Electronic Journal of Statistics
Vol. 15 (2021) 2483–2516
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1838

Sparse random tensors: Concentration, regularization and applications*

Zhixin Zhou

City University of Hong Kong
e-mail: zhixzhou@cityu.edu.hk

Yizhe Zhu

University of California, San Diego
e-mail: yiz084@ucsd.edu

Abstract: We prove a non-asymptotic concentration inequality for the spectral norm of sparse inhomogeneous random tensors with Bernoulli entries. For an order-k inhomogeneous random tensor T with sparsity $p_{\max} \geq \frac{c \log n}{n}$, we show that $\|T - \mathbb{E}T\| = O(\sqrt{np_{\max}} \log^{k-2}(n))$ with high probability. The optimality of this bound up to polylog factors is provided by an information theoretic lower bound. By tensor unfolding, we extend the range of sparsity to $p_{\max} \geq \frac{c \log n}{n^m}$ with $1 \leq m \leq k-1$ and obtain concentration inequalities for different sparsity regimes. We also provide a simple way to regularize T such that $O(\sqrt{n^m p_{\max}})$ concentration still holds down to sparsity $p_{\max} \geq \frac{c}{n^m}$ with $k/2 \leq m \leq k-1$. We present our concentration and regularization results with two applications: (i) a randomized construction of hypergraphs of bounded degrees with good expander mixing properties, (ii) concentration of sparsified tensors under uniform sampling.

MSC2020 subject classifications: Primary 15B52; secondary 60C05.
Keywords and phrases: Sparse random tensor, spectral norm, hypergraph expander, tensor sparsification.

Received February 2020.

Contents

*Y.Z. is partially supported by NSF DMS-1949617.

1. Introduction

Tensors have been popular data formats in machine learning and network analysis. The statistical models on tensors and the related algorithms have been widely studied in the last ten years, including tensor decomposition [1, 29], tensor completion [34, 48], tensor sketching [49], tensor PCA [53, 17, 7], and community detection on hypergraphs [36, 30, 50, 20]. This raises the urgent demand for random tensor theory, especially the concentration inequalities in a non-asymptotic point of view. There are several concentration results of sub-Gaussian random tensors [55] and Gaussian tensors [3, 53, 49]. Recently concentration inequalities for rank-1 tensor were also studied in [59] with application to the capacity of polynomial threshold functions [5]. In many of the applications in data science, the sparsity of the random tensor is an important aspect. However, there are only a few results for the concentration of order-3 sparse random tensors [34, 41], and not much is known for general order-k sparse random tensors.

Inspired by discrepancy properties in random hypergraph theory, we prove concentration inequalities on sparse random tensors in the measurement of the tensor spectral norm. Previous results for tensors include the concentration of sub-Gaussian tensors and expectation bound on the spectral norm for general random tensors [55, 49]. The sparsity parameter does not appear in those bounds and directly applying those results would not get the desired concentration for sparse random tensors.

To simplify our presentation, we focus on real-valued order-k $n \times \cdots \times n$ tensors, while the results can be extended to tensors with other dimensions. We denote the set of these tensors by $\mathbb{R}^{n^k}$. We first define the Frobenius inner product and spectral norm for tensors.

Definition 1.1 (Frobenius inner product and spectral norm)**.** *For order-k $n \times \cdots \times n$ tensors T and A, the Frobenius inner product is defined by the sum of entrywise products:*

$$\langle T, A\rangle := \sum_{i_1,\ldots,i_k\in[n]} t_{i_1,\ldots,i_k} a_{i_1,\ldots,i_k},$$

and the Frobenius norm is defined by $\|T\|_F := \sqrt{\langle T, T\rangle}$. *Let* $x_1 \otimes \cdots \otimes x_k \in \mathbb{R}^{n^k}$ *be the outer product of vectors* $x_1, \ldots, x_k \in \mathbb{R}^n$, *i.e.,*

$$(x_1 \otimes \cdots \otimes x_k)_{i_1,\ldots,i_k} = x_{1,i_1} \cdots x_{k,i_k}$$

for $i_1, \ldots, i_k \in [n]$. *Then the spectral norm of* T *is defined by*

$$\begin{aligned}\|T\| :&= \sup_{\|x_1\|_2=\cdots=\|x_k\|_2=1} |\langle T, x_1 \otimes \cdots \otimes x_k\rangle| \\ &= \sup_{\|x_1\|_2=\cdots=\|x_k\|_2=1} \left| \sum_{i_1,\ldots,i_k\in[n]} t_{i_1,\ldots,i_k} x_{1,i_1} \cdots x_{k,i_k} \right|.\end{aligned}$$

For the ease of notation, we denote the Frobenius inner product between a tensor T and a tensor $x_1 \otimes \cdots \otimes x_k$ by

$$T(x_1, \ldots, x_k) := \langle T, x_1 \otimes \cdots \otimes x_k\rangle,$$

which can be seen as a multi-linear form on $x_1, \ldots, x_k$. It is worth noting that $\|T\| \le \|T\|_F$ since the following inequality holds:

$$\|T\| = \sup_{\|x_1\|_2=\cdots=\|x_k\|_2=1} |T(x_1, \ldots, x_k)| \le \sup_{A:\|A\|_F\le 1} |\langle T, A\rangle| = \|T\|_F. \tag{1.1}$$

In general, it is NP-hard to compute the spectral norm of tensors for $k \ge 3$ [33]. However, it would be possible to show the concentration of sparse random tensors in the measurement of the spectral norm with high probability.

There have been many fruitful results on the concentration of random matrices, including the sparse ones. We briefly discuss different proof techniques and their difficulty and limitation for generalization to random tensors. For sub-Gaussian matrices, an ε-net argument will quickly give a desired spectral norm bound [58]. For Gaussian matrices, one could relate the spectral norm to the maximal of a certain Gaussian process [57]. Another powerful way to derive a good spectral norm bound for random matrices is called the high moment method. Considering a centered $n \times n$ Hermitian random matrix A, for any integer k, its spectral norm satisfies $\mathbb{E}[\|A\|^{2k}] \le \mathbb{E}[\mathrm{tr}(A^{2k})]$. By taking k growing with n, if one can have a good estimate of $\mathbb{E}[\mathrm{tr}(A^{2k})]$, it implies a good concentration bound on $\|A\|$. It's well-known that computing the trace of a random matrix is equivalent to counting a certain class of cycles in a graph. This type of argument, together with some more refined modifications and variants (e.g. estimating high moments for the corresponding non-backtracking operator), is particularly useful for bounding the spectral norm of sparse random matrices, see [60, 6, 9, 39, 12].

A different approach is called the Kahn-Szemerédi argument, which was first applied to obtain the spectral gap of random regular graphs [26], and was extensively applied to other random graph models [15, 46, 23, 21, 54, 63]. In particular, this argument was used in [25, 42] to estimate the largest eigenvalue of sparse Erdős-Rényi graphs. Although using this method one cannot obtain

the exact constant of the spectral norm, it does capture the right order on n and the sparsity parameter p.

A natural question is how those methods can be applied to study the spectral norm of random tensors. For sub-Gaussian random tensors of order k, the ε-net argument would give us a spectral norm bound $O(\sqrt{n})$ [55]. However, the dependence on the order k might not be optimal, and it cannot capture the sparsity in the sparse random tensor case. For Gaussian random tensors, surprisingly, none of the above approaches could obtain a sharp spectral norm bound with the correct constant. Instead, the exact asymptotic spectral norm was given in [3] using techniques from spin glasses. This is also the starting point for a line of further research: tensor PCA and spiked tensor models under Gaussian noise, see for example [53, 44, 17, 7]. The tools from spin glasses rely heavily on the assumption of Gaussian distribution and cannot be easily adapted to non-Gaussian cases.

One might try to develop a high moment method for random tensors. Unfortunately, there is no natural generalization of the trace or eigenvalues for tensors that match our cycle counting interpretation in the random matrix case. Instead, by projecting the random tensor into a matrix form (including the adjacency matrix, self-avoiding matrix, and the non-backtracking matrix of a hypergraph), one could still apply the moment method to obtain some information of the original tensor or hypergraph, see [45, 50, 24, 2]. This approach is particularly useful for the study of community detection problems on random hypergraphs. However, after reducing the adjacency tensor into an adjacency matrix, there is a strict information loss and one could not get the exact spectral norm information of the original tensor. Due to the barrier of extending other methods to sparse random tensors, we generalize the Kahn-Szemerédi argument to obtain a good spectral norm bound when $p \geq \frac{c \log n}{n}$.

2. Main results

2.1. Concentration

Let $P = (p_{i_1,\dots,i_k}) \in [0,1]^{n^k}$ be an order-k tensor and T be a random tensor with independent entries such that

$$t_{i_1,\dots,i_k} \sim \text{Bernoulli}(p_{i_1,\dots,i_k}), \quad \text{where in particular, } P = \mathbb{E}T. \tag{2.1}$$

To control the sparsity of random tensor, we introduce the parameter for maximal probability

$$p := p_{\max} := \max_{i_1,\dots,i_k \in [n]} p_{i_1,\dots,i_k}.$$

Note that when $k = 2$, np is the maximal expected degree parameter in [25, 42, 40]. For two order-k tensors $A, B \in \mathbb{R}^{n^k}$, define the Hadamard product $A \circ B$ as

$$(A \circ B)_{i_1,\dots,i_k} := A_{i_1,\dots,i_k} B_{i_1,\dots,i_k}.$$

Now we are ready to state our first main result, which is a generalization of the case when $k = 2$ in [25, 42] to all $k \geq 2$.

Theorem 2.1. *Let $k \geq 2$ be fixed. Let A be a deterministic tensor of order k, and T be a random tensor of order k with Bernoulli entries. Assume $p \geq \frac{c \log n}{n}$ for some constant $c > 0$. Then for any $r > 0$, there is a constant $C > 0$ depending only on r, c, k such that with probability at least $1 - n^{-r}$,*

$$\|A \circ T - \mathbb{E}[A \circ T]\| \leq C\sqrt{np} \log^{k-2}(n) \max_{i_1,\ldots,i_k} |A_{i_1,\ldots,i_k}|. \tag{2.2}$$

Remark 2.1. *From the proof of Theorem 2.1 in Section 4, the constant C in (2.2) depends exponentially on k and linearly on r.*

In fact, we can provide a lower bound on the spectral norm as follows, which shows (2.2) is tight up to a polylog factor.

Theorem 2.2. *Let $k \geq 2$ be fixed and T be a random tensor of order k with Bernoulli entries and $\mathbb{E}T_{i_1,\ldots,i_k} = p$ for $i_1, \ldots, i_k \in [n]$. Assume $p = o(1)$ and $np \to \infty$ as $n \to \infty$. Then with high probability,*

$$\|T - \mathbb{E}T\| \geq \sqrt{np}.$$

From (1.1), when $p \geq \frac{c \log n}{n^{k-1}}$, a concentration bound by Bernstein inequality on $\|T - \mathbb{E}T\|_F$ implies $\|T - \mathbb{E}T\| = O(\sqrt{n^k p})$ with high probability. Applying tensor unfolding, we can improve this bound and obtain concentration inequalities for different sparsity ranges as follows.

Theorem 2.3. *Let $k \geq 2$ be fixed and T be a random tensor defined in (2.1). Assume $p \geq \frac{c \log n}{n^m}$ for some constant $c > 0$ and an integer m such that $k/2 \leq m \leq k - 1$. Then for any $r > 0$, there is a constant $C > 0$ depending only on r, c, k such that with probability at least $1 - n^{-r}$,*

$$\|T - \mathbb{E}T\| \leq C\sqrt{n^m p}. \tag{2.3}$$

Assume $p \geq \frac{c \log n}{n^m}$ with $1 \leq m < k/2$. Then there is a constant $C > 0$ depending on r, c, k such that with probability at least $1 - n^{-r}$,

$$\|T - \mathbb{E}T\| \leq \begin{cases} C\sqrt{n^m p} \log^{\frac{k-1}{m}-1}(n) & \text{if } k/m \notin \mathbb{Z}, \\ C\sqrt{n^m p} \log^{\frac{k}{m}-2}(n) & \text{if } k/m \in \mathbb{Z}. \end{cases} \tag{2.4}$$

Remark 2.2. *In (2.2), the factor $\sqrt{np}$ corresponds to the Euclidean norm of a fiber (a vector obtained by fixing all but one indices in a tensor) of T. The same applies in Theorem 2.3 after tensor unfolding. The factor $\sqrt{n^m p}$ in (2.3) corresponds to the Euclidean norm of a row in the unfolded $n^{k-m} \times n^m$ matrix form of T. Similarly, the same factor in (2.4) corresponds to the Euclidean norm of an n^m-dimensional fiber in an unfolded tensor form of T.*

Remark 2.3. *When $p = \frac{c \log n}{n^m}$ with $1 \leq m \leq k - 1$, the inequalities in Theorem 2.3 are tight up to polylog factors. Note that (2.3) and (2.4) imply $\|T - \mathbb{E}T\| \leq C\sqrt{c} \log^{k-1/2}(n)$. On the other hand, when $\frac{c \log n}{n^{k-1}} \leq p \leq \frac{1}{2}$, with high probability the random tensor T has at least one non-zero entry. By the inequality $\|T - \mathbb{E}T\| \geq \max_{i_1,\ldots,i_k} |T_{i_1,\ldots,i_k} - p_{i_1,\ldots,i_k}|$, we have $\|T - \mathbb{E}T\| \geq 1/2$ with high probability.*

2.2. Minimax lower bound

Consider the problem of constructing an estimator of $\mathbb{E}T$ under the spectral norm based on T. We show that the high probability bound in Theorem 2.1 is optimal up to the logarithm term in the minimax sense.

Theorem 2.4. *Suppose we observe a Bernoulli random tensor T with independent entries and $\mathbb{E}T = \theta$ for $\theta \in [0,p]^{n^k}$, where $p \in (0,1]$ and $n \geq 16$. Then there exists constants $c_1, c_2 > 0$ only depending on k such that*

$$\inf_{\hat{\theta}} \sup_{\theta \in [0,p]^{n^k}} \mathbb{P}\left(\|\hat{\theta} - \theta\| \geq (c_1\sqrt{np}) \wedge (c_2 n^{k/2} p) \right) \geq \frac{1}{3},$$

where the infimum is taken over all functions $\hat{\theta} : \mathbb{R}^{n^k} \to \mathbb{R}^{n^k}$, $T \mapsto \hat{\theta}(T)$. In particular, if $p \geq \frac{c \log n}{n}$, then there exists constant $c_3 > 0$ only depending on k and c such that

$$\inf_{\hat{\theta}} \sup_{\theta \in [0,p]^{n^k}} \mathbb{P}\left(\|\hat{\theta} - \theta\| \geq c_3\sqrt{np} \right) \geq \frac{1}{3}.$$

This theorem implies that in Theorem 2.1, $\sqrt{np}\log^{k-2}(n)$ cannot be replaced by other terms with order $o(\sqrt{np})$ at least when all the entries of A are one. Hence, the upper bound is tight when $k = 2$ and tight up to a logarithm factor when $k > 2$. More generally, even if we consider all functions $\hat{\theta} : \mathbb{R}^{n^k} \to \mathbb{R}^{n^k}$, $T \mapsto \hat{\theta}(T)$, $\|\hat{\theta}(T) - \mathbb{E}T\|$ has no high probability bound tighter than $O(\sqrt{np})$. Therefore it is stronger than Theorem 2.2.

2.3. Regularization

Regularization of random graphs was first studied in [25]. It was proved in [25] that by removing high-degree vertices from a random graph, one can recover the concentration under the spectral norm because the $O(\sqrt{np})$ concentration breaks down when $np = O(1)$, due to the appearance of high-degree vertices, see [38, 40, 8]. A data-driven threshold for finding high degree vertices for the stochastic block model can be found in [62]. A different regularization analysis was given in [40] by decomposing the adjacency matrix into several parts and modify a small submatrix. This method was later generalized to other random matrices in [52, 51]. In this section we consider the regularization for Bernoulli random tensors.

Let $d_{i_1,\ldots,i_{k-1}} := \sum_{i_k \in [n]} t_{i_1,i_2,\ldots,i_k}$ be the degree of the tuple $(i_1,\ldots,i_{k-1})$. When $p = \frac{c \log n}{n}$, the maximal degree of all possible $(i_1,\ldots,i_{k-1})$ tuples is $O(np)$, and the bounded degree property of the random hypergraph holds (see Lemma 4.3). When $p = o\left(\frac{\log n}{n}\right)$, the maximal degree of all possible $(k-1)$-tuples is no longer $O(np)$, and our proof techniques of Theorem 2.1 will fail in this regime without the bounded degree property. In fact, when $p = \frac{c}{n}$ and

$\mathbb{E}T = pJ$, define a matrix $A \in \mathbb{R}^{n\times n}$ such that $a_{i_1,i_2} = t_{i_1,i_2,1,\ldots,1} - p$. According to [25], we have $\|T - \mathbb{E}T\| \geq \|A - \mathbb{E}A\| = \omega(\sqrt{np})$. We can see that $\|T - \mathbb{E}T\|$ can not be bounded by $O(\sqrt{np})$ with high probability in this regime, and the high $(k-1)$-order degrees destroy the concentration. In general, define

$$d_{i_1,\ldots,i_{k-m}} := \sum_{i_{k-m+1},\ldots,i_k\in[n]} t_{i_1,i_2,\ldots,i_k} \tag{2.5}$$

to be the degree of the tuple $(i_1,\ldots,i_{k-m})$, when $p = o\left(\frac{\log n}{n^m}\right)$ with $1 \leq m < k/2$, the degrees of all $(k-m)$-tuples fail to concentrate and our proof for (2.4) will not work in this regime. We conjecture that removing the $(k-m)$-tuples could be a possible way to recover the concentration of the spectral norm, but the proof of this conjecture is beyond the scope of this paper. One obstacle is that too many tuples are removed, which creates a large error in $\|\mathbb{E}T - \mathbb{E}\hat{T}\|$, where $\hat{T}$ is the tensor after removing tuples of high degrees. Another obstacle is the probability estimate. similar to [25, 40], we need to take a union bound to estimate failure probability over $2^{n^{k-m}}$ many possible sets of removed tuples. With the union bound argument, we fail to bound the spectral norm with high probability.

When $p \geq \frac{c\log n}{n^m}$ for $k/2 \leq m \leq k-1$, we can still analyze the regularization procedure by using tensor unfolding. In the proof of Theorem 2.3, we have obtained $\|T - \mathbb{E}T\| = O(\sqrt{n^m p})$ with high probability. The main proof idea is that the spectral norm of $(T - \mathbb{E}T)$ can be bounded by the spectral norm of an $n^m \times n^m$ matrix representation of $(T - \mathbb{E}T)$, which we denote it M for now. When $p = o\left(\frac{\log n}{n^m}\right)$, this tensor unfolding argument would fail because the random matrix M does not have spectral norm $O(\sqrt{n^m p})$. If we regularize this matrix M by removing high degree vertices, it will still provide an upper bound on the spectral norm for $T - \mathbb{E}T$. This is a sufficient way to obtain a spectral norm bound when $p \geq \frac{c}{n^m}$. In terms of the tensor structure, before regularization, the $(k-m)$-order degrees fail to concentrate. But after regularization, each degree $\hat{d}_{i_1,\ldots,i_{k-m}}$ in the regularized tensor $\hat{T}$ is bounded by $2\sqrt{n^m p}$, which guarantees that the unfolded tensor is concentrated under the spectral norm.

We adapt the techniques from [25, 40], together with the tensor unfolding operation, and apply it to an inhomogeneous random directed hypergraph, whose adjacency tensor has independent entries. This allows us to generalize the concentration inequalities in [25, 40] for regularized inhomogeneous random directed graphs with the same probability estimate. Based on different ranges of sparsity, our regularization procedures are slightly different, which depend on the boundedness property for different orders of degrees.

We now introduce some definitions for hypergraphs before stating the regularization procedure.

Definition 2.1 (Hypergraph)**.** *A hypergraph H consists of a set V of vertices and a set E of hyperedges such that each hyperedge is a nonempty set of V. H is k-uniform if every hyperedge $e \in E$ contains exactly k vertices. The degree of a vertex i is the number of all hyperedges incident to i.*

Let us index the vertices by $V = \{1, \dots, n\}$. A k-uniform hypergraph can be represented by order-k tensor with dimension $n \times \cdots \times n$.

Definition 2.2 (Adjacency tensor)**.** *Given a k-uniform hypergraph H, an order-k tensor T is the adjacency tensor of $H = (V, E)$ if*

$$t_{i_1,\dots,i_k} = \begin{cases} 1, & \text{if } \{i_1, \dots, i_k\} \in E, \\ 0, & \text{otherwise.} \end{cases}$$

For any adjacency tensor T, $t_{i_{\sigma(1)},\dots,i_{\sigma(k)}} = t_{i_1,\dots,i_k}$ for any permutation σ. We may abuse notation and write t_e in place of $t_{i_1,\dots,i_k}$, where $e = \{i_1, \dots, i_k\}$. A k-uniform directed hypergraph $H = (V, E)$ consists of a set V of vertices and a set E of directed hyperedges such that each directed hyperedge is an element in $V \times \cdots \times V = V^k$. Define T to be the adjacency tensor of H such that

$$t_{i_1,\dots,i_k} = \begin{cases} 1, & \text{if } (i_1, \dots, i_k) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

For different orders of sparsity in terms of m, our regularization procedure is based on the boundedness of $(k-m)$-th order degrees in the random directed hypergraph. Assume $p \geq \frac{c}{n^m}$ with $k/2 \leq m \leq k-1$. For any order-k tensor A indexed by $[n]$, let $S \subset [n]^{k-m}$ be a subset of indices. We define the regularized tensor A^S as

$$a^S_{i_1,\dots,i_k} = \begin{cases} 0 & \text{if } (i_1, \dots, i_{k-m}) \in S, \\ a_{i_1,\dots,i_k} & \text{otherwise.} \end{cases}$$

When we observe a random tensor T, we regularize T as follows. Suppose the degree of a $(k-m)$-tuple $(i_1, \dots, i_{k-m})$ is greater than $2n^m p$, then we remove all directed hyperedges containing this tuple. In other words, we zero out the corresponding hyperedges in the adjacency tensor. We call the resulting tensor $\hat{T}$. Let $\tilde{S} \subset [n]^{k-m}$ be the set of $(k-m)$-tuples with degree greater than $2n^m p$. Then with our notation,

$$\hat{T} = T^{\tilde{S}}. \tag{2.6}$$

Since from our Theorem 2.3, when $p \geq \frac{c \log n}{n^m}$ for some $c > 0$, the regularization is not needed, below we assume $p < \frac{\log n}{n^m}$ for simplicity. The following lemma shows that with high probability, not many $(k-m)$-tuples are removed.

Lemma 2.1. *Let $\frac{c}{n^m} \leq p < \frac{\log n}{n^m}$ for a sufficiently large $c > 1$ and an integer m with $k/2 \leq m \leq k-1$. Then the number of regularized $(k-m)$-tuples $|\tilde{S}|$ is at most $\frac{1}{n^{2m-k}p}$ with probability at least $1 - \exp\left(-\frac{n^{k-m}}{6 \log n}\right)$.*

After regularization, the following holds.

Theorem 2.5. *Let* $\frac{c}{n^m} \le p < \frac{\log n}{n^m}$ *for a sufficiently large* $c > 1$ *and an integer* m *with* $k/2 \le m \le k-1$. *Let* $\hat{T}$ *be the random order-*k *tensor* T *after regularization. Then for any* $r > 0$, *there exists a constant* C_2 *depending on* k, r *such that*

$$\mathbb{P}\left(\|\hat{T} - \mathbb{E}T\| \le C_2\sqrt{n^m p}\right) \ge 1 - n^{-r}. \tag{2.7}$$

3. Applications

To demonstrate the usefulness of our concentration and regularization results, we highlight two applications.

3.1. Sparse hypergraph expanders

The expander mixing lemma for a d-regular graph (the degree of each vertex is d) states the following: Let G be a d-regular graph on n vertices with the second largest eigenvalue in absolute value of its adjacency matrix satisfying $\lambda := \max\{\lambda_2, |\lambda_n|\} < d$. For any two subsets $V_1, V_2 \subseteq V(G)$, let

$$e(V_1, V_2) = |\{(v_1, v_2) \in V_1 \times V_2 : \{v_1, v_2\} \in E(G)\}|$$

be the number of edges between V_1 and V_2. Then

$$\left|e(V_1, V_2) - \frac{d|V_1||V_2|}{n}\right| \le \lambda\sqrt{|V_1||V_2|\left(1 - \frac{|V_1|}{n}\right)\left(1 - \frac{|V_2|}{n}\right)}. \tag{3.1}$$

(3.1) shows that d-regular graphs with small λ have a good mixing property, where the number of edges between any two vertex subsets is approximated by the number of edges we would expect if they were drawn at random. Such graphs are called expanders, and the quality of such an approximation is controlled by λ, which is also the mixing rate of simple random walks on G [18].

Hypergraph expanders have recently received significant attention in combinatorics and theoretical computer science [47, 22]. Many different definitions have been proposed for hypergraph expanders, each with their own strengths and weaknesses. In this section, we only consider hypergraph models that have a generalized version of expander mixing lemma (3.1).

There are several hypergraph expander mixing lemmas in the literature based on the spectral norm of tensors [27, 43, 19]. However, for deterministic tensors, the spectral norm is NP-hard to compute [33], hence those estimates might not be applicable in practice. In [11, 31], the authors obtained a weaker expander version mixing lemma for a sparse deterministic hypergraph model where the mixing property depends on the second eigenvalue of a regular graph. Friedman and Widgerson [27] obtained the following spectral norm bound for a random hypergraph model: Consider a k-uniform hypergraph model on n vertices where $n^k p$ hyperedges are chosen independently at random. Let J be the order-k tensor with all entries taking value 1. If $p \ge Ck\log n/n$, then with high probability

$\|T - pJ\| \le (C \log n)^{k/2}\sqrt{np}$. Combining with their expander mixing lemma in [27], it provides a random hypergraph model with a good control of the mixing property. This is a random hypergraph model with expected degrees $n^{k-1}p$, which is not bounded. To the best of our knowledge, our Theorem 3.1 below is the first construction of a sparse random hypergraph model with bounded degrees that satisfies a k-subset expander mixing lemma with high probability. The idea of applying expander mixing lemma and spectral gap results of sparse expanders to analyze matrix completion and tensor completion has been developed in [32, 10, 14, 28, 31]. We believe our result could also be useful for tensor completion or other related problems.

Let H be a k-uniform Erdős-Rényi hypergraph (recall Definition 2.1) on n vertices with sparsity $p = \frac{c}{n^{k-1}}$, where each hyperedge is generated independently with probability p. Its adjacency tensor is then a symmetric tensor, denoted by T. We construct a regularized hypergraph H' as follows:

1. Construct $\tilde{T}$ such that
$$\tilde{t}_{i_1,\dots,i_k} = \begin{cases} t_{i_1,\dots,i_k} & \text{if } 1 \le i_1 < i_2 < \cdots < i_k \le n, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$
2. Compute $\tilde{d}_i := \sum_{i_1,\dots,i_{k-1}\in[n]} \tilde{t}_{i,i_1,\dots,i_{k-1}}$ for all $i \in [n]$. If $\tilde{d}_i > 2n^{k-1}p$, zero out all entries $\tilde{t}_{i,i_1,\dots,i_{k-1}}, i_1, \dots, i_{k-1} \in [n]$. We then obtain a new tensor $\hat{T}$.
3. Define T' such that $(t')_{i_1,i_2,\dots,i_k} = \sum_{\sigma\in\mathfrak{S}_k} \hat{t}_{i_{\sigma(1)},\dots,i_{\sigma(k)}}$, where $\mathfrak{S}_k$ is the symmetric group of order k. We then obtain a regularized hypergraph H' with adjacency tensor T'.

Note that this regularization procedure is applicable to inhomogeneous random hypergraphs by taking $p = \max_{i_1,\dots,i_k\in[n]} p_{i_1,\dots,i_k}$. By our construction, H' is a k-uniform hypergraph with degrees at most $2k!n^{k-1}p = 2k!c$. Let $J \in \mathbb{R}^{n^k}$ be an order k tensor with all entries taking value 1. From Theorem 2.5, for some constant $C > 0$, with high probability its adjacency tensor T' satisfies

$$\|T' - pJ\| \le C\sqrt{n^{k-1}p}. \tag{3.3}$$

In the next theorem we use (3.3) to show that H' satisfies an expander mixing lemma with high probability.

Definition 3.1. *If $V_1, \dots, V_k$ are subsets of $V(H)$ for a k-uniform hypergraph H, define*

$$e_H(V_1, \dots V_k) := |\{(v_1, \dots, v_k) \in V_1 \times \cdots \times V_k : \{v_1, \dots, v_k\} \in E(H)\}| \tag{3.4}$$

to be the number of hyperedges between $V_1, \dots, V_k$.

Theorem 3.1. *Let H be a k-uniform Erdős-Rényi hypergraph on n vertices with sparsity $p = \frac{c}{n^{k-1}}$ for some sufficiently large constant $c > 1$. Let H' be the hypergraph H after regularization, then there exists a constant $C > 0$ depending*

on k such that with high probability for any non-empty subsets $V_1, \dots, V_k \subset [n]$, we have the following expander mixing inequality:

$$|e_{H'}(V_1, \dots, V_k) - p|V_1| \cdots |V_k|| \le C\sqrt{c} \cdot \sqrt{|V_1| \cdots |V_k|}. \tag{3.5}$$

3.2. Tensor sparsification

In the *tensor completion* problem, one aims to estimate a low-rank tensor based on a random sample of observed entries. A commonly used definition of the rank for tensors is called canonical polyadic (CP) rank. We refer to [37] for more details. In order to solve a tensor completion problems, there are two main steps. First, one needs to sample some entries from a low-rank tensor T. Then, based on the observed entries, one solves an optimization problem and justifies that the solutions to this problem will be exactly or nearly the original tensor T. A fundamental question is: how many observations are required to guarantee that the solution of the optimization problem provides a good recovery of the original tensor T?

After a random sampling from the original tensor T, we obtain a random tensor $\tilde{T}$. If we require the sample size to be small, $\tilde{T}$ then will be random and sparse. In the next step, the optimization procedure is then based on $\tilde{T}$. In the matrix or tensor completion algorithm, especially for the non-convex optimization algorithm, we need some stability guarantee on the initial data, see for example [35, 34, 16]. Therefore, it is important to have concentration guarantee such that $\tilde{T}$ is close to T under the spectral norm.

Another related problem is called *tensor sparsification*. Given a tensor T, through some sampling algorithm, one wants to construct a sparsified version $\tilde{T}$ of T such that $\|T - \tilde{T}\|$ is relatively small with high probability. In [49], a non-uniform sampling algorithm was proposed and the probability of sampling each entry is chosen based on the magnitude of the entry in T. However, without knowing the exact value of the original tensor T, a reasonable way to output a sparsified tensor $\tilde{T}$ is through uniform sampling.

We obtain a concentration inequality of the spectral norm for tensors under uniform sampling, which is useful to both of the problems above. It improves the sparsity assumption in the analysis of the initialization step for the tensor completion algorithm proposed in [34]. Let T be a deterministic tensor. We obtain a new tensor $\tilde{T}$ by uniformly sampling entries in T with probability p. Namely,

$$\tilde{t}_{i_1,\dots,i_k} = \begin{cases} t_{i_1,\dots,i_k} & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

By our definition, $\mathbb{E}\tilde{T} = pT$. The following is an estimate about the concentration of $\tilde{T}$ under the spectral norm when $p \ge \frac{c \log n}{n^m}$, $1 \le m \le k-1$. It is a quick corollary from Theorem 2.1 and Theorem 2.3.

Corollary 3.1. *Let $p \geq \frac{c\log n}{n^m}$ for some constant $c > 0$ and some integer $1 \leq m \leq k-1$. Denote $t_{\max} := \max_{i_1,\ldots,i_k\in[n]} |t_{i_1,\ldots,i_k}|$. For any $r > 0$, there exists a constant $C > 0$ depending on r, k, c such that with probability $1 - n^{-r}$,*

$$\|\tilde{T} - pT\| \leq \begin{cases} Ct_{\max}\sqrt{n^m p} & \text{if} \quad k/2 \leq m \leq k-1, \\ Ct_{\max}\sqrt{n^m p}\log^{\frac{k-1}{m}-1}(n) & \text{if} \quad 1 \leq m < k/2, \quad k/m \notin \mathbb{Z}, \\ Ct_{\max}\sqrt{n^m p}\log^{\frac{k}{m}-2}(n) & \text{if} \quad 1 \leq m < k/2, \quad k/m \in \mathbb{Z}. \end{cases}$$

Remark 3.1. *Theorem 2.1 in [34] provides an estimate of $\|\tilde{T} - pT\|$ for a symmetric tensor T with symmetric sampling, assuming $k = 3$ and $p \geq \frac{\log n}{n^{3/2}}$. When $k = 3$, our result allows the sparsity down to $p \geq \frac{c\log n}{n^2}$ and covers non-symmetric tensors with uniform sampling.*

4. Proof of Theorem 2.1

The proof is a generalization of [25, 42] and is suitable for sparse random tensors. This type of method is known as the Kahn-Szemerédi argument originally introduced in [26]. Without loss of generality, we may assume

$$\max_{i_1,\ldots,i_k} |A_{i_1,\ldots,i_k}| = 1 \tag{4.1}$$

in our proof for simplicity.

4.1. Discretization

Fix $\delta \in (0,1)$, define the n-dimensional ball of radius t as

$$S_t := \{v \in \mathbb{R}^n : \|v\|_2 \leq t\}.$$

We introduce a set of lattice points in S_1 as follows:

$$\mathcal{T} = \left\{x = (x_1, \ldots x_n) \in S_1 : \frac{\sqrt{n}x_i}{\delta} \in \mathbb{Z}, \forall i \in [n]\right\}. \tag{4.2}$$

By the Lipschitz property of spectral norms, we have the following upper bound, which reduces the problem of bounding the spectral norm of T to an optimization problem over $\mathcal{T}$.

Lemma 4.1. *For any tensor $T \in \mathbb{R}^{n^k}$ and any fixed $\delta \in (0,1)$, we have*

$$\|T\| \leq (1-\delta)^{-k} \sup_{y_1,\ldots,y_k\in\mathcal{T}} |T(y_1, \ldots, y_k)|.$$

Proof. The proof follows from Lemma 2.1 in the supplement of [42]. For completeness, we provide the proof here. For any $v \in S_{1-\delta}$, consider the cube in $\mathbb{R}^n$ of edge length $\delta/\sqrt{n}$ that contains v, with all its vertices in $\left(\frac{\delta}{\sqrt{n}}\mathbb{Z}\right)^n$. The diameter of the cube is δ, so the entire cube is contained in S_1. Hence all vertices of this

cube are in $\mathcal{T}$ and $S_{1-\delta} \subset \text{convhull}(\mathcal{T})$. Therefore for each $u_i \in S_1, 1 \le i \le k$, we can find some sequence $\{x_{i_j}\}_{j=1}^{N_i} \subset \mathcal{T}$ such that $(1-\delta)u_i$ is a linear combination of those $\{x_{i_j}\}$, namely,

$$(1-\delta)u_i = \sum_{j=1}^{N_i} a_j^{(i)} x_{i_j},$$

for some $a_j^{(i)} \in [0,1]$ satisfying $\sum_{j=1}^{N_i} a_j^{(i)} = 1$. Then

$$\begin{aligned}
|T(u_1,\dots,u_k)| &= (1-\delta)^{-k}|T((1-\delta)u_1,\dots,(1-\delta)u_k)| \\
&\le (1-\delta)^{-k} \sum_{j_1=1}^{N_1} \cdots \sum_{j_k=1}^{N_k} a_{j_1}^{(1)} \cdots a_{j_k}^{(k)} |T(x_{1_{j_1}},\dots,x_{k_{j_k}})| \\
&\le (1-\delta)^{-k} \sup_{y_1,\dots,y_k \in \mathcal{T}} |T(y_1,\dots,y_k)|,
\end{aligned}$$

where the last inequality is due to

$$\sum_{j_1=1}^{N_1} \cdots \sum_{j_k=1}^{N_k} a_{j_1}^{(1)} \cdots a_{j_k}^{(k)} = \prod_{i=1}^{k} \left(\sum_{j_i=1}^{N_i} a_{j_i}^{(i)} \right) = 1.$$

This completes the proof. □

Now for any fixed k-tuples $(y_1,\dots,y_k) \in \mathcal{T} \times \cdots \times \mathcal{T}$, we decompose its index set. Define the set of *light tuples* as

$$\mathcal{L} = \mathcal{L}(y_1,\dots,y_k) := \left\{ (i_1,\dots,i_k) \in [n]^k : |y_{1,i_1} \cdots y_{k,i_k}| \le \frac{\sqrt{np}}{n} \right\}, \tag{4.3}$$

and *heavy tuples* as

$$\overline{\mathcal{L}} = \overline{\mathcal{L}}(y_1,\dots,y_k) := \left\{ (i_1,\dots,i_k) \in [n]^k : |y_{1,i_1} \cdots y_{k,i_k}| > \frac{\sqrt{np}}{n} \right\}. \tag{4.4}$$

In the remaining part of our proof, we control the contributions of light and heavy tuples to the spectral norm respectively.

4.2. Light tuples

Let $W := A \circ T - \mathbb{E}[A \circ T]$ be the centered random tensor and we denote the entries of W by $w_{i_1,\dots,i_k}$ for $i_1,\dots,i_k \in [n]$. We have the following concentration bound for the contribution of light tuples.

Lemma 4.2. *For any constant $c > 0$,*

$$\mathbb{P}\left(\sup_{y_1,\dots,y_k \in \mathcal{T}} \left| \sum_{(i_1,\dots,i_k) \in \mathcal{L}} y_{1,i_1} \cdots y_{k,i_k} w_{i_1,\dots,i_k} \right| \ge c\sqrt{np} \right)$$

$$\leq 2\exp\left[-n\left(\frac{c^2}{2(1+c/3)}-k\log\left(\frac{7}{\delta}\right)\right)\right],$$

where $\delta\in(0,1)$ *on the right hand side is determined by the definition of* $\mathcal{T}$ *in* (4.2).

Proof. Denote

$$u_{i_1,\ldots,i_k}:=y_{1,i_1}\cdots y_{k,i_k}\mathbf{1}\{|y_{1,i_1}\cdots y_{k,i_k}|\leq\sqrt{np}/n\}. \tag{4.5}$$

Then the contribution from light tuples can be written as

$$\sum_{i_1,\ldots,i_k\in[n]} w_{i_1,\ldots,i_k}u_{i_1,\ldots,i_k}.$$

Since from (4.1), each term in the sum has mean zero and is bounded by $\sqrt{np}/n$, we are ready to apply Bernstein's inequality to get for any constant $c>0$,

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{i_1,\ldots,i_k\in[n]} w_{i_1,\ldots,i_k}u_{i_1,\ldots,i_k}\right|\geq c\sqrt{np}\right) &(4.6)\\ \leq&2\exp\left(-\frac{c^2np/2}{\sum_{i_1,\ldots,i_k\in[n]}p_{i_1,\ldots,i_k}(1-p_{i_1,\ldots,i_k})u^2_{i_1,\ldots,i_k}+\frac{1}{3}\frac{\sqrt{np}}{n}c\sqrt{np}}\right)\\ \leq&2\exp\left(-\frac{c^2np/2}{p\sum_{i_1,\ldots,i_k\in[n]}u^2_{i_1,\ldots,i_k}+\frac{cp}{3}}\right). &(4.7)\end{aligned}$$

From (4.5) we have

$$\sum_{i_1,\ldots,i_k\in[n]}u^2_{i_1,\ldots,i_k}\leq\sum_{i_1,\ldots,i_k\in[n]}y^2_{1,i_1}\cdots y^2_{k,i_k}=\prod_{j=1}^k\|y_j\|_2^2=1.$$

Then (4.7) is bounded by $2\exp\left(\frac{-c^2n}{2+\frac{2c}{3}}\right)$. By the volume argument (see for example [58]) we have $|\mathcal{T}|\leq\exp(n\log(7/\delta))$, hence the k-th product of $\mathcal{T}$ satisfies

$$|\mathcal{T}\times\cdots\times\mathcal{T}|\leq\exp(kn\log(7/\delta)).$$

Then taking a union bound over all possible $y_1,\ldots,y_k\in\mathcal{T}$, we have

$$\sup_{y_1,\ldots,y_k\in\mathcal{T}}\left|\sum_{(i_1,\ldots,i_k)\in\mathcal{L}}y_{1,i_1}\cdots y_{k,i_k}w_{i_1,\ldots,i_k}\right|\leq c\sqrt{np}$$

with probability at least $1-2\exp\left[-\frac{c^2n}{2(1+c/3)}+kn\log(7/\delta)\right]$. This completes the proof. □

By Lemma 4.2, for any $r > 0$, we can take the constant c in Lemma 4.2 large enough depending on k and r (for example, taking $c = 6r + 6k\log(7/\delta)$ suffices) such that with probability at least $1 - n^{-r}$,

$$\sup_{y_1,\dots,y_k\in\mathcal{T}}\left|\sum_{i_1,\dots,i_k\in\mathcal{L}} y_{1,i_1}\cdots y_{k,i_k}w_{i_1,\dots,i_k}\right| \le c\sqrt{np}.$$

Therefore to prove Theorem 2.3, it remains to control the contribution from heavy tuples.

4.3. Heavy tuples

Next, we show the contribution from heavy tuples is bounded by $c\sqrt{np}\log^{k-2}(n)$ for some constant $c > 0$ depending on k with high enough probability. Namely,

$$\sup_{y_1,\dots,y_k\in\mathcal{T}}\left|\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}} y_{1,i_1}\cdots y_{k,i_k}\cdot w_{i_1,\dots,i_k}\right| \le c\sqrt{np}\log^{k-2}(n).$$

Note that from our definition of heavy tuples in (4.4), we have

$$\begin{aligned}&\left|\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}} y_{1,i_1}\cdots y_{k,i_k}\cdot(a_{i_1,\dots,i_t}p_{i_1,\dots,i_k})\right|\\ \le&\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}}\frac{y_{1,i_1}^2\cdots y_{k,i_k}^2}{|y_{1,i_1}\cdots y_{k,i_k}|}\cdot p_{i_1,\dots,i_k} \le \sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}}\frac{n}{\sqrt{np}}y_{1,i_1}^2\cdots y_{k,i_k}^2\cdot p\\ \le&\sqrt{np}\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}} y_{1,i_1}^2\cdots y_{k,i_k}^2 \le \sqrt{np}.\end{aligned} \tag{4.8}$$

Then it suffices to show that with high enough probability for all $y_1,\dots,y_k\in\mathcal{T}$,

$$\left|\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}} y_{1,i_1}\cdots y_{k,i_k}\cdot(a_{i_1,\dots,i_k}t_{i_1,\dots,i_k})\right| \le C_k\sqrt{np}\log^{k-2}(n) \tag{4.9}$$

for a constant C_k depending on k. We will focus on the heavy tuples $(i_1,\dots,i_k)$ such that $y_{1,i_1},\dots,y_{k,i_k} > 0$. We denote this set by $\overline{\mathcal{L}}^+$. The remaining cases will be similar and there are 2^k different cases in total. Note that

$$\begin{aligned}&\left|\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}^+} y_{1,i_1}\cdots y_{k,i_k}\cdot(a_{i_1,\dots,i_k}t_{i_1,\dots,i_k})\right|\\ \le&\sum_{(i_1,\dots,i_k)\in\overline{\mathcal{L}}^+} y_{1,i_1}\cdots y_{k,i_k}\cdot t_{i_1,\dots,i_k}.\end{aligned} \tag{4.10}$$

For the rest of the proof we will bound the right hand side of (4.10). We now define the following index sets for a fixed tuple $(y_1, \dots, y_k) \in \mathcal{T} \times \cdots \times \mathcal{T}$:

$$D_j^s = \left\{ i : \frac{2^{s-1}\delta}{\sqrt{n}} \le y_{j,i} \le \frac{2^s \delta}{\sqrt{n}} \right\} \text{ for } s = 1, \dots, \left\lceil \log_2(\sqrt{n}/\delta) \right\rceil \text{ and } 1 \le j \le k. \tag{4.11}$$

By our definition of $\mathcal{T}$ in (4.2), for any $(y_1, \dots, y_k) \in (\mathcal{T} \times \cdots \times \mathcal{T})$ and $(i_1, \dots, i_k) \in \overline{\mathcal{L}}^+$, we have $y_{j,i_j} \ge \delta/\sqrt{n}$ for all $1 \le j \le k$. Therefore each i_j is in D_j^s for some s. Recall the definition of degree for a $(k-1)$-tuple in (2.5). Also the following definitions are needed:

1. $$e(I_1, \dots, I_k) := \left| \{ (i_1, \dots, i_k) : t_{i_1, \dots, i_k} = 1, i_1 \in I_1, \dots, i_k \in I_k \} \right|.$$
2. $$\mu(I_1, \dots, I_k) = \mathbb{E} e(I_1, \dots, I_k), \quad \overline{\mu}(I_1, \dots, I_k) = p|I_1| \cdots |I_k|.$$
3. For $1 \le s_1, \dots, s_k \le \log_2(\sqrt{n}/\delta)$,
$$\lambda_{s_1, \dots, s_k} = \frac{e(D_1^{s_1}, \dots, D_k^{s_k})}{\overline{\mu}_{s_1, \dots, s_k}}, \quad \overline{\mu}_{s_1, \dots, s_k} = \overline{\mu}(D_1^{s_1}, \dots, D_k^{s_k}).$$
4. $\alpha_{j,s} = |D_j^s| \cdot 2^{2s}/n, 1 \le j \le k, 1 \le s \le \log_2(\sqrt{n}/\delta)$.
5. $\sigma_{s_1, \dots, s_k} = \lambda_{s_1, \dots, s_k} n^{k/2-1} \sqrt{np} \cdot 2^{-(s_1 + \cdots + s_k)}, 1 \le s_1, \dots, s_k \le \log_2(\sqrt{n}/\delta)$.

The following two lemmas are about the properties of the sparse directed random hypergraphs (recall Definition 2.2), which are important for the rest of our proof.

Lemma 4.3 (Bounded degree)**.** *Assume $p \ge c \log n/n$ for some constant $c > 0$. Then for any $r > 0$, there exists a constant $c_1 > 1$ depending on c, r, k such that with probability at least $1 - n^{-r}$, for all $i_1, \dots, i_{k-1} \in [n]$, $d_{i_1, \dots, i_{k-1}} \le c_1 np$.*

Proof. For a fixed $(k-1)$-tuple $(i_1, \dots, i_{k-1})$, by Bernstein's inequality,

$$\begin{aligned} \mathbb{P}(d_{i_1, \dots, i_{k-1}} \ge c_1 np) &= \mathbb{P}\left(\sum_{i_k \in [n]} t_{i_1, \dots, i_k} \ge c_1 np \right) \\ &\le \mathbb{P}\left(\sum_{i_k \in [n]} (t_{i_1, \dots, i_k} - p_{i_1, \dots, i_k}) \ge (c_1 - 1) np \right) \\ &\le \exp\left[-\frac{\frac{1}{2}(c_1 - 1)^2 n^2 p^2}{np + \frac{1}{3}(c_1 - 1) np} \right] \le n^{-\frac{3(c_1-1)^2 c}{4 + 2c_1}}, \end{aligned} \tag{4.12}$$

where in the last inequality we use the assumption $p \ge c \log n/n$. Then taking a union bound over $i_1, \dots, i_{k-1} \in [n]$ implies

$$\mathbb{P}\left(\sup_{i_1, \dots, i_{k-1} \in [n]} d_{i_1, \dots, i_{k-1}} \ge c_1 np \right) \le n^{-\frac{3(c_1-1)^2 c}{4+2c_1} + k - 1}.$$

Therefore for any $r, c > 0$ we can take c_1 sufficiently large (depending linearly on k, r) to make Lemma (4.3) hold. □

Lemma 4.4 (Bounded discrepancy)**.** *Assume $p \geq c \log n/n$ for some constant $c > 0$. For any $r > 0$, there exist constants $c_2, c_3 > 1$ depending on c, r, k such that with probability at least $1 - 2n^{-r}$, for any nonempty sets $I_1, \dots, I_k \subset [n]$ with $1 \leq |I_1| \leq \cdots \leq |I_k|$, at least one of the following events hold:*

1. $\frac{e(I_1,\dots,I_k)}{\bar{\mu}(I_1,\dots,I_k)} \leq ec_2$,
2. $e(I_1, \dots, I_k) \log\left(\frac{e(I_1,\dots,I_k)}{\bar{\mu}(I_1,\dots,I_k)}\right) \leq c_3 |I_k| \log\left(\frac{n}{|I_k|}\right)$.

We will use the following Chernoff's inequality (see [13]).

Lemma 4.5 (Chernoff bound)**.** *Let $X_1, \dots, X_n$ be independent Bernoulli random variables. Let $X = \sum_{i=1}^n X_n$ and $\mu = \mathbb{E}X$. Then for any $\delta > 0$,*

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp(-\mu((1+\delta)\ln(1+\delta) - \delta)). \tag{4.13}$$

In particular, we have a weaker version of (4.13): for any $\delta > 0$,

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp\left(\frac{-\delta^2\mu}{2+\delta}\right). \tag{4.14}$$

Proof of Lemma 4.4. In this proof, we assume the event in Lemma 4.3 that all degrees of vertices are bounded by $c_1 np$ holds. If $|I_k| \geq n/e$, then the bounded degree property implies $e(I_1, \dots, I_k) \leq |I_1| \cdots |I_{k-1}| c_1 np$. Hence

$$\frac{e(I_1, \dots, I_k)}{\bar{\mu}(I_1, \dots, I_k)} = \frac{e(I_1, \dots, I_k)}{p|I_1| \cdots |I_k|} \leq \frac{|I_1| \cdots |I_{k-1}| c_1 np}{p|I_1| \cdots |I_{k-1}| n/e} \leq c_1 e.$$

This completes the proof for Case (1). Now we consider the case when $|I_k| < n/e$. Let $s(I_1, \dots, I_k)$ be the set of all possible distinct hyperedges between $I_1, \dots, I_k$. We have for any $\tau > 1$ and any fixed $I_1, \dots, I_k$,

$$\begin{aligned}
&\mathbb{P}(e(I_1, \dots, I_k) \geq \tau \bar{\mu}(I_1, \dots, I_k)) \\
=&\mathbb{P}\left(\sum_{(i_1,\dots,i_k)\in s(I_1,\dots,I_k)} t_{i_1,\dots,i_k} \geq \tau \bar{\mu}(I_1, \dots, I_k)\right) \\
\leq&\mathbb{P}\left(\sum_{(i_1,\dots,i_k)\in s(I_1,\dots,I_k)} (t_{i_1,\dots,i_k} - p_{i_1,\dots,i_k}) \geq (\tau-1)\bar{\mu}(I_1, \dots, I_k)\right).
\end{aligned}$$

By Chernoff's inequality (4.13), the last line above is bounded by

$$\begin{aligned}
&\exp\left((\tau-1)\overline{\mu}(I_1, \dots, I_k) - \tau\overline{\mu}(I_1, \dots, I_k)\log\tau\right) \\
\leq &\exp\left(-\frac{1}{2}\bar{\mu}(I_1, \dots, I_k)\tau\log\tau\right),
\end{aligned}$$

where the last inequality holds when $\tau \geq 8$. This implies for $\tau \geq 8$,

$$\mathbb{P}(e(I_1,\dots,I_k) \geq \tau\bar{\mu}(I_1,\dots,I_k)) \leq \exp\left(-\frac{1}{2}\bar{\mu}(I_1,\dots,I_k)\tau\log\tau\right). \tag{4.15}$$

For a given number $c_3 > 0$, define $\gamma(I_1,\dots,I_k)$ to be the unique value of γ such that

$$\gamma\log\gamma = \frac{c_3|I_k|}{\bar{\mu}(I_1,\dots,I_k)}\log\left(\frac{n}{|I_k|}\right). \tag{4.16}$$

Let $\kappa(I_1,\dots,I_k) = \max\{8,\gamma(I_1,\dots,I_k)\}$. Then by (4.15) and (4.16),

$$\begin{aligned}
&\mathbb{P}(e(I_1,\dots,I_k) \geq \gamma(I_1,\dots,I_k)\bar{\mu}(I_1,\dots,I_k))\\
&\leq \exp\left(-\frac{1}{2}\bar{\mu}(I_1,\dots,I_k)\kappa(I_1,\dots,I_k)\log\kappa(I_1,\dots,I_k)\right)\\
&\leq \exp\left[-\frac{1}{2}c_3|I_k|\log\left(\frac{n}{|I_k|}\right)\right].
\end{aligned} \tag{4.17}$$

Let $\Omega = \{(I_1,\dots,I_k) : |I_1| \leq \cdots \leq |I_k| \leq \frac{n}{e}\}$ and

$$S(h_1,\dots,h_k) := \{(I_1,\dots,I_k) : \forall i \in [k], |I_i| = h_i\}.$$

By a union bound and (4.17), we have

$$\begin{aligned}
&\mathbb{P}\Big(\exists(I_1,\dots,I_k)\in\Omega : e(I_1,\dots,I_k) \geq \gamma(I_1,\dots,I_k)\bar{\mu}(I_1,\dots,I_k)\Big)\\
&\leq \sum_{(I_1,\dots,I_k)\in\Omega} \exp\left[-\frac{1}{2}c_3|I_k|\log\left(\frac{n}{|I_k|}\right)\right]\\
&= \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \sum_{(I_1,\dots,I_k)\in S(h_1,\dots,h_k)} \exp\Big[-\frac{1}{2}c_3h_k\log\Big(\frac{n}{h_k}\Big)\Big]\\
&= \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \binom{n}{h_1}\dots\binom{n}{h_k}\exp\Big[-\frac{1}{2}c_3h_k\log\Big(\frac{n}{h_k}\Big)\Big].
\end{aligned}$$

Since $\binom{n}{k} \leq (\frac{ne}{k})^k$ for any integer $1 \leq k \leq n$, we have the last line above is bounded by

$$\begin{aligned}
&\sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \Big(\frac{ne}{h_1}\Big)^{h_1}\dots\Big(\frac{ne}{h_k}\Big)^{h_k}\exp\Big[-\frac{1}{2}c_3h_k\log\Big(\frac{n}{h_k}\Big)\Big]\\
&\leq \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \exp\Big[-\frac{1}{2}c_3h_k\log\Big(\frac{n}{h_k}\Big) + kh_k\log\Big(\frac{n}{h_k}\Big) + kh_k\Big]\\
&\leq \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \exp\Big[-\frac{1}{2}(c_3-4k)h_k\log\Big(\frac{n}{h_k}\Big)\Big]
\end{aligned}$$

$$\leq \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} \exp\left[-\frac{1}{2}(c_3-4k)\log n\right]$$
$$= \sum_{1\leq h_1\leq\cdots\leq h_k\leq\frac{n}{e}} n^{-\frac{1}{2}(c_3-4k)} \leq n^{-\frac{1}{2}(c_3-6k)}.$$

As a result, $e(I_1,\ldots,I_k) < \kappa(I_1,\ldots,I_k)\bar{\mu}(I_1,\ldots,I_k)$ for all $(I_1,\ldots,I_k)\in\Omega$ with probability at least $1-n^{-\frac{1}{2}(c_3-6k)}$. For any $r>0$, we can choose c_3 large enough depending linearly on k,r such that $1-n^{-\frac{1}{2}(c_3-6k)} \leq 1-n^{-r}$. Suppose $\kappa(I_1,\ldots,I_k)=8$, then $e(I_1,\ldots,I_k)<8\bar{\mu}(I_1,\ldots,I_k)$ as desired. Otherwise suppose $\kappa(I_1,\ldots,I_k)=\gamma(I_1,\ldots,I_k)>8$, then

$$\frac{e(I_1,\ldots,I_k)}{\bar{\mu}(I_1,\ldots,I_k)} < \gamma(I_1,\ldots,I_k).$$

Since $x\mapsto x\log x$ is an increasing function for $x\geq 1$, we have

$$\begin{aligned}\frac{e(I_1,\ldots,I_k)}{\bar{\mu}(I_1,\ldots,I_k)}\log\frac{e(I_1,\ldots,I_k)}{\bar{\mu}(I_1,\ldots,I_k)} <& \gamma(I_1,\ldots,I_k)\log\gamma(I_1,\ldots,I_k)\\ =& \frac{c_3|I_k|}{\bar{\mu}(I_1,\ldots,I_k)}\log\left(\frac{n}{|I_k|}\right),\end{aligned}$$

which gives the desired result for Case (2). □

With Lemma 4.3 and Lemma 4.4, we prove our estimates (4.9) for all heavy tuples. Recall we are dealing with the tuples over $\overline{\mathcal{L}}^+$, we then have

$$\begin{aligned}&\sum_{(i_1,\ldots,i_k)\in\overline{\mathcal{L}}^+} y_{1,i_1}\cdots y_{k,i_k}\cdot t_{i_1,\ldots,i_k}\\ \leq& \sum_{\substack{(s_1,\ldots,s_k):\\ 2^{s_1+\cdots+s_k}\geq\sqrt{np}(2/\delta)^k n^{k/2-1}}} e(D_1^{s_1},\ldots,D_k^{s_k})\frac{2^{s_1}\delta}{\sqrt{n}}\cdots\frac{2^{s_k}\delta}{\sqrt{n}}\\ \leq& \delta^k\sqrt{np}\sum_{\substack{(s_1,\ldots,s_k):\\ 2^{s_1+\cdots+s_k}\geq\sqrt{np}n^{k/2-1}}} \alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\ldots,s_k}. \qquad (4.18)\end{aligned}$$

The last equality follows directly from definitions in (5). (4.18) implies that it suffices to estimate the contribution of heavy tuples through its index sets. We then bound the contribution of heavy tuples by splitting the indices $(s_1,\ldots,s_k)$ into 6 different categories. Let

$$\mathcal{C} := \left\{(s_1,\cdots,s_k): 2^{s_1+\cdots+s_k}\geq\sqrt{np}n^{k/2-1}, |D_1^{s_1}|\leq\cdots\leq|D_k^{s_k}|\right\} \qquad (4.19)$$

be the ordered index set for heavy tuples where we assume $|D_1^{s_1}|\leq\cdots\leq|D_k^{s_k}|$. For the case where the sequence $\{|D_i^{s_i}|, 1\leq i\leq k\}$ have different orders can be

treated similarly, and there are $k!$ many in total. We then define the following 6 categories in $\mathcal{C}$:

$$\begin{aligned}
\mathcal{C}_1 &= \{(s_1,\dots,s_k)\in\mathcal{C} : \sigma_{s_1,\dots,s_k}\le 1\},\\
\mathcal{C}_2 &= \{(s_1,\dots,s_k)\in\mathcal{C}\setminus\mathcal{C}_1 : \lambda_{s_1,\dots,s_k}\le ec_2\},\\
\mathcal{C}_3 &= \left\{(s_1,\dots,s_k)\in\mathcal{C}\setminus(\mathcal{C}_1\cup\mathcal{C}_2) : 2^{s_1+s_2+\cdots+s_{k-1}-s_k}\ge n^{k/2-1}\sqrt{np}\right\},\\
\mathcal{C}_4 &= \{(s_1,\dots,s_k)\in\mathcal{C}\setminus(\mathcal{C}_1\cup\mathcal{C}_2\cup\mathcal{C}_3) : \log\lambda_{s_1,\dots,s_k} > \frac{1}{2}s_k\log 2+\frac{1}{4}\log(\alpha_{k,s_k}^{-1})\},\\
\mathcal{C}_5 &= \{(s_1,\dots,s_k)\in\mathcal{C}\setminus(\mathcal{C}_1\cup\mathcal{C}_2\cup\mathcal{C}_3\cup\mathcal{C}_4) : 2s_k\log 2\ge\log(1/\alpha_{k,s_k})\},\\
\mathcal{C}_6 &= \mathcal{C}\setminus(\mathcal{C}_1\cup\mathcal{C}_2\cup\mathcal{C}_3\cup\mathcal{C}_4\cup\mathcal{C}_5).
\end{aligned}$$

For the rest of the proof, we will show for all 6 categories $\{\mathcal{C}_t, 1\le t\le 6\}$,

$$\sum_{(s_1,\dots,s_k)\in\mathcal{C}}\alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\dots,s_k}\mathbf{1}\{(s_1,\dots,s_k)\in\mathcal{C}_t\}\le C_k\log^{k-2}(n) \tag{4.20}$$

for some constant C_k depending only on k, c_1, c_2, c_3 and δ, where the constants c_1, c_2, c_3 are the same ones as in Lemma 4.3 and Lemma 4.4. Here C_k depends exponentially on k and linearly on r. We will repeatedly use the following estimate which follows from (4):

$$\sum_{s_i=1}^{\lceil\log_2(\sqrt{n}/\delta)\rceil}\alpha_{i,s_i}\le\sum_{j\in[n]}|2y_{i,j}/\delta|^2\le(2/\delta)^2,\quad\forall 1\le i\le k. \tag{4.21}$$

From now on, for simplicity, whenever we are summing over s_i for some $1\le i\le k$, the range of s_i is understood as $1\le s_i\le\lceil\log_2(\sqrt{n}/\delta)\rceil$.

Tuples in $\mathcal{C}_1$

In this case we get

$$\begin{aligned}
&\sum_{(s_1,\dots,s_k)\in\mathcal{C}}\alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\dots,s_k}\mathbf{1}\{(s_1,\dots,s_k)\in\mathcal{C}_1\}\\
\le&\sum_{(s_1,\dots,s_k)\in\mathcal{C}}\alpha_{1,s_1}\cdots\alpha_{k,s_k}\le(2/\delta)^{2k},
\end{aligned}$$

where the last inequality comes from (4.21).

Tuples in $\mathcal{C}_2$

The constraint on $\mathcal{C}_2$ is the same as the condition in Case (1) of Lemma 4.4. Recall Definition (5) and (4.19). We have

$$\sigma_{s_1,\dots,s_k}=\lambda_{s_1,\dots,s_k}n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)}\le\lambda_{s_1,\dots,s_k}\le ec_2.$$

Therefore,

$$\begin{aligned}&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}} \alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\ldots,s_k}\mathbf{1}\left\{(s_1,\ldots,s_k)\in\mathcal{C}_2\right\}\\ \leq& ec_2\sum_{s_1,\ldots,s_k}\alpha_{1,s_1}\cdots\alpha_{k,s_k}\leq ec_2(2/\delta)^{2k}.\end{aligned}$$

Tuples in $\mathcal{C}_3$

By Lemma 4.3, all $(k-1)$-tuples have bounded degrees. Therefore we have

$$e(D_1^{s_1},\ldots,D_k^{s_k})\leq c_1|D_1^{s_1}|\cdots|D_{k-1}^{s_{k-1}}|np.$$

Hence by Definition (3),

$$\lambda_{s_1,\ldots,s_k}=\frac{e(D_1^{s_1},\ldots,D_k^{s_k})}{p|D_1^{s_1}|\cdots|D_k^{s_k}|}\leq\frac{c_1 n}{|D_k^{s_k}|}. \tag{4.22}$$

Therefore we have

$$\begin{aligned}&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}} \alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\ldots,s_k}\mathbf{1}\left\{(s_1,\ldots,s_k)\in\mathcal{C}_3\right\}\\ =&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}_3}\alpha_{1,s_1}\alpha_{2,s_2}\cdots\alpha_{k,s_k}\lambda_{s_1,\ldots,s_k}n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)}\\ \leq&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}_3}\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\frac{|D_k^{s_k}|2^{2s_k}}{n}\frac{c_1 n}{|D_k^{s_k}|}n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)}\\ =&c_1 n^{k/2-1}\sqrt{np}\sum_{(s_1,\ldots,s_k)\in\mathcal{C}_3}\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}2^{s_k-(s_2+\cdots+s_{k-1})}\\ =&c_1\sum_{s_1,\ldots,s_k}\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}n^{k/2-1}\sqrt{np}\cdot 2^{s_k-(s_2+\cdots+s_{k-1})}\mathbf{1}\left\{(s_1,\ldots,s_k)\in\mathcal{C}_3\right\},\end{aligned} \tag{4.23}$$

where the inequality in the third line is from (4.22). Since for all $(s_1,\ldots,s_k)\in\mathcal{C}_3$ we have $n^{k/2-1}\sqrt{np}\cdot 2^{s_k-(s_2+\cdots+s_{k-1})}\leq 1$, it implies that

$$\sum_{s_k} n^{k/2-1}\sqrt{np}\cdot 2^{s_k-(s_1+\cdots+s_{k-1})}\mathbf{1}\left\{(s_1,\ldots,s_k)\in\mathcal{C}_3\right\}\leq\sum_{i=0}^{\infty}2^{-i}\leq 2.$$

In view of (4.21), we can bound (4.23) by

$$2c_1\sum_{s_1,\ldots,s_{k-1}}\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\leq 2c_1(2/\delta)^{2k-2}.$$

This completes the proof for the case of $\mathcal{C}_3$. For the remaining categories $\mathcal{C}_4,\mathcal{C}_5$ and $\mathcal{C}_6$, we rely on the Case (2) in the bounded discrepancy lemma. Recall $\mathcal{C}_2$

corresponds to Case (1) in Lemma 4.4. Therefore Case (2) must hold in $\mathcal{C}_4, \mathcal{C}_5$ and $\mathcal{C}_6$. Case (2) in Lemma 4.4 can be written as

$$\lambda_{s_1,\ldots,s_k}|D_1^{s_1}|\cdots|D_k^{s_k}|\cdot p\log\lambda_{s_1,\ldots,s_k}\le c_3|D_k^{s_k}|\log\left(\frac{n}{|D_k^{s_k}|}\right).$$

By definitions in (5), the inequality above is equivalent to

$$\begin{aligned}&\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\sigma_{s_1,\ldots,s_k}\log\lambda_{s_1,\ldots,s_k}\\ \le& c_3\frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}}\left(2s_k\log 2+\log\alpha_{k,s_k}^{-1}\right).\end{aligned}\tag{4.24}$$

For the remaining of our proof, we will repeatedly use (4.24).

Tuples in $\mathcal{C}_4$

The inequality $\log\lambda_{s_1,\ldots,s_k}>\frac{1}{4}(2s_k\log 2+\log(1/\alpha_{k,s_k}))$ in the assumption of $\mathcal{C}_4$ and (4.24) imply that

$$\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\sigma_{s_1,\ldots,s_k}\le 4c_3n^{1-k/2}\cdot 2^{s_1+\cdots+s_{k-1}-s_k}/\sqrt{np}.$$

Then we have

$$\begin{aligned}&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}}\alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\ldots,s_k}\mathbf{1}\{(s_1,\ldots,s_k)\in\mathcal{C}_4\}\\ =&\sum_{s_k}\alpha_{k,s_k}\sum_{s_1,\ldots,s_{k-1}}\alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\sigma_{s_1,\ldots,s_k}\mathbf{1}\{(s_1,\ldots,s_k)\in\mathcal{C}_4\}\\ \le&4c_3\sum_{s_k}\alpha_{k,s_k}\sum_{s_1,\ldots,s_{k-1}}\frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}}\mathbf{1}\{(s_1,\ldots,s_k)\in\mathcal{C}_4\}.\end{aligned}\tag{4.25}$$

Since $(s_1,\ldots,s_k)\notin\mathcal{C}_3$, we have $\frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}}\le 1$ for all $(s_1,\ldots,s_k)\in\mathcal{C}_4$. Therefore (4.25) is bounded by

$$\begin{aligned}&4c_3\sum_{s_k}\alpha_{k,s_k}\sum_{s_1,\ldots,s_{k-2}}\sum_{s_{k-1}}\frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}}\mathbf{1}\{(s_1,\ldots,s_k)\in\mathcal{C}_4\}\\ \le&4c_3\sum_{s_k}\alpha_{k,s_k}\sum_{s_1,\ldots,s_{k-2}}2\le 8c_3\sum_{s_k}\alpha_{k,s_k}(\log_2(\sqrt{n}/\delta)+1)^{k-2},\end{aligned}\tag{4.26}$$

where the last inequality is from the fact that s_i satisfies $1\le s_i\le\lceil\log_2(\sqrt{n}/\delta)\rceil$ for $i\in[k]$ (see (4.11)). Therefore (4.26) can be bounded by

$$8c_3\left(\frac{1}{2}\log_2 n-\log_2(\delta)+1\right)^{k-2}(2/\delta)^2\le C\log^{k-2}(n)\tag{4.27}$$

for a constant C depending only on δ, k and c_3.

Tuples in $\mathcal{C}_5$

In this case we have $2s_k \log 2 \geq \log(\alpha_{k,s_k}^{-1})$. Also because $(s_1, \ldots, s_k) \notin \mathcal{C}_4$, we have

$$\log \lambda_{s_1,\ldots,s_k} \leq \frac{1}{4}\left(2s_k \log 2 + \log(1/\alpha_{k,s_k})\right) \leq s_k \log 2, \tag{4.28}$$

thus $\lambda_{s_1,\ldots,s_k} \leq 2^{s_k}$. On the other hand, because $(s_1, \ldots, s_k) \notin \mathcal{C}_1$,

$$1 < \sigma_{s_1,\ldots,s_k} = \lambda_{s_1,\ldots,s_k} n^{k/2-1}\sqrt{np} \cdot 2^{-(s_1+\cdots+s_k)} \leq n^{k/2-1}\sqrt{np} \cdot 2^{-(s_1+\cdots+s_{k-1})}.$$

Therefore we have

$$2^{s_1+\cdots+s_{k-1}} \leq n^{k/2-1}\sqrt{np}. \tag{4.29}$$

In addition, since $(s_1, \ldots, s_k) \notin \mathcal{C}_2$, we have $\lambda_{s_1,\ldots,s_k} > ec_2 > e$, which implies $\log \lambda_{s_1,\ldots,s_k} \geq 1$. Recall (4.24), together with (4.28), we then have

$$\begin{aligned}
\alpha_{1,s_1} \cdots \alpha_{k-1,s_{k-1}} \sigma_{s_1,\ldots,s_k} &\leq \alpha_{1,s_1} \cdots \alpha_{k-1,s_{k-1}} \sigma_{s_1,\ldots,s_k} \log \lambda_{s_1,\ldots,s_k} \\
&\leq c_3 \frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}} \left(2s_k \log 2 + \log \alpha_{k,s_k}^{-1}\right) \\
&\leq 4c_3 \log 2 \cdot s_k \frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{(s_1,\ldots,s_k)\in\mathcal{C}} \alpha_{1,s_1} \cdots \alpha_{k,s_k} \sigma_{s_1,\ldots,s_k} \mathbf{1}\left\{(s_1, \ldots s_k) \in \mathcal{C}_5\right\} \\
=&\sum_{s_k} \alpha_{k,s_k} \sum_{s_1,\ldots,s_{k-1}} \alpha_{1,s_1} \cdots \alpha_{k-1,s_{k-1}} \sigma_{s_1,\ldots,s_k} \mathbf{1}\left\{(s_1, \ldots s_k) \in \mathcal{C}_5\right\} \\
\leq&\sum_{s_k} \alpha_{k,s_k} \sum_{s_1,\ldots,s_{k-1}} 4c_3 \log 2 \cdot s_k \frac{2^{s_1+\cdots+s_{k-1}-s_k}}{n^{k/2-1}\sqrt{np}} \mathbf{1}\left\{(s_1, \ldots s_k) \in \mathcal{C}_5\right\} \\
\leq&4c_3 \sum_{s_k} \alpha_{k,s_k} \cdot s_k 2^{-s_k} \sum_{s_1,\ldots,s_{k-1}} \frac{2^{s_1+\cdots+s_{k-1}}}{n^{k/2-1}\sqrt{np}} \mathbf{1}\left\{(s_1, \ldots s_k) \in \mathcal{C}_5\right\}.
\end{aligned} \tag{4.30}$$

From (4.29), we have $\frac{2^{s_1+\cdots+s_{k-1}}}{n^{k/2-1}\sqrt{np}} \leq 1$ for any $(s_1, \ldots, s_k) \in \mathcal{C}_5$. Note that $s_k \cdot 2^{-s_k} \leq \frac{1}{2}$, therefore there exists a constant C depending only on δ, k and c_3 such that (4.30) can be bounded by

$$\begin{aligned}
&2c_3 \cdot \sum_{s_k} \alpha_{k,s_k} \sum_{s_1,\ldots,s_{k-1}} \frac{2^{s_1+\cdots+s_{k-1}}}{n^{k/2-1}\sqrt{np}} \mathbf{1}\left\{(s_1, \ldots s_k) \in \mathcal{C}_5\right\} \\
\leq&2c_3(2/\delta)^2(\log_2(\sqrt{n}/\delta) + 1)^{k-2} \leq C \log^{k-2}(n),
\end{aligned}$$

where the inequality above follows in the same way as in (4.26) and (4.27).

Tuples in $\mathcal{C}_6$

In this case we have $2s_k \log 2 < \log(\alpha_{k,s_k}^{-1})$. Because $(s_1, \dots s_k) \notin (C_4 \cup \mathcal{C}_2)$, we have

$$1 \le \log \lambda_{s_1,\dots,s_k} \le \frac{1}{4}\left[2s_k \log 2 + \log(1/\alpha_{k,s_k})\right] \le \frac{1}{2}\log \alpha_{k,s_k}^{-1} \le \log \alpha_{k,s_k}^{-1},$$

which implies $\lambda_{s_1,\dots,s_k}\alpha_{k,s_k} \le 1$. Recall Definition (5). We obtain

$$\begin{aligned}
&\sum_{(s_1,\dots,s_k)\in\mathcal{C}} \alpha_{1,s_1}\cdots\alpha_{k,s_k}\sigma_{s_1,\dots,s_k}\mathbf{1}\left\{(s_1,\dots,s_k)\in\mathcal{C}_6\right\} \\
=&\sum_{(s_1,\dots,s_{k-1},s_k)\in\mathcal{C}_6} \alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\alpha_{k,s_k}\lambda_{s_1,\dots,s_k} n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)} \\
\le&\sum_{s_1,\dots,s_{k-1}} \alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}}\sum_{s_k} n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)}\mathbf{1}\left\{(s_1,\dots,s_k)\in\mathcal{C}_6\right\}.
\end{aligned} \tag{4.31}$$

Recall from (4.19), $2^{s_1+\cdots+s_k} \ge \sqrt{np}\cdot n^{k/2-1}$, we have

$$\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)} \le n^{1-k/2}.$$

for all $(s_1,\dots,s_k)\in\mathcal{C}_5$. Hence

$$\sum_{s_k} n^{k/2-1}\sqrt{np}\cdot 2^{-(s_1+\cdots+s_k)}\mathbf{1}\left\{(s_1,\dots,s_k)\in\mathcal{C}_6\right\} \le 2.$$

Therefore (4.31) can be bounded by

$$2\sum_{s_1,\dots,s_{k-1}} \alpha_{1,s_1}\cdots\alpha_{k-1,s_{k-1}} \le 2(2/\delta)^{2k-2}.$$

Combining all the estimates from $\mathcal{C}_1$ to $\mathcal{C}_6$, we have (4.20) holds. This completes the proof of Theorem 2.1.

5. Proof of Theorem 2.2

Let $e_1 = (1,0,\dots,0)$ be a unit vector in $\mathbb{R}^n$ and denote $W = T - \mathbb{E}T$. Define a matrix $A \in \mathbb{R}^{n\times n}$ such that $a_{i_1,i_2} = w_{i_1,i_2,1,\dots,1}$. By the definition of the spectral norm,

$$\begin{aligned}
\|T-\mathbb{E}T\| &\ge \max_{\|x\|_2=\|y\|_2=1} |\langle W, x\otimes y\otimes e_1\otimes\cdots\otimes e_1\rangle| \\
&= \max_{\|x\|_2=\|y\|_2=1}\left|\sum_{i_1,i_2} w_{i_1,i_2,1,\dots,1}\cdot x_{i_1}y_{i_2}\right| = \|A\|.
\end{aligned}$$

Note that A is an $n\times n$ sparse random matrix with independent centered Bernoulli entries. Since the limiting singular value distribution of $\frac{A}{\sqrt{np}}$ is known (see for example Theorem 3.6 in [4]), the largest singular value of A is at least $(2-o(1))\sqrt{np}$ with high probability. Therefore $\|T-\mathbb{E}T\| \ge \sqrt{np}$ with high probability. This completes the proof.

6. Proof of Theorem 2.3

When $p = o(\log n/n)$, a direct application of the Kahn-Szemerédi argument would fail since the bounded degree and bounded discrepancy properties in the proof of Theorem 2.1 do not hold with high probability. We will use the following operation called tensor unfolding, in the proof of Theorem 2.3. For more details, see [37, 61].

Definition 6.1 (Partition)**.** *For any n, the l-partition π of $[k]$ is a collection $\{B_1^\pi, \dots, B_l^\pi\}$ of l disjoint non-empty subsets $B_i^\pi, 1 \le i \le l$ such that $\cup_{i=1}^l B_i^\pi = [k]$.*

Definition 6.2 (Tensor unfolding)**.** *Let $T \in \mathbb{R}^{n\times\cdots\times n}$ be an order-k tensor and π be an l-partition of $[k]$. The partition of π defines a map*

$$\phi_\pi : [n]^k \to \left[n^{|B_1^\pi|}\right] \times \cdots \times \left[n^{|B_l^\pi|}\right], \qquad \phi_\pi(i_1, \dots, i_k) = (m_1, \dots, m_l),$$

where

$$m_j = 1 + \prod_{r\in B_j^\pi} (i_r - 1)J_r, \quad J_r = \prod_{\substack{l\in B_j^\pi,\\ l<r}} n.$$

The map ϕ_π induces an unfolding action $T \to \mathrm{Unfold}_\pi(T)$, where $\mathrm{Unfold}_\pi(T)$ is a tensor of order l such that $\mathrm{Unfold}_\pi(T)_{m_1,\dots,m_l} = T_{i_1,\dots,i_k}$.

We will use the following inequality between the spectral norms of the original tensor T and the unfolded tensor $\mathrm{Unfold}_\pi(T)$.

Lemma 6.1 (Proposition 4.1 in [61])**.** *For any order-k tensor T and any partition π of $[k]$, we have $\|T\| \le \|\mathrm{Unfold}_\pi(T)\|$.*

With Lemma 6.1, we are ready to prove Theorem 2.3.

Proof of Theorem 2.3. (1) Assume $p \ge \frac{c\log n}{n^m}$ with an integer m such that $k/2 \le m \le k-1$. Consider a 2-partition of $[k]$ denoted by $\pi_1 = \{\{1, 2, \dots, m\}, \{m+1, \dots, k\}\}$. From Lemma 6.1, we have

$$\|T - \mathbb{E}T\| \le \|\mathrm{Unfold}_{\pi_1}(T - \mathbb{E}T)\|. \tag{6.1}$$

Here $\mathrm{Unfold}_{\pi_1}(T - \mathbb{E}T)$ is an $n^{k-m} \times n^m$ random matrix whose entries are one-to-one correspondent to entries in $T - \mathbb{E}T$. Let $A \in \mathbb{R}^{n^m} \times \mathbb{R}^{n^m}$ be a matrix such that

$$A_{i,j} = \begin{cases} (\mathrm{Unfold}_{\pi_1}(T))_{i,j} & \text{if } i \in [n^{k-m}], j \in [n^m], \\ 0 & \text{otherwise.} \end{cases}$$

Then A is an adjacency matrix of a random directed graph on n^m many vertices with

$$p \ge \frac{c\log n}{n^m} = \frac{c}{m} \cdot \frac{\log(n^m)}{n^m}.$$

Then we apply Theorem 2.1 with the matrix case. For any $r > 0$, there is a constant $C > 0$ depending on r and $\frac{c}{m}$ such that $\|A - \mathbb{E}A\| \le C\sqrt{n^m p}$ with probability at least $1 - n^{-rm}$. Then from (6.1), with probability at least $1 - n^{-r}$,

$$\|T - \mathbb{E}T\| \le \|\mathrm{Unfold}_{\pi_1}(T - \mathbb{E}T)\| \le \|A - \mathbb{E}A\| \le C\sqrt{n^m p}.$$

This completes the proof of (2.3).

(2) Assume $p \ge \frac{c \log n}{n^m}$ with an integer m such that $1 \le m < k/2$. Denote $k = ml + s$ for some $l \ge 1, 0 \le s \le m - 1$.

When $s \ne 0$, let π_2 be a $(l+1)$-partition of $[k]$ into $l+1$ parts given by

$$\pi_2 = \{\{1, \dots, m\}, \{m+1, \dots, 2m\}, \dots, \{ml+1, \dots, ml+s\}\}.$$

Then $\mathrm{Unfold}_{\pi_2}(T - \mathbb{E}T) \in \mathbb{R}^{n^m} \times \cdots \times \mathbb{R}^{n^m} \times \mathbb{R}^{n^s}$ is a tensor of order $(l+1)$. Since $p \ge \frac{c}{m} \frac{\log(n^m)}{n^m}$, we can apply Theorem 2.1 and Lemma 6.1. Then with probability at least $1 - n^{-r}$, we have

$$\begin{aligned} \|T - \mathbb{E}T\| &\le \|\mathrm{Unfold}_{\pi_2}(T - \mathbb{E}T)\| \\ &\le C\sqrt{n^m p} \log^{l-1}(n^m) \le C_1 \sqrt{n^m p} \log^{\frac{k-1}{m} - 1}(n), \end{aligned}$$

where C_1 is a constant depending on k, r, c.

When $s = 0$, we can similarly define π_2 as a l-partition of $[k]$ into l blocks such that $\mathrm{Unfold}_{\pi_2}(T - \mathbb{E}T) \in \mathbb{R}^{n^m} \times \cdots \times \mathbb{R}^{n^m}$ is a tensor of order l. With probability at least $1 - n^{-r}$, for a constant C_2 depending on k, r, c, we have

$$\|T - \mathbb{E}T\| \le C_2 \sqrt{n^m p} \log^{k/m - 2}(n).$$

This completes the proof of Theorem 2.3. □

7. Proof of Theorem 2.4

In this section, we will prove Theorem 2.4. We first compute the packing number over the parameter space under the spectral norm, then apply Fano's inequality. We first introduce several useful lemmas for showing this result. We will use the version in [56].

Lemma 7.1 (Varshamov-Gilbert bound)**.** *For $n \ge 8$, there exists a subset $S \subset \{0,1\}^n$ such that $|S| \ge 2^{n/8} + 1$ and for every distinct pair of $\omega, \omega' \in S$, the Hamming distance satisfies*

$$H(\omega, \omega') := \|\omega - \omega'\|_1 > n/8.$$

Lemma 7.2 (Fano's inequality)**.** *Assume $N \ge 3$ and suppose $\{\theta_1, \dots, \theta_N\} \subset \Theta$ such that*

(i) for all $1 \le i < j \le N$, $d(\theta_i, \theta_j) \ge 2\alpha$, where d is a metric on Θ;
(ii) let P_i be the distribution with respect to parameter θ_i, then for all $i, j \in [N]$, P_i is absolutely continuous with respect to P_j;

(iii) for all $i, j \in N$, the Kullback-Leibler divergence $D_{\mathrm{KL}}(P_i \| P_j) \le \beta \log(N-1)$ for some $0 < \beta < 1/8$.

Then

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{P}(d(\hat{\theta}, \theta) \ge \alpha) \ge \frac{\sqrt{N-1}}{1+\sqrt{N-1}} \left(1 - 2\beta - \sqrt{\frac{2\beta}{\log(N-1)}}\right).$$

Since we will apply Fano's inequality associated with Kullback-Leibler divergence, it requires the following lemma about random tensor with independent Bernoulli entries.

Lemma 7.3. *For $0 \le a < b \le 1$, we consider parameters $\theta, \theta' \in [a,b]^{n^k}$ for $0 \le a < b \le 1$, and let P and P' be the corresponding distributions, then the Kullback-Leibler divergence satisfies*

$$D_{\mathrm{KL}}(P \| P') \le \frac{\|\theta - \theta'\|_F^2}{a(1-b)}.$$

Proof. We firstly consider entrywise KL-divergence. For $p, q \in [a,b]$,

$$\begin{aligned} D_{\mathrm{KL}}(\mathrm{Ber}(p) \| \mathrm{Ber}(q)) &= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \\ &= p \log \left(1 + \frac{p-q}{q}\right) + (1-p) \log \left(1 - \frac{p-q}{1-q}\right) \\ &\le p\Big(\frac{p-q}{q}\Big) + (1-p)\Big(-\frac{p-q}{1-q}\Big) = \frac{(p-q)^2}{q(1-q)} \le \frac{(p-q)^2}{a(1-b)}. \end{aligned}$$

By independence of each entry, we have $D_{\mathrm{KL}}(P \| P') \le \frac{\|\theta-\theta'\|_F^2}{a(1-b)}$. □

Now we are ready to prove Theorem 2.4.

Proof of Theorem 2.4. By Lemma 7.1, there exists a subset $\{\omega^{(1)}, \dots, \omega^{(N)}\}$ of $\{0,1\}^n$ such that

$$\min_{1 \le i < j \le N} H(\omega^{(i)}, \omega^{(j)}) > \frac{n}{8} \quad \text{and} \quad N \ge 2^{n/8} + 1 \ge e^{n/12} + 1.$$

We note that $H(\omega^{(i)}, \omega^{(j)}) = \|\omega^{(i)} - \omega^{(j)}\|_2^2$. Let W be a fixed order-$(k-1)$ tensor with entries either 0 or 1 and dimension $n \times \cdots \times n$. The entries of W is designed as follows. Let $m = \lfloor p^{-\frac{1}{k-1}} \rfloor \wedge n$, so $1 \le m^{k-1} \le 1/p$. We assign 1's to an $m \times \cdots \times m$ subtensor of W and assign 0's to the rest entries. Then the rank of W is 1 and $\|W\| = \|W\|_F = m^{(k-1)/2}$. Now we define for $i \in [N]$,

$$\theta^{(i)} := \frac{p}{2} J + \frac{p}{30} \omega^{(i)} \otimes W,$$

where $J \in \mathbb{R}^{n^k}$ is an order-k tensor with all ones, and

$$(\omega^{(i)} \otimes W)_{i_1, \dots, i_k} = \omega^{(i)}_{i_1} w_{i_2, \dots, i_k}.$$

Then for all $i, j \in [N]$, $\theta^{(i)} - \theta^{(j)} = \frac{p}{30}(\omega^{(i)} - \omega^{(j)}) \otimes W$. By the choice of $\theta^{(i)}$'s,

$$\min_{1\le i<j\le N} \|\theta^{(i)} - \theta^{(j)}\|^2 = \min_{1\le i<j\le n} \frac{\|\omega^{(i)} - \omega^{(j)}\|_2^2 \|W\|^2 p^2}{900} \ge \frac{nm^{k-1}p^2}{7200}.$$

On the other hand, $\|\omega^{(i)} - \omega^{(j)}\|_2^2 \le n$, so

$$\max_{1\le i<j\le N} \|\theta^{(i)} - \theta^{(j)}\|^2 = \max_{1\le i<j\le N} \frac{\|\omega^{(i)} - \omega^{(j)}\|_2^2 \|W\|^2 p^2}{900} \le \frac{nm^{k-1}p^2}{900}.$$

Let P_i be the distribution of a random tensor T associated with parameter $\theta^{(i)}$ for $i \in [N]$. Since $\theta^{(i)} \in [\frac{p}{2}, \frac{8p}{15}]^{n^k}$, by Lemma 7.3, we have

$$\begin{aligned}\max_{1\le i<j\le N} D_{\mathrm{KL}}(P_i \| P_j) &\le \max_{1\le i<j\le N} \frac{\|\theta^{(i)} - \theta^{(j)}\|_F^2}{\left(\frac{p}{2}\right)\left(1 - \frac{8p}{15}\right)} \\ &\le \frac{nm^{k-1}p^2}{900\left(\frac{p}{2}\right)\left(1 - \frac{8p}{15}\right)} \le \frac{nm^{k-1}p}{210} \le \frac{n}{210},\end{aligned}$$

where the last inequality is due to the choice $m = \lfloor p^{-\frac{1}{k-1}} \rfloor \wedge n \le p^{-\frac{1}{k-1}}$. To apply Fano's inequality, we let $\alpha = \frac{nm^{k-1}p^2}{14400}$ and verify that for $i, j \in [N]$,

$$D_{\mathrm{KL}}(\theta^{(i)}, \theta^{(j)}) \le \frac{n}{210} \le \beta \log e^{n/12}$$

for $\beta = \frac{1}{9}$. Then by Lemma 7.2, we have

$$\inf_{\hat\theta} \sup_{\theta\in[0,p]^{n^k}} \mathbb{P}\left(\|\hat\theta - \theta\|^2 \ge \frac{nm^{k-1}p^2}{14400}\right) \ge \frac{2^{n/16}}{1 + 2^{n/16}} \left(1 - \frac{2}{9} - \sqrt{\frac{2/9}{n/12}}\right) \ge \frac{1}{3}$$

when $n \ge 16$. By the choice of m, we have

$$nm^{k-1}p^2 = n(\lfloor p^{-\frac{1}{k-1}} \rfloor \wedge n)^{k-1}p^2 \ge (2^{1-k}np) \wedge (n^k p^2),$$

which gives the desired result. □

8. Proof of Lemma 2.1

Similar to (4.12), by Bernstein's inequality, we have for each $(i_1, \ldots, i_{k-m}) \in [n]^{k-m}$,

$$\mathbb{P}(d_{i_1,\ldots,i_{k-m}} > 2n^m p) \le \exp\left(-\frac{3n^m p}{8}\right).$$

Then $\mathbf{1}\{d_{i_1,\ldots,i_{k-m}} > 2n^m p\}$ is a Bernoulli random variable with mean at most $\mu := \exp\left(-\frac{3n^m p}{8}\right)$. Since $d_{i_1,\ldots,i_{k-m}}$ are independent for all $(i_1, \ldots, i_{k-m}) \in [n]^{k-m}$, by Chernoff's inequality (4.14), for any $\lambda \ge 0$,

$$\mathbb{P}\left(|\tilde S| \ge (1+\lambda)n^{k-m}\mu\right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\sum_{i_1,\dots,i_{k-m}\in[n]} \mathbf{1}\{d_{i_1,\dots,i_{k-m}} > 2n^m p\} \geq (1+\lambda) n^{k-m}\mu \right) \\
&\leq \exp\left(-\frac{\lambda^2 n^{k-m}\mu}{2+\lambda} \right). \qquad (8.1)
\end{aligned}
$$

Since $n^m p \geq c$, we can choose a constant $c = 9$ and take

$$
\lambda = \frac{1}{n^m p \mu} - 1 = \frac{\exp\left(\frac{3n^m p}{8}\right)}{n^m p} - 1 \geq 1, \qquad (8.2)
$$

so that $2 + \lambda \leq 3\lambda$, and from (8.2) we know

$$
n^{k-m} \exp\left(-\frac{3n^m p}{8} \right) \leq \frac{1}{2n^{2m-k}p}. \qquad (8.3)
$$

Then (8.1) implies

$$
\begin{aligned}
\mathbb{P}\left(|S| \geq \frac{1}{n^{2m-k}p} \right) &\leq \exp\left(-\frac{\lambda n^{k-m}\mu}{3} \right) \\
&= \exp\left(-\frac{1}{3} n^{k-m}\mu \left(\frac{1}{n^m p\mu} - 1 \right) \right) \\
&= \exp\left(-\frac{1}{3n^{2m-k}p} + \frac{1}{3} n^{k-m} \exp\left(-\frac{3n^m p}{8} \right) \right) \\
&\leq \exp\left(-\frac{1}{6n^{2m-k}p} \right) \leq \exp\left(-\frac{n^{k-m}}{6\log n} \right),
\end{aligned}
$$

where the last line of inequalities follows from (8.3) and our assumption that $n^m p < \log n$.

9. Proof of Theorem 2.5

Let T be the adjacency tensor of a k-uniform random directed hypergraph and $P = \mathbb{E}T$. Recall Definition 6.2. Let $\pi = \{\{1, 2, \dots, k-m\}, \{k-m+1, \dots, k\}\}$ be a 2-partition of $[k]$. Then $\mathrm{Unfold}_\pi(T - \mathbb{E}T)$ is an $n^{k-m} \times n^m$ random matrix whose entries are one-to-one correspondent to entries in $(T - \mathbb{E}T)$. Let $A \in \mathbb{R}^{n^m} \times \mathbb{R}^{n^m}$ be a matrix such that

$$
A_{i,j} = \begin{cases} \mathrm{Unfold}_\pi(T)_{i,j} & \text{if } i \in [n^{k-m}], j \in [n^m], \\ 0 & \text{otherwise.} \end{cases}
$$

Then A is an adjacency matrix of a random directed graph on n^m vertices with $p \geq \frac{c}{n^m}$. Regularizing A by removing vertices of degrees greater than $2n^m p$, from Theorem 2.1 in [40], we have with probability at least $1 - n^{-rm}$, $\|\hat{A} - \mathbb{E}A\| \leq C\sqrt{n^m p}$. By the way we regularize an order-k random tensor T

in (2.6), we have $\mathrm{Unfold}_\pi(\hat{T}-P)$ is a submatrix of $(\hat{A}-\mathbb{E}A)$ with other entries being 0. Therefore by Lemma 6.1, with probability at least $1-n^{-r}$,

$$\|\hat{T}-P\| \le \|\mathrm{Unfold}_\pi(\hat{T}-P)\| \le \|\hat{A}-\mathbb{E}A\| \le C\sqrt{n^m p}.$$

This completes the proof of (2.7) in Theorem 2.5.

10. Proof of Theorem 3.1

Let 1_{V_i} be the indicator vector of $V_i \neq \emptyset, 1 \le i \le k$ such that the j-th entry of 1_{V_i} is 1 if $j \in V_i$ and 0 if $j \notin V_i$. We then have

$$\begin{aligned}
&\frac{|e_{H'}(V_1,\dots,V_k) - p|V_1|\cdots|V_k||}{\sqrt{|V_1|\cdots|V_k|}} \\
=&\frac{|T'(1_{V_1},\dots,1_{V_k}) - p\cdot J(1_{V_1},\dots,1_{V_k})|}{\sqrt{|V_1|\cdots|V_k|}} \\
=&\left|T'\left(\frac{1_{V_1}}{\sqrt{|V_1|}},\dots,\frac{1_{V_k}}{\sqrt{|V_k|}}\right) - p\cdot J\left(\frac{1_{V_1}}{\sqrt{|V_1|}},\dots,\frac{1_{V_k}}{\sqrt{|V_k|}}\right)\right| \\
\le& \|T'-pJ\| \le C\sqrt{n^{k-1}p} = C\sqrt{c},
\end{aligned}$$

where the last line is from the definition of the spectral norm for tensors and (3.3). Then (3.5) follows.

11. Conclusions

In this paper, we considered the concentration of sparse random tensors under spectral norm in different sparsity regimes. When $p \ge \frac{c\log n}{n}$, the extra log factor in Theorem 2.1 is due to the proof techniques, which only appears when bounding the contribution from $\mathcal{C}_4$ and $\mathcal{C}_5$ in Section 4.3. The Kahn-Szemerédi argument was specially designed to control random quadratic forms in the matrix case, but not for random multi-linear forms in the tensor case. An improvement to remove the extra log factors (which we conjecture should not appear) will require a new argument.

The regularization step we used is a sufficient way to recover the concentration of spectral norms, and it relies on the tensor unfolding inequality in Lemma 6.1. It remains to extend the analysis to other sparsity regimes. Proving a lower bound on the spectral norm without regularization will involve a more combinatorial argument similar to [38, 8]. We leave it as a future direction.

It would also be interesting to discuss the dependence on k for the constant C in (2.2). Using the ε-net argument, we obtain a constant C depending exponentially on k. It is possible that by using different arguments, the dependence can be improved.

Acknowledgments

We thank anonymous referees for their detailed comments and suggestions, which have improved the quality of this paper. We also thank Arash A. Amini, Nicholas Cook, Ioana Dumitriu, Kameron Decker Harris, and Roman Vershynin for helpful comments.

References

[1] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014. MR3270750

[2] Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *53rd Annual Allerton Conference on Communication, Control, and Computing, 2015 (Allerton)*, pages 66–73. IEEE, 2015.

[3] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013. MR2999295

[4] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010. MR2567175

[5] Pierre Baldi and Roman Vershynin. Polynomial threshold functions, hyperplane arrangements, and random tensors. *SIAM Journal on Mathematics of Data Science*, 2019. MR4016132

[6] Afonso S Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016. MR3531673

[7] Gérard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019. MR4011861

[8] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *Annals of Probability*, 47(3):1653–1676, 2019. MR3945756

[9] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 56(3):2141–2161, 2020. MR4116720

[10] Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *International Conference on Machine Learning*, pages 1881–1889, 2014.

[11] Yonatan Bilu and Shlomo Hoory. On codes from hypergraphs. *European Journal of Combinatorics*, 25(3):339–354, 2004. MR2036471

[12] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular ramanujan graphs. *The Annals of Probability*, 46(1):1–71, 2018. MR3758726

[13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration*

inequalities: A nonasymptotic theory of independence. Oxford university press, 2013. MR3185193

[14] Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and its applications. *arXiv preprint arXiv:1804.07808*, 2018. MR3768465

[15] Andrei Z Broder, Alan M Frieze, Stephen Suen, and Eli Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM Journal on Computing*, 28(2):541–573, 1998. MR1634360

[16] Changxiao Cai, Gen Li, H Vincent Poor, and Yuxin Chen. Nonconvex low-rank symmetric tensor completion from noisy data. *Advances in Neural Information Processing Systems*, 32, 2019.

[17] Wei-Kuo Chen. Phase transition in the spiked random tensor with Rademacher prior. *The Annals of Statistics*, 47(5):2734–2756, 2019. MR3988771

[18] Fan Chung. *Spectral Graph Theory*. No. 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences, 1997. MR1421568

[19] Emma Cohen, Dhruv Mubayi, Peter Ralli, and Prasad Tetali. Inverse expander mixing for hypergraphs. *The Electronic Journal of Combinatorics*, 23(2):2–20, 2016. MR3512642

[20] Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593:45–73, 2020. MR4066151

[21] Nicholas Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. *The Annals of Probability*, 46(1):72–125, 2018. MR3758727

[22] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 974–985. IEEE, 2017. MR3734297

[23] Ioana Dumitriu, Tobias Johnson, Soumik Pal, and Elliot Paquette. Functional limit theorems for random regular graphs. *Probability Theory and Related Fields*, 156(3-4):921–975, 2013. MR3078290

[24] Ioana Dumitriu and Yizhe Zhu. Spectra of random regular hypergraphs. *arXiv preprint arXiv:1905.06487*, 2019.

[25] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005. MR2155709

[26] Joel Friedman, Jeff Kahn, and Endre Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598. ACM, 1989.

[27] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995. MR1325271

[28] David Gamarnik, Quan Li, and Hongyi Zhang. Matrix completion from $O(n)$ samples in linear time. In *Conference on Learning Theory*, pages 940–947, 2017.

[29] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.

[30] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, 2017. MR3611493

[31] Kameron Decker Harris and Yizhe Zhu. Deterministic tensor completion with hypergraph expanders. *arXiv preprint arXiv:1910.10692*, 2019.

[32] Eyal Heiman, Gideon Schechtman, and Adi Shraibman. Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 45(2):306–317, 2014. MR3245293

[33] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013. MR3144915

[34] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems*, pages 1431–1439, 2014.

[35] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE transactions on information theory*, 56(6):2980–2998, 2010. MR2683452

[36] Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages 124–128. IEEE, 2017.

[37] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. MR2535056

[38] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12(1):61–72, 2003. MR1967486

[39] Rafał Latała, Ramon van Handel, and Pierre Youssef. The dimension-free structure of nonhomogeneous random matrices. *Inventiones mathematicae*, 214(3):1031–1080, 2018. MR3878726

[40] Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017. MR3689343

[41] Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73, 2020. MR4064140

[42] Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015. MR3285605

[43] John Lenz and Dhruv Mubayi. Eigenvalues and linear quasirandom hypergraphs. In *Forum of Mathematics, Sigma*, volume 3. Cambridge University Press, 2015. MR3324939

[44] Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017. MR3683819

[45] Linyuan Lu and Xing Peng. Loose Laplacian spectra of random hypergraphs. *Random Structures & Algorithms*, 41(4):521–545, 2012. MR2993134

[46] Eyal Lubetzky, Benny Sudakov, and Van Vu. Spectra of lifted Ramanujan graphs. *Advances in Mathematics*, 227(4):1612–1645, 2011. MR2799807

[47] Alexander Lubotzky. High dimensional expanders. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 705–730, 2019. MR3966743

[48] Andrea Montanari and Nike Sun. Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425, 2018. MR3862094

[49] Nam H Nguyen, Petros Drineas, and Trac D Tran. Tensor sparsification via a bound on the spectral norm of random tensors. *Information and Inference: A Journal of the IMA*, 4(3):195–229, 2015. MR3394292

[50] Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, pages 1–57, 2021.

[51] Elizaveta Rebrova. Constructive regularization of the random matrix norm. *Journal of Theoretical Probability*, 33:1768–1790, 2020. MR4125975

[52] Elizaveta Rebrova and Roman Vershynin. Norms of random matrices: local and global problems. *Advances in Mathematics*, 324:40–83, 2018. MR3733881

[53] Emile Richard and Andrea Montanari. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

[54] Konstantin Tikhomirov and Pierre Youssef. The spectral gap of dense random regular graphs. *The Annals of Probability*, 47(1):362–419, 2019. MR3909972

[55] Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors. *arXiv preprint arXiv:1407.1870*, 2014.

[56] Alexandre B Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition, 2008. MR2724359

[57] Ramon van Handel. On the spectral norm of Gaussian random matrices. *Transactions of the American Mathematical Society*, 369(11):8161–8178, 2017. MR3695857

[58] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018. MR3837109

[59] Roman Vershynin. Concentration inequalities for random tensors. *Bernoulli*, 26(4):3139–3162, 2020. MR4140540

[60] Van H Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007. MR2384414

[61] Miaoyan Wang, Khanh Dao Duc, Jonathan Fischer, and Yun S Song. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear algebra and its applications*, 520:44–66, 2017. MR3611456

[62] Zhixin Zhou and Arash A Amini. Analysis of spectral clustering algorithms for community detection: the general bipartite setting. *Journal of Machine Learning Research*, 20(47):1–47, 2019. MR3948087

[63] Yizhe Zhu. On the second eigenvalue of random bipartite biregular graphs. *arXiv preprint arXiv:2005.08103*, 2020.