Exact Inference in High-order Structured Prediction

Chuyang Ke [1]   **Jean Honorio** [1]

Abstract

In this paper, we study the problem of inference in high-order structured prediction tasks. In the context of Markov random fields, the goal of a high-order inference task is to maximize a score function on the space of labels, and the score function can be decomposed into sum of unary and high-order potentials. We apply a generative model approach to study the problem of high-order inference, and provide a two-stage convex optimization algorithm for exact label recovery. We also provide a new class of hypergraph structural properties related to hyperedge expansion that drives the success in general high-order inference problems. Finally, we connect the performance of our algorithm and the hyperedge expansion property using a novel hypergraph Cheeger-type inequality.

1. Introduction

Structured prediction has been widely used in various machine learning fields in the past 20 years, including applications like social network analysis, computer vision, molecular biology, natural language processing (NLP), among others. A common objective in these tasks is assigning / recovering labels, that is, given some possibly noisy observation, the goal is to output a group label for each entity in the task. In social network analysis, this could be detecting communities based on user profiles and preferences (Ke & Honorio, 2022b; Kelley et al., 2012). In computer vision, researchers want the AI to decide whether a pixel is in the foreground or background (Nowozin et al., 2011). In biology, it is sometimes desirable to cluster molecules by structural similarity (Nugent & Meila, 2010). In NLP, part-of-speech tagging is probably one of the most well-known structured prediction task (Weiss & Taskar, 2010).

From a methodological point of view, a standard approach in

the structured prediction tasks above, is to recover the *global structure* by exploiting many *local structures*. Take social networks as an example. A widely used assumption in social network analysis is *homophily* — users with similar profiles and preferences are more likely to become friends (Ke & Honorio, 2018; 2019). Intuitively, a structured prediction algorithm tends to assign two users the same label, if they have a higher affinity score. Similarly, the same idea can be motivated in the context of Markov random fields (MRFs). Assume all entities form an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, structured prediction can be viewed as the task of solving the following inference problem (Bello & Honorio, 2019):

$$
\begin{aligned}
\underset{y \in \mathcal{L}^{|\mathcal{V}|}}{\text{maximize}} \quad & \sum_{v \in \mathcal{V}, l \in \mathcal{L}} c_v(l) \cdot \mathbb{1}[y_v = l] \\
& + \sum_{\substack{(v_1, v_2) \in \mathcal{E} \\ l_1, l_2 \in \mathcal{L}}} c_{v_1, v_2}(l_1, l_2) \cdot \mathbb{1}[y_{v_1} = l_1, y_{v_2} = l_2], \quad (1)
\end{aligned}
$$

where $\mathcal{L}$ is the space of labels, $c_v(l)$ is the score of assigning label l to node v, and $c_{v_1, v_2}(l_1, l_2)$ is the score of assigning labels l_1 and l_2 to neighboring nodes v_1 and v_2. In the MRF and inference literature, the two terms in (1) are often referred to as *unary* and *pairwise potentials*, respectively. The inference formulation above allows one to recover the global structure, by finding a configuration that maximizes the summation of unary and pairwise local scores.

However, entities in many real-world problems could interact beyond the pairwise fashion. Take the social network example again, but this time let us focus on the academia co-authorship network: many published papers are written by more than two authors (Liu et al., 2005). Such high-order interactions cannot be captured by pairwise structures. Geometrically, the co-authorship network can no longer be represented by a graph. As a result, the introduction of hypergraphs is necessary to model high-order structured prediction problems.

In this paper, we study the problem of high-order structured prediction, in which instead of using pairwise potentials, *high-order potentials* are considered. Using the MRF formulation, we are interested in inference problems of the

---

[1]Department of Computer Science, Purdue University, West Lafayette, IN, USA. Correspondence to: Chuyang Ke <cke@purdue.edu>.

following form:

$$\begin{aligned}
\underset{y \in \mathcal{L}^{|\mathcal{V}|}}{\text{maximize}} \quad & \sum_{v \in \mathcal{V}, l \in \mathcal{L}} c_v(l) \cdot \mathbb{1}[y_v = l] \\
+ & \sum_{\substack{e \in \mathcal{E} \\ l_1, \ldots, l_m \in \mathcal{L} \\ e = (v_1, \ldots, v_m)}} c_e(l_1, \ldots, l_m) \cdot \mathbb{1}[y_{v_1} = l_1, \ldots, y_{v_m} = l_m],
\end{aligned}$$

$$(2)$$

where m is the order of the inference problem as well as the hypergraph (each hyperedge connects m vertices), and $c_e(l_1, \ldots, l_m)$ is the score of assigning labels l_1 through l_m to neighboring nodes v_1 through v_m connected by hyperedge $e \in \mathcal{E}$.

1.1. Inference as a Recovery Task

Structured prediction and inference problems with unary and pairwise potentials in the form of (1) have been studied in prior literature. Globerson et al. (2015) introduced the problem of label recovery in the case of two-dimensional grid lattices, and analyzed the conditions for approximate inference. Along the same line of work, Foster et al. (2018) generalized the model by allowing tree decompositions. Another flavor is the problem of exact inference, for which Bello & Honorio (2019) proposed a convex semidefinite programming (SDP) approach. In these works, the problem of label recovery is motivated by a generative model, which assumes the existence of a ground truth label vector y^*, and generates (possibly noisy) unary and pairwise observations based on label interactions.

Unfortunately, little is known for structured prediction with high-order potentials and hypergraphs. In recent years, there have been various attempts to generalize some graph properties (including hypergraph Laplacian, Rayleigh quotient, hyperedge expansion, Cheeger constant, among others) to hypergraphs (Li & Milenkovic, 2018; Mulas, 2021; Yoshida, 2019; Chan et al., 2018; Chen et al., 2017; Chang et al., 2020). However, it took a long time for us to find out that due to the nature of structured prediction tasks, the hypergraph definitions must fulfill certain properties, and none of the definitions in the aforementioned works fulfill those. This makes it challenging to design hypergraph-based label recovery algorithms. Furthermore, none of the aforementioned works provide guarantees of either approximate or exact inference.

In this work, we apply a generative model approach to study the problem of high-order inference. We analyze the task of label recovery, and answer the following question:

Problem 1 (Label Recovery). *Is there any algorithm that takes noisy unary and high-order local observation as the input, and correctly recovers the underlying true labels?*

1.2. Inference and Structural Properties

In the MRF inference literature, a central and longtime discussion focuses on inference solvability versus certain structural properties of the problem.

To see this, we first revisit various classes of structural properties in pairwise inference problems (i.e., in the form of (1)). Chandrasekaran et al. (2012) studied treewidth in graphs as a structural property, and showed that graphs with low treewidths are solvable. Schraudolph & Kamenetsky (2008) demonstrated that planar graphs can be solved. Boykov & Veksler (2006) analyzed graphs with binary labels and sub-modular pairwise potentials. Bello & Honorio (2019) showed that inference with graphs that are "good" expanders, or "bad" expanders plus an Erdos-Renyi random graph, can be achieved. It is worth highlighting that the structural properties above are not directly comparable or reducible. Instead, they characterize the difficulty of an inference problem from different angles, or in other words, for different classes of graphs.

Similar discussions about inference versus structural properties exist in high-order MRF inference literature. For example, Komodakis & Paragios (2009) investigated hypergraphs fulfilling the property of one sub-hypergraph per clique, and proved that high-order MRF inference can be achieved through solving linear programs (LPs). Fix et al. (2014) studied inference in hypergraphs with the property of local completeness. Gallagher et al. (2011) analyzed the performance of high-order MRF inference through order reduction versus the number of non-submodular edges. However, these works do not provide theoretical guarantees of either approximate or exact inference.

In this paper, we provide a new class of hypergraph structural properties by analyzing *hyperedge expansion*. In order to get some intuition, let us consider a social network with two disconnected sub-networks. With an ideal algorithm, we may recover the user communities in subnet 1 and the those in subnet 2, but since there is no interaction between the two subnets at all, we will not be able to infer the global community structure (e.g., the relationship between the recovered communities in subnet 1 and subnet 2). A less extreme case is networks with "bottlenecks," i.e., removing these bottleneck edges will disconnect the network. For similar reasons, one can imagine that inference in networks with bottlenecks can be hard if noise is present. See Figure 1 for an illustration. In pairwise graphs (2-graphs), such connectivity / bottleneck property can be characterized by the edge expansion (i.e., the Cheeger constant) of the graph. Characterizing similar expansion properties in high-order hypergraphs poses a challenge, especially if one wants to relate such topological properties to the conditions of exact inference.

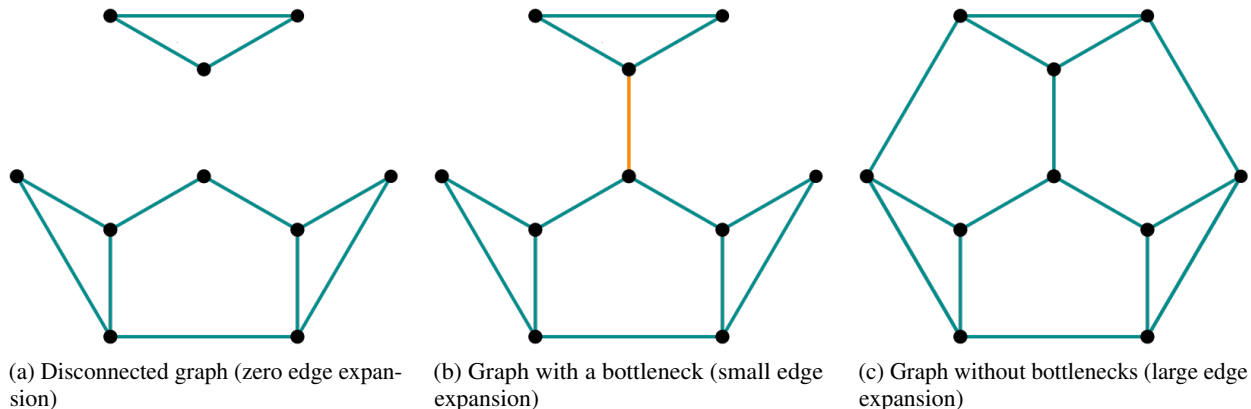**Problem 2** (Structural Property). *Under what topological*

(a) Disconnected graph (zero edge expansion)

(b) Graph with a bottleneck (small edge expansion)

(c) Graph without bottlenecks (large edge expansion)

Figure 1. Graph expansion examples. Figure 1a shows a disconnected graph. With an ideal algorithm we may be able to recover the user communities in the top subgraph and in the bottom subgraph, but since there is no observed interaction between the two, we will not be able to infer the global structure. Figure 1b connects the two components in Figure 1a using a single orange edge. In this case the orange edge is the "bottleneck." Removing the orange edge disconnects the graph. Figure 1c adds two more edges to Figure 1b. In this case every component is connected with no weak "bottleneck." Edge expansion is a structural property, which characterizes how connected the components in a graph are.

conditions will our label recovery algorithm work correctly with high probability?

Summary of our contribution. Our work is highly theoretical. We provide a series of novel definitions and results in this paper:

- We provide a new class of hypergraph structural properties for high-order inference problems. We derive a novel Cheeger-type inequality, which relates the tensor spectral gap of a hypergraph Laplacian to a Cheeger-type hypergraph expansion property. These hypergraph results are not only limited to the scope of the model in this paper, but also can be helpful to researchers working on high-order inference problems.

- We propose a two-stage approach to solve the problem of high-order structured prediction. We formulate the label recovery problem as a high-order combinatorial optimization problem, and further relax it to a novel convex conic form optimization problem.

- We carefully analyze the Karush–Kuhn–Tucker (KKT) conditions of the conic form optimization problem, and derive the sufficient statistical and topological conditions for exact inference. Our KKT analysis guarantees the solution to be optimal with a high probability, as long as the conditions are fulfilled.

2. Preliminaries

In this section, we formally define the high-order exact inference problem and introduce the notations that will be used throughout the paper.

We use lowercase font (e.g., a, b, u, v) to denote scalars and vectors, and uppercase font (e.g., A, B, C) to denote tensors. We denote the set of real numbers by $\mathbb{R}$.

For any natural number n, we use $[n]$ to denote the set $\{1, \ldots, n\}$.

We use $\mathbf{1}$ to denote the all-ones vector.

For clarity we use superscripts (i) to denote the i-th object in a sequence of objects, and subscripts j to denote the j-th entry. We use $\circ$ to denote the Hadamard product, and $\otimes$ to denote the outer product. Let $v^{(1)}, \ldots, v^{(m)} \in \mathbb{R}^n$ be a sequence of m vectors of dimension n. Then $v^{(1)} \otimes \ldots \otimes v^{(m)}$ is a tensor of order m and dimension n (or equivalently, of shape $n^{\otimes m}$), such that $(v^{(1)} \otimes \ldots \otimes v^{(m)})_{i_1, \ldots, i_m} = v_{i_1}^{(1)} \ldots v_{i_m}^{(m)}$.

2.1. Tensor Definitions

Let $A \in \mathbb{R}^{n^{\otimes m}}$ be an m-th order, n-dimensional real tensor. Throughout the paper, we limit our discussion to $m = 2, 6, 10, 14, \ldots$ for clarity of exposition. While other even orders ($m = 4, 8, 12 \ldots$) are possible and a similar analysis will follow, the hypergraph definitions will be involving many more terms and the paper will be less readable. See Remark 3.18 for discussion.

A symmetric tensor is invariant under any permutation of the indices. In other words, A is symmetric if for any permutation $\sigma : [m] \to [m]$, we have $A_{\sigma(i_1, \ldots, i_m)} = A_{i_1, \ldots, i_m}$.

We define the inner product of two tensors A, B of the same shape as $\langle A, B \rangle := \sum_{i_1, \ldots, i_m = 1}^{n} A_{i_1, \ldots, i_m} B_{i_1, \ldots, i_m}$. We define the tensor Frobenius norm as $\|A\|_{\mathrm{F}} := \sqrt{\langle A, A \rangle}$.

A symmetric tensor A is positive semidefinite (PSD), if for all $v \in \mathbb{R}^n$, we have $\langle A, v^{\otimes m} \rangle \geq 0$. We use $\mathcal{S}_+^{n,m}$ to denote the convex cone of all m-order, n-dimensional PSD tensors.

The dual cone of $\mathcal{S}_+^{n,m}$ is the Caratheodory tensor cone $\mathcal{S}_+^{*,n,m}$, which is defined as $\mathcal{S}_+^{*,n,m} := \left\{ \sum_{i=1}^{\binom{m+n-1}{m}} v^{(i)\otimes m} \mid v^{(i)} \in \mathbb{R}^n \right\}$. In other words, every tensor in $\mathcal{S}_+^{*,n,m}$ is the summation of at most $\binom{m+n-1}{m}$ rank-one tensors. $\mathcal{S}_+^{n,m}$ and $\mathcal{S}_+^{*,n,m}$ are dual to each other (Ke & Honorio, 2022a).

For any tensor $A \in \mathbb{R}^{n^{\otimes m}}$, we define its minimum tensor eigenvalue $\lambda_{\min}(A)$ (or equivalently $\lambda_1(A)$) using a variational characterization, such that $\lambda_{\min}(A) := \min_{v \in \mathbb{R}^n, \|v\|=1} \langle A, v^{\otimes m} \rangle$. Similarly we define its second minimum tensor eigenvalue $\lambda_2(A)$ as $\lambda_2(A) := \min_{v \in \mathbb{R}^n, \|v\|=1, v \perp v^{(1)}} \langle A, v^{\otimes m} \rangle$, where $v^{(1)}$ is the eigenvector corresponding to $\lambda_{\min}(A)$.

We denote $\sigma_2^{n,m}$ as the index set of m-tuples in the shape of $\sigma(i_1, i_1, i_2, i_2, \ldots, i_{m/2}, i_{m/2})$, for any permutation $\sigma : [m] \to [m]$ and $i_j \in [n]$. Intuitively, in every tuple of $\sigma_2^{n,m}$, every index repeats an even number of times. We use $\bar{\sigma}_2^{n,m}$ to denote the set $\{(i_1, \ldots, i_m) \mid (i_1, \ldots, i_m) \notin \sigma_2^{n,m}$, and at least one index repeats twice$\}$. In other words, $\bar{\sigma}_2^{n,m}$ is the complement of $\sigma_2^{n,m}$, subtracting cases with all unique indices.

2.2. High-order Inference Task

We consider the task of predicting a set of n vertex labels $y^* = (y_1^*, \ldots, y_n^*)$, where $y_i^* \in \{+1, -1\}$, from a set of observations X and z. X and z are noisy observations generated from some underlying m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. In particular, $\mathcal{V}$ is the set of vertices (nodes) with $|\mathcal{V}| = n$, and $\mathcal{E}$ is the set of hyperedges.

For every possible m-vertex tuple $e = (i_1, \ldots, i_m)$, if $e \in \mathcal{E}$, the hyperedge observation $X_{i_1, \ldots, i_m}$ (and all corresponding symmetric entries $X_{\sigma(i_1, \ldots, i_m)}$) is sampled to be $y_1^* \cdots y_m^*$ with probability $1 - p$, and $-y_1^* \cdots y_m^*$ with probability p independently. If $e \notin \mathcal{E}$, $X_{i_1, \ldots, i_m}$ is set to 0.

For every node $v_i \in \mathcal{V}$, the node observation z_i is sampled to be y_i^* with probability $1 - q$, and $-y_i^*$ with probability q independently.

We now summarize the generative model.

Definition 2.1 (High-order Structured Prediction with Partial Observation). **Unknown**: True node labeling vector $y^* = (y_1^*, \ldots, y_n^*)$. **Observation**: Partial and noisy hyperedge observation tensor $X \in \{-1, 0, +1\}^{n^{\otimes m}}$. Noisy node label observation vector $z \in \{-1, +1\}^n$. **Task**: Infer and recover the correct node labeling vector y^* from the observation X and z.

3. Hypergraph Structural Properties and Cheeger-type Inequality

In this section, we introduce a series of novel Cheeger-type analysis for hypergraphs. This allows us to characterize the spectral gap of a hypergraph Laplacian, via the topological hyperedge expansion of the graph itself. Hypergraph theorems in this section are general, and are not limited to the specific model covered in our inference task. To the best of our knowledge, the following high-order definitions and results are novel. All missing proofs of the lemmas and theorems can be found in Appendix A.

3.1. Hypergraph Topology

We first introduce the necessary hypergraph topological definitions.

Definition 3.1 (Induced Hypervertices). Given an m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we use

$$\mathcal{H} := \{\{i_1, \ldots, i_{m/2}\} \mid i_1, \ldots, i_{m/2} \in [n]\}$$

to denote the set of induced hypervertices. We denote its cardinality by $N := |\mathcal{H}| = \binom{n}{m/2}$.

Definition 3.2 (Boundary of a Hypervertex Set). For any hypervertex set $S \subset \mathcal{H}$, we denote its boundary set as

$$\partial S := \{h_1 \cup h_2 \mid h_1 \in S, h_2 \notin S, h_1 \cap h_2 = \emptyset, h_1 \cup h_2 \in \mathcal{E}\}.$$

Note that ∂S is a set of m-th order hyperedges.

Definition 3.3 (Hyperedge Expansion of a Hypervertex Set). For any hypervertex set $S \subset \mathcal{H}$, we denote the hyperedge expansion of the set S as

$$\phi_S := \frac{|\partial S|}{|S|}.$$

Definition 3.4 (Hyperedge Expansion of a Hypergraph). Given an m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with induced hypervertices $\mathcal{H}$, we denote the hyperedge expansion of the hypergraph $\mathcal{G}$ as

$$\phi_{\mathcal{G}} := \min_{S \subset \mathcal{H}, 0 < |S| \leq N/2} \phi_S = \min_{S \subset \mathcal{H}, 0 < |S| \leq N/2} \frac{|\partial S|}{|S|}.$$

We also call $\phi_{\mathcal{G}}$ the Cheeger constant of the hypergraph.

3.2. Hypergraph Laplacian

In this section, we introduce our hypergraph Laplacian related definitions.

Definition 3.5 (ζ-function). $\zeta : \mathbb{R}^m \to \mathbb{R}$ is a function defined as

$$\zeta(v_1, \ldots, v_m) = \sum_{\mathcal{I} \subset [m], |\mathcal{I}| = m/2} \left(\sum_{i \in \mathcal{I}} v_i - \sum_{j \notin \mathcal{I}} v_j \right)^m.$$

Definition 3.6 (Hypergraph Laplacian). Given an m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we use $L \in \mathbb{R}^{n^{\otimes m}}$ to denote its Laplacian tensor, which fulfills

$$\langle L, v^{\otimes m} \rangle = \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \zeta(v_{i_1}, \ldots, v_{i_m}) .$$

Remark 3.7. Definition 3.5 and 3.6 are the generalization of graph Laplacians to hypergraph cases. To see this, note that for any pairwise graph ($m = 2$), its graph Laplacian fulfills the following quadratic form property $v^\top L v = \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} (v_i - v_j)^2$. This is exactly equivalent to setting $m = 2$ in the definitions above.

Definition 3.8 (Rayleigh Quotient). For any hypergraph Laplacian L and non-zero vector $v \in \mathbb{R}^n$, the Rayleigh quotient $R_L(v)$ is defined as

$$R_L(v) := \frac{\langle L, v^{\otimes m} \rangle}{\|v\|^m} .$$

Definition 3.9 (Signed Laplacian Tensor). Given an m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a sign vector $y \in \{-1, +1\}^n$, we use $L_y \in \mathbb{R}^{n^{\otimes m}}$ to denote its signed Laplacian tensor, which fulfills

$$\langle L_y, v^{\otimes m} \rangle = \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \zeta(y_{i_1} v_{i_1}, \ldots, y_{i_m} v_{i_m}) .$$

Remark 3.10. The hypergraph Laplacian tensor L can be viewed as a signed Laplacian tensor L_y, by taking $y = \mathbf{1}$.

Here we provide some important properties of hypergraph Laplacians and Rayleigh quotients.

Lemma 3.11 (Laplacian Eigenpair). *For any hypergraph Laplacian L, $\mathbf{1}$ is an eigenvector of L with a minimum eigenvalue of 0. Similarly, for any signed hypergraph Laplacian L_y, y is an eigenvector of L_y with an eigenvalue of 0.*

Proof of Lemma 3.11. This follows directly from the definition of ζ-function, and Definition 3.6, 3.9. □

Lemma 3.12 (Invariant under Scaling). *For any non-zero $\alpha \in \mathbb{R}$, we have*

$$R_L(v) = R_L(\alpha v) .$$

Proof of Lemma 3.12.

$$
\begin{aligned}
R_L(\alpha v) &= \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\zeta(\alpha v_{i_1}, \ldots, \alpha v_{i_m})}{\|\alpha v\|^m} \\
&= \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\alpha^m \zeta(v_{i_1}, \ldots, v_{i_m})}{\alpha^m \|v\|^m} \\
&= \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\zeta(v_{i_1}, \ldots, v_{i_m})}{\|v\|^m} \\
&= R_L(v) .
\end{aligned}
$$

□

Lemma 3.13 (Invariant under Scaling). *For any $\delta \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $v \perp \mathbf{1}$, we have*

$$R_L(v) \geq R_L(v + \delta \mathbf{1}) .$$

Proof of Lemma 3.13. Note that

$$
\begin{aligned}
R_L(v + \delta \mathbf{1}) &= \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\zeta(v_{i_1} + \delta, \ldots, v_{i_m} + \delta)}{\|v + \delta \mathbf{1}\|^m} \\
&= \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\zeta(v_{i_1}, \ldots, v_{i_m})}{\|v + \delta \mathbf{1}\|^m} \\
&\leq \frac{1}{m! \binom{m}{m/2}} \sum_{(i_1, \ldots, i_m) \in \mathcal{E}} \frac{\zeta(v_{i_1}, \ldots, v_{i_m})}{\|v\|^m} \\
&= R_L(v) ,
\end{aligned}
$$

where the inequality follows from the fact that

$$
\begin{aligned}
\|v + \delta \mathbf{1}\|^m &= \left(\sum_i (v_i + \delta)^2 \right)^{m/2} \\
&= \left(\sum_i v_i^2 + n\delta^2 + 2\delta \sum_i v_i \right)^{m/2} \\
&= \left(\sum_i v_i^2 + n\delta^2 \right)^{m/2} \\
&\geq \left(\sum_i v_i^2 \right)^{m/2} \\
&= \|v\|^m .
\end{aligned}
$$

□

Lemma 3.14 (Signed Rayleigh Quotient Lower Bound). *For any hypergraph with Laplacian L and signed Laplacian L_y, and for any $\delta \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $v \perp y$, we have*

$$R_{L_y}(v) \geq R_L(v \circ y + \delta \mathbf{1}) .$$

Lemma 3.15 (Existence of Degree Tensor). *For any hypergraph Laplacian L, let A denote the corresponding symmetric adjacency tensor, such that $A_{i_1,\ldots,i_m} = 1$ if and only if $(i_1,\ldots,i_m) \in \mathcal{E}$. Then, there exists a high-order degree tensor D fulfilling: 1) $D_{i_1,\ldots,i_m} = 0$ if $i_1,\ldots,i_m$ are all different, and 2) $\langle D - A, v^{\otimes m}\rangle = \langle L, v^{\otimes m}\rangle$ for every vector $v \in \mathbb{R}^n$.*

3.3. High-order Cheeger-type Inequality

Here, we connect the previous two sections and provide our novel high-order Cheeger-type inequality.

Theorem 3.16 (High-order Cheeger-type Inequality). *For any m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with Laplacian L, we have*

$$\lambda_2(L) \geq \phi_{\mathcal{G}}^m.$$

Remark 3.17. Here we would like to discuss our choice of hypergraph Laplacian as well as the related high-order definitions. Despite the fact that there have been multiple works trying to generalize some graph properties (including hypergraph Laplacian, Rayleigh quotient, hyperedge expansion, Cheeger constant, among others) to hypergraphs (Li & Milenkovic, 2018; Mulas, 2021; Yoshida, 2019; Chan et al., 2018; Chen et al., 2017; Chang et al., 2020), we want to highlight that none above fulfill the requirement of our high-order inference tasks.

To obtain a set of hypergraph definitions that are useful for our inference task, we need to fulfill all the lemmas in this section. In particular, here are the necessary conditions:

- To fulfill Lemma 3.12, each term in the parenthesis in the ζ-function (Definition 3.5) must have a coefficient of 1 or -1.

- To fulfill Lemma 3.13, the ζ-function (Definition 3.5) and the norm in the denominator of the Rayleigh quotient (Definition 3.8) must have the same power. Additionally, the summation of coefficients of the terms in the parenthesis in the ζ-function (Definition 3.5) must be 0. Combining with the last condition, it requires the number of plus and minus terms to be equal.

- To fulfill Lemma 3.15, the number of minus terms in the parenthesis in the ζ-function (Definition 3.5) must be odd, so that $\langle -A, v^{\otimes m}\rangle$ can be canceled by $\langle L, v^{\otimes m}\rangle$ in those entries without repeating indices.

- To fulfill Lemma 3.11, the Rayleigh quotient must achieve minimum with $v = \mathbf{1}$.

At this point it should be clear that we cannot simply use some arbitrary definition from the prior literature. Our hypergraph definitions are carefully constructed and chosen, to fulfill all the necessary conditions above.

Remark 3.18. Regarding the choice of order m, note that if m is odd, the task of label recovery will not be feasible in the current formulation, since the Rayleigh quotient will be unbounded below (instead of a minimum 0). From Remark 3.17 it can be noticed that if we keep the current definition of ζ-functions, the only possible m's are $2, 6, 10, 14, \ldots$ The definition of ζ-functions can be generalized for other even orders, such that for every m-tuple $\mathcal{I}$, we take the average of many m-power terms by iterating through the permutation of all possible signs. For clarity of exposition we keep the current simpler definition, and focus on the case of $m = 2, 6, 10, 14, \ldots$

4. Exact Recovery of True Labels

In this section, we present an inference algorithm, which recovers the underlying true labels in our model. To do so, we take a two stage approach. In the first stage we only utilize the hyperedge information observed from X, and this allows us to narrow down our solution space to two possible solutions. In the second stage, with the help of the node information observed from z, we are able to infer the correct labeling of the nodes.

4.1. Stage One: Inference up to Permutation

We start by considering the following combinatorial problem

$$\begin{aligned}
\underset{y}{\text{maximize}} \quad & \langle X, y^{\otimes m}\rangle \\
\text{subject to} \quad & y \in \{-1, +1\}^n.
\end{aligned} \tag{3}$$

The issue with this optimization formulation is that the problem is not convex, which makes the analysis hard and intractable. Instead, we consider the following relaxed version

$$\begin{aligned}
\underset{Y}{\text{maximize}} \quad & \langle X, Y\rangle \\
\text{subject to} \quad & Y \in \mathcal{S}_+^{*,n,m} \\
& -1 \leq Y_{\text{odd}} \leq 1, \forall \text{odd} \in \bar{\sigma}_2^{n,m} \\
& Y_{\text{even}} = 1, \forall \text{even} \in \sigma_2^{n,m}.
\end{aligned} \tag{4}$$

Alternatively, we represent the last two constraints as $-1 \leq Y_{\bar{\sigma}_2^{n,m}} \leq 1$, and $Y_{\sigma_2^{n,m}} = 1$. Recall that $\mathcal{S}_+^{*,n,m}$ is the convex cone of rank-one tensors. The motivation is that we are using a rank-one tensor Y instead of the outer product $y^{\otimes m}$, so that the problem becomes convex in the objective function and the constraints. We have $-1 \leq Y_{\bar{\sigma}_2^{n,m}} \leq 1$ because the product of y's is either -1 or $+1$, and we have $Y_{\sigma_2^{n,m}} = 1$ because if every y_i repeats an even number of times, we know the product must be $+1$.

It remains to prove correctness of program (4). We are interested in identifying the regime, in which (4) returns the

exact rank-one tensor solution $Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$. We now present our main theorem on inference. The proof can be found in Appendix B.

Theorem 4.1 (Inference from Hyperedge Observation). *For an m-order structured prediction model with underlying hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and hyperedge observation X, the rank-one tensor solution $Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$ can be recovered from the convex optimization program* (4) *with probability at least $1 - \epsilon_1(\phi_{\mathcal{G}}, n, p)$, where*

$$\epsilon_1(\phi_{\mathcal{G}}, n, p) = 2n^m \exp\left(-\frac{(1-2p)^2 \phi_{\mathcal{G}}^{2m}}{8n^m \cdot \max(|\mathcal{E}|, n^{m-1})}\right) + \frac{16(1-p)|\mathcal{E}|}{(1-2p)^2 \phi_{\mathcal{G}}^{2m}}. \quad (5)$$

Remark 4.2. A natural question to ask is, under what topological and statistical conditions can we obtain a high probability guarantee from Theorem 4.1. An observation is that if the Cheeger constant of the underlying hypergraph is large, or the noise level is small, recovery is more likely to succeed. In Section 5, we analyze at some interesting classes of hypergraphs with good expansion property (large Cheeger constant), so that high probability recovery can be guaranteed.

4.2. Stage Two: Exact Inference

In the previous section, we established the high probability inference guarantee for the rank-one tensor solution $Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$. By taking a factorization step, we know either y^* or $-y^*$ is the correct label vector. In this section, our goal is to decide the correct labeling using the node observation z.

Theorem 4.3. *Let $y \in \{y^*, -y^*\}$. The correct label vector y^* can be recovered from program*

$$z^\top y^* = \max_{y \in \{y^*, -y^*\}} z^\top y. \quad (6)$$

with probability at least $1 - \epsilon_2(n, q)$, where

$$\epsilon_2(n, q) = e^{-(1-2q)^2 n/2}. \quad (7)$$

Proof of Theorem 4.3. Applying Hoeffding's inequality, we obtain that

$$\mathbb{P}\left\{z^\top y^* \leq -z^\top y^*\right\} = \mathbb{P}\left\{z^\top y^* \leq 0\right\}$$
$$\leq e^{-(1-2q)^2 n/2}.$$

$\square$

Corollary 4.4. *Combining the results in Theorem 4.1 and 4.3, we obtain that exact inference of the correct label vector $y = y^*$ can be achieved with probability at least*

$$1 - \epsilon_1(\phi_{\mathcal{G}}, n, p) - \epsilon_2(n, q). \quad (8)$$

Proof of Corollary 4.4. Apply a union bound to Theorem 4.1 and 4.3. $\square$

Remark 4.5. Given $p, q < 0.5$, we observe that as long as $n \to \infty$, we know $\epsilon_2(n, q) \to 0$ (an exponential decay). As a result, whether one can obtain a high probability guarantee in the shape of $1 - O(n^{-1})$ depends only on the order of $\phi_{\mathcal{G}}$, that is, the topological structure of the underlying hypergraph. In Section 5, we investigate hypergraphs with good expansion properties, that allow us to achieve a high probability guarantee of exact inference.

5. Examples of Hypergraphs with Good Expansion Property

In this section, we consider some example classes of hypergraphs with good expansion property, and demonstrate that they lead to high probability guarantees in our exact inference algorithm.

We first analyze complete hypergraphs.

Definition 5.1 (Complete Hypergraphs). An m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is complete, if for every m-vertex tuple $e = (i_1, \ldots, i_m)$, we have $e \in \mathcal{E}$.

Proposition 5.2 (Expansion Property of Complete Hypergraphs). *For any complete hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we have*

$$\phi_{\mathcal{G}} = \frac{N}{2}.$$

Proof. Recall that $\mathcal{H}$ is the set of induced hypervertices. By definition 3.3, with any $S \subset \mathcal{H}$ with $0 < |S| \leq N/2$, we have

$$\phi_S = \frac{|S| \cdot |\mathcal{H} \setminus S|}{|S|} = |\mathcal{H} \setminus S|.$$

Taking a minimum as in definition 3.4, we obtain $\phi_{\mathcal{G}} = \frac{N}{2}$. $\square$

Corollary 5.3 (Exact Inference in Complete Hypergraphs). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a complete hypergraph. Assume $p, q < 0.5$. Then exact inference can be achieved using the proposed two-stage approach with probability tending to 1 with a sufficiently large n.*

Proof. Note that for complete hypergraphs, we have

$$\phi_{\mathcal{G}} = \frac{N}{2} = \frac{1}{2}\binom{n}{m/2} \geq \frac{1}{2}\left(\frac{n}{m/2}\right)^{m/2} = \frac{2^{m/2-1}}{m^{m/2}} \cdot n^{m/2}.$$

Substituting $\phi_{\mathcal{G}}$ into (8), as long as n is greater than some constant c_0, we have

$$1 - \epsilon_1(\phi_{\mathcal{G}}, n, p) - \epsilon_2(n, q) \geq 1 - O(n^{-1}) \,.$$

$\square$

Next, we focus on the case of regular expanders.

Definition 5.4 (d-regular Expander). An m-uniform hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a d-regular expander hypergraph with constant $c > 0$, if for any induced hypervertex set $S \subset \mathcal{H}$ with $0 < |S| \leq N/2$, we have

$$|\partial S| \geq c \cdot d \cdot |S| \,.$$

Proposition 5.5 (Expander Hypergraphs). *For any d-regular expander hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with constant $c > 0$, we have*

$$\phi_{\mathcal{G}} = cd \,.$$

Proof. By definition 3.3, with any $S \subset \mathcal{H}$ with $0 < |S| \leq N/2$, we have

$$\phi_S \geq \frac{c \cdot d \cdot |S|}{|S|} = cd \,.$$

Taking a minimum as in definition 3.4, we obtain $\phi_{\mathcal{G}} = cd$. $\square$

Corollary 5.6 (Exact Inference in Expander Hypergraphs). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a d-regular expander hypergraph with constant $c > 0$. Assume $p, q < 0.5$. Then exact inference can be achieved using the proposed two-stage approach with probability tending to 1 with a sufficiently large n, if*

$$d \in \Omega(n \cdot (\log n)^{1/2m}) \,.$$

Proof. First note that the exponential term in (5) is the dominating factor. Substituting $\phi_{\mathcal{G}} = cd$ into the first term of (5), we want to ensure

$$2n^m \exp\left(-\frac{(1 - 2p)^2 (cd)^{2m}}{8n^m \cdot \max(|\mathcal{E}|, n^{m-1})}\right) \leq c_0 n^{-1} \,,$$

for some constant $c_0 > 0$. Note that $|\mathcal{E}| \in O(n^m)$. As a result, a sufficient condition for d is $d \in \Omega(n \cdot (\log n)^{1/2m})$. Substituting d into the other term also fulfills the $O(n^{-1})$ bound. $\square$

5.1. Simulation Results

We test the proposed method on a hypergraph with order $m = 6$; see Figure 2. We fix the number of nodes $n = 10$. We focus on the Stage One, and check how many labels can be recovered up to permutation of the signs. We implement a tensor projected gradient descent solver motivated by Ke

& Honorio (2022a); Han (2013). For each setting we run 20 iterations. Our results suggest that if the noise level p is small and the hypergraph Cheeger constant $\phi_{\mathcal{G}}$ is large, the proposed algorithm performs well and recovers the underlying group structure. This matches our theoretic findings in Theorem 4.1.
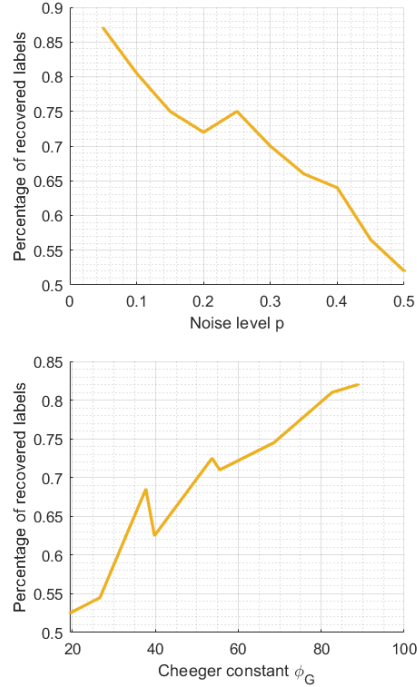

Figure 2. Simulations with different noise levels p (top) and Cheeger constant $\phi_{\mathcal{G}}$ (down). Our results suggest that if the noise level p is small and the hypergraph Cheeger constant $\phi_{\mathcal{G}}$ is large, the proposed algorithm performs well and recovers the underlying group structure.

Additionally, we implement a local-search based, hypergraph label propagation algorithm as a relevant baseline approximation method (Raghavan et al., 2007). We use the same setting with $n = 10$ and $m = 6$. To initialize the propagation, we make the label of five nodes to be visible to the algorithm. The algorithm proceeds as follows: for every unvisited node i, the algorithm will check its interaction with all known nodes repetitively to obtain enough samples (we set the number of repetitions to be 1001 in our script). For example, suppose i_1 through i_k are visited nodes, then the algorithm will repetitively check $X(i, i_1, i_2, i_3, i_4, i_5), X(i, i_1, i_2, i_3, i_4, i_6), \ldots$ in a stochastic fashion to collect evidence. After that, if the sum of these samples is positive, the label of i is assigned to be $+1$, and otherwise -1. The algorithm then marks the node as visited and continues, until all nodes are visited. We ran the experiment with $p = 0.1$ for 20 trials. The average accuracy of the label propagation algorithm above is 0.75 in total, and 0.5

on the initially unknown parts. In particular, only 1 out of 20 trials perfect recovery is achieved. This performs much worse than our results (perfect recovery in 80% trials).

5.2. Simulation on Real-world Datasets

We run the first stage of the proposed algorithm on a fourth-order hypergraph generated from the real-world dataset *email-Eu-core* (Yin et al., 2017; Leskovec et al., 2007). We pick the largest two communities in the network. The generated hypergraph has 200 nodes in total (91 and 109 in each group). We run our algorithm with a tensor projected gradient descent solver on the hypergraph, and exact inference of the 200 labels was achieved perfectly up to permutation of the two groups.

6. Concluding Remarks

We analyzed the problem of high-order label recovery in this paper. To achieve this, we provided a novel class of hypergraph structural properties, and derived a high-order version Cheeger-type inequality. We also proposed a two-stage algorithm to solve the high-order structured prediction problem, and provided KKT analysis for its optimality.

We gave the first results with theoretical guarantees of inference in hypergraphs. For some context, (1) is the structured prediction inference problem in graph cases. This inference setting has been studied by Globerson et al. (2015); Foster et al. (2018); Bello & Honorio (2019), among others. These works have focused on binary label cases. And these works studied various classes of graphs, including grid lattices, graphs allowing tree decompositions, and expander graphs. (2) is the structured prediction inference problem generalized to hypergraph cases. If we set $m = 2$, (2) reduces to (1). We focused on hypergraph and binary label cases. Our work studied a class of hypergraphs: those with good expansion property, or hyper-expanders.

A limitation in our setting is that the observation of each hyperedge is binary: either $+1$ or -1. As a future direction, it can be interesting to think about more general cases, in which the observations are weighted (that is, real values instead of 1's).

References

Bello, K. and Honorio, J. Exact inference in structured prediction. *Advances in Neural Information Processing Systems*, 32, 2019.

Boykov, Y. and Veksler, O. Graph cuts in vision and graphics: Theories and applications. In *Handbook of mathematical models in computer vision*, pp. 79–96. Springer, 2006.

Chan, T.-H. H., Louis, A., Tang, Z. G., and Zhang, C. Spectral properties of hypergraph laplacian and approximation algorithms. *Journal of the ACM (JACM)*, 65(3):1–48, 2018.

Chandrasekaran, V., Srebro, N., and Harsha, P. Complexity of inference in graphical models. *arXiv preprint arXiv:1206.3240*, 2012.

Chang, J., Chen, Y., Qi, L., and Yan, H. Hypergraph clustering using a new laplacian tensor with applications in image processing. *SIAM Journal on Imaging Sciences*, 13(3):1157–1178, 2020.

Chen, Y., Qi, L., and Zhang, X. The fiedler vector of a laplacian tensor for hypergraph partitioning. *SIAM Journal on Scientific Computing*, 39(6):A2508–A2537, 2017.

Fix, A., Gruber, A., Boros, E., and Zabih, R. A hypergraph-based reduction for higher-order binary markov random fields. *IEEE transactions on pattern analysis and machine intelligence*, 37(7):1387–1395, 2014.

Foster, D., Sridharan, K., and Reichman, D. Inference in sparse graphs with pairwise measurements and side information. In *International Conference on Artificial Intelligence and Statistics*, pp. 1810–1818. PMLR, 2018.

Gallagher, A. C., Batra, D., and Parikh, D. Inference for order reduction in markov random fields. In *CVPR 2011*, pp. 1857–1864. IEEE, 2011.

Globerson, A., Roughgarden, T., Sontag, D., and Yildirim, C. How hard is inference for structured prediction? In *International Conference on Machine Learning*, pp. 2181–2190. PMLR, 2015.

Han, L. An unconstrained optimization approach for finding real eigenvalues of even order symmetric tensors. *Numerical Algebra, Control and Optimization*, 3(3):583–599, 2013.

Ke, C. and Honorio, J. Information-theoretic limits for community detection in network models. *Advances in Neural Information Processing Systems*, 31, 2018.

Ke, C. and Honorio, J. Exact inference with latent variables in an arbitrary domain. *arXiv preprint arXiv:1902.03099*, 2019.

Ke, C. and Honorio, J. Exact partitioning of high-order models with a novel convex tensor cone relaxation. *Journal of Machine Learning Research*, 23(284):1–28, 2022a.

Ke, C. and Honorio, J. Federated myopic community detection with one-shot communication. In *International Conference on Artificial Intelligence and Statistics*, pp. 4937–4954. PMLR, 2022b.

Kelley, S., Goldberg, M., Magdon-Ismail, M., Mertsalov, K., and Wallace, A. Defining and discovering communities in social networks. In *Handbook of Optimization in Complex Networks*, pp. 139–168. Springer, 2012.

Komodakis, N. and Paragios, N. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2985–2992. IEEE, 2009.

Leskovec, J., Kleinberg, J., and Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1 (1):2–es, 2007.

Li, P. and Milenkovic, O. Submodular hypergraphs: p-laplacians, cheeger inequalities and spectral clustering. In *International Conference on Machine Learning*, pp. 3014–3023. PMLR, 2018.

Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H. Co-authorship networks in the digital library research community. *Information processing & management*, 41 (6):1462–1480, 2005.

Mulas, R. A cheeger cut for uniform hypergraphs. *Graphs and Combinatorics*, 37(6):2265–2286, 2021.

Nowozin, S., Lampert, C. H., et al. Structured learning and prediction in computer vision. *Foundations and Trends® in Computer Graphics and Vision*, 6(3–4):185–365, 2011.

Nugent, R. and Meila, M. An overview of clustering applied to molecular biology. *Statistical methods in molecular biology*, pp. 369–404, 2010.

Raghavan, U. N., Albert, R., and Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

Schraudolph, N. and Kamenetsky, D. Efficient exact inference in planar ising models. *Advances in Neural Information Processing Systems*, 21, 2008.

Weiss, D. and Taskar, B. Structured prediction cascades. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 916–923. JMLR Workshop and Conference Proceedings, 2010.

Yin, H., Benson, A. R., Leskovec, J., and Gleich, D. F. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 555–564, 2017.

Yoshida, Y. Cheeger inequalities for submodular transformations. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2582–2601. SIAM, 2019.

A. Proofs of Hypergraph Structural Properties and Cheeger-type Inequality

For clarity of presentation, in the following proofs, we use

$$RN_L(v) := \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \zeta(v_{i_1},\ldots,v_{i_m})$$

to denote the numerator part of the Rayleigh quotient of Laplacian L, and

$$RD_L(v) := \|v\|^m$$

to denote the denominator part.

Proof of Lemma 3.14. Regarding the numerator, we have

$$
\begin{aligned}
RN_L(v \circ y + \delta\mathbf{1}) &= \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \left[(v_{i_1}y_{i_1} + \delta + \cdots + v_{i_{m/2}}y_{i_{m/2}} + \delta - v_{i_{m/2+1}}y_{i_{m/2+1}} - \delta - \cdots - v_{i_m}y_{i_m} - \delta)^m + \ldots \right] \\
&= \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \left[(v_{i_1}y_{i_1} + \cdots + v_{i_{m/2}}y_{i_{m/2}} - v_{i_{m/2+1}}y_{i_{m/2+1}} - \cdots - v_{i_m}y_{i_m})^m + \ldots \right] \\
&= RN_{L_y}(v).
\end{aligned}
$$

Regarding the denominator, we have

$$
\begin{aligned}
RD_L(v \circ y + \delta\mathbf{1}) &= \|v \circ y + \delta\mathbf{1}\|^m \\
&= \left(\sum_i (v_i y_i + \delta)^2 \right)^{m/2} \\
&= \left(\sum_i v_i^2 y_i^2 + n\delta^2 + 2\delta \sum_i v_i y_i \right)^{m/2} \\
&= \left(\sum_i v_i^2 + n\delta^2 \right)^{m/2} \\
&\geq \left(\sum_i v_i^2 \right)^{m/2} \\
&= \|v\|^m \\
&= RD_{L_y}(v).
\end{aligned}
$$

$\square$

Proof of Lemma 3.15. Our goal is to construct a tensor D fulfilling $\langle D - A, v^{\otimes m} \rangle = \langle L, v^{\otimes m} \rangle$ for any vector $v \in \mathbb{R}^n$. In other words, we require

$$\sum_{i_1,\ldots,i_m} D_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m} - \sum_{i_1,\ldots,i_m} A_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m} = \sum_{i_1,\ldots,i_m} L_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m}.$$

A key observation is that on the left-hand side, D and A control different entries: $D_{i_1,\ldots,i_m}$ is equal to 0 in those entries without repeating indices, while $A_{i_1,\ldots,i_m}$ is equal to 0 in those entries with repeating indices. Recall that the union $\sigma_2^{n,m} \cup \bar{\sigma}_2^{n,m}$ contains all entry indices without repeating indices. We can rewrite the equation above as

$$\sum_{(i_1,\ldots,i_m)\in\sigma_2^{n,m}\cup\bar{\sigma}_2^{n,m}} D_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m} - \sum_{(i_1,\ldots,i_m)\notin\sigma_2^{n,m}\cup\bar{\sigma}_2^{n,m}} A_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m} = \sum_{i_1,\ldots,i_m} L_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m}.$$

Recall that

$$\langle L, v^{\otimes m}\rangle = \frac{1}{m!\binom{m}{m/2}} \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \zeta(v_{i_1},\ldots,v_{i_m})$$

$$= \frac{1}{m!\binom{m}{m/2}} \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \left[(v_{i_1} + v_{i_2} + \cdots + v_{i_{m/2}} - v_{i_{m/2+1}} - \cdots - v_{i_m})^m + \ldots \right] . \tag{9}$$

The next observation is that, by expanding the m-power terms in the bracket, for each summand $(i_1,\ldots,i_m)$ we will obtain a sequence of monomials consisting of v_{i_1} through v_{i_m}. For example, this includes $C_0 \cdot v_{i_1} v_{i_2} \ldots v_{i_{m-1}} v_{i_m}$, $C_1 \cdot v_{i_1}^2 v_{i_2} \ldots v_{i_{m-1}}$, etc., where C's are coefficients. Note that all these monomials have an order of m.

We now analyze the monomial terms above by grouping the monomials with the same power pattern. Formally, we group all terms in the shape of $C \cdot v_{i_1}^{d_1} v_{i_2}^{d_2} \ldots v_{i_q}^{d_q}$ together, where $d_1 \geq d_2 \geq \cdots \geq d_q \geq 1$, $d_1 + \cdots + d_q = m$. All permutation of subscripts are enumerated in the parenthesis. We use Q to denote the number of terms inside this group. Here we provide some examples of monomial groups:

- $C \cdot (v_{i_1}^m + v_{i_2}^m + \cdots + v_{i_m}^m)$, with pattern $d_1 = m$. In this case $Q = m$.

- $m = 6$, $C \cdot (v_{i_1}^5 v_{i_2}^1 + v_{i_1}^5 v_{i_3}^1 + \cdots + v_{i_6}^5 v_{i_5}^1)$, with pattern $d_1 = 5$, $d_2 = 1$. In this case $Q = 6 \cdot 5 = 30$.

- $m = 6$, $C \cdot (v_{i_1}^3 v_{i_2}^2 v_{i_3}^1 + v_{i_1}^3 v_{i_2}^2 v_{i_4}^1 + \cdots + v_{i_6}^3 v_{i_5}^2 v_{i_4}^1)$, with pattern $d_1 = 3$, $d_2 = 2$, $d_3 = 1$. In this case $Q = 6 \cdot 5 \cdot 4 = 120$.

Next we discuss the power pattern in each group.

If the power pattern is $d_1 = \cdots = d_m = 1$ (i.e., all 1-power components), we know the coefficient C is equal to $-\frac{1}{m!\binom{m}{m/2}}$ by counting. Thus, these terms with all 1-power components are already balanced by $-\sum_{(i_1,\ldots,i_m)\notin\sigma_2^{n,m}\cup\bar\sigma_2^{n,m}} A_{i_1,\ldots,i_m} v_{i_1} \ldots v_{i_m}$ on the left-hand side. This also shows the necessity of introducing the factor $\frac{1}{m!\binom{m}{m/2}}$ in the definition of hypergraph Laplacians: it ensures that the term $v_{i_1} v_{i_2} \ldots v_{i_m}$ in the expanded form of $\langle L, v^{\otimes m}\rangle$ has a coefficient of 1.

For groups with other power patterns ($d_1 \geq \ldots d_q \geq 1$, $q < m$), we balance them by setting D entries with indices in the permutation of $(i_1,\ldots,i_1,i_2,\ldots,i_2,\cdots\cdots,i_q,\ldots,i_q)$, in which i_1 repeats d_1 times, i_2 repeats d_2 times, etc. We set the value of these D entries to: $CQ \cdot \sum_{i_{q+1},\ldots,i_m} A_{i_1,\ldots,i_m} = CQ \cdot \sum_{i_{q+1},\ldots,i_m} \mathbb{1}[(i_1,\ldots,i_m) \in \mathcal{E}]$. In other words, for D entries containing index i_1 through i_q, we set them to be equal to those A entries containing index $\{i_1,\ldots,i_m\}\setminus\{i_1,\ldots,i_q\}$. This can be illustrated using the same examples above:

- Set $D_{i_1,\ldots,i_1} = CQ\sum_{i_2,\ldots,i_m} A_{i_1,\ldots,i_m}$.

- Set $D_{i_1,i_1,i_1,i_1,i_1,i_2} = CQ\sum_{i_3,i_4,i_5,i_6} A_{i_1,\ldots,i_m}$, and the same for all symmetric entries in $\sigma(i_1,i_1,i_1,i_1,i_1,i_2)$.

- Set $D_{i_1,i_1,i_1,i_2,i_2,i_3} = CQ\sum_{i_4,i_5,i_6} A_{i_1,\ldots,i_m}$, and the same for all symmetric entries in $\sigma(i_1,i_1,i_1,i_1,i_1,i_2)$.

After the procedure is done, we have $\langle L, v^{\otimes m}\rangle = \langle D - A, v^{\otimes m}\rangle$, and $D_{i_1,i_2,\ldots,i_m} = 0$ if $i_1, i_2, \ldots, i_m$ are all different. $\qquad\square$

Proof of Theorem 3.16. Suppose v is the eigenvector associated with the second smallest eigenvalue $\lambda_2(L)$. By Lemma 3.12, we assume $\|v\| = 1$ without loss of generality. We also sort the entries of v, i.e., $v_1 \leq \cdots \leq v_n$.

Our first step is to set up a helper vector u based on the second smallest eigenvector v. Recall that $\mathcal{H}$ is the set of induced hypervertices (see Definition 3.1). For any hypervertex $h_j = \{i_1,\ldots,i_{m/2}\} \in \mathcal{H}$, based on the second eigenvector v, we define the v-value of h_j by

$$v(h_j) := \frac{1}{m/2}\left(v_{i_1} + \cdots + v_{i_{m/2}}\right) .$$

This allows us to sort all the hypervertices by their v-value, such that the hypervertex indices fulfill

$$v(h_1) \leq v(h_2) \leq \cdots \leq v(h_N),$$

and recall that $N := |\mathcal{H}|$. We use M to denote the smallest integer fulfilling $M \geq \frac{1}{2}|\mathcal{H}|$, and we introduce a shifting operation by defining $u := v - v(h_M) \cdot \mathbf{1}$. We further find a constant $c_u > 0$, and we scale u by multiplying u with c_u so that it fulfills $(u_1 + \cdots + u_{m/2})^2 + (u_{n-m/2+1} + \cdots + u_n)^2 = 1$. Note that u fulfills the following properties:

- $(u_1 + \cdots + u_{m/2})^2 + (u_{n-m/2+1} + \cdots + u_n)^2 = 1$.

- $u_1 \leq \cdots \leq u_n$.

- $u_M = 0$.

- $R_L(v) \geq R_L(u)$, by Lemma 3.12 and 3.13.

The second step of our proof is to construct a random set of hypervertices S_t. To do so, we define t to be a random variable on the support $[u_1 + \cdots + u_{m/2}, u_{n-m/2+1} + \cdots + u_n]$, with probability density function $f(t) = 2|t|$. It can be verified that t is a valid random variable, because

$$\int_{u_1 + \cdots + u_{m/2}}^{u_{n-m/2+1} + \cdots + u_n} 2|t| = (u_1 + \cdots + u_{m/2})^2 + (u_{n-m/2+1} + \cdots + u_n)^2 = 1\,.$$

Based on t, we can construct a random set of hypervertices S_t as follows

$$S_t := \{h_j \mid h_j = \{i_1, \ldots, i_{m/2}\}, u_{i_1} + \cdots + u_{i_{m/2}} \leq t\}\,.$$

Here we consider the size of S_t in the average case. We have

$$\mathbb{E}_t[|S_t|] = \sum_j \mathbb{P}_t\{u_{h_j} \leq t\}\,,$$

and

$$\mathbb{E}_t[|\mathcal{H} \setminus S_t|] = \sum_j \mathbb{P}_t\{u_{h_j} > t\}\,.$$

Combining the two leads to

$$\mathbb{E}_t[\min(|S_t|, |\mathcal{H} \setminus S_t|)] = \sum_j \mathbb{P}_t\{u_{h_j} \leq t < 0\} + \sum_j \mathbb{P}_t\{u_{h_j} > t \geq 0\}$$
$$= \sum_j u_{h_j}^2\,.$$

We also consider the boundary set ∂S_t. By Definition 3.2, a hyperedge $e = h_1 \cup h_2$ belongs to ∂S_t, if $h_1 \in S_t$, $h_2 \notin S_t$, and $h_1 \cap h_2 = \emptyset$. Define the shorthand notation $u_h := \sum_{i \in h} u_i$, and assume $u_{h_1} \leq u_{h_2}$. Then

$$\mathbb{P}\{e \in \partial S_t\} = \mathbb{P}\{u_{h_1} \leq t \leq u_{h_2}\} \leq |u_{h_1} - u_{h_2}|(|u_{h_1}| + |u_{h_2}|)\,.$$

Finally, we analyze the expectation of $|\partial S_t|$. It follows that

$$
\begin{aligned}
\mathbb{E}_t\left[|\partial S_t|\right] &= \sum_{e \in \mathcal{E}, e = h_1 \cup h_2} \mathbb{E}_t\left[\mathbb{1}[e \in \partial S_t]\right] \\
&= \sum_{e \in \mathcal{E}, e = h_1 \cup h_2} \mathbb{P}_t\left\{e \in \partial S_t\right\} \\
&\leq \sum_{e \in \mathcal{E}, e = h_1 \cup h_2} |u_{h_1} - u_{h_2}| \left(|u_{h_1}| + |u_{h_2}|\right) \\
&\overset{(a)}{\leq} \left(\sum_{e \in \mathcal{E}, e = h_1 \cup h_2} (u_{h_1} - u_{h_2})^m\right)^{\frac{1}{m}} \left(\sum_{e \in \mathcal{E}, e = h_1 \cup h_2} (|u_{h_1}| + |u_{h_2}|)^{\frac{m}{m-1}}\right)^{\frac{m-1}{m}} \\
&= RN^{\frac{1}{m}} \cdot \left(\sum_{e \in \mathcal{E}, e = h_1 \cup h_2} (|u_{h_1}| + |u_{h_2}|)^{\frac{m}{m-1}}\right)^{\frac{m-1}{m}} \\
&= R_L(u)^{\frac{1}{m}} \cdot \|u\| \cdot \left(\sum_{e \in \mathcal{E}, e = h_1 \cup h_2} (|u_{h_1}| + |u_{h_2}|)^{\frac{m}{m-1}}\right)^{\frac{m-1}{m}} \\
&\leq R_L(u)^{\frac{1}{m}} \cdot \|u\| \cdot \left(\sum_{e \in \mathcal{E}, e = h_1 \cup h_2} (|u_{h_1}| + |u_{h_2}|)^2\right)^{\frac{m-1}{m}} \\
&\leq R_L(u)^{\frac{1}{m}} \cdot \mathbb{E}_t\left[\min(|S_t|, |\mathcal{H} \setminus S_t|)\right]^{\frac{m-1}{m}} \\
&\leq R_L(u)^{\frac{1}{m}} \cdot \mathbb{E}_t\left[\min(|S_t|, |\mathcal{H} \setminus S_t|)\right],
\end{aligned}
$$

where (a) follows from Holder's inequality. Rearranging the terms above leads to

$$
\mathbb{E}_t\left[R_L(u)^{\frac{1}{m}} \cdot \min(|S_t|, |\mathcal{H} \setminus S_t|) - |\partial S_t|\right] \geq 0.
$$

Thus there exists some t fulfilling

$$
R_L(u)^{\frac{1}{m}} \cdot \min(|S_t|, |\mathcal{H} \setminus S_t|) - |\partial S_t| \geq 0,
$$

or equivalently,

$$
R_L(u) \geq \left(\frac{|\partial S_t|}{\min(|S_t|, |V \setminus S_t|)}\right)^m.
$$

Plugging in $R_L(v) \geq R_L(u)$ on the left-hand side, and Definition 3.4 on the right-hand side, we obtain

$$
\lambda_2(L) = R_L(v) \geq R_L(u) \geq \left(\frac{|\partial S_t|}{\min(|S_t|, |V \setminus S_t|)}\right)^m \geq \phi_{\mathcal{G}}^m.
$$

$\square$

B. Proof of Exact Recovery of True Labels

Proof of Theorem 4.1. Our goal is to recover the true labels $Y^* := (\pm y^*)^{\otimes m}$ (up to flipping signs) using (4). For readers' convenience, here we restate the formulation:

$$
\begin{aligned}
\underset{Y}{\text{maximize}} \quad & \langle X, Y \rangle \\
\text{subject to} \quad & Y \in \mathcal{S}_+^{*,n,m} \\
& -1 \leq Y_{\bar{\sigma}_2^{n,m}} \leq 1 \\
& Y_{\sigma_2^{n,m}} = 1.
\end{aligned}
$$

The Lagrangian dual problem of (4) is

$$\underset{V,V^+,V^-,A}{\text{minimize}} \quad \sum_{(i_1,\ldots,i_m)\in\sigma_2^{n,m}} V_{i_1,\ldots,i_m} + \sum_{(i_1,\ldots,i_m)\in\bar{\sigma}_2^{n,m}} (V_{i_1,\ldots,i_m}^+ + V_{i_1,\ldots,i_m}^-)$$

subject to $\quad V_{i_1,\ldots,i_m} = 0$, if $i_1,\ldots,i_m$ are all different

$$V_{\bar{\sigma}_2^{n,m}} = V_{\bar{\sigma}_2^{n,m}}^+ - V_{\bar{\sigma}_2^{n,m}}^-$$

$$V_{\bar{\sigma}_2^{n,m}}^+ \geq 0$$

$$V_{\bar{\sigma}_2^{n,m}}^- \geq 0$$

$$-X - A + \sum_{(i_1,\ldots,i_m)\in\sigma_2^{n,m}} V_{i_1,\ldots,i_m} \cdot E_{i_1,\ldots,i_m} + \sum_{(i_1,\ldots,i_m)\in\bar{\sigma}_2^{n,m}} (V_{i_1,\ldots,i_m}^+ - V_{i_1,\ldots,i_m}^-) \cdot E_{i_1,\ldots,i_m} = 0$$

$$A \in \mathcal{S}_+^{n,m},$$

where $E_{i_1,\ldots,i_m}$ is a tensor with 1 in entry $(i_1,\ldots,i_m)$, and 0 everywhere else. From the primal and dual problems, we obtain the following KKT conditions:

$$V - X - A = 0 \qquad\qquad\qquad\qquad \text{(Stationarity)}$$

$$Y_{\sigma_2^{n,m}} = 1$$

$$Y_{\bar{\sigma}_2^{n,m}} \leq 1$$

$$-Y_{\bar{\sigma}_2^{n,m}} \leq 1$$

$$Y \in \mathcal{S}_+^{*,n,m} \qquad\qquad\qquad\qquad \text{(Primal Feasibility)}$$

$$V_{i_1,\ldots,i_m} = 0$$

$$V_{\bar{\sigma}_2^{n,m}} = V_{\bar{\sigma}_2^{n,m}}^+ - V_{\bar{\sigma}_2^{n,m}}^-$$

$$V_{\bar{\sigma}_2^{n,m}}^+ \geq 0$$

$$V_{\bar{\sigma}_2^{n,m}}^- \geq 0$$

$$A \in \mathcal{S}_+^{n,m} \qquad\qquad\qquad\qquad \text{(Dual Feasibility)}$$

$$V_{\bar{\sigma}_2^{n,m}}^+ (Y_{\bar{\sigma}_2^{n,m}} - 1) = 0$$

$$V_{\bar{\sigma}_2^{n,m}}^- (Y_{\bar{\sigma}_2^{n,m}} + 1) = 0$$

$$\langle A, Y \rangle = 0. \qquad\qquad\qquad\qquad \text{(Complementary Slackness)}$$

Here we construct primal and dual variables to fulfill all KKT conditions above. For the primal variable we set $Y = Y^*$. For the dual variable we define $V = D_{y^*}$, where D_{y^*} is the high-order degree tensor constructed from the signed Laplacian L_{y^*}, with the procedure defined in Lemma 3.15. We then define $V_{\sigma_2^{n,m}} = (D_{y^*})_{\sigma_2^{n,m}}$, $V_{\bar{\sigma}_2^{n,m}}^+ = \max((D_{y^*})_{\bar{\sigma}_2^{n,m}}, 0)$, $V_{\bar{\sigma}_2^{n,m}}^- = \min((D_{y^*})_{\bar{\sigma}_2^{n,m}}, 0)$. From the stationarity condition, we set $A = L_{y^*} = D_{y^*} - X$.

At this point, our construction of primal and dual variables have fulfilled every KKT conditions above except the positive semidefinite condition $A = L_{y^*} \in \mathcal{S}_+^{n,m}$. From Lemma 3.11, we know y^* is an eigenvector of L_{y^*} with an eigenvalue of 0, or equivalently, $\langle L_{y^*}, Y^* \rangle = 0$. It remains to ensure for all orthogonal vectors, we have $\min_{u \perp y^*} R_{L_{y^*}}(u) \geq 0$. On top of that, we want to ensure the solution $Y = Y^*$ is unique. This further requires that

$$\min_{u \perp y^*} R_{L_{y^*}}(u) = \min_{u \perp y^*} \frac{\langle D_{y^*} - X, u^{\otimes m} \rangle}{\|u\|^m} > 0.$$

Without loss of generality, we fix $\|u\| = 1$ in the following discussion. We split the terms above into

$$\min_{u \perp y^*, \|u\|=1} \langle D_{y^*} - X, u^{\otimes m} \rangle \geq \min_{u \perp y^*, \|u\|=1} \langle D_{y^*} - \mathbb{E}[D_{y^*}], u^{\otimes m} \rangle \qquad\qquad (10)$$

$$+ \min_{u \perp y^*, \|u\|=1} \langle \mathbb{E}[X] - X, u^{\otimes m} \rangle \qquad\qquad (11)$$

$$+ \min_{u \perp y^*, \|u\|=1} \langle \mathbb{E}[D_{y^*} - X], u^{\otimes m} \rangle. \qquad\qquad (12)$$

First we bound the expectation term (12). Note that

$$
\begin{aligned}
\min_{u \perp y^*, \|u\|=1} \left\langle \mathbb{E}\left[D_{y^*} - X\right], u^{\otimes m}\right\rangle &= \min_{u \perp y^*, \|u\|=1} \left\langle \mathbb{E}\left[L_{y^*}\right], u^{\otimes m}\right\rangle \\
&= (1 - 2p) \cdot \min_{u \perp y^*, \|u\|=1} \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \zeta(u_{i_1} y_{i_1}^*, \ldots, u_{i_m} y_{i_m}^*) \\
&= (1 - 2p) \cdot R_{L_{y^*}}(u) \\
&\geq (1 - 2p) \cdot R_L(u \circ y^* + \delta\mathbf{1}) \\
&\geq (1 - 2p) \cdot \phi_{\mathcal{G}}^m ,
\end{aligned}
$$

where the first inequality follows from Lemma 3.14, and the second inequality follows from Theorem 3.16.

Next we bound (11) using concentration inequalities. We have

$$
\begin{aligned}
\mathbb{P}\left\{-\lambda_{\max}(\mathbb{E}\left[X\right] - X) \leq -t\right\} &\leq \mathbb{P}\left\{-\|\mathbb{E}\left[X\right] - X\|_{\mathrm{F}}^2 \leq -t^2\right\} \\
&\leq \frac{1}{t^2} \cdot \mathbb{E}\left[\|\mathbb{E}\left[X\right] - X\|_{\mathrm{F}}^2\right] \\
&= \frac{1}{t^2} \cdot 4p(1-p)\,|\mathcal{E}| .
\end{aligned}
$$

Setting $t = \frac{1-2p}{2}\phi_{\mathcal{G}}^m$ leads to

$$
\mathbb{P}\left\{-\lambda_{\max}(\mathbb{E}\left[X\right] - X) \leq -\frac{1-2p}{2}\phi_{\mathcal{G}}^m\right\} \leq \frac{16p(1-p)\,|\mathcal{E}|}{(1-2p)^2\phi_{\mathcal{G}}^{2m}} . \tag{13}
$$

Finally we bound (10). Using Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
-\left|\left\langle \mathbb{E}\left[D_{y^*}\right] - D_{y^*}, u^{\otimes m}\right\rangle\right| &\geq -\|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\| \cdot \|\mathrm{vec}\left(u^{\otimes m}\right)\| \\
&= -\|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\| \cdot \|u\|^m \\
&= -\|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\| \\
&\geq -n^{m/2} \cdot \|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\|_{\infty} ,
\end{aligned}
$$

where $\mathrm{vec}\left(\cdot\right)$ is the vectorization operator. We then use Hoeffding's inequality for every entry. By the construction procedure defined in the proof of Lemma 3.15, every entry of D_{y^*} is the summation of at most $\max(|\mathcal{E}|, n^{m-1})$ Rademacher random variables. We obtain

$$
\begin{aligned}
\mathbb{P}\left\{\mathrm{vec}\left(\mathbb{E}\left[D_{y^*,i_1,\ldots,i_m}\right] - D_{y^*,i_1,\ldots,i_m}\right) \geq t\right\} &\leq 2\exp\left(-\frac{2t^2}{2^2\max(|\mathcal{E}|, n^{m-1})}\right) \\
&\leq 2\exp\left(-\frac{t^2}{2\max(|\mathcal{E}|, n^{m-1})}\right) .
\end{aligned}
$$

By a union bound, we obtain

$$
\begin{aligned}
\mathbb{P}\left\{\|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\|_{\infty} \geq t\right\} &\leq 2\exp\left(-\frac{2t^2}{2^2\max(|\mathcal{E}|, n^{m-1})}\right) \\
&\leq 2n^m\exp\left(-\frac{t^2}{2\max(|\mathcal{E}|, n^{m-1})}\right) .
\end{aligned}
$$

Setting $t = \frac{(1-2p)\phi_{\mathcal{G}}^m}{2n^{m/2}}$ leads to

$$
\mathbb{P}\left\{-\left|\left\langle \mathbb{E}\left[D_{y^*}\right] - D_{y^*}, u^{\otimes m}\right\rangle\right| \leq -\frac{1-2p}{2}\phi_{\mathcal{G}}^m\right\} \leq 2n^m\exp\left(-\frac{(1-2p)^2\phi_{\mathcal{G}}^{2m}}{8n^m\max(|\mathcal{E}|, n^{m-1})}\right) . \tag{14}
$$

Combining the results of (13) and (14) with a union bound completes the proof. $\qquad\square$