Sharp Impossibility Results for Hypergraph Testing

Jiashun Jin
Carnegie Mellon University
jiashun@stat.cmu.edu

Zheng Tracy Ke
Harvard University
zke@fas.harvard.edu

Jiajun Liang
Purdue University
liangjj@purdue.edu

Abstract

In a broad Degree-Corrected Mixed-Membership (DCMM) setting, we test whether a non-uniform hypergraph has only one community or has multiple communities. Since both the null and alternative hypotheses have many unknown parameters, the challenge is, given an alternative, how to identify the null that is hardest to separate from the alternative. We approach this by proposing a degree matching strategy where the main idea is leveraging the theory for tensor scaling to create a least favorable pair of hypotheses. We present a result on standard minimax lower bound theory and a result on Region of Impossibility (which is more informative than the minimax lower bound). We show that our lower bounds are tight by introducing a new test that attains the lower bound up to a logarithmic factor. We also discuss the case where the hypergraphs may have mixed-memberships.

1 Introduction

The hypergraph is a useful representation of social relationships beyond pairwise interactions [5, 11]. For example, the co-authorship hypergraph is often used to analyze the co-authorship topology of authors, and it provides more information than a co-authorship graph (where an \( m \)-author paper is treated as an \( m \)-clique). The community detection on a hypergraph [10] is a problem of great interest (communities in a hypergraph are clusters of nodes that have more hyperedges within than across). It has many applications in social network analysis [15] and machine learning [1, 17, 18, 23]. We are interested in the problem of global testing, where we test whether the hypergraph has one community or multiple communities. It has applications in measuring co-authorship and citation diversity [12] and discovering non-obvious social groups and patterns [4]. It is also useful for understanding other problems such as community detection and change-point detection in dynamic hypergraphs.

For instructional purpose only, we start with the 3-uniform hypergraphs (i.e., each hyperedge consists of 3 nodes), but our results cover both higher-order hypergraphs and non-uniform hypergraphs. Let \( A \) be the adjacency tensor of a uniform and symmetric order-3 hypergraph with \( n \) nodes, where

\[
A_{i_1, i_2, i_3} = 1 \text{ if } i_1, i_2, i_3 \text{ share a hyperedge and equals 0 otherwise},
\]

for \( 1 \leq i_1, i_2, i_3 \) (distinct) \( \leq n \). Since the tensor is symmetric, \( A_{i_1, i_2, i_3} = A_{j_1, j_2, j_3} \) for two sets of indices \( \{i_1, i_2, i_3\} \) and \( \{j_1, j_2, j_3\} \) if one is a permutation of the other. We do not consider self hyperedges, hence, \( A_{i_1, i_2, i_3} = 0 \) whenever \( i_1, i_2, i_3 \) are non-distinct.

Real world hypergraphs have several noteworthy features. First, there may be severe degree heterogeneity (i.e., the degree of one node is many times higher than that of another). Second, the overall sparsity levels may vary significantly from one hypergraph to another. Last, a node may have mixed-memberships across multiple communities (i.e., nonzero weights on more than one community). To accommodate these features, we adopt the Degree-Corrected Mixed-Membership (tensor-DCMM) model. The notations below are frequently used in tensor analysis.

Definition 1.1. (matricization, slicing, and slice aggregation). Let \( A \) be a 3-symmetric tensor of \( n \)-dimension. First, we call the \( n \times n^2 \) matrix \( A \) the matricization of \( A \), defined by \( A_{i,j+n(k-1)} = A_{ijk} \).
1 \leq i, j, k \leq n. Second, for 1 \leq k \leq n, we use \( A_{i,k} \) to denote the \( n \times n \) matrix whose row-i-and-column-j is \( A_{ijk} \), 1 \leq i, j \leq n, and call it the k-th slice of \( A \). Last, for any \( n \times 1 \) vector \( x \), we use \( (Ax) \) to denote the matrix \( \sum_{k=1}^{n} x_k A_{ik} \), which is an aggregation of the slices.

Now, suppose the tensor has \( K \) perceivable communities. Let \( P \) be a symmetric (3-uniform) tensor of \( K \)-dimension that models the community structure, let \( \theta = (\theta_1, \theta_2, \ldots, \theta_3)^T \) be positive parameters that model the degree heterogeneity of nodes, and \( \pi_1, \pi_2, \ldots, \pi_n \) be \( K \)-dimensional membership vectors where \( \pi_i(k) \) is the weight node \( i \) puts on community \( k \), 1 \leq k \leq K. We assume for all 1 \leq i_1, i_2, i_3 \leq n that are three distinct indices, \( A_{i_1i_2i_3} \) are independent Bernoulli random variables with \( P(A_{i_1i_2i_3} = 1) = \theta_1 \theta_2 \theta_3 \pi_{i_1}^T \pi_{i_2} \pi_{i_3} \). Introduce a non-stochastic 3-uniform tensor \( Q \) of \( n \)-dimension where \( Q_{i_1i_2i_3} = \theta_1 \theta_2 \theta_3 \pi_{i_1}^T (P \pi_{i_2}) \pi_{i_3} \). Let \( \text{diag}(Q) \) be the tensor with the same size of \( Q \) where \( (\text{diag}(Q))_{i_1i_2i_3} = Q_{i_1i_2i_3} \) if \( i_1, i_2, i_3 \) are non-distinct, and \( (\text{diag}(Q))_{i_1i_2i_3} = 0 \) otherwise. It follows that

\[
\text{E}(A) = Q - \text{diag}(Q), \quad \text{where} \quad Q_{i_1i_2i_3} = \theta_1 \theta_2 \theta_3 \pi_{i_1}^T (P \pi_{i_2}) \pi_{i_3}.
\] (1.2)

For identifiability, let \( P \in \mathbb{R}^{K \times K^2} \) be the matricization of \( P \) (see Definition 1.1). We assume

\[
\text{rank}(P) = K, \quad \text{and} \quad \pi_{ii} = 1 \text{ for all } 1 \leq i \leq K.
\] (1.3)

**Definition 1.2.** We call (1.1)-(1.3) the tensor-DCMM model for 3-uniform hypergraphs. We call \( Q \) and \( P \) the Bernoulli probability tensor and community structure tensor for DCMM, respectively.

Later in Section 3, we introduce the non-uniform tensor-DCMM as a more sophisticated model. In tensor-DCMM, if we require all \( \pi_i \) to be degenerate (i.e., one entry is 1, all other entries are 0), then tensor-DCMM reduces to the Degree-Corrected Block Model (tensor-DCBM) [15]. If we further require \( \theta_1 = \ldots = \theta_n \) (but the second condition in (1.3) can be removed), then tensor-DCBM further reduces to the Stochastic Block Model (tensor-SBM) [9]. For simplicity, we may drop “tensor” in these terms if there is no confusion. The global testing problem above is then to test

\[
H_0 : K = 1 \quad \text{vs.} \quad H_1 : K > 1.
\] (1.4)

Our primary goals are (a) to find a sharp information lower bound for 3-uniform DCMM, and especially, to fully characterize the lower bound by a simple quantity to be discovered, and (b) extend the results to more sophisticated non-uniform hypergraph (see Section 3). A good understanding of the problem greatly helps us understand the fundamental limits of many other problems (e.g., community detection, determining the number of communities \( K \), dynamic hypergraphs). For example, for community detection (e.g., [10]), we either assume \( K \) as known or estimate it first. Note that in parameter regions where we cannot tell whether \( K = 1 \) or \( K > 1 \), we cannot estimate \( K \) consistently, so we cannot have consistent community detection either. Therefore, a lower bound for global testing is always a valid lower bound for estimating \( K \) and for community detection.

To facilitate the lower bound study, we frequently adopt a Random Mixed-Membership (RMM) model. Introduce a subset \( V_0 = \{ x \in \mathbb{R}^K : x_k \geq 0, \sum_{k=1}^{K} x_k = 1 \} \), and let \( F \) be a \( K \)-variate distribution with support contained in \( V_0 \). We assume

\[
\pi_i \overset{iid}{\sim} F; \quad \text{(let } h = \mathbb{E}_F[\pi_i]).
\] (1.5)

Moreover, let \( V_0^* = \{ e_1, e_2, \ldots, e_K \} \subset V_0 \), where \( e_k \) is the \( k \)-th basis vector of \( \mathbb{R}^K \). Similarly, RMM-DCMM reduces to RMM-DCBM if we require \( \text{supp}(F) = V_0^* \), and reduces to RMM-SBM if we further require \( \theta_1 = \theta_2 = \ldots = \theta_n \) (but the second condition in (1.3) can be removed). Let \( \theta = (\theta_1, \theta_2, \ldots, \theta_n)^T \). We allow \( (\theta, P, h, F) \) to vary with \( n \) to cover a variety of settings where we allow for severe degree heterogeneity, mixed-memberships, flexible sparsity levels, and weak signals.

**Example 1 (2-parameter SBM [2, 22, 19]).** This model is a special case of DCMM, where \( \theta_1 = \ldots = \theta_n = \alpha_n \) (no degree heterogeneity), all \( \pi_i \) are degenerate (no mixed membership), and \( P_{ijk} = 1 \) if \( i = j = k \) and \( P_{ijk} = \rho_n \) otherwise, for two parameters \( (\alpha_n, \rho_n) \). Also, Lin et. al. [20] studied a 3-parameter SBM, which is the same as above except that they assume a different form of \( P \), where \( P_{ijk} \) equals to 1, \( \rho_n \), or \( \tau_n \) if \( i, j, k \) take 1, 2, or 3 distinct values, respectively, for three parameters \( (\alpha_n, \rho_n, \tau_n) \). Compared to DCMM, these models are much narrower: they do not accommodate severe degree heterogeneity or mixed-memberships, and \( P \) is parametrized by 2 or 3 distinct values.
We propose a new test statistic and show that the sum of Type I error and Type II error

with existing results on minimax lower bounds, these results are more informative and theoretically

we can pair it with a null such that the

θ, we have many parameters

there exists an alternative model

The standard minimax theory only claims that

max

Region of Impossibility for the hypothesis testing

for a logarithmic factor here, our lower bounds are tight.

(1.4)

max

m

ℓ

Liang et al. [19] and Kim et al. [16] for related settings) with a similar model. Lin et al. [20] and Chien et al. [7] used the 3-parameter SBM in Example 1 for study of the lower bounds for community detection. While these papers are very interesting, their lower bounds are characterized by only 2 or 3 parameters (i.e., αₙ, ρₙ, τₙ) assumed in their models. For the much broader tensor-DCMM model considered here, we have many parameters θ, P, h, F (θ is an n-vector and P is a tensor), and how to extend existing results to the tensor-DCMM setting here is a challenging problem.

Our contributions. The main challenge in lower bound study is that, since each DCMM model (no matter what K is) has a large number of unknown parameters, so given an alternative hypothesis (K > 1), it is hard to identify the null hypothesis (K = 1) that is most difficult to distinguish from the alternative hypothesis. As our main contribution, we approach this by proposing a degree matching strategy, where for any given DCMM model with K > 1, we pair it with a DCMM model with K = 1 in a way so that for each node, the expected degree under the null matches with that under the alternative. This way, it is hard to separate the two hypotheses by a naive degree-based statistic. We show (a) the degree matching is always possible by using a tensor scaling technique [3, 8], and (b) the pair of hypotheses we construct this way lead to sharp results on lower bounds. See Section 2.

We have the following results. Consider the 3-hyperparameter hypergraph first. We present the current standard minimax theory. Define a class of RMM-DCMM models with K > 1 and µ²∥θ∥²∥θ∥₁ → 0 (µ₂ is the second singular value of P). In this class, we can find an RMM-DCMM model (the alternative), and pair it with a DCMM model with K = 1 (the null), such that the χ²-divergence between the pair converges to 0 as n → ∞. Therefore, in this class, there exists an alternative model that is asymptotically inseparable from the null.

The standard minimax theory only claims that there exists an alternative model within a specified class that is inseparable from the null. It is desirable to show a much stronger result where for any alternative in the class, we can pair it with a null so that the χ²-divergence of the pair goes to 0 as n → ∞. In detail, we show that in the parameter space (θ, P, h, F) of RMM-DCBM, there is a Region of Impossibility defined by µ²∥θ∥²∥θ∥₁ → 0; for any alternative in Region of Impossibility, we can pair it with a null such that the χ²-divergence between the pair goes to 0 as n → ∞. Compared with existing results on minimax lower bounds, these results are more informative and theoretically more satisfactory. The proof is also different from the proof of minimax lower bounds: we have used the tensor scaling theory [3, 8] and the “degree matching strategy” aforementioned. We also extend such results to the broader RMM-DCMM case (Section 2.3) and discuss some major differences on the Region of Impossibility between DCBM and the more restrictive SBM (Section 2.4).

Next, we generalize the results to higher-order and non-uniform hypergraphs. Fix M ≥ 2 and consider a non-uniform hypergraph (e.g., see [10]) that consists of m-uniform hypergraphs for all m = 2, . . . , M, each following a DCMM model with individual (g(m), P(m)) but the common π₁, . . . , πₙ. Let ℓₘ = ||g(m)||₁∥g(m)||₂∥(µ₂(m))² (to be defined in Section 3). We show that (a) for the m-uniform hypergraph case, the Region of Impossibility for the hypothesis testing (1.4) is fully characterized by the condition of ℓₘ → 0, and (b) for the non-uniform hypergraph case, the Region of Impossibility for the hypothesis testing (1.4) is fully characterized by the condition of max₂≤m≤M {ℓₘ} → 0.

Last, we show that our lower bounds are tight. Consider the non-uniform hypergraph above. We propose a new test statistic and show that the sum of Type I error and Type II error → 0 if max₂≤m≤M {ℓₘ} ≥ log(n)¹⁺δ for a constant δ > 0 (taking δ = 0.1 will work). Therefore, except for a logarithmic factor here, our lower bounds are tight.

In summary, existing results on lower bounds are largely focused on more restrictive settings (e.g., uniform hypergraphs without degree heterogeneity or mixed membership). We provide sharp lower bounds for a much broader setting. Our study is highly non-trivial because we need (i) a novel degree
matching strategy to construct least favorable hypothesis pairs, (ii) to identify a simple quantity that is able to fully characterize the lower bounds, (iii) delicate analysis of the $\chi^2$-divergence between the null and alternative, (iv) a carefully designed test that leads to tight upper bounds.

2 Sharp Lower Bounds for 3-Uniform Hypergraphs

For notational simplicity, we focus on 3-uniform hypergraphs in this section. The study of higher-order and non-uniform hypergraphs is deferred to Section 3. Consider a (3-uniform) DCMM model. Recall that $h = E_P[\pi_i]$ and that $P \in \mathbb{R}^{K,K^2}$ is the matricization of the tensor $P$. In this paper, we use $C > 0$ as a generic constant which may vary from occurrence to occurrence. We assume

$$\|P\| \leq C, \quad \theta_{\max} \equiv \max\{\theta_1, \ldots, \theta_n\} \leq C, \quad \max_{1 \leq k \leq K} \{h_k\} \leq C, \quad \min_{1 \leq k \leq K} \{h_k\}, \quad \|\theta\|^2 \|\theta_1\| \to \infty. \quad (2.6)$$

The first condition is mild, because the model identifiability already requires that all diagonal entries of $P$ are 1. The second one is also mild (note that while the largest possible value of $\theta_{\max}$ is $O(1)$, $\theta_{\max}$ is allowed to tend to 0 relatively fast). The third one assumes that the community memberships are balanced, which is also mild. For the last condition, we will see soon that if $\|\theta\|^2 \|\theta_1\| \to 0$, then the signal is so weak that successful global testing is impossible, so this condition is also mild.

2.1 Standard minimax lower bounds (RMM-DCMM)

We start with the least favorable configuration. The goal is to find a pair of models (a null model with $K = 1$, and an alternative model with $K > 1$) which are hard to distinguish from each other. We use the following degree-matching technique: we choose the pair in a way so that for each node, the expected degree under the null matches that under the alternative, approximately. The idea is, if the degrees are not matching, then we may separate the two hypotheses by a simple degree-based statistic, so we should not expect the $\chi^2$-divergence between the pair to be small.

In detail, consider a pair of models, a DCMM model with $K = 1$ and an RMM-DCMM model with $K > 1$, where for all $1 \leq i_1, i_2, i_3 \leq n$, the Bernoulli probability tensors $Q$ and $Q^*$ satisfy

$$Q_{i_1i_2i_3} = \theta_{i_1} \theta_{i_2} \theta_{i_3}, \quad Q^*_{i_1i_2i_3} = \theta^*_{i_1} \theta^*_{i_2} \theta^*_{i_3} \cdot \pi_{i_1}(P \pi_{i_2}) \pi_{i_3}; \quad (2.7)$$

here we recall that $(P \pi_{i_1})$ is a $K \times K$ matrix (see Definition 1.1). We call the two models the null and the alternative, respectively. In (2.7), the community structure tensor $P$ is as in (1.2), and $\pi_i$ and $h = E_P[\pi_i]$ are as in (1.5). Recall that $\theta = (\theta_1, \theta_2, \ldots, \theta_n)^\top$. Similarly, let $\theta^* = (\theta^*_1, \theta^*_2, \ldots, \theta^*_n)^\top$.

Fix $1 \leq i_1 \leq n$. By definitions and elementary statistics, the leading term of the expected degree of node $i_1$, conditional on its own membership $\pi_{i_1}$, under the null and alternative are

$$\theta_{i_1} \|\theta\|^2, \quad \text{and} \quad \theta_{i_1}^* \|\theta^*\|^2, \quad \text{respectively, where } a = (P h) h \in \mathbb{R}^K. \quad (2.8)$$

For least favorable construction, we choose $(P, h)$ in a way so that

$$a = (P h) h = \frac{c_0^2}{1} \mathbf{1}_K, \quad \text{for a scalar } c_0 > 0. \quad (2.9)$$

For broadness, we allow $c_0$ to depend on $n$. There are many $(P, h)$ that satisfy (2.9). For example, in the 2-parameter SBM model as in Example 1 with $h = (1/K, \ldots, 1/K)^\top$ and $\alpha_k = (1/K^2)[1 + (K^2 - 1)\rho_n]$ for $1 \leq k \leq K$, (2.9) is satisfied with $c_0^2 = (1/K^2)[1 + (K^2 - 1)\rho_n]$. Moreover, we choose $\theta^*_i$ in the alternative model such that

$$\theta^*_i = (1/c_0) \theta_i, \quad 1 \leq i \leq n. \quad (2.10)$$

Now, by (2.8)-(2.10), for all $1 \leq i_1 \leq n$, $\theta_{i_1}^* \|\theta^*\|^2 \cdot \pi_{i_1} a = \theta_{i_1}^* \|\theta^*\|^2 \cdot \pi_{i_1} c_0 = \theta_{i_1} \|\theta\|^2$. Therefore, for each node, the expected degree under the alternative matches that under the null (at least in the leading term), making it hard to separate the null and alternative by any naive degree-based statistics. Only when such a degree-matching holds, we can hope two models are asymptotically indistinguishable. This is the key for our least favorable configuration. Recall that $P \in \mathbb{R}^{K,K^2}$ is the matricization of the community tensor $P$ and $\mu_k$ is the $k$-th largest singular value of $P$.

**Theorem 2.1 (Least favorable configuration).** Fix $K > 1$ and consider a pair of models, a null and an alternative with $K$ communities, given in (2.7), where (2.9)-(2.10) hold. Assume (2.6) holds and
We aim to select a matrix \( \minimax_{\text{lower bound}} \). In our setting, the alternative model define a longer works, and we must find a new approach to constructing the model pair. \( \theta \)

It is crucial to pick an alternative model where the other approximately. Recall that in Section 2.2, in order to have a desired degree matching, it is \( \theta \)

Consider again a pair of models, a DCBM null model and an RMM-DCBM model with inseparable. A much more satisfactory result is to show that, for any alternative \( \theta \)

Remark (How degree matching affects the \( \chi^2 \)-divergence). Intuitively speaking, the \( \chi^2 \)-divergence has many terms, each being the sum (or a function) of terms in a Taylor expansion. We can roughly call these terms the first-order term, second-order term, and so on. When the expected node degrees are not matched between the null and the alternative, the first-order term dominates, and correspondingly, a degree-based \( \chi^2 \)-test may have power; see Section 2.4. When the expected degrees are matched, the first-order term vanishes as we have hoped, and the degree-based \( \chi^2 \)-test loses power. As a result, the second-order term in the \( \chi^2 \)-divergence now dominates and gives the sharp lower bound.

The above least favorable configuration gives rise to the standard minimax theorem. Fix \( K \geq 1 \) and consider a DCMM model (where \( \pi_i \) are non-random). Introduce a vector \( g \in \mathbb{R}^K \) by \( g_k = (1/||\theta||_1) \sum_{i=1}^{n} \theta_i \pi_i(k), 1 \leq k \leq K \). For a constant \( 0 < c_0 < 1 \), and two positive sequences \( \{\alpha_n\}_{n=1}^{\infty} \) and \( \{\beta_n\}_{n=1}^{\infty} \), we define a class of DCMM models by

\[
M_n(K, c_0, \alpha_n, \beta_n) = \left\{ (\theta, \Pi, P): \left\| P \right\| \leq c_0, \left\| \theta \right\|_{\max} \leq c_0, \max_{1 \leq k \leq K} g_k (1/\beta_n), \left\| \theta \right\|_1^{2} \left\| \theta \right\|^2 \leq \alpha_n \right\}
\]

Here, we assume \( \max_{1 \leq k \leq K} g_k (1/\beta_n) \rightarrow C \min_{1 \leq k \leq K} g_k \), so the tensor is balanced. This is similar to the third condition in (2.6), except that \( \pi_i \)'s are random there. For the null case, \( K = \pi = \pi_1 = 1 \), and the above defines a class of \( \theta \), which we write for short by \( M_n(1, c_0, \alpha_n, \beta_n) = M_n^*(\beta_n) \). The following theorem is proved in the supplement.

Theorem 2.2 (Minimax lower bound). Fix \( K \geq 2 \), a constant \( c_0 > 0 \), and any sequences \( \{\alpha_n\}_{n=1}^{\infty} \) and \( \{\beta_n\}_{n=1}^{\infty} \) where \( \alpha_n = o(1) \) and \( \beta_n = o(1) \). As \( n \rightarrow \infty \), \( \inf_{\psi} \left\{ \sup_{\theta \in M_n^*(\beta_n)} P(\psi = 1) + \sup_{\theta \in M_n^*(\beta_n)} P(\psi = 0) \right\} = 1 - o(1) \), where the infimum is over all possible tests \( \psi \).

2.2 Region of Impulsibility for RMM-DCBM

The standard minimax theorem in Section 2.1 only says that in the class of all alternative models with \( \left\| \theta \right\|^2 \left\| \theta \right\|_1^{2} \rightarrow 0 \), there exists one where we can pair it with a null so the pair are asymptotically inseparable. A much more satisfactory result is to show that, for any alternative in the same class, we can pair it with a null such that the pair are asymptotically inseparable. In this section, we prove this for the RMM-DCBM case (mixed-memberships are not allowed). The discussion for the RMM-DCMM (mixed-memberships allowed) is in Section 2.3 and the supplement.

Consider again a pair of models, a DCBM null model and an RMM-DCBM model with \( K > 1 \), where the Bernoulli probability tensors are \( Q \) and \( Q^* \), respectively. We assume for all \( 1 \leq i_1, i_2, i_3 \leq n \),

\[
Q_{i_1 i_2 i_3} = \theta_{i_1} \theta_{i_2} \theta_{i_3}, \quad Q^*_{i_1 i_2 i_3} = \theta^*_{i_1} \theta^*_{i_2} \theta^*_{i_3} \pi_{i_1} \pi_{i_2} \pi_{i_3} \quad (P \pi i_3) \pi_{i_2} i_3.
\]

Here, the community structure tensor \( P \) is as in (1.2), \( \pi_i \) and \( h = E_F[\pi_i] \) are as in (1.5), and \( \supp(F) = \{e_1, \ldots, e_K\} \). Similarly, the goal is degree-matching: for any \( (\theta, P, h, F) \), we construct \( \theta^* \) in a way so that for each node, the expected degrees under the null and the alternative match with each other approximately. Recall that in Section 2.2, in order to have a desired degree matching, it is crucial to pick an alternative model where the \( (P, h) \) satisfies \( a = (P h) h = c_0^0(1)_K \) for some scalar \( c_0 > 0 \); once this holds, we have the desired degree matching by taking \( \theta^* = (1/c_0) \theta \). Unfortunately, for general \( (P, h) \), we don’t have \( a = (P h) h \propto \chi^2_K \), so we should not expect to have the desired degree matching by taking \( \theta^* \propto \theta \). In short, for our purpose here, the approach in Section 2.1 no longer works, and we must find a new approach to constructing the model pair.

Our proposal is as follows. For any diagonal matrix \( D = diag(d_1, d_2, \ldots, d_K) \) with \( d_k > 0, 1 \leq k \leq K \), define a \( K \)-dimensional vector \( a^D \) by \( (P^D) \) is a \( 3 \)-tensor in dimension \( K \)

\[
a^D = (P^D h) h \quad \text{with} \quad P^D_{k_1 k_2 k_3} = d_{k_1} d_{k_2} d_{k_3} P_{k_1 k_2 k_3}, \quad 1 \leq k_1, k_2, k_3 \leq K.
\]

We aim to select a matrix \( D \) such that \( a^D = 1_K \). The next lemma states that such \( D \) always exists and is unique. It leverages classic results on tensor scaling (e.g., [3, 8]) and is proved in the supplement.

---

1 The \( \chi^2 \)-divergence between two models, \( f_0(A) \) and \( f_k(A) \), is defined as \( \int[\{f_0(A) - f_k(A)\}^2/f_0(A)]dA \). In our setting, the alternative model \( f_k(A) \) alone involves an integral over the distribution of \( \pi_i \) in (1.5).
Lemma 2.1. Fix $K > 1$ and let $P$, $h$, $D$, and $a^D$ be as above. Suppose $\min\{h_1, h_2, \ldots, h_K\} \geq C$. There exists a unique diagonal matrix $D = \text{diag}(d_1, d_2, \ldots, d_K)$ such that $a^D = 1_K$.

Now, given $(\theta, P, h, F)$ and the two models in (2.11), let $D = \text{diag}(d_1, d_2, \ldots, d_K)$ be as in Lemma 2.1. Moreover, in (2.11), we choose $\theta^*_i$ as follows:

$$\theta^*_i = d_k \theta_i, \quad \text{if node } i \text{ belongs to community } k. \quad (2.12)$$

Combining it with (2.11), we have $Q_{11}^i = \theta^*_i \theta^*_i \pi_{i}^i, \pi_{i}^i (P \pi_i) \pi_i = \theta_i' \theta_i \pi_{i}^i (P^D \pi_i) \pi_i$.

By similar calculations as in (2.8), for $1 \leq i \leq n$, in the null and the alternative, the leading terms of the expected degrees of node $i$ are

$$\theta_i \|\theta\|_1 \quad \text{and} \quad \theta_i \|\pi_i^D\| \|\theta\|^2, \quad \text{respectively, where } a^D = (P^D h) h.$$ 

By Lemma 2.1, $a^D = 1_K$. Hence, for each node, the expected degrees match under the null and alternative, so it is hard to separate two models by a naive degree-based statistic. Recall that $P$ is the matricization of $P$ and $\mu_k$ is the $k$-th singular value of $P$. Theorem 2.3 is proved in the supplement.

Theorem 2.3 (Impossibility for RMM-DCBM). Fix $K > 1$. For any given $(\theta, P, h, F)$, consider a pair of models, a null and an alternative with $K$ communities, as in (2.11), where $\theta^*_i$ are given by (2.12) with the matrix $D$ as in Lemma 2.1. Suppose (2.6) holds and $||\theta||_1 ||\theta||^2 \mu^2 / \log n \to \infty$. As $n \to \infty$, the $\chi^2$-divergence between the pair tends to 0. Therefore, the two models are asymptotically indistinguishable: for any test, the sum of Type I and Type II errors is no smaller than $1 + o(1)$.

Region of Possibility, Region of Impossibility, and tightness. In the parameter space $(\theta, P, h, F)$ for DCBM, we call the region prescribed by $||\theta||_1 ||\theta||^2 \mu^2 \to 0$ the Region of Possibility: by Theorem 2.3, for any model in this region, we can pair it with a null so that the pair are asymptotically inseparable. We call the region prescribed by $||\theta||_1 ||\theta||^2 \mu^2 / \log n \to \infty$ the Region of Possibility: for any alternative model in this region, there is a method that can separate it from any null with asymptotically full power (this follows from Theorem 3.2 as a special case). Comparing Region of Possibility and Region of Possibility, except for a $\log(n)$ term, our lower bounds are tight.

2.3 Region of Possibility for RMM-DCM

The discussion for RMM-DCM is similar, so for reasons of space, we leave it to Section A of the supplement. In that section, we present a similar theorem for RMM-DCM, where the hypergraphs may have mixed-memberships. The proofs are largely similar, where again the key is to construct a pair of null and alternative using the degree matching strategy. Since the model is more complicated than RMM-DCBM, we need an extra (but mild) condition. See details therein.

2.4 Major differences on Region of Possibility for the more restrictive RMM-SBM

For an alternative RMM-DCBM, we can always pair it with a null using tensor scaling technique: for each node, the expected degrees under the null and the alternative match with each other. For the more restrictive RMM-SBM (where $\theta_1 = \ldots = \theta_h$), such a degree matching is not always possible: A null SBM has only 1 parameter, so we have insufficient flexibility in choosing the null for degree matching. A consequence is that a naive degree-based test may have non-trivial power for SBM.

Consider a pair of models, where the Bernoulli probability tensors $Q$ and $Q^*$ under two hypotheses are such that, for $1 \leq i_1, i_2, i_3 \leq n$,

$$Q_{i_1 i_2 i_3} = \alpha_{i_1}, \quad Q_{i_1 i_2 i_3}^* = \pi_{i_1}^i (P \pi_{i_3}) \pi_{i_3}.$$  

(2.13)

Same as before, $\pi_i$ are iid generated from a distribution $F$ on $\{e_1, e_2, \ldots, e_K\}$ with $h = \text{E}_F[\pi_i]$. Let $\tilde{\alpha}_n = (n(n-1)(n-2))^{-1/2}(A_{1i}^n) i_1, i_2, i_3 \leq n, \eta = (1/2)(A_{1i}^n) i_1, i_2, i_3 \leq n,$ and $\bar{\eta}$ be the mean of $\eta_1, \eta_2, \ldots, \eta_K$. Consider the centered-$\chi^2$-statistic

$$\psi_n = (2n)^{-1/2} \sum_{1 \leq i \leq n} \left( (\eta_i - \bar{\eta})^2 / \left( \left(1 - \tilde{\alpha}_n \right) - 1 \right) \right).$$

Lemma 2.2. Consider the global testing problem (1.4) under the SBM model (2.13) for $H_0$ and $H_1$, respectively. Let $\tilde{\alpha}_n = \text{E}[\tilde{\alpha}_n], h = (1/n) \sum_{i=1}^n \pi_i$ and $\Sigma = (1/2) \sum_{i=1}^n (\pi_i - h)(\pi_i - h)^\top$. Let $\lambda_{K-1}(\Sigma)$ be the $(K-1)$-th largest eigenvalue (in magnitude) of $\Sigma$. Assume $\alpha_n \leq c_0, \max_i \alpha_n, \Sigma \leq c_0, n^2 \alpha_n \rightarrow \infty$, and $n^2 \alpha_n \rightarrow \infty$. Also, assume $\min\{h_1, h_2, \ldots, h_K\} \geq C$ and $\lambda_{K-1}(\Sigma) \geq C$. Write $\delta_n = \|\tilde{\alpha}_n - (I_K - H_K)(P^D h) h\|$, where $H_K = (1/K) 1_K 1_K^\top$ and $I_K$ is the identity matrix of the same size. As $n \to \infty, \psi_n \to N(0, 1)$ if $H_0$ holds, and $\psi_n \to \infty$ if $H_1$ holds and $n^{3/2} \tilde{\alpha}_n^{-1/2} \delta_n^2 \to \infty$. 


By Lemma 2.2, the power of the $\chi^2$-test hinges on $\delta_n$

$$\delta_n = \|\tilde{\alpha}_n^{-1}(I_K - H_K)(\mathcal{P}h)h\|.$$ 

Note that $\delta_n \neq 0$ if and only if $(\mathcal{P}h)h \propto 1_K$. We call the cases of $(\mathcal{P}h)h \propto 1_K$ and $(\mathcal{P}h)h \npropto 1_K$ the symmetric case and the asymmetric case, respectively. For symmetric SBM, $\delta_n = 0$, and we do not expect the $\chi^2$-test to have power. However, for asymmetric SBM, the $\chi^2$-test may have non-trivial power, implying a potential shift of the lower bound.

**Example 2.** Consider an SBM setting where we either have $K = 1$ (null) or $K = 2$ (alternative). Also, when $K = 2$, we assume $h = (a, 1 - a)'$ for some $0 < a < 1$, $\mathcal{P}_{ijk}$ is equal to $\rho_0$ if $i = j = k$ and $\rho_1$ otherwise. Suppose $n^2\rho_0 \to 0$ and $\rho_1/\rho_0 \to 1$. This can be viewed as a special DCBM with $\theta_i \equiv \rho_0^{1/3}$ and off-diagonals of $\mathcal{P}$ being $\rho_1/\rho_0$. In this case, we have $\|\theta_1\|/\|\theta_2\| \approx n^2\rho_0$, $\tau_n \equiv \|\theta_1\|/\|\theta_2\| \approx n^2\rho_0\rho_1 - \rho_0)^2$, and $n^{3/2}\tilde{\alpha}_n\delta_n^2 \approx (n\rho_0)^{-1/2}(2a - 1)^2\tau_n$. In the symmetric case where $a = 1/2$, the $\chi^2$-test is powerless, and the Region of Impossibility is given by $\tau_n \to 0$ (same as in the DCBM case). In the asymmetric case where $|a - 1/2| \geq c_0$, by direct calculations, we have that even when $\tau_n \to 0$, we may have $n^{3/2}\tilde{\alpha}_n\delta_n^2 \to \infty$ (so $\chi^2$-test has asymptotically full power). Here, the interesting range of $\rho_0$ is $(n^2, 1)$, so we may have $(n\rho_0)^{-1/2} \to \infty$. Therefore, for the asymmetric case, the Region of Impossibility is different from that of DCBM.

In most lower bound results for SBM [2, 7, 20, 22], they focused on the symmetric case $(\mathcal{P}h)h \propto 1_K$. Our lower bound restricted to symmetric SBM agrees with those in the literature. The asymmetric case $(\mathcal{P}h)h \npropto 1_K$ is less studied, except for [14, 21] which focused on the network setting ($m = 2$). We discover: (i) the Region of Impossibility for symmetric SBM is similar to that of DCBM (see Example 2), and (ii) the Region of Impossibility for asymmetric SBM is quite different from that of DCBM. This is because DCBM is much broader than SBM, where the problem of global testing is much harder, and a naïve degree-based test statistic may lose power.

## 3 Sharp lower bound for non-uniform hypergraphs

Section 2 discusses lower bounds for uniform 3-hypergraph. We now first extend the results to more general non-uniform hypergraphs, and then present a tight upper bound. Note that our results include those for $m$-uniform hypergraphs as special cases (see the paragraph behind Theorem 3.1). The notation below is useful:

**Definition 3.1.** Given an order-$m$ tensor $\mathcal{M}$ in dimension $K$ and vectors $b_1, b_2, \ldots, b_m \in \mathbb{R}^K$, let $[\mathcal{M}; b_1, \ldots, b_m]$ denote the summation $\sum_{k_1, k_2, \ldots, k_m \leq K} [\mathcal{M}_{k_1k_2\ldots k_m}b_1(k_1)b_2(k_2)\ldots b_m(k_m)]$.

Fix $M \geq 2$. Consider a general non-uniform hypergraph that consists of $m$-uniform hypergraphs for all $2 \leq m \leq M$. Fixing $2 \leq m \leq M$, let $A^{(m)}$ be the adjacency tensor of the order-$m$ hypergraph (i.e., $A^{(m)}_{i_1i_2\ldots i_m} = 1$ if $\{i_1, i_2, \ldots, i_m\}$ is a hyper-edge and 0 otherwise). As before, we model $A^{(m)}$ with the tensor-DCMM model. Let $\mathcal{P}^{(m)}$ be a symmetric order-$m$ tensor in dimension $K$, and let $\theta^{(m)} = (\theta^{(m)}_1, \theta^{(m)}_2, \ldots, \theta^{(m)}_m)'$ be a positive vector of degree parameters. Let $\pi_1, \pi_2, \ldots, \pi_n$ be the $K$-dimensional membership vectors (which do not depend on $m$). We assume $\{A^{(m)}_{i_1i_2\ldots i_m}\}_{1 \leq i_1 < i_2 < \ldots < i_m \leq n}$ are independent Bernoulli variables, where the Bernoulli probabilities are specified by the tensor $Q^{(m)}$, given by

$$Q^{(m)}_{i_1i_2\ldots i_m} = \theta^{(m)}_{i_1} \cdots \theta^{(m)}_{i_m} \times [\mathcal{P}^{(m)}; \pi_{i_1}, \ldots, \pi_{i_m}], \quad 1 \leq i_1, i_2, \ldots, i_m \leq n. \quad (3.14)$$

Similar to (1.2), we have $\mathbb{E}[A^{(m)}] = Q^{(m)} - \text{diag}(Q^{(m)})$. This extends the tensor-DCMM model to $m$-uniform hypergraphs for a general $m$. Finally, we denote the non-uniform hypergraph by $A[\mathcal{M}] = \{A^{(2)}, A^{(3)}, \ldots, A^{(M)}\}$.

**Definition 3.2.** We say that $A[\mathcal{M}] = \{A^{(2)}, \ldots, A^{(M)}\}$ follows a (general non-uniform) tensor-DCMM model if $\{A^{(m)}\}_{2 \leq m \leq M}$ are independent of each other and each $A^{(m)}$ follows an $m$-uniform tensor-DCMM model as in (3.14), where $\pi_1, \pi_2, \ldots, \pi_n$ are shared by all $2 \leq m \leq M$.

A similar model is introduced in [10], but is more restrictive for it assumes $\theta_{i_1}^{(m)} = \theta_{i_2}^{(m)} = \cdots = \theta_{i_m}^{(m)}$ for each $2 \leq m \leq M$. Note also the focus of [10] is on community detection, while the focus here is.
on global testing. As argued before, since in parameter regions where we cannot tell whether \( K = 1 \) or \( K > 1 \), it is impossible to estimate \( K \) and community labels consistently. Therefore, our lower bound is also a valid lower bound for estimating \( K \) and for community detection.

We present the Region of Impossibility for testing \( H_0: K = 1 \) versus \( H_1: K > 1 \). Similarly as in Section 2.2, we focus on the special case of tensor-DCBM models (i.e., each \( \pi_i \) is degenerate); the study of tensor-DCMM is similar. Fix \( K > 1 \). Consider a DCBM null model with probability tensors \( Q[M] = \{Q^{(1)}, \ldots, Q^{(M)}\} \) and an RMM-DCBM model with probability tensors \( Q'[M] = \{Q'^{(1)}, \ldots, Q'^{(M)}\} \), where for every \( 2 \leq m \leq M \) and \( 1 \leq i, i_2, \ldots, i_m \leq n \),

\[
Q^{(m)}_{i_1, i_2, \ldots, i_m} = \theta^{(m)}_{i_1} \theta^{(m)}_{i_2} \cdots \theta^{(m)}_{i_m} \quad \text{and} \quad Q'^{(m)}_{i_1, i_2, \ldots, i_m} = \theta'^{(m)}_{i_1} \theta'^{(m)}_{i_2} \cdots \theta'^{(m)}_{i_m}.
\]

The support of \( F \) is in \( V_n^* = \{e_1, e_2, \ldots, e_K\} \). Let \( h = E_F[\pi_i] \) and suppose \( \min\{h_1, \ldots, h_K\} \geq C \). In the supplemental material, we provide a lemma analogous to Lemma 2.1: For each \( 2 \leq m \leq M \), there exists a unique diagonal matrix \( D(m) = \text{diag}(d^{(m)}_1, d^{(m)}_2, \ldots, d^{(m)}_K) \) such that

\[
\sum_{1 \leq i_2, \ldots, i_m \leq K} d^{(m)}_{i_1} \cdot \mathcal{P}^{(m)}_{i_1, i_2, \ldots, i_m} \cdot (d^{(m)}_{i_2} P_{i_2}) \cdots (d^{(m)}_{i_m} P_{i_m}) = 1, \quad \text{for every } 1 \leq i_1 \leq K.
\]

Its proof leverages the Sinkhorn theorems for higher-order tensors \([3, 8]\). We choose \( \theta^{(m)} \) in (3.16) by

\[
\theta^{(m)}_i = d^{(m)}_k \theta^{(m)}_k, \quad \text{if node } i \text{ belongs to community } k.
\]

This is analogous to the degree matching strategy in (2.12), and it is conducted for each \( m \) separately. Let \( P^{(m)} \in \mathbb{R}^{K \times K^{m-1}} \) be the matricization of \( \mathcal{P}^{(m)} \) and let \( \mu^{(m)}_2 \) be the second singular value of \( P^{(m)} \). For short, let

\[
\ell_m = \|\theta^{(m)}\|_1 \cdot \|\theta^{(m)}\|_2^2 \cdot \mu^{(m)}_2.
\]

Theorem 3.1 is proved in the supplement.

**Theorem 3.1 (Impossibility for non-uniform RMM-DCBM).** Fix \( K > 1 \) and \( M \geq 2 \). For any given \( (h, F) \) and \( \{(\theta^{(m)}, P^{(m)})\}_{2 \leq m \leq M} \), consider a pair of models, a null as in (3.15) and an alternative with \( K \) communities as in (3.16), where \( \{\theta^{(m)}_i\}_{1 \leq i \leq n, 2 \leq m \leq M} \) are as in (3.17). Suppose \( \|P^{(m)}\| \leq C, \max_{1 \leq i \leq n} \theta^{(m)}_i \leq C \), and \( \min_{1 \leq k \leq K} h_k \geq C \). If \( \max_{2 \leq m \leq M} \ell_m = o(1) \), then as \( n \to \infty \), the \( \chi^2 \)-divergence between the pair tends to 0.

**Remark (Comparison with [22]).** Yuan et al. [22] gave a nice impossibility result for the 2-parameter symmetric SBM. Their model is a special case of our model where \( \theta^{(m)}_i = \alpha^{(m)}_n \) and \( P^{(m)} \) has equal off-diagonal entries \( \rho^{(m)} \) (see Example 1 for \( m = 3 \)). In this case, \( |\mu^{(m)}_2| = |1 - \rho^{(m)}| \), and \( \ell_m \asymp n^{m-1} \alpha^{(m)}_n (1 - \rho^{(m)})^2 \). Hence, \( \ell_m \to 0 \) is equivalent to \( n^{m-1} \alpha^{(m)}_n (1 - \rho^{(m)})^2 \to 0 \), which matches with results in [22].

Below in Section 3.1, we show that the lower bounds are tight. Theorem 3.1 includes \( m \)-uniform hypergraphs as a special case (e.g., to apply the theorem to an \( m_0 \)-uniform hypergraph, we set \( \theta^{(m)} \) a zero vector for all \( m \neq m_0 \)). For an \( m \)-uniform hypergraph, the Region of Impossibility is given by \( \ell_m \to 0 \), which is the same as that in Theorem 2.3 when \( m = 3 \). While many lower bound results are available for uniform hypergraphs [2, 7, 20, 22], non-uniform hypergraphs are less studied. Our lower bound in Theorem 3.1 leads to two notable discoveries: (i) the Regions of Possibility/Impossibility for \( \mathcal{A}^{(m)} \) are fully characterized by the simple quantity of \( \ell_m \); (ii) the Regions of Possibility/Impossibility for non-uniform hypergraphs are fully characterized by the simple quantity of \( \max_{2 \leq m \leq M} \{\ell_m\} \).

### 3.1 Tightness of the lower bounds

We propose a test for the global testing problem (1.4) and show that it attains the lower bounds in Theorems 2.3 and Theorem 3.1. For simplicity, we only discuss this for DCBM with moderate degree heterogeneity but the tightness holds in much broader settings (e.g., DCMM).
For each $2 \leq m \leq M$, we first compute a vector $\eta^{(m)} \in \mathbb{R}^n$, which serves as an estimate of $\theta^{(m)}$ when the null hypothesis is true. Fix $m$. Given a positive vector $u \in \mathbb{R}^n$, define $L(u) \in \mathbb{R}^n$ by

$$L_i(u) = \frac{\sum_{i_1, \ldots, i_m = 1}^{n} A^{(m)}_{i_1 \ldots i_m} u_{i_1} u_{i_2} \cdots u_{i_m}}{\left(\sum_{i_1, \ldots, i_m = 1}^{n} A^{(m)}_{i_1 \ldots i_m} \right)^{(m-1)/m}}.$$ \hfill (3.18)

Let $N_m = \left\lceil \frac{m^2}{2} \right\rceil$. Initialize at $u(0) = \mathbf{0}_n$, compute $u(k) = L(u(k-1))$ iteratively for $k = 1, \ldots, N_m$, and output $u(N_m)$ as $\eta^{(m)}$. We note that each $u_i^{(1)}$ is a simple function of node degrees, $1 \leq i \leq n$. For $m \in \{2, 3\}$, since $\eta^{(m)} = u^{(1)}$, we estimate $\theta^{(m)}$ directly from node degrees. However, for $m \geq 4$, $u^{(1)}$ is a biased estimator of $\theta^{(m)}$ under the null, and the bias is caused by $\text{diag}(Q^{(m)})$. The iteration serves to reduce this bias. The required number of iterations depends on $m$ explicitly.

Next, we construct a statistic $Q_n^{(m)}$ to capture the difference between $A^{(m)}$ and a rank-1 estimate of the Bernoulli probability tensor. Let $A^{(m)}$ be a tensor of the same size as $A^{(m)}$, where $A^{(m)}_{i_1 \ldots i_m} = A^{(m)}_{i_1 \ldots i_m} - \eta^{(m)}_{i_1} \cdots \eta^{(m)}_{i_m}$ for $1 \leq i_1, \ldots, i_m \leq n$. We say that $S = (S_1, S_2, \ldots, S_{m+1})$ is an $(m+1)$-partition of $\{1, 2, \ldots, n\}$ if $S_1, \ldots, S_{m+1}$ are disjoint and their union is $\{1, 2, \ldots, n\}$. Let $B$ be the set of all $(m+1)$-partitions. For each $S = (S_1, \ldots, S_{m+1}) \in B$ and $1 \leq k_1, \ldots, k_m \leq m+1$, let

$$Q_n^{(m)} = \max_{S=(S_1, \ldots, S_{m+1}) \in B} \max_{1 \leq k_1, \ldots, k_m \leq m+1} \left\{ \sum_{i_1 \in S_{k_1}, \ldots, i_m \in S_{k_m}} A^{(m)}_{i_1 \ldots i_m} \right\}.$$ \hfill (3.19)

Finally, we combine $Q_n^{(2)}, \ldots, Q_n^{(M)}$. For each $m$, let $V_n^{(m)} = \frac{n}{m} \hat{\alpha}_n (1 - \hat{\alpha}_n)$, where $\hat{\alpha}_n = \lfloor (n-m)/n \rfloor \sum_{1 \leq i_1, \ldots, i_m = 1} A^{(m)}_{i_1 \ldots i_m}$. The test statistic is

$$\phi_n = \max_{2 \leq m \leq M} \left\{ Q_n^{(m)} / \left[ n \log(n) \right]^{1/2} \right\}.$$ \hfill (3.20)

Recall that by Theorem 3.1, for any alternative with $\max_{2 \leq m \leq M} \{ \ell_m \} = o(1)$, we can pair it with a null so that the pair are asymptotically indistinguishable. The next theorem says that the proposed test statistic $\phi_n$ can successfully separate any alternative satisfying $\max_{2 \leq m \leq M} \{ \ell_m \} \geq \log^{1+\delta}(n)$ (for some $\delta > 0$; taking $\delta = 0.1$ is adequate) from the null. Therefore, except for a logarithmic factor, our lower bounds are tight. Recall that $\ell_m = \| \theta^{(m)} \|_1^{-1} - \| \theta^{(m)} \|_2^{-2} / \| \mu_2^{(m)} \|^2 \log(n) \to \infty$, for every $2 \leq m \leq M$. Then, $\phi_n \to 0$ in probability, if $H_0$ holds, and $\phi_n \to \infty$ in probability, if $H_1$ holds and $\max_{2 \leq m \leq M} \{ \ell_m \} / [\log^{1+\delta}(n)] \to \infty$.

To speed up the computation of $\phi_n$ for large $n$, we introduce a proxy statistic by replacing the search over $B$ by a specific $\hat{B}$ as follows: conduct SVD on the matricization of $A^{(m)}$, take the first $(m+1)$ left singular vectors, and apply the spectral algorithm [13] to partition nodes into $(m+1)$ groups.

We use this partition $\hat{B}$ to replace the maximization over $B$ in (3.19), to get a proxy to $Q_n^{(m)}$; the test statistic $\hat{\phi}_n$ is then defined similarly as in (3.20). As long as the hypergraph is not too sparse, this proxy works well and is computationally much faster.

4 Numerical study

We use simulated data to validate our theoretical results. Fix $(n, K, m) = (500, 2, 3)$. In Experiment 1, we consider the SBM model and verify that the Regions of Impossibility are different for symmetric and asymmetric SBM (see Section 2.4). Let $\theta_i = n^{-1/2}$ for $1 \leq i \leq n$, and $P_{ij} = 1$ if $i = j = k$ and $P_{ijk} = 1/4$ otherwise. We consider a symmetric case where each communities have 250 nodes and an asymmetric case where two communities have 375 and 125 nodes, respectively. For each setting, we randomly generate the hypergraphs, apply the degree-based $\chi^2$-statistic $\psi_n$ in Section 2.4, and repeat for 500 times. The histograms of $\psi_n$ for two cases are on the left panel of the figure below.
We consider the problem of global testing for non-uniform hypergraphs in a broad DCMM setting. We see that $\phi_n$ indeed has no power. Let $(n, K, m)$ be the same as above. In the null, we let $\theta_i$ be iid drawn from Pareto$(0.5, 5)$ and then re-normalize the vector of $\theta$ so that $n^{-2}\|\theta\|_1^2 = c_n$, for $c_n = n^{-1}$; in the alternative, let $(P^*, h)$ be the same as in the asymmetric case in Experiment 1 and generate $\theta^*_n$ as in (2.12) ($D$ is obtained by treating $D(P(Dh))Dh = I_K$ as a nonlinear equation and apply the Matlab function \texttt{solve}). The histograms of $\psi_n$ under the null (green) and alternative (red) are shown on the middle panel of the figure below. The two histograms are inseparable from each other.

In Experiment 3, we study the test statistic $\phi_n$ proposed in Section 3.1. The simulation setting is the same as that in Experiment 2, except that $c_n = n^{-1/2}$. To save computing time, we use the proxy of $\phi_n$ by plugging in a $\hat{B}$ from spectral clustering (see Section 3.1). The histograms of the test statistic under the null (green) and alternative (red) are shown on the right panel of the figure above. We see that $\phi_n$ successfully distinguishes the alternative from the null. This validates our result in Theorem 3.2. The distribution of $\phi_n$ in the alternative has two modes, due to that the proxy $\hat{B}$ we plug in has two most frequent realizations. However, this does not affect the testing performance.

5 Conclusion

We consider the problem of global testing for non-uniform hypergraphs in a broad DCMM setting. Given an alternative, how to identify the null that is hardest to separate from the alternative is a challenging problem. We solve this by proposing a degree matching strategy, and use it to derive a tight lower bound by tensor scaling techniques and delicate analysis of the $\chi^2$-divergence. We discover that for an $m$-uniform hypergraph, the Regions of Impossibility/Possibility are governed by the simple quantity of $\ell_m = ||\theta^{(m)}||_2^{-2}||\varphi^{(m)}||_2^{-2}(\mu_2^{(m)})^{-2}$ (and so those for a non-uniform hypergraph are governed by $\max_{2 \leq m \leq M}\{\ell_m\}$). We also propose a new test that attains the lower bounds, so our lower bounds are tight. For future work, we notice that the test in Section 3.1 is computationally expensive. It is desirable to find some fast algorithms that also achieve the lower bounds. The signed-cycle statistics [6, 14] are polynomial-time statistics that have shown appealing performance for network global testing. It is possible that these statistics can be generalized to hypergraphs to provide polynomial-time tests that are also theoretically optimal. We leave it to future study.

Acknowledgments and Disclosure of Funding

The authors thank the anonymous reviewers for useful comments. J. Jin is supported in part by NSF Grant DMS-2015469. Z. Ke is supported in part by NSF CAREER Grant DMS-1943902.

References


Supplement of “Sharp Impossibility Results for Hyper-graph Global Testing”

In this supplement file, we first present the impossibility results for RMM-DCMM, which is omitted from the main text due to space limit. Then, we prove all the theorems and lemmas. Note that in this paper, $C$ is a generic constant that may vary from occurrence to occurrence.

A The region of impossibility for RMM-DCMM

For RMM-DCMM models, we allow mixed-memberships. The discussion is quite similar, and the impossibility result in Section 2.2 continues to hold under a mild condition.

Similarly, consider a model pair, where we have a null DCMM model and an RMM-DCMM model with $K$ communities as the alternative. Denote the Bernoulli probability tensors by $Q$ and $Q^*$, respectively. Similarly, for any matrix $D = \text{diag}(d_1, d_2, \ldots, d_K)$ with $d_k > 0$, $1 \leq k \leq K$, let $P^D$ be the tensor with the same size of $P$ satisfying $P^D_{k_1k_2k_3} = d_{k_1}d_{k_2}d_{k_3}P_{k_1k_2k_3}$. Also, let $h^D = E[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$ and $\tilde{a}^D = (P^Dh^D)h^D$. We assume that there is a matrix $D$ such that

$$\tilde{a}^D = 1_K, \quad \min_{1 \leq k \leq K} \{h^D_k\} \geq C. \tag{A.3}$$

Recall that in Lemma 2.1, we have shown that such a matrix $D$ always exists for DCBM. To see the point, note that if we do not allow mixed-memberships, then each realized $\pi_i$ is degenerate (i.e., only one entry is 1, all other entries are 0). In this case, $h^D = E_F[\pi_i] = h$, and $\tilde{a}^D = a^D$. Therefore, (A.3) always holds, by Lemma 2.1. For this reason, (A.3) is only a mild condition.

Suppose now (A.3) holds for a matrix $D = D_0$. Let $P^*$ and $\tilde{a}^*$ be $P^D$ and $\tilde{a}^D$ evaluated at $D = D_0$, respectively. By definitions, $\tilde{a}^* = 1_K$. For $1 \leq i \leq n$, let

$$\theta_i^* = \theta_i/\|D_0^{-1}\pi_i\|_1, \quad \pi_i^* = D_0^{-1}\pi_i/\|D_0^{-1}\pi_i\|_1. \tag{A.4}$$

Combining them with (A.2), for all $1 \leq i_1, i_2, i_3 \leq n$, we have $Q^*_{i_1i_2i_3} = \theta^*_{i_1}\theta^*_{i_2}\theta^*_{i_3}(P^\pi_{i_3})\pi_{i_2} = \theta_{i_1}\theta_{i_2}\theta_{i_3}(P^\pi_{i_3})\pi^*_{i_2}$. By similar calculations, for $1 \leq i_1 \leq n$, the leading term of the expected degree of node $i_1$ under the alternative is $\theta_{i_1}\|\theta\|_1^2(\pi^*_{i_3})'\tilde{a}^* = \theta_{i_1}\|\theta\|_1^2$, where the right hand side is the leading term of the expected degree of node $i_1$ under the null. Therefore, we have the desired degree matching as before. The following theorem is proved in Section D.

Theorem A.1 (Impossibility for DCMM). Fix $K > 1$. Given $(\theta, P, h, F)$, consider a pair of models, an alternative with $K$ communities and a null, as in (A.2) and (A.1) respectively, where (A.3) holds and $\theta^*$ is given by (A.4). Suppose (2.6) hold and $\|\theta\|_1\|\theta\|^2\mu_2^2 = o(1)$. As $n \to \infty$, ...
the $\chi^2$-divergence between the pair tends to 0. Therefore, the two models are asymptotically indistinguishable in the sense that the sum of Type I and Type II errors of any test is no smaller than $1 + o(1)$.

Similarly, in the parameter space ($\theta, \mathcal{P}, h, F$) for DCCM, we call the region prescribed by $\|\theta\|_1 \|\theta\|_2 \mu_2^2 \to 0$ the Region of Impossibility. For any model in this region, we can pair it with a null so they are asymptotically inseparable.

We next generalize the result to non-uniform DCMM. Consider a DCMM null model with probability tensors $Q[M] = \{Q^{(1)}, \ldots, Q^{(M)}\}$ and an RMM-DCMM model with probability tensors $Q^*[M] = \{Q^{*(1)}, \ldots, Q^{*(M)}\}$, where for every $2 \leq m \leq M$ and $1 \leq i_1, i_2, \ldots, i_m \leq n$,

$$Q_{i_1, i_2, \ldots, i_m}^{(m)} = \theta^{(m)}_{i_1} \theta^{(m)}_{i_2} \cdots \theta^{(m)}_{i_m}, \quad Q_{i_1, i_2, \ldots, i_m}^{*(m)} = \theta^{*(m)}_{i_1} \theta^{*(m)}_{i_2} \cdots \theta^{*(m)}_{i_m} \|P^{(m)}; \pi_{i_1}, \ldots, \pi_{i_m}\|, \quad \pi_i \sim F. \quad (A.5)$$

For any matrix $D^{(m)} = \text{diag}(d_{11}^{(m)}, d_{22}^{(m)}, \ldots, d_{Kk}^{(m)})$ with $d_{kk}^{(m)} > 0$, $1 \leq k \leq K$, let $\overline{P}^{(m)}$ be the tensor with the same size of $P^{(m)}$ satisfying $\overline{P}_{k_1k_2\cdots k_m}^{(m)} = d_{k_1}^{(m)} d_{k_2}^{(m)} \cdots d_{k_m}^{(m)} P_{k_1k_2\cdots k_m}^{(m)}$. Also, let $\overline{h}^{(m)} = E[D^{(m)}]^{-1} - \min_{1 \leq i \leq K} h_i^{(m)}$ and $\overline{a}^{(m)} = \sum_{1 \leq i_1, \ldots, i_m \leq K} d_{i_11}^{(m)} \cdots d_{i_m, i_m}^{(m)} h_{i_1}^{(m)} \cdots h_{i_m}^{(m)}$, for every $1 \leq i_1 \leq K$. We assume that there are matrices $D^{(2)}, \ldots, D^{(m)}$ such that for $m = 2, \ldots, M$

$$\overline{a}^{(m)} = 1_K, \quad \min_{1 \leq i \leq K} \{\overline{h}^{(m)}_i\} \geq C. \quad (A.7)$$

Note that (A.7) always holds for non-uniform DCBM, by Lemma C.1 in Section C below. For this reason, (A.7) is only a mild condition.

Suppose now (A.7) holds for a matrix $D^{(m)} = D^{(m)}_0$, for $m = 2, \ldots, M$. Let $P^{*(m)}$ and $\overline{a}^{*(m)}$ be $P^{(m)}$ and $\overline{a}^{(m)}$ evaluated at $D^{(m)} = D^{(m)}_0$, respectively. By definitions, $\overline{a}^{*(m)} = 1_K$. For $1 \leq i \leq n$, $2 \leq m \leq M$, let

$$\theta^{(m)}_i = \theta_i^{(m)} / \|D_0^{(m)}\|^{-1} \|\pi_i\|_1, \quad \overline{\theta}^{(m)}_i = \theta_i^{(m)} / \|D_0^{(m)}\|^{-1} \|\pi_i\|_1. \quad (A.8)$$

This is analogous to the degree matching strategy in (A.4), and it is conducted for each $m$ separately. Let $\mu_2^{(m)}$ be the second singular value of $P^{(m)}$. For short, let $\ell_m = \|\theta^{(m)}_i\|_1 - \|\theta^{(m)}_i\|_2^2 (\mu_2^{(m)})^2$. The following Theorem is for non-uniform DCMM.

**Theorem A.2** (Impossibility for non-uniform RMM-DCMM). Fix $K > 1$ and $M \geq 2$. For any given $(h, F)$ and $\{\theta^{(m)}_i, P^{(m)}\}_{2 \leq m \leq M}$, consider a pair of models, a null as in (A.6) and an alternative with $K$ communities as in (A.5), where (A.7) hold and $\{\theta^{(m)}_i\}_{1 \leq i \leq n, 2 \leq m \leq M}$ are as in (A.8). Suppose $\|P^{(m)}\| \leq C$ and $\max_{1 \leq i \leq n} \theta_i^{(m)} \leq C$. If $\max_{2 \leq m \leq M} \{\ell_m\} = o(1)$, then as $n \to \infty$, the $\chi^2$-divergence between the pair tends to 0.

## B Proof of Theorem 2.2

Fix an arbitrary $(\theta, \mathcal{P}, h, F)$ that satisfies the requirement of Theorem A.1. We consider a pair of models: a null model where $Q_{i_1i_2\cdots i_m} = \theta_i^{(m)} \theta_i^{(m)} \cdots \theta_i^{(m)}$ and a $K$-community uniform RMM-DCMM model as in Theorem A.1. Let $\mathcal{P}_0^{(n)}$ and $\mathcal{P}_1^{(n)}$ denote the probability measures associated with these two models, respectively. We further modify $\mathcal{P}_1^{(n)}$ as follows. In this RMM-DCMM, the membership matrix $\Pi$ is randomly generated. Let $\Pi_0$ be a non-random membership matrix such that $(\theta, \Pi_0, \mathcal{P}) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n)$. We define

$$\Pi = \begin{cases} \Pi, & \text{if } (\theta, \Pi, \mathcal{P}) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n), \\ \Pi_0, & \text{otherwise.} \end{cases} \quad (B.9)$$
We construct a similar RMM-DCMM by replacing $\Pi$ with $\tilde{\Pi}$ and denote $\tilde{\mathcal{P}}_1^{(n)}$ the probability measure associated with this new RMM-DCMM.

Consider a pair of hypotheses, where $\mathcal{A}$ is generated from $\mathcal{P}_0^{(n)}$ under the null hypothesis and it is generated from $\mathcal{P}_1^{(n)}$ under the alternative hypothesis. Given any $\psi$, its sum of type I and type II errors is equal to

$$\mathcal{P}_0^{(n)}(\psi = 1) + \mathcal{P}_1^{(n)}(\psi = 0)$$

$$= \mathbb{P}_0(\psi = 1) + \mathbb{E}_\tilde{\Pi}[\mathbb{P}_1(\psi = 0|\tilde{\Pi})]$$

$$\leq \sup_{\theta \in \mathcal{M}_n(K,\alpha_n,\beta_n)} \mathbb{P}(\psi = 1) + \sup_{(\theta,\Pi,\mathcal{P}) \in \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)} \mathcal{P}(\psi = 0).$$

In the last inequality, we have used the fact that $(\theta,\tilde{\Pi},\mathcal{P}) \in \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)$ for any realization of $\tilde{\Pi}$ (this is guaranteed by the construction in (B.9)). At the same time, by Neyman-Pearson lemma,

$$\mathcal{P}_0^{(n)}(\psi = 1) + \mathcal{P}_1^{(n)}(\psi = 0) \geq 1 - \|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1,$$

where $\|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1$ is the $L_1$-distance between two probability measures. Therefore, to show the claim, it suffices to show that

$$\|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1 = o(1). \quad \text{(B.10)}$$

We now show (B.10). Recall that in Theorem A.1 we have seen that the $\chi^2$-divergence between $\mathcal{P}_0^{(n)}$ and $\mathcal{P}_1^{(n)}$ tends to 0. Using the triangle inequality and the connection between $L_1$-distance and $\chi^2$-divergence (e.g., equation (2.27) of [5]), we have

$$\|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1 \leq \|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1 + \|\mathcal{P}_1^{(n)} - \mathcal{P}_1^{(n)}\|_1$$

$$\leq \sqrt{\chi^2(\mathcal{P}_0^{(n)},\mathcal{P}_1^{(n)})} + \|\mathcal{P}_1^{(n)} - \mathcal{P}_1^{(n)}\|_1$$

$$\leq o(1) + \|\mathcal{P}_1^{(n)} - \mathcal{P}_1^{(n)}\|_1. \quad \text{(B.11)}$$

It suffices to show that $\|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1 \to 0$. By (B.9), $\mathcal{P}_1^{(n)}$ is obtained from $\mathcal{P}_1^{(n)}$ by modifying those realizations of $\Pi$ where $(\theta,\Pi,\mathcal{P}) \notin \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)$. By some elementary calculations, we have

$$\|\mathcal{P}_1^{(n)} - \mathcal{P}_1^{(n)}\|_1 \leq 2 \mathbb{P}(\{\theta,\Pi,\mathcal{P} \notin \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)})$$

where $\mathbb{P}$ is with respect to the randomness of $\Pi$. In the definition of $\mathcal{M}_n(K,c_0,\alpha_n,\beta_n)$, the only requirement involving $\Pi$ is that $\max_{1 \leq k \leq K} \{g_k\} = c_0^{-1} \min_{1 \leq k \leq K} \{g_k\}$. The following lemma is proved below:

**Lemma B.1.** Fix a constant $c_0 \geq 1$. As $n \to \infty$, suppose $\|P\| \leq c_0$, $\theta_{\text{max}} \leq c_0$, and $\|\theta\|_1 \to \infty$. Write $h = \mathbb{E}[D^{-1}\pi_i/D^{-1}\pi_i]$. If $\min_{1 \leq k \leq K} \{h_k\} \geq c_1$, for an appropriate constant $c_1 > 0$, then as $n \to \infty$, with probability $1 - o(1)$, the following condition is satisfied,

$$\frac{\max_{1 \leq k \leq K} \{g_k\}}{\min_{1 \leq k \leq K} \{g_k\}} \leq c_0^{-1}.$$ 

By Lemma B.1, the probability of $(\theta,\Pi,\mathcal{P}) \notin \mathcal{M}_n(K,c_0,\alpha_n,\beta_n)$ tends to 0 as $n \to \infty$. It follows that $\|\mathcal{P}_0^{(n)} - \mathcal{P}_1^{(n)}\|_1 \to 0$. We plug it into (B.11) to get (B.10). This completes the proof.

**B.1 Proof of Lemma B.1**

Recall that $g_k = (1/\|\theta\|_1) \sum_{i=1}^n \theta_i \pi_i(k)$, for $1 \leq k \leq K$. Since $\max_k \{\sum_{i=1}^n \theta_i \pi_i(k)\} \leq \|\theta\|_1$, it suffices to show that

$$\min_k \{\sum_{i=1}^n \theta_i \pi_i(k)\} \geq c_0 \|\theta\|_1. \quad \text{(B.12)}$$
Let \( c_1 \) be a constant such that \( c_1 > c_0 \). Our assumptions say that \( \min_{1 \leq k \leq K} \{ h_k \} \geq c_1 \), where \( h = \mathbb{E}[D^{-1} \pi_i/|D^{-1} \pi_i|] \). Let \( h^* = \mathbb{E}[\pi_i] \). We first show that \( \min_{1 \leq k \leq K} \{ h_k \} \geq c_1 \) implies \( \min_{1 \leq k \leq K} \{ h_k^* \} \geq c_1 \cdot [1 + o(1)] \). By Lemma E.5 in section E, we have

\[
\max_{1 \leq i \leq K} \{ |d_i| - 1 \} \leq C \mu_2 \quad \text{with} \quad \mu_2 = o(1),
\]

and so \( d_i = 1 + o(1), 1 \leq i \leq K \). By definitions, it follows that

\[
h_k \leq \mathbb{E}[\min_{1 \leq k \leq K} \{ d_k \}^{-1} \pi_i(k)/\max_{1 \leq k \leq K} \{ d_k \}^{-1}] \leq h_k^* \cdot [1 + o(1)].
\]

Combining this with \( \min_{1 \leq k \leq K} \{ h_k \} \geq c_1 \), we have \( \min_{1 \leq k \leq K} \{ h_k^* \} \geq c_1 \cdot [1 + o(1)] \).

Now we are going to show (B.12). Note that \( X = \sum_{i=1}^n \theta_i (\pi_i(k) - h_k^* \cdot 1) \) is a sum of independent mean-zero random variables, where \( \theta_i (\pi_i(k) - h_k^*) \leq C\theta_{\max} \) and \( \sum_{i=1}^n \text{Var}(\theta_i(\pi_i(k) - h_k^*)) \leq C\|\theta\|^2 \). By Bernstein’s inequality,

\[
P(|X| > t) \leq \exp\left(-\frac{t^2}{C\|\theta\|^2 + C\theta_{\max} t}\right), \quad \text{for any} \ t > 0.
\]

Taking \( t = C\|\theta\| \sqrt{\log(||\theta||_1 + C\theta_{\max} \log(||\theta||_1))} \), it follows that, with probability at least \( 1 - ||\theta||_1^{-1} \),

\[
|\sum_i \theta_i (\pi_i(k) - h_k^*)||\theta||_1| = |X| \leq C\|\theta\| \sqrt{\log(||\theta||_1) + C\theta_{\max} \log(||\theta||_1)},
\]

where by \( ||\theta||^2 \leq ||\theta||_1 \), the RHS is \( o(||\theta||_1) \). Combining this with \( \min_k \{ h_k^* \} \geq c_1 \cdot [1 + o(1)] \),

\[
\sum_i \theta_i \pi_i(k) = h_k^* ||\theta||_1 \cdot [1 + o(1)] \geq c_1 \|\theta\|_1 \cdot [1 + o(1)]
\]

where \( c_1 \) is a constant strictly larger than \( c_0 \). This proves (B.12). The claim follows.

\section*{C Proof of Lemma 2.1}

We prove a version of this lemma for \( m \)-uniform hypergraph below where the desired result is by letting \( m = 3 \).

\textbf{Lemma C.1} (Lemma 2.1 for \( m \)-uniform hypergraph). Fix \( K > 1 \) and \( m > 1 \). Let \( \mathcal{P} \) be a nonnegative \( m \)-uniform tensor of dimension \( K \) and \( h \) be a vector in \( \mathbb{R}^K \), where we assume \( \mathcal{P}_{i_1 \ldots i_m} = 1 \), for \( i_1 = 1, \ldots, K \) and \( \min \{ h_1, h_2, \ldots, h_K \} \geq C \). There exists an unique diagonal matrix \( D = \text{diag}(d_1, d_2, \ldots, d_K) \) such that

\[
\sum_{i_1, \ldots, i_m=1}^K d_{i_1} \mathcal{P}_{i_1 \ldots i_m} \cdot (d_{i_2} h_{i_2}) \cdots (d_{i_m} h_{i_m}) = 1, \quad \text{for all} \ i_1 = 1, \ldots, K. \tag{C.13}
\]

To begin with, we transform the problem (C.13) into an equivalent form (C.14).

Multiplying \( h_{i_1} \) on both sides of (C.13) and let \( \bar{d}_{i_1} = d_{i_1} h_{i_1} \) for \( i_1 = 1, \ldots, K \). It is equivalent to find an unique diagonal matrix \( \bar{D} = \text{diag}(\bar{d}_1, \ldots, \bar{d}_K) \) with strictly positive entries such that

\[
\sum_{i_1, \ldots, i_m=1}^K \bar{d}_{i_1} \mathcal{P}_{i_1 \ldots i_m} \bar{d}_{i_2} \cdots \bar{d}_{i_m} = h_{i_1}, \quad \text{for all} \ i_1 = 1, \ldots, K. \tag{C.14}
\]

Now, by the Theorem 6 in [1], for a nonnegative order-\( m \) tensor \( \mathcal{P} \) of dimension \( K \) (not necessarily symmetric) such that \( \mathcal{P}_{i_1 \ldots i_m} > 0, \ i_1 = 1, \ldots, K, \) and \( K \) positive numbers \( h_1, \ldots, h_K \), there exist positive numbers \( x_1, \ldots, x_K \) such that

\[
\sum_{i_1, \ldots, i_m=1}^K x_{i_1} \mathcal{P}_{i_1 \ldots i_m} x_{i_2} \cdots x_{i_m} = h_{i_1}, \quad \text{for all} \ i_1 = 1, \ldots, K. \tag{C.15}
\]
which gives the existence of such $\tilde{D}$ satisfying (C.14).

The uniqueness of such $\tilde{D}$ is given by the Theorem 1.1 in [2] which states that there is a unique tensor $A$ that is defined by $A_{i_1...i_m} = \hat{d}_{i_1} \cdots \hat{d}_{i_m} P_{i_1...i_m}$ for $i_1, \ldots, i_m = 1, \ldots, K$ and satisfies

$$\sum_{i_2, \ldots, i_m = 1}^{K} A_{i_1...i_m} = h_{i_1}, \quad \text{for all } i_1 = 1, \ldots, K.$$  \hfill (C.16)

Therefore, $\tilde{D}$ is unique since $A$ is unique and one-to-one correspondence with $\tilde{D}$. This completes the proof.

## D Proof of Theorem 2.1, Theorem 2.3 and Theorem A.1-A.2

Theorem 2.1 and Theorem 2.3 are the special cases of Theorem 3.1, which do not need separate proofs. Furthermore, in the proof of Theorem 3.1 below, we actually consider the more general setting of non-uniform DCMM where $\theta^r$ is constructed as $\theta^r = \theta_i / \| D^{-1} \pi_i \|_1$ (note that when $\pi_i$ is degenerate, this reduces to the construction of $\theta^r = \theta_i d_k$ for DCBM). Therefore, the proof of Theorem 3.1 (for non-uniform DCMM) already includes the proof of Theorem A.1 (for 3-uniform DCMM) and Theorem A.2 (for non-uniform DCMM). It remains to prove Theorem 3.1, which is contained in Section E.

## E Proof of Theorem 3.1

We first state the preliminary lemmas, Lemmas E.1-E.5, needed for the proof of Theorem 3.1. Next, we prove this theorem. Finally, we prove all the preliminary lemmas.

### E.1 Preliminary lemmas

The following lemmas are used in the main proof and proved after the main proof.

**Lemma E.1.** Let $\mathcal{P}$ be a $m$-way symmetric $K$ dimensional tensor, $\mathcal{P}_0$ be the tensor with the same size as $\mathcal{P}$ where all entries are 1, and introduce a tensor $\mathcal{M}$ by $\mathcal{M} = \mathcal{P} - \mathcal{P}_0$. Let $h, \pi_1$ be weight vectors in $\mathbb{R}^K$ and $y_i = \pi_i - h$, for $1 \leq i \leq n$. Then

$$[\mathcal{P}; \pi_1, \ldots, \pi_m] = 1 + x^{(m)} + z^{(m)}, \quad \text{holds for any } m > 1,$$

where

$$x^{(m)} = \langle \mathcal{M}; h, \ldots, h \rangle + \sum_{s=1}^{m} \langle \mathcal{M}; h, \ldots, h, y_s, h, \ldots, h \rangle,$$

$$z^{(m)} = \sum_{s_1 = 1}^{m-1} \sum_{s_2 = s_1+1}^{m} \sum_{s_1-1}^{m-1} \langle \mathcal{M}; h, \ldots, h, y_{s_1}, h, \ldots, h, y_{s_2}, \pi_{s_2+1}, \ldots, \pi_m \rangle.$$

**Lemma E.2.** With the same notations as in Section E.2, let $\{w^{(j)}_i : 1 \leq i \leq n, 1 \leq j \leq m\}$ be a set of weight vectors in $\mathbb{R}^K$ and $\{w^{(j)}_i\}$ be an independent copy of $\{w^{(j)}_i\}$. Assume that for distinct $i_1, \ldots, i_m$, vectors $y_{i_1}, y_{i_2}, w^{(3)}_{i_3}, \ldots, w^{(m)}_{i_m}$ are mutually independent and that $\| \mathcal{M} ; k_3 \ldots k_m \| \leq C \mu$, for $1 \leq k_3, \ldots, k_m \leq K$. Denote

$$S = \sum_{i_1, \ldots, i_m}^{(dxt)} \frac{\langle \theta_1 \cdots \theta_m \rangle^t}{a_i} [\mathcal{M}; y_{i_1}, y_{i_2}, w^{(3)}_{i_3}, \ldots, w^{(m)}_{i_m}] [\mathcal{M}; \tilde{y}_{i_1}, \tilde{y}_{i_2}, \tilde{w}^{(3)}_{i_3}, \ldots, \tilde{w}^{(m)}_{i_m}].$$

5
Then, for any constant \( c \) independent of \( n \),
\[
E \left[ \exp(cS) \right] \leq E \left[ \exp \left( C \mu_2^2 \|\theta^t\|_t^{(m-2)} |T| / a_t \right) \right] \cdot \exp(C \mu_2^2 \|\theta^t\|_t^{(m-2)} \|\theta\|_{2t}^2 / a_t),
\]
where \( T \) is a random variable satisfying \( P(|T| > x) \leq 2 \exp(-x/(2K^2 \|\theta\|_{2t}^2)) \), for \( x > 0 \).

**Lemma E.3.** With the same setting in Lemma E.2, denote
\[
S = \sum_{i_1, \ldots, i_m(\text{dist})} \frac{(\theta_{i_1} \cdots \theta_{i_m})^t}{a_t} [M; y_{i_1}, y_{i_2}, w_{i_3}^{(3)}, \ldots, w_{i_m}^{(m)} | M; \bar{y}_{i_1}, \bar{y}_{i_2}, \bar{w}_{i_3}^{(4)}, \ldots, \bar{w}_{i_m}^{(m)}].
\]
Then, for any constant \( c \) independent of \( n \),
\[
E \left[ \exp(cS) \right] \leq E \left[ \exp \left( C \mu_2^2 \|\theta^t\|_t^{(m-2)} |T| / a_t \right) \right] \cdot \exp(C \mu_2^2 \|\theta^t\|_t^{(m-2)} \|\theta\|_{2t}^2 / a_t),
\]
where \( T \) is a random variable satisfying \( P(|T| > x) \leq 4 \exp(-x/(2K^2 \|\theta\|_{2t}^2)) \), for \( x > 0 \).

**Lemma E.4.** With the same setting in Lemma E.2, denote
\[
S = \sum_{i_1, \ldots, i_m(\text{dist})} \frac{(\theta_{i_1} \cdots \theta_{i_m})^t}{a_t} [M; y_{i_1}, y_{i_2}, w_{i_3}^{(3)}, \ldots, w_{i_m}^{(m)} | M; h, h, \bar{y}_{i_3}, \bar{y}_{i_4}, \bar{w}_{i_5}^{(5)}, \ldots, \bar{w}_{i_m}^{(m)}].
\]
Then, for any constant \( c \) independent of \( n \),
\[
E \left[ \exp(cS) \right] \leq E \left[ \exp \left( C \mu_2^2 \|\theta^t\|_t^{(m-2)} |T| / a_t \right) \right] \cdot \exp(C \mu_2^2 \|\theta^t\|_t^{(m-2)} \|\theta\|_{2t}^2 / a_t),
\]
where \( T \) is a random variable satisfying \( P(|T| > x) \leq 4 \exp(-x/(2K^2 \|\theta\|_{2t}^2)) \), for \( x > 0 \).

**Lemma E.5.** Under the conditions of Theorem 3.1, for \( m = 2, \ldots, M \) we have
\[
\max_{1 \leq k_1, \ldots, k_m \leq K} \|M^{(m)}_{k_1 \cdots k_m}\| \leq C \mu_2^{(m)}, \quad \max_{1 \leq i \leq K} |d^{(m)}_{i} - 1| \leq C |\mu_2^{(m)}|,
\]
where \( M^{(m)} \) is a \( m \)-way symmetric tensor defined by \( M_{k_1 \cdots k_m}^{(m)} = (P_{k_1 \cdots k_m}^{(m)} - 1) d_{k_1}^{(m)} \cdots d_{k_m}^{(m)} \), \( 1 \leq k_1, \ldots, k_m \leq K \).

**E.2 Proof of Theorem 3.1**

Let \( P_0^{(n)} \) and \( P_1^{(n)} \) denote the probability measures associated with the null and alternative hypotheses, respectively, and let \( \chi^2(P_0^{(n)}, P_1^{(n)}) \) be the \( \chi^2 \) divergence between the two probability measures. By definitions,
\[
\chi^2(P_0^{(n)}, P_1^{(n)}) = \int_A \left[ \frac{dP_1^{(n)}(A)}{dP_0^{(n)}(A)} \right]^2 dP_0^{(n)}(A) - 1.
\]

To show the claim, it suffices to show that when \( (\mu_2^{(m)})^2 \|\theta^{(m)}\|_1^{m-2} \|\theta^{(m)}\|_2^2 \to 0 \), \( m = 1, \ldots, M \), we have
\[
\int_A \left[ \frac{dP_1^{(n)}(A)}{dP_0^{(n)}(A)} \right]^2 dP_0^{(n)}(A) = 1 + o(1). \tag{E.17}
\]

By definitions,
\[
dP_0^{(n)}(A) = \prod_{m=2}^M \prod_{i_1 < \cdots < i_m} dP_0^{(n,m)}(A_{i_1 \cdots i_m}^{(m)}),
\]
\[
dP_1^{(n)}(A) = \prod_{m=2}^M \prod_{i_1 < \cdots < i_m} dP_1^{(n,m)}(A_{i_1 \cdots i_m}^{(m)} | \Pi).
\]
Let \( \Pi \) be an independent copy of \( \Pi \). Putting the above two equations into (E.17) gives

\[
\int_{A} \left[ \frac{dP_{1}^{(n)}(A)}{dP_{0}^{(n)}(A)} \right]^{2} dP_{0}^{(n)}(A) = \int_{A} \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}}) | \Pi) dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi) \right]
\]
\[
= \int_{A} \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} \frac{dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi) dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi)}{dP_{0}^{(n,m)}(A_{i_{1} \cdots i_{m}})} \right].
\]

Exchanging the order of integral and expectation in the last equation and by elementary probability,

\[
\int_{A} \left[ \frac{dP_{1}^{(n)}(A)}{dP_{0}^{(n)}(A)} \right]^{2} dP_{0}^{(n)}(A) = \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \int_{A} \prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi) dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi) \right]
\]
\[
= \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} \int_{A_{i_{1} \cdots i_{m}}} \frac{dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi) dP_{1}^{(n,m)}(A_{i_{1} \cdots i_{m}} | \Pi)}{dP_{0}^{(n,m)}(A_{i_{1} \cdots i_{m}})} \right].
\]

Let \( \chi_{1 \cdots i_{m}}(\Pi, \tilde{\Pi}) \) denote \( \int_{A_{1 \cdots i_{m}}} dP_{1}^{(n,m)}(A_{1 \cdots i_{m}} | \Pi) dP_{1}^{(n,m)}(A_{1 \cdots i_{m}} | \Pi)/dP_{0}^{(n,m)}(A_{1 \cdots i_{m}}) - 1 \). Hence

\[
\int_{A} \left[ \frac{dP_{1}^{(n)}(A)}{dP_{0}^{(n)}(A)} \right]^{2} dP_{0}^{(n)}(A) = \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} (\chi_{1 \cdots i_{m}}(\Pi, \tilde{\Pi}) + 1) \right]. \tag{E.18}
\]

Note that by inequality \( \prod_{i=1}^{M} (1 + x_{i}) \leq \exp(\sum_{i=1}^{M} x_{i}) \), for all \( x_{i} \) such that \( 1 + x_{i} \geq 0 \), we have

\[
\prod_{m=2}^{M} \prod_{i_{1} < \cdots < i_{m}} (\chi_{1 \cdots i_{m}}(\Pi, \tilde{\Pi}) + 1) \leq \exp \left( \sum_{m=2}^{M} \sum_{i_{1} < \cdots < i_{m}} \chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) \right). \tag{E.19}
\]

Further, by Jensen’s inequality, \( \exp(\sum_{i=1}^{M} x_{i}) \leq \sum_{i=1}^{M} \exp(x_{i}) \). It follows that

\[
\exp \left( \sum_{m=2}^{M} \sum_{i_{1} < \cdots < i_{m}} \chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) \right) \leq \sum_{m=2}^{M} \frac{1}{M-1} \exp \left( (M-1) \sum_{i_{1} < \cdots < i_{m}} \chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) \right). \tag{E.20}
\]

Combining (E.18)-(E.20) gives

\[
\int_{A} \left[ \frac{dP_{1}^{(n)}(A)}{dP_{0}^{(n)}(A)} \right]^{2} dP_{0}^{(n)}(A) \leq \sum_{m=2}^{M} \frac{1}{M-1} \mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \exp \left( (M-1) \sum_{i_{1} < \cdots < i_{m}} \chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) \right) \right].
\]

Therefore, to show (E.17), it is sufficient to show that when the conditions hold, for each \( m = 2, \ldots, M \) we have

\[
\mathbb{E}_{\Pi, \tilde{\Pi}} \left[ \exp \left( (M-1) \sum_{i_{1} < \cdots < i_{m}} \chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) \right) \right] = 1 + o_{n}(1). \tag{E.21}
\]

Fix \( m \), recall that

\[
\chi_{1 \cdots i_{m}}^{2}(\Pi, \tilde{\Pi}) = \int_{A} dP_{1}^{(n,m)}(A_{1 \cdots i_{m}} | \Pi) dP_{1}^{(n,m)}(A_{1 \cdots i_{m}} | \Pi) - 1. \tag{E.22}
\]

By definitions,

\[
dP_{0}^{(n,m)}(A_{1 \cdots i_{m}}) = (Q_{1 \cdots i_{m}}^{(m)})(A_{1 \cdots i_{m}} - Q_{1 \cdots i_{m}}^{(m)})^{-1} A_{1 \cdots i_{m}},
\]
\[
dP_{1}^{(n,m)}(A_{1 \cdots i_{m}} | \Pi) = (Q_{1 \cdots i_{m}}^{(m)} | \Pi)(A_{1 \cdots i_{m}} - Q_{1 \cdots i_{m}}^{(m)} | \Pi)^{-1} A_{1 \cdots i_{m}}.
\]
Putting the above two equations into (E.22) gives

\[ \chi^2_{i_1 \cdots i_m}(\Pi, \bar{\Pi}) = \frac{Q^{(m)}_{i_1 \cdots i_m}(\Pi)Q^{(m)}_{i_1 \cdots i_m}(\bar{\Pi}) + (1 - Q^{(m)}_{i_1 \cdots i_m}(\Pi))(1 - Q^{(m)}_{i_1 \cdots i_m}(\bar{\Pi}))}{1 - Q^{(m)}_{i_1 \cdots i_m}} \]

\[ = \left( \frac{Q^{(m)}_{i_1 \cdots i_m}(\Pi) - Q^{(m)}_{i_1 \cdots i_m}}{1 - Q^{(m)}_{i_1 \cdots i_m}} \right) \left( \frac{Q^{(m)}_{i_1 \cdots i_m}(\bar{\Pi}) - Q^{(m)}_{i_1 \cdots i_m}}{1 - Q^{(m)}_{i_1 \cdots i_m}} \right) \]  

(E.23)

Based on the expression of \( \chi^2_{i_1 \cdots i_m}(\Pi, \bar{\Pi}) \), it is seen that the LHS of (E.21) only relates to the variables in \( m \)-uniform tensor DCMM (e.g., \( A^{(m)}, Q^{(m)}, P^{(m)}, \theta^{(m)} \)), for ease of notations, we remove the superscript \((m)\) whenever it is clear from the context.

Next we continue to simplify \( \chi^2_{i_1 \cdots i_m}(\Pi, \bar{\Pi}) \). According to the constructions of our model,

\[ Q_{i_1 \cdots i_m} = \theta_{i_1} \cdots \theta_{i_m} \quad \text{and} \quad Q_{i_1 \cdots i_m}^* = \theta_{i_1} \cdots \theta_{i_m} [P^*, \pi_{i_1}^*, \ldots, \pi_{i_m}^*], \]

where we recall that \( P^* \) is the \( m \)-uniform tensor defined by \( P_{k_1 \cdots k_m}^* = d_{k_1} \cdots d_{k_m} P_{k_1 \cdots k_m}, 1 \leq k_1, \ldots, k_m \leq K, \pi_i^* = D^{-1} \pi_i / ||D^{-1} \pi||_1, 1 \leq i \leq n \) and \( D = \text{diag}(d_1, d_2, \ldots, d_K) \) is the scaling matrix given by degree matching.

Let \( P_0 \) the tensor with the same size as \( P^* \) and where all entries are 1, and introduce a tensor \( M \) by \( M = P^* - P_0 \). Let \( h = E_P[\pi_i^*] \), and \( y_i = \pi_i^* - h, 1 \leq i \leq n \). By Lemma E.1, we can write the Bernoulli probability tensor for the alternative \( Q^* \) by

\[ Q_{i_1 \cdots i_m}^* = \theta_{i_1} \cdots \theta_{i_m} (1 + x_{i_1 \cdots i_m} + z_{i_1 \cdots i_m}), \quad 1 \leq i_1, \ldots, i_m \leq n, \]  

(E.24)

where

\[ x_{i_1 \cdots i_m} = [M; h, \ldots, h] + \sum_{s=1}^{m} \left[ M; h, \ldots, h, y_{i_s}, h, \ldots, h \right], \]

\[ z_{i_1 \cdots i_m} = \sum_{s_1=1}^{m-1} \sum_{s_2=s_1+1}^{m} \left[ M; h_{i_{s_1}}, h, \ldots, h, y_{i_{s_2}}, \pi_{i_{s_2+1}}^*, \ldots, \pi_{i_m}^* \right]. \]

Let \( e_{i_1} \) be the \( i_1 \)-th standard basis vector of the Euclidean space \( \mathbb{R}^K \), \( 1 \leq i_1 \leq K \). Note that by definitions and symmetry,

\[ [M; h, \ldots, h, e_{i_1}, h, \ldots, h] = \sum_{i_2, \ldots, i_m=1}^{K} (P_{i_1 \cdots i_m}^* - 1) \cdot h_{i_2} \cdots h_{i_m} \]

\[ = \sum_{i_2, \ldots, i_m=1}^{K} P_{i_1 \cdots i_m}^* \cdot h_{i_2} \cdots h_{i_m} - 1 \]

(By degree matching) \( = 0 \)

This indicates that any linear combination of elements in \( \{[M; h, \ldots, h, e_{i}, h, \ldots, h]; 1 \leq i \leq K\} \) equals to 0. It follows that the term \( x_{i_1 \cdots i_m} \) in the RHS of (E.24) equals to 0.

Write for short \( z_{i_1 \cdots i_m}(s_1, s_2) = [M; h, \ldots, h, y_{i_{s_1}}, h, \ldots, h, y_{i_{s_2}}, \pi_{i_{s_2+1}}^*, \ldots, \pi_{i_m}^*] \), we get

\[ Q_{i_1 \cdots i_m}^* = \theta_{i_1} \cdots \theta_{i_m} \left( 1 + \sum_{s_1=1}^{m} \sum_{s_2=s_1+1}^{m} z_{i_1 \cdots i_m}(s_1, s_2) \right) \]

(E.25)

Let \( \bar{z}_{i_1 \cdots i_m}(s_1, s_2) \) be \( z_{i_1 \cdots i_m}(s_1, s_2) \) evaluated at \( \bar{\Pi} \). Inserting (E.25) into (E.23) gives

\[ \chi^2_{i_1 \cdots i_m}(\Pi, \bar{\Pi}) = \frac{\theta_{i_1} \cdots \theta_{i_m}}{1 - \theta_{i_1} \cdots \theta_{i_m}} \sum_{s_1=1}^{m-1} \sum_{s_2=s_1+1}^{m} z_{i_1 \cdots i_m}(s_1, s_2) \bar{z}_{i_1 \cdots i_m}(s_1, s_2). \]
Note that $\frac{x}{1-x} = \sum_{i=1}^{\infty} x^i$ for any $x \in [0, 1)$, we have $\frac{\theta_1 \cdots \theta_m}{1-\theta_1 \cdots \theta_m} = \sum_{i=1}^{\infty} (\theta_1 \cdots \theta_m)^i$ and so

$$\chi^2_{i_1 \cdots i_m} (\Pi, \Pi') = \sum_{t=1}^{\infty} (\theta_1 \cdots \theta_m)^t \sum_{s_1=1, s_2=s_1+1}^{m-1} \sum_{s_1=1, s_2=s_1+1}^{m} z_{i_1 \cdots i_m} (s_1, s_2) \tilde{z}_{i_1 \cdots i_m} (\tilde{s}_1, \tilde{s}_2).$$

Introduce

$$a_t = \theta_m^{(t-1)} (1 - \theta_m^2),$$

$$S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) = (M - 1) 4^m \sum_{i_1 < \cdots < i_m} \frac{(\theta_1 \cdots \theta_m)^t}{a_t} z_{i_1 \cdots i_m} (s_1, s_2) \tilde{z}_{i_1 \cdots i_m} (\tilde{s}_1, \tilde{s}_2).$$

(E.26)

Exchanging the order of summation, we then can write

$$(M - 1) \sum_{i_1 < \cdots < i_m} \chi^2_{i_1 \cdots i_m} (\Pi, \Pi') = \sum_{t=1}^{\infty} \sum_{s_1=1, s_2=s_1+1}^{m-1} \sum_{s_1=1, s_2=s_1+1}^{m} \frac{a_t}{4^m} S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2).$$

Note that $\sum_{t=1}^{\infty} \sum_{s_1=1, s_2=s_1+1}^{m-1} \sum_{s_1=1, s_2=s_1+1}^{m} a_t/4^m = 1$ and $\exp(\cdot)$ is convex, by Jensen’s inequality

$$\exp\left( (M - 1) \sum_{i_1 < \cdots < i_m} \chi^2_{i_1 \cdots i_m} (\Pi, \Pi') \right) \leq \sum_{t=1}^{\infty} \sum_{s_1=1, s_2=s_1+1}^{m-1} \sum_{s_1=1, s_2=s_1+1}^{m} \frac{a_t}{4^m} \exp\left( S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right).$$

Therefore, to prove (E.21), it is sufficient to show that

$$\max_{t, s_1, s_2, \tilde{s}_1, \tilde{s}_2} \left\{ \mathbb{E} \left[ \exp\left( S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right] \right\} \leq 1 + o_n(1).$$

(E.27)

Fix $t, s_1, s_2, \tilde{s}_1, \tilde{s}_2$, we are going to bound $\mathbb{E} \left[ \exp\left( S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right]$. Recall that by construction, $s_1 < s_2$ and $\bar{s}_1 < \bar{s}_2$. By symmetry, without loss of generality, assume $s_2 \leq \bar{s}_2$. Now, we can separate the situations into three cases. Case 1: $s_1 = \bar{s}_1, s_2 = \bar{s}_2$; Case 2: Only one of $\{s_1, s_2\}$ matches any one of $\{\bar{s}_1, \bar{s}_2\}$ (e.g., $s_1 = \bar{s}_1 < s_2 < \bar{s}_2$ or $s_1 < s_2 = \bar{s}_1 < \bar{s}_2$ or $s_1 \neq \bar{s}_1, s_2 = \bar{s}_2$); Case 3: None of $\{s_1, s_2\}$ matches one of $\{\bar{s}_1, \bar{s}_2\}$.

Remark: Case 2 only exists for $m \geq 3$ and Case 3 only exists for $m \geq 4$. They require much tricky and delicate analysis to resolve extra random effects induced by $\Pi$. This indicates one of the differences on the calculations of the $\chi^2$-divergence between hypergraph and network.

By symmetry of $\mathcal{M}$, we summarized the derivation of the bounds on $\mathbb{E} \left[ \exp\left( S(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right]$ for Case 1,2,3 into Lemma E.2, E.3, E.4, respectively. Take Case 1 for example,

**Case 1** ($s_1 = \bar{s}_1, s_2 = \bar{s}_2$): By definitions and symmetry of $\mathcal{M}$, we can rewrite

$$S(t, s_1, s_2, \bar{s}_1, \bar{s}_2) := 4^m (M - 1) \sum_{i_1 < \cdots < i_m} \frac{(\theta_1 \cdots \theta_m)^t}{a_t} \left[ \mathcal{M}; h, \ldots, h, y_{i_1}, h, \ldots, h, y_{i_2}, \pi_{i_2+1}^*, \ldots, \pi_{i_m}^* \right]$$

$$\cdot \left[ \mathcal{M}; h, \ldots, h, \bar{y}_{i_1}, h, \ldots, h, \bar{y}_{i_2}, \pi_{i_2+1}^*, \ldots, \pi_{i_m}^* \right],$$

$$= \frac{4^m (M - 1)}{m!} \sum_{i_1, \ldots, i_m (\text{dist})} \frac{(\theta_1 \cdots \theta_m)^t}{a_t} \left[ \mathcal{M}; \bar{y}_{i_1}, \bar{y}_{i_2}, h, \ldots, h, \pi_{i_2+1}^*, \ldots, \pi_{i_m}^* \right],$$

which is implied by the standard forms discussed in Lemma E.2. Similarly, **Case 2** is implied by Lemma E.3 and **Case 3** is implied by Lemma E.4.
Combining Lemmas E.2-E.4 with Lemma E.5, we have

\[ \mathbb{E}\left[ \exp\left( \mathcal{S}(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right] \leq \mathbb{E}\left[ \exp\left( C \frac{\mu_2^2 \|	heta\|_t^{(m-2)} ||\theta||_t^2}{a_t} \right) \right] \cdot \exp\left( C \frac{\mu_2^2 \|	heta\|_t^{(m-2)} ||\theta||_t^2}{a_t} \right), \quad (E.28) \]

where \( \mu_2 \) is the second singular value of the matricization of the tensor \( \mathcal{P}^{(m)} \) and \( T \) is a random variable satisfying \( \mathbb{P}(|T| > x) \leq 4 \exp(-x/(2K^2 ||\theta||_t^2)) \), for any \( x > 0 \).

Now, we are ready to calculate a bound for \( \mathbb{E}\left[ \exp\left( \mathcal{S}(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right] \). By direct calculations,

\[ \mathbb{E}\left[ \exp\left( C \frac{\|	heta\|_t^{(m-2)}}{a_t} \right) \right] = \left( 1 + \int_0^\infty e^x \cdot \mathbb{P}\left( C \frac{\|	heta\|_t^{(m-2)}}{a_t} \right) > x) dx \right) \]
\[ \leq \left( 1 + \int_0^\infty e^{a_t x} \cdot 4 \exp\left(- \frac{a_t x}{2CK^2 \mu_2^2 \|	heta\|_t^{(m-2)} ||\theta||_t^2} \right) dx \right) \]

By \( \theta_{\text{max}} \leq c_0, ||\theta||_t \leq ||\theta||_1 \theta_{\text{max}}^{-1} \) and \( ||\theta||_t^{2t} \leq ||\theta||_1^{2t} \theta_{\text{max}}^{-2} \), we have

\[ \frac{a_t}{\mu_2^2 \|	heta\|_t^{(m-2)} ||\theta||_t^2} = c_{\text{max}}^{m(t-1)} (1 - \theta_{\text{max}}^m) \geq 1 - c_{0}^m \mu_2^2 \|	heta\|_1^{m-2} \|	heta\|_2^2 \]

Combining this with (E.28)-(E.29), we get

\[ \mathbb{E}\left[ \exp\left( \mathcal{S}(t, s_1, s_2, \tilde{s}_1, \tilde{s}_2) \right) \right] \leq \left( 1 + \int_0^\infty e^{a_t x} \cdot 4 \exp\left(- \frac{(1 - c_{0}^m)x}{2CK^2 \mu_2^2 \|	heta\|_1^{(m-2)} ||\theta||_2^2} \right) dx \right) e^{\frac{c_{0}^m}{2CK^2 \mu_2^2 \|	heta\|_1^{m-2} ||\theta||_2^2}} \]

where the RHS on the last inequality goes 1 as \( \mu_2^2 \|	heta\|_1^{m-2} ||\theta||_2^2 \to 0 \). This proves (E.27) and finishes the proof.

### E.3 Proof of Lemma E.1

Recall the definition of \( \mathcal{P}; \pi_1, \ldots, \pi_m \)

\[ \mathcal{P}; \pi_1, \ldots, \pi_m := \sum_{k_1, \ldots, k_m = 1}^K \mathcal{P}_{k_1 \ldots k_m} \pi_1(k_1) \cdots \pi_m(k_m). \]

Note that \( \mathcal{P} = \mathcal{M} + \mathcal{P}_0 \) and \( \sum_{k=1}^K \pi_i(k) = 1 \), for \( 1 \leq i \leq n \). By direct calculations

\[ \mathcal{P}; \pi_1, \ldots, \pi_m = \sum_{k_1, \ldots, k_m = 1}^K \mathcal{M}_{k_1 \ldots k_m} \pi_1(k_1) \cdots \pi_m(k_m) + \sum_{k_1, \ldots, k_m = 1}^K 1 \cdot \pi_1(k_1) \cdots \pi_m(k_m) \]
\[ = [\mathcal{M}; \pi_1, \ldots, \pi_m] + 1. \]

Therefore, we are left to show for \( m > 1 \)

\[ [\mathcal{M}; \pi_1, \ldots, \pi_m] = x^{(m)} + z^{(m)}. \]

We prove it by induction. When \( m = 2 \), \( \mathcal{M} \in \mathbb{R}^{K \times K} \). By definitions and elementary algebra,

\[ [\mathcal{M}; \pi_1, \pi_2] = \pi_1^T \mathcal{M} \pi_2 \]
\[ = h' \mathcal{M} h + y_1' \mathcal{M} y_2 + h' \mathcal{M} y_2 + y_1' \mathcal{M} y_2 \]
\[ = [\mathcal{M}; h, h] + [\mathcal{M}; y_1, h] + [\mathcal{M}; h, y_2] + [\mathcal{M}; y_1, y_2]. \]
Hence, the claim holds for \( m = 2 \).

Assume that for \( m = r \), the claim holds. Note that for each \( k_{r+1} \in \{1, \ldots, K\} \), \( \{\mathcal{M}_{k_1 \ldots k_r k_{r+1}} : 1 \leq k_1, \ldots, k_r \leq K\} \) forms a \( r \)-way symmetric tensor of \( K \) dimensions. It follows that

\[
[\mathcal{M}; \pi_1, \ldots, \pi_{r+1}] = [\mathcal{M}; h, \ldots, h, \pi_{r+1}] + \sum_{s=1}^{r} [\mathcal{M}; h, \ldots, h, y_s, h, \ldots, h, \pi_{r+1}]
\]

\[
+ \sum_{s_1=1}^{r-1} \sum_{s_2=s_1+1}^{r} [\mathcal{M}; h, \ldots, h, y_{s_1}, h, \ldots, h, y_{s_2}, \pi_{s_2+1}, \ldots, \pi_{r+1}].
\]

Further, decompose \( \pi_{r+1} \) into \( h + y_{r+1} \). By direct calculations

\[
[\mathcal{M}; \pi_1, \ldots, \pi_r, \pi_{r+1}] = \left( [\mathcal{M}; h, \ldots, h, h] + [\mathcal{M}; h, \ldots, h, y_{r+1}] \right)
\]

\[
+ \left( \sum_{s=1}^{m-1} [\mathcal{M}; h, \ldots, h, y_s, h, \ldots, h] + \sum_{s=1}^{m} [\mathcal{M}; h, \ldots, h, y_s, h, \ldots, h, y_{r+1}] \right)
\]

\[
+ \sum_{s=1}^{r-1} \sum_{s_1=1}^{r} \sum_{s_2=s_1+1}^{r} [\mathcal{M}; h, \ldots, h, y_{s_1}, h, \ldots, h, y_{s_2}, \pi_{s_2+1}, \ldots, \pi_{r+1}]
\]

\[
= [\mathcal{M}; h, \ldots, h] + \sum_{s=1}^{r+1} [\mathcal{M}; h, \ldots, h, y_s, h, \ldots, h]
\]

\[
+ \sum_{s_1=1}^{r} \sum_{s_2=s_1+1}^{r+1} [\mathcal{M}; h, \ldots, h, y_{s_1}, h, \ldots, h, y_{s_2}, \pi_{s_2+1}, \ldots, \pi_{r+1}],
\]

which suggests that the claim also holds for \( m = r + 1 \). By induction, (E.30) is proved.

**E.4 Proof of Lemma E.2**

Introduce \( N_0 = \sum_{i_3, \ldots, i_m \in \{\text{dist}\}} (\theta_{i_3} \cdots \theta_{i_m})^t \) and \( I^{(i)} \) be the shorthand notation for set \( \{1, \ldots, n\} \setminus \{i_3, \ldots, i_m\} \). Here, for convenience, we misuse the superscript \( (i) \) to indicate that this element depends on the choice of \( (i_3, \ldots, i_m) \) whenever it is clear from the context.

By definitions and elementary algebra,

\[
S = \sum_{i_3, \ldots, i_m \in \{\text{dist}\}} \frac{(\theta_{i_3} \cdots \theta_{i_m})^t}{N_0} \sum_{k_3, \ldots, k_m=1}^{K} \prod_{s=3}^{m} u_i(s)(k_s) \bar{w}_i(s)(k_s) \frac{N_0}{a_t}
\]

\[
\times \left[ \sum_{i_1, i_2 \in \{\text{dist}\} \cap I^{(i)}} (\theta_{i_1} \theta_{i_2})^t(y_{i_1}^t \mathcal{M}_{i_1:k_3 \ldots k_m} y_{i_2})(\bar{y}_{i_1}^t \mathcal{M}_{i_2:k_3'} \cdots k_m' \bar{y}_{i_2}) \right].
\]

(E.31)

Let \( \mathcal{M}_{i_1:k_3 \ldots k_m} = \sum_{j=1}^{K} b_j^{(k)} b_j^{(k')} \delta_j^{(k)} \delta_j^{(k')} \) and \( \mathcal{M}_{i_2:k_3' \ldots k_m'} = \sum_{j'=1}^{K} b_j^{(k)} b_j^{(k')} \delta_j^{(k)} \delta_j^{(k')} \) be the eigen-decomposition of the matrices \( \mathcal{M}_{i_1:k_3 \ldots k_m} \) and \( \mathcal{M}_{i_2:k_3' \ldots k_m'} \), respectively. Introduce

\[
X(i, j, j', k, k') = \sum_{i_1, i_2 \in \{\text{dist}\} \cap I^{(i)}} (\theta_{i_1} \theta_{i_2})^t y_{i_1}^t \delta_j^{(k)} \delta_j^{(k')} (y_{i_2}^t \delta_j^{(k)} \delta_j^{(k')}) (\bar{y}_{i_1}^t \delta_{j'}^{(k')} \delta_{j'}^{(k')}) (\bar{y}_{i_2}^t \delta_{j'}^{(k')} \delta_{j'}^{(k')}).
\]

Then we can write

\[
\sum_{i_1, i_2 \in \{\text{dist}\} \cap I^{(i)}} (\theta_{i_1} \theta_{i_2})^t (y_{i_1}^t \mathcal{M}_{i_1:k_3 \ldots k_m} y_{i_2})(\bar{y}_{i_1}^t \mathcal{M}_{i_2:k_3' \ldots k_m'} \bar{y}_{i_2}) = \sum_{j, j'=1}^{K} X(i, j, j', k, k').
\]
Inserting this into (E.31) gives
\[ S = \sum_{i_3, \ldots, i_m \text{ (dist)}} \frac{(\theta_{i_3} \cdots \theta_{i_m})^t}{N_\theta} \sum_{k_3, \ldots, k_m = 1}^K \prod_{s=3}^m w_{i_s}(k_s) \bar{w}_{i_s}(s) (k_s) \sum_{j,j'=1}^K \frac{1}{K^2} \left( \frac{K^2 N_\theta}{a_t} X(i, j, j', k, k') \right). \]

Note that \( \sum_{i_3, \ldots, i_m \text{ (dist)}} \frac{(\theta_{i_3} \cdots \theta_{i_m})^t}{N_\theta} \sum_{k_3, \ldots, k_m = 1}^K \prod_{s=3}^m w_{i_s}(k_s) \bar{w}_{i_s}(s) (k_s) \sum_{j,j'=1}^K \frac{1}{K^2} = 1 \) and that \( \exp(\cdot) \) is convex. By Jensen’s inequality,
\[ \exp(cS) \leq \sum_{i_3, \ldots, i_m \text{ (dist)}} \frac{(\theta_{i_3} \cdots \theta_{i_m})^t}{N_\theta} \sum_{k_3, \ldots, k_m = 1}^K \prod_{s=3}^m w_{i_s}(k_s) \bar{w}_{i_s}(s) (k_s) \sum_{j,j'=1}^K \frac{1}{K^2} \exp\left( \frac{cK^2 N_\theta}{a_t} X(i, j, j', k, k') \right). \]

By assumptions \( w_{i_s}(s), \bar{w}_{i_s}(s) \) are independent of \( y_{i_1}, y_{i_2}, \bar{y}_{i_1}, \bar{y}_{i_2}, \), \( 3 \leq s \leq m \). Taking expectation on both sides gives
\[ \mathbb{E}[\exp(cS)] \leq \sum_{i_3, \ldots, i_m \text{ (dist)}} \frac{(\theta_{i_3} \cdots \theta_{i_m})^t}{N_\theta} \sum_{k_3, \ldots, k_m = 1}^K \prod_{s=3}^m \mathbb{E}[w_{i_s}(k_s)] \mathbb{E}[\bar{w}_{i_s}(s) (k_s)] \sum_{j,j'=1}^K \frac{1}{K^2} \max_{i,j,j',k,k'} \mathbb{E} \left[ \exp\left( \frac{cK^2 N_\theta}{a_t} X(i, j, j', k, k') \right) \right]. \]

Now, to show the claim, note that \( N_\theta := \sum_{i_3, \ldots, i_m \text{ (dist)}} (\theta_{i_3} \cdots \theta_{i_m})^t \leq \|\theta\|_{t}^{(m-2)} \), we are sufficient to show that
\[ X(i, j, j', k, k') \leq C \mu^2 |T| + C \mu^2 \|\theta\|_2^2, \quad \text{for } x > 0. \]

To see this, we rewrite
\[ X(i, j, j', k, k') := \sum_{i_1, i_2 \in I^{(i)}} (1 - \mathbb{I}_{i_1 = i_2}) (\theta_{i_1} \theta_{i_2}) (x_j (k_j (y_{i_1} b_i^{(k)}))(y_{i_2} b_j^{(k)}))(\bar{y}_{i_1} b_i^{(k)}))(\bar{y}_{i_2} b_j^{(k)})), \]
where \( T_1 = \left( \sum_{i_1 \in I^{(i)}} \theta_{i_1} (y_{i_1} b_i^{(k)}))(\bar{y}_{i_1} b_i^{(k)})) \right)^2, \quad T_2 = \sum_{i_1 \in I^{(i)}} \left( \theta_{i_1} (y_{i_1} b_i^{(k)}))(\bar{y}_{i_1} b_i^{(k)})) \right)^2. \]

Consider \( T_2 \) first. Note that \( \max_{i} \{ \|y_{i_1}\|, \|\bar{y}_{i_1}\| \} \leq \sqrt{K} \) and that \( \|\delta_j^{(k)}\| = \|b_j^{(k)}\| = 1, \) \( j, j', k, k' \). By direct calculations
\[ |T_2| \leq (K)^2 \sum_{i_1} \theta_{i_1}^2 \leq C \|\theta\|_{2t}^2. \]

Next, consider \( T_1 \). Let \( Z = \sum_{i_1 \in I^{(i)}} \theta_{i_1} (y_{i_1} b_i^{(k)}))(\bar{y}_{i_1} b_i^{(k)})). \) Note that \( Z \) is a sum of \( n - (m - 2) \) independent random variables with \( |\theta_{i_1} (y_{i_1} b_i^{(k)}))(\bar{y}_{i_1} b_i^{(k)})) | \leq \sqrt{K} \theta_{i_1}^2 \). By Hoeffding’s inequality
\[ \mathbb{P}(|Z| > x) \leq 2 \exp \left( -2x^2/(\sum_{i_1 \in I^{(i)}} (2\sqrt{K} \theta_{i_1}^2)^2) \right), \quad \text{for } x > 0. \]
Combining this with $\sum_{i_t \in I^{(i)}} (2\sqrt{K}^2 \theta_{i_t}^T)^2 \leq 4K^2 \|\theta\|_2^2$ and $T_1 = Z^2$, it follows that
\[
P(|T_1| > x) \leq 2 \exp(-x/(2K^2\|\theta\|_2^2)), \quad \text{for } x > 0.
\] (E.33)

At the same time, recall that $\delta_j^{(k)} \delta_j^{(k')}$ are the eigenvalues of the matrices $M_{k, \ldots, k_m}$ and $M_{i, k_1' \ldots k_m'}$. By the assumption $\|M_{k, \ldots, k_m}\| \leq C\mu$, for $1 \leq k_3, \ldots, k_m \leq K$, max$_{j,k} \{\|\delta_j^{(k)}\|\} \leq C\mu$. It is seen that
\[
X(i, j, j', k, k') := \delta_j^{(k)} \delta_j^{(k')}(T_1 - T_2) \leq C\mu^2 |T_1| + C\mu^2 \|\theta\|_2^2,
\]
with $T_1$ satisfying (E.33).

This shows (E.32) and finishes the proof.

### E.5 Proof of Lemma E.3

Similarly, let $N_\theta = \sum_{i_3, \ldots, i_m (\text{dist})} (\theta_{i_3} \cdots \theta_{i_m})^T$ and $I^{(i)}$ be the shorthand notation for set $\{1, \ldots, n\} \backslash \{i_3, \ldots, i_m\}$. Here, for convenience, we misuse the superscript $(i)$ to indicate that this element depends on the choice of $(i_3, \ldots, i_m)$ whenever it is clear from the context. Let $M_{k, \ldots, k_m} = \sum_{j=1}^K b_j^{(k)} b_j^{(k')}$ and $M_{i, k_1', \ldots, k_m'} = \sum_{j'=1}^K b_j^{(k')} b_j^{(k')}$ be the eigen-decomposition of the matrices $M_{i, k_1' \ldots k_m'}$ and $M_{k_1', \ldots, k_m'}$, respectively. Following the procedures in the proof of Lemma E.2, we can obtain
\[
\exp(cS) \leq \sum_{i_3, \ldots, i_m (\text{dist})} \frac{(\theta_{i_3} \cdots \theta_{i_m})^T}{N_\theta} \sum_{k_1, \ldots, k_m = 1}^K h(k_1') \overline{w}_{i_3}^{(3)} (k_3) \prod_{s=4}^m u_{i_s}^{(s)}(k_s) \overline{w}_{i_s}^{(s)}(k_s) \sum_{j, j'=1}^K \frac{1}{K^2} \exp\left(\frac{cK^2 N_{\theta}}{a_t} X(i, j, j', k, k')\right),
\] (E.34)

where
\[
X(i, j, j', k, k') = \sum_{i_3, i_3 (\text{dist}) \in I^{(i)}} (\theta_{i_3} \theta_{i_3})^T \delta_j^{(k)} \delta_j^{(k')} (y_{i_3}^{(i)} b_j^{(k)}) (y_{i_3}^{(i)} b_j^{(k')}) (\overline{y}_{i_3} b_j^{(k)}) (\overline{y}_{i_3} b_j^{(k')}).
\]

Note that $\overline{w}_{i_3}^{(3)}$ may not be independent of $y_{i_3}$ which exists in $X(i, j, j', k, k')$. Consequently, we cannot directly take expectation on both sides of (E.34) like that in Lemma E.2 to eliminate weight vectors $\{w_{i_j}^{(j)}\}$ by a maximum bound. To resolve this, we first derive a bound on $X(i, j, j', k, k')$ to eliminate $\overline{y}_{i_3}$. We rewrite
\[
X(i, j, j', k, k') := \sum_{i_3, i_3 (\text{dist}) \in I^{(i)}} (1 - \mathbb{I}_{i_3 = i_3}) (\theta_{i_3} \theta_{i_3})^T \delta_j^{(k)} \delta_j^{(k')} (y_{i_3}^{(i)} b_j^{(k)}) (y_{i_3}^{(i)} b_j^{(k')}) (\overline{y}_{i_3} b_j^{(k)}) (\overline{y}_{i_3} b_j^{(k')}),
\]
where
\[
T_1 = \left( \sum_{i_t \in I^{(i)}} \theta_{i_t} (y_{i_t} b_j^{(k)}) (\overline{y}_{i_t} b_j^{(k)}) \right) \left( \sum_{i_t' \in I^{(i)'}} \theta_{i_t'} (y_{i_t'} b_j^{(k)}) \right), \quad T_2 = \left( \sum_{i_t \in I^{(i)}} \theta_{i_t} (y_{i_t} b_j^{(k)}) \right)^2 (\overline{y}_{i_t} b_j^{(k)}).
\]

Recall that $\delta_j^{(k)} \delta_j^{(k')}$ are the eigenvalues of the matrices $M_{i, k_1' \ldots k_m'}$ and $M_{i, k_1' \ldots k_m'}$. By the assumption $\|M_{i, k_1' \ldots k_m'}\| \leq C\mu$, for $1 \leq k_3, \ldots, k_m \leq K$, max$_{j,k} \{\|\delta_j^{(k)}\|\} \leq C\mu$. Combining this with $\|b_j^{(k)}\| = 1$ and $\|y_{i_3}\| \leq \sqrt{K}$, we have
\[
X(i, j, j', k, k') := \delta_j^{(k)} \delta_j^{(k')} (T_1 - T_2) (\overline{y}_{i_3} b_j^{(k)}) \leq C\mu^2 (|T_1| + |T_2|).
\]
Note that $T_1, T_2$ (and so the bound) are independent of $w_i^{(s)}, \bar{w}_i^{(s)}$, $3 \leq s \leq m$. Applying this inequality to the RHS of (E.34) and taking expectation on both sides give

$$
\mathbb{E}[\exp(cS)] \leq \sum_{i_3, \ldots, i_m (\text{dist})} \frac{\langle \theta_{i_3} \cdots \theta_{i_m} \rangle^t}{N_{\theta}} \sum_{k_3, \ldots, k_m = 1}^K h(k'_2) \mathbb{E}[\bar{w}_{i_3}^{(3)}(k_3)] \prod_{j=4}^m \mathbb{E}[w_{i_j}^{(s)}(k_s)] \sum_{j, j' = 1}^K \frac{1}{K^2} \\
\cdot \mathbb{E} \left[ \exp \left( \frac{CN_{\theta}}{a_t} \mu^2(|T_1| + |T_2|) \right) \right] \\
\leq \max_{i, j, j', k, k'} \mathbb{E} \left[ \exp \left( \frac{CN_{\theta}}{a_t} \mu^2(|T_1| + |T_2|) \right) \right].
$$

Now, to show the claim, note that $N_{\theta} := \sum_{i_3, \ldots, i_m (\text{dist})} \langle \theta_{i_3} \cdots \theta_{i_m} \rangle^t \leq \|\theta\|_{t(t-2)}^t$, it is then sufficient to show that

$$(I) : \mathbb{P}(|T_1| > x) \leq 4 \exp(-x/(2K^2\|\theta\|_{2t}^2)), \quad \forall x > 0, \quad (II) : |T_2| \leq C\|\theta\|_{2t}^2, \quad (E.35)$$

Consider (I) first. Let $Z_1 = \sum_{i \in I^{(i)}} \theta_i^{(k_i)}(y_i^{(k_i)})(\bar{y}_i^{(k_i)})$, $Z_2 = \sum_{i \in I^{(i)}} \theta_i^{(k_i)}(y_i^{(k_i)})$ and so $T_1 = Z_1 \cdot Z_2$. Note that $Z_1$ and $Z_2$ are the sum of $n - (m - 2)$ independent random variables. Similarly, by Hoeffding’s inequality, for any $x > 0$

$$
\mathbb{P}(|Z_1| > x) \leq 2 \exp(-2x/(2K^2\|\theta\|_{2t}^2)), \quad \mathbb{P}(|Z_2| > x) \leq 2 \exp(-2x/(2\sqrt{K})^2\|\theta\|_{2t}^2)).
$$

Combining this with $|T_1| = |Z_1| \cdot |Z_2|$ and union bound $\mathbb{P}(|Z_1| \cdot |Z_2| > x) \leq \mathbb{P}(|Z_1| > \sqrt{x}) + \mathbb{P}(|Z_2| > \sqrt{x})$, we get

$$
\mathbb{P}(|T_1| > x) \leq 2 \exp(-x/(2K^2\|\theta\|_{2t}^2)) + 2 \exp(-x/(2K\|\theta\|_{2t}^2)) \leq 4 \exp(-x/(2K^2\|\theta\|_{2t}^2)),$$

which proves the first claim in (E.35).

Next, consider (II) in (E.35). By max$_i \|y_i\|, \|\bar{y}_i\| \leq \sqrt{K}, \|\theta_j^{(k)}\| = \|b_j^{(k)}\| = 1$, $\forall j, j', k, k'$

$$
|T_2| := \sum_{i \in I^{(i)}} \left( \theta_i^{(k_i)}(y_i^{(k_i)})(\bar{y}_i^{(k_i)}) \right)^2 \leq \sum_{i} \theta_i^{2t}(\sqrt{\bar{K}^k})^2 \sqrt{\bar{K}} \leq C\|\theta\|_{2t}^2,
$$

which proves (II) and finishes the whole proof.

### E.6 Proof of Lemma E.4

The proof is similar to that in Lemma E.3. Similarly, let $N_{\theta} = \sum_{i_3, \ldots, i_m (\text{dist})} \langle \theta_{i_3} \cdots \theta_{i_m} \rangle^t$ and $I^{(i)}$ be the shorthand notation for the set $\{1, \ldots, n\} \setminus \{i_3, \ldots, i_m\}$. Here, for convenience, we misuse the superscript $(i)$ to indicate that this element depends on the choice of $(i_3, \ldots, i_m)$ whenever it is clear from the context. Let $M_{i_3 \cdots i_m} = \sum_{j=1}^K \delta_j^{(k_i)}(b_j^{(k_i)})^t$, and $M_{k_3 \cdots k_m} = \sum_{j=1}^K \delta_j^{(k_i)}(b_j^{(k_i)})^t$ be the eigen-decomposition of the matrices $M_{i_3 \cdots i_m}$ and $M_{k_3 \cdots k_m}$, respectively. Following the procedures in the proof of Lemma E.2, we can obtain

$$
\exp(cS) \leq \sum_{i_3, \ldots, i_m (\text{dist})} \frac{\langle \theta_{i_3} \cdots \theta_{i_m} \rangle^t}{N_{\theta}} \sum_{k_3, \ldots, k_m = 1}^K h(k'_2) h(k'_2) \prod_{s=3}^m w_{i_s}^{(s)}(k_s) \bar{w}_{i_s}^{(s)}(k_s) \sum_{j, j' = 1}^K \frac{1}{K^2} \cdot \exp \left( \frac{cK^2 N_{\theta}}{a_t} X(i, j, j', k, k') \right),
$$

(E.36)
where
\[ X(i, j, j', k, k') = \sum_{i_1, i_2(d_{i_1} i_2) \in I^{(i)}} (\theta_{i_1} \theta_{i_2})^t \delta_j^{(k)} \delta_{j'}^{(k')} (y_{i_1 j}^b(b_j^{(k)}))(y_{i_2 j'}^b(b_{j'}^{(k')}))(y_{i_3 j}^b(b_j^{(k)}))(y_{i_4 j'}^b(b_{j'}^{(k')})). \]

Note that \( w_{i_3}^{(3)} \) and \( w_{i_4}^{(4)} \) may not be independent of \( y_{i_3} \) and \( y_{i_4} \) which exist in \( X(i, j, j', k, k') \). Similar to the proof of Lemma E.3, we rewrite
\[ X(i, j, j', k, k') := \sum_{i_1, i_2 \in I^{(i)}} (1 - \mathbb{I}_{\{i_1 = i_2\}}) (\theta_{i_1} \theta_{i_2})^t \delta_j^{(k)} \delta_{j'}^{(k')} (y_{i_1 j}^b(b_j^{(k)}))(y_{i_2 j'}^b(b_{j'}^{(k')}))(y_{i_3 j}^b(b_j^{(k)}))(y_{i_4 j'}^b(b_{j'}^{(k')})) \]
\[ = \delta_j^{(k)} \delta_{j'}^{(k')} (T_1 - T_2)(y_{i_3}^b(b_j^{(k)}))(y_{i_4}^b(b_{j'}^{(k')})), \]
where
\[ T_1 = \left( \sum_{i_1 \in I^{(i)}} \theta_{i_1}^t (y_{i_1 j}^b(b_j^{(k)}))^2 \right), \quad T_2 = \sum_{i_1 \in I^{(i)}} \left( \theta_{i_1}^t (y_{i_1 j}^b(b_j^{(k)}))^2 \right). \]

Recall that \( \delta_j^{(k)} \delta_{j'}^{(k')} \) are the eigenvalues of the matrices \( M_{3,k_3,\cdots,k_m} \) and \( M_{k_3',k_4',\cdots,k_m'} \). By the assumption \( \|M_{3,k_3,\cdots,k_m}\| \leq C\mu \), for \( 1 \leq k_3, \ldots, k_m \leq K \), \( \max_{j,k}\{|\delta_j^{(k)}|\} \leq C\mu \). Combining this with \( \|b_j^{(k')}\| = 1 \) and \( \|y_{i_3}\| \leq \sqrt{K} \), we have
\[ X(i, j, j', k, k') := \delta_j^{(k)} \delta_{j'}^{(k')} (T_1 - T_2)(y_{i_3}^b(b_j^{(k)}))(y_{i_4}^b(b_{j'}^{(k')})) \leq C\mu^2 (|T_1| + |T_2|). \]

Note that \( T_1, T_2 \) (and so the bound) are independent of \( w_{i_3}^{(s)}, w_{i_4}^{(s)} \), \( 3 \leq s \leq m \). Applying this inequality to the RHS of (E.36) and taking expectation on both sides give
\[ \mathbb{E}[\exp(cS)] \leq \max_{i,j,j',k,k'} \mathbb{E}\left[ \exp\left( \frac{CN_\theta}{\delta_i^2} \|\theta\|^2_{\mathbb{I}^{t(m-2)}} \right) \right]. \]

Now, to show the claim, note that \( N_\theta := \sum_{i_3, \ldots, i_m(d_{i_3} \cdots d_{i_m})^t} (\theta_{i_3} \cdots \theta_{i_m})^t \leq \|\theta\|_{\mathbb{I}^t}^{t(m-2)} \), it is then sufficient to show that
\[ (I) : \mathbb{P}(|T_1| > x) \leq 2 \exp(-x/(2K\|\theta\|_{\mathbb{I}^{2t}}^2)), \quad \forall x > 0, \quad (II) : |T_2| \leq C\|\theta\|_{\mathbb{I}^{2t}}^2. \]

The procedures to show them are the same as that in the proof of Lemma E.2. So we omit them.

### E.7 Proof of Lemma E.5

The following lemma is used in this proof and we prove it below.

**Lemma E.6** (Each element of community structure tensor is close to one). *Using the same notations of Theorem 3.1, for each \( m \in \{2, \ldots, M\}, \)
\[ \max_{1 \leq i_1, \ldots, i_m \leq K} \{ |P_{i_1, \ldots, i_m}^{(m)} - 1| \} \propto |\mu_2^{(m)}|, \quad (E.37) \]

Fix \( m \), for simplicity of notation, we remove the superscript \( (m) \) whenever it is clear from the context. Recall that \( D = \text{diag}(d_1, \cdots, d_K) \) and \( h = \mathbb{E}[D^{-1} \pi_i/\|D^{-1} \pi_i\|] \). Write for short \( s = \sum_{k=1}^{K} d_k\pi_k \) and \( v = (d_1, \ldots, d_K)' \). With these notations and direct calculations, for \( 1 \leq k_3, \ldots, k_m \leq K \)
\[ M_{k_3, \ldots, k_m} = D(P_{k_3, \ldots, k_m} - 1_K 1_K') D^{-1} \big( \prod_{j=3}^{m} d_k \big) + \left( \prod_{j=3}^{m} d_k - s^{m-2}\right) v v' \big( s^{m-2} v v' - 1_K 1_K' \big). \]
Therefore, to prove the first claim of this lemma, by elementary algebra, it is sufficient to show that

\[(a): \max_{1 \leq k_1, \ldots, k_m \leq K} \left\{ |P_{k_1 \ldots k_m} - 1| \right\} \leq C|\mu_2|, \]
\[(b): \max_{1 \leq k \leq K} \left\{ d_k \right\} \leq C, \]
\[(c): \max_{1 \leq i, j \leq K} \left\{ |(s^{m-2}v^v - 1_K 1'_K)_{ij}| \right\} \leq C|\mu_2|, \]
\[(d): \max_{1 \leq k \leq K} \left\{ |d_k - s| \right\} \leq C|\mu_2|, \]

where we note that \((a)\) is implied by Lemma E.6.

Consider \((b)\). Recall that by degree matching

\[
\sum_{k_2, \ldots, k_m = 1}^{K} DP_{k_2 \ldots k_m} \prod_{j=2}^{m} (d_{k_j} h_{k_j}) = 1_K. \tag{E.38}
\]

Note that each element of \(P\) is non-negative and \(P_{k_1 \ldots k_1} = 1\) for \(1 \leq k_1 \leq K\). It follows that

\[d_{k_1}(d_{k_1} h_{k_1})^{m-1} \leq \sum_{k_2, \ldots, k_m = 1}^{K} d_{k_1} P_{k_1 \ldots k_m} \prod_{j=2}^{m} (d_{k_j} h_{k_j}) = 1, \quad 1 \leq k_1 \leq K. \]

Combining this with our assumption \(\min_{1 \leq k \leq K} \{h_k\} \geq C,\)

\[d_k \leq h_k^{-(m-1)/m} \leq C, \quad 1 \leq k \leq K, \tag{E.39}\]

which proves \((b)\).

Next consider \((c)\). Let \(H\) be a tensor defined by \(H_{k_1 \ldots k_m} = P_{k_1 \ldots k_m} - 1,\) for \(1 \leq k_1, \ldots, k_m \leq K\) and introduce \(w\) as the vector \(\sum_{k_2, \ldots, k_m = 1}^{K} DH_{k_2 \ldots k_m} \prod_{j=2}^{m} (d_{k_j} h_{k_j})\). Recall that \(s = \sum_{k=1}^{K} d_k h_k\). By definitions and calculations, \((E.38)\) can be written as

\[w + s^{m-1} v = 1_K. \tag{E.40}\]

Note that \(h'v = s\). Left multiplying \(h'\) on both sides gives

\[h'w + s^m = 1. \tag{E.41}\]

At the same time, inserting \((E.40)\) into \(s^{m-2}v^v - 1_K 1'_K\) through \(1_K\) gives

\[s^{m-2}v^v - 1_K 1'_K = s^{m-2}v^v - (w + s^{-1} v)(w + s^{-1} v)' \]
\[= s^{m-2}(1 - s^m)v^v - s^{-1} wv' - s^{-1} wv' - wv'. \]

Note that by \((E.41), 1 - s^m = h'w.\) It follows that

\[s^{m-2}v^v - 1_K 1'_K = s^{m-2}h'wv' - s^{-1} wv' - s^{-1} wv' - wv'. \]

By \((E.39), \max_{1 \leq k \leq K} \{h_k\} \leq 1\) and elementary algebra

\[\max_{1 \leq i, j \leq K} \left\{ |(s^{m-2}v^v - 1_K 1'_K)_{ij}| \right\} \leq C\|h\|_{\text{max}} \cdot \|v\|_{\text{max}} \cdot \|w\|_{\text{max}} \leq C\|H\|_{\text{max}}, \]

where \(\cdot\|_{\text{max}}\) is the element-wise maximum norm and \(\|H\|_{\text{max}} := \max_{k_1, \ldots, k_m} \{|P_{k_1, \ldots, k_m} - 1|\} \leq C|\mu_2|\). This proves \((c)\).

On the other hand, by elementary algebra, \(|(s^{m-2}v^v - 1_K 1'_K)_{ii}| \leq \|s^{m-2}v^v - 1_K 1'_K\|,\) for all \(1 \leq i \leq K\) and so

\[|s^{m-2} d_i d_i - 1| \leq C|\mu_2|. \]
Transforming the above formula gives,
\[ d_i = s^{-(m-2)/2} + O(\|\mu_2\|). \] (E.42)

Summing up with weight \( h_i \) in terms of \( i \) on two sides and noting that \( \sum_i h_i = 1 \), it gives
\[ s = s^{-(m-2)/2} + O(\|\mu_2\|). \] (E.43)

Combining this with (E.42) gives (d).

Next we consider the second claim of this lemma i.e. \( \max_{1 \leq i \leq K} \{|d_i| - 1\} \leq C|\mu_2| \). By elementary algebra, (E.43) can be rewritten as
\[ s = 1 + \frac{\sqrt{s^{m-1}} + \sqrt{s^{m-2}}}{\sum_{j=0}^{m-2} \sqrt{s^j}} \cdot O(\|\mu_2\|), \]
where we note that \( \frac{\sqrt{s^{m-1}} + \sqrt{s^{m-2}}}{\sum_{j=0}^{m-2} \sqrt{s^j}} \leq 1 \). Combining this with (E.42) proves the second claim.

\section*{E.8 Proof of Lemma E.6}

Since the claim is argued for each \( m \)-uniform tensor \( \mathcal{P}^{(m)} \) separately, fixing \( m \), we remove the superscript \( (m) \) whenever it is clear from the context.

Let the \( K \times K^{m-1} \) matrix \( P \) denote the matricization of \( \mathcal{P}^{(m)} \). Let \( U \Sigma V' \) be the SVD of \( P \), where \( U = (u_1, \ldots, u_K) \), \( V = (v_1, \ldots, v_{K^{m-1}}) \) and \( \Sigma = (\text{diag}(\mu_1, \ldots, \mu_K), 0_K \times (K^{m-1} - K)) \).

To show that claim, it is sufficient to show that
\[ |\mu_2| \leq \max_{1 \leq i_1, \ldots, i_m \leq K} \{|P_{i_1 \cdots i_m} - 1|\}, \quad (I) : \max_{1 \leq i_1, \ldots, i_m \leq K} \{|P_{i_1 \cdots i_m} - 1|\} \leq C|\mu_2|. \]

Consider (I) first. Let \( P_0 \) be the \( K \times K^{m-1} \) matrix of ones. Recall that \( \mu_2 \) is the second singular value of \( P \), and note that the second singular value of \( P_0 \) is 0. By [4, Corollary 7.3.5, Page 451],
\[ |\mu_2| \leq \|P - P_0\|. \]
At the same time, by elementary algebra, \( \|P - P_0\| \leq C\max_{1 \leq i_1, \ldots, i_m \leq K} \{|P_{i_1 \cdots i_m} - 1|\} \). Combining these proves (I).

Next we consider (II).

By our assumption \( \|P\| \leq C \) and elementary algebra,
\[ \max_{1 \leq i_1, \ldots, i_m \leq K} \{|P_{i_1 \cdots i_m}|\} = \|P\|_{\text{max}} \leq \|P\| \leq C, \]
where \( \|\cdot\|_{\text{max}} \) is the element-wise maximum norm. Therefore, (II) directly holds for the case that \( |\mu_2| \geq \epsilon \) for some positive constants \( \epsilon < 1 \). It is then sufficient to consider the case when \( |\mu_2| < \epsilon \).

By definitions,
\[ (PP')_{ii} \geq P_{i \cdots i}^2 = 1, \quad (PP')_{ij} \geq 0, \quad 1 \leq i, j \leq K. \]

Therefore, by Perron’s theorem [4], the first eigenvalue (in magnitude) and each entry of the first eigenvector of \( PP' \) are positive. Note that \( PP' = U \Sigma^2 U' \), it follows that
\[ \mu_1 > 0, \quad u_1(i) > 0, \quad 1 \leq i \leq K. \]

Let \( a = u_1(i)^{\frac{1}{m}} \) and \( b = v_1(i)^{\frac{1}{K^{m-1}}} \) be the scaled version of \( u_1 \) and \( v_1 \), where \( a_i > 0 \) since \( u_1(i) > 0, 1 \leq i \leq K \). Introduce \( \tilde{P} = ab' \). For simplicity, we misuse the notation \( b_{i_2 \cdots i_m} \).
for $b_{i_2} + \sum_{m=3}^{\infty} x^{m-2} (i_3 - 1)$. To show (II), by triangle inequality, it is sufficient to show that for $1 \leq i_1, \ldots, i_m \leq K$,

$\text{(IIa)}: |P_{i_1 \ldots i_m} - a_{i_1} b_{i_2 \ldots i_m}| \leq C |\mu_2|, \quad \text{(IIb)}: |a_{i_1} b_{i_2 \ldots i_m} - 1| \leq C |\mu_2|.$

Note that by elementary algebra

$$|P_{i_1 \ldots i_m} - a_{i_1} b_{i_2 \ldots i_m}| \leq \|P - \tilde{P}\|_{\max} \leq \|P - \tilde{P}\| = |\mu_2|,$$

(E.44)

This proves (IIa).

It is left to show (IIb). We start by showing that $a$ is a vector with elements are almost the same. By equality $x^m - y^m = (x - y) \sum_{j=0}^{m-1} x^{m-1-j} y^j$, we have,

$$|a_{i_1} - a_{i_2}| = \frac{|a_{i_1}^m - a_{i_2}^m|}{\sum_{j=0}^{m-1} a_{i_1}^{m-j-1} a_{i_2}^j} = \frac{|a_{i_1} / a_{i_2} - (a_{i_1} / a_{i_2})^{m-1}|}{\sum_{j=0}^{m-1} a_{i_1}^{-j} a_{i_2}^j}, \quad 1 \leq i_1, i_2 \leq K.
$$

Combining this with triangle's inequality $|a_{i_1} / a_{i_2} - (a_{i_2} / a_{i_1})^{m-1}| \leq |a_{i_1} b_{i_2 \ldots i_m} - a_{i_1} / a_{i_2}| + |a_{i_1} b_{i_2 \ldots i_m} - (a_{i_2} / a_{i_1})^{m-1}|$,

$$|a_{i_1} - a_{i_2}| \leq \frac{|a_{i_1} b_{i_2 \ldots i_m} - a_{i_1} / a_{i_2}| + |a_{i_1} b_{i_2 \ldots i_m} - (a_{i_2} / a_{i_1})^{m-1}|}{\sum_{j=0}^{m-1} a_{i_1}^{-j} a_{i_2}^j}, \quad 1 \leq i_1, i_2 \leq K. \quad (E.45)
$$

We claim that for $1 \leq k \leq m$ the following holds and prove it later.

$$|a_{i_1} b_{i_2 \ldots i_k \ldots i_1} - \frac{\prod_{j=1}^{k} a_{i_j}}{a_{i_1}^{k-1}}| \leq \left(2 \sum_{s=2}^{k} \prod_{j=s+1}^{k} a_{i_j} + \prod_{j=1}^{k} a_{i_j} / a_{i_1}^{k-1}\right) |\mu_2|, \quad 1 \leq i_1, \ldots, i_m \leq K. \quad (E.46)
$$

By setting $k = m; i_3, \ldots, i_m = i_2$ and $k = 1, i_1 = i_2$ separately in the above inequality, we obtain

$$|a_{i_1} b_{i_2 \ldots i_m} - \frac{a_{i_2}^{m-1}}{a_{i_1}^{m-1}}| \leq \left(2 \sum_{s=2}^{m} \frac{a_{i_2}^{m-s}}{a_{i_1}^{m-s}} + \frac{a_{i_2}^{m-1}}{a_{i_1}^{m-1}}\right) |\mu_2|, \quad |a_{i_1} b_{i_2 \ldots i_m} - \frac{a_{i_1}}{a_{i_2}}| \leq \frac{a_{i_1}}{a_{i_2}} |\mu_2|.
$$

Inserting the above into the RHS of (E.45) and by direct calculations

$$|a_{i_1} - a_{i_2}| \leq \frac{1}{\sum_{j=0}^{m-1} a_{i_1}^{-j} a_{i_2}^j} \left(2 \sum_{s=2}^{m} \frac{a_{i_2}^{m-s}}{a_{i_1}^{m-s}} |\mu_2| + \frac{a_{i_2}^{m-1}}{a_{i_1}^{m-1}} |\mu_2| + \frac{a_{i_1}}{a_{i_2}} |\mu_2|\right) = (a_{i_1} + a_{i_2}) |\mu_2|.
$$

Combining this inequality with $\sum_{j=1}^{K} |a_i - a_j| \leq \sum_{j=1}^{K} a_j \leq \sum_{j=1}^{K} (a_i + |a_i - a_j|)$ give

$$\sum_{i_2=1}^{K} \left(a_{i_1} - (a_{i_1} + a_{i_2}) |\mu_2|\right) \leq \sum_{i_2=1}^{K} a_{i_2} \leq \sum_{i_2=1}^{K} \left(a_{i_1} + (a_{i_1} + a_{i_2}) |\mu_2|\right).
$$

By $\sum_{i_2=1}^{K} a_{i_2} = \|a\|_1$, we can rewrite it as

$$\frac{\|a\|_1}{K} 1 - |\mu_2| \leq a_{i_1} \leq \frac{\|a\|_1}{K} 1 + |\mu_2|.
$$

Note that $|\mu_2| < \epsilon < 1$, it is seen that

$$a_{i_1} = \frac{\|a\|_1}{K} (1 + O(|\mu_2|)), \quad 1 \leq i_1 \leq K. \quad (E.47)
$$

Now we are ready to show (IIb). By triangle inequality

$$|a_{i_1} b_{i_2 \ldots i_m} - 1| \leq |a_{i_1} b_{i_2 \ldots i_m} - \frac{\prod_{j=1}^{m} a_{i_j}}{a_{i_1}^{m-1}}| + |\prod_{j=1}^{m} a_{i_j} / a_{i_1}^{m-1} - 1|. \quad (E.48)$$
Note that setting $k = m$ in (E.46) gives
\[
\left| a_{i_1} b_{i_2} \cdots i_m - \frac{\prod_{j=1}^m a_{i_j}}{a_{i_{i_1}}^m} \right| \leq \left( 2 \sum_{s=2}^m \frac{\prod_{j=s+1}^m a_{i_j}}{a_{i_{i_1}}^m} + \frac{\prod_{j=1}^m a_{i_j}}{a_{i_{i_1}}^m} \right) |\mu_2|.
\]
Inserting this into (E.48). By direct calculations and (E.47)
\[
|a_{i_1} b_{i_2} \cdots i_m - 1| \leq \left( 2 \sum_{s=2}^m \frac{\prod_{j=s+1}^m a_{i_j}}{a_{i_{i_1}}^m} + \frac{\prod_{j=1}^m a_{i_j}}{a_{i_{i_1}}^m} \right) |\mu_2| + \left| \frac{\prod_{j=1}^m a_{i_j}}{a_{i_{i_1}}^m} - 1 \right| = O(|\mu_2|).
\]
which holds proves (IIb) and finishes the main proof of this lemma.

Lastly, we prove the claim (E.46), which is done by induction. Consider $k = 1$, the goal is to show
\[
|a_{i_1} b_{i_1} \cdots i_1| \leq |\mu_2|,
\]
which is exactly (E.49) and so the claim (E.46) holds for $k = 1$.

Now, assume that the claim holds for $k = k_0$ and the goal is to show that this implies that the claim holds for $k = k_0 + 1$. By triangle’s inequality,
\[
\left| a_{i_1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1 - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_{i_1}}^{k_0+1}} \right| \\
\leq |a_{i_1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1 - P_{i_1 \cdots i_{k_0+1} i_{i_1} \cdots i_1} - P_{i_{k_0+1} i_{i_1} \cdots i_{k_0+1} i_{i_1} \cdots i_1}| \\
+ |P_{i_{k_0+1} i_{i_1} \cdots i_{k_0+1} i_{i_1} \cdots i_1} - a_{i_{k_0+1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1}| + |a_{i_{k_0+1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1} - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_{i_1}}^{k_0+1}}|.
\]
By (E.44), the first term and the third is bounded by $|\mu_2|$. Also, by symmetry of $P$, the second term is 0. Moving a factor $a_{i_{k_0+1}}/a_{i_1}$ from the last term, it follows that
\[
\left| a_{i_1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1 - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_{i_1}}^{k_0+1}} \right| \leq 2|\mu_2| + \frac{a_{i_{k_0+1}}}{a_{i_1}} |a_{i_1} b_{i_2} \cdots i_{k_0+1} i_{i_1} \cdots i_1| - \frac{\prod_{j=1}^{k_0} a_{i_j}}{a_{i_{i_1}}^{k_0}}
\]
(By the assumption for $k = k_0$) \leq 2|\mu_2| + \frac{a_{i_{k_0+1}}}{a_{i_1}} \left( 2 \sum_{s=2}^{k_0} \frac{\prod_{j=s+1}^{k_0} a_{i_j}}{a_{i_{i_1}}^{k_0-s}} + \frac{\prod_{j=1}^{k_0} a_{i_j}}{a_{i_{i_1}}^{k_0}} \right) |\mu_2|
\]
\[
= \left( 2 \sum_{s=2}^{k_0+1} \frac{\prod_{j=s+1}^{k_0+1} a_{i_j}}{a_{i_{i_1}}^{k_0+1-s}} + \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_{i_1}}^{k_0+1}} \right) |\mu_2|,
\]
which shows (E.46) also holds for $k = k_0 + 1$. Hence, by induction, (E.46) holds for $1 \leq k \leq m$.

**F  Proof of Lemma 2.2**

We have the following lemma which is used in the proof of Lemma 2.2 and prove it below.

**Lemma F.1.** Under the conditions of Lemma 2.2, as $n \to \infty$, with probability at least $1 - O(1/n)$,
\[ \bullet \ (a) \ Under \ both \ the \ null \ and \ the \ alternative, \ |\hat{\alpha}_n - \tilde{\alpha}_n| \leq C \log(n) (\tilde{\alpha}_n/n^3)^{1/2}. \]
\[ \bullet \ (b) \ Under \ the \ alternative, \ \tilde{\alpha}_n \leq \max_{1 \leq k_1, k_2, k_3 \leq K} \{ P_{k_1 k_2 k_3} \} \leq C \tilde{\alpha}_n \text{ and } \hat{\alpha}_n = h'(P_0) h + O(\tilde{\alpha}_n^2). \]
Similarly, we have
\[ S \]
\[
\text{Recall that } M, \quad x \]
\[
\text{Consider } (i, 1, 2, 3) \text{ first. Write } (I). \]
\[
\sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) = \sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) \]
\[
= (I) + (II), \]
where
\[
(I) = \sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3), \quad (II) = \sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) \]
\[
\text{Consider } (I) \text{ first. Write } (Ia), (Ib), \quad \text{where}
\]
\[
(Ia) = \sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3), \quad (Ib) = \sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) \]
\[
\text{Now, by direct calculations,}
\]
\[
(Ib) = (\alpha_n - \bar{\alpha}_n) \sum_{x \in \mathcal{S}} (\mathcal{A}^*_x(i, 1, 2, 3) - \bar{\mathcal{A}}^*_x(i, 1, 2, 3)). \]
\[
\text{Note that for each tuple } (i_1, i_2, i_3), \text{ there are } \binom{n-1}{2} - 1 \text{ different } x = (i_1, i_2, i_3, i_4, i_5) \text{ in } \mathcal{S} \text{ with the same } (i_1, i_2, i_3). \text{ It follows}
\]
\[
\sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) = \binom{n-1}{2} - 1 \sum_{i_1, i_2, i_3 (\text{dist})} \mathcal{A}^*_x(i, 1, 2, 3) = \frac{n^2(n-1)(n-2)(n-3)}{4} \alpha_n. \]
\[
\text{Similarly, we have}
\]
\[
\sum_{x \in \mathcal{S}} \mathcal{A}^*_x(i, 1, 2, 3) = \frac{n^2(n-1)(n-2)(n-3)}{4} \alpha_n. \]
\[
\text{Inserting (F.55)-(F.56) into (F.54) gives}
\]
\[
(Ib) = -\frac{n^2(n-1)(n-2)(n-3)}{4} (\alpha_n - \bar{\alpha}_n)^2.
\]
Combining this with (F.53) gives
\[
(I) = (Ia) - \frac{n^2(n-1)(n-2)(n-3)}{4} (\hat{\alpha}_n - \hat{\alpha}_n)^2. \tag{F.57}
\]
Next consider (II). Note that for any \( x \in S_0 \setminus S \), \( i_1 < i_2 \) and \((i_1, i_2) = (i_4, i_5)\). By direct calculations
\[
\sum_{x \in S_0 \setminus S} A_{i_1i_2i_3}^* A_{i_3i_4i_5} = \frac{1}{2} \sum_{i_1, i_2, i_3 (\text{dist})} (A_{i_1i_2i_3}^*)^2 = \frac{1}{2} \sum_{i_1, i_2, i_3 (\text{dist})} (A_{i_1i_2i_3}^2 - 2\hat{\alpha}_n A_{i_1i_2i_3} + \hat{\alpha}_n^2). \tag{F.58}
\]
Since \( A_{i_1i_2i_3} \in \{0, 1\} \), we have \( A_{i_1i_2i_3}^2 = A_{i_1i_2i_3} \). Combining this with definitions, the RHS of (F.58) reduces to
\[
\frac{n(n-1)(n-2)}{2} \hat{\alpha}_n (1 - \hat{\alpha}_n).
\]
It follows that
\[
(II) = 0. \tag{F.59}
\]
Combining (F.52), (F.57), and (F.59), it follows from (F.51) that
\[
\psi_n = \frac{(Ia) - (1/4)n^2(n-1)(n-2)(n-3)(\hat{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n(n-1)}\hat{\alpha}_n (1 - \hat{\alpha}_n)}.
\]
Now, by Lemma F.1, \( |\hat{\alpha}_n - \hat{\alpha}_n| \leq C \log(n)/\hat{\alpha}_n n^{3/2} \) except for a probability of \( 1 - O(1/n) \). It is seen that except for a probability of \( 1 - O(1/n) \)
\[
\left| \frac{\hat{\alpha}_n}{\alpha_n} - 1 \right| \leq C \frac{\log(n)}{\sqrt{n^3 \alpha_n}} , \quad \left| \frac{(1/4)n^2(n-1)(n-2)(n-3)(\hat{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n(n-1)}\hat{\alpha}_n (1 - \hat{\alpha}_n)} \right| \leq C \frac{\log^2(n)}{n^{1/2}} .
\]
By \( n^2 \hat{\alpha}_n \to \infty \), we have that in probability,
\[
\frac{\hat{\alpha}_n}{\alpha_n} \to 1 , \quad \frac{(1/4)n^2(n-1)(n-2)(n-3)(\hat{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n(n-1)}\hat{\alpha}_n (1 - \hat{\alpha}_n)} \to 0.
\]
Let
\[
Z_n = \frac{(Ia)}{\sqrt{2n(n-1)}\hat{\alpha}_n (1 - \hat{\alpha}_n)}.
\]
To show (F.60), it is sufficient to show that as \( n \to \infty \),
\[
Z_n \to N(0, 1), \quad \text{under the null}, \tag{F.60}
\]
and
\[
\PP(|Z_n| > M) \to 1 \text{ for any } M > 0, \quad \text{under the alternative}. \tag{F.61}
\]
We now show (F.60)-(F.61). We consider (F.61) first since the proof is shorter. The following lemma is proved below.

**Lemma F.2.** Under the conditions of Lemma 2.2, if the alternative hypothesis is true, then as \( n \to \infty \)
\[
\mathbb{E}[Z_n] \geq C n^{2.5} \hat{\alpha}_n \delta_n^2, \quad \text{Var}(Z_n) \leq C n^2 \hat{\alpha}_n .
\]
Now, suppose the alternative hypothesis is true. Note that by triangle inequality
\[
\PP(|Z_n| \leq M) \leq \PP \left( \left| \mathbb{E}[Z_n] \right| - |Z_n - \mathbb{E}[Z_n]| \leq M \right) = \mathbb{P}(M - \mathbb{E}[Z_n] \geq |\mathbb{E}[Z_n]| - M),
\]
and
where by Chebyshev’s inequality,
\[
P(|Z_n - \mathbb{E}[Z_n]| \geq |\mathbb{E}[Z_n] - M|) \leq \frac{\text{Var}(Z_n)}{(\mathbb{E}|Z_n| - M)^2}.
\]

At the same time, by Lemma F.2 and our assumptions of \(n^2\bar{\alpha}_n \to \infty\) and \(n^{3/2}\bar{\alpha}_n^{1/2}\delta_n^2 \to \infty\),
\[
\frac{\text{Var}(Z_n)}{(\mathbb{E}|Z_n| - M)^2} \leq \frac{Cn^2\bar{\alpha}_n}{(Cn^{2.5}\bar{\alpha}_n\delta_n^2 - M)^2} \leq \frac{1}{C(n^{3/2}\bar{\alpha}_n^{1/2}\delta_n^2)^2} \to 0.
\]

Combining these proves (F.61).

We now consider (F.60). For \(1 \leq m \leq n\), introduce a subset of \(S\) by
\[
S^{(m)} = \{x = (i_1, i_2, i_3, i_4, i_5) \in S : \max\{i_1, i_2, i_3, i_4, i_5\} \leq m\}.
\]
Introduce
\[
\bar{T}_{n,m} = \sum_{x \in S^{(m)}} \bar{A}_{i_1i_2i_3} \bar{A}_{i_3i_4i_5}, \quad Z_{n,m} = \frac{\bar{T}_{n,m}}{\sqrt{2n(n-1)\bar{\alpha}_n(1-\bar{\alpha}_n)}}, \quad (\bar{T}_{n,0} = Z_{n,0} = 0),
\]
and
\[
X_{n,m} = Z_{n,m} - Z_{n,m-1}.
\]
It is seen that
\[
(Ia) = \bar{T}_{n,n}, \quad \text{and} \quad Z_n = Z_{n,n} = \sum_{m=1}^{n} X_{n,m}, \quad (\text{F.62})
\]
Consider the filtration \(\mathcal{F}_{n,m}\) \(1 \leq m \leq n\) with \(\mathcal{F}_{n,m} = \sigma\left(\{\bar{A}_{i_1i_2i_3} : 1 \leq i_1, i_2, i_3 \leq m\}\right)\). It is seen that for all \(1 \leq m \leq n\),
\[
\mathbb{E}[X_{n,m} | \mathcal{F}_{n,m-1}] = \mathbb{E}[Z_{n,m} | \mathcal{F}_{n,m-1}] - Z_{n,m-1} = 0,
\]
so \(\{X_{n,m}\}_{m=1}^{n}\) is a martingale difference sequence with respect to \(\{\mathcal{F}_{n,m}\}_{1 \leq m \leq n}\). We have the following lemma which is proved below.

**Lemma F.3.** Under the conditions of Lemma 2.2, if the null hypothesis is true, then as \(n \to \infty\),
\[
(a) \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] \to 1, \quad \text{in probability},
\]
\[
(b) \forall \epsilon > 0, \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 \mathbb{I}\{|X_{n,m}| > \epsilon\} | \mathcal{F}_{n,m-1}] \to 0, \quad \text{in probability}.
\]

By Lemma F.3 and [3, Corollary 3.1], it follows from (F.62) that under the null,
\[
Z_n = Z_{n,n} \to N(0, 1).
\]
This proves (F.60).

**F.1 Proof of Lemma F.1**

We first prove the claim (b). By definitions
\[
\bar{\alpha}_n = \mathbb{E}[\hat{\alpha}_n] = \frac{\sum_{i_1, i_2, i_3(i\text{dist})} Q_{i_1i_2i_3}}{n(n-1)(n-2)},
\]
Recall that under alternative
\[ Q_{i_1i_2i_3} = \sum_{1 \leq k_1, k_2, k_3 \leq K} \pi_{i_1}(k_1)\pi_{i_2}(k_2)\pi_{i_3}(k_3)P_{k_1k_2k_3}, \quad 1 \leq i_1, i_2, i_3 \leq n. \]
It is seen that \( Q_{i_1i_2i_3} \leq \max_{1 \leq k_1, k_2, k_3 \leq K} \{ P_{k_1k_2k_3} \}, \quad 1 \leq i_1, i_2, i_3 \leq n \) and so
\[ \tilde{\alpha}_n \leq \max_{1 \leq k_1, k_2, k_3 \leq K} \{ P_{k_1k_2k_3} \}, \quad \bar{\alpha}_n = h'(Ph)h + O\left( \frac{\max_{1 \leq k_1, k_2, k_3 \leq K} \{ P_{k_1k_2k_3} \}}{n} \right). \]
At the same time, by our assumption \( \min_{K} \{ h_k \} \geq c_0 \) and elementary calculations
\[ \max_{1 \leq k_1, k_2, k_3 \leq K} \{ P_{k_1k_2k_3} \} \leq C \sum_{1 \leq k_1, k_2, k_3 \leq K} h_{k_1}h_{k_2}h_{k_3}P_{k_1k_2k_3} \leq C\bar{\alpha}_n. \]
These prove the claims in (b). Now we show the claim (a).

Note that, \( \hat{\alpha}_n \) is the average of \( \binom{n}{3} \) independent Bernoulli random variables with parameter bounded by \( C\bar{\alpha}_n \) under both null and alternative hypothesis. By Bernstein’s inequality,
\[ \mathbb{P}(\binom{n}{3}|\hat{\alpha}_n - \bar{\alpha}_n| \geq t) \leq 2 \exp\left( -\frac{t^2}{\binom{n}{3}C\bar{\alpha}_n(1-C\bar{\alpha}_n) + \frac{t}{3}} \right). \]
Let \( t = C(\binom{n}{3})^{\frac{\log(n)\bar{\alpha}_n^{1/2}}{n^{3/2}}} \), by elementary calculations, we get
\[ \mathbb{P}\left(|\hat{\alpha}_n - \bar{\alpha}_n| \geq C\frac{\log(n)\bar{\alpha}_n^{1/2}}{n^{3/2}} \right) \leq O(1/n). \]  
(F.63)
This is equivalent to the claim in (a).

### F.2 Proof of Lemma F.2

Recall that
\[ Z_n = (2n)^{-1/2} \frac{(Ia)}{(n-1)\bar{\alpha}_n(1-\bar{\alpha}_n)}, \quad \text{with } (Ia) = \sum_{x \in S}(A_{i_1i_2i_3} - \bar{\alpha}_n)(A_{i_3i_4i_5} - \bar{\alpha}_n). \]
Therefore, to show the claims, it is sufficient to show that as \( n \to \infty \)
\[ \mathbb{E}[(Ia)] \geq Cn^2\bar{\alpha}_n^2\bar{\alpha}_n^2, \quad \text{(F.64)} \]
and
\[ \text{Var}((Ia)) \leq Cn^7\bar{\alpha}_n^3. \quad \text{(F.65)} \]
Consider (F.64) first. Since for each \( x = (i_1, i_2, i_3, i_4, i_5) \in S \), \( A_{i_1i_2i_3} \) is independent of \( A_{i_3i_4i_5} \), by direct calculations,
\[ \mathbb{E}[(Ia)] = \sum_{x \in S}(Q_{i_1i_2i_3} - \bar{\alpha}_n)(Q_{i_3i_4i_5} - \bar{\alpha}_n). \]
Let \( \tilde{Q}_{i_1i_2i_3} = Q_{i_1i_2i_3} - \bar{\alpha}_n \), by definitions,
\[ \mathbb{E}[(Ia)] = \frac{1}{4} \left( \sum_{x \in S_0} \tilde{Q}_{i_1i_2i_3} \tilde{Q}_{i_3i_4i_5} \right) - \sum_{x \in (S_0 \backslash S'_1)} \tilde{Q}_{i_1i_2i_3} \tilde{Q}_{i_3i_4i_5}, \]
where
\[ S_0' = \{x : 1 \leq i_1, i_2, i_3, i_4, i_5 \leq n\} \]
\[ S'_1 = \{x \in S_0' : i_1, i_2, i_3, i_4, i_5 \leq n; (i_1, i_2) \neq (i_4, i_5)\}. \]
To show (F.64), it is sufficient to show that
\[
\sum_{x \in S_0} \widetilde{Q}_{i_1 i_2 i_3 i_4} \geq C n^5 \alpha_n^2 \delta_n^2, \quad \text{and} \quad \sum_{x \in (S_0 \setminus S_1')} \widetilde{Q}_{i_1 i_2 i_3} \widetilde{Q}_{i_4 i_5} = o(\sum_{x \in S_0} \widetilde{Q}_{i_1 i_2 i_3} \widetilde{Q}_{i_4 i_5}). \tag{F.66}
\]

Consider the first claim in (F.66). Recall that
\[
\widetilde{Q}_{i_1 i_2 i_3} = Q_{i_1 i_2 i_3} - \alpha_n = \sum_{k_1, k_2, k_3} \pi_{i_1}(k_1) \pi_{i_2}(k_2) \pi_{i_3}(k_3) P_{k_1 k_2 k_3} - \alpha_n, \quad \text{and} \quad h = \frac{\sum_{i=1}^n \pi_i}{n}.
\]

By direct calculations and elementary algebra,
\[
\sum_{x \in S_0} \widetilde{Q}_{i_1 i_2 i_3} \widetilde{Q}_{i_4 i_5} = n^4 \| (P_h) h - \alpha_n 1_n \|^2.
\]

By triangle inequality, we have \( \| (P_h) h - \alpha_n 1_n \| \geq \| (P_h) h - h' (P_h) h 1_n \| - \| (h' (P_h) h - \alpha_n) 1_n \|. \)

It follows that
\[
\sum_{x \in S_0} \widetilde{Q}_{i_1 i_2 i_3} \widetilde{Q}_{i_4 i_5} \geq n^4 \| (P_h) h - h' (P_h) h 1_n \| - \| (h' (P_h) h - \alpha_n) 1_n \|^2. \tag{F.67}
\]

Recall that \( \Sigma = \Pi /n - h h' \) and note that \( \Sigma 1_K = 0 \). Also, recall that \( H_K = K^{-1} 1_K 1_K \) and note that \( I_K - H_K \) is a projection matrix. By elementary algebra,
\[
\Sigma = (I_K - H_K) \Sigma (I_K - H_K).
\]

First, by elementary algebra,
\[
\| (P_h) h - h' (P_h) h 1_n \|^2 = n \| (P_h) - h' (P_h) h h' (P_h) h \| = n \| (P_h) h \| \Sigma (P_h) h, \tag{F.68}
\]

where the RHS equals to
\[
n \| (P_h) h \| (I_K - H_K) \Sigma (I_K - H_K) (P_h) h. \tag{F.69}
\]

By our assumption \( \lambda_{K-1}(\Sigma) = \min_{\|v\|=1, v \perp 1_K} v' \Sigma v \geq c_0 \), it is seen that
\[
n \| (P_h) h \| (I_K - H_K) \Sigma (I_K - H_K) (P_h) h \geq c_0 n \alpha_n^2 \| \alpha_n^{-1} (I_K - H_K) (P_h) h \|^2. \tag{F.70}
\]

Recall that \( \delta_n = \| \alpha_n^{-1} (I_K - H_K) (P_h) h \| \), combining with (F.68)-(F.70), we get
\[
\| (P_h) h - h' (P_h) h 1_n \|^2 \geq c_0 n \alpha_n^2 \delta_n^2. \tag{F.71}
\]

At the same time, by Lemma F.1,
\[
\alpha_n = h' (P_h) h + O(\frac{\alpha_n}{n}). \tag{F.72}
\]

By direct calculations,
\[
\| (h' (P_h) h - \alpha_n) 1_n \|^2 = n \| (h' (P_h) h - \alpha_n) \|^2 = O(\frac{\alpha_n^2}{n}), \tag{F.73}
\]

where by \( \alpha_n \leq \max_{1 \leq i_1, i_2, i_3 \leq n} \{ P_{i_1 i_2 i_3} \} \leq c_0 \) and our condition \( n^{3/2} \alpha_n^{-1/2} \delta_n^2 \rightarrow \infty \),
\[
\frac{\alpha_n^2}{n} = o(1) \cdot (n \alpha_n^2 \delta_n^2), \tag{F.74}
\]
Combining (F.72)-(F.74),
\[ ||(h'(\mathcal{P}h) - \tilde{\alpha}_n)1_n||^2 = o(n^{\tilde{\alpha}_n^2} \bar{\alpha}_n^2). \]  \hspace{1cm} (F.75)

Inserting (F.71) and (F.75) into (F.67) proves the first claim in (F.66).

Next, we consider the second claim in (F.66). Notice that by symmetry, the two leading terms of \( \sum_{x \in (S'_0 \setminus S'_1)} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \) are the following:
\[ O\left( \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq n} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \right), \quad \text{and} \quad O\left( \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq n} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \right). \]  \hspace{1cm} (F.76)

The other terms are \( O(n^3 \tilde{\alpha}_n^2) = o(n^5 \tilde{\alpha}_n^2) \) and thus are negligible. It is therefore adequate to consider the two terms in (F.76).

Consider the first term in (F.76). By Cauchy-Schwarz inequality,
\[ \left| \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq n} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \right| \leq \sqrt{\sum_{1 \leq i_1 \leq n} \left( \sum_{1 \leq i_3 \leq n} \bar{Q}_{i_1i_3i_5} \right)^2} \sqrt{\sum_{1 \leq i_3 \leq n} \left( \sum_{1 \leq i_1, i_2 \leq n} \bar{Q}_{i_1i_2i_3} \right)^2}. \]  \hspace{1cm} (F.77)

Note that by definitions and Lemma F.1, \( |\bar{Q}_{i_1i_2i_3}| \leq C \alpha_n \). It is seen that
\[ \sum_{1 \leq i_3 \leq n} \left( \sum_{1 \leq i_1, i_2 \leq n} \bar{Q}_{i_1i_2i_3} \right)^2 \leq Cn^3 \alpha_n^2. \]  \hspace{1cm} (F.78)

By our condition \( n^{3/2} \alpha_n^{1/2} \delta_n^2 \to \infty \), we have \( n^2 \delta_n^2 \to \infty \). Comparing the RHS of (F.78) with the first claim of (F.66), the RHS is at a smaller order of \( \sum_{x \in S'_0} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \). At the same time,
\[ \sum_{1 \leq i_3 \leq n} \left( \sum_{1 \leq i_1, i_2 \leq n} \bar{Q}_{i_1i_2i_3} \right)^2 = \sum_{x \in S'_0} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6}. \]  \hspace{1cm} (F.79)

Inserting (F.78)-(F.79) into (F.77), we have
\[ \left| \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq n} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \right| = o\left( \sum_{x \in S'_0} \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6} \right). \]

For the second term in (F.76), the analysis is similar, so we omit the details. These prove the second claim of (F.66), and so complete the proof of (F.64).

Next we consider (F.65). Let \( \mathcal{W} \) be the tensor with the same size as \( \mathcal{A} \), where \( \mathcal{W}_{i_1i_2i_3} = A_{i_1i_2i_3} - \bar{Q}_{i_1i_2i_3} \) if \( i_1, i_2, i_3 \) are distinct, and \( \mathcal{W}_{i_1i_2i_3} = 0 \) otherwise. By symmetry and definitions,
\[ (Ia) = \sum_{x \in S} (\mathcal{W}_{i_1i_2i_3} - \bar{Q}_{i_1i_2i_3})(\mathcal{W}_{i_4i_5i_6} - \bar{Q}_{i_4i_5i_6}) = \sum_{x \in S} (\mathcal{W}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} - 2 \bar{Q}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} + \bar{Q}_{i_1i_2i_3} \bar{Q}_{i_4i_5i_6}). \]  \hspace{1cm} (F.80)

Since for any random variables \( X \) and \( Y \), \( \text{Var}(X + Y) \leq 2\text{Var}(X) + 2\text{Var}(Y) \), we have
\[ \text{Var}((Ia)) \leq 2\text{Var}\left( \sum_{x \in S} \mathcal{W}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} \right) + 2\text{Var}\left( \sum_{x \in S} 2 \bar{Q}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} \right). \]

Here, we note that \( \bar{Q} \) is non-random, so the variance of the last term in (F.80) is 0. By direct calculations,
\[ \text{Var}\left( \sum_{x \in S} \mathcal{W}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} \right) = \sum_{x \in S} \text{Var}(\mathcal{W}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6}) = O(n^5 \tilde{\alpha}_n^2), \]
\[ \text{Var}\left( \sum_{x \in S} 2 \bar{Q}_{i_1i_2i_3} \mathcal{W}_{i_4i_5i_6} \right) = \frac{1}{4} \sum_{i_3,i_5,i_6 \in \text{dist}} (\sum_{i_1,i_2} \mathcal{Q}_{i_1i_2i_3})^2 \text{Var}(\mathcal{W}_{i_4i_5i_6}) = O(n^7 \tilde{\alpha}_n^3). \]
By our assumptions, \( n^2 \tilde{\alpha}_n \to \infty \), and so \( n^5 \tilde{\alpha}_n = o(1) \cdot n^7 \tilde{\alpha}_n^2 \). Combining these gives that
\[
\Var((Ia)) \leq C n^7 \tilde{\alpha}_n^2.
\]
This proves (F.65).

F.3 Proof of Lemma F.3

We first show claim (a). By Chebyshev’s inequality, it is sufficient to show that
\[
\mathbb{E}\left[ \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] \right] \to 1, \quad \Var\left( \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] \right) \to 0. \tag{F.81}
\]
Introduce
\[
T^{(m)} = \mathbb{E}\left[ \sum_{x \in S^{(m)} \setminus S^{(m-1)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} | \mathcal{F}_{n,m-1} \right].
\]
By definitions,
\[
\mathbb{E}[X_{n,m}^2 | \mathcal{F}_{n,m-1}] = \frac{\mathbb{E}\left[ \left( \sum_{x \in S^{(m)} \setminus S^{(m-1)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \right)^2 | \mathcal{F}_{n,m-1} \right]}{(\sqrt{2n}(n-1) \alpha_n(1 - \alpha_n))^2} = \frac{T^{(m)}}{(\sqrt{2n}(n-1) \alpha_n(1 - \alpha_n))^2}.
\]
To show (F.81), it is sufficient to show that
\[
\mathbb{E}\left[ \sum_{m=1}^{n} T^{(m)} \right] = \frac{n^5 \alpha_n^2(1 - \alpha_n)^2}{2} (1 + o(1)), \tag{F.82}
\]
and that
\[
\Var\left( \sum_{m=1}^{n} T^{(m)} \right) = o(n^{10} \alpha_n^4). \tag{F.83}
\]

Consider (F.82) first. Recall that \( S^{(m)} = \{ x = (i_1, i_2, i_3, i_4, i_5) \in S : \max\{i_1, i_2, i_3, i_4, i_5\} \leq m \} \) and \( x = (i_1, i_2, i_3, i_4, i_5) \) for short. Similarly, for short, we write \( x' = (i'_1, i'_2, i'_3, i'_4, i'_5) \) and let \( (S^{(m)} \setminus S^{(m-1)})^2 = \{ (x, x') : x \in S^{(m)} \setminus S^{(m-1)}, x' \in S^{(m)} \setminus S^{(m-1)} \} \).

Let
\[
S_1^{(m)} = \{(x, x') \in (S^{(m)} \setminus S^{(m-1)})^2 : i_3 = i'_3, \{i_1, i_2, i_4, i_5\} = \{i'_1, i'_2, i'_4, i'_5\}\},
\]
\[
S_2^{(m)} = (S^{(m)} \setminus S^{(m-1)})^2 \setminus S_1^{(m)}.
\]
It is seen that the LHS of (F.82) equals to
\[
(I) + (II),
\]
where
\[
(I) = \mathbb{E}\left[ \sum_{m=1}^{n} \mathbb{E}\left[ \sum_{(x, x') \in S_1^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} | \mathcal{F}_{n,m-1} \right] \right],
\]
and
\[
(II) = \mathbb{E}\left[ \sum_{m=1}^{n} \mathbb{E}\left[ \sum_{(x, x') \in S_2^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i'_1 i'_2 i'_3} \tilde{A}_{i'_3 i'_4 i'_5} | \mathcal{F}_{n,m-1} \right] \right].
\]
Notice that for any \( (x, x') \in S_2^{(m)} \), each \( \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i'_1 i'_2 i'_3} \tilde{A}_{i'_3 i'_4 i'_5} \) is a mean-zero random variable. It follows that
\[
(II) = 0.
\]
At the same time, note that for any \((x, x') \in SS_1^{(m)}\) (where \(x = (i_1, i_2, i_3, i_4, i_5)\) and \(x' = (i_1', i_2', i_3', i_4', i_5')\)), there are two possibilities: \((i_1, i_2, i_4, i_5) = (i_1', i_2', i_4', i_5')\) and \((i_1, i_2, i_4, i_5) = (i_1', i_3', i_2', i_5')\). By symmetry,

\[
(I) = 2 \sum_{m=1}^{n} \sum_{x \in S^{(m)}} \mathbb{E} \left[ \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \right] = 2 \sum_{x \in S} \alpha_n^2 (1 - \alpha_n)^2 = 12n \left( \frac{n}{4} \right) \alpha_n^2 (1 - \alpha_n)^2.
\]

Combining these gives (F.82).

Next, consider (F.83). In \(S^{(m)} \setminus S^{(m-1)}\), we have \(i_2 = m\) or \(i_2 = m\) or \(i_5 = m\). Let

\[
S_1^{(m)} = \left\{ x \in S^{(m)} \setminus S^{(m-1)} : \text{either } i_2 = m, i_5 < m \text{ or } i_5 = m, i_2 < m \right\},
\]

\[
S_2^{(m)} = (S^{(m)} \setminus S^{(m-1)}) \setminus S_1^{(m)}.
\]

Write

\[
T^{(m)} = T_1^{(m)} + 2T_2^{(m)} + T_3^{(m)},
\]

where

\[
T_1^{(m)} = \mathbb{E} \left[ \sum_{x, x' \in S_1^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} | F_{n,m-1} \right],
\]

\[
T_2^{(m)} = \mathbb{E} \left[ \sum_{x \in S_1^{(m)}, x' \in S_2^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} | F_{n,m-1} \right],
\]

\[
T_3^{(m)} = \mathbb{E} \left[ \sum_{x, x' \in S_2^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} | F_{n,m-1} \right].
\]

Notice that for \(x \in S_1^{(m)}, x' \in S_2^{(m)}\), \(\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'}\) is mean-zero conditional on \(F_{n,m-1}\). It follows directly that

\[
T_2^{(m)} = 0.
\]

Also, by definitions, for each \(x \in S_2^{(m)}\), we must have \(i_3 = m\) or \(i_2 = i_5 = m\). Let \(E_m = \{(x, x') \in S_2^{(m)} \times S_2^{(m)} : \{i_1, i_2, i_3, i_4, i_5\} = \{i_1', i_2', i_3', i_4', i_5'\}\}\), by direct calculations

\[
T_3^{(m)} = |E_m| \alpha_n^2 (1 - \alpha_n)^2.
\]

It is seen that \(T_3^{(m)}\) is non-random. Therefore,

\[
T^{(m)} = T_1^{(m)} + |E_m| \alpha_n^2 (1 - \alpha_n)^2, \quad \text{and} \quad \text{Var}(\sum_{m=1}^{n} T^{(m)}) = \text{Var}(\sum_{m=1}^{n} T_1^{(m)}),
\]

and to show (F.83), it is sufficient to show that

\[
\text{Var}(\sum_{m=1}^{n} T_1^{(m)}) = o(n^{10} \alpha_n^4).
\]

By definitions and symmetry

\[
T_1^{(m)} = \mathbb{E} \left[ 4 \sum_{1 \leq i_1, i_2, i_3, i_4, i_5, i_1', i_2', i_3', i_4', i_5' \leq m-1} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} | F_{n,m-1} \right].
\]

If \(\{i_1, i_4\} \neq \{i_1', i_4'\}\), then \(\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_1' i_2' i_3'}\) has a conditional mean of zero. Therefore, we have

\[
T_1^{(m)} = T_{11}^{(m)} + T_{12}^{(m)}.
\]
where
\[
T_{11}^{(m)} = E[4 \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq m-1} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i_1' i_2' i_3'} | F_{n, m-1}],
\]
\[
T_{12}^{(m)} = E[4 \sum_{1 \leq i_1, i_2, i_3, i_4, i_5 \leq m-1} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i_1' i_2' i_3'} | F_{n, m-1}] .
\]

Since for any random variables \(X\) and \(Y\), \(\text{Var}(X + Y) \leq 2\text{Var}(X) + 2\text{Var}(Y)\), to show (F.84), it is sufficient to show that
\[
\text{Var} \left( \sum_{m=1}^{n} T_{11}^{(m)} \right) = o(n^{10} \alpha_n^4), \quad \text{and} \quad \text{Var} \left( \sum_{m=1}^{n} T_{12}^{(m)} \right) = o(n^{10} \alpha_n^4). \tag{F.85}
\]

Consider the first claim in (F.85). Recall that
\[
\bar{T}_{n, m} = \sum_{x \in S^{(m)}} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} = \sum_{1 \leq i_1, \ldots, i_5 \leq m} \tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} .
\]

By elementary calculations
\[
T_{11}^{(m)} = 4(m - 2) \alpha_n (1 - \alpha_n) \bar{T}_{n, m-1} + 4(m - 2) \alpha_n (1 - \alpha_n) \sum_{1 \leq i_1, i_2, i_3 \leq m-1} \tilde{A}_{i_1 i_2 i_3}^2 .
\]

By inequality \(\text{Var}(X + Y) \leq 2\text{Var}(X) + 2\text{Var}(Y)\), to show the first claim in (F.85), it is sufficient to show that
\[
\text{Var} \left( \sum_{m=1}^{n} 4(m - 2) \alpha_n (1 - \alpha_n) \bar{T}_{n, m-1} \right) = o(n^{10} \alpha_n^4), \tag{F.86}
\]
and
\[
\text{Var} \left( \sum_{m=1}^{n} 4(m - 2) \alpha_n (1 - \alpha_n) \sum_{1 \leq i_1, i_2, i_3 \leq m-1} \tilde{A}_{i_1 i_2 i_3}^2 \right) = o(n^{10} \alpha_n^4). \tag{F.87}
\]

Consider the LHS of (F.86), by definitions,
\[
\text{Var} \left( \sum_{m=1}^{n} 4(m - 2) \alpha_n (1 - \alpha_n) \bar{T}_{n, m-1} \right) = \sum_{m, m'=1}^{n} 16(m - 2)(m'-2) \alpha_n^2 (1 - \alpha_n)^2 \text{Cov}(\bar{T}_{n, m-1}, \bar{T}_{n, m'-1}) . \tag{F.88}
\]

Notice that
\[
\text{Cov}(\bar{T}_{n, m-1}, \bar{T}_{n, m'-1}) = \sum_{1 \leq i_1, \ldots, i_5 \leq m} \sum_{1 \leq i_1', \ldots, i_5' \leq m} \text{E}[\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_3' i_4' i_5'}].
\]

Only if \(\{i_1, i_2, i_3, i_4, i_5\} = \{i_1', i_2', i_3', i_4', i_5'\}\), \(\text{E}[\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_3 i_4 i_5} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_3' i_4' i_5'}]\) will be non-zero. Since there are only a bounded number of ways to pair the indexes, by direct calculations
\[
\text{Cov}(\bar{T}_{n, m-1}, \bar{T}_{n, m'-1}) = O \left( \sum_{1 \leq i_1, \ldots, i_5 \leq m} \text{E}[\tilde{A}_{i_1 i_2 i_3}^2 \tilde{A}_{i_3 i_4 i_5}^2] \right) = O(n^5 \alpha_n^2). \]
Combining this with (F.88), it is seen that

\[ \text{Var}(\sum_{m=1}^{n} 4(m-2)\alpha_n(1-\alpha_n)\tilde{F}_{n,m-1}) = O(n^4 n^{3/2} \alpha_n^3) = o(n^{10} \alpha_n^4). \]

This proves (F.86).

Next consider the LHS of (F.87), by direct calculations,

\[
\text{Var}(\sum_{m=1}^{n} 4(m-2)\alpha_n(1-\alpha_n) \sum_{1 \leq i_1, i_2, i_3 \leq m-1 \atop i_1 < i_2} \tilde{A}_{i_1 i_2 i_3}^2) \leq 16n^4 \alpha_n^2 (1-\alpha_n)^2 \text{Var}(\sum_{1 \leq i_1, i_2, i_3 \leq n \atop i_1 < i_2} \tilde{A}_{i_1 i_2 i_3}^2) \\
= 16n^4 \alpha_n^2 (1-\alpha_n)^2 \cdot \sum_{1 \leq i_1, i_2, i_3 \leq n \atop i_1 < i_2} 3 \cdot \text{Var}(\tilde{A}_{i_1 i_2 i_3}^2) \\
= O(n^7 \alpha_n^4).
\]

By our assumption $n^2 \tilde{\alpha}_n \to \infty$ (i.e., $n^2 \alpha_n \to \infty$), the RHS of the above inequality is $o(n^{10} \alpha_n^4)$.

This proves (F.87) and completes the first claim of (F.85).

Next consider the second claim in (F.85), by definitions,

\[
\text{Var}(\sum_{m=1}^{n} T_{12}^{(m)}) = \sum_{m, m'=1}^{n} 16n^2 (1-\alpha_n)^2 \sum_{1 \leq i_1, \ldots, i_6 \leq m \atop i_1 < i_2 < i_3} \sum_{1 \leq i_1', \ldots, i_6' \leq m \atop i_1' < i_2' < i_3'} E[\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_4 i_5 i_6} \tilde{A}_{i_1' i_2' i_3'} \tilde{A}_{i_4' i_5' i_6'}].
\]

Similarly, it is sufficient to consider terms that satisfy \( \{i_1, \ldots, i_6\} = \{i_1', \ldots, i_6'\} \), hence

\[
\text{Var}(\sum_{m=1}^{n} T_{12}^{(m)}) = O(\sum_{m, m'=1}^{n} 16n^2 (1-\alpha_n)^2 \sum_{1 \leq i_1, \ldots, i_6 \leq m \atop i_1 < i_2 < i_3} \sum_{1 \leq i_1', \ldots, i_6' \leq m \atop i_1' < i_2' < i_3'} E[\tilde{A}_{i_1 i_2 i_3} \tilde{A}_{i_4 i_5 i_6}]) = O(n^8 \alpha_n^4).
\]

Note that the RHS above is $o(n^{10} \alpha_n^4)$. This proves the second claim in (F.85) and completes the proof of claim (a) of (F.81).

Now we consider the claim (b), where the goal is to show that

\[ \forall \epsilon > 0, \sum_{m=1}^{n} E[X_{n,m}^2 \mathbb{I}\{|X_{n,m}| > \epsilon\}|\mathcal{F}_{n,m-1}] \to 0, \quad \text{in probability.} \quad \text{(F.89)} \]

By Cauchy-Schwarz inequality

\[ \left| \sum_{m=1}^{n} E[X_{n,m}^2 \mathbb{I}\{|X_{n,m}| > \epsilon\}|\mathcal{F}_{n,m-1}] \right| \leq \sum_{m=1}^{n} \sqrt{E[X_{n,m}^4 |\mathcal{F}_{n,m-1}]} \sqrt{\mathbb{P}(|X_{n,m}| > \epsilon |\mathcal{F}_{n,m-1})}. \quad \text{(F.90)} \]

At the same time, by Markov’s inequality,

\[ \sqrt{\mathbb{P}(|X_{n,m}| > \epsilon |\mathcal{F}_{n,m-1})} \leq \sqrt{E[X_{n,m}^4 |\mathcal{F}_{n,m-1}]/\epsilon^4}. \quad \text{(F.91)} \]

Combining (F.90) and (F.91) gives

\[ \left| \sum_{m=1}^{n} E[X_{n,m}^2 \mathbb{I}\{|X_{n,m}| > \epsilon\}|\mathcal{F}_{n,m-1}] \right| \leq \sum_{m=1}^{n} E[X_{n,m}^4 |\mathcal{F}_{n,m-1}]/\epsilon^2. \]
To show (F.89), by Markov’s inequality, it is sufficient to show that
\[
E \left[ \sum_{m=1}^{n} E[X_{n,m}^4 | \mathcal{F}_{n,m-1}] \right] \rightarrow 0.
\] (F.92)

Recall that
\[
X_{n,m} = \frac{\sum_{x \in S^{(m)}} A_{ij} i_{23} A_{i_{34}i_5} A_{i_{45}i_6}}{\sqrt{2n^2} \alpha_n (1 - \alpha_n)}.
\]

Write for short \(y = (i_1, i_2, i_3, i_4, i_5, j_1, j_2, j_3, j_4, j_5)\), similarly, \(y' = (i'_1, i'_2, i'_3, i'_4, j'_1, j'_2, j'_3, j'_4, j'_5)\).

To show (F.92), it is sufficient to show that
\[
E \left[ \sum_{m=1}^{n} \sum_{y,y' \in (S^{(m)})^2} A_{i_1i_2i_3} A_{i_3i_4i_5} A_{j_1j_2j_3} A_{j_3j_4j_5} A_{i'_1i'_2i'_3} A_{i'_4i'_5i'_6} A_{j'_1j'_2j'_3} A_{j'_4j'_5j'_6} \right] = o(n^{10} \alpha_n^4).
\]

Similarly, to have non-zero expected value, \(A_{i_1i_2i_3} A_{i_3i_4i_5} A_{j_1j_2j_3} A_{j_3j_4j_5} A_{i'_1i'_2i'_3} A_{i'_4i'_5i'_6} A_{j'_1j'_2j'_3} A_{j'_4j'_5j'_6}\) must be in quadratic form. Since there are only a bounded number of ways to pair them into quadratic forms, it is sufficient to show that
\[
\sum_{m=1}^{n} \sum_{y \in (S^{(m)})^2} E[A_{i_1i_2i_3} A_{i_3i_4i_5} A_{j_1j_2j_3} A_{j_3j_4j_5} A_{i'_1i'_2i'_3} A_{i'_4i'_5i'_6} A_{j'_1j'_2j'_3} A_{j'_4j'_5j'_6}] = o(n^{10} \alpha_n^4).
\]

Recall that for each \(x \in S^{(m)} \setminus S^{(m-1)}\), there are at least one index of \((i_1, i_2, i_3, i_4, i_5)\) is \(m\). It is seen that
\[
\sum_{m=1}^{n} \sum_{y \in (S^{(m)})^2} E[A_{i_1i_2i_3} A_{i_3i_4i_5} A_{j_1j_2j_3} A_{j_3j_4j_5} A_{i'_1i'_2i'_3} A_{i'_4i'_5i'_6} A_{j'_1j'_2j'_3} A_{j'_4j'_5j'_6}] \leq \sum_{m=1}^{n} n^{10-2} (\alpha_n (1 - \alpha_n))^4 = o(n^{10} \alpha_n^4).
\]

This finishes the proof.

### G Proof of Theorem 3.2

Recall that \(\phi_n = \max_{2 \leq m \leq M} \{\phi^{(m)}\}\). To prove this theorem, it is sufficient to show that if there is a \(m \in \{2, \ldots, M\}\) such that \(||\theta^{(m)}||_{m-2}^2 ||g^{(m)}||^2 (\mu_2^{(m)})^2 \gg \log(n)\), we will have
\[
\phi^{(m)}_n \rightarrow 0 \text{ under } H_0, \quad \text{and } \phi^{(m)}_n \rightarrow \infty \text{ under } H_1.
\]

Fix \(m\). For simplicity, we remove the superscript \((m)\) whenever it is clear from the context.

Let
\[
\alpha_n = E[\bar{\alpha}_n], \quad \beta = \sum_{k_2, \ldots, k_m=1}^{K} P_{k_2, \ldots, k_m} g_{k_2} \cdots g_{k_m} / (|\mathcal{P}; g, \ldots, g|^{(m-1)/m}),
\]
where \(g \in \mathbb{R}^K\) is defined by \(g_k = (1/||\theta||_1) \sum_{i=1}^{n} \theta_i \pi_i(k), 1 \leq k \leq K\).

Introduce ideal counterparts of \(V_n\) and \(\eta\) by
\[
\bar{V}_n = \binom{n}{m} \alpha_n (1 - \alpha_n) \quad \text{and} \quad \eta^* = \Theta \Pi \beta, \quad \text{respectively.} \quad \text{(G.93)}
\]

The following lemma is used in this proof and we prove it after the main proof.

**Lemma G.1.** With the conditions of Theorem 3.2, as \(n \rightarrow \infty\),

- (a) Under both the null and alternative, \(\bar{V}_n/V_n \rightarrow 1\) in probability.
• (b) Under the null, with a probability at least $1 - O(1/n)$, 
$$ \max_{1 \leq i \leq n} \{|\eta_i/\eta_i^* - 1|\} \leq C(n^{m-1} \theta_{\max}^m / \log(n))^{-1/2}.$$  

• (c) Under the alternative, with a probability at least $1 - O(1/n)$, 
$$ \max_{1 \leq i \leq n} \{|\eta_i/\eta_i^* - 1|\} \leq C(n^{m-1} \theta_{\max}^m / \log(n))^{-1/2} + C_{\gamma_n} \theta_{\max} \gamma_n / (n^{m+1} \theta_{\max}^m \log(n))^{1/2} \to \infty, $$  
where $\gamma_n = \max_{1 \leq k_1, \ldots, k_m \leq K} \{ |P_{k_1 \ldots k_m} - \beta_{k_1} \cdots \beta_{k_m}| \}$. 

G.1 Main Proof of Theorem 3.2

Recall that $\phi_n^{(m)} = Q_n / \sqrt{n \log(n)^{1.1} V_n}$. The goal is to show that with probability $1 - o(1)$
$$ Q_n \leq (n \log(n)^{1.1} V_n)^{1/2} \text{ under } H_0^{(n)}, \quad Q_n \geq (n \log(n)^{1.1} V_n)^{1/2} \text{ under } H_1^{(n)}, $$  
(G.94)

By (a) in Lemma G.1, $V_n / V_1 \to 1$ in probability. Hence to show (G.94), it is sufficient to show that with probability $1 - o(1)$
$$ Q_n \leq 0.5(n \log(n)^{1.1} V_n)^{1/2} \text{ under } H_0^{(n)}, \quad Q_n \geq 1.5(n \log(n)^{1.1} V_n)^{1/2} \text{ under } H_1^{(n)}. $$  
(G.95)

Recall that
$$ Q_n = \max_{S = (S_1, \ldots, S_{m+1}) \in B} \max_{1 \leq k_1, \ldots, k_m \leq m+1} \{|X_{S,k_1 \ldots k_m}|\}, $$  
where
$$ X_{S,k_1 \ldots k_m} = \sum_{i_t \in S_{k_t}, \ldots, i_m \in S_{k_m}} (A_{i_1 \ldots i_m} - \eta_{i_1} \cdots \eta_{i_m}). $$

Also, recall that $\eta^*$ is the ideal counterparts of $\eta$, defined in (G.93). Introduce a counterpart of $X_{S,k_1 \ldots k_m}$ by replacing $\eta$ with $\eta^*$
$$ \tilde{X}_{S,k_1 \ldots k_m} = \sum_{i_t \in S_{k_t}, \ldots, i_m \in S_{k_m}} (A_{i_1 \ldots i_m} - \eta_{i_1}^* \cdots \eta_{i_m}^*). $$

Let
$$ \tilde{Q}_n = \max_{S = (S_1, \ldots, S_{m+1}) \in B} \max_{1 \leq k_1, \ldots, k_m \leq m+1} \{|\tilde{X}_{S,k_1 \ldots k_m}|\}. $$

Note that for any number $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$,
$$ |\max\{x_1, x_2, \ldots, x_n\} - \max\{y_1, y_2, \ldots, y_n\}| \leq \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\}. $$

It is seen that
$$ |Q_n - \tilde{Q}_n| \leq \max_{S \in B} \max_{1 \leq k_1, \ldots, k_m \leq m+1} \{|X_{S,k_1 \ldots k_m} - \tilde{X}_{S,k_1 \ldots k_m}|\}. $$  
(G.96)

At the same time, by definitions and direct calculations, for all $S = (S_1, \ldots, S_{m+1}) \in B$ and $1 \leq k_1, \ldots, k_m \leq m + 1$
$$ |X_{S,k_1 \ldots k_m} - \tilde{X}_{S,k_1 \ldots k_m}| \leq |S_{k_1}| \cdots |S_{k_m}| \max_{1 \leq i_1, \ldots, i_m \leq n} |\eta_{i_1} \cdots \eta_{i_m} - \eta_{i_1}^* \cdots \eta_{i_m}^*|, $$  
(G.97)

where by (b) and (c) in Lemma G.1, except for a probability $O(1/n)$
$$ \max_{1 \leq i \leq n} \{|\eta_i/\eta_i^* - 1|\} \leq C \left( \frac{\log(n)}{n^{m-1} \theta_{\max}^m} \right)^{1/2} \quad \text{under } H_0, $$  
(G.98)

and
$$ \max_{1 \leq i \leq n} \{|\eta_i/\eta_i^* - 1|\} \leq C \left( \frac{\log(n)}{n^{m-1} \theta_{\max}^m} \right)^{1/2} + \frac{C_{\gamma_n}}{n} \quad \text{under } H_1. $$  
(G.99)
Here $\gamma_n$ denotes $\max_{1 \leq k_1, \ldots, k_m \leq K} \{|P_{k_1 \ldots k_m} - \beta_{k_1} \cdots \beta_{k_m}|\}$ under $H_1$. Note that by our regular conditions and elementary calculations, $\log(n)/(n_{m-1}\theta_{\max}^m) = o(1)$ and $\gamma_n/n = O(1/n)$.

Therefore, $\max_{1 \leq i \leq n} \left\{ \frac{\eta_i}{\gamma_i^*} \right\} = o(1)$ under both hypotheses. By Taylor’s expansion, for $1 \leq i_1, \ldots, i_m \leq n$

$$|\eta_{i_1} \cdots \eta_{i_m} - \eta_{i_1}^* \cdots \eta_{i_m}^*| \leq C \eta_{i_1}^* \cdots \eta_{i_m}^* \max_{1 \leq i \leq n} \left\{ \frac{\eta_i}{\gamma_i^*} - 1 \right\}. \quad (G.100)$$

Combining (G.96)-(G.100) and observe that $\eta_{i}^* \leq C\theta_{\max}$ and $|S_{k_j}| \leq n$, $1 \leq j \leq m$, with probability $1 - o(1)$

$$|Q_n - \bar{Q}_n| \leq C \left( \log(n)n^{m+1}\theta_{\max}^m \right)^{1/2} \quad \text{under } H_0, \quad (G.101)$$

and

$$|Q_n - \bar{Q}_n| \leq C \left( \log(n)n^{m+1}\theta_{\max}^m \right)^{1/2} + C\gamma_n n^{-1}\theta_{\max}^m \quad \text{under } H_1 \quad (G.102)$$

Note that by direct calculations, we have $\bar{V}_n \approx n^m\theta_{\max}^m$. Therefore, to show (G.95), it is sufficient to show that with probability $1 - o(1)$

$$\begin{align*}
(I) & : \bar{Q}_n \leq 0.5(n \log(n)^{1+1}\bar{V}_n)\frac{1}{2} \quad \text{under } H_0^{(n)}, \\
(II) & : \bar{Q}_n \geq 2(n \log(n)^{1+1}\bar{V}_n)\frac{1}{2} + C\gamma_n n^{-1}\theta_{\max}^m \quad \text{under } H_1^{(n)}.
\end{align*}$$

Consider (I) first. Recall that

$$\bar{Q}_n = \max_{S = (S_1, \ldots, S_{m+1}) \in B} \max_{1 \leq k_1, \ldots, k_m \leq m+1} \{ |\bar{X}_{S,k_1 \ldots k_m}| \},$$

where the RHS is the maximum of

$$\leq m^n m^m = m^{n+m}$$

random variables. By union bound, it is sufficient to show that for every $S = (S_1, \ldots, S_{m+1}) \in B$ and $1 \leq k_1, \ldots, k_m \leq m+1$, except for a probability of $O(m^{-(n+m)}n^{-1})$

$$|\bar{X}_{S,k_1 \ldots k_m}| \leq 0.5(n \log(n)^{1+1}\bar{V}_n)\frac{1}{2}. \quad (G.103)$$

Now we are going to prove (G.103). Note that under null hypothesis, $\theta^* = \theta$. By definitions

$$\bar{X}_{S,k_1 \ldots k_m} = \sum_{i_1 \in S_1, \ldots, i_m \in S_m} (A_{i_1 \ldots i_m} - \theta_{i_1} \cdots \theta_{i_m}),$$

where by symmetry the RHS is a sum of no more than $\binom{n}{m}$ unique independent random variables, each of which has mean 0 and variance $\leq (m!)^2\theta_{i_1} \cdots \theta_{i_m}(1 - \theta_{i_1} \cdots \theta_{i_m})$. By Bernstein’s inequality, for any $t > 0$,

$$\mathbb{P}(|\bar{X}_{S,k_1 \ldots k_m}| \geq t) \leq 2\exp\left( -\frac{t^2}{\sum_{i_1 \in S_1, \ldots, i_m \in S_m} (m!)^2\theta_{i_1} \cdots \theta_{i_m}(1 - \theta_{i_1} \cdots \theta_{i_m}) + t/3} \right).$$

Since $\sum_{i_1 \in S_1, \ldots, i_m \in S_m} (m!)^2\theta_{i_1} \cdots \theta_{i_m}(1 - \theta_{i_1} \cdots \theta_{i_m}) \leq Cn^m \theta_{\max}^m$, it follows that

$$\mathbb{P}(|\bar{X}_{S,k_1 \ldots k_m}| \geq t) \leq 2\exp\left( -\frac{t^2}{Cn^m \theta_{\max}^m + t/3} \right). \quad (G.104)$$
Taking \( t = (n \log(n) \bar{V}_n)^{1/2} \) and noting that \((1/C) \sqrt{n^{m+1} \log(n) \theta_{\text{max}}^m} \leq t \leq \sqrt{n^{m+1} \log(n) \theta_{\text{max}}^m}\),

\[
\exp\left(-\frac{t^2}{C n^m \theta_{\text{max}}^m + t/3}\right) \leq \exp\left(-\frac{(1/C) \sqrt{n^{m+1} \log(n) \theta_{\text{max}}^m}^2}{C n^m \theta_{\text{max}}^m + \sqrt{n^{m+1} \log(n) \theta_{\text{max}}^m}^3/3}\right).
\]

Combining this with our assumption \( \|\theta\|^{m-2} \|\theta\|^2 / \log(n) \to \infty \) and \( \theta_{\text{max}} \leq C \theta_{\text{min}} \), by elementary calculations, the RHS of (G.104) is \( O(\exp(-C n \log(n))) \). This proves (G.103).

Next, consider (II) for the alternative case. Let \( S_k^\ast \) denote the true partition set \{1 \leq i \leq n : \text{node } i \text{ is in community } k\}, 1 \leq k \leq K. \) Also, recall that

\[
\gamma_n = \max_{1 \leq k_1, \ldots, k_m \leq K} \{|\mathcal{P}_{k_1 \cdots k_m} - \beta_{k_1} \cdots \beta_{k_m} |\}.
\]

Suppose the maximum on the right hand side is assumed at \((k_1, \ldots, k_m) = (k_1^\ast, \ldots, k_m^\ast)\) and so

\[
\gamma_n = |\mathcal{P}_{k_1^\ast \cdots k_m^\ast} - \beta_{k_1^\ast} \cdots \beta_{k_m^\ast} |.
\]

Without loss of generality, assume \( k_1^\ast, \ldots, k_m^\ast \) are distinct. The proofs for the cases that \( k_1^\ast, \ldots, k_m^\ast \) are not distinct are similar, so we omit them.

Now let \( S^\ast = (S_{k_1^\ast}, \ldots, S_{k_m^\ast}, \{1, \ldots, n\} \setminus (S_{k_1^\ast} \cup \cdots \cup S_{k_m^\ast}) \). It follows that \( S^\ast \in B \). By definitions,

\[
\bar{Q}_n \geq |\bar{X}_{S^\ast, k_1^\ast \cdots k_m^\ast}|.
\]

Therefore, to show (II), it is sufficient to show that except for a probability of \( 1 - O(1/n) \),

\[
|\bar{X}_{S^\ast, k_1^\ast \cdots k_m^\ast}| \geq C(n \log(n))^{1/2} \bar{V}_n^{1/2} + C \gamma_n n^{m-1} \theta_{\text{max}}^m. \tag{G.105}
\]

Write

\[
\bar{X}_{S^\ast, k_1^\ast \cdots k_m^\ast} := \sum_{i_1 \in S_{k_1^\ast}, \ldots, i_m \in S_{k_m^\ast}} (A_{i_1 \cdots i_m} - \eta_{i_1}^* \cdots \eta_{i_m}^*) = (I) + (II), \tag{G.106}
\]

where

\[
(I) = \sum_{i_1 \in S_{k_1^\ast}, \ldots, i_m \in S_{k_m^\ast}} (\theta_{i_1} \cdots \theta_{i_m} P_{k_1^\ast \cdots k_m^\ast} - \eta_{i_1}^* \cdots \eta_{i_m}^*),
\]

and

\[
(II) = \sum_{i_1 \in S_{k_1^\ast}, \ldots, i_m \in S_{k_m^\ast}} (A_{i_1 \cdots i_m} - \theta_{i_1} \cdots \theta_{i_m} P_{k_1^\ast \cdots k_m^\ast}).
\]

By definitions, \( \eta_{i_1}^* \cdots \eta_{i_m}^* = \theta_{i_1} \cdots \theta_{i_m} \beta_{k_1^\ast} \cdots \beta_{k_m^\ast} \), for \( i_1 \in S_{k_1^\ast}, \ldots, i_m \in S_{k_m^\ast} \). It is seen that

\[
\| (I) \| = \| \theta \|_{1} g_{k_1^\ast}^* \cdots g_{k_m^\ast}^* \gamma_n.
\]

By our assumption \( \max_{k=1}^K \{h_k\} \leq C \min_{k=1}^K \{h_k\} \) and \( \theta_{\text{max}} \leq C \theta_{\text{min}} \),

\[
\| \theta \|_{1} g_{k_1^\ast}^* \cdots g_{k_m^\ast} \geq C n^{m} \theta_{\text{max}}^m,
\]

and so

\[
\| (I) \| \geq C n^{m} \theta_{\text{max}}^m \gamma_n. \tag{G.107}
\]

Write for short

\[
N = |S_{k_1^\ast}^*| \cdots |S_{k_m^\ast}^*|.
\]

Note that (II) is a sum of no more than \( N \) independent random variables, each with a mean of 0 and a variance less than \( C \theta_{\text{max}}^m \). By Bernstein’s Lemma, for any \( t > 0 \),

\[
\mathbb{P}(| (II) | \geq t) \leq \exp(-\frac{t^2}{NC \theta_{\text{max}}^m + t/3}). \tag{G.108}
\]
Taking $t = (\log(n)\bar{V}_n)^{1/2}$. Note that $t \asymp (\log(n)n^m\theta_{\max}^m)^{1/2}$ and $N \leq n^m$, by direct calculations

$$\exp(-\frac{t^2}{NC\theta_{\max}^m + t/3}) = O(1/n).$$

Putting this into (G.108), gives except for a probability of $O(1/n)$,

$$|(II)| \leq (\log(n)\bar{V}_n)^{1/2}. \quad \text{(G.109)}$$

Inserting (G.107)-(G.109) into (G.106) gives that except for a probability of $O(1/n)$,

$$|\bar{X}_{s^*k^*_1\cdots k^*_m}| \geq Cn^m\theta_{\max}^m \gamma_n - (\log(n)\bar{V}_n)^{1/2}, \quad \text{(G.110)}$$

where we note that by Lemma G.1, $n^m\theta_{\max}^m \gamma_n / (n \log(n)^{1/2} \bar{V}_n)^{1/2} \to \infty$. This proves (G.105) and finishes the proof.

### G.2 Proof of Lemma G.1

Consider the claim (a). By definitions

$$\frac{V_n}{\bar{V}_n} - 1 = \frac{(\hat{\alpha}_n - \bar{\alpha}_n)(1 - \hat{\alpha}_n - \bar{\alpha}_n)}{\hat{\alpha}_n(1 - \bar{\alpha}_n)}. \quad \text{(G.111)}$$

Note that $\hat{\alpha}_n$ is the average of $\binom{n}{m}$ independent Bernoulli random variables with parameters bounded by $C\theta_{\max}^m$ under both null and alternative hypothesis. By Bernstein’s inequality,

$$\mathbb{P}(|(\binom{n}{m})|\hat{\alpha}_n - \bar{\alpha}_n| \geq t) \leq 2 \exp(-\frac{t^2}{C\theta_{\max}^m + t/3}).$$

Let $t = C\log(n)(\binom{n}{m})^{1/2}$, by elementary calculations, we get

$$\mathbb{P}\left(|\hat{\alpha}_n - \bar{\alpha}_n| \geq C\log(n)(\theta_{\max}^m / \binom{n}{m})^{1/2}\right) \leq o(1/n).$$

Combining this with (G.111) and $\bar{\alpha}_n \leq C\theta_{\max}^m \leq C\theta_0^m < 1$, by elementary calculations,

$$\left|\frac{V_n}{\bar{V}_n} - 1\right| \leq C\log(n)(\binom{n}{m})^{1/2} \theta_{\max}^m, \quad \text{except for a probability of } O(1/n),$$

where by our conditions $n^{m-1}\theta_{\max}^m / \log(n) \to \infty$ (implied by $\|\theta\|^{m-2}\|\theta\|^2\mu^2_2 / \log(n) \to \infty$), the RHS is $o(1)$. Therefore $V_n/\bar{V}_n \to 1$ in probability.

Combining this with Slutsky’s Lemma, we get $\bar{V}_n/V_n \to 1$ in probability and finish the proof of (a).

Next we consider the claim (b) and the first claim in (c). Our goal is to show that except for a probability $O(1/n)$

$$\max_{1 \leq i \leq n} \left\{ \frac{\eta_i}{\eta_{i^*}} - 1 \right\} \leq C\left(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\right)^{1/2}, \quad \text{under } H_0 \quad \text{(G.112)}$$

and

$$\max_{1 \leq i \leq n} \left\{ \frac{\eta_i}{\eta_{i^*}} - 1 \right\} \leq C\left(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\right)^{1/2} + C\gamma_n / n, \quad \text{under } H_1. \quad \text{(G.113)}$$

Recall that

$$\eta = u^{(\frac{m-1}{2})} \quad \text{and} \quad u^{(k)} = g(u^{(k-1)}), \quad 1 \leq k \leq m,$$
where for \(1 \leq i \leq n\)

\[
L_i(u) = \frac{\sum_{i_2, \ldots, i_m \text{ (distinct)}} A_{i_1 \cdots i_m} + \sum_{i_2, \ldots, i_m \text{ (non-distinct)}} u_{i_2} \cdots u_{i_m}}{\left(\sum_{i_1, \ldots, i_m \text{ (distinct)}} A_{i_1 \cdots i_m} + \sum_{i_1, \ldots, i_m \text{ (non-distinct)}} u_{i_1} \cdots u_{i_m}\right)^{(m-1)/m}}.
\]

Let \(I^{(i_1)}\) denote \(\{1, \ldots, n\} \setminus \{i_1\}\). We claim that if the following events

\[
E_1 : \max_{1 \leq i_1 \leq n} \left\{ \left| \sum_{i_2, \ldots, i_m \in I^{(i_1)}} (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}) \right| \right\} \leq (n^{m-1} \theta_{\max}^m \log(n))^{1/2},
\]

\[
E_2 : \left| \sum_{i_2, \ldots, i_m \in I^{(i_1)}} (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}) \right| \leq (n^m \theta_{\max}^m)^{1/2}
\]

hold then for \(1 \leq k \leq m\)

\[
\max_{1 \leq i \leq n} \left\{ \left| \frac{L_i(u^{(k)})}{\eta_i^*} - 1 \right| \right\} \leq C \left( \frac{\log(n)}{n^{m-1} \theta_{\max}^m} \right)^{1/2} + C \frac{\log(n)}{n} \max_{1 \leq i \leq n} \left\{ \left| \frac{u^{(k-1)}_i}{\eta_i^*} - 1 \right| \right\} + C \frac{\gamma_n}{n},
\]

where by definitions \(\gamma_n\) is 0 under \(H_0\).

Note that inequality (G.115) implies the claims (G.112)-(G.113). To see this, recall that \(u^{(k)} = g(u^{(k-1)})\). If inequality (G.115) holds, then

\[
\max_{1 \leq i \leq n} \left\{ \left| \frac{u^{(k)}_i}{\eta_i^*} - 1 \right| \right\} = \max_{1 \leq i \leq n} \left\{ \left| \frac{L_i(u^{(k-1)})}{\eta_i^*} - 1 \right| \right\} \leq C \left( \frac{\log(n)}{n^{m-1} \theta_{\max}^m} \right)^{1/2} + C \frac{\log(n)}{n} \max_{1 \leq i \leq n} \left\{ \left| \frac{u^{(k-1)}_i}{\eta_i^*} - 1 \right| \right\} + C \frac{\gamma_n}{n}
\]

(Note that \(u^{(0)} = 0\))

\[
\leq C \left( \frac{\log(n)}{n^{m-1} \theta_{\max}^m} \right)^{1/2} + C \frac{\gamma_n}{n},
\]

Combining this with \(\eta = u^{(\lceil m-1 \rceil)}\), it follows that \(n^{-k} (k = \lceil \frac{m-1}{2} \rceil)\) is a minor term and so \(\max_{1 \leq i \leq n} \left\{ \left| \eta_i/\eta_i^* - 1 \right| \right\} \leq C \log(n)/n^{m-1} \theta_{\max}^m \log(n) + C \gamma_n/n\) (i.e., the claims (G.112)-(G.113)).

Therefore, it is sufficient to show that events (G.114) hold except for a probability \(O(1/n)\) and that inequality (G.115) holds for \(1 \leq k \leq m\) given these events.

First, we show that the events \(E_1\) and \(E_2\) hold with a probability of \(1 - O(1/n)\).

Consider event \(E_1\) first. For \(1 \leq i \leq n\), note that by symmetry,

\[
\sum_{i_2, \ldots, i_m \in I^{(i_1)} \text{ (dist)}} (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}) = \sum_{i_2 < \cdots < i_m \in I^{(i_1)}} (m-1)! (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}),
\]

where the RHS is a sum of \(\left( \binom{n-1}{m-1} \right)\) independent centered Bernoulli random variables with parameters bounded by \(C \theta_{\max}^m\). By Bernstein’s inequality, for any \(t_1 > 0\)

\[
P \left( \sum_{i_2, \ldots, i_m \in I^{(i_1)} \text{ (dist)}} (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}) > t_1 \right) \leq \exp \left( - \frac{t_1^2}{C n^{m-1} \theta_{\max}^m + t_1/3} \right).
\]

Similarly, for event \(E_2\), we have for any \(t_2 > 0\)

\[
P \left( \sum_{i_2, \ldots, i_m \in I^{(i_1)} \text{ (dist)}} (A_{i_1 \cdots i_m} - Q_{i_1 \cdots i_m}) > t_2 \right) \leq \exp \left( - \frac{t_2^2}{C n^m \theta_{\max}^m + t_2/3} \right).
\]
Letting \( t_1 = \sqrt{2C(n^{m-1}\theta_{\text{max}} \log(n))^{1/2}} \) and \( t_2 = (n^{m}\theta_{\text{max}})^{1/2} \) and by direct calculations

\[
P \left( \sum_{i_2, \ldots, i_m \in I^{(t_1)}} (A_{i_1 \ldots i_m} - Q_{i_1 \ldots i_m}) > \sqrt{2C(n^{m-1}\theta_{\text{max}} \log(n))^{1/2}} \right) \leq \exp(-2\log(n)) = O(1/n^2).
\]

and

\[
P \left( \sum_{i_1, \ldots, i_m} (A_{i_1 \ldots i_m} - Q_{i_1 \ldots i_m}) > (n^{m}\theta_{\text{max}})^{1/2} \right) \leq \exp(-n/C) = o(1/n^2).
\]

Combining these with union bound over \( 1 \leq i_1 \leq n \), we see that events \( E_1 \) and \( E_2 \) hold except for a probability \( O(1/n) \).

Next, we show inequality (G.115) when (G.114) is given.

By definitions (G.93) and elementary algebra, \( \eta^* \) can be written as

\[
\eta^* = \frac{\sum_{i_2, \ldots, i_m=1}^{n} Q_{i_1 \ldots i_m}}{(\sum_{i_1, \ldots, i_m=1}^{n} Q_{i_1 \ldots i_m})^{m-1}}.
\]

For \( 1 \leq i_1 \leq n \) and \( 0 \leq k \leq m \), we can then write

\[
\frac{L_{i_1}(u^{(k)})}{\eta^*_i} = (I^{(k)})_{i_1} (II^{(k)})_{i_1}^{-m-1},
\]

where

\[
(I^{(k)})_{i_1} = \frac{\sum_{i_2, \ldots, i_m=1}^{n} A_{i_1 \ldots i_m}}{\sum_{i_1, \ldots, i_m=1}^{n} Q_{i_1 \ldots i_m}} + \sum_{i_2, \ldots, i_m=1}^{n} u^{(k)}_{i_1 \ldots i_m} u^{(k)}_{i_1 \ldots i_m} \frac{u^{(k)}_{i_1 \ldots i_m}}{\eta^*_i}
\]

and

\[
(II^{(k)})_{i_1} = \frac{\sum_{i_1, \ldots, i_m=1}^{n} A_{i_1 \ldots i_m}}{\sum_{i_1, \ldots, i_m=1}^{n} Q_{i_1 \ldots i_m}} + \sum_{i_1, \ldots, i_m=1}^{n} u^{(k)}_{i_1 \ldots i_m} u^{(k)}_{i_1 \ldots i_m} \frac{u^{(k)}_{i_1 \ldots i_m}}{\eta^*_i}
\]

Therefore to show (G.115), by Taylor’s expansion, it is sufficient to show that

\[
\max_{1 \leq i_1 \leq n} \{|(I^{(k)})_{i_1} - 1|\} = o(1), \quad \max_{1 \leq i_1 \leq n} \{|(II^{(k)})_{i_1} - 1|\} = o(1), \quad \text{(G.116)}
\]

\[
\max_{1 \leq i_1 \leq n} \{|(I^{(k)})_{i_1} - 1|\} \leq C \left( \frac{\log(n)}{n^{m-1}\theta_{\text{max}}} \right)^{1/2} + \frac{C}{n} \max_{1 \leq i_1 \leq n} \{|\frac{u^{(k)}_{i_1 \ldots i_m}}{\eta^*_i} - 1|\} + C \frac{\gamma n}{n} \quad \text{(G.117)}
\]

and that

\[
\max_{1 \leq i_1 \leq n} \{|(II^{(k)})_{i_1} - 1|\} \leq C \left( \frac{\log(n)}{n^{m-1}\theta_{\text{max}}} \right)^{1/2} + \frac{C}{n} \max_{1 \leq i_1 \leq n} \{|\frac{u^{(k)}_{i_1 \ldots i_m}}{\eta^*_i} - 1|\} + C \frac{\gamma n}{n} \quad \text{(G.118)}
\]

Note that by triangle’s inequality,

\[
|(I^{(k)})_{i_1} - 1| \leq \left| \sum_{i_2, \ldots, i_m \in I^{(t_1)}} (A_{i_1 \ldots i_m} - Q_{i_1 \ldots i_m}) \right| + \left| \sum_{i_2, \ldots, i_m = 1}^{n} (u^{(k)}_{i_1 \ldots i_m} - \eta^*_i \ldots \eta^*_m) \right| + \left( \sum_{i_2, \ldots, i_m = 1}^{n} (\eta^*_i \ldots \eta^*_m - Q_{i_1 \ldots i_m}) \right).
\]
By event $E_1$ and $Q_{i_1\cdots i_m} \sim \theta_{\max}^m$, the first term on the RHS is $\leq C(n^{m-1}\theta_{\max}^m / \log(n))^{-1/2}$. At the same time, by definitions and elementary algebra, $|\eta_{i_1}^* \cdots \eta_{i_m}^* - Q_{i_1\cdots i_m}| \leq \theta_{i_1} \cdots \theta_{i_m} \gamma_n$. It follows that

$$|(I^{(k)})_{i_1} - 1| \leq C\left(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\right)^{1/2} + C\max_{1 \leq i_1, \ldots, i_m \leq n} \left\{\frac{|u_{i_1}^{(k)} \cdots u_{i_m}^{(k)} - 1|}{\eta_{i_1}^* \cdots \eta_{i_m}^*}\right\} + C\frac{\gamma_n}{n}.$$ (G.119)

Similarly, by event $E_2$ and elementary calculations, we have

$$|(II^{(k)})_{i_1} - 1| \leq C\left(\frac{1}{n^{m-1}\theta_{\max}^m}\right)^{1/2} + C\max_{1 \leq i_1, \ldots, i_m \leq n} \left\{\frac{|u_{i_1}^{(k)} \cdots u_{i_m}^{(k)} - 1|}{\eta_{i_1}^* \cdots \eta_{i_m}^*}\right\} + C\frac{\gamma_n}{n}.$$ (G.120)

Therefore, using Taylor’s expansion on $u_{i_1}^{(k)} \cdots u_{i_m}^{(k)} / (\eta_{i_1}^* \cdots \eta_{i_m}^*)$, to show (G.116)-(G.118), it is sufficient to show that

$$\max_{1 \leq i \leq n} \left\{\frac{|u_i^{(k)} - 1|}{\eta_i^*}\right\} = o(1), \quad 1 \leq k \leq K,$$

where we recall that our original goal is to show

$$\max_{1 \leq i \leq n} \left\{\frac{|L_i(u^{(k)}) - 1|}{\eta_i^*}\right\} \leq C\left(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\right)^{1/2} + C\max_{1 \leq i \leq n} \left\{\frac{|u_i^{(k)} - 1|}{\eta_i^*}\right\} + C\frac{\gamma_n}{n},$$

Noting that $u^{(k)} = g(u^{(k-1)})$. Using induction, we only need to verify that $\max_{1 \leq i \leq n} \{|L_i(u^{(0)}) - \eta_i^* - 1| = o(1)\}. To see this, by $u^{(0)} = 0$, we have

$$\max_{1 \leq i_1, \ldots, i_m \leq n} \left\{\frac{|u_{i_1}^{(0)} \cdots u_{i_m}^{(0)} - 1|}{\eta_{i_1}^* \cdots \eta_{i_m}^*}\right\} = 1 = \max_{1 \leq i \leq n} \left\{\frac{|u_i^{(0)} - 1|}{\eta_i^*}\right\}.$$ Combining this with (G.119)-(G.120), we get (G.116)-(G.118) hold for $k = 0$. It follows that

$$\max_{1 \leq i \leq n} \left\{\frac{|L_i(u^{(0)}) - 1|}{\eta_i^*}\right\} \leq C\max_{1 \leq i \leq n} \{|I^{(0)}_{i_1} - 1\| + C\max_{1 \leq i \leq n} \{|II^{(0)}_{i_1} - 1\| = o(1).$$

This finishes the proof of the claim (b) and the first claim in (c).

Lastly, consider the second claim of (c). Let $G$ be a $m$--way symmetric tensor of dimension $K$ defined by

$$G_{k_1 \cdots k_m} = \beta_{k_1} \cdots \beta_{k_m}, \quad 1 \leq k_1, \ldots, k_m \leq K,$$

and $G$ be the matricization of $G$. By [4, Corollary 7.3.5, Page 451],

$$|\sigma_2(P) - \sigma_2(G)| \leq \|P - G\|,$$ (G.121)

where $\sigma_2(B)$ denotes the second largest singular value of matrix $B$. Note that by definitions, the $k_2 + \sum_{j=3}^{m} K^{j-1}(k_j - 1)$-th column of the matrix $G$ can be written as the following form

$$G_{k_2 + \sum_{j=3}^{m} K^{j-1}(k_j - 1)} = \beta \cdot (\beta_{k_2} \cdots \beta_{k_m}), \quad 1 \leq k_2, \ldots, k_m \leq K.$$ It is seen that $G$ is a rank-one matrix and so $\sigma_2(G) = 0$. Also, by the definition $\sigma_2(P) = |\mu_2|$. Combining these with (G.121) and noting that $\|P - G\| \leq C\max_{1 \leq k_1, \ldots, k_m \leq K} \{|P_{k_1 \cdots k_m} - \beta_{k_1} \cdots \beta_{k_m}|\} = C\gamma_n$, we obtain

$$|\mu_2| \leq \|P - G\| \leq C\gamma_n.$$ By our assumption $|\theta|_{L_1}^m \|\theta\|^2 \|P_{\theta}\|^2 \log(n) \rightarrow \infty$ and $\theta_{\max} \leq C\theta_{\min}$, the above inequality implies $n^{m-1}\theta_{\max}^m \gamma_n \rightarrow \infty$. It follows that

$$n^{m}\theta_{\max}^m \gamma_n \log(n) \rightarrow \infty.$$ This proves the last claim in (c).
References


