_1))^2} \geq \frac{s((n_{r_1}))^{s-1}}{B_1^2} \end{aligned}$$

where the inequality follows from Assumption 2.5(2). It is straightforward to check that when ϕ is given by (56) and (60) hold if (58) holds.

To check (61), note that

$$\begin{aligned} &\sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [k]^s} \sum_{(j_1, \dots, j_s) \in [k]^s} (\phi(i_1, \dots, i_s) - \phi(j_1, \dots, j_s))^2 \times \left(\prod_{r=1}^s t_{i_s j_s}(x, y) \right) \\ &\geq \sum_{s=s_1}^{s_2} \sum_{g \in [s]} \sum_{j \in [s]} \sum_{i \in [k], i \neq j} (\phi(w(i), \dots, w(i)) - \phi(i, \dots, i, j, i, \dots, i))^2 \\ &\quad \times t_{w(i)j}(x, y) \prod_{r=[s] \setminus \{g\}} t_{w(i),i}(x, y) \end{aligned}$$

When (10) holds, the following cases might occur

- w is not a bijection from $[k]$ to $[k]$. In this case, there exists $i, j \in [k]$, such that $w(i) = w(j)$, then when (10) holds, we obtain

$$t_{w(i)j} = t_{w(j)j} \geq \frac{n_j}{k}$$

- w is a bijection from $[k]$ to $[k]$. However, there exists $i \in [k]^2$, such that

$$t_{w(i)j} \leq n_i - \varepsilon n.$$

Let

$$i := w^{-1}(\operatorname{argmax}_{j \in [k] \setminus \{w(i)\}} t_{ij}),$$

then $i \neq j$ and

$$t_{w(i)j}(x, y) \geq \frac{\varepsilon n}{k - 1}$$

When (10) holds and $y \in \Omega_{c^*}$, we have

$$\begin{aligned} &\sum_{s=s_1}^{s_2} \sum_{(i_1, \dots, i_s) \in [k]^s} \sum_{(j_1, \dots, j_s) \in [k]^s} (\phi(i_1, \dots, i_s) - \phi(j_1, \dots, j_s))^2 \times \left(\prod_{r=1}^s t_{i_s j_s}(x, y) \right) \\ &\geq \sum_{s=s_1}^{s_2} s 2^{2s} \left(\frac{n_i}{k} \right)^{s-1} \min \left\{ \frac{n_j}{k}, \frac{\varepsilon n}{k - 1} \right\} \geq \sum_{s=s_1}^{s_2} \frac{2^{2s} s n^s}{(ck)^{s-1} \max \left\{ ck, \frac{k-1}{\varepsilon} \right\}} \end{aligned}$$

Let

$$T(n) := \sum_{s=s_1}^{s_2} \frac{2^{2s} s n^s}{(ck)^{s-1} \max \left\{ ck, \frac{k-1}{\varepsilon} \right\}}$$

Then we obtain (61).

5. Community detection on Gaussian mixture models with fixed number of vertices in each community

In this section, we consider the MLE restricted to the sample space consisting of all the mappings satisfying the condition that the number of vertices in each community is the same as that of the true community assignment mapping $y \in \Omega_{n_1, \dots, n_k}$. Again we shall prove a sufficient condition for the occurrences of exact recovery.

Let $x \in \Omega_{n_1, \dots, n_k}$. By (28),

$$\check{y} := \operatorname{argmin}_{x \in \Omega_{n_1, \dots, n_k}} \|\Phi * (\mathbf{K}_y - \mathbf{A}_x)\|^2 = \operatorname{argmin}_{x \in \Omega_{n_1, \dots, n_k}} f(x)$$

Recall that $f(x)$ is defined as in (29). Recall also that $f(x) - f(y)$ is a Gaussian random variable with mean value $L_\Phi(x, y)$ and variance $4L_\Phi(x, y)$.

For each $x \in \Omega_{n_1, \dots, n_k}$, let

$$C^*(x) := C(x) \cap \Omega_{n_1, \dots, n_k};$$

that is, $C^*(x)$ consists of all the community assignment mappings in Ω_{n_1, \dots, n_k} that are equivalent to x in the sense of Definition 1. Let

$$\bar{\Omega}_{n_1, \dots, n_k} := \{C^*(x) : x \in \Omega_{n_1, \dots, n_k}\};$$

that is, $\bar{\Omega}_{n_1, \dots, n_k}$ consists of all the equivalence classes in Ω_{n_1, \dots, n_k} .

Lemma 5.1. *For $x, z \in \Omega_{n_1, \dots, n_k}$. If $x \in C^*(z)$, then*

$$f(x) = f(z).$$

Proof. The lemma follows from Lemma 3.1. □

Define

$$p(\check{y}; \sigma) := \Pr(\check{y} \in C(y)) = \Pr\left(f(\check{y}) < \min_{C^*(x) \in (\bar{\Omega}_{n_1, \dots, n_k} \setminus \{C^*(y)\})} f(x)\right)$$

Then

$$\begin{aligned} 1 - p(\check{y}; \sigma) &\leq \sum_{C^*(x) \in (\bar{\Omega}_{n_1, \dots, n_k} \setminus \{C^*(y)\})} \Pr(f(x) - f(y) \leq 0) \\ &= \sum_{C^*(x) \in (\bar{\Omega}_{n_1, \dots, n_k} \setminus \{C^*(y)\})} \Pr_{\xi \in \mathcal{N}(0,1)}\left(\xi \geq \frac{L_\Phi(x, y)}{2}\right) \\ &\leq \sum_{C^*(x) \in (\bar{\Omega}_{n_1, \dots, n_k} \setminus \{C^*(y)\})} e^{-\frac{L_\Phi(x, y)}{8}} \end{aligned} \tag{62}$$

Lemma 5.2. *Let $y \in \Omega_{n_1, \dots, n_k} \cap \Omega_c$ be the true community assignment mapping. Let $x \in \Omega_{n_1, \dots, n_k}$. For $i \in [k]$, let $w(i) \in [k]$ be defined as in (34). Then*

1. When $\varepsilon \in (0, \frac{\varepsilon}{k})$ and $(t_{1,1}(x, y), \dots, t_{k,k}(x, y)) \in \mathcal{B}_\varepsilon$, w is a bijection from $[k]$ to $[k]$.
2. Assume there exist $i, j \in [k]$, such that $n_i \neq n_j$. If

$$\varepsilon < \min_{i, j \in [k] : n_i \neq n_j} \left| \frac{n_i - n_j}{n} \right| \tag{63}$$

Then for any $i \in [k]$,

$$n_i = |y^{-1}(i)| = |y^{-1}(w(i))| = n_{w(i)}. \tag{64}$$

Proof. See Lemma 6.6 of [13]. □

Definition 3. Let $l \geq 2$ be a positive integer. Let $x, y \in \Omega_{n_1, \dots, n_k}$. We say l distinct communities $(i_1, \dots, i_l) \in [k]^l$ is an l -cycle for (x, y) , if $t_{i_{s-1}, i_s}(x, y) > 0$ for all $2 \leq s \leq l + 1$, where $i_{l+1} := i_1$.

Lemma 5.3. Let $x, y \in \Omega_{n_1, \dots, n_k}$ and $x \neq y$. Then there exists an l -cycle for (x, y) with $2 \leq l \leq k$.

Proof. See Lemma 3.3 of [13]. □

Lemma 5.4. For any $x, y \in \Omega_{n_1, \dots, n_k}$, $L_\Phi(x, y) \geq 0$, where the equality holds if and only if $x \in C^*(y)$.

Proof. The lemma follows from Lemma 3.3. □

Lemma 5.5. Suppose that Assumption 2.4 holds. Then Assumption 2.7 holds.

Proof. Let $z_0 = y_m$ and $z_j = y_h$. For $i \in [j - 1]$, define $z_i \in \Omega$ by

$$z_i(v) = \begin{cases} z_{i-1}(v) & \text{if } v \in [n] \setminus \{u_i\} \\ x(u_i) & \text{if } v = u_i \end{cases}$$

Then for any $i \in [j]$,

$$D_\Omega(z_i, z_{i-1}) = 1.$$

by Assumption 2.4, we obtain

$$L_\Phi(x, z_{i-1}) - L_\Phi(x, z_i) \geq \Delta(1 + o(1)), \quad \forall i \in [j] \tag{65}$$

summing over all the $i \in [j]$, we obtain (18). □

Proof of Theorem 2.8. Let

$$\Gamma := \sum_{C^*(x) \in (\overline{\Omega}_{n_1, \dots, n_k} \setminus \{C^*(y)\})} e^{-\frac{L_\Phi(x,y)}{8}}.$$

By (62), it suffices to show that $\lim_{n \rightarrow \infty} \Gamma = 0$.

Let

$$0 < \varepsilon < \min \left(\frac{2c}{3k}, \min_{i,j \in [k], n_i \neq n_j} \left| \frac{n_i - n_j}{n} \right| \right)$$

Note that

$$\Gamma \leq \Gamma_1 + \Gamma_2;$$

where

$$\Gamma_1 = \sum_{C^*(x) \in \overline{\Omega}_{n_1, \dots, n_k} : (t_{1,1}(x,y), \dots, t_{k,k}(x,y)) \in (\mathcal{B} \setminus \mathcal{B}_\varepsilon), C^*(x) \neq C^*(y)} e^{-\frac{L_\Phi(x,y)}{8}}$$

and

$$\Gamma_2 = \sum_{C^*(x) \in \overline{\Omega}_{n_1, \dots, n_k} : (t_{1,1}(x,y), \dots, t_{k,k}(x,y)) \in \mathcal{B}_\varepsilon, C(x) \neq C(y)} e^{-\frac{L_\Phi(x,y)}{8}}. \tag{66}$$

Under Assumption 2.5, by Lemma 3.5 we have

$$0 \leq \Gamma_1 \leq k^n e^{-\frac{T(n)}{B_1^2}}$$

By (15), we have

$$\lim_{n \rightarrow \infty} \Gamma_1 = 0. \tag{67}$$

Now let us consider Γ_2 . Recall that $y \in \Omega_{n_1, \dots, n_k} \cap \Omega_c$ is the true community assignment mapping. Let w be the bijection from $[k]$ to $[k]$ as defined in (34). Let $y^* \in \Omega$ be defined by:

$$y^*(z) = w(y(z)), \quad \forall z \in [n].$$

Then $y^* \in C(y)$. By Part (2) of Lemma 5.2, we obtain that for $i \in [k]$

$$|(y^*)^{-1}(i)| = |y^{-1}(w^{-1}(i))| = |y^{-1}(i)|;$$

therefore $y^* \in \Omega_{n_1, \dots, n_k}$. Moreover, x and y^* satisfies

$$t_{i,i}(x, y^*) \geq n_i(y^*) - n\varepsilon, \quad \forall i \in [k]. \tag{68}$$

If $x \neq y^*$, by Lemma 5.3, there exists an l -cycle (i_1, \dots, i_l) for (x, y^*) with $2 \leq l \leq k$. Then for each $2 \leq a \leq (l + 1)$, choose an arbitrary vertex u_m in $S_{i_{m-1}, i_m}(x, y^*)$, and let $y_1(u_m) = i_{m-1}$, where $i_{l+1} := i_1$. For any vertex $z \in [n] \setminus \{u_2, \dots, u_{l+1}\}$, let $y_1(z) = y^*(z)$.

Note that $y_1 \in \Omega_{n_1, \dots, n_k}$. Moreover, for $1 \leq m \leq l$, we have

$$\begin{aligned} t_{i_m, i_m}(x, y^*) + 1 &= t_{i_m, i_m}(x, y_1); \\ t_{i_m, i_{m+1}}(x, y^*) - 1 &= t_{i_m, i_{m+1}}(x, y_1) \end{aligned}$$

and

$$t_{a,b}(x, y^*) = t_{a,b}(x, y_1), \quad \forall (a, b) \notin \{(i_m, i_m), (i_m, i_{m+1})\}_{s=1}^l.$$

When $(t_{1,1}(x, y), \dots, t_{k,k}(x, y)) \in \mathcal{B}_\varepsilon$, From Assumption 2.4 and Lemma 5.5, we obtain

$$L_\Phi(x, y_1) - L_\Phi(x, y) = L_\Phi(x, y_1) - L_\Phi(x, y^*) \leq -l\Delta(1 + o(1))$$

Therefore,

$$e^{-\frac{L_\Phi(x,y)}{8}} \leq e^{-\frac{L_\Phi(x,y_1)}{8}} e^{-\frac{l\Delta(1+o(1))}{8}} \tag{69}$$

If $y_1 \neq x$, we find an l_2 -cycle ($2 \leq l_2 \leq k$) for (x, y_1) , change community assignments along the l_2 -cycle as above, and obtain another community assignment mapping $y_2 \in \Omega_{n_1, \dots, n_k}$, and so on. Let $y_0 := y$, and note that for each $r \geq 1$, if y_r is obtained from y_{r-1} by changing colours along an l_r cycle for (x, y_{r-1}) , we have

$$D_\Omega(x, y_r) = D_\Omega(x, y_{r-1}) - l_r$$

Therefore, finally we can obtain x from y by changing colours along at most $\lfloor \frac{n}{2} \rfloor$ cycles. By similar arguments as those used to derive (69), we obtain that for each r

$$e^{-\frac{L_\Phi(x,y_{r-1})}{8}} \leq e^{-\frac{L_\Phi(x,y_r)}{8}} e^{-\frac{l_r\Delta(1+o(1))}{8}}$$

Therefore, if $y_h = x$ for some $1 \leq h \leq \lfloor \frac{n}{2} \rfloor$, we have

$$e^{-\frac{L_\Phi(x,y)}{8}} \leq e^{-\frac{L_\Phi(x,y_{h-1})}{8}} e^{-\frac{(\sum_{r=1}^{h-1} l_r)\Delta(1+o(1))}{8}}.$$

By Assumption 2.4 and Lemma 5.5, we obtain

$$L_\Phi(x, y_{h-1})^2 \geq l_h\Delta(1 + o(1))$$

Therefore,

$$e^{-\frac{L_\Phi(x,y)}{8}} \leq \prod_{i \in [h]} e^{-\frac{l_i\Delta(1+o(1))}{8}}.$$

Note also that for any $r_1 \neq r_2$, in the process of obtaining y_{r_1} from y_{r_1-1} and the process of obtaining y_{r_2} from y_{r_2-1} , we change community assignments on disjoint sets of vertices. Hence, the order of these steps of changing community assignments along cycles does not affect the final community assignment mapping we obtain. From (66), we have

$$\Gamma_2 \leq \prod_{l=2}^k \left(\sum_{m_l=0}^{\infty} (nk)^{m_l} e^{-\frac{(1+o(1))\Delta l m_l}{8}} \right) - 1. \tag{70}$$

On the right-hand side of (70), when expanding the product, each summand has the form:

$$\left[(nk)^{2m_2} e^{-\frac{(1+o(1))\Delta 2m_2}{8}} \right] \cdot \left[(nk)^{3m_3} e^{-\frac{(1+o(1))\Delta 3m_3}{8}} \right] \cdot \dots \cdot \left[(nk)^{km_k} e^{-\frac{(1+o(1))\Delta km_k}{8}} \right]$$

where the factor $\left[(nk)^{2m_2} e^{-\frac{(1+\alpha(1))\Delta 2m_2}{8}} \right]$ represents that we changed along 2-cycles m_2 times, the factor $\left[(nk)^{3m_3} e^{-\frac{(1+\alpha(1))\Delta 3m_3}{8}} \right]$ represents that we changed along 3-cycles m_3 times, and so on. Moreover, each time we changed along a l -cycle, we need to first determine the l different colours involved in the l -cycle, and there are at most k^l different l -cycles; we then need to choose l vertices to change colours, and there are at most n^l choices. It is straightforward to check that if σ satisfies (17), then

$$\lim_{n \rightarrow \infty} nke^{-\frac{(1+\alpha(1))\Delta}{8}} = 0.$$

Therefore, we have

$$\sum_{m_l=0}^{\infty} (nk)^{m_l} e^{-\frac{(1+\alpha(1))\Delta m_l}{8}} \leq \frac{1}{1 - e^{\log k + \log n - \frac{(1+\alpha(1))\Delta}{8}}};$$

when n is sufficiently large and ε is sufficiently small. Let

$$\Psi := \prod_{l=2}^k \left(\sum_{m_l=0}^{\infty} (nk)^{m_l} e^{-\frac{(1+\alpha(1))m_l\Delta}{8}} \right).$$

Since $\log(1+x) \leq x$ for $x \geq 0$, we have

$$\begin{aligned} 0 \leq \log \Psi &= \sum_{l=2}^k \log \left(1 + \sum_{m_l=1}^{\infty} (nk)^{m_l} e^{-\frac{(1+\alpha(1))\Delta m_l}{8}} \right) \\ &\leq \sum_{l=2}^k \sum_{m_l=1}^{\infty} (nk)^{m_l} e^{-\frac{(1-\delta)\Delta m_l}{8}} \\ &\leq \sum_{l=2}^k \frac{\left(nke^{-\frac{(1-\delta)\Delta}{8}} \right)^l}{1 - \left(nke^{-\frac{(1-\delta)\Delta}{8}} \right)^l} \\ &\leq \frac{\left(nke^{-\frac{(1-\delta)\Delta}{8}} \right)^2}{\left[1 - \left(nke^{-\frac{(1-\delta)\Delta}{8}} \right)^2 \right] \left[1 - \left(nke^{-\frac{(1-\delta)\Delta}{8}} \right) \right]} \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Then

$$0 \leq \lim_{n \rightarrow \infty} \Gamma_2 \leq \lim_{n \rightarrow \infty} e^{\log \Sigma} - 1 = 0. \tag{71}$$

Then the proposition follows from (67) and (71).

6. Community detection on hypergraphs with fixed number of vertices in each community

In this section, we study community detection on hypergraphs under the assumption that the number of vertices in each community is known and fixed. We shall prove a condition when exact recovery does not occur.

Recall that $y \in \Omega_{n_1, \dots, n_k}$ is the true community assignment mapping.

Proof of Theorem 2.10. When $y^{(ab)} \in \Omega_{n_1, \dots, n_k}$ is defined by (23):

$$\begin{aligned} t_{y^{(ab)}(a), y^{(a)}}(y^{(ab)}, y) - 1 &= t_{y^{(ab)}(a), y^{(a)}}(y, y) = t_{y^{(b)}, y^{(a)}}(y, y) = 0 \\ t_{y^{(b)}, y^{(b)}}(y^{(ab)}, y) + 1 &= t_{y^{(b)}, y^{(b)}}(y, y) = n_{y^{(b)}} \\ t_{y^{(ab)}(b), y^{(b)}}(y^{(ab)}, y) - 1 &= t_{y^{(a)}, y^{(b)}}(y^{(ab)}, y) - 1 = t_{y^{(a)}, y^{(b)}}(y, y) = 0 \\ t_{y^{(a)}, y^{(a)}}(y^{(ab)}, y) + 1 &= t_{y^{(a)}, y^{(a)}}(y, y) = n_{y^{(a)}}. \end{aligned}$$

and

$$t_{ij}(y^{(ab)}, y) = t_{ij}(y), \quad \forall (i, j) \in ([k]^2 \setminus \{(y(a), y(a)), (y(a), y(b)), (y(b), y(a)), (y(b), y(b))\})$$

Note that

$$1 - p(\check{y}; \sigma) \geq \Pr \left(\bigcup_{a,b \in [n], y(a) \neq y(b)} (f(y^{(ab)}) - f(y) < 0) \right),$$

since any of the event $(f(y^{(ab)}) - f(y) < 0)$ implies $\check{y} \neq y$. By (45), we obtain that $f(y^{(ab)}) - f(y)$ is a Gaussian random variable with mean value $\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2$ and variance $4\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2$. So $1 - p(\check{y}; \sigma)$ is at least

$$\begin{aligned} & \Pr \left(\bigcup_{a,b \in [n], y(a) \neq y(b)} (f(y^{(ab)}) - f(y) < 0) \right) \\ & \geq \Pr \left(\max_{a,b \in [n], y(a) \neq y(b)} \frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2} > 1 \right) \end{aligned}$$

Let H_1 and H_2 be given as in the assumptions of the proposition. Then

$$1 - p(\check{y}; \sigma) \geq \Pr \left(\max_{a \in H_1, b \in H_2} \frac{2\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2} > 1 \right)$$

Let $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ be a partition of $[n]^s$ defined by:

$$\begin{aligned} \mathcal{X} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in [n]^s : s \in \{s_1, s_1 + 1, \dots, s_2\}, \{\alpha_1, \dots, \alpha_s\} \cap (H_1 \cup H_2) = \emptyset\} \\ \mathcal{Y} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in [n]^s : s \in \{s_1, s_1 + 1, \dots, s_2\}, |r \in [s] : \alpha_r \in (H_1 \cup H_2)| = 1\} \\ \mathcal{Z} &= \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in [n]^s : s \in \{s_1, s_1 + 1, \dots, s_2\}, |r \in [s] : \alpha_r \in (H_1 \cup H_2)| \geq 2\} \end{aligned}$$

For $\eta \in \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, define a random tensor \mathbf{W}_η from the entries of \mathbf{W} as follows:

$$(\mathbf{W}_\eta)_{(a_1, \dots, a_s)} = \begin{cases} 0 & \text{if } (a_1, \dots, a_s) \notin \eta \\ \mathbf{W}_{(a_1, \dots, a_s)}, & \text{if } (a_1, \dots, a_s) \in \eta \end{cases}$$

For each $u \in H_1$ and $v \in H_2$, let

$$\begin{aligned} \mathcal{X}_{uv} &= \langle \mathbf{W}_{\mathcal{X}}, \Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y) \rangle \\ \mathcal{Y}_{uv} &= \langle \mathbf{W}_{\mathcal{Y}}, \Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y) \rangle \\ \mathcal{Z}_{uv} &= \langle \mathbf{W}_{\mathcal{Z}}, \Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y) \rangle \end{aligned}$$

Lemma 6.1. *The followings are true:*

1. $\mathcal{X}_{uv} = 0$ for $u \in H_1$ and $v \in H_2$.
2. For each $u \in H_1$ and $v \in H_2$, the variables \mathcal{Y}_{uv} and \mathcal{Z}_{uv} are independent.
3. Each \mathcal{Y}_{uv} can be decomposed into $Y_u + Y_v$ where $\{Y_u\}_{u \in H_1} \cup \{Y_v\}_{v \in H_2}$ is a collection of i.i.d. Gaussian random variables.

Proof. Note that for $J_s = (j_1, j_2, \dots, j_s) \in [n]^s$,

$$\begin{aligned} & (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y)_{J_s} \\ &= \begin{cases} \phi(y^{(uv)}(j_1), y^{(uv)}(j_2), \dots, y^{(uv)}(j_s)) - \phi(y(j_1), y(j_2), \dots, y(j_s)) & \text{if } \{a, b\} \cap \{j_1, \dots, j_s\} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{72}$$

It is straightforward to check (1). (2) holds because $\mathcal{Y} \cap \mathcal{Z} = \emptyset$.

For $g \in H_1 \cup H_2$, let $\mathcal{Y}_g \subseteq \mathcal{Y}$ be defined by:

$$\mathcal{Y}_g = \{\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s) \in \mathcal{Y} : g \in \{\alpha_1, \dots, \alpha_s\}\}.$$

Note that for $g_1, g_2 \in H_1 \cup H_2$ and $g_1 \neq g_2$, $\mathcal{Y}_{g_1} \cap \mathcal{Y}_{g_2} = \emptyset$. Moreover, $\mathcal{Y} = \bigcup_{g \in H_1 \cup H_2} \mathcal{Y}_g$. Therefore,

$$\mathcal{Y}_{ab} = \sum_{g \in H_1 \cup H_2} \langle \mathbf{W}_{\mathcal{Y}_g}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle$$

Note also that $\langle \mathbf{W}_{\mathcal{Y}_g}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle = 0$, if $g \notin \{a, b\}$. Hence,

$$\mathcal{Y}_{ab} = \sum_{\alpha \in \mathcal{Y}_a \cup \mathcal{Y}_b} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha}$$

So, we can define

$$Y_a := \sum_{\alpha \in \mathcal{Y}_a} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha}$$

By (72), we obtain

$$\begin{aligned} Y_a &= \sum_{\alpha \in \mathcal{Y}_a} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha} \\ &= \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i}_j, \dots, i_s) \in ([n] \setminus (H_1 \cup H_2))^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, a, i_{j+1}, \dots, i_s)}} \\ &\quad \left\{ (\phi(y(i_1), \dots, y^{(ab)}(a), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s))) (\mathbf{W})_{(i_1, \dots, i_{j-1}, a, i_{j+1}, \dots, i_s)} \right\} \end{aligned}$$

Similarly, define

$$\begin{aligned} Y_b &:= \sum_{\alpha \in \mathcal{Y}_b} \frac{(\mathbf{W})_\alpha \cdot \{(\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)_\alpha\}}{\sigma_\alpha} \\ &= \sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i}_j, \dots, i_s) \in ([n] \setminus (H_1 \cup H_2))^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, b, i_{j+1}, \dots, i_s)}} \\ &\quad \left\{ (\phi(y(i_1), \dots, y^{(ab)}(b), \dots, y(i_s)) - \phi(y(i_1), \dots, y(b), \dots, y(i_s))) (\mathbf{W})_{(i_1, \dots, i_{j-1}, b, i_{j+1}, \dots, i_s)} \right\} \end{aligned}$$

Then $\mathcal{Y}_{ab} = Y_a + Y_b$ and $\{Y_g\}_{g \in H_1 \cup H_2}$ is a collection of independent Gaussian random variables. Moreover, the variance of Y_g is

$$\begin{aligned} &\sum_{s=s_1}^{s_2} \sum_{j=1}^s \sum_{(i_1, \dots, \widehat{i}_j, \dots, i_s) \in ([n] \setminus (H_1 \cup H_2))^{s-1}} \frac{1}{\sigma_{(i_1, \dots, i_{j-1}, g, i_{j+1}, \dots, i_s)}^2} \\ &\quad (\phi(y(i_1), \dots, y(b), \dots, y(i_s)) - \phi(y(i_1), \dots, y(a), \dots, y(i_s)))^2 \end{aligned}$$

By Assumption (6) of the proposition, this is independent of g . □

By Lemma 6.1, we obtain

$$\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle = Y_a + Y_b + \mathcal{L}_{ab}$$

Moreover,

$$\begin{aligned} \max_{u \in H_1, v \in H_2} Y_u + Y_v + \mathcal{L}_{uv} &\geq \max_{u \in H_1, v \in H_2} (Y_u + Y_v) - \max_{u \in H_1, v \in H_2} (-\mathcal{L}_{uv}) \\ &= \max_{u \in H_1} Y_u + \max_{v \in H_2} Y_v - \max_{u \in H_1, v \in H_2} (-\mathcal{L}_{uv}) \end{aligned}$$

By Lemma A.1, we obtain that when ε, h satisfy (A1) with N replaced by h , each one of the following two events

$$\begin{aligned} F_1 &:= \left\{ \max_{u \in H_1} \frac{Y_u}{\|\Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y)\|^2} \geq (1 - \varepsilon) \sqrt{2 \log h \cdot \min_{u \in H_1} \frac{\text{Var}(Y_u)}{(L_\Phi(y^{(u,v)}, y))^2}} \right\} \\ F_2 &:= \left\{ \max_{v \in H_2} \frac{Y_v}{\|\Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y)\|^2} \geq (1 - \varepsilon) \sqrt{2 \log h \cdot \min_{v \in H_2} \frac{\text{Var}(Y_v)}{(L_\Phi(y^{(u,v)}, y))^2}} \right\} \end{aligned}$$

has probability at least $1 - e^{-h^\varepsilon}$. Moreover, the event

$$F_3 := \left\{ \max_{u \in H_1, v \in H_2} \frac{\mathcal{L}_{uv}}{\|\Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y)\|^2} \leq (1 + \varepsilon) \sqrt{2 \log(2h) \cdot \max_{u \in H_1, v \in H_2} \frac{\text{Var}(Z_{uv})}{(L_\Phi(y^{(uv)}, y))^2}} \right\}$$

occurs with probability at least $1 - h^{-2\varepsilon}$. Then by Assumption (4) of the proposition, we have

$$\begin{aligned} \text{Var} \mathcal{L}_{uv} &= \|\Phi * (\mathbf{A}_{y^{(uv)}} - \mathbf{A}_y)\|^2 - \text{Var}(Y_u) - \text{Var}(Y_v) \\ &= L_\Phi(y^{(uv)}, y) - (1 + o(1))L_\Phi(y^{(uv)}, y) \\ &= o(1)L_\Phi(y^{(uv)}, y). \end{aligned}$$

By Assumption (5) of the proposition, for any $u \in H_1$ and $v \in H_2$, we have

$$\text{Var}(Y_u) = \text{Var}(Y_v).$$

Moreover, by Assumption (4) of the proposition,

$$\text{Var}(Y_u) + \text{Var}(Y_v) = (1 + o(1))L_\Phi(y^{(uv)}, y).$$

Hence, the probability of the event

$$\begin{aligned} F &:= \left\{ \max_{a \in H_1, b \in H_2} \frac{\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2} \geq \frac{(1 - \varepsilon)\sqrt{2 \log h}}{\max_{u \in H_1, v \in H_2} L_\Phi(y^{(u,v)}, y)} \right. \\ &\quad \left. \left(\sqrt{\min_{u \in H_1} \text{Var}(Y_u)} + \sqrt{\min_{v \in H_2} \text{Var}(Y_v)} - (1 + o(1))\sqrt{\max_{u \in H_1, v \in H_2} \text{Var}(Z_{uv})} \right) \right\} \\ &= \left\{ \max_{a \in H_1, b \in H_2} \frac{\langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2} \geq \frac{2(1 - \varepsilon)\sqrt{\log h}}{\sqrt{\max_{u \in H_1, v \in H_2} L_\Phi(y^{(u,v)}, y)}} (1 + o(1)) \right\} \end{aligned}$$

is at least

$$\begin{aligned} \Pr(F_1 \cap F_2 \cap F_3) &= 1 - \Pr((F_1)^c \cup (F_2)^c \cup (F_3)^c) \\ &\geq 1 - \Pr((F_1)^c) - \Pr((F_2)^c) - \Pr((F_3)^c) \\ &\geq 1 - 2e^{-h^\varepsilon} - h^{-2\varepsilon}. \end{aligned}$$

When (25) holds, we have

$$\Pr \left(\max_{a, b \in [n], y^{(a)} \neq y^{(b)}} \frac{2 \langle \mathbf{W}, \Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y) \rangle}{\|\Phi * (\mathbf{A}_{y^{(ab)}} - \mathbf{A}_y)\|^2} > 1 \right) \geq \Pr(F) \rightarrow 1,$$

as $n \rightarrow \infty$. Then the proposition follows.

7. An example

Assume n vertices are divided into group I and group II such that group I contains $\lfloor \alpha n \rfloor$ vertices and group II contains $n - \lfloor \alpha n \rfloor$ vertices, where $\alpha \in (0, 1)$. Let $y : [n] \rightarrow [2]$ be the true community assignment mapping. For any $x \in \Omega$, define

$$(\mathbf{A}_x)_{i,j} := \begin{cases} 1 & \text{if } x(i) = x(j) = 1 \\ 0 & \text{if } x(i) \neq x(j) \\ 2 & \text{if } x(i) = x(j) = 2 \end{cases}$$

and

$$\sigma_{i,j} = \begin{cases} \delta_1 & \text{if } y(i) = y(j) = 1 \\ \delta_0 & \text{if } y(i) \neq y(j) \\ \delta_2 & \text{if } y(i) = y(j) = 2. \end{cases}$$

Then

$$L_\Phi(x, y) = \sum_{i \in [n]} \sum_{j \in [n]} \frac{1}{\sigma_{ij}^2} [(\mathbf{A}_x)_{ij} - (\mathbf{A}_y)_{ij}]^2$$

$$= \frac{[t_{1,2}(x, y)]^2}{\delta_2^2} + \frac{[t_{2,1}(x, y)]^2}{\delta_1^2} + \frac{2t_{1,1}(x, y)t_{1,2}(x, y)}{\delta_0^2} + \frac{2t_{1,1}(x, y)t_{2,1}(x, y)}{\delta_1^2}$$

$$+ \frac{8t_{1,2}(x, y)t_{2,2}(x, y)}{\delta_2^2} + \frac{8t_{2,1}(x, y)t_{2,2}(x, y)}{\delta_0^2}$$

Assume

$$\frac{\alpha}{\delta_0^2} + \frac{4(1 - \alpha)}{\delta_2^2} = \frac{\alpha}{\delta_1^2} + \frac{4(1 - \alpha)}{\delta_0^2} := Q \tag{73}$$

Note that (73) holds whenever $\delta_0 = \delta_1 = \delta_2$, which corresponds to the case when the Gaussian perturbations for each pair of vertices are i.i.d.; however, (73) may also hold even when $\delta_0, \delta_1, \delta_2$ are not all equal.

When (73) holds, one may check that Assumption 2.4 holds with

$$\Delta = 2nQ$$

and Assumption 2.3 holds with

$$R(n) = 2Q\epsilon n^2$$

We obtain that if

$$Q > \frac{4(1 + \delta) \log n}{n}$$

then a.s. exact recovery occurs; if

$$Q < \frac{4(1 - \delta) \log n}{n}$$

Then a.s. exact recovery does not occur by Theorem 2.9.

In the following tables, we let $(n_1, n_2) = (10, 6)$, $n = 16$ and $\alpha = 0.625$. In this case, we have

$$\frac{4 \log n}{n} = \log(2) \approx 0.6931$$

For each triple $(\delta_0, \delta_1, \delta_2)$ satisfying (73), we perform the MLE for $m = 200$ times to compute \hat{y} , and let m_+ be the number of experiments with $\hat{y} = y$. The values $\frac{m_+}{m}$ are shown in the following tables.

$(\delta_0, \delta_1, \delta_2)$	(1,1,1)	(1.2,1.2,1.2)	(1.26,1.26,1.26)	(1.3,1.3,1.3)	(1.5,1.5,1.5)	(1.6,1.6,1.6)	(1.7,1.7,1.7)
Q	2.125	1.4757	1.3385	1.2574	0.9444	0.8301	0.7353
$\frac{Qn}{4 \log n}$	3.0657	2.1290	1.9310	1.8140	1.3625	1.1975	1.0608
$\frac{m_+}{m}$	1	1	1	0.995	0.960	0.955	0.930

$(\delta_0, \delta_1, \delta_2)$	(1.8,1.8,1.8)	(2,2,2)	(2.5,2.5,2.5)	(3,3,3)	(5,5,5)	(7,7,7)	(8,8,8)
Q	0.6559	0.5313	0.4390	0.2361	0.0850	0.0434	0.0332
$\frac{Qn}{4 \log n}$	0.9462	0.7664	0.6334	0.3406	0.1226	0.0626	0.0479
$\frac{m_+}{m}$	0.910	0.810	0.660	0.395	0.06	0	0

$(\delta_0, \delta_1, \delta_2)$	(0.6769,1.2,0.8)	(0.8069,2,1)	(1.0938,2,1.3)	(1.3016,3,1.6)
Q	3.7078	2.4598	1.4099	0.9549
$\frac{Qn}{4 \log n}$	5.3493	3.5488	2.0341	1.3776
$\frac{m+}{m}$	1	1	0.995	0.980
$(\delta_0, \delta_1, \delta_2)$	(1.9997,1.6,1.8)	(3.6228,2.5,3)	(7.2457,5,6)	(12.4864,5,7)
Q	0.6193	0.2143	0.0536	0.0346
$\frac{Qn}{4 \log n}$	0.8934	0.3091	0.0773	0.0499
$\frac{m+}{m}$	0.910	0.270	0.025	0

Financial support. ZL's research is supported by National Science Foundation grant 1608896 and Simons Foundation grant 638143.

Competing interests. None.

References

- [1] Abbe, E. (2018) Community detection and stochastic block models: Recent developments. *J Mach Learn Res* **18**, 1–86.
- [2] Abbe, E. & Sandon, C. (2015). Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 670–688.
- [3] Abbe, E., Bandeira, A. S. & Hall, G. (2016) Exact recovery in the stochastic block model. *IEEE Trans Info Theory* **62**, 471–487.
- [4] Chen, X. & Yang, Y. (2020) Cutoff for exact recovery of gaussian mixture models, arXiv: 2001.01194.
- [5] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *J Royal Stat. Soci. Ser. B* **39**(1), 1–38.
- [6] Dyer, M. E. & Frieze, A. M. (1989) The solution of some random np-hard problems in polynomial expected time. *J Algorithm* **10**, 451–489.
- [7] Fraley, C. & Raftery, A. (2002) Model-based clustering, discriminant analysis, and density estimation. *J. American Stat. Assoc.* **97**, 611–631.
- [8] Giraud, C. & Verzelen, N. (2019) Partial recovery bounds for clustering with the relaxed k means, arXiv: 1807.07547.
- [9] Holland, P. W., Laskey, K. B. & Leinhardt, S. (1983) Stochastic blockmodels: First steps. *Soc. Networks* **5**, 109–137.
- [10] Kannan, R., Salmasian, H. & Vempala, S. (2008) The spectral method for general mixture models. *SIAM J. Comput.* **38**, 1141–1156.
- [11] Kim, C., Bandeira, A. & Goemans, M. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In: *12th International Conference on Sampling Theory and Applications*, pp. 124–128.
- [12] Löffler, M., Zhang, A. Y. & Zhou, H. (2020). Optimality of spectral clustering in the gaussian mixture model.
- [13] Li, Z. (2020) Exact recovery of community detection in k-partite graph models, arXiv:1910.04320.
- [14] Massoulié, L. (2014). Community detection thresholds and the weak Ramanujan property. In: *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pp. 694–703.
- [15] McQueen, J. B. (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Statistics, University of California Press, Berkeley, pp. 281–297.
- [16] Mossel, El., Neeman, J. & Sly, A. (2018) A proof of the blockmodel threshold conjecture. *Combinatorica* **38**, 665–708.
- [17] Peng, J. & Wei, Y. (2015) Approximating k-means-type clustering via semidefinite programming. *SIAM J. Optim* **18**, 186–205.
- [18] Pollard, D. (1981) Strong consistency of k-means clustering. *Ann. Statist* **9**, 135–140.
- [19] Vempala, G. & Wang, S. (2004) A spectral algorithm for learning mixture models. *J. Comput. Syst. Sci* **68**, 841–860.

Appendix

A. Maximum of Gaussian Random Variables

Lemma A.1. Let G_1, \dots, G_N be Gaussian random variables with mean 0. Let $\varepsilon \in (0, 1)$. Then

$$\Pr \left(\max_{i=1, \dots, N} G_i > (1 + \varepsilon) \sqrt{2 \max_{i \in [N]} \text{Var}(G_i) \log N} \right) \leq N^{-\varepsilon}$$

and moreover, if G_i 's are independent, and ε, N satisfy

$$\frac{N^{\varepsilon-\varepsilon^2}(1-\varepsilon)\sqrt{2\log N}}{\sqrt{2\pi}(1+2(1-\varepsilon)^2\log N)} > 1 \tag{A1}$$

Then

$$\Pr\left(\max_{i=1,\dots,N} G_i < (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right) \leq \exp(-N^\varepsilon)$$

Proof. It is known that for a Gaussian random variable G_i and $x > 0$,

$$\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}(1+x^2)} \leq \Pr\left(\frac{G_i}{\sqrt{\text{Var}(G_i)}} > x\right) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} \tag{A2}$$

Let G_1, \dots, G_N be N Gaussian random variables. Then by (A2), we have

$$\begin{aligned} & \Pr\left(\max_{i\in[N]} G_i \geq (1+\varepsilon)\sqrt{2\max_{i\in[N]} \text{Var}(G_i)\log N}\right) \\ & \leq \sum_{i\in[N]} \Pr\left(\frac{G_i}{\sqrt{\text{Var}(G_i)}} \geq (1+\varepsilon)\sqrt{2\log N}\right) \\ & \leq \frac{Ne^{-(1+\varepsilon)^2\log N}}{2(1+\varepsilon)\sqrt{\pi\log N}} \\ & \leq N^{-\varepsilon} \end{aligned}$$

If we further assume that G_i 's are independent, then

$$\begin{aligned} & \Pr\left(\max_{i\in[N]} G_i < (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right) \\ & = \prod_{i\in[N]} \Pr\left(G_i < (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right) \\ & = \prod_{i\in[N]} \left[1 - \Pr\left(G_i > (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right)\right] \\ & \leq \prod_{i\in[N]} \left[1 - \Pr\left(\frac{G_i}{\sqrt{\text{Var}(G_i)}} > (1-\varepsilon)\sqrt{2\log N}\right)\right] \end{aligned}$$

By (A2), we obtain

$$\Pr\left(\max_{i\in[N]} G_i < (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right) \leq \left(1 - \frac{(1-\varepsilon)\sqrt{2\log N}}{\sqrt{2\pi}(1+2(1-\varepsilon)^2\log N)} \frac{1}{N^{(1-\varepsilon)^2}}\right)^N$$

When (A1) holds, we have

$$\Pr\left(\max_{i\in[N]} G_i < (1-\varepsilon)\sqrt{2\min_{j\in[N]} \text{Var}(G_j)\log N}\right) \leq \left(1 - \frac{1}{N^{1-\varepsilon}}\right)^{N^{1-\varepsilon}N^\varepsilon} \leq e^{-N^\varepsilon}$$

Then the lemma follows. □