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 $\mathcal{A}$ be the adjacency tensor of a uniform and symmetric order-3 hypergraph with n nodes, where

$$\mathcal{A}_{i_1 i_2 i_3} \text{ equals 1 if } i_1, i_2, i_3 \text{ share a hyperedge and equals } 0 \text{ otherwise,} \tag{1.1}$$

for $1 \leq i_1, i_2, i_3$ (distinct) $\leq n$. Since the tensor is symmetric, $\mathcal{A}_{i_1 i_2 i_3} = \mathcal{A}_{j_1 j_2 j_3}$ 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, $\mathcal{A}_{i_1 i_2 i_3} = 0$ 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 $\mathcal{A}$ be a 3-symmetric tensor of n-dimension. First, we call the $n \times n^2$ matrix A the matricization of $\mathcal{A}$, defined by $A_{i,j+n(k-1)} = \mathcal{A}_{ijk}$,*

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

$1 \le i, j, k \le n$. *Second, for* $1 \le k \le n$*, we use* $\mathcal{A}_{::k}$ *to denote the* $n \times n$ *matrix whose row-*i*-and-column-*j *is* $\mathcal{A}_{ijk}$*,* $1 \le i, j \le n$*, and call it the* k*-th slice of* $\mathcal{A}$*. Last, for any* $n \times 1$ *vector* x*, we use* $(\mathcal{A}x)$ *to denote the matrix* $\sum_{k=1}^n x_k \mathcal{A}_{::k}$*, which is an aggregation of the slices.*

Now, suppose the tensor has K perceivable communities. Let $\mathcal{P}$ be a symmetric (3-uniform) tensor of K-dimension that models the community structure, let $\theta = (\theta_1, \theta_2, \ldots, \theta_n)'$ be positive parameters that model the degree heterogeneity of nodes, and $\pi_1, \pi_2, \ldots, \pi_n$ be K-dimensional membership vectors where $\pi_i(k) =$ the weight node i puts on community k, $1 \le k \le K$. We assume for all $1 \le i_1, i_2, i_3 \le n$ that are three distinct indicies, $\mathcal{A}_{i_1 i_2 i_3}$ are independent Bernoulli random variables with $\mathbb{P}(\mathcal{A}_{i_1 i_2 i_3} = 1) = \theta_{i_1}\theta_{i_2}\theta_{i_3} \sum_{k_1,k_2,k_3=1}^K \mathcal{P}_{k_1,k_2,k_3}\pi_{i_1}(k_1)\pi_{i_2}(k_2)\pi_{i_3}(k_3)$, where by Definition 1.1, the right hand side equals to $\theta_{i_1}\theta_{i_2}\theta_{i_3}\pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2}$. Introduce a non-stochastic 3-uniform tensor $\mathcal{Q}$ of n-dimension where $\mathcal{Q}_{i_1 i_2 i_3} = \theta_{i_1}\theta_{i_2}\theta_{i_3}\pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2}$ for $1 \le i_1, i_2, i_3 \le n$. Let $\mathrm{diag}(\mathcal{Q})$ be the tensor with the same size of $\mathcal{Q}$ where $(\mathrm{diag}(\mathcal{Q}))_{i_1 i_2 i_3} = \mathcal{Q}_{i_1 i_2 i_3}$ if i_1, i_2, i_3 are non-distinct, and $(\mathrm{diag}(\mathcal{Q}))_{i_1 i_2 i_3} = 0$ otherwise. It follows that

$$\mathbb{E}(\mathcal{A}) = \mathcal{Q} - \mathrm{diag}(\mathcal{Q}), \qquad \text{where} \quad \mathcal{Q}_{i_1 i_2 i_3} = \theta_{i_1}\theta_{i_2}\theta_{i_3}\pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2}. \tag{1.2}$$

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

$$\mathrm{rank}(P) = K, \quad \text{and} \quad \mathcal{P}_{iii} = 1 \text{ for all } 1 \le i \le K. \tag{1.3}$$

Definition 1.2. *We call (1.1)-(1.3) the tensor-DCMM model for* 3*-uniform hypergraphs. We call* $\mathcal{Q}$ *and* $\mathcal{P}$ *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 π_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

$$H_0 : K = 1 \qquad \text{vs.} \qquad H_1 : K > 1. \tag{1.4}$$

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 hypergraphse (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 $V_0 = \{x \in \mathbb{R}^K : x_k \ge 0, \sum_{k=1}^K x_k = 1\}$, and let F be a K-variate distribution with support contained in V_0. We assume

$$\pi_i \stackrel{iid}{\sim} F; \qquad (\text{let } h = \mathbb{E}_F[\pi_i]). \tag{1.5}$$

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 $\mathrm{supp}(F) = V_0^*$, 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 $\theta = (\theta_1, \theta_2, \ldots, \theta_n)'$. We allow $(\theta, \mathcal{P}, h, F)$ 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 π_i are degenerate (no mixed membership), and $\mathcal{P}_{ijk} = 1$ if $i = j = k$ and $\mathcal{P}_{ijk} = \rho_n$ otherwise, for two parameters (α_n, ρ_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 $\mathcal{P}$, where $\mathcal{P}_{ijk}$ equals to 1, ρ_n, or τ_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 $\mathcal{P}$ is parametrized by 2 or 3 different values.

How to derive a sharp lower bound for global testing (and especially, to identify a *simple quantity* that fully characterizes the lower bound) in our setting is a rather challenging problem. Our model is a non-uniform hypergraph model (see Section 3), which consists of hypergraphs of order $2, 3, \ldots, M$, and each layer consists of many unknown parameters $(\theta, \mathcal{P}, h, F)$. Existing works on lower bounds have been focused on uniform hypergraphs, and non-uniform hypergraphs are much less studied. Even for uniform hypergraphs, existing works have been focused on the the special SBM as in Example 1, not the more general DCMM model. For example, Yuan et al. [22] derived the lower bounds for global testing with the 2-parameter SBM, focusing on the extremely sparse case. Ahn et al. [2] provided lower bound results for exactly recovering the communities (see 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., α_n, ρ_n, τ_n) assumed in their models. For the much broader tensor-DCMM model considered here, we have many parameters $\theta, \mathcal{P}, h, F$ (θ is an n-vector and $\mathcal{P}$ 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-uniform hypergraph first. We first present the standard minimax theory. Define a class of RMM-DCMM models with $K > 1$ and $\mu_2^2\|\theta\|^2\|\theta\|_1 \to 0$ (μ_2 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 χ^2-divergence between the pair converges to 0 as $n \to \infty$. 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 χ^2-divergence of the pair goes to 0 as $n \to \infty$. In detail, we show that in the parameter space $(\theta, \mathcal{P}, h, F)$ of RMM-DCBM, there is a *Region of Impossibility* defined by $\mu_2^2\|\theta\|^2\|\theta\|_1 \to 0$; for *any alternative* in Region of Impossibility, we can pair it with a null such that the χ^2-divergence between the pair goes 0 as $n \to \infty$. 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 $M \geq 2$ and consider a non-uniform hypergraph (e.g., see [10]) that consists of m-uniform hypergraphs for all $m = 2, \ldots, M$, each following a DCMM model with individual $(\theta^{(m)}, \mathcal{P}^{(m)})$ but the common $\pi_1, \ldots, \pi_n$. Let $\ell_m = \|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2(\mu_2^{(m)})^2$ (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 $\ell_m \to 0$, 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 $\max_{2 \leq m \leq M}\{\ell_m\} \to 0$.

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 $\to 0$ if $\max_{2 \leq m \leq M}\{\ell_m\} \geq \log(n)^{1+\delta}$ for a constant $\delta > 0$ (taking $\delta = 0.1$ 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 χ^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 $h = \mathbb{E}_F[\pi_i]$ and that $P \in \mathbb{R}^{K,K^2}$ is the matricization of the tensor $\mathcal{P}$. In this paper, we use $C > 0$ as a generic constant which may vary from occurrence to occurrence. We assume

$$\|P\| \leq C,\ \theta_{\max} \equiv \max\{\theta_1, \ldots, \theta_n\} \leq C,\ \max_{1\leq k\leq K}\{h_k\} \leq C \min_{1\leq k\leq K}\{h_k\},\ \|\theta\|^2\|\theta\|_1 \to \infty. \quad (2.6)$$

The first condition is mild, because the model identifiability already requires that all diagonal entries of $\mathcal{P}$ 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 with 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 χ^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 $\mathcal{Q}$ and $\mathcal{Q}^*$ satisfy

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

here we recall that $(\mathcal{P}\pi_{i_3})$ 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 $\mathcal{P}$ is as in (1.2), and π_i and $h = \mathbb{E}_F[\pi_i]$ are as in (1.5). Recall that $\theta = (\theta_1, \theta_2, \ldots, \theta_n)'$. Similarly, let $\theta^* = (\theta^*_1, \theta^*_2, \ldots, \theta^*_n)'$. 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 π_{i_1}, under the null and alternative are

$$\theta_{i_1}\|\theta\|_1^2 \text{ and } \theta^*_{i_1} \cdot \|\theta^*\|_1^2 \cdot \pi'_{i_1}a, \text{ respectively, where } a = (\mathcal{P}h)h \in \mathbb{R}^K. \quad (2.8)$$

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

$$a = (\mathcal{P}h)h = c_0^3\mathbf{1}_K, \qquad \text{for a scalar } c_0 > 0. \quad (2.9)$$

For broadness, we allow c_0 to depend on n. There are many $(\mathcal{P}, h)$ that satisfy (2.9). For example, in the 2-parameter SBM model as in Example 1 with $h = (1/K, \ldots, 1/K)'$ and $a_k = (1/K^2)[1 + (K^2-1)\rho_n]$ for $1 \leq k \leq K$, (2.9) is satisfied with $c_0^3 = (1/K^2)[1 + (K^2-1)\rho_n]$. Moreover, we choose θ^*_i in the alternative model such that

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

Now, by (2.8)-(2.10), for all $1 \leq i_1 \leq n$, $\theta^*_{i_1}\|\theta^*\|_1^2 \cdot \pi'_{i_1}a = \theta^*_{i_1}\|\theta^*\|_1^2 \cdot c_0^3 = \theta_{i_1}\|\theta\|_1^2$. 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 $\mathcal{P}$ and μ_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*

$\|\theta\|_1\|\theta\|^2\mu_2^2 = o(1)$*. As* $n \to \infty$*, the* χ^2*-divergence* [1] *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* $1 + o(1)$.

Remark *(How degree matching affects the* χ^2*-divergence).* Intuitively speaking, the χ^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 χ^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 χ^2-test loses power. As a result, the second-order term in the χ^2-divergence now dominates and gives the sharp lower bound.

The above least favorable configuration gives rises to the standard minimax theorem. Fix $K \geq 1$ and consider a DCMM model (where π_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

$$\mathcal{M}_n(K, c_0, \alpha_n, \beta_n) = \left\{ \begin{array}{ll} (\theta, \Pi, P): & \|P\| \leq c_0,\ \theta_{\max} \leq c_0,\ \max_{1\leq k\leq K} g_k \leq c_0^{-1}\min_{1\leq k\leq K} g_k, \\ & \|\theta\|_1\|\theta\|^2 \geq 1/\beta_n,\ \|\theta\|_1\|\theta\|^2\mu_2^2 \leq \alpha_n \end{array} \right\}.$$

Here, we assume $\max_{1\leq k\leq K}\{g_k\} \leq C\min_{1\leq k\leq K}\{g_k\}$, so the tensor is balanced. This is similar to the third condition in (2.6), except that π_i's are random there. For the null case, $K = P = \pi_i = 1$, and the above defines a class of θ, which we write for short by $\mathcal{M}_n(1, c_0, \alpha_n, \beta_n) = \mathcal{M}_n^*(\beta_n)$. 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* $n \to \infty$*,* $\inf_\psi\big\{\sup_{\theta\in\mathcal{M}_n^*(\beta_n)}\mathbb{P}(\psi = 1) + \sup_{(\theta,\Pi,P)\in\mathcal{M}_n(K,c_0,\alpha_n,\beta_n)}\mathbb{P}(\psi = 0)\big\} = 1 - o(1)$*, where the infimum is over all possible tests* ψ*.*

2.2 Region of Impossibility for RMM-DCBM

The standard minimax theorem in Section 2.1 only says that in the class of all alternative models with $\|\theta\|^2\|\theta\|_1\mu_2^2 \to 0$, *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 $\mathcal{Q}$ and $\mathcal{Q}^*$, respectively. We assume for all $1 \leq i_1, i_2, i_3 \leq n$,

$$\mathcal{Q}_{i_1i_2i_3} = \theta_{i_1}\theta_{i_2}\theta_{i_3}, \qquad \mathcal{Q}^*_{i_1i_2i_3} = \theta^*_{i_1}\theta^*_{i_2}\theta^*_{i_3}\pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2}. \tag{2.11}$$

Here, the community structure tensor $\mathcal{P}$ is as in (1.2), π_i and $h = \mathbb{E}_F[\pi_i]$ are as in (1.5), and $\mathrm{supp}(F) = \{e_1, \ldots, e_K\}$. Similarly, the goal is degree-matching: for any $(\theta, \mathcal{P}, h, F)$, we construct θ^* 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 $(\mathcal{P}, h)$ satisfies $a \equiv (\mathcal{P}h)h = c_0^3\mathbf{1}_K$ 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 $(\mathcal{P}, h)$, we don't have $a \equiv (\mathcal{P}h)h \propto \mathbf{1}_K$, 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 $D = \mathrm{diag}(d_1, \ldots, d_K)$ with $d_k > 0, 1 \leq k \leq K$, define a K-dimensional vector a^D by ($\mathcal{P}^D$ is a 3-tensor in dimension K)

$$a^D = (\mathcal{P}^D h)h, \qquad \text{with} \quad \mathcal{P}^D_{k_1k_2k_3} = d_{k_1}d_{k_2}d_{k_3}\mathcal{P}_{k_1k_2k_3}, \quad 1 \leq k_1, k_2, k_3 \leq K.$$

We aim to select a matrix D such that $a^D = \mathbf{1}_K$. 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 χ^2-divergence between two models, $f_0(\mathcal{A})$ and $f_1(\mathcal{A})$, is defined as $\int[(f_0(\mathcal{A}) - f_1(\mathcal{A}))^2/f_0(\mathcal{A})]d\mathcal{A}$. In our setting, the alternative model $f_1(\mathcal{A})$ alone involves an integral over the distribution of π_i in (1.5)

Lemma 2.1. *Fix $K > 1$ and let $\mathcal{P}$, 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 $D = \mathrm{diag}(d_1, d_2, \ldots, d_K)$ such that $a^D = \mathbf{1}_K$.*

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

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

Combining it with (2.11), we have $\mathcal{Q}^*_{i_1 i_2 i_3} = \theta^*_{i_1}\theta^*_{i_2}\theta^*_{i_3} \cdot \pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2} = \theta_{i_1}\theta_{i_2}\theta_{i_3} \cdot \pi'_{i_1}(\mathcal{P}^D\pi_{i_3})\pi_{i_2}$. 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

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

By Lemma 2.1, $a^D = \mathbf{1}_K$. 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 $\mathcal{P}$ and μ_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 $(\theta, \mathcal{P}, h, F)$, consider a pair of models, a null and an alternative with K communities, as in (2.11), where θ_i^* are given by (2.12) with the matrix D as in Lemma 2.1. Suppose (2.6) holds and $\|\theta\|_1\|\theta\|^2\mu_2^2 = o(1)$. As $n \to \infty$, the χ^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 $(\theta, \mathcal{P}, h, F)$ for DCBM, we call the region prescribed by $\|\theta\|_1\|\theta\|^2\mu_2^2 \to 0$ the *Region of Impossibility*: 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 $\|\theta\|_1\|\theta\|^2\mu_2^2/\log^{1.1}(n) \to \infty$ 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 Impossibility and Region of Possibility, except for a $\log(n)$ term, our lower bounds are tight.

2.3 Region of Impossibility for RMM-DCMM

The discussion for RMM-DCMM 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-DCMM, 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 Impossibility 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 $\theta_1 = \ldots = \theta_n$), 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 $\mathcal{Q}$ and $\mathcal{Q}^*$ under two hypotheses are such that, for $1 \leq i_1, i_2, i_3 \leq n$,

$$\mathcal{Q}_{i_1 i_2 i_3} = \alpha_n, \qquad \mathcal{Q}^*_{i_1 i_2 i_3} = \pi'_{i_1}(\mathcal{P}\pi_{i_3})\pi_{i_2}. \tag{2.13}$$

Same as before, π_i are iid generated from a distribution F on $\{e_1, e_2, \ldots, e_K\}$ with $h = \mathbb{E}_F[\pi_i]$. Let $\hat{\alpha}_n = (n(n-1)(n-2))^{-1}\mathbf{1}_n'(\mathcal{A}\mathbf{1}_n)\mathbf{1}_n$, $\eta = (1/2)(\mathcal{A}\mathbf{1}_n)\mathbf{1}_n$, and $\bar{\eta}$ be the mean of $\eta_1, \eta_2, \ldots, \eta_K$. Consider the centered-$\chi^2$-statistic

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

Lemma 2.2. *Consider the global testing problem (1.4) under the SBM model (2.13) for H_0 and H_1, respectively. Let $\widetilde{\alpha}_n = \mathbb{E}[\hat{\alpha}_n]$, $h = (1/n)\sum_{i=1}^n \pi_i$ and $\Sigma = \frac{1}{n}\sum_{i=1}^n (\pi_i - h)(\pi_i - h)'$. Let $\lambda_{K-1}(\Sigma)$ be the $(K-1)$-th largest eigenvalue (in magnitude) of Σ. Assume $\alpha_n \leq c_0$, $\max_{1 \leq i,j,k \leq K}\{\mathcal{P}_{ijk}\} \leq c_0$, $n^2\alpha_n \to \infty$, and $n^2\widetilde{\alpha}_n \to \infty$. Also, assume $\min\{h_1, h_2, \ldots, h_K\} \geq C$ and $\lambda_{K-1}(\Sigma) \geq C$. Write $\delta_n = \|\widetilde{\alpha}_n^{-1}(I_K - H_K)(\mathcal{P}h)h\|$, where $H_K = (1/K)\mathbf{1}_K\mathbf{1}_K'$ 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 $n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2 \to \infty$*

By Lemma 2.2, the power of the χ^2-test hinges on δ_n

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

Note that $\delta_n = 0$ if and only if $(\mathcal{P}h)h \propto \mathbf{1}_K$. We call the cases of $(\mathcal{P}h)h \propto \mathbf{1}_K$ and $(\mathcal{P}h)h \not\propto \mathbf{1}_K$ the *symmetric* case and the *asymmetric* case, respectively. For symmetric SBM, $\delta_n = 0$, and we do not expect the χ^2-test to have power. However, for asymmetric SBM, the χ^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 ρ_0 if $i = j = k$ and ρ_1 otherwise. Suppose $n^2\rho_0 \to \infty$ 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 ρ_1/ρ_0. In this case, we have $\|\theta\|_1\|\theta\|^2 \asymp n^2\rho_0$, $\tau_n \equiv \|\theta\|_1\|\theta\|^2\mu_2^2 \asymp n^2\rho_0^{-1}(\rho_1 - \rho_0)^2$, and $n^{3/2}\widetilde{\alpha}_n\delta_n^2 \asymp (n\rho_0)^{-1/2}(2a - 1)^2\tau_n$. In the symmetric case where $a = 1/2$, the χ^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 $n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2 \to \infty$ (so χ^2-test has asymptotically full power). Here, the interesting range of ρ_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 $(\mathcal{P}h)h \propto \mathbf{1}_K$. Our lower bound restricted to symmetric SBM agrees with those in the literature. The asymmetric case $(\mathcal{P}h)h \not\propto \mathbf{1}_K$ 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 naive 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 $[\mathcal{M}; b_1, \cdots, b_m]$ denote the summation $\sum_{1 \leq k_1, k_2, \ldots, k_m \leq K}[\mathcal{M}_{k_1k_2\ldots k_m}b_1(k_1)b_2(k_2)\cdots b_m(k_m)]$.*

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 $\mathcal{A}^{(m)}$ be the adjacency tensor of the order-m hypergraph (i.e., $\mathcal{A}^{(m)}_{i_1i_2\cdots i_m} = 1$ if $\{i_1, i_2, \ldots, i_m\}$ is a hyper-edge and 0 otherwise). As before, we model $\mathcal{A}^{(m)}$ with the tensor-DCMM model. Let $\mathcal{P}^{(m)}$ be a symmetric order-m tensor in dimension K, and let $\theta^{(m)} = (\theta^{(m)}_1, \theta^{(m)}_2, \ldots, \theta^{(m)}_n)'$ 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 $\{\mathcal{A}^{(m)}_{i_1i_2\cdots i_m}\}_{1\leq i_1<i_2<\ldots<i_m\leq n}$ are independent Bernoulli variables, where the Bernoulli probabilities are specified by the tensor $\mathcal{Q}^{(m)}$, given by

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

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

Definition 3.2. *We say that $\mathcal{A}[M] = \{\mathcal{A}^{(2)}, \ldots, \mathcal{A}^{(M)}\}$ follows a (general non-uniform) tensor-DCMM model if $\{\mathcal{A}^{(m)}\}_{2\leq m\leq M}$ are independent of each other and each $\mathcal{A}^{(m)}$ 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 $\theta_1^{(m)} = \theta_2^{(m)} = \cdots = \theta_n^{(m)}$ 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 π_i is degenerate); the study of tensor-DCMM is similar. Fix $K > 1$. Consider a DCBM null model with probability tensors $\mathcal{Q}[M] = \{\mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(M)}\}$ and an RMM-DCBM model with probability tensors $\mathcal{Q}^*[M] = \{\mathcal{Q}^{*(2)}, \ldots, \mathcal{Q}^{*(M)}\}$, where for every $2 \leq m \leq M$ and $1 \leq i_1, i_2, \ldots, i_m \leq n$,

$$\mathcal{Q}^{(m)}_{i_1,i_2,\ldots,i_m} = \theta^{(m)}_{i_1}\theta^{(m)}_{i_2}\cdots\theta^{(m)}_{i_m}, \tag{3.15}$$

$$\mathcal{Q}^{*(m)}_{i_1,i_2,\ldots,i_m} = \theta^{*(m)}_{i_1}\cdots\theta^{*(m)}_{i_m} \times [\mathcal{P}^{(m)}; \pi_{i_1}, \ldots, \pi_{i_m}], \qquad \pi_i \overset{iid}{\sim} F. \tag{3.16}$$

The support of F is in $V_0^* = \{e_1, e_2, \ldots, e_K\}$. Let $h = \mathbb{E}_F[\pi_i]$ 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 $D^{(m)} = \mathrm{diag}(d^{(m)}_1, d^{(m)}_2, \ldots, d^{(m)}_K)$ such that

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

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

$$\theta^{*(m)}_i = d^{(m)}_k\theta^{(m)}_i, \quad \text{if node } i \text{ belongs to community } k. \tag{3.17}$$

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

$$\ell_m = \|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2(\mu^{(m)}_2)^2.$$

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 $\{(\theta^{(m)}, \mathcal{P}^{(m)})\}_{2\leq m\leq M}$, 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 χ^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 $\theta^{(m)}_i \equiv \alpha^{(m)}_n$ and $\mathcal{P}^{(m)}$ 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 ℓ_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 $2 \le m \le M$, 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

$$L_i(u) = \frac{\sum_{i_2,\dots,i_m(\text{distinct})} \mathcal{A}^{(m)}_{i\,i_2\dots i_m} + \sum_{i_2,\dots,i_m(\text{non-distinct})} u_i u_{i_2}\cdots u_{i_m}}{\left(\sum_{i_1,\dots,i_m(\text{distinct})} \mathcal{A}^{(m)}_{i_1 i_2\dots i_m} + \sum_{i_1,\dots,i_m(\text{non-distinct})} u_{i_1}\cdots u_{i_m}\right)^{(m-1)/m}}. \tag{3.18}$$

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

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

$$Q^{(m)}_n = \max_{S=(S_1,\dots,S_{m+1})\in B}\ \max_{1\le k_1,\dots,k_m\le m+1} \left\{ \Big| \sum_{i_1\in S_{k_1},\dots,i_m\in S_{k_m}(\text{distinct})} \mathcal{A}^{*(m)}_{i_1\dots i_m} \Big| \right\}. \tag{3.19}$$

Finally, we combine $Q^{(2)}_n, \dots, Q^{(M)}_n$. For each m, let $V^{(m)}_n = \binom{n}{m}\hat{\alpha}_n(1 - \hat{\alpha}_n)$, where $\hat{\alpha}_n = [(n-m)!/n!] \sum_{i_1,\dots,i_m=1}^n \mathcal{A}^{(m)}_{i_1\dots i_m}$. The test statistic is

$$\phi_n = \max_{2\le m\le M} \{ Q^{(m)}_n / [n \log(n)^{1.1} V^{(m)}_n]^{1/2} \}. \tag{3.20}$$

Recall that by Theorem 3.1, for any alternative with $\max_{2\le m\le M}\{\ell_m\} = o(1)$, we can pair it with a null so that the pair are asymptotically indistinguishable. The next theorem says that the proposed test statistic ϕ_n can successfully separate any alternative satisfying $\max_{2\le m\le M}\{\ell_m\} \gg \log^{1+\delta}(n)$ (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 $\ell_m = \|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2(\mu^{(m)}_2)^2$. Let $\theta^{(m)}_{\max}$ and $\theta^{(m)}_{\min}$ be the maximum and minimum entry of the vector $\theta^{(m)}$, respectively.

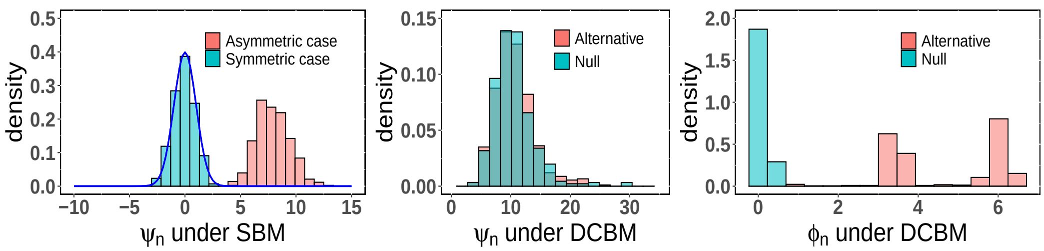
Theorem 3.2 (Tightness of lower bounds)**.** *Consider the general tensor-DCBM model with $M \ge 2$. Let $h = \frac{1}{n}\sum_{i=1}^n \pi_i$, and let $P^{(m)} \in \mathbb{R}^{K, K^{m-1}}$ be the matricization of $\mathcal{P}^{(m)}$. Suppose $\|P^{(m)}\| \le C$, $\max_{1\le k\le K}\{h_k\} \le C \min_{1\le k\le K}\{h_k\}$, $\theta^{(m)}_{\max} \le C\theta^{(m)}_{\min}$, $\theta^{(m)}_{\max} \le c_0$ for a constant $c_0 < 1$, and $\|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2/\log(n) \to \infty$, for every $2 \le m \le M$. Then, $\phi_n \to 0$ in probability, if H_0 holds, and $\phi_n \to \infty$ in probability, if H_1 holds and $\max_{2\le m\le M}\{\ell_m\}/[\log^{1.1}(n)] \to \infty$.*

To speed up the computation of ϕ_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 $\mathcal{A}^{(m)}$, 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 $Q^{(m)}_n$; the test statistic ϕ_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 $1 \le i \le n$, and $\mathcal{P}_{ijk} = 1$ if $i = j = k$ and $\mathcal{P}_{ijk} = 1/4$ 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 χ^2-statistic ψ_n in Section 2.4, and repeat for 500 times. The histograms of ψ_n for two cases are on the left panel of the figure below (green: symmetric alternative; red: asymmetric alternative; blue: density of $N(0,1)$). By Lemma 2.2, $\psi_n \approx N(0,1)$ in the null. Hence, the results suggest that ψ_n is unable to distinguish the symmetric alternative from the null but can distinguish the asymmetric alternative from the null.

In **Experiment 2**, we consider DCBM and use the least-favorable configuration in Section 2.2 where the degree matching strategy is employed. We verify that a degree-based test such as ψ_n indeed has no power. Let (n, K, m) be the same as above. In the null, we let θ_i be iid drawn from Pareto$(0.5, 5)$ and then re-normalize the vector of θ so that $n^{-2}\|\theta\|_1\|\theta\|^2 = c_n$, for $c_n = n^{-1}$; in the alternative, let $(\mathcal{P}^*, h)$ be the same as in the asymmetric case in Experiment 1 and generate θ_i^* is as in (2.12) (D is obtained by treating $D(\mathcal{P}(Dh))Dh = 1_K$ as a nonlinear equation and apply the Matlab function `solve`). The histograms of ψ_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 ϕ_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 ϕ_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 ϕ_n successfully distinguishes the alternative from the null. This validates our result in Theorem 3.2. The distribution of ϕ_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 χ^2-divergence. We discover that for an m-uniform hypergraph, the Regions of Impossibility/Possibility are governed by the simple quantity of $\ell_m = \|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2(\mu_2^{(m)})^2$ (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

[1] Sameer Agarwal, Jongwoo Lim, Lihi Zelnik-Manor, Pietro Perona, David Kriegman, and Serge Belongie. Beyond pairwise clustering. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 838–845. IEEE, 2005.

[2] Kwangjun Ahn, Kangwook Lee, and Changho Suh. Community recovery in hypergraphs. *IEEE Transactions on Information Theory*, 65(10):6561–6579, 2019.

[3] Ravindra Bapat. D_1AD_2 theorems for multidimensional matrices. *Linear Algebra and its Applications*, 48:437–442, 1982.

[4] Javier Béjar, Sergio Álvarez, Dario García, Ignasi Gómez, Luis Oliva, Arturo Tejeda, and Javier Vázquez-Salceda. Discovery of spatio-temporal patterns from location-based social networks. *J. Exp. Theor. Artif. Intell.*, 28(1-2):313–329, 2016.

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

[6] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

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

[8] Shmuel Friedland. Positive diagonal scaling of a nonnegative tensor to one with prescribed slice sums. *Linear Algebra and its Applications*, 434(7):1615–1619, 2011.

[9] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. *Advances in Neural Information Processing Systems*, 27:397–405, 2014.

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

[11] Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210, 2017.

[12] Pengsheng Ji, Jiashun Jin, Zheng Tracy Ke, and Wanshan Li. Co-citation and co-authorship networks of statisticians. *Journal of Business & Economic Statistics, to appear*, 2021.

[13] Jiashun Jin. Fast community detection by SCORE. *The Annals of Statistics*, 43(1):57–89, 2015.

[14] Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Optimal adaptivity of signed-polygon statistics for network testing. *Annals of Statistics, to appear*, 2021.

[15] Zheng Tracy Ke, Feng Shi, and Dong Xia. Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*, 2019.

[16] Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach. *arXiv:1807.02884*, 2018.

[17] Sungwoong Kim, Sebastian Nowozin, Pushmeet Kohli, and Chang Yoo. Higher-order correlation clustering for image segmentation. *Advances in Neural Information Processing Systems*, 24:1530–1538, 2011.

[18] Lei Li and Tao Li. News recommendation via hypergraph learning: encapsulation of user behavior and news content. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pages 305–314, 2013.

[19] Jiajun Liang, Chuyang Ke, and Jean Honorio. Information theoretic limits of exact recovery in sub-hypergraph models for community detection. *IEEE International Symposium on Information Theory*, 2021.

[20] Chung-Yi Lin, I Eli Chien, and I-Hsiang Wang. On the fundamental statistical limit of community detection in random hypergraphs. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2178–2182. IEEE, 2017.

[21] Cammarata Louis and Zheng Tracy Ke. Power enhancement and phase transitions for global testing of the mixed membership stochastic block model. *Manuscript*, 2021.

[22] Mingao Yuan, Ruiqi Liu, Yang Feng, and Zuofeng Shang. Testing community structures for hypergraphs. *Annals of Statistics, to appear*, 2021.

[23] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. *Advances in Neural Information Processing Systems*, 19:1601–1608, 2006.

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 $\mathcal{Q}$ and $\mathcal{Q}^*$, respectively. Similarly, for $1 \le i_1, i_2, i_3 \le n$, we assume

$$\mathcal{Q}_{i_1 i_2 i_3} = \theta_{i_1}\theta_{i_2}\theta_{i_3}, \tag{A.1}$$

$$\mathcal{Q}^*_{i_1 i_2 i_3} = \theta^*_{i_1}\theta^*_{i_2}\theta^*_{i_3} \cdot \pi_{i_1}'(\mathcal{P}\pi_{i_3})\pi_{i_2}, \tag{A.2}$$

where the community structure tensor $\mathcal{P}$ is as in (1.2), and π_i and $h = \mathbb{E}_F[\pi_i]$ are as in (1.5). Similarly, for any matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_K)$ with $d_k > 0$, $1 \le k \le K$, let $\mathcal{P}^D$ be the tensor with the same size of $\mathcal{P}$ satisfying $\mathcal{P}^D_{k_1k_2k_3} = d_{k_1}d_{k_2}d_{k_3}\mathcal{P}_{k_1k_2k_3}$. Also, let $h^D = \mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$ and $\tilde{a}^D = (\mathcal{P}^D h^D)h^D$. We assume that there is a matrix D such that

$$\tilde{a}^D = \mathbf{1}_K, \qquad \min_{1\le k\le K}\{h^D_k\} \ge C. \tag{A.3}$$

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 π_i is degenerate (i.e., only one entry is 1, all other entries are 0). In this case, $h^D = \mathbb{E}_F[\pi_i] = h$, 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 $\mathcal{P}^*$ and $\tilde{a}^*$ be $\mathcal{P}^D$ and $\tilde{a}^D$ evaluated at $D = D_0$, respectively. By definitions, $\tilde{a}^* = \mathbf{1}_K$. For $1 \le i \le n$, 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 $1 \le i_1, i_2, i_3 \le n$, we have $\mathcal{Q}^*_{i_1i_2i_3} = \theta^*_{i_1}\theta^*_{i_2}\theta^*_{i_3}\cdot\pi_{i_1}'(\mathcal{P}\pi_{i_3})\pi_{i_2} = \theta_{i_1}\theta_{i_2}\theta_{i_3}{\pi^*_{i_1}}'(\mathcal{P}^*\pi^*_{i_3})\pi^*_{i_2}$. By similar calculations, for $1 \le i_1 \le n$, the leading term of the expected degree of node i_1 under the alternative is $\theta_{i_1}\|\theta\|_1^2(\pi^*_{i_1})'\tilde{a}^* = \theta_{i_1}\|\theta\|_1^2$, 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 $(\theta, \mathcal{P}, h, F)$, 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 θ^* is given by (A.4). Suppose (2.6) hold and $\|\theta\|_1\|\theta\|^2\mu_2^2 = o(1)$. As $n \to \infty$,* *the χ^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 DCMM, 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 $\mathcal{Q}[M] = \{\mathcal{Q}^{(2)}, \ldots, \mathcal{Q}^{(M)}\}$ and an RMM-DCMM model with probability tensors $\mathcal{Q}^*[M] = \{\mathcal{Q}^{*(2)}, \ldots, \mathcal{Q}^{*(M)}\}$, where for every $2 \le m \le M$ and $1 \le i_1, i_2, \ldots, i_m \le n$,

$$\mathcal{Q}^{(m)}_{i_1,i_2,\ldots,i_m} = \theta^{(m)}_{i_1}\theta^{(m)}_{i_2}\cdots\theta^{(m)}_{i_m}, \tag{A.5}$$

$$\mathcal{Q}^{*(m)}_{i_1,i_2,\ldots,i_m} = \theta^{*(m)}_{i_1}\cdots\theta^{*(m)}_{i_m} \times [\mathcal{P}^{(m)}; \pi_{i_1}, \ldots, \pi_{i_m}], \qquad \pi_i \overset{iid}{\sim} F. \tag{A.6}$$

For any matrix $D^{(m)} = \mathrm{diag}(d^{(m)}_1, d^{(m)}_2, \ldots, d^{(m)}_K)$ with $d^{(m)}_k > 0$, $1 \le k \le K$, let $\widetilde{\mathcal{P}}^{(m)}$ be the tensor with the same size of $\mathcal{P}^{(m)}$ satisfying $\widetilde{\mathcal{P}}^{(m)}_{k_1k_2\cdots k_m} = d^{(m)}_{k_1}d^{(m)}_{k_2}\cdots d^{(m)}_{k_m}\mathcal{P}^{(m)}_{k_1k_2\cdots k_m}$. Also, let $\widetilde{h}^{(m)} = \mathbb{E}[{D^{(m)}}^{-1}\pi_i/\|{D^{(m)}}^{-1}\pi_i\|_1]$ and $\widetilde{a}^{(m)} = \sum_{1\le i_2,\ldots,i_m\le K} d^{(m)}_{i_1}\cdot\mathcal{P}^{(m)}_{i_1\cdots i_m}\cdot(d^{(m)}_{i_2}\widetilde{h}^{(m)}_{i_2})\cdots(d^{(m)}_{i_m}\widetilde{h}^{(m)}_{i_m})$, for every $1 \le i_1 \le K$. We assume that there are matrices $D^{(2)}, \ldots, D^{(m)}$ such that for $m = 2, \ldots, M$

$$\widetilde{a}^{(m)} = \mathbf{1}_K, \qquad \min_{1\le k\le K}\{\widetilde{h}^{(m)}_k\} \ge C. \tag{A.7}$$

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 $\mathcal{P}^{*(m)}$ and $\widetilde{a}^{*(m)}$ be $\widetilde{\mathcal{P}}^{(m)}$ and $\widetilde{a}^{(m)}$ evaluated at $D^{(m)} = D^{(m)}_0$, respectively. By definitions, $\widetilde{a}^{*(m)} = \mathbf{1}_K$. For $1 \le i \le n$, $2 \le m \le M$, let

$$\theta^{*(m)}_i = \theta^{(m)}_i/\|{D^{(m)}_0}^{-1}\pi_i\|_1, \quad \pi^{*(m)}_i = {D^{(m)}_0}^{-1}\pi_i/\|{D^{(m)}_0}^{-1}\pi_i\|_1. \tag{A.8}$$

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

Theorem A.2 (Impossibility for non-uniform RMM-DCMM)**.** *Fix $K > 1$ and $M \ge 2$. For any given (h, F) and $\{(\theta^{(m)}, \mathcal{P}^{(m)})\}_{2\le m\le M}$, 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 $\{\theta^{*(m)}_i\}_{1\le i\le n, 2\le m\le M}$ are as in (A.8). Suppose $\|P^{(m)}\| \le C$ and $\max_{1\le i\le n}\theta^{(m)}_i \le C$. If $\max_{2\le m\le M}\{\ell_m\} = o(1)$, then as $n \to \infty$, the χ^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 $\mathcal{Q}_{i_1i_2i_3} = \theta_{i_1}\theta_{i_2}\theta_{i_3}$ and a K-community uniform RMM-DCMM model as in Theorem A.1. Let $\mathcal{P}^{(n)}_0$ and $\mathcal{P}^{(n)}_1$ denote the probability measures associated with these two models, respectively. We further modify $\mathcal{P}^{(n)}_1$ as follows. In this RMM-DCMM, the membership matrix Π is randomly generated. Let Π_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

$$\widetilde{\Pi} = \begin{cases} \Pi, & \text{if } (\theta, \Pi, \mathcal{P}) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n), \\ \Pi_0, & \text{otherwise.} \end{cases}, \qquad \text{where } \pi_i \overset{iid}{\sim} F. \tag{B.9}$$

We construct a similar RMM-DCMM by replacing Π with $\widetilde{\Pi}$ and denote $\widetilde{P}_1^{(n)}$ 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 $\widetilde{P}_1^{(n)}$ under the alternative hypothesis. Given any test ψ, its sum of type I and type II errors is equal to

$$\begin{aligned} & \mathcal{P}_0^{(n)}(\psi=1)+\widetilde{\mathcal{P}}_1^{(n)}(\psi=0) \\ = \; & \mathbb{P}_0(\psi=1)+\mathbb{E}_{\widetilde{\Pi}}\big[\mathbb{P}_1\big(\psi=0|\widetilde{\Pi}\big)\big] \\ \leq \; & \sup_{\theta\in\mathcal{M}_n^*(\beta_n)} \mathbb{P}(\psi=1)+\sup_{(\theta,\Pi,\mathcal{P})\in\mathcal{M}_n(K,c_0,\alpha_n,\beta_n)} \mathbb{P}(\psi=0). \end{aligned}$$

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

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

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

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

We now show (B.10). Recall that in Theorem A.1 we have seen that the χ^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 χ^2-divergence (e.g., equation (2.27) of [5]), we have

$$\begin{aligned} \|\mathcal{P}_0^{(n)}-\widetilde{\mathcal{P}}_1^{(n)}\|_1 &\leq \|\mathcal{P}_0^{(n)}-\mathcal{P}_1^{(n)}\|_1+\|\mathcal{P}_1^{(n)}-\widetilde{\mathcal{P}}_1^{(n)}\|_1 \\ &\leq \sqrt{\chi^2(\mathcal{P}_0^{(n)},\mathcal{P}_1^{(n)})}+\|\mathcal{P}_1^{(n)}-\widetilde{\mathcal{P}}_1^{(n)}\|_1 \\ &\leq o(1)+\|\mathcal{P}_1^{(n)}-\widetilde{\mathcal{P}}_1^{(n)}\|_1. \end{aligned} \tag{B.11}$$

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

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

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

Lemma B.1. *Fix a constant $c_0\geq 1$. As $n\to\infty$, suppose $\|P\|\leq c_0$, $\theta_{\max}\leq c_0$, and $\|\theta\|_1\to\infty$. Write $h=\mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$. 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 $\|\mathcal{P}_1^{(n)}-\widetilde{\mathcal{P}}_1^{(n)}\|_1\to 0$. 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

$$\min_k\{\sum_{i=1}^n\theta_i\pi_i(k)\}\geq c_0\,\|\theta\|_1. \tag{B.12}$$

Let c_1 be a constant such that $c_1 > c_0$. Our assumptions say that $\min_{1\le k\le K}\{h_k\} \ge c_1$, where $h = \mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$. Let $h^* = \mathbb{E}[\pi_i]$. We first show that $\min_{1\le k\le K}\{h_k\} \ge c_1$ implies $\min_{1\le k\le K}\{h_k^*\} \ge c_1 \cdot [1 + o(1)]$. By Lemma E.5 in section E, we have

$$\max_{1\le i\le K}\{|d_i - 1|\} \le C\mu_2 \qquad \text{with } \mu_2 = o(1),$$

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

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

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

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

$$\mathbb{P}(|X| > t) \le \exp\Big(-\frac{t^2}{C\|\theta\|^2 + C\theta_{\max}t}\Big), \qquad \text{for any } t > 0.$$

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

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

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

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

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

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 $m = 3$.

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 $\mathcal{P}_{i\ldots i} = 1$, for $i = 1, \ldots, K$ and $\min\{h_1, h_2, \ldots, h_K\} \ge C$. There exists an unique diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_K)$ such that*

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

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 $\widetilde{d}_i = d_ih_i$ for $i = 1, \ldots, K$. It is equivalent to find an unique diagonal matrix $\widetilde{D} = diag(\widetilde{d}_1, \ldots, \widetilde{d}_K)$ with strictly positive entries such that

$$\sum_{i_2,\ldots,i_m=1}^K \widetilde{d}_{i_1}\mathcal{P}_{i_1\cdots i_m}\widetilde{d}_{i_2}\cdots\widetilde{d}_{i_m} = h_{i_1}, \qquad \text{for all } i_1 = 1, \ldots, K. \tag{C.14}$$

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

$$\sum_{i_2,\ldots,i_m=1}^K x_{i_1}\mathcal{P}_{i_1\cdots i_m}x_{i_2}\cdots x_{i_m} = h_{i_1}, \qquad \text{for all } i_1 = 1, \ldots, K. \tag{C.15}$$

which gives the existence of such $\widetilde{D}$ satisfying (C.14).

The uniqueness of such $\widetilde{D}$ is given by the Theorem 1.1 in [2] which states that there is an unique tensor $\mathcal{A}$ that is defined by $\mathcal{A}_{i_1\cdots i_m} = \widetilde{d}_{i_1}\mathcal{P}_{i_1\cdots i_m}\widetilde{d}_{i_2}\cdots\widetilde{d}_{i_m}$ for $i_1,\ldots,i_m = 1,\ldots,K$ and satisfies

$$\sum_{i_2,\ldots,i_m=1}^{K} \mathcal{A}_{i_1\cdots i_m} = h_{i_1}, \qquad \text{for all } i_1 = 1,\ldots,K. \tag{C.16}$$

Therefore, $\widetilde{D}$ is unique since $\mathcal{A}$ is unique and one-to-one correpondence with $\widetilde{D}$. 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 θ_i^* is constructed as $\theta_i^* = \theta_i/\|D^{-1}\pi_i\|_1$ (note that when π_i is degenerate, this reduces to the construction of $\theta_i^* = \theta_i d_k$ 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,π_i be weight vectors in $\mathbb{R}^K$ and $y_i = \pi_i - h$, for $1 \le i \le n$. Then*

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

where

$$x^{(m)} = [\mathcal{M};h,\ldots,h] + \sum_{s=1}^{m}[\mathcal{M};\underbrace{h,\ldots,h}_{s-1},y_s,\underbrace{h,\ldots,h}_{m-s}],$$

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

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

$$S = \sum_{i_1,\ldots,i_m(dist)} \frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a_t}[\mathcal{M};y_{i_1},y_{i_2},w_{i_3}^{(3)},\ldots,w_{i_m}^{(m)}][\mathcal{M};\widetilde{y}_{i_1},\widetilde{y}_{i_2},\widetilde{w}_{i_3}^{(3)},\ldots,\widetilde{w}_{i_m}^{(m)}].$$

Then, for any constant c independent of n,

$$\mathbb{E}\Big[\exp(cS)\Big] \le \mathbb{E}\Big[\exp\Big(C\mu^2\|\theta\|_t^{t(m-2)}|T|/a_t\Big)\Big]\cdot\exp(C\mu^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}/a_t),$$

where T is a random variable satisfying $\mathbb{P}(|T|>x)\le 2\exp(-x/(2K^2\|\theta\|_{2t}^{2t}))$, for $x>0$.

Lemma E.3. *With the same setting in Lemma E.2, denote*

$$S=\sum_{i_1,\dots,i_m(dist)}\frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a_t}[\mathcal{M};y_{i_1},y_{i_2},w_{i_3}^{(3)},\dots,w_{i_m}^{(m)}][\mathcal{M};\widetilde{y}_{i_1},h,\widetilde{y}_{i_3},\widetilde{w}_{i_4}^{(4)},\dots,\widetilde{w}_{i_m}^{(m)}].$$

Then, for any constant c independent of n,

$$\mathbb{E}\Big[\exp(cS)\Big] \le \mathbb{E}\Big[\exp\Big(C\mu^2\|\theta\|_t^{t(m-2)}|T|/a_t\Big)\Big]\cdot\exp(C\mu^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}/a_t),$$

where T is a random variable satisfying $\mathbb{P}(|T|>x)\le 4\exp(-x/(2K^2\|\theta\|_{2t}^{2t}))$, for $x>0$.

Lemma E.4. *With the same setting in Lemma E.2, denote*

$$S=\sum_{i_1,\dots,i_m(dist)}\frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a_t}[\mathcal{M};y_{i_1},y_{i_2},w_{i_3}^{(3)},\dots,w_{i_m}^{(m)}][\mathcal{M};h,h,\widetilde{y}_{i_3},\widetilde{y}_{i_4},\widetilde{w}_{i_5}^{(5)},\dots,\widetilde{w}_{i_m}^{(m)}].$$

Then, for any constant c independent of n,

$$\mathbb{E}\Big[\exp(cS)\Big] \le \mathbb{E}\Big[\exp\Big(C\mu^2\|\theta\|_t^{t(m-2)}|T|/a_t\Big)\Big]\cdot\exp(C\mu^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}/a_t),$$

where T is a random variable satisfying $\mathbb{P}(|T|>x)\le 4\exp(-x/(2K\|\theta\|_{2t}^{2t}))$, for $x>0$.

Lemma E.5. *Under the conditions of Theorem 3.1, for $m=2,\dots,M$ we have*

$$\max_{1\le k_3,\dots,k_m\le K}\|\mathcal{M}^{(m)}_{::k_3\cdots k_m}\|\le C|\mu_2^{(m)}|,\qquad \max_{1\le i\le K}|d_i^{(m)}-1|\le C|\mu_2^{(m)}|,$$

where $\mathcal{M}^{(m)}$ is a $m-$way symmetric tensor defined by $\mathcal{M}^{(m)}_{k_1\cdots k_m}=(\mathcal{P}^{(m)}_{k_1\cdots k_m}-1)d_{k_1}^{(m)}\cdots d_{k_m}^{(m)}$, $1\le k_1,\dots,k_m\le K$.

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 χ^2 divergence between the two probability measures. By definitions,

$$\chi^2(P_0^{(n)},P_1^{(n)})=\int_{\mathcal{A}}\left[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\right]^2 dP_0^{(n)}(\mathcal{A})-1.$$

To show the claim, it suffices to show that when $(\mu_2^{(m)})^2\|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|_2^2\to 0$, $m=1,\dots,M$, we have

$$\int_{\mathcal{A}}\left[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\right]^2 dP_0^{(n)}(\mathcal{A})=1+o(1). \tag{E.17}$$

By definitions,

$$dP_0^{(n)}(\mathcal{A})=\prod_{m=2}^{M}\prod_{i_1<\cdots<i_m}dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}),$$

$$dP_1^{(n)}(\mathcal{A})=\mathbb{E}_{\Pi}\Big[\prod_{m=2}^{M}\prod_{i_1<\cdots<i_m}dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)\Big],$$

Let $\widetilde{\Pi}$ be an independent copy of Π. Putting the above two equations into (E.17) gives

$$\int_{\mathcal{A}}\Big[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\Big]^2 dP_0^{(n)}(\mathcal{A}) = \int_{\mathcal{A}} \frac{\mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\prod_{m=2}^M \prod_{i_1<\cdots<i_m} dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})\Big]}{\prod_{m=2}^M \prod_{i_1<\cdots<i_m} dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m})}$$
$$= \int_{\mathcal{A}} \mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\prod_{m=2}^M \prod_{i_1<\cdots<i_m} \frac{dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})}{dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m})}\Big].$$

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

$$\int_{\mathcal{A}}\Big[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\Big]^2 dP_0^{(n)}(\mathcal{A}) = \mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\int_{\mathcal{A}} \prod_{m=2}^M \prod_{i_1<\cdots<i_m} \frac{dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})}{dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m})}\Big]$$
$$= \mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\prod_{m=2}^M \prod_{i_1<\cdots<i_m} \int_{\mathcal{A}^{(m)}_{i_1\cdots i_m}} \frac{dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})}{dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m})}\Big].$$

Let $\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})$ denote $\int_{\mathcal{A}^{(m)}_{i_1\cdots i_m}} dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})/dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}) - 1$. Hence

$$\int_{\mathcal{A}}\Big[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\Big]^2 dP_0^{(n)}(\mathcal{A}) = \mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\prod_{m=2}^M \prod_{i_1<\cdots<i_m} (\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})+1)\Big]. \tag{E.18}$$

Note that by inequality $\prod_{i=1}^n(1+x_i) \le \exp(\sum_{i=1}^n x_i)$, for all x_i such that $1+x_i \ge 0$, we have

$$\prod_{m=2}^M \prod_{i_1<\cdots<i_m} (\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})+1) \le \exp\Big(\sum_{m=2}^M \sum_{i_1<\cdots<i_m} \chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big), \tag{E.19}$$

Further, by Jensen's inequality, $\exp(\sum_{i=2}^M x_i) \le \frac{1}{M-1}\sum_{i=2}^M \exp(x_i)$. It follows that

$$\exp\Big(\sum_{m=2}^M \sum_{i_1<\cdots<i_m} \chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big) \le \sum_{m=2}^M \frac{1}{M-1}\exp\Big((M-1)\sum_{i_1<\cdots<i_m} \chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big). \tag{E.20}$$

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

$$\int_{\mathcal{A}}\Big[\frac{dP_1^{(n)}(\mathcal{A})}{dP_0^{(n)}(\mathcal{A})}\Big]^2 dP_0^{(n)}(\mathcal{A}) \le \sum_{m=2}^M \frac{1}{M-1}\mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\exp\Big((M-1)\sum_{i_1<\cdots<i_m} \chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big)\Big].$$

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

$$\mathbb{E}_{\Pi,\widetilde{\Pi}}\Big[\exp\Big((M-1)\sum_{i_1<\cdots<i_m} \chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big)\Big] = 1 + o_n(1). \tag{E.21}$$

Fix m, recall that

$$\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi}) = \int_{\mathcal{A}} \frac{dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi)dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\widetilde{\Pi})}{dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m})} - 1. \tag{E.22}$$

By definitions,

$$dP_0^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}) = (\mathcal{Q}^{(m)}_{i_1\cdots i_m})^{\mathcal{A}^{(m)}_{i_1\cdots i_m}}(1-\mathcal{Q}^{(m)}_{i_1\cdots i_m})^{1-\mathcal{A}^{(m)}_{i_1\cdots i_m}},$$
$$dP_1^{(n,m)}(\mathcal{A}^{(m)}_{i_1\cdots i_m}|\Pi) = (\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\Pi))^{\mathcal{A}^{(m)}_{i_1\cdots i_m}}(1-\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\Pi))^{1-\mathcal{A}^{(m)}_{i_1\cdots i_m}}.$$

Putting the above two equations into (E.22) gives

$$
\begin{aligned}
\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi}) =& \frac{\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\Pi)\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\widetilde{\Pi})}{\mathcal{Q}^{(m)}_{i_1\cdots i_m}} + \frac{(1-\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\Pi))(1-\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\widetilde{\Pi}))}{1-\mathcal{Q}^{(m)}_{i_1\cdots i_m}} - 1 \\
=& \frac{\left(\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\Pi) - \mathcal{Q}^{(m)}_{i_1\cdots i_m}\right)\left(\mathcal{Q}^{*(m)}_{i_1\cdots i_m}(\widetilde{\Pi}) - \mathcal{Q}^{(m)}_{i_1\cdots i_m}\right)}{\mathcal{Q}^{(m)}_{i_1\cdots i_m}(1-\mathcal{Q}^{(m)}_{i_1\cdots i_m})}.
\end{aligned}
\tag{E.23}
$$

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

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

$$
\mathcal{Q}_{i_1\cdots i_m} = \theta_{i_1}\cdots\theta_{i_m} \qquad \text{and} \qquad \mathcal{Q}^*_{i_1\cdots i_m} = \theta_{i_1}\cdots\theta_{i_m}[\mathcal{P}^*;\pi^*_{i_1},\ldots,\pi^*_{i_m}],
$$

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

Let $\mathcal{P}_0$ the tensor with the same size as $\mathcal{P}^*$ and where all entries are 1, and introduce a tensor $\mathcal{M}$ by $\mathcal{M} = \mathcal{P}^* - \mathcal{P}_0$. Let $h = \mathbb{E}_F[\pi^*_i]$, and $y_i = \pi^*_i - h$, $1\le i\le n$. By Lemma E.1, we can write the Bernoulli probability tensor for the alternative $\mathcal{Q}^*$ by

$$
\mathcal{Q}^*_{i_1\cdots i_m} = \theta_{i_1}\cdots\theta_{i_m}(1 + x_{i_1\cdots i_m} + z_{i_1\cdots i_m}), \qquad 1\le i_1,\ldots,i_m\le n, \tag{E.24}
$$

where

$$
\begin{aligned}
x_{i_1\cdots i_m} =& [\mathcal{M}; h,\ldots,h] + \sum_{s=1}^{m}[\mathcal{M};\underbrace{h,\ldots,h}_{s-1}, y_{i_s}, \underbrace{h,\ldots,h}_{m-s}], \\
z_{i_1\cdots i_m} =& \sum_{s_1=1}^{m-1}\sum_{s_2=s_1+1}^{m}[\mathcal{M};\underbrace{h,\ldots,h}_{s_1-1}, y_{i_{s_1}}, \underbrace{h,\ldots,h}_{s_2-s_1-1}, y_{i_{s_2}}, \underbrace{\pi^*_{i_{s_2+1}}\ldots,\pi^*_{i_m}}_{m-s_2}].
\end{aligned}
$$

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

$$
\begin{aligned}
[\mathcal{M}; h,\ldots,h,e_{i_1},h,\ldots,h] =& \sum_{i_2,\ldots,i_m=1}^{K}(\mathcal{P}^*_{i_1\cdots i_m} - 1)\cdot h_{i_2}\cdots h_{i_m} \\
=& \sum_{i_2,\ldots,i_m=1}^{K}\mathcal{P}^*_{i_1\cdots i_m}\cdot h_{i_2}\cdots h_{i_m} - 1 \\
\text{(By degree matching)} =& 0
\end{aligned}
$$

This indicates that any linear combination of elements in $\{[\mathcal{M}; h,\ldots,h,e_i,h,\ldots,h] :, 1\le i\le K\}$ 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 $z_{i_1\cdots i_m}(s_1,s_2) = [\mathcal{M}; h,\ldots,h,y_{i_{s_1}},h,\ldots,h,y_{i_{s_2}},\pi^*_{i_{s_2+1}}\ldots,\pi^*_{i_m}]$, we get

$$
\mathcal{Q}^*_{i_1\cdots i_m} = \theta_{i_1}\cdots\theta_{i_m}\Big(1 + \sum_{s_1=1}^{m-1}\sum_{s_2=s_1+1}^{m} z_{i_1\cdots i_m}(s_1,s_2)\Big), \tag{E.25}
$$

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

$$
\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi}) = \frac{\theta_{i_1}\cdots\theta_{i_m}}{1-\theta_{i_1}\cdots\theta_{i_m}}\sum_{\substack{s_1=1,\\ \widetilde{s}_1=1}}^{m-1}\sum_{\substack{s_2=s_1+1\\ \widetilde{s}_2=\widetilde{s}_1+1}}^{m} z_{i_1\cdots i_m}(s_1,s_2)\widetilde{z}_{i_1\cdots i_m}(\widetilde{s}_1,\widetilde{s}_2).
$$

Note that $\frac{x}{1-x} = \sum_{i=1}^{\infty} x^i$ for any $x \in [0,1)$, we have $\frac{\theta_{i_1}\cdots\theta_{i_m}}{1-\theta_{i_1}\cdots\theta_{i_m}} = \sum_{i=1}^{\infty}(\theta_{i_1}\cdots\theta_{i_m})^t$ and so

$$\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi}) = \sum_{t=1}^{\infty}(\theta_{i_1}\cdots\theta_{i_m})^t \sum_{\substack{s_1=1,\\ \widetilde{s}_1=1}}^{m-1} \sum_{\substack{s_2=s_1+1\\ \widetilde{s}_2=\widetilde{s}_1+1}}^{m} z_{i_1\cdots i_m}(s_1,s_2)\widetilde{z}_{i_1\cdots i_m}(\widetilde{s}_1,\widetilde{s}_2).$$

Introduce

$$\begin{aligned} a_t =& \theta_{\max}^{m(t-1)}(1-\theta_{\max}^m), \\ S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2) =& (M-1)4^m \sum_{i_1<\cdots<i_m} \frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a^t} z_{i_1\cdots i_m}(s_1,s_2)\widetilde{z}_{i_1\cdots i_m}(\widetilde{s}_1,\widetilde{s}_2). \end{aligned} \tag{E.26}$$

Exchanging the order of summation, we then can write

$$(M-1)\sum_{i_1<\cdots<i_m}\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi}) = \sum_{t=1}^{\infty}\sum_{\substack{s_1=1,\\ \widetilde{s}_1=1}}^{m-1} \sum_{\substack{s_2=s_1+1\\ \widetilde{s}_2=\widetilde{s}_1+1}}^{m} \frac{a_t}{4^m} S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2).$$

Note that $\sum_{t=1}^{\infty}\sum_{s_1,\widetilde{s}_1=1}^{m-1}\sum_{s_2=s_1+1,\widetilde{s}_2=\widetilde{s}_1+1}^{m} a_t/4^m = 1$ and $\exp(\cdot)$ is convex, by Jensen's inequality

$$\exp\Big((M-1)\sum_{i_1<\cdots<i_m}\chi^2_{i_1\cdots i_m}(\Pi,\widetilde{\Pi})\Big) \le \sum_{t=1}^{\infty}\sum_{\substack{s_1=1,\\ \widetilde{s}_1=1}}^{m-1} \sum_{\substack{s_2=s_1+1\\ \widetilde{s}_2=\widetilde{s}_1+1}}^{m} \frac{a_t}{4^m}\exp\Big(S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2)\Big).$$

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

$$\max_{t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2}\Big\{\mathbb{E}\Big[\exp\Big(S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2)\Big)\Big]\Big\} \le 1 + o_n(1). \tag{E.27}$$

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

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

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

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

$$\begin{aligned} S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2) :=& 4^m(M-1)\sum_{i_1<\cdots<i_m}\frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a^t}[\mathcal{M}; h,\ldots,h,y_{i_{s_1}},h,\ldots,h,y_{i_{s_2}},\pi^*_{i_{s_2+1}}\ldots,\pi^*_{i_m}] \\ &\cdot [\mathcal{M}; h,\ldots,h,\widetilde{y}_{i_{s_1}},h,\ldots,h,\widetilde{y}_{i_{s_2}},\widetilde{\pi}^*_{i_{s_2+1}}\ldots,\widetilde{\pi}^*_{i_m}]. \\ =& \frac{4^m(M-1)}{m!}\sum_{i_1,\ldots,i_m(dist)}\frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{a^t}[\mathcal{M}; y_{i_1},y_{i_2},h\ldots,h,\pi^*_{i_{s_2+1}}\ldots,\pi^*_{i_m}] \\ &\cdot [\mathcal{M}; \widetilde{y}_{i_1},\widetilde{y}_{i_2},h\ldots,h,\widetilde{\pi}^*_{i_{s_2+1}}\ldots,\widetilde{\pi}^*_{i_m}]. \end{aligned}$$

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

$$\mathbb{E}\Big[\exp\Big(S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2)\Big)\Big] \le \mathbb{E}\Big[\exp\Big(C\frac{\mu_2^2\|\theta\|_t^{t(m-2)}}{a_t}|T|\Big)\Big]\cdot\exp\Big(C\frac{\mu_2^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}}{a_t}\Big), \quad \text{(E.28)}$$

where μ_2 is the second singular value of the matricization of the tensor $\mathcal{P}^{(m)}$ and T is a random variable satisfying $\mathbb{P}(|T|>x)\le 4\exp(-x/(2K^2\|\theta\|_{2t}^{2t}))$, for any $x>0$.

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

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(C\frac{\|\theta\|_t^{t(m-2)}}{a_t}\mu_2^2|T|\Big)\Big] &= \Big(1+\int_0^\infty e^x\cdot\mathbb{P}(C\frac{\|\theta\|_t^{t(m-2)}}{a_t}\mu_2^2|T|>x)dx\Big)\\ &\le\Big(1+\int_0^\infty e^x\cdot 4\exp(-\frac{a_t x}{2CK^2\mu_2^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}})dx\Big)\end{aligned} \quad \text{(E.29)}$$

By $\theta_{\max}\le c_0$, $\|\theta\|_t^t\le\|\theta\|_1\theta_{\max}^{t-1}$ and $\|\theta\|_{2t}^{2t}\le\|\theta\|^2\theta_{\max}^{t-2}$, we have

$$\frac{a_t}{\mu_2^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}} = \frac{\theta_{\max}^{m(t-1)}(1-\theta_{\max}^m)}{\mu_2^2\|\theta\|_t^{t(m-2)}\|\theta\|_{2t}^{2t}} \ge \frac{1-c_0^m}{\mu_2^2\|\theta\|_1^{m-2}\|\theta\|_2^2}$$

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

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(S(t,s_1,s_2,\widetilde{s}_1,\widetilde{s}_2)\Big)\Big] \le&\Big(1+\int_0^\infty e^x\cdot 4\exp(-\frac{(1-c_0^m)x}{2CK^2\mu_2^2\|\theta\|_1^{(m-2)}\|\theta\|_2^2})dx\Big)e^{\frac{C}{1-c_0^m}\mu_2^2\|\theta\|_1^{m-2}\|\theta\|^2}\\ =&e^{\frac{C}{1-c_0^m}\mu_2^2\|\theta\|_1^{m-2}\|\theta\|^2}\Big(1+4\Big(\frac{(1-c_0^m)}{2CK^2\mu_2^2\|\theta\|_1^{m-2}\|\theta\|_2^2}-1\Big)^{-1}\Big),\end{aligned}$$

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

E.3 Proof of Lemma E.1

Recall the definition of $[\mathcal{P};\pi_1,\dots,\pi_m]$

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

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

$$\begin{aligned}[\mathcal{P};\pi_1,\dots,\pi_m] =& \sum_{k_1,\dots,k_m=1}^K \mathcal{M}_{k_1\dots k_m}\pi_1(k_1)\cdots\pi_m(k_m) + \sum_{k_1,\dots,k_m=1}^K 1\cdot\pi_1(k_1)\cdots\pi_m(k_m)\\ =&[\mathcal{M};\pi_1,\dots,\pi_m]+1.\end{aligned}$$

Therefore, we are left to show for $m>1$

$$[\mathcal{M};\pi_1,\dots,\pi_m] = x^{(m)}+z^{(m)}. \quad \text{(E.30)}$$

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

$$\begin{aligned}[\mathcal{M};\pi_1,\pi_2] =&\pi_1'\mathcal{M}\pi_2\\ =&h'\mathcal{M}h+y_1'\mathcal{M}h+h'\mathcal{M}y_2+y_1'\mathcal{M}y_2\\ =&\underbrace{[\mathcal{M};h,h]+[\mathcal{M};y_1,h]+[\mathcal{M};h,y_2]}_{x^{(2)}}+\underbrace{[\mathcal{M};y_1,y_2]}_{z^{(2)}}.\end{aligned}$$

Hence, the claim holds for $m = 2$.

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

$$\begin{aligned}[\mathcal{M}; \pi_1, \dots, \pi_{r+1}] =& [\mathcal{M}; h, \dots, h, \pi_{r+1}] + \sum_{s=1}^{r} [\mathcal{M}; h, \dots, h, y_s, h, \dots, h, \pi_{r+1}] \\ &+ \sum_{s_1=1}^{r-1} \sum_{s_2=s_1+1}^{r} [\mathcal{M}; h, \dots, h, y_{s_1}, h, \dots, h, y_{s_2}, \pi_{s_2+1} \dots, \pi_{r+1}].\end{aligned}$$

Further, decompose π_{r+1} into $h + y_{r+1}$. By direct calculations

$$\begin{aligned}[\mathcal{M}; \pi_1, \dots, \pi_r, \pi_{r+1}] =& \Big([\mathcal{M}; h, \dots, h, h] + [\mathcal{M}; h, \dots, h, y_{r+1}] \Big) \\ &+ \Big(\sum_{s=1}^{r} [\mathcal{M}; h, \dots, h, y_s, h, \dots, h, h] + \sum_{s=1}^{r} [\mathcal{M}; h, \dots, h, y_s, h, \dots, h, y_{r+1}] \Big) \\ &+ \sum_{s_1=1}^{m-1} \sum_{s_2=s_1+1}^{m} [\mathcal{M}; h, \dots, h, y_{s_1}, h, \dots, h, y_{s_2}, \pi_{s_2+1} \dots, \pi_{r+1}] \\ =& [\mathcal{M}; h, \dots, h] + \sum_{s=1}^{r+1} [\mathcal{M}; h, \dots, h, y_s, h, \dots, h] \\ &+ \sum_{s_1=1}^{r} \sum_{s_2=s_1+1}^{r+1} [\mathcal{M}; h, \dots, h, y_{s_1}, h, \dots, h, y_{s_2}, \pi_{s_2+1} \dots, \pi_{r+1}], \\ =& x^{r+1} + z^{r+1},\end{aligned}$$

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

By definitions and elementary algebra,

$$\begin{aligned}S = \sum_{i_3, \dots, i_m (dist)} \frac{(\theta_{i_1} \cdots \theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3, \dots, k_m = 1 \\ k_3', \dots, k_m' = 1}}^{K} \prod_{s=3}^{m} w_{i_s}^{(s)}(k_s) \widetilde{w}_{i_s}^{(s)}(k_s) \frac{N_\theta}{a_t} \\ \cdot \Big[\sum_{i_1, i_2 (dist) \in I^{(i)}} (\theta_{i_1} \theta_{i_2})^t (y_{i_1}' \mathcal{M}_{::k_3 \cdots k_m} y_{i_2}) (\widetilde{y}_{i_1}' \mathcal{M}_{::k_3' \cdots k_m'} \widetilde{y}_{i_2}) \Big],\end{aligned} \tag{E.31}$$

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

$$X(i, j, j', k, k') = \sum_{i_1, i_2 (dist) \in I^{(i)}} (\theta_{i_1} \theta_{i_2})^t \delta_j^{(k)} \delta_{j'}^{(k')} (y_{i_1}' b_j^{(k)}) (y_{i_2}' b_j^{(k)}) (\widetilde{y}_{i_1}' b_{j'}^{(k')}) (\widetilde{y}_{i_2}' b_{j'}^{(k')}).$$

Then we can write

$$\sum_{i_1, i_2 (dist) \in I^{(i)}} (\theta_{i_1} \theta_{i_2})^t (y_{i_1}' \mathcal{M}_{::k_3 \cdots k_m} y_{i_2}) (\widetilde{y}_{i_1}' \mathcal{M}_{::k_3' \cdots k_m'} \widetilde{y}_{i_2}) = \sum_{j, j'=1}^{K} X(i, j, j', k, k').$$

Inserting this into (E.31) gives

$$S = \sum_{i_3,\dots,i_m(dist)} \frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\dots,k_m=1\\k'_3,\dots,k'_m=1}}^{K} \prod_{s=3}^{m} w_{i_s}^{(s)}(k_s)\widetilde{w}_{i_s}^{(s)}(k_s) \sum_{j,j'=1}^{K} \frac{1}{K^2}\Big(\frac{K^2 N_\theta}{a_t} X(i,j,j',k,k')\Big).$$

Note that $\sum_{i_3,\dots,i_m(dist)} \frac{(\theta_{i_1}\cdots\theta_{i_m})^t}{N_\theta} \sum_{k_3,k'_3,\dots,k_m,k'_m=1}^{K} \prod_{s=3}^{m} w_{i_s}^{(s)}(k_s)\widetilde{w}_{i_s}^{(s)}(k_s) \sum_{j,j'=1}^{K} \frac{1}{K^2} = 1$ and that $\exp(\cdot)$ is convex. By Jensen's inequality,

$$\exp(cS) \le \sum_{\substack{i_3,\dots,i_m\\(dist)}} \frac{(\theta_{i_3}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\dots,k_m=1\\k'_3,\dots,k'_m=1}}^{K} \prod_{s=3}^{m} w_{i_s}^{(s)}(k_s)\widetilde{w}_{i_s}^{(s)}(k_s) \sum_{j,j'=1}^{K} \frac{1}{K^2} \exp\Big(\frac{cK^2 N_\theta}{a_t} X(i,j,j',k,k')\Big)$$

By assumptions $w_{i_s}^{(s)}, \widetilde{w}_{i_s}^{(s)}$ are independent of $y_{i_1}, y_{i_2}, \widetilde{y}_{i_1}, \widetilde{y}_{i_2}$, $3 \le s \le m$. Taking expectation on both sides gives

$$\begin{aligned}\mathbb{E}[\exp(cS)] \le& \sum_{\substack{i_3,\dots,i_m\\(dist)}} \frac{(\theta_{i_3}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\dots,k_m=1\\k'_3,\dots,k'_m=1}}^{K} \prod_{s=3}^{m} \mathbb{E}[w_{i_s}^{(s)}(k_s)]\mathbb{E}[\widetilde{w}_{i_s}^{(s)}(k_s)] \sum_{j,j'=1}^{K} \frac{1}{K^2}\\ &\cdot \mathbb{E}\Big[\exp\Big(\frac{cK^2 N_\theta}{a_t} X(i,j,j',k,k')\Big)\Big]\\ \le& \max_{i,j,j',k,k'} \mathbb{E}\Big[\exp\Big(\frac{cK^2 N_\theta}{a_t} X(i,j,j',k,k')\Big)\Big].\end{aligned}$$

Now, to show the claim, note that $N_\theta := \sum_{i_3,\dots,i_m(dist)} (\theta_{i_3}\cdots\theta_{i_m})^t \le \|\theta\|_t^{t(m-2)}$, we are sufficient to show that

$$X(i,j,j',k,k') \le C\mu^2|T| + C\mu^2\|\theta\|_{2t}^{2t}, \tag{E.32}$$

where T is a random variable satisfying $\mathbb{P}(|T| > x) \le 2\exp(-x/(2K^2\|\theta\|_{2t}^{2t}))$, for $x > 0$.

To see this, we rewrite

$$\begin{aligned}X(i,j,j',k,k') :=& \sum_{i_1,i_2\in I^{(i)}} (1 - \mathbb{I}_{\{i_1=i_2\}})(\theta_{i_1}\theta_{i_2})^t \delta_j^{(k)}\delta_{j'}^{(k')}(y'_{i_1}b_j^{(k)})(y'_{i_2}b_j^{(k)})(\widetilde{y}'_{i_1}b_{j'}^{(k')})(\widetilde{y}'_{i_2}b_{j'}^{(k')})\\ =&\delta_j^{(k)}\delta_{j'}^{(k')}(T_1 - T_2),\end{aligned}$$

where

$$T_1 = \Big(\sum_{i_1\in I^{(i)}} \theta_{i_1}^t (y'_{i_1}b_j^{(k)})(\widetilde{y}'_{i_1}b_{j'}^{(k')})\Big)^2, \qquad T_2 = \sum_{i_1\in I^{(i)}} \Big(\theta_{i_1}^t (y'_{i_1}b_j^{(k)})(\widetilde{y}'_{i_1}b_{j'}^{(k')})\Big)^2.$$

Consider T_2 first. Note that $\max_{i_1}\{\|y_{i_1}\|, \|\widetilde{y}_{i_1}\|\} \le \sqrt{K}$ and that $\|b_j^{(k)}\| = \|b_{j'}^{(k')}\| = 1$, $\forall$ j, j', k, k'. By direct calculations

$$|T_2| \le (K)^2 \sum_{i_1} \theta_{i_1}^{2t} \le C\|\theta\|_{2t}^{2t}.$$

Next, consider T_1. Let $Z = \sum_{i_1\in I^{(i)}} \theta_{i_1}^t (y'_{i_1}b_j^{(k)})(\widetilde{y}'_{i_1}b_{j'}^{(k')})$. Note that Z is a sum of $n-(m-2)$ independent random variables with $|\theta_{i_1}^t (y'_{i_1}b_j^{(k)})(\widetilde{y}'_{i_1}b_{j'}^{(k')})| \le \sqrt{K}^2\theta_{i_1}^t$. By Hoeffding's inequality

$$\mathbb{P}(|Z| > x) \le 2\exp\Big(-2x^2/(\sum_{i_1\in I^{(i)}} (2\sqrt{K}^2\theta_{i_1}^t)^2)\Big), \qquad \text{for } x > 0.$$

Combining this with $\sum_{i_1\in I^{(i)}}(2\sqrt{K}^2\theta_{i_1}^t)^2 \le 4K^2\|\theta\|_{2t}^{2t}$ and $T_1 = Z^2$, it follows that

$$\mathbb{P}(|T_1| > x) \le 2\exp(-x/(2K^2\|\theta\|_{2t}^{2t})), \qquad \text{for } x > 0. \tag{E.33}$$

At the same time, recall that $\delta_j^{(k)}$,$\delta_{j'}^{(k')}$ are the eigenvalues of the matrices $\mathcal{M}_{::k_3\cdots k_m}$ and $\mathcal{M}_{::k_3'\cdots k_m'}$. By the assumption $\|\mathcal{M}_{::k_3\cdots k_m}\| \le C\mu$, for $1 \le k_3, \ldots, k_m \le K$, $\max_{j,k}\{|\delta_j^{(k)}|\} \le C\mu$. It is seen that

$$X(i,j,j',k,k') := \delta_j^{(k)}\delta_{j'}^{(k')}(T_1 - T_2) \le C\mu^2|T_1| + C\mu^2\|\theta\|_{2t}^{2t}, \qquad \text{with } T_1 \text{ satisfying (E.33)}.$$

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

E.5 Proof of Lemma E.3

Similarly, let $N_\theta = \sum_{i_3,\ldots,i_m(dist)}(\theta_{i_3}\cdots\theta_{i_m})^t$ 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. Let $\mathcal{M}_{::k_3\cdots k_m} = \sum_{j=1}^K b_j^{(k)}b_j^{(k)'}\delta_j^{(k)}$, and $\mathcal{M}_{:k_2':k_4'\cdots k_m'} = \sum_{j'=1}^K b_{j'}^{(k')}b_{j'}^{(k')'}\delta_j^{(k')}$ be the eigen-decomposition of the matrices $\mathcal{M}_{::k_3\cdots k_m}$ and $\mathcal{M}_{:k_2':k_4'\cdots k_m'}$, respectively. Following the procedures in the proof of Lemma E.2, we can obtain

$$\begin{aligned}\exp(cS) \le \sum_{\substack{i_3,\ldots,i_m\\(dist)}} \frac{(\theta_{i_3}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\ldots,k_m=1\\k_2',k_4',\ldots,k_m'=1}}^K h(k_2')\widetilde{w}_{i_3}^{(3)}(k_3)\prod_{s=4}^m w_{i_s}^{(s)}(k_s)\widetilde{w}_{i_s}^{(s)}(k_s)\sum_{j,j'=1}^K\frac{1}{K^2}\\ \cdot\exp\Big(\frac{cK^2N_\theta}{a_t}X(i,j,j',k,k')\Big),\end{aligned} \tag{E.34}$$

where

$$X(i,j,j',k,k') = \sum_{i_1,i_2(dist)\in I^{(i)}}(\theta_{i_1}\theta_{i_2})^t\delta_j^{(k)}\delta_{j'}^{(k')}(y_{i_1}'b_j^{(k)})(y_{i_2}'b_j^{(k)})(\widetilde{y}_{i_1}'b_{j'}^{(k')})(\widetilde{y}_{i_3}'b_{j'}^{(k')}).$$

Note that $\widetilde{w}_{i_3}^{(3)}$ may not be independent of $\widetilde{y}_{i_3}$ which exists in $X(i,j,j',k,k')$. Consequently, we can not 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 $\widetilde{y}_{i_3}$.We rewrite

$$\begin{aligned}X(i,j,j',k,k') :=& \sum_{i_1,i_2\in I^{(i)}}(1-\mathbb{I}_{\{i_1=i_2\}})(\theta_{i_1}\theta_{i_2})^t\delta_j^{(k)}\delta_{j'}^{(k')}(y_{i_1}'b_j^{(k)})(y_{i_2}'b_j^{(k)})(\widetilde{y}_{i_1}'b_{j'}^{(k')})(\widetilde{y}_{i_3}'b_{j'}^{(k')})\\ =&\delta_j^{(k)}\delta_{j'}^{(k')}(T_1-T_2)(\widetilde{y}_{i_3}'b_{j'}^{(k')}),\end{aligned}$$

where

$$T_1 = \Big(\sum_{i_1\in I^{(i)}}\theta_{i_1}^t(y_{i_1}'b_j^{(k)})(\widetilde{y}_{i_1}'b_{j'}^{(k')})\Big)\Big(\sum_{i_2\in I^{(i)}}\theta_{i_2}^t(y_{i_2}'b_j^{(k)})\Big), \qquad T_2 = \sum_{i_1\in I^{(i)}}\Big(\theta_{i_1}^t(y_{i_1}'b_j^{(k)})\Big)^2(\widetilde{y}_{i_1}'b_{j'}^{(k')}).$$

Recall that $\delta_j^{(k)}$,$\delta_{j'}^{(k')}$ are the eigenvalues of the matrices $\mathcal{M}_{::k_3\cdots k_m}$ and $\mathcal{M}_{:k_2':k_4'\cdots k_m'}$. By the assumption $\|\mathcal{M}_{::k_3\cdots k_m}\| \le C\mu$, for $1 \le k_3,\ldots,k_m \le K$, $\max_{j,k}\{|\delta_j^{(k)}|\} \le C\mu$. Combining this with $\|b_{j'}^{(k')}\| = 1$ and $\|y_{i_3}\| \le \sqrt{K}$, we have

$$X(i,j,j',k,k') := \delta_j^{(k)}\delta_{j'}^{(k')}(T_1-T_2)(\widetilde{y}_{i_3}'b_{j'}^{(k')}) \le C\mu^2(|T_1|+|T_2|).$$

Note that T_1, T_2 (and so the bound) are independent of $w_{i_s}^{(s)}, \widetilde{w}_{i_s}^{(s)}$, $3 \le s \le m$. Applying this inequality to the RHS of (E.34) and taking expectation on both sides give

$$\begin{aligned}\mathbb{E}[\exp(cS)] \le & \sum_{\substack{i_3,\dots,i_m \\ (dist)}} \frac{(\theta_{i_3}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\dots,k_m=1 \\ k_2',k_4',\dots,k_m'=1}}^{K} h(k_2')\mathbb{E}[\widetilde{w}_{i_3}^{(3)}(k_3)] \prod_{j=4}^{m} \mathbb{E}[w_{i_s}^{(s)}(k_s)]\mathbb{E}[\widetilde{w}_{i_s}^{(s)}(k_s)] \sum_{j,j'=1}^{K} \frac{1}{K^2} \\ & \cdot \mathbb{E}\Big[\exp\Big(\frac{CN_\theta}{a_t}\mu^2(|T_1|+|T_2|)\Big)\Big] \\ \le & \max_{i,j,j',k,k'} \mathbb{E}\Big[\exp\Big(\frac{CN_\theta}{a_t}\mu^2(|T_1|+|T_2|)\Big)\Big].\end{aligned}$$

Now, to show the claim, note that $N_\theta := \sum_{i_3,\dots,i_m(dist)}(\theta_{i_3}\cdots\theta_{i_m})^t \le \|\theta\|_t^{t(m-2)}$, it is then sufficient to show that

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

Consider (I) first. Let $Z_1 = \sum_{i_1 \in I^{(i)}} \theta_{i_1}^t (y_{i_1}' b_j^{(k)})(\widetilde{y}_{i_1}' b_{j'}^{(k')})$, $Z_2 = \sum_{i_2 \in I^{(i)}} \theta_{i_2}^t (y_{i_2}' b_j^{(k)})$ 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 $x > 0$

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

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

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

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

Next, consider (II) in (E.35). By $\max_{i_1}\{\|y_{i_1}\|, \|\widetilde{y}_{i_1}\|\} \le \sqrt{K}$, $\|b_j^{(k)}\| = \|b_{j'}^{(k')}\| = 1, \forall\, j, j', k, k'$

$$|T_2| := \sum_{i_1 \in I^{(i)}} \Big(\theta_{i_1}^t (y_{i_1}' b_j^{(k)})\Big)^2 (\widetilde{y}_{i_1}' b_{j'}^{(k')}) \le \sum_{i_1} \theta_{i_1}^{2t} (\sqrt{K})^2 \sqrt{K} \le C\|\theta\|_{2t}^{2t},$$

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 $N_\theta = \sum_{i_3,\dots,i_m(dist)}(\theta_{i_3}\cdots\theta_{i_m})^t$ and $I^{(i)}$ be the shorthand notation for set $\{1,\dots,n\} \setminus \{i_3,\dots,i_m\}$. Here, for convenience, we misuse the superscript (i) to indicate that this element depends on the choice of $(i_3,\dots,i_m)$ whenever it is clear from the context. Let $\mathcal{M}_{::k_3\cdots k_m} = \sum_{j=1}^K b_j^{(k)} b_j^{(k)'} \delta_j^{(k)}$, and $\mathcal{M}_{k_1'k_2'::k_5'\dots k_m'} = \sum_{j'=1}^K b_{j'}^{(k')} b_{j'}^{(k')'} \delta_j^{(k')}$ be the eigen-decomposition of the matrices $\mathcal{M}_{::k_3\cdots k_m}$ and $\mathcal{M}_{k_1'k_2'::k_5'\dots k_m'}$, respectively. Following the procedures in the proof of Lemma E.2, we can obtain

$$\begin{aligned}\exp(cS) \le & \sum_{\substack{i_3,\dots,i_m \\ (dist)}} \frac{(\theta_{i_3}\cdots\theta_{i_m})^t}{N_\theta} \sum_{\substack{k_3,\dots,k_m=1 \\ k_1'k_2',k_5',\dots,k_m'=1}}^{K} h(k_1')h(k_2') \prod_{s=5}^{m} w_{i_s}^{(s)}(k_s)\widetilde{w}_{i_s}^{(s)}(k_s) \\ & \cdot \widetilde{w}_{i_3}^{(3)}(k_3)\widetilde{w}_{i_4}^{(4)}(k_4) \sum_{j,j'=1}^{K} \frac{1}{K^2} \cdot \exp\Big(\frac{cK^2N_\theta}{a_t} X(i,j,j',k,k')\Big),\end{aligned} \tag{E.36}$$

where

$$X(i,j,j',k,k') = \sum_{i_1,i_2(dist)\in I^{(i)}} (\theta_{i_1}\theta_{i_2})^t \delta_j^{(k)} \delta_{j'}^{(k')} (y_{i_1}' b_j^{(k)})(y_{i_2}' b_j^{(k)})(\widetilde{y}_{i_3}' b_{j'}^{(k')})(\widetilde{y}_{i_4}' b_{j'}^{(k')}).$$

Note that $\widetilde{w}_{i_3}^{(3)}$ and $\widetilde{w}_{i_4}^{(4)}$ may not be independent of $\widetilde{y}_{i_3}$ and $\widetilde{y}_{i_4}$ which exist in $X(i,j,j',k,k')$. Similar to the proof of Lemma E.3, we rewrite

$$\begin{aligned} X(i,j,j',k,k') &:= \sum_{i_1,i_2\in I^{(i)}} (1-\mathbb{I}_{\{i_1=i_2\}})(\theta_{i_1}\theta_{i_2})^t \delta_j^{(k)} \delta_{j'}^{(k')} (y_{i_1}' b_j^{(k)})(y_{i_2}' b_j^{(k)})(\widetilde{y}_{i_3}' b_{j'}^{(k')})(\widetilde{y}_{i_4}' b_{j'}^{(k')}) \\ &= \delta_j^{(k)} \delta_{j'}^{(k')} (T_1 - T_2)(\widetilde{y}_{i_3}' b_{j'}^{(k')})(\widetilde{y}_{i_4}' b_{j'}^{(k')}), \end{aligned}$$

where

$$T_1 = \Big(\sum_{i_1\in I^{(i)}} \theta_{i_1}^t (y_{i_1}' b_j^{(k)})\Big)^2, \qquad T_2 = \sum_{i_1\in I^{(i)}} \Big(\theta_{i_1}^t (y_{i_1}' b_j^{(k)})\Big)^2.$$

Recall that $\delta_j^{(k)}$,$\delta_{j'}^{(k')}$ are the eigenvalues of the matrices $\mathcal{M}_{::k_3\cdots k_m}$ and $\mathcal{M}_{:k_2':k_4'\cdots k_m'}$. By the assumption $\|\mathcal{M}_{::k_3\cdots k_m}\| \le C\mu$, for $1 \le k_3,\dots,k_m \le K$, $\max_{j,k}\{|\delta_j^{(k)}|\} \le C\mu$. Combining this with $\|b_{j'}^{(k')}\| = 1$ and $\|y_{i_3}\| \le \sqrt{K}$, we have

$$X(i,j,j',k,k') := \delta_j^{(k)} \delta_{j'}^{(k')} (T_1 - T_2)(\widetilde{y}_{i_3}' b_{j'}^{(k')})(\widetilde{y}_{i_4}' b_{j'}^{(k')}) \le C\mu^2(|T_1| + |T_2|).$$

Note that T_1, T_2 (and so the bound) are independent of $w_{i_s}^{(s)}, \widetilde{w}_{i_s}^{(s)}$, $3 \le s \le m$. Applying this inequality to the RHS of (E.36) and taking expectation on both sides give

$$\mathbb{E}[\exp(cS)] \le \max_{i,j,j',k,k'} \mathbb{E}\Big[\exp\Big(\frac{CN_\theta}{a_t}\mu^2(|T_1| + |T_2|)\Big)\Big].$$

Now, to show the claim, note that $N_\theta := \sum_{i_3,\dots,i_m(dist)} (\theta_{i_3}\cdots\theta_{i_m})^t \le \|\theta\|_t^{t(m-2)}$, it is then sufficient to show that

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

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 $m \in \{2,\dots,M\}$,*

$$\max_{1\le i_1,\dots,i_m\le K}\{|\mathcal{P}_{i_1\cdots i_m}^{(m)} - 1|\} \asymp |\mu_2^{(m)}|. \tag{E.37}$$

Fix m, for simplicity of notation, we remove the superscript (m) whenever it is clear from the context. Recall that $D = \mathrm{diag}(d_1,\cdots,d_K)$ and $h = \mathbb{E}[D^{-1}\pi_i/\|D^{-1}\pi_i\|_1]$. Write for short $s = \sum_{k=1}^K d_k h_k$ and $v = (d_1,\dots,d_K)'$. With these notations and direct calculations, for $1 \le k_3,\dots,k_m \le K$

$$\mathcal{M}_{::k_3\cdots k_m} = D(\mathcal{P}_{::k_3\cdots k_m} - \mathbf{1}_K\mathbf{1}_K')D\prod_{j=3}^m d_{k_j} + (\prod_{j=3}^m d_{k_j} - s^{m-2})vv' + (s^{m-2}vv' - \mathbf{1}_K\mathbf{1}_K').$$

Therefore, to prove the first claim of this lemma, by elementary algebra, it is sufficient to show that

$$
\begin{aligned}
&(a): \max_{1\le k_1,\dots,k_m\le K}\{|\mathcal{P}_{k_1\cdots k_m}-1|\}\le C|\mu_2|,\\
&(b): \max_{1\le k\le K}\{d_k\}\le C,\\
&(c): \max_{1\le i,j\le K}\{|(s^{m-2}vv'-\mathbf{1}_K\mathbf{1}_K')_{ij}|\}\le C|\mu_2|,\\
&(d): \max_{1\le k\le K}\{|d_k-s|\}\le C|\mu_2|,
\end{aligned}
$$

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

Consider (b). Recall that by degree matching

$$\sum_{k_2,\dots,k_m=1}^{K} D\mathcal{P}_{:k_2\cdots k_m}\prod_{j=2}^{m}(d_{k_j}h_{k_j})=\mathbf{1}_K. \tag{E.38}$$

Note that each element of $\mathcal{P}$ is non-negative and $\mathcal{P}_{k_1\cdots k_1}=1$ for $1\le k_1\le K$. It follows that

$$d_{k_1}(d_{k_1}h_{k_1})^{m-1}\le \sum_{k_2,\dots,k_m=1}^{K} d_{k_1}\mathcal{P}_{k_1\cdots k_m}\prod_{j=2}^{m}(d_{k_j}h_{k_j})=1,\qquad 1\le k_1\le K.$$

Combining this with our assumption $\min_{1\le k\le K}\{h_k\}\ge C$,

$$d_k\le h_k^{-(m-1)/m}\le C,\qquad 1\le k\le K, \tag{E.39}$$

which proves (b).

Next consider (c). Let $\mathcal{H}$ be a tensor defined by $\mathcal{H}_{k_1\cdots k_m}=\mathcal{P}_{k_1\cdots k_m}-1$, for all $1\le k_1,\dots,k_m\le K$ and introduce w as the vector $\sum_{k_2\cdots k_m=1}^{K} D\mathcal{H}_{:k_2\cdots k_m}\prod_{j=2}^{m}(d_{k_j}h_{k_j})$. Recall that $s=\sum_{k=1}^{K} d_k h_k$. By definitions and calculations, (E.38) can be written as

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

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 $s^{m-2}vv'-\mathbf{1}_K\mathbf{1}_K'$ through $\mathbf{1}_K$ gives

$$
\begin{aligned}
s^{m-2}vv'-\mathbf{1}_K\mathbf{1}_K' =& s^{m-2}vv'-(w+s^{m-1}v)(w+s^{m-1}v)'\\
=& s^{m-2}(1-s^m)vv'-s^{m-1}wv'-s^{m-1}vw'-ww'.
\end{aligned}
$$

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

$$s^{m-2}vv'-\mathbf{1}_K\mathbf{1}_K' = s^{m-2}h'wvv'-s^{m-1}wv'-s^{m-1}vw'-ww'.$$

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

$$\max_{1\le i,j\le K}\{|(s^{m-2}vv'-\mathbf{1}_K\mathbf{1}_K')_{ij}|\}\le C\|h\|_{\max}\cdot\|v\|_{\max}\cdot\|w\|_{\max}\le C\|\mathcal{H}\|_{\max},$$

where $\|\cdot\|_{\max}$ is the element-wise maximum norm and $\|\mathcal{H}\|_{\max}:=\max_{k_1,\dots,k_m}\{|\mathcal{P}_{k_1,\dots,k_m}-1|\}\le C|\mu_2|$. This proves (c).

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

$$|s^{m-2}d_id_i-1|\le C|\mu_2|.$$

Transforming the above formula gives,

$$d_i = s^{-(m-2)/2} + O(|\mu_2|). \tag{E.42}$$

Summing up with weight h_i in terms of i on two sides and noting that $\sum_i h_i = 1$, it gives

$$s = s^{-(m-2)/2} + O(|\mu_2|). \tag{E.43}$$

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

Next we consider the second claim of this lemma i.e. $\max_{1\le i\le K}\{|d_i - 1|\} \le C|\mu_2|$. By elementary algebra, (E.43) can be rewritten as

$$s = 1 + \frac{\sqrt{s}^{m-1} + \sqrt{s}^{m-2}}{\sum_{j=0}^{m-1}\sqrt{s}^j}\cdot O(|\mu_2|),$$

where we note that $\frac{\sqrt{s}^{m-1}+\sqrt{s}^{m-2}}{\sum_{j=0}^{m-1}\sqrt{s}^j} \le 1$. Combining this with (E.42) proves the second claim.

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 $\Sigma = (\mathrm{diag}(\mu_1,\ldots,\mu_K), \mathbf{0}_{K\times(K^{m-1}-K)})$.

To show that claim, it is sufficient to show that

$$(I): |\mu_2| \le C\max_{1\le i_1,\ldots,i_m\le K}\{|\mathcal{P}_{i_1\cdots i_m} - 1|\}, \qquad (II): \max_{1\le i_1,\ldots,i_m\le K}\{|\mathcal{P}_{i_1\cdots i_m} - 1|\} \le C|\mu_2|.$$

Consider (I) first. Let P_0 be the $K\times K^{m-1}$ matrix of ones. Recall that μ_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],

$$|\mu_2| \le \|P - P_0\|.$$

At the same time, by elementary algebra, $\|P - P_0\| \le C\max_{1\le i_1,\ldots,i_m\le K}\{|\mathcal{P}_{i_1\cdots i_m} - 1|\}$. Combining these proves (I).

Next we consider (II).

By our assumption $\|P\|\le C$ and elemantary algebra,

$$\max_{1\le i_1,\ldots,i_m\le K}\{|\mathcal{P}_{i_1\cdots i_m}|\} = \|P\|_{\max} \le \|P\| \le C,$$

where $\|\cdot\|_{\max}$ is the element-wise maximum norm. Therefore, (II) directly holds for the case that $|\mu_2|\ge\epsilon$ for some positive constants $\epsilon < 1$. It is then sufficient to consider the case when $|\mu_2| < \epsilon$.

By definitions,

$$(PP')_{ii} \ge \mathcal{P}^2_{i\cdots i} = 1, \qquad (PP')_{ij} \ge 0, \qquad 1\le i,j\le K.$$

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

$$\mu_1 > 0, \qquad u_1(i) > 0, \qquad 1\le i\le K.$$

Let $a = u_1\mu_1^{\frac{1}{m}}$ and $b = v_1\mu_1^{\frac{m-1}{m}}$ be the scaled version of u_1 and v_1, where $a_i > 0$ since $u_1(i) > 0$, $1\le i\le K$. Introduce $\widetilde{P} = ab'$. For simplicity, we misuse the notation $b_{i_2\cdots i_m}$

for $b_{i_2+\sum_{s=3}^m K^{s-2}(i_s-1)}$. To show (II), by triangle inequality, it is sufficient to show that for $1 \le i_1, \dots, i_m \le K$,

$$(IIa): |\mathcal{P}_{i_1\cdots i_m} - a_{i_1} b_{i_2\cdots i_m}| \le C|\mu_2|, \qquad (IIb): |a_{i_1} b_{i_2\cdots i_m} - 1| \le C|\mu_2|.$$

Note that by elementary algebra

$$|\mathcal{P}_{i_1\cdots i_m} - a_{i_1} b_{i_2\cdots i_m}| \le \|P - \widetilde{P}\|_{\max} \le \|P - \widetilde{P}\| = |\mu_2|, \tag{E.44}$$

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,

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

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

$$|a_{i_1} - a_{i_2}| \le \frac{|a_{i_1} b_{i_2\cdots i_2} - a_{i_1}/a_{i_2}| + |a_{i_1} b_{i_2\cdots i_2} - (a_{i_2}/a_{i_1})^{m-1}|}{\sum_{j=0}^{m-1} a_{i_1}^{-j} a_{i_2}^{j-1}}, \qquad 1 \le i_1, i_2 \le K. \tag{E.45}$$

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

$$\Big| a_{i_1} b_{i_2\cdots i_k i_1 \cdots i_1} - \frac{\prod_{j=1}^k a_{i_j}}{a_{i_1}^k} \Big| \le \Big(2\sum_{s=2}^k \frac{\prod_{j=s+1}^k a_{i_j}}{a_{i_1}^{k-s}} + \frac{\prod_{j=1}^k a_{i_j}}{a_{i_1}^k} \Big) |\mu_2|, \quad 1 \le i_1, \dots, i_m \le K. \tag{E.46}$$

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

$$\Big| a_{i_1} b_{i_2\cdots i_2} - \frac{a_{i_2}^{m-1}}{a_{i_1}^{m-1}} \Big| \le \Big(2\sum_{s=2}^m \frac{a_{i_2}^{m-s}}{a_{i_1}^{m-s}} + \frac{a_{i_2}^{m-1}}{a_{i_1}^{m-1}} \Big) |\mu_2|, \qquad \Big| a_{i_1} b_{i_2\cdots i_2} - \frac{a_{i_1}}{a_{i_2}} \Big| \le \frac{a_{i_1}}{a_{i_2}} |\mu_2|.$$

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

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

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

$$\sum_{i_2=1}^K \Big(a_{i_1} - (a_{i_1} + a_{i_2})|\mu_2| \Big) \le \sum_{i_2=1}^K a_{i_2} \le \sum_{i_2=1}^K \Big(a_{i_1} + (a_{i_1} + a_{i_2})|\mu_2| \Big).$$

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

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

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

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

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

$$|a_{i_1} b_{i_2\cdots i_m} - 1| \le |a_{i_1} b_{i_2\cdots i_m} - \frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m}| + |\frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m} - 1|. \tag{E.48}$$

Note that setting $k = m$ in (E.46) gives

$$\Big|a_{i_1} b_{i_2\cdots i_m} - \frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m}\Big| \leq \Big(2\sum_{s=2}^m \frac{\prod_{j=s+1}^m a_{i_j}}{a_{i_1}^{m-s}} + \frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m}\Big)|\mu_2|.$$

Inserting this into (E.48). By direct calculations and (E.47)

$$|a_{i_1} b_{i_2\cdots i_m} - 1| \leq \Big(2\sum_{s=2}^m \frac{\prod_{j=s+1}^m a_{i_j}}{a_{i_1}^{m-s}} + \frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m}\Big)|\mu_2| + |\frac{\prod_{j=1}^m a_{i_j}}{a_{i_1}^m} - 1| = O(|\mu_2|).$$

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

$$|a_{i_1} b_{i_1\cdots i_1} - 1| \leq |\mu_2|, \qquad 1 \leq i_1 \leq K \tag{E.49}$$

Since $\mathcal{P}_{i_1\cdots i_1} = 1$, for $1 \leq i_1 \leq K$. By (E.44), we have

$$|a_{i_1} b_{i_1\cdots i_1} - 1| \leq |\mu_2|,$$

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,

$$\begin{aligned}
&\Big|a_{i_1} b_{i_2\cdots i_{k_0+1} i_1\cdots i_1} - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_1}^{k_0+1}}\Big| \\
\leq &\Big|a_{i_1} b_{i_2\cdots i_{k_0+1} i_1\cdots i_1} - \mathcal{P}_{i_1\cdots i_k i_{k_0+1} i_1\cdots i_1}\Big| + \Big|\mathcal{P}_{i_1\cdots i_k i_{k_0+1} i_1\cdots i_1} - \mathcal{P}_{i_{k_0+1} i_1\cdots i_{k_0} i_1\cdots i_1}\Big| \\
&+ \Big|\mathcal{P}_{i_{k_0+1} i_1\cdots i_{k_0} i_1\cdots i_1} - a_{i_{k_0+1}} b_{i_2\cdots i_{k_0} i_1\cdots i_1}\Big| + \Big|a_{i_{k_0+1}} b_{i_2\cdots i_{k_0} i_1\cdots i_1} - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_1}^{k_0+1}}\Big|
\end{aligned}$$

By (E.44), the first term and the third is bounded by $|\mu_2|$. Also, by symmetry of $\mathcal{P}$, the second term is 0. Moving a factor $a_{i_{k_0+1}}/a_{i_1}$ from the last term, it follows that

$$\begin{aligned}
\Big|a_{i_1} b_{i_2\cdots i_{k_0+1} i_1\cdots i_1} - \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_1}^{k_0+1}}\Big| &\leq 2|\mu_2| + \frac{a_{i_{k_0+1}}}{a_{i_1}}\Big|a_{i_1} b_{i_2\cdots i_{k_0} i_1\cdots i_1} - \frac{\prod_{j=1}^{k_0} a_{i_j}}{a_{i_1}^{k_0}}\Big| \\
\text{(By the assumption for } k = k_0\text{)} \quad &\leq 2|\mu_2| + \frac{a_{i_{k_0+1}}}{a_{i_1}}\Big(2\sum_{s=2}^{k_0} \frac{\prod_{j=s+1}^{k} a_{i_j}}{a_{i_1}^{k_0-s}} + \frac{\prod_{j=1}^{k} a_{i_j}}{a_{i_1}^{k_0}}\Big)|\mu_2| \\
&= \Big(2\sum_{s=2}^{k_0+1} \frac{\prod_{j=s+1}^{k_0+1} a_{i_j}}{a_{i_1}^{k_0+1-s}} + \frac{\prod_{j=1}^{k_0+1} a_{i_j}}{a_{i_1}^{k_0+1}}\Big)|\mu_2|,
\end{aligned}$$

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 $1 - O(1/n)$,*

- *(a) Under both the null and under the alternative, $|\hat{\alpha}_n - \widetilde{\alpha}_n| \leq C\log(n)(\widetilde{\alpha}_n/n^3)^{1/2}$.*
- *(b) Under the alternative, $\widetilde{\alpha}_n \leq \max_{1\leq k_1,k_2,k_3\leq K}\{\mathcal{P}_{k_1k_2k_3}\} \leq C\widetilde{\alpha}_n$ and $\widetilde{\alpha}_n = h'(\mathcal{P}h)h + O(\frac{\widetilde{\alpha}_n}{n})$.*

To show the claims of Lemma 2.2, it is sufficient to show as $n \to \infty$, for any positive constant M,

$$\psi_n \to N(0,1) \text{ under the null, } \qquad \text{and} \qquad \mathbb{P}(|\psi_n| \le M) \to 0 \text{ under the alternative.} \tag{F.50}$$

Recall that

$$\widetilde{\alpha}_n = \mathbb{E}[\hat{\alpha}_n],$$

Let $\mathcal{A}^*$ and $\widetilde{\mathcal{A}}$ be two tensors with the same size as $\mathcal{A}$, where $\mathcal{A}^*_{i_1i_2i_3} = \mathcal{A}_{i_1i_2i_3} - \hat{\alpha}_n$ and $\widetilde{\mathcal{A}}_{i_1i_2i_3} = \mathcal{A}_{i_1i_2i_3} - \widetilde{\alpha}_n$ if i_1, i_2, i_3 are distinct, and $\mathcal{A}^*_{i_1i_2i_3} = \widetilde{\mathcal{A}}_{i_1i_2i_3} = 0$ otherwise. By definitions,

$$\sqrt{2n}\psi_n = \frac{\sum_{1\le i\le n}\left(\sum_{j<k}\mathcal{A}^*_{ijk}\right)^2 - n\binom{n-1}{2}\hat{\alpha}_n(1-\hat{\alpha}_n)}{\binom{n-1}{2}\hat{\alpha}_n(1-\hat{\alpha}_n)}. \tag{F.51}$$

Let $S_0 = \{(i_1,i_2,i_3,i_4,i_5) : 1 \le i_1,i_2,i_3,i_4,i_5 \le n; i_1 < i_2; i_4 < i_5; i_1,i_2,i_4,i_5 \ne i_3\}$, and write for short $x = (i_1,i_2,i_3,i_4,i_5)$. Introduce a subset of S_0 by $S = \{x \in S_0 : (i_1,i_2) \ne (i_4,i_5)\}$. Note that for any $x \in S_0 \setminus S$, $(i_1,i_2) = (i_4,i_5)$. It is seen that the numerator on the RHS of (F.51) is

$$\begin{aligned}
&\sum_{x\in S_0}\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5} - n\binom{n-1}{2}\hat{\alpha}_n(1-\hat{\alpha}_n)\\
=&\sum_{x\in S}\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5} + \sum_{x\in S_0\setminus S}\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5} - n\binom{n-1}{2}\hat{\alpha}_n(1-\hat{\alpha}_n)\\
=&(I) + (II),
\end{aligned} \tag{F.52}$$

where

$$(I) = \sum_{x\in S}\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5}, \qquad (II) = \sum_{x\in S_0\setminus S}\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5} - n\binom{n-1}{2}\hat{\alpha}_n(1-\hat{\alpha}_n).$$

Consider (I) first. Write

$$(I) = (Ia) + (Ib), \tag{F.53}$$

where

$$(Ia) = \sum_{x\in S}\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}, \qquad (Ib) = \sum_{x\in S}(\mathcal{A}^*_{i_1i_2i_3}\mathcal{A}^*_{i_3i_4i_5} - \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}).$$

Now, by direct calculations,

$$(Ib) = (\widetilde{\alpha}_n - \hat{\alpha}_n)\sum_{x\in S}(\mathcal{A}_{i_1i_2i_3} + \mathcal{A}_{i_3i_4i_5} - \hat{\alpha}_n - \widetilde{\alpha}_n). \tag{F.54}$$

Note that for each tuple (i_1,i_2,i_3), there are $\binom{n-1}{2} - 1$ different $x = (i_1,i_2,i_3,i_4,i_5)$ in S with the same (i_1,i_2,i_3). It follows

$$\sum_{x\in S}\mathcal{A}_{i_1i_2i_3} = \left(\binom{n-1}{2} - 1\right)\sum_{\substack{i_1,i_2,i_3(dist)\\ i_1<i_2}}\mathcal{A}_{i_1i_2i_3} = \frac{n^2(n-1)(n-2)(n-3)}{4}\hat{\alpha}_n. \tag{F.55}$$

Similarly, we have

$$\sum_{x\in S}\mathcal{A}_{i_3i_4i_5} = \frac{n^2(n-1)(n-2)(n-3)}{4}\hat{\alpha}_n. \tag{F.56}$$

Inserting (F.55)-(F.56) into (F.54) gives

$$(Ib) = -\frac{n^2(n-1)(n-2)(n-3)}{4}(\widetilde{\alpha}_n - \hat{\alpha}_n)^2.$$

Combining this with (F.53) gives

$$(I) = (Ia) - \frac{n^2(n-1)(n-2)(n-3)}{4}(\widetilde{\alpha}_n - \hat{\alpha}_n)^2. \tag{F.57}$$

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

$$\sum_{x \in S_0 \setminus S} \mathcal{A}^*_{i_1 i_2 i_3} \mathcal{A}^*_{i_3 i_4 i_5} = \frac{1}{2} \sum_{i_1, i_2, i_3 (dist)} (\mathcal{A}^*_{i_1 i_2 i_3})^2 = \frac{1}{2} \sum_{i_1, i_2, i_3 (dist)} (\mathcal{A}^2_{i_1 i_2 i_3} - 2\hat{\alpha}_n \mathcal{A}_{i_1 i_2 i_3} + \hat{\alpha}_n^2). \tag{F.58}$$

Since $\mathcal{A}_{i_1 i_2 i_3} \in \{0, 1\}$, we have $\mathcal{A}^2_{i_1 i_2 i_3} = \mathcal{A}_{i_1 i_2 i_3}$. 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

$$\psi_n = \frac{(Ia) - (1/4)n^2(n-1)(n-2)(n-3)(\widetilde{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n}\binom{n-1}{2}\hat{\alpha}_n(1 - \hat{\alpha}_n)}.$$

Now, by Lemma F.1, $|\hat{\alpha}_n - \widetilde{\alpha}_n| \leq C \log(n)(\widetilde{\alpha}_n/n^3)^{1/2}$ except for a probability of $1 - O(1/n)$. It is seen that except for a probability of $1 - O(1/n)$

$$\left|\frac{\hat{\alpha}_n}{\widetilde{\alpha}_n} - 1\right| \leq C\frac{\log(n)}{\sqrt{n^3 \widetilde{\alpha}_n}}, \qquad \left|\frac{(1/4)n^2(n-1)(n-2)(n-3)(\widetilde{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n}\binom{n-1}{2}\hat{\alpha}_n(1 - \hat{\alpha}_n)}\right| \leq C\frac{\log^2(n)}{n^{1/2}}.$$

By $n^2\widetilde{\alpha}_n \to \infty$, we have that in probability,

$$\frac{\hat{\alpha}_n}{\widetilde{\alpha}_n} \to 1, \qquad \frac{(1/4)n^2(n-1)(n-2)(n-3)(\widetilde{\alpha}_n - \hat{\alpha}_n)^2}{\sqrt{2n}\binom{n-1}{2}\hat{\alpha}_n(1 - \hat{\alpha}_n)} \to 0.$$

Let

$$Z_n = \frac{(Ia)}{\sqrt{2n}\binom{n-1}{2}\widetilde{\alpha}_n(1 - \widetilde{\alpha}_n)}.$$

To show (F.50), it is sufficient to show that as $n \to \infty$,

$$Z_n \to N(0, 1), \qquad \text{under the null}, \tag{F.60}$$

and

$$\mathbb{P}(|Z_n| > M) \to 1 \text{ for any } M > 0, \qquad \text{under the alternative}. \tag{F.61}$$

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* $n \to \infty$

$$\mathbb{E}[Z_n] \geq C n^{2.5} \widetilde{\alpha}_n \delta_n^2, \qquad Var(Z_n) \leq C n^2 \widetilde{\alpha}_n.$$

Now, suppose the alternative hypothesis is true. Note that by triangle inequality

$$\mathbb{P}(|Z_n| \leq M) \leq \mathbb{P}\Big(\big|\mathbb{E}[Z_n]\big| - \big|Z_n - \mathbb{E}[Z_n]\big| \leq M\Big) = \mathbb{P}(\big|Z_n - \mathbb{E}[Z_n]\big| \geq \big|\mathbb{E}[Z_n]\big| - M),$$

where by Chebyshev's inequality,

$$\mathbb{P}(\big|Z_n - \mathbb{E}[Z_n]\big| \geq \big|\mathbb{E}[Z_n]\big| - M) \leq \frac{\mathrm{Var}(Z_n)}{(\mathbb{E}[Z_n] - M)^2}.$$

At the same time, by Lemma F.2 and our assumptions of $n^2\widetilde{\alpha}_n \to \infty$ and $n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2 \to \infty$,

$$\frac{\mathrm{Var}(Z_n)}{(\mathbb{E}[Z_n] - M)^2} \leq \frac{Cn^2\widetilde{\alpha}_n}{(Cn^{2.5}\widetilde{\alpha}_n\delta_n^2 - M)^2} \leq \frac{1}{C(n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2)^2} \to 0.$$

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

$$\widetilde{T}_{n,m} = \sum_{x \in S^{(m)}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}, \qquad Z_{n,m} = \frac{\widetilde{T}_{n,m}}{\sqrt{2n\binom{n-1}{2}}\widetilde{\alpha}_n(1-\widetilde{\alpha}_n)}, \qquad (\widetilde{T}_{n,0} = Z_{n,0} = 0),$$

and

$$X_{n,m} = Z_{n,m} - Z_{n,m-1}.$$

It is seen that

$$(Ia) = \widetilde{T}_{n,n}, \qquad \text{and} \qquad Z_n = Z_{n,n} = \sum_{m=1}^{n} X_{n,m}. \tag{F.62}$$

Consider the filtration $\{\mathcal{F}_{n,m}\}_{1\leq m\leq n}$ with $\mathcal{F}_{n,m} = \sigma\Big(\{\widetilde{\mathcal{A}}_{i_1i_2i_3} : 1 \leq i_1, i_2, i_3 \leq m\}\Big)$. 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 $n \to \infty$,*

$$(a) \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] \to 1, \qquad \textit{in probability},$$

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

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

$$\widetilde{\alpha}_n = \mathbb{E}[\hat{\alpha}_n] = \frac{\sum_{i_1,i_2,i_3(dist)} \mathcal{Q}_{i_1i_2i_3}}{n(n-1)(n-2)}.$$

Recall that under alternative

$$\mathcal{Q}_{i_1 i_2 i_3} = \sum_{1 \le k_1, k_2, k_3 \le K} \pi_{i_1}(k_1)\pi_{i_2}(k_2)\pi_{i_3}(k_3)\mathcal{P}_{k_1 k_2 k_3}, \qquad 1 \le i_1, i_2, i_3 \le n.$$

It is seen that $\mathcal{Q}_{i_1 i_2 i_3} \le \max_{1 \le k_1, k_2, k_3 \le K}\{\mathcal{P}_{k_1 k_2 k_3}\}$, $1 \le i_1, i_2, i_3 \le n$ and so

$$\widetilde{\alpha}_n \le \max_{1 \le k_1, k_2, k_3 \le K}\{\mathcal{P}_{k_1 k_2 k_3}\}, \qquad \widetilde{\alpha}_n = h'(\mathcal{P}h)h + O(\frac{\max_{1 \le k_1, k_2, k_3 \le K}\{\mathcal{P}_{k_1 k_2 k_3}\}}{n}).$$

At the same time, by our assumption $\min_{k=1}^{K}\{h_k\} \ge c_0$ and elementary calculations

$$\max_{1 \le k_1, k_2, k_3 \le K}\{\mathcal{P}_{k_1 k_2 k_3}\} \le C \sum_{1 \le k_1, k_2, k_3 \le K} h_{k_1} h_{k_2} h_{k_3} \mathcal{P}_{k_1 k_2 k_3} \le C\widetilde{\alpha}_n.$$

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 $C\widetilde{\alpha}_n$ under both null and alternative hypothesis. By Bernstein's inequality,

$$\mathbb{P}((\binom{n}{3})|\hat{\alpha}_n - \widetilde{\alpha}_n| \ge t\Big) \le 2\exp(-\frac{t^2}{\binom{n}{3}C\widetilde{\alpha}_n(1 - C\widetilde{\alpha}_n) + \frac{t}{3}}).$$

Let $t = C\binom{n}{3}\frac{\log(n)\widetilde{\alpha}_n^{1/2}}{n^{3/2}}$, by elementary calculations, we get

$$\mathbb{P}\Big(|\hat{\alpha}_n - \widetilde{\alpha}_n| \ge C\frac{\log(n)\widetilde{\alpha}_n^{1/2}}{n^{3/2}}\Big) \le O(1/n). \tag{F.63}$$

This is equivalent to the claim in (a).

F.2 Proof of Lemma F.2

Recall that

$$Z_n = (2n)^{-1/2}\frac{(Ia)}{\binom{n-1}{2}\widetilde{\alpha}_n(1 - \widetilde{\alpha}_n)}, \qquad \text{with } (Ia) = \textstyle\sum_{x \in S}(\mathcal{A}_{i_1 i_2 i_3} - \widetilde{\alpha}_n)(\mathcal{A}_{i_3 i_4 i_5} - \widetilde{\alpha}_n).$$

Therefore, to show the claims, it is sufficient to show that as $n \to \infty$

$$\mathbb{E}[(Ia)] \ge Cn^5\widetilde{\alpha}_n^2\delta_n^2, \tag{F.64}$$

and

$$\mathrm{Var}((Ia)) \le Cn^7\widetilde{\alpha}_n^3. \tag{F.65}$$

Consider (F.64) first. Since for each $x = (i_1, i_2, i_3, i_4, i_5) \in S$, $\mathcal{A}_{i_1 i_2 i_3}$ is independent of $\mathcal{A}_{i_3 i_4 i_5}$, by direct calculations,

$$\mathbb{E}[(Ia)] = \sum_{x \in S}(\mathcal{Q}_{i_1 i_2 i_3} - \widetilde{\alpha}_n)(\mathcal{Q}_{i_3 i_4 i_5} - \widetilde{\alpha}_n).$$

Let $\widetilde{Q}_{i_1 i_2 i_3} = \mathcal{Q}_{i_1 i_2 i_3} - \widetilde{\alpha}_n$, by definitions,

$$\mathbb{E}[(Ia)] = \frac{1}{4}(\sum_{x \in S_0'}\widetilde{Q}_{i_1 i_2 i_3}\widetilde{Q}_{i_3 i_4 i_5} - \sum_{x \in (S_0' \backslash S_1')}\widetilde{Q}_{i_1 i_2 i_3}\widetilde{Q}_{i_3 i_4 i_5}),$$

where

$$\begin{aligned} S_0' &= \{x : 1 \le i_1, i_2, i_3, i_4, i_5 \le n\} \\ S_1' &= \{x \in S_0' : i_1, i_2, i_3(dist); i_3, i_4, i_5(dist); (i_1, i_2) \neq (i_4, i_5)\}. \end{aligned}$$

To show (F.64), it is sufficient to show that

$$\sum_{x\in S_0'} \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5} \geq Cn^5\widetilde{\alpha}_n^2\delta_n^2, \qquad \text{and} \qquad \sum_{x\in (S_0'\setminus S_1')} \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5} = o(\sum_{x\in S_0'} \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}). \tag{F.66}$$

Consider the first claim in (F.66). Recall that

$$\widetilde{Q}_{i_1i_2i_3} = \mathcal{Q}_{i_1i_2i_3} - \widetilde{\alpha}_n = \sum_{k_1,k_2,k_3} \pi_{i_1}(k_1)\pi_{i_2}(k_2)\pi_{i_3}(k_3)\mathcal{P}_{k_1k_2k_3} - \widetilde{\alpha}_n, \qquad \text{and} \qquad h = \sum_{i=1}^n \pi_i/n.$$

By direct calculations and elementary algebra,

$$\sum_{x\in S_0'} \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5} = n^4\|\Pi(\mathcal{P}h)h - \widetilde{\alpha}_n\mathbf{1}_n\|^2.$$

By triangle inequality, we have $\|\Pi(\mathcal{P}h)h - \widetilde{\alpha}_n\mathbf{1}_n\| \geq \big|\|\Pi(\mathcal{P}h)h - h'(\mathcal{P}h)h\mathbf{1}_n\| - \|(h'(\mathcal{P}h)h - \widetilde{\alpha}_n)\mathbf{1}_n\|\big|$. It follows that

$$\sum_{x\in S_0'} \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5} \geq n^4(\|\Pi(\mathcal{P}h)h - h'(\mathcal{P}h)h\mathbf{1}_n\| - \|(h'(\mathcal{P}h)h - \widetilde{\alpha}_n)\mathbf{1}_n\|)^2. \tag{F.67}$$

Recall that $\Sigma = \Pi'\Pi/n - hh'$ and note that $\Sigma\mathbf{1}_K = 0$. Also, recall that $H_K = K^{-1}\mathbf{1}_K\mathbf{1}_K'$ 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,

$$\|\Pi(\mathcal{P}h)h - h'(\mathcal{P}h)h\mathbf{1}_n\|^2 = n\big(h'(\mathcal{P}h)\frac{\Pi'\Pi}{n}(\mathcal{P}h)h - h'(\mathcal{P}h)hh'(\mathcal{P}h)h\big) = n((\mathcal{P}h)h)'\Sigma((\mathcal{P}h)h), \tag{F.68}$$

where the RHS equals to

$$n((\mathcal{P}h)h)'(I_K - H_K)\Sigma(I_K - H_K)(\mathcal{P}h)h. \tag{F.69}$$

By our assumption $\lambda_{K-1}(\Sigma) = \min_{\|v\|=1, v\perp\mathbf{1}_K} v'\Sigma v \geq c_0$, it is seen that

$$n((\mathcal{P}h)h)'(I_K - H_K)\Sigma(I_K - H_K)(\mathcal{P}h)h \geq c_0 n\widetilde{\alpha}_n^2\|\widetilde{\alpha}_n^{-1}(I_K - H_K)(\mathcal{P}h)h\|^2. \tag{F.70}$$

Recall that $\delta_n = \|\widetilde{\alpha}_n^{-1}(I_K - H_K)(\mathcal{P}h)h\|$, combining with (F.68)-(F.70), we get

$$\|\Pi(\mathcal{P}h)h - h'(\mathcal{P}h)h\mathbf{1}_n\|^2 \geq c_0 n\widetilde{\alpha}_n^2\delta_n^2. \tag{F.71}$$

At the same time, by Lemma F.1,

$$\widetilde{\alpha}_n = h'(\mathcal{P}h)h + O(\frac{\widetilde{\alpha}_n}{n}). \tag{F.72}$$

By direct calculations,

$$\|(h'(\mathcal{P}h)h - \widetilde{\alpha}_n)\mathbf{1}_n\|^2 = n(h'(\mathcal{P}h)h - \widetilde{\alpha}_n)^2 = O(\frac{\widetilde{\alpha}_n^2}{n}), \tag{F.73}$$

where by $\widetilde{\alpha}_n \leq \max_{1\leq i_1,i_2,i_3\leq n}\{\mathcal{P}_{i_1i_2i_3}\} \leq c_0$ and our condition $n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2 \to \infty$,

$$\frac{\widetilde{\alpha}_n^2}{n} = o(1)\cdot(n\widetilde{\alpha}_n^2\delta_n^2). \tag{F.74}$$

Combining (F.72)-(F.74),

$$\|(h'(\mathcal{P}h)h - \widetilde{\alpha}_n)\mathbf{1}_n\|^2 = o(n\widetilde{\alpha}_n^2\delta_n^2). \tag{F.75}$$

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 $\sum_{x\in(S_0'\backslash S_1')}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}$ are the following:

$$O(\sum_{\substack{1\le i_1,i_2,i_3,i_4,i_5\le n\\ i_3=i_4}}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}), \qquad \text{and} \qquad O(\sum_{\substack{1\le i_1,i_2,i_3,i_4,i_5\le n\\ i_4=i_5}}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}). \tag{F.76}$$

The other terms are $O(n^3\widetilde{\alpha}_n^2) = o(n^5\widetilde{\alpha}_n^2\delta_n^2)$ 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,

$$|\sum_{\substack{1\le i_1,i_2,i_3,i_4,i_5\le n\\ i_3=i_4}}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}| \le \sqrt{\sum_{1\le i_3\le n}(\sum_{1\le i_5\le n}\widetilde{Q}_{i_3i_3i_5})^2}\sqrt{\sum_{1\le i_3\le n}(\sum_{1\le i_1,i_2\le n}\widetilde{Q}_{i_1i_2i_3})^2}. \tag{F.77}$$

Note that by definitions and Lemma F.1, $|\widetilde{Q}_{i_3i_3i_5}| \le C\widetilde{\alpha}_n$. It is seen that

$$\sum_{1\le i_3\le n}(\sum_{1\le i_5\le n}\widetilde{Q}_{i_3i_3i_5})^2 \le Cn^3\widetilde{\alpha}_n^2. \tag{F.78}$$

By our condition $n^{3/2}\widetilde{\alpha}_n^{1/2}\delta_n^2 \to \infty$, 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 $\sum_{x\in S_0'}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}$. At the same time,

$$\sum_{1\le i_3\le n}(\sum_{1\le i_1,i_2\le n}\widetilde{Q}_{i_1i_2i_3})^2 = \sum_{x\in S_0'}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}. \tag{F.79}$$

Inserting (F.78)-(F.79) into (F.77), we have

$$|\sum_{\substack{1\le i_1,i_2,i_3,i_4,i_5\le n\\ i_3=i_4}}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}| = o(\sum_{x\in S_0'}\widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}).$$

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 $\mathcal{W}_{i_1i_2i_3} = \mathcal{A}_{i_1i_2i_3} - \mathcal{Q}_{i_1i_2i_3}$ if i_1, i_2, i_3 are distinct, and $\mathcal{W}_{i_1i_2i_3} = 0$ otherwise. By symmetry and definitions,

$$(Ia) = \sum_{x\in S}(\mathcal{W}_{i_1i_2i_3} - \widetilde{Q}_{i_1i_2i_3})(\mathcal{W}_{i_3i_4i_5} - \widetilde{Q}_{i_3i_4i_5}) = \sum_{x\in S}(\mathcal{W}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5} - 2\widetilde{Q}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5} + \widetilde{Q}_{i_1i_2i_3}\widetilde{Q}_{i_3i_4i_5}). \tag{F.80}$$

Since for any random variables X and Y, $\mathrm{Var}(X+Y) \le 2\mathrm{Var}(X) + 2\mathrm{Var}(Y)$, we have

$$\mathrm{Var}((Ia)) \le 2\mathrm{Var}(\sum_{x\in S}\mathcal{W}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5}) + 2\mathrm{Var}(\sum_{x\in S}2\widetilde{Q}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5}).$$

Here, we note that $\widetilde{Q}$ is non-random, so the variance of the last term in (F.80) is 0. By direct calculations,

$$\mathrm{Var}(\sum_{x\in S}\mathcal{W}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5}) = \sum_{x\in S}\mathrm{Var}(\mathcal{W}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5}) = O(n^5\widetilde{\alpha}_n^2),$$

$$\mathrm{Var}(\sum_{x\in S}2\widetilde{Q}_{i_1i_2i_3}\mathcal{W}_{i_3i_4i_5}) = \frac{1}{4}\sum_{i_3i_4i_5(dist)}(\sum_{\substack{i_1i_2(dist)\\ \{i_1,i_2\}\neq\{i_4,i_5\}\\ i_1,i_2\neq i_3}}\mathcal{Q}_{i_1i_2i_3})^2\mathrm{Var}(W_{i_3i_4i_5}) = O(n^7\widetilde{\alpha}_n^3).$$

By our assumptions, $n^2\widetilde{\alpha}_n \to \infty$, and so $n^5\widetilde{\alpha}_n = o(1)\cdot n^7\widetilde{\alpha}_n^3$. Combining these gives that

$$\mathrm{Var}((Ia)) \le Cn^7\widetilde{\alpha}_n^3.$$

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

$$\mathbb{E}\Big[\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]\Big] \to 1, \qquad \mathrm{Var}(\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}]) \to 0. \tag{F.81}$$

Introduce

$$T^{(m)} = \mathbb{E}[(\sum_{x\in S^{(m)}\backslash S^{(m-1)}}\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5})^2|\mathcal{F}_{n,m-1}].$$

By definitions,

$$\mathbb{E}[X_{n,m}^2|\mathcal{F}_{n,m-1}] = \frac{\mathbb{E}[(\sum_{x\in S^{(m)}\backslash S^{(m-1)}}\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5})^2|\mathcal{F}_{n,m-1}]}{(\sqrt{2n}\binom{n-1}{2}\alpha_n(1-\alpha_n))^2} = \frac{T^{(m)}}{(\sqrt{2n}\binom{n-1}{2}\alpha_n(1-\alpha_n))^2}.$$

To show (F.81), it is sufficient to show that

$$\mathbb{E}[\sum_{m=1}^{n}T^{(m)}] = \frac{n^5\alpha_n^2(1-\alpha_n)^2}{2}(1+o(1)), \tag{F.82}$$

and that

$$\mathrm{Var}(\sum_{m=1}^{n}T^{(m)}) = o(n^{10}\alpha_n^4). \tag{F.83}$$

Consider (F.82) first. Recall that $S^{(m)} = \{x = (i_1,i_2,i_3,i_4,i_5) \in S : \max\{i_1,i_2,i_3,i_4,i_5\} \le m\}$ 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

$$(S^{(m)}\backslash S^{(m-1)})^2 = \{(x,x') : x \in S^{(m)}\backslash S^{(m-1)}, x' \in S^{(m)}\backslash S^{(m-1)}\}.$$

Let

$$SS_1^{(m)} = \{(x,x') \in (S^{(m)}\backslash S^{(m-1)})^2 : i_3 = i_3', \{i_1,i_2,i_4,i_5\} = \{i_1',i_2',i_4',i_5'\}\},$$
$$SS_2^{(m)} = (S^{(m)}\backslash S^{(m-1)})^2\backslash SS_1^{(m)}.$$

It is seen that the LHS of (F.82) equals to

$$(I) + (II),$$

where

$$(I) = \mathbb{E}\Big[\sum_{m=1}^{n}\mathbb{E}[\sum_{(x,x')\in SS_1^{(m)}}\widetilde{\mathcal{A}}_{i_1i_2i_3}^2\widetilde{\mathcal{A}}_{i_3i_4i_5}^2|\mathcal{F}_{n,m-1}]\Big],$$

and

$$(II) = \mathbb{E}\Big[\sum_{m=1}^{n}\mathbb{E}[\sum_{(x,x')\in SS_2^{(m)}}\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{i_1'i_2'i_3'}\widetilde{\mathcal{A}}_{i_3'i_4'i_5'}|\mathcal{F}_{n,m-1}]\Big].$$

Notice that for any $(x,x') \in SS_2^m$, each $\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{i_1'i_2'i_3'}\widetilde{\mathcal{A}}_{i_3'i_4'i_5'}$ 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 $(i_1, i_2, i_4, i_5) = (i_4', i_5', i_1', i_2')$. By symmetry,

$$(I) = 2\sum_{m=1}^{n} \sum_{x \in S^{(m)}\setminus S^{(m-1)}} \mathbb{E}\Big[\widetilde{\mathcal{A}}_{i_1 i_2 i_3}^2 \widetilde{\mathcal{A}}_{i_3 i_4 i_5}^2\Big] = 2\sum_{x\in S} \alpha_n^2(1-\alpha_n)^2 = 12n\binom{n}{4}\alpha_n^2(1-\alpha_n^2).$$

Combining these gives (F.82).

Next, consider (F.83). In $S^{(m)}\setminus S^{(m-1)}$, we have $i_3 = m$ or $i_2 = m$ or $i_5 = m$. Let

$$\begin{aligned} S_1^{(m)} =& \big\{x \in S^{(m)}\setminus S^{(m-1)} : \text{either } i_2 = m, i_5 < m \text{ or } i_5 = m, i_2 < m\big\}, \\ S_2^{(m)} =& (S^{(m)}\setminus S^{(m-1)})\setminus S_1^{(m)}. \end{aligned}$$

Write

$$T^{(m)} = T_1^{(m)} + 2T_2^{(m)} + T_3^{(m)},$$

where

$$\begin{aligned} T_1^{(m)} =& \mathbb{E}[\sum_{x,x'\in S_1^{(m)}} \widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 i_5}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' i_5'}|\mathcal{F}_{n,m-1}], \\ T_2^{(m)} =& \mathbb{E}[\sum_{x\in S_1^{(m)}, x'\in S_2^{(m)}} \widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 i_5}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' i_5'}|\mathcal{F}_{n,m-1}], \\ T_3^{(m)} =& \mathbb{E}[\sum_{x,x'\in S_2^{(m)}} \widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 i_5}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' i_5'}|\mathcal{F}_{n,m-1}]. \end{aligned}$$

Notice that for $x \in S_1^{(m)}, x' \in S_2^{(m)}$, $\widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 i_5}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' i_5'}$ is mean-zero conditional on $\mathcal{F}_{n,m-1}$. 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 $E_m = \big\{(x,x') \in S_2^{(m)} \times S_2^{(m)} : \{i_1, i_2, i_3, i_4, i_5\} = \{i_1', i_2', i_3', i_4', i_5'\}\big\}$, 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,

$$T^{(m)} = T_1^{(m)} + |E_m|\alpha_n^2(1-\alpha_n)^2, \qquad \text{and} \qquad \mathrm{Var}(\sum_{m=1}^{n} T^{(m)}) = \mathrm{Var}(\sum_{m=1}^{n} T_1^{(m)}),$$

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

$$\mathrm{Var}(\sum_{m=1}^{n} T_1^{(m)}) = o(n^{10}\alpha_n^4). \tag{F.84}$$

By definitions and symmetry

$$T_1^{(m)} = \mathbb{E}[4 \sum_{\substack{1\le i_1,i_2,i_3,i_4,i_1',i_2',i_3',i_4'\le m-1 \\ i_1<i_2; i_1'<i_2' \\ i_1,i_2,i_4\neq i_3; i_1',i_2',i_4'\neq i_3'}} \widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 m}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' m}|\mathcal{F}_{n,m-1}].$$

If $\{i_3, i_4\} \neq \{i_3', i_4'\}$, then $\widetilde{\mathcal{A}}_{i_1 i_2 i_3}\widetilde{\mathcal{A}}_{i_3 i_4 m}\widetilde{\mathcal{A}}_{i_1' i_2' i_3'}\widetilde{\mathcal{A}}_{i_3' i_4' m}$ has a conditional mean of zero. Therefore, we have

$$T_1^{(m)} = T_{11}^{(m)} + T_{12}^{(m)},$$

where

$$T_{11}^{(m)} = \mathbb{E}[4 \sum_{\substack{1\le i_1,i_2,i_3,i_4,i_1',i_2'\le m-1 \\ i_1<i_2; i_1'<i_2' \\ i_1,i_2,i_1',i_2',i_4\neq i_3}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4m}^2\widetilde{\mathcal{A}}_{i_1'i_2'i_3}|\mathcal{F}_{n,m-1}],$$

$$T_{12}^{(m)} = \mathbb{E}[4 \sum_{\substack{1\le i_1,i_2,i_3,i_4,i_1',i_2'\le m-1 \\ i_1<i_2; i_1'<i_2' \\ i_1,i_2,i_4\neq i_3; i_1'i_2'\neq i_4}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4m}^2\widetilde{\mathcal{A}}_{i_1'i_2'i_4}|\mathcal{F}_{n,m-1}].$$

Since for any random variables X and Y, $\mathrm{Var}(X+Y)\le 2\mathrm{Var}(X)+2\mathrm{Var}(Y)$, to show (F.84), it is sufficient to show that

$$\mathrm{Var}(\sum_{m=1}^n T_{11}^{(m)}) = o(n^{10}\alpha_n^4), \qquad \text{and} \qquad \mathrm{Var}(\sum_{m=1}^n T_{12}^{(m)}) = o(n^{10}\alpha_n^4). \tag{F.85}$$

Consider the first claim in (F.85). Recall that

$$\widetilde{T}_{n,m} = \sum_{x\in S^{(m)}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5} = \sum_{\substack{1\le i_1,\cdots,i_5\le m \\ i_1<i_2; i_4<i_5 \\ i_1,i_2,i_4,i_5\neq i_3 \\ (i_1,i_2)\neq(i_4,i_5)}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}.$$

By elementary calculations

$$T_{11}^{(m)} = 4(m-2)\alpha_n(1-\alpha_n)\widetilde{T}_{n,m-1} + 4(m-2)\alpha_n(1-\alpha_n) \sum_{\substack{1\le i_1,i_2,i_3\le m-1 \\ i_1<i_2 \\ i_1,i_2\neq i_3}} \widetilde{\mathcal{A}}_{i_1i_2i_3}^2.$$

By inequality $\mathrm{Var}(X+Y)\le 2\mathrm{Var}(X)+2\mathrm{Var}(Y)$, to show the first claim in (F.85), it is sufficient to show that

$$\mathrm{Var}(\sum_{m=1}^n 4(m-2)\alpha_n(1-\alpha_n)\widetilde{T}_{n,m-1}) = o(n^{10}\alpha_n^4), \tag{F.86}$$

and

$$\mathrm{Var}(\sum_{m=1}^n 4(m-2)\alpha_n(1-\alpha_n) \sum_{\substack{1\le i_1,i_2,i_3\le m-1 \\ i_1<i_2 \\ i_1,i_2\neq i_3}} \widetilde{\mathcal{A}}_{i_1i_2i_3}^2) = o(n^{10}\alpha_n^4). \tag{F.87}$$

Consider the LHS of (F.86), by definitions,

$$\mathrm{Var}(\sum_{m=1}^n 4(m-2)\alpha_n(1-\alpha_n)\widetilde{T}_{n,m-1}) = \sum_{m,m'=1}^n 16(m-2)(m'-2)\alpha_n^2(1-\alpha_n)^2\mathrm{Cov}(\widetilde{T}_{n,m-1},\widetilde{T}_{n,m'-1}). \tag{F.88}$$

Notice that

$$\mathrm{Cov}(\widetilde{T}_{n,m-1},\widetilde{T}_{n,m'-1}) = \sum_{\substack{1\le i_1,\cdots,i_5\le m \\ i_1<i_2; i_4<i_5 \\ i_1,i_2,i_4,i_5\neq i_3 \\ (i_1,i_2)\neq(i_4,i_5)}} \sum_{\substack{1\le i_1',\cdots,i_5'\le m \\ i_1'<i_2'; i_4'<i_5' \\ i_1',i_2',i_4',i_5'\neq i_3' \\ (i_1',i_2')\neq(i_4',i_5')}} \mathbb{E}[\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{i_1'i_2'i_3'}\widetilde{\mathcal{A}}_{i_3'i_4'i_5'}].$$

Only if $\{i_1,i_2,i_3,i_4,i_5\} = \{i_1',i_2',i_3',i_4',i_5'\}$, $\mathbb{E}[\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{i_1'i_2'i_3'}\widetilde{\mathcal{A}}_{i_3'i_4'i_5'}]$ will be non-zero. Since there are only a bounded number of ways to pair the indexes, by direct calculations

$$\mathrm{Cov}(\widetilde{T}_{n,m-1},\widetilde{T}_{n,m'-1}) = O(\sum_{\substack{1\le i_1,\cdots,i_5\le m \\ i_1<i_2; i_4<i_5 \\ i_1,i_2,i_4,i_5\neq i_3 \\ (i_1,i_2)\neq(i_4,i_5)}} \mathbb{E}[(\widetilde{\mathcal{A}}_{i_1i_2i_3}^2\widetilde{\mathcal{A}}_{i_3i_4i_5}^2]) = O(n^5\alpha_n^2).$$

Combining this with (F.88), it is seen that

$$\text{Var}(\sum_{m=1}^{n} 4(m-2)\alpha_n(1-\alpha_n)\widetilde{T}_{n,m-1}) = O(n^4 n^5 \alpha_n^4) = o(n^{10}\alpha_n^4).$$

This proves (F.86).

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

$$\begin{aligned}\text{Var}(\sum_{m=1}^{n} 4(m-2)\alpha_n(1-\alpha_n) \sum_{\substack{1\le i_1,i_2,i_3\le m-1\\ i_1<i_2\\ i_1,i_2\neq i_3}} \widetilde{\mathcal{A}}^2_{i_1i_2i_3}) \le& 16n^4\alpha_n^2(1-\alpha_n)^2 \text{Var}(\sum_{\substack{1\le i_1,i_2,i_3\le n\\ i_1<i_2\\ i_1,i_2\neq i_3}} \widetilde{\mathcal{A}}^2_{i_1i_2i_3})\\ =& 16n^4\alpha_n^2(1-\alpha_n)^2 \cdot \sum_{\substack{1\le i_1,i_2,i_3\le n\\ i_1<i_2\\ i_1,i_2\neq i_3}} 3\cdot \text{Var}(\widetilde{\mathcal{A}}^2_{i_1i_2i_3})\\ =& O(n^7\alpha_n^3).\end{aligned}$$

By our assumption $n^2\widetilde{\alpha}_n \to \infty$ (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,

$$\text{Var}(\sum_{m=1}^{n} T_{12}^{(m)}) = \sum_{m,m'=1}^{n} 16\alpha_n^2(1-\alpha_n)^2 \sum_{\substack{1\le i_1,\cdots,i_6\le m\\ i_1<i_2;i_4<i_5\\ i_1,i_2\neq i_3;i_4,i_5\neq i_6\\ i_3\neq i_6}} \sum_{\substack{1\le i_1',\cdots,i_6'\le m\\ i_1'<i_2';i_4'<i_5'\\ i_1',i_2'\neq i_3';i_4',i_5'\neq i_6'\\ i_3'\neq i_6'}} \mathbb{E}[\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_4i_5i_6}\widetilde{\mathcal{A}}_{i_1'i_2'i_3'}\widetilde{\mathcal{A}}_{i_4'i_5'i_6'}].$$

Similarly, it is sufficient to consider terms that satisfy $\{i_1,\cdots,i_6\} = \{i_1',\cdots,i_6'\}$, hence

$$\text{Var}(\sum_{m=1}^{n} T_{12}^{(m)}) = O(\sum_{m,m'=1}^{n} 16\alpha_n^2(1-\alpha_n)^2 \sum_{\substack{1\le i_1,\cdots,i_6\le m\\ i_1<i_2;i_4<i_5\\ i_1,i_2\neq i_3;i_4,i_5\neq i_6\\ i_3\neq i_6}} \mathbb{E}[\widetilde{\mathcal{A}}^2_{i_1i_2i_3}\widetilde{\mathcal{A}}^2_{i_4i_5i_6}]) = O(n^8\alpha_n^4).$$

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

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

By Cauchy-Schwarz inequality

$$\big|\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2\mathbb{I}\{|X_{n,m}|>\epsilon\}|\mathcal{F}_{n,m-1}]\big| \le \sum_{m=1}^{n} \sqrt{\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]}\sqrt{\mathbb{P}(|X_{n,m}|>\epsilon|\mathcal{F}_{n,m-1})}. \tag{F.90}$$

At the same time, by Markov's inequality,

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

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

$$\big|\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2\mathbb{I}\{|X_{n,m}|>\epsilon\}|\mathcal{F}_{n,m-1}]\big| \le \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]/\epsilon^2.$$

To show (F.89), by Markov's inequality, it is sufficient to show that

$$\mathbb{E}\Big[\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^4|\mathcal{F}_{n,m-1}]\Big] \to 0. \tag{F.92}$$

Recall that

$$X_{n,m} = \frac{\sum_{x\in S^{(m)}\backslash S^{(m-1)}} \widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}}{\sqrt{2n}\binom{n-1}{2}\widetilde{\alpha}_n(1-\widetilde{\alpha}_n)}.$$

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, $y' = (i'_1, i'_2, i'_3, i'_4, i'_5, j'_1, j'_2, j'_3, j'_4, j'_5)$. To show (F.92), it is sufficient to show that

$$\mathbb{E}[\sum_{m=1}^{n}\sum_{y,y'\in(S^{(m)}\backslash S^{(m-1)})^2}\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{j_1j_2j_3}\widetilde{\mathcal{A}}_{j_3j_4j_5}\widetilde{\mathcal{A}}_{i'_1i'_2i'_3}\widetilde{\mathcal{A}}_{i'_3i'_4i'_5}\widetilde{\mathcal{A}}_{j'_1j'_2j'_3}\widetilde{\mathcal{A}}_{j'_3j'_4j'_5}] = o(n^{10}\alpha_n^4).$$

Similarly, to have non-zero expected value, $\widetilde{\mathcal{A}}_{i_1i_2i_3}\widetilde{\mathcal{A}}_{i_3i_4i_5}\widetilde{\mathcal{A}}_{j_1j_2j_3}\widetilde{\mathcal{A}}_{j_3j_4j_5}\widetilde{\mathcal{A}}_{i'_1i'_2i'_3}\widetilde{\mathcal{A}}_{i'_3i'_4i'_5}\widetilde{\mathcal{A}}_{j'_1j'_2j'_3}\widetilde{\mathcal{A}}_{j'_3j'_4j'_5}$ 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

$$\sum_{m=1}^{n}\sum_{y\in(S^{(m)}\backslash S^{(m-1)})^2}\mathbb{E}[\widetilde{\mathcal{A}}^2_{i_1i_2i_3}\widetilde{\mathcal{A}}^2_{i_3i_4i_5}\widetilde{\mathcal{A}}^2_{j_1j_2j_3}\widetilde{\mathcal{A}}^2_{j_3j_4j_5}] = o(n^{10}\alpha_n^4).$$

Recall that for each $x \in S^{(m)}\backslash S^{(m-1)}$, there are at least one index of $(i_1, i_2, i_3, i_4, i_5)$ is m. It is seen that

$$\sum_{m=1}^{n}\sum_{y\in(S^{(m)}\backslash S^{(m-1)})^2}\mathbb{E}[\widetilde{\mathcal{A}}^2_{i_1i_2i_3}\widetilde{\mathcal{A}}^2_{i_3i_4i_5}\widetilde{\mathcal{A}}^2_{j_1j_2j_3}\widetilde{\mathcal{A}}^2_{j_3j_4j_5}] \le \sum_{m=1}^{n} n^{10-2}\Big(\alpha_n(1-\alpha_n)\Big)^4 = o(n^{10}\alpha_n^4).$$

This finishes the proof.

G Proof of Theorem 3.2

Recall that $\phi_n = \max_{2\le m\le M}\{\phi_n^{(m)}\}$. To prove this theorem, it is sufficient to show that if there is a $m \in \{2,\dots,M\}$ such that $\|\theta^{(m)}\|_1^{m-2}\|\theta^{(m)}\|^2(\mu_2^{(m)})^2 \gg \log(n)$, we will have

$$\phi_n^{(m)} \to 0 \text{ under } H_0, \qquad \text{and } \phi_n^{(m)} \to \infty \text{ under } H_1.$$

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

Let

$$\widetilde{\alpha}_n = \mathbb{E}[\hat{\alpha}_n], \qquad \beta = \sum_{k_2,\dots,k_m=1}^{K}\mathcal{P}_{:k_2\cdots k_m}g_{k_2}\cdots g_{k_m}/([\mathcal{P};g,\dots,g])^{(m-1)/m}.,$$

where $g \in \mathbb{R}^K$ is defined by $g_k = (1/\|\theta\|_1)\sum_{i=1}^n \theta_i\pi_i(k)$, $1 \le k \le K$.

Introduce ideal counterparts of V_n and η by

$$\widetilde{V}_n = \binom{n}{m}\widetilde{\alpha}_n(1-\widetilde{\alpha}_n) \qquad \text{and} \qquad \eta^* = \Theta\Pi\beta, \qquad \text{respectively.} \tag{G.93}$$

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 $n \to \infty$,*

- *(a) Under both the null and alternative, $\widetilde{V}_n/V_n \to 1$ in probability.*

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

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 $1 - o(1)$

$$Q_n \le (n\log(n)^{1.1}V_n)^{1/2} \;\text{ under } H_0^{(n)}, \qquad Q_n \ge (n\log(n)^{1.1}V_n)^{1/2} \;\text{ under } H_1^{(n)}, \tag{G.94}$$

By (a) in Lemma G.1, $\widetilde{V}_n/V_n \to 1$ in probability. Hence to show (G.94), it is sufficient to show that with probability $1 - o(1)$

$$Q_n \le 0.5(n\log(n)^{1.1}\widetilde{V}_n)^{1/2} \;\text{ under } H_0^{(n)}, \qquad Q_n \ge 1.5(n\log(n)^{1.1}\widetilde{V}_n)^{1/2} \;\text{ under } H_1^{(n)}. \tag{G.95}$$

Recall that

$$Q_n = \max_{S=(S_1,\ldots,S_{m+1})\in B}\ \max_{1\le k_1,\ldots,k_m\le m+1}\{|X_{S,k_1\cdots k_m}|\},$$

where

$$X_{S,k_1\cdots k_m} = \sum_{\substack{i_1\in S_{k_1},\ldots,i_m\in S_{k_m}\\ i_1,\ldots,i_m(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \eta_{i_1}\cdots\eta_{i_m}).$$

Also, recall that η^* is the ideal counterparts of η, defined in (G.93). Introduce a counterpart of $X_{S,k_1\cdots k_m}$ by replacing η with η^*

$$\widetilde{X}_{S,k_1\cdots k_m} = \sum_{\substack{i_1\in S_{k_1},\ldots,i_m\in S_{k_m}\\ i_1,\ldots,i_m(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \eta^*_{i_1}\cdots\eta^*_{i_m}).$$

Let

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

Note that for any number $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$,

$$|\max\{x_1, x_2, \ldots, x_n\} - \max\{y_1, y_2, \ldots, y_n\}| \le \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\},$$

It is seen that

$$|Q_n - \widetilde{Q}_n| \le \max_S\ \max_{1\le k_1,\ldots,k_m\le m+1}\{|X_{S,k_1\cdots k_m} - \widetilde{X}_{S,k_1\cdots k_m}|\}. \tag{G.96}$$

At the same time, by definitions and direct calculations, for all $S = (S_1, \ldots, S_{m+1}) \in B$ and $1 \le k_1, \ldots, k_m \le m+1$

$$|X_{S,k_1\cdots k_m} - \widetilde{X}_{S,k_1\cdots k_m}| \le |S_{k_1}|\cdots|S_{k_m}| \max_{1\le i_1,\ldots,i_m\le n}|\eta_{i_1}\cdots\eta_{i_m} - \eta^*_{i_1}\cdots\eta^*_{i_m}|, \tag{G.97}$$

where by (b) and (c) in Lemma G.1, except for a probability $O(1/n)$

$$\max_{1\le i\le n}\Big\{\Big|\frac{\eta_i}{\eta_i^*} - 1\Big|\Big\} \le C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} \qquad \text{under } H_0, \tag{G.98}$$

and

$$\max_{1\le i\le n}\Big\{\Big|\frac{\eta_i}{\eta_i^*} - 1\Big|\Big\} \le C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C\gamma_n}{n} \qquad \text{under } H_1. \tag{G.99}$$

Here γ_n denotes $\max_{1\le k_1,\ldots,k_m\le K}\{|\mathcal{P}_{k_1\cdots k_m}-\beta_{k_1}\cdots\beta_{k_m}|\}$ 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, $\max_{1\le i\le n}\Big\{|\frac{\eta_i}{\eta_i^*}-1|\Big\}=o(1)$ under both hypotheses. By Taylor's expansion, for $1\le i_1,\ldots,i_m\le n$

$$|\eta_{i_1}\cdots\eta_{i_m}-\eta_{i_1}^*\cdots\eta_{i_m}^*|\le C\eta_{i_1}^*\cdots\eta_{i_m}^*\max_{1\le i\le n}\Big\{|\frac{\eta_i}{\eta_i^*}-1|\Big\}. \tag{G.100}$$

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

$$|Q_n-\widetilde{Q}_n|\le C\Big(\log(n)n^{m+1}\theta_{\max}^m\Big)^{1/2}\qquad\text{under } H_0, \tag{G.101}$$

and

$$|Q_n-\widetilde{Q}_n|\le C\Big(\log(n)n^{m+1}\theta_{\max}^m\Big)^{1/2}+C\gamma_n n^{m-1}\theta_{\max}^m\qquad\text{under } H_1 \tag{G.102}$$

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

$$(I):\widetilde{Q}_n\le 0.5(n\log(n)^{1.1}\widetilde{V}_n)^{1/2}\qquad\text{under } H_0^{(n)},$$
$$(II):\widetilde{Q}_n\ge 2(n\log(n)^{1.1}\widetilde{V}_n)^{1/2}+C\gamma_n n^{m-1}\theta_{\max}^m\qquad\text{under } H_1^{(n)}.$$

Consider (I) first. Recall that

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

where the RHS is the maximum of

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

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

$$\big|\widetilde{X}_{S,k_1\cdots k_m}\big|\le 0.5(n\log(n)^{1.1}\widetilde{V}_n)^{1/2}. \tag{G.103}$$

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

$$\widetilde{X}_{S,k_1\cdots k_m}=\sum_{\substack{i_1\in S_{k_1},\ldots,i_m\in S_{k_m}\\ i_1,\ldots,i_m(dist)}}(\mathcal{A}_{i_1\cdots i_m}-\theta_{i_1}\cdots\theta_{i_m}),$$

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 $\le(m!)^2\theta_{i_1}\cdots\theta_{i_m}(1-\theta_{i_1}\cdots\theta_{i_m})$. By Bernstein's inequality, for any $t>0$,

$$\mathbb{P}\big(\big|\widetilde{X}_{S,k_1\cdots k_m}\big|\ge t\big)\le 2\exp\Bigg(-\frac{t^2}{\sum_{\substack{i_1\in S_{k_1},\ldots,i_m\in S_{k_m}\\ i_1,\ldots,i_m(unique)}}(m!)^2\theta_{i_1}\cdots\theta_{i_m}(1-\theta_{i_1}\cdots\theta_{i_m})+t/3}\Bigg).$$

Since $\sum_{\substack{i_1\in S_{k_1},\ldots,i_m\in S_{k_m}\\ i_1,\ldots,i_m(unique)}}(m!)^2\theta_{i_1}\cdots\theta_{i_m}(1-\theta_{i_1}\cdots\theta_{i_m})\le Cn^m\theta_{\max}^m$, it follows that

$$\mathbb{P}\big(\big|\widetilde{X}_{S,k_1\cdots k_m}\big|\ge t\big)\le 2\exp\Big(-\frac{t^2}{Cn^m\theta_{\max}^m+t/3}\Big). \tag{G.104}$$

Taking $t = (n\log(n)\widetilde{V}_n)^{1/2}$ and noting that $(1/C)\sqrt{n^{m+1}\log(n)\theta_{\max}^m} \leq t \leq \sqrt{n^{m+1}\log(n)\theta_{\max}^m}$,

$$\exp\Big(-\frac{t^2}{Cn^m\theta_{\max}^m + t/3}\Big) \leq \exp\Big(-\frac{\Big((1/C)\sqrt{n^{m+1}\log(n)\theta_{\max}^m}\Big)^2}{Cn^m\theta_{\max}^m + \sqrt{n^{m+1}\log(n)\theta_{\max}^m}/3}\Big).$$

Combining this with our assumption $\|\theta\|_1^{m-2}\|\theta\|^2/\log(n) \to \infty$ and $\theta_{\max} \leq C\theta_{\min}$, by elementary calculations, the RHS of (G.104) is $O(\exp(-Cn\log(n)))$. This proves (G.103).

Next, consider (II) for the alternative case. Let S_k^* 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 $(k_1, \ldots, k_m) = (k_1^*, \ldots, k_m^*)$ and so

$$\gamma_n = |\mathcal{P}_{k_1^*\cdots k_m^*} - \beta_{k_1^*}\cdots\beta_{k_m^*}|.$$

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

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

$$\widetilde{Q}_n \geq |\widetilde{X}_{S^*,k_1^*\cdots k_m^*}|.$$

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

$$|\widetilde{X}_{S^*,k_1^*\cdots k_m^*}| \geq C(n\log(n)^{1.1}\widetilde{V}_n)^{1/2} + C\gamma_n n^{m-1}\theta_{\max}^m. \tag{G.105}$$

Write

$$\widetilde{X}_{S^*,k_1^*\cdots k_m^*} := \sum_{i_1\in S_{k_1^*},\ldots,i_m\in S_{k_m^*}} (\mathcal{A}_{i_1\cdots i_m} - \eta_{i_1}^*\cdots\eta_{i_m}^*) = (I) + (II), \tag{G.106}$$

where

$$(I) = \sum_{i_1\in S_{k_1^*},\ldots,i_m\in S_{k_m^*}} (\theta_{i_1}\cdots\theta_{i_m}\mathcal{P}_{k_1^*\ldots k_m^*} - \eta_{i_1}^*\cdots\eta_{i_m}^*),$$

and

$$(II) = \sum_{i_1\in S_{k_1^*},\ldots,i_m\in S_{k_m^*}} (\mathcal{A}_{i_1\cdots i_m} - \theta_{i_1}\cdots\theta_{i_m}\mathcal{P}_{k_1^*\cdots k_m^*}).$$

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

$$|(I)| = \|\theta\|_1^m g_{k_1^*}\cdots g_{k_m^*}\gamma_n.$$

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

$$\|\theta\|_1^m g_{k_1^*}\cdots g_{k_m^*} \geq Cn^m\theta_{\max}^m,$$

and so

$$|(I)| \geq Cn^m\theta_{\max}^m\gamma_n. \tag{G.107}$$

Write for short

$$N = |S_{k_1^*}^*|\cdots|S_{k_m^*}^*|.$$

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 $C\theta_{\max}^m$. By Bernstein's Lemma, for any $t > 0$,

$$\mathbb{P}(|(II)| \geq t) \leq \exp(-\frac{t^2}{NC\theta_{\max}^m + t/3}). \tag{G.108}$$

Taking $t = (\log(n)\widetilde{V}_n)^{1/2}$. 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 $O(1/n)$,

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

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

$$|\widetilde{X}_{S^*,k_1^*\cdots k_m^*}| \geq Cn^m\theta_{\max}^m\gamma_n - (\log(n)\widetilde{V}_n)^{1/2}, \tag{G.110}$$

where we note that by Lemma G.1, $n^m\theta_{\max}^m\gamma_n/(n\log(n)^{1.1}\widetilde{V}_n)^{1/2} \to \infty$. This proves (G.105) and finishes the proof.

G.2 Proof of Lemma G.1

Consider the claim (a). By definitions

$$\frac{V_n}{\widetilde{V}_n} - 1 = \frac{(\hat{\alpha}_n - \widetilde{\alpha}_n)(1 - \hat{\alpha}_n - \widetilde{\alpha}_n)}{\widetilde{\alpha}_n(1 - \widetilde{\alpha}_n)}. \tag{G.111}$$

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,

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

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

$$\mathbb{P}\Big(|\hat{\alpha}_n - \widetilde{\alpha}_n| \geq C\log(n)(\theta_{\max}^m/\tbinom{n}{m})^{1/2}\Big) \leq o(1/n).$$

Combining this with (G.111) and $\widetilde{\alpha}_n \leq C\theta_{\max}^m \leq Cc_0^m < 1$, by elementary calculations,

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

where by our conditions $n^{m-1}\theta_{\max}^m/\log(n) \to \infty$ (implied by $\|\theta\|_1^{m-2}\|\theta\|^2\mu_2^2/\log(n) \to \infty$), the RHS is $o(1)$. Therefore $V_n/\widetilde{V}_n \to 1$ in probability.

Combining this with Slutsky's Lemma, we get $\widetilde{V}_n/V_n \to 1$ 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 $O(1/n)$

$$\max_{1\leq i\leq n}\Big\{|\frac{\eta_i}{\eta_i^*} - 1|\Big\} \leq C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2}, \qquad \text{under } H_0 \tag{G.112}$$

and

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

Recall that

$$\eta = u^{(\lceil\frac{m-1}{2}\rceil)} \qquad \text{and} \qquad u^{(k)} = g(u^{(k-1)}), \qquad 1 \leq k \leq m,$$

where for $1 \le i \le n$

$$L_{i_1}(u) = \frac{\sum_{i_2,\dots,i_m(\text{distinct})} \mathcal{A}_{i_1\cdots i_m} + \sum_{i_2,\dots,i_m(\text{non-distinct})} u_{i_1}\cdots u_{i_m}}{\left(\sum_{i_1,\dots,i_m(\text{distinct})} \mathcal{A}_{i_1\cdots i_m} + \sum_{i_1,\dots,i_m(\text{non-distinct})} u_{i_1}\cdots u_{i_m}\right)^{(m-1)/m}}.$$

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

$$\begin{aligned} E_1 :& \max_{1\le i_1\le n}\Big\{\Big|\sum_{\substack{i_2,\dots,i_m\in I^{(i_1)}\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m})\Big|\Big\} \le (n^{m-1}\theta_{\max}^m \log(n))^{1/2}, \\ E_2 :& \qquad \Big|\sum_{\substack{i_1,\dots,i_m\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m})\Big| \le (n^m\theta_{\max}^m)^{1/2} \end{aligned} \tag{G.114}$$

hold then for $1 \le k \le m$

$$\max_{1\le i\le n}\{|\frac{L_i(u^{(k)})}{\eta_i^*} - 1|\} \le C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\le i\le n}\{|\frac{u_i^{(k)}}{\eta_i^*} - 1|\} + C\frac{\gamma_n}{n}, \tag{G.115}$$

where by definitions γ_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

$$\begin{aligned} \max_{1\le i\le n}\{|\frac{u_i^{(k)}}{\eta_i^*} - 1|\} &= \max_{1\le i\le n}\{|\frac{L_i(u^{(k-1)})}{\eta_i^*} - 1|\} \\ &\le C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\le i\le n}\{|\frac{u_i^{(k-1)}}{\eta_i^*} - 1|\} + \frac{C\gamma_n}{n} \\ &\overset{\cdots}{\le} C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2}(1+o(1)) + \frac{C}{n^k}\max_{1\le i\le n}\{|\frac{u_i^{(0)}}{\eta_i^*} - 1|\} + \frac{C\gamma_n}{n}(1+o(1)) \\ \text{(Note that } u^{(0)} = 0) \qquad &\le C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n^k} + \frac{C\gamma_n}{n}. \end{aligned}$$

Combining this with $\eta = u^{(\lceil\frac{m-1}{2}\rceil)}$, it follows that n^{-k} $(k = \lceil\frac{m-1}{2}\rceil)$ is a minor term and so $\max_{1\le i\le n}\{|\eta_i/\eta_i^* - 1|\} \le C(\log(n)/n^{m-1}\theta_{\max}^m)^{1/2} + C\gamma_n/n$ (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 $1 \le k \le m$ 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 $1 \le i_1 \le n$, note that by symmetry,

$$\sum_{\substack{i_2,\dots,i_m\in I^{(i_1)}\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}) = \sum_{i_2<\cdots<i_m\in I^{(i_1)}} (m-1)!(\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}),$$

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

$$\mathbb{P}\Big(\sum_{\substack{i_2,\dots,i_m\in I^{(i_1)}\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}) > t_1\Big) \le \exp(-\frac{t_1^2}{Cn^{m-1}\theta_{\max}^m + t_1/3}).$$

Similarly, for event E_2, we have for any $t_2 > 0$

$$\mathbb{P}\Big(\sum_{\substack{i_1,\dots,i_m\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}) > t_2\Big) \le \exp(-\frac{t_2^2}{Cn^m\theta_{\max}^m + t_2/3}).$$

Letting $t_1 = \sqrt{2C}(n^{m-1}\theta_{\max}^m \log(n))^{1/2}$ and $t_2 = (n^m\theta_{\max}^m)^{1/2}$ and by direct calculations

$$\mathbb{P}\Big(\sum_{\substack{i_2,\ldots,i_m\in I^{(i_1)}\\(dist)}} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}) > \sqrt{2C}(n^{m-1}\theta_{\max}^m \log(n))^{1/2}\Big) \leq \exp(-2\log(n)) = O(1/n^2).$$

and

$$\mathbb{P}\Big(\sum_{i_1,\ldots,i_m} (\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m}) > (n^m\theta_{\max}^m)^{1/2}\Big) \leq \exp(-n/C) = o(1/n^2).$$

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, η^* can be written as

$$\eta^* = \frac{\sum_{i_2,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}{(\sum_{i_1,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m})^{\frac{m-1}{m}}}.$$

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

$$\frac{L_{i_1}(u^{(k)})}{\eta_{i_1}^*} = (I^{(k)})_{i_1}(II^{(k)})_{i_1}^{-\frac{m-1}{m}},$$

where

$$(I^{(k)})_{i_1} = \frac{\sum_{i_2,\ldots,i_m(\text{distinct})} \mathcal{A}_{i_1\cdots i_m} + \sum_{i_2,\ldots,i_m(\text{non-distinct})} u_{i_1}^{(k)}\cdots u_{i_m}^{(k)}}{\sum_{i_2,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}$$

and

$$(II^{(k)})_{i_1} = \frac{\sum_{i_1,\ldots,i_m(\text{distinct})} \mathcal{A}_{i_1\cdots i_m} + \sum_{i_1,\ldots,i_m(\text{non-distinct})} u_{i_1}^{(k)}\cdots u_{i_m}^{(k)}}{\sum_{i_1,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}.$$

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

$$\max_{1\leq i\leq n}\{|(I^{(k)})_i - 1|\} = o(1), \qquad \max_{1\leq i\leq n}\{|(II^{(k)})_i - 1|\} = o(1), \tag{G.116}$$

$$\max_{1\leq i\leq n}\{|(I^{(k)})_i - 1|\} \leq C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\leq i\leq n}\{|\frac{u_i^{(k)}}{\eta_i^*} - 1|\} + C\frac{\gamma_n}{n} \tag{G.117}$$

and that

$$\max_{1\leq i\leq n}\{|(II^{(k)})_i - 1|\} \leq C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\leq i\leq n}\{|\frac{u_i^{(k)}}{\eta_i^*} - 1|\} + C\frac{\gamma_n}{n}. \tag{G.118}$$

Note that by triangle's inequality,

$$\begin{aligned}|(I^{(k)})_{i_1} - 1| \leq & \Big|\frac{\sum_{\substack{i_2,\ldots,i_m\in I^{(i_1)}\\(\text{distinct})}}(\mathcal{A}_{i_1\cdots i_m} - \mathcal{Q}_{i_1\cdots i_m})}{\sum_{i_2,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}\Big| \\ &+ \Big|\frac{\sum_{\substack{i_2,\ldots,i_m\\(\text{non-distinct})}}(u_{i_1}^{(k)}\cdots u_{i_m}^{(k)} - \eta_{i_1}^*\cdots\eta_{i_m}^*)}{\sum_{i_2,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}\Big| \\ &+ \Big|\frac{\sum_{\substack{i_2,\ldots,i_m\\(\text{non-distinct})}}(\eta_{i_1}^*\cdots\eta_{i_m}^* - \mathcal{Q}_{i_1\cdots i_m})}{\sum_{i_2,\ldots,i_m=1}^n \mathcal{Q}_{i_1\cdots i_m}}\Big|.\end{aligned}$$

By event E_1 and $\mathcal{Q}_{i_1\cdots i_m} \asymp \theta_{\max}^m$, 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, $|\eta_{i_1}^*\cdots\eta_{i_m}^* - \mathcal{Q}_{i_1\cdots i_m}| \leq \theta_{i_1}\cdots\theta_{i_m}\gamma_n$. It follows that

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

Similarly, by event E_2 and elementary calculations, we have

$$\begin{aligned}|(II^{(k)})_{i_1} - 1| \leq& C\Big(\frac{1}{n^m\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\leq i_1,\ldots,i_m\leq n}\Big\{\Big|\frac{u_{i_1}^{(k)}\cdots u_{i_m}^{(k)}}{\eta_{i_1}^*\cdots\eta_{i_m}^*} - 1\Big|\Big\} + C\frac{\gamma_n}{n} \\ \leq& C\Big(\frac{\log(n)}{n^{m-1}\theta_{\max}^m}\Big)^{1/2} + \frac{C}{n}\max_{1\leq i_1,\ldots,i_m\leq n}\Big\{\Big|\frac{u_{i_1}^{(k)}\cdots u_{i_m}^{(k)}}{\eta_{i_1}^*\cdots\eta_{i_m}^*} - 1\Big|\Big\} + C\frac{\gamma_n}{n}.\end{aligned} \qquad \text{(G.120)}$$

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

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

where we recall that our original goal is to show

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

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

$$\max_{1\leq i_1,\ldots,i_m\leq n}\Big|\frac{u_{i_1}^{(0)}\cdots u_{i_m}^{(0)}}{\eta_{i_1}^*\cdots\eta_{i_m}^*} - 1\Big| = 1 = \max_{1\leq i\leq n}\Big\{\Big|\frac{u_i^{(0)}}{\eta_i^*} - 1\Big|\Big\}.$$

Combining this with (G.119)-(G.120), we get (G.116)-(G.118) hold for $k = 0$. It follows that

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

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

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

$$\mathcal{G}_{k_1\cdots k_m} = \beta_{k_1}\cdots\beta_{k_m}, \qquad 1\leq k_1,\ldots,k_m\leq K,$$

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

$$|\sigma_2(P) - \sigma_2(G)| \leq \|P - G\|, \qquad \text{(G.121)}$$

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

$$G_{:,k_2+\sum_{j=3}^m K^{k_j-1}(k_j-1)} = \beta\cdot(\beta_{k_2}\cdots\beta_{k_m}), \qquad 1\leq k_2,\ldots,k_m\leq K.$$

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 $\|P - G\| \leq C\max_{1\leq k_1,\ldots,k_m\leq K}\{|\mathcal{P}_{k_1\cdots k_m} - \beta_{k_1}\cdots\beta_{k_m}|\} = C\gamma_n$, we obtain

$$|\mu_2| \leq \|P - G\| \leq C\gamma_n.$$

By our assumption $\|\theta\|_1^{m-2}\|\theta\|^2\mu_2^2/\log(n)^{1.1} \to \infty$ and $\theta_{\max} \leq C\theta_{\min}$, the above inequality implies $n^{m-1}\theta_{\max}^m\gamma_n^2/\log(n)^{1.1} \to \infty$. It follows that

$$n^m\theta_{\max}^m\gamma_n/(n^{m+1}\theta_{\max}^m\log(n)^{1.1})^{1/2} = C(n^{m-1}\theta_{\max}^m\gamma_n^2/\log(n)^{1.1})^{1/2} \to \infty.$$

This proves the last claim in (c).

References

[1] Ravindra Bapat. D1ad2 theorems for multidimensional matrices. *Linear Algebra and its Applications*, 48:437–442, 1982.

[2] Shmuel Friedland. Positive diagonal scaling of a nonnegative tensor to one with prescribed slice sums. *Linear algebra and its applications*, 434(7):1615–1619, 2011.

[3] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 1980.

[4] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[5] Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.