Exact Partitioning of High-order Models with a Novel Convex Tensor Cone Relaxation

Chuyang Ke   CKE@PURDUE.EDU
Department of Computer Science
Purdue University
West Lafayette, IN 47907, USA

Jean Honorio   JHONORIO@PURDUE.EDU
Department of Computer Science
Purdue University
West Lafayette, IN 47907, USA

Abstract

In this paper we propose an algorithm for exact partitioning of high-order models. We define a general class of m-degree Homogeneous Polynomial Models, which subsumes several examples motivated from prior literature. Exact partitioning can be formulated as a tensor optimization problem. We relax this high-order combinatorial problem to a convex conic form problem. To this end, we carefully define the Carathéodory symmetric tensor cone, and show its convexity, and the convexity of its dual cone. This allows us to construct a primal-dual certificate to show that the solution of the convex relaxation is correct (equal to the unobserved true group assignment) and to analyze the statistical upper bound of exact partitioning.

Keywords: Tensor Cone, Hypergraph, Exact Partitioning

1. Introduction

Partitioning and clustering algorithms have been favored by researchers from various fields, including machine learning, data mining, molecular biology, and network analysis (Xu and Tian, 2015; Cai et al., 2015; Nugent and Meila, 2010; Berkhin, 2006). Although there is no identical criterion, partitioning algorithms often aim to find a group labeling for a set of entities in a dataset equipped with some *pairwise* metric. In general, the goal is to maximize in-group *affinity*, that is, the entities from the same group are more similar to those from different groups (Liu et al., 2010; Huang et al., 2012). However in many complex real-world networks, pairwise metrics are not expressive enough to capture all the information. One common assumption is that entities interact in groups instead of pairs. For instance, in a co-authorship network, researchers collaborate in small groups and publish papers (Liu et al., 2005). Another example is the air traffic network (Rosvall et al., 2014), such that a flight may follow a triangular route A-B-C-A. In these scenarios, pairwise metrics are not sufficient to handle high-order relationships between entities. Thus it is important to develop a general high-order partitioning algorithm that can better characterize multi-entity interactions in complex networks.

Recent years witnessed a growing amount of literature on high-order problems, most of them investigating hypergraphs and related applications (Pal and Zhu, 2021; Papa and Markov, 2007; Agarwal et al., 2005; Gibson et al., 2000; Hagen and Kahng, 1992). A common approach used in hypergraph-related works, is to transform the hypergraph to a pairwise graph by embedding high-order interactions into pairwise affinities, and then apply traditional graph-based partitioning algorithms (Leordeanu and Sminchisescu, 2012; Zhou et al., 2007). Some works applied tensor methods to classic community detection problems with overlapping membership including the Stochastic Block Model (Anandkumar et al., 2013; Huang et al., 2015). The related problem of tensor clustering is also analyzed by some recent works through power iteration methods (Ke et al., 2019; Luo and Zhang, 2022).

In this paper we propose a novel high-order model class, namely *m-degree Homogeneous Polynomial Models (m-HPMs).* Our m-HPM class definition employs the use of homogeneous polynomials to carefully construct an m-order tensor, which captures the multi-entity affinities in underlying high-order networks. We also provide an *exact partitioning* algorithm with statistical guarantees, which aims at recovering the true underlying group structure. It is worth mentioning that in the case of second order ($m = 2$), the partitioning problem reduces to the Minimum Bisection problem, which is known to be NP-hard (Garey et al., 1976). We relax a high-order combinatorial problem to a convex *conic form problem*, and analyze the Karush–Kuhn–Tucker (KKT) conditions for the optimal solution. Conic form problems are a highly general class of convex optimization problems. For example, *semidefinite programming* is a special case of conic form programs, when the cone is the set of positive semidefinite matrices. We prove that as long as certain statistical conditions are fulfilled, exact partitioning in m-HPMs can be achieved.

Summary of Our Contributions. We provide a series of novel results in this paper:

- Our definition of m-HPMs is a contribution. We are providing the first general model class which characterizes multi-entity interactions in various high-order models. Our definition is highly general, subsumes a wide range of high-order models studied in prior literature, and is amenable to analysis. We show that several high-order problems, including high-order counting models, hypergraph cuts / cliques / volumes / conductance, and motif models, belong to the class of m-HPMs.

- We formulate exact partitioning as a high-order combinatorial optimization problem, and relax it to a convex conic form problem using carefully-defined novel tensor primal and dual cones.

- We construct a novel primal-dual certificate that leads to the optimal solution of the exact partitioning problem. KKT conditions guarantee our solution to be optimal, as long as the statistical conditions are satisfied. We furthermore characterize the statistical upper bound of exact partitioning by analyzing the tensor eigenvalues associated with the optimal solution.

2. Problem Setting and Notation

In this section, we introduce the notations that will be used in the paper. For any positive integer n, we use $[n]$ to denote the set $\{1, \ldots, n\}$. For clarity when dealing with a sequence

of objects, we use the superscript (i) to denote the i-th object in the sequence, and subscript j to denote the j-th entry. For example, for a sequence of vectors $\{x^{(i)}\}_{i\in[n]}$, $x_2^{(1)}$ represents the second entry of vector $x^{(1)}$. The notation $\otimes$ is used to denote outer product of vectors, for example, $x^{(1)} \otimes \ldots \otimes x^{(m)}$ is a tensor of order m, such that $(x^{(1)} \otimes \ldots \otimes x^{(m)})_{i_1,\ldots,i_m} = x_{i_1}^{(1)} \ldots x_{i_m}^{(m)}$. We use $\mathbf{1}$ to denote the all-one vector.

Let A be an m-th order n-dimensional real tensor, such that $A_{i_1,\ldots,i_m} \in \mathbb{R}$, where $i_j \in [n]$ for every $j \in [m]$. Throughout the paper we require m to be a positive even integer; see Section 4 for discussion.

A tensor is *symmetric* if it is invariant under any permutation of its indices, i.e., $A_{\sigma(i_1),\ldots,\sigma(i_m)} = A_{i_1,\ldots,i_m}$ for any permutation $\sigma : [m] \to [m]$. We denote the space of all m-th order n-dimensional symmetric tensors as $\mathcal{S}^{n,m} := \{A \mid A_{i_1,\ldots,i_m} \in \mathbb{R}, i_j \in [n], j \in [m], A \text{ is symmetric}\}$. Note that $\mathcal{S}^{n,m}$ is a vector space, with dimension (which in this case coincides with the maximum number of different entries) equal to $\dim \mathcal{S}^{n,m} = \binom{m+n-1}{m}$. We denote the constant

$$M := \dim \mathcal{S}^{n,m} + 1 = \binom{m+n-1}{m} + 1\,.$$

We use $\sigma_2^{n,m}$ to denote the set of m-tuples in the form 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 $i_j \in [n]$. We use $\bar{\sigma}_2^{n,m}$ to denote the complement set $\{(i_1,\ldots,i_m) \mid (i_1,\ldots,i_m) \notin \sigma_2^{n,m}\}$.

For symmetric tensors $A, B \in \mathcal{S}^{n,m}$, we define the *inner product* $\langle A, B \rangle$, the *tensor Frobenius norm* $\|A\|_{\mathrm{F}}$, and the *tensor trace* $\mathrm{tr}\,(A)$ respectively as $\langle A, B \rangle = \sum_{i_1,\ldots,i_m=1}^n A_{i_1,\ldots,i_m} B_{i_1,\ldots,i_m}$, $\|A\|_{\mathrm{F}} = \sqrt{\langle A, A \rangle}$, and $\mathrm{tr}\,(A) = \sum_{i=1}^n A_{i,\ldots,i}$.

For any cone $\mathcal{K}$, we use $\mathcal{K}^*$ to denote its *dual cone*.

For any vector $u \in \mathbb{R}^n$, we denote the corresponding m-th order *rank-one* tensor as $u^{\otimes m}$, where $(u^{\otimes m})_{i_1,\ldots,i_m} = u_{i_1} \ldots u_{i_m}$, and we denote the set of all m-th order n-dimensional rank-one tensors as $\mathcal{U}^{n,m} := \{u^{\otimes m} \mid u \in \mathbb{R}^n\}$.

For any tensor $A \in \mathcal{S}^{n,m}$, we call A a *positive semidefinite (PSD)* tensor, if for every $B \in \mathcal{U}^{n,m}$, $\langle A, B \rangle \geq 0$. Similarly, A is called *positive definite*, if for every $B \in \mathcal{U}^{n,m}$, $\langle A, B \rangle > 0$. We denote the set of all m-th order n-dimensional positive semidefinite tensors as $\mathcal{S}_+^{n,m} := \{A \mid A \in \mathcal{S}^{n,m}, \langle A, B \rangle \geq 0, \forall B \in \mathcal{U}^{n,m}\}$.

We also introduce the *Carathéodory symmetric tensor cone* $\mathcal{V}^{n,m}$, which is defined as $\mathcal{V}^{n,m} := \left\{\sum_{i=1}^M A^{(i)} \mid A^{(i)} \in \mathcal{U}^{n,m}, i \in [M]\right\}$. In particular, $\mathcal{S}_+^{n,m}$ and $\mathcal{V}^{n,m}$ are well-defined convex cones and are dual to each other (see Section 2.1).

We define the *maximum tensor eigenvalue* and the *minimum tensor eigenvalue* of A, by its variational characterization, as $\lambda_{\mathrm{tmax}}(A) = \max_{u \in \mathbb{R}^n, \|u\|=1} \langle A, u^{\otimes m} \rangle$, $\lambda_{\mathrm{tmin}}(A) = \min_{u \in \mathbb{R}^n, \|u\|=1} \langle A, u^{\otimes m} \rangle$, where $\|u\|$ is the Euclidean norm of vector u.

We now introduce the definition of m-degree Homogeneous Polynomial Models, or m-HPM for short.

Definition 1 (m-degree Homogeneous Polynomial Model) *For a high-order random model $\mathcal{M}$, let n be the number of entities (each of them belonging to either one of the two groups), $y^* \in \{+1, -1\}^n$ be the unobserved true group assignment, m be the order of the model, $p = (p_0, \ldots, p_m)$ be the coefficient parameter, σ^2 be the variance, B be the entrywise bound, and W be the random affinity tensor associated with the model. We say model $\mathcal{M}$ belongs to the class of m-HPM(n, p, σ^2, B), if $\mathcal{M}$ satisfies the following properties:*

(P1) *Expectation Decomposition:* $\mathbb{E}\left[W\right] = \sum_{k=0}^{m} p_k \sum_{\substack{z \in \{0,1\}^m \\ \mathbf{1}^\top z = k}} \bigotimes_{i=1}^{m} (z_i \mathbf{1} + (1 - z_i)y^*);$

(P2) *Variance Boundedness:* $\mathbb{E}\left[\|W - \mathbb{E}\left[W\right]\|_F^2\right] \leq \sigma^2;$

(P3) *Entrywise Boundedness:* $|W_{i_1,\dots,i_m}| \leq B,$ *for all* $i_1,\dots,i_m \in [n].$

The goal is to identify the group membership y^* from the observed affinity tensor W.

Our definition of m-degree Homogeneous Polynomial Model is highly general. Informally speaking, (P1) states that one could set the expectation of W for each group arbitrarily by choosing proper p's (see Lemma 16). (P2) and (P3) only require the variance and absolute value to be bounded above. In Section 5, we show that several high-order examples motivated from prior literature, such as high-order counting models, hypergraph cuts models, minimum bisection models, and motif models, belong to the class of m-HPMs.

2.1 Related Tensor Lemmas

Here we provide a series of tensor lemmas that will be used in our analysis. Proofs of the lemmas can be found in Appendix A.

Lemma 2 (Tensor Inner Product) *For any tensor $X = x^{(1)} \otimes \dots \otimes x^{(m)}$ and $Y = y^{(1)} \otimes \dots \otimes y^{(m)}$ in $\mathcal{S}^{n,m}$, we have $\langle X, Y \rangle = \prod_{i=1}^{m} x^{(i)\top} y^{(i)}$.*

Lemma 3 (Tensor Norm Inequality) *For any tensor $A \in \mathcal{S}^{n,m}$, $\lambda_{\text{tmax}}(A) \leq \|A\|_F$.*

Lemma 4 (Positive Semidefinite Tensor Cone) $\mathcal{S}_+^{n,m}$ *is a convex cone.*

Lemma 5 (Carathéodory symmetric tensor cone) $\mathcal{V}^{n,m}$ *is a convex cone.*

Lemma 6 (Rank-one Tensors) $\mathcal{U}^{n,m} \subset \mathcal{S}_+^{n,m}$, *and the dual cone of $\mathcal{U}^{n,m}$ is $\mathcal{S}_+^{n,m}$, i.e.,* $(\mathcal{U}^{n,m})^* = \mathcal{S}_+^{n,m}$.

Lemma 7 (Dual of Positive Semidefinite Tensor Cone) *The dual cone of $\mathcal{S}_+^{n,m}$ is $\mathcal{V}^{n,m}$, i.e., $(\mathcal{S}_+^{n,m})^* = \mathcal{V}^{n,m}$.*

Lemma 8 (Dual of Carathéodory symmetric tensor cone) *The dual cone of $\mathcal{V}^{n,m}$ is $\mathcal{S}_+^{n,m}$, i.e., $(\mathcal{V}^{n,m})^* = \mathcal{S}_+^{n,m}$.*

3. Convex Relaxation and Analysis

In this section we investigate the conditions for exact partitioning the m-degree Homogeneous Polynomial Model into two groups of equal size. We say an algorithm achieves *exact partitioning* if the recovered node labels y is identical to the true labels y^*.

Our analysis consists of two parts. First we show the exact partitioning problem for m-HPMs can be relaxed to a *conic form problem*, a class of convex optimization problems containing *semidefinite programming* as a specific case. In the second part we use *primal-dual*

certificates and statistical concentration inequalities to analyze the sufficient conditions of the problem.

Our algorithm does not require rounding of the solution. Our proof states that if the statistical conditions are satisfied, our optimization problem will always return the integral ground truth y^* as the solution.

The balanced clusters assumption is for clarity of presentation. We can relax the last constraint in (1) in the following subsection to $\sum_i y_i = k$, to allow for different cluster sizes. This does not break our analysis with the novel tensor primal and dual cones.

3.1 Conic Relaxation

We first consider a greedy approach to partition a m-HPM. Given an observed affinity tensor W, we try to find a labeling vector y, such that $\sum_{i_1,\dots,i_m} W_{i_1,\dots,i_m} y_{i_1} \dots y_{i_m}$ is maximized. Using tensor notations introduced in the previous sections, this can be cast as the following optimization problem

$$\underset{y}{\text{maximize}} \quad \langle W, y^{\otimes m} \rangle ,$$
$$\text{subject to} \quad y \in \{+1, -1\}^n, \quad \sum_i y_i = 0 . \tag{1}$$

Problem (1) is nonconvex because of the constraint on y. The size of the space of possible y's is exponential in terms of n. In fact, in the case of second order ($m = 2$), the problem reduces to the Minimum Bisection problem, which is known to be NP-hard (Garey et al., 1976).

To relax the problem we denote $Y = y^{\otimes m}$. Note that every tensor diagonal element $Y_{i,\dots,i}$ is always 1 since m is even. By Lemma 2, $\langle Y, \mathbf{1}^{\otimes m} \rangle = (\mathbf{1}^\top y)^m = 0$. Thus (1) can be rewritten in the following tensor form

$$\underset{Y}{\text{maximize}} \quad \langle W, Y \rangle ,$$
$$\text{subject to} \quad Y_{\sigma_2^{n,m}} = 1, \quad \langle Y, \mathbf{1}^{\otimes m} \rangle = 0, \quad Y = y^{\otimes m} , \tag{2}$$

and recall that $\sigma_2^{n,m}$ is a set of indices defined in Section 2.

The first constraint above ensures that all entries in tensor Y with even number of repeating indices are set to 1. For example, in the $m = 4$ case, this leads to $Y_{1,1,1,1} = Y_{2,2,2,2} = Y_{1,1,2,2} = Y_{1,2,1,2} = \dots = 1$. On the other hand, the last constraint in (2) is still nonconvex. We then substitute it with a tensor cone constraint

$$\underset{Y}{\text{maximize}} \quad \langle W, Y \rangle ,$$
$$\text{subject to} \quad Y_{\sigma_2^{n,m}} = 1, \quad \langle Y, \mathbf{1}^{\otimes m} \rangle = 0, \quad Y \succeq_{\mathcal{V}^{n,m}} 0 , \tag{3}$$

where $\mathcal{V}^{n,m}$ is the Carathéodory symmetric tensor cone as defined in the previous section.

Lemma 5 tells that $\mathcal{V}^{n,m}$ is a convex cone, thus (3) is a convex conic form problem. Furthermore it can be seen that $\mathcal{V}^{n,m}$ has a non-empty interior, and there always exists some strictly feasible Y's for the problem. The Lagrangian of (3) is

$$L(Y, V, \eta, A) = -\langle W, Y \rangle + \langle V, Y - \mathbf{1}^{\otimes m} \rangle + \eta \langle Y, \mathbf{1}^{\otimes m} \rangle - \langle Y, A \rangle$$
$$= \langle -W + V + \eta \mathbf{1}^{\otimes m} - A, Y \rangle - \langle V, \mathbf{1}^{\otimes m} \rangle ,$$

where $V \in \mathcal{S}^{n,m}, \eta \in \mathbb{R}, A \succeq_{\mathcal{S}_+^{n,m}} 0$ are Lagrangian multipliers, subject to the constraint that $V_{\bar{\sigma}_2^{n,m}} = 0$. Note that $\mathcal{S}_+^{n,m}$ is the dual cone of $\mathcal{V}^{n,m}$ by Lemma 8. Taking the derivative of L with respect to Y, we obtain

$$\nabla_Y L = -W + V + \eta \mathbf{1}^{\otimes m} - A.$$

Setting the derivative to 0, we obtain $-W + V + \eta \mathbf{1}^{\otimes m} = A$, and consequently $-W + V + \eta \mathbf{1}^{\otimes m} \succeq_{\mathcal{S}_+^{n,m}} 0$. This gives us the lower bound $\min L = -\langle V, \mathbf{1}^{\otimes m} \rangle$, which is independent of the choice of Y. Otherwise, the value of L will be unbounded. This leads to the following dual problem

$$\begin{aligned} \underset{V,\eta}{\text{minimize}} \quad & \langle V, \mathbf{1}^{\otimes m} \rangle\,, \\ \text{subject to} \quad & V_{\bar{\sigma}_2^{n,m}} = 0\,, \quad -W + V + \eta \mathbf{1}^{\otimes m} \succeq_{\mathcal{S}_+^{n,m}} 0\,. \end{aligned} \tag{4}$$

Lemma 4 tells that $\mathcal{S}_+^{n,m}$ is a convex cone, thus (4) is also a convex conic form problem.

We now examine the optimality condition of the primal problem (3) and the dual problem (4). We first list the Karush–Kuhn–Tucker (KKT) conditions for a primal and dual pair (Y, V, η, A) to be optimal.

$$-W + V + \eta \mathbf{1}^{\otimes m} - A = 0\,, \qquad \text{(Stationarity)}$$
$$Y_{\sigma_2^{n,m}} = 1\,, \quad \langle Y, \mathbf{1}^{\otimes m} \rangle = 0\,, \quad Y \succeq_{\mathcal{V}^{n,m}} 0\,, \qquad \text{(Primal Feasibility)}$$
$$V_{\bar{\sigma}_2^{n,m}} = 0\,, \quad A \succeq_{\mathcal{S}_+^{n,m}} 0\,, \qquad \text{(Dual Feasibility)}$$
$$\langle A, Y \rangle = 0\,. \qquad \text{(Complementary Slackness)}$$

To guarantee $Y^* = y^{*\otimes m}$ is an optimal solution to the primal problem (3), all KKT conditions need to be fulfilled. First note that Y^* fulfills (Primal Feasibility) trivially because $Y^*_{\sigma_2^{n,m}} = 1$, $\langle Y^*, \mathbf{1}^{\otimes m} \rangle = (y^{*\top}\mathbf{1})^m = 0$, and $Y^* = y^{*\otimes m}$ and thus in $\mathcal{V}^{n,m}$. Next, combining (Stationarity) and (Complementary Slackness), we obtain that an optimal solution must fulfill

$$\langle -W + V + \eta \mathbf{1}^{\otimes m}, Y^* \rangle = 0\,. \tag{5}$$

To fulfill (5), we can construct the dual variables V^*, A^*, η^* as follows: $V^*_{i,\ldots,i} = \sum_{i_2,\ldots,i_m} W_{i,i_2,\ldots,i_m} y_i^* y_{i_2}^* \cdots y_{i_m}^*$ for every $i \in [n]$, $V^*_{i_1,\ldots,i_m} = 0$ for all other entries, $A^* = V^* - W + \eta^* \mathbf{1}^{\otimes m}$, and $\eta^* \to \infty$. It remains to prove that our construction (Y^*, V^*, η^*, A^*) fulfills (Dual Feasibility) and (Complementary Slackness). This gives us the following optimality condition. Proofs in this section can be found in Appendix B.

Lemma 9 (Optimality Condition) *The primal problem* (3) *achieves KKT optimality, if*

$$\min_{\|u\|=1, u \perp \mathbf{1}} \langle V^* - W, u^{\otimes m} \rangle \geq 0\,.$$

The KKT conditions, once fulfilled, guarantee that $Y^* = y^{*\otimes m}$ is an optimal solution to the primal problem. However, there could exist other sets of primal and dual variables that satisfy all KKT conditions above. To illustrate this, we construct a set of example

primal and dual variables $(\tilde{Y}, \tilde{V}, \tilde{\eta}, \tilde{A})$ as follows: $\tilde{Y} := \mathbf{1}^{\otimes m}$ is the all-one tensor, $\tilde{V}_{i,\dots,i} := \sum_{i_2,\dots,i_m} W_{i,i_2,\dots,i_m}$ for every $i \in [n]$, $\tilde{\eta} \to \infty$, and $\tilde{A} := \tilde{V} - W + \tilde{\eta}\mathbf{1}^{\otimes m}$. One can verify that $(\tilde{Y}, \tilde{V}, \tilde{\eta}, \tilde{A})$ fulfill all KKT conditions above, and as a result, $\tilde{Y} = \mathbf{1}^{\otimes m}$ is an optimal solution to the primal problem, which is undesirable from the perspective of recovery.

In the following analysis we denote $\lambda_{\mathbf{1}}(A) = \min_{\|u\|=1, u \perp \mathbf{1}, u \nparallel y^*} \langle A, u^{\otimes m} \rangle$, where u cannot be a multiple of y^*. To ensure that $Y^* = y^{*\otimes m}$ is the *unique* optimal solution to (3) and eliminate all other undesirable solutions, we present the following lemma about uniqueness.

Lemma 10 (Uniqueness Condition) *The primal problem* (3) *achieves exact recovery and returns the unique optimal solution $Y^* = y^{*\otimes m}$, if*

$$\lambda_{\mathbf{1}}(V^* - W) > 0 \,. \tag{6}$$

3.2 Statistical Conditions of Exact Partitioning

In this section we analyse the regime in which (6) holds with probability tending to 1. For simplicity, we define the combinatorial function

$$f(m, l, k) := \sum_{s=\max(0, k-l)}^{\min(k, m-l)} (-1)^s \binom{l}{k-s} \binom{m-l}{s},$$

and

$$F(m, p) := \min \begin{cases} \sum_{l=1}^{m} \binom{m-1}{l-1} \sum_{k=0}^{m} p_k f(m, l, k) \\ \sum_{l=0}^{m-1} \binom{m-1}{l} \sum_{k=0}^{m} p_k f(m, l, k) \,. \end{cases}$$

In particular, note that $F(m, p)$ is a function of model order m and the parameter vector p, and it characterizes the signal / noise level of the underlying model.

Here we present our main theorem on the statistical conditions of exact partitioning.

Theorem 11 *Consider any model $\mathcal{M}$ sampled from class m-HPM(n, p, σ^2, B) with the assumption of $F(m, p) > 0$. If $\frac{(2^{1-m} F(m,p) - p_0)^2}{B^2} = \Omega\left(\frac{\log n}{n}\right)$, and $\frac{(2^{1-m} F(m,p) - p_0)^2}{\sigma^2} = \Omega\left(n^{1-m}\right)$, then conic form problem* (3) *partitions the groups correctly, i.e., the true group assignment y^* is the optimal solution of* (3), *with probability at least $1 - O(1/n)$.*

Proof [Proof Sketch] The proof can be broken down into two parts. Starting from our dual construction as in Lemma 9 and 10, the random tensor $V^* - W$ can be rewritten as $(V^* - \mathbb{E}[V^*]) - (W - \mathbb{E}[W]) + (\mathbb{E}[V^*] - \mathbb{E}[W])$. In the first part, we analyze the variational characterization of the expected tensor $\mathbb{E}[V^* - W]$. We show that $\lambda_{\mathbf{1}}(\mathbb{E}[V^* - W])$ can be bounded below by a quantity. In the second part, we characterize the eigenvalues of the deviation tensor $V^* - \mathbb{E}[V^*]$ and $W - \mathbb{E}[W]$. Since the dual variable V^* is constructed to be a diagonal tensor, the minimum tensor eigenvalue of $V^* - \mathbb{E}[V^*]$ is related to the smallest element in its diagonal. For $W - \mathbb{E}[W]$, by Lemma 3, the maximum tensor eigenvalue is related to its Frobenius norm. At the end, our goal is to ensure that $\lambda_{\mathbf{1}}(V^* - W)$ is greater than 0 with high probability. This gives us the statistical conditions in terms of n, p, σ and B, and since $\lambda_{\mathbf{1}}(V^* - W) > 0$, Lemma 9 and 10 guarantee exact recovery through solving

7

the convex primal problem (3). ■

Theorem 11 provides the sufficient statistical conditions for the high-order exact partitioning problem. Our proof in Theorem 11 states that once the statistical conditions are satisfied, optimization problem (3) will return the integral ground truth $Y^* = y^{*\otimes m}$ with high probability, where y^* is the groundtruth in Definition 1, as well as the optimal solution to problem (1). This means our tensor cone relaxation from (1) to (3) works effectively.

4. Discussions

It is now the best time to discuss the requirement of m being a positive even integer in the previous section. It is known that there exists no odd-degree nonnegative homogeneous polynomial, and there exists no non-zero odd-order positive semidefinite tensor (Yuan and You, 2014). That is, if m is odd, the cone $\mathcal{S}_+^{n,m} = \{\mathbf{0}\}$ and therefore its dual cone $\mathcal{V}^{n,m} = (\mathcal{S}_+^{n,m})^* = \{\mathbf{0}\}^* = \mathcal{S}^{n,m}$, the space of all symmetric tensors. Thus, the requirement of m being a positive even integer is necessary for the convex relaxation.

Our analysis in Section 3 requires the optimal solution to be unique. A natural question is: why is uniqueness important? The reason is that our models are *generative*. In other words, the true group assignment vector y^*, selected by nature, generates the observed affinity tensor W. From an optimization point of view, it is possible that there exists multiple distinct optimal solutions to problem (3), however we are only interested in the groundtruth y^* which generates the model.

Our analysis in Section 2.1 and 3 focuses on the Carathéodory symmetric tensor cone $\mathcal{V}^{n,m}$ and the positive semidefinite tensor cone $\mathcal{S}_+^{n,m}$. In the tensor literature another commonly used cone is the *Sum-of-Squares (SoS)* cone, and recently SoS relaxation approaches have been used in some combinatorial inference problems (Barak and Moitra, 2016; Bello et al., 2022). It is known that for any $m \geq 4$ and $n \geq 2$, the three cones fulfill $\mathcal{V}^{n,m} \subset \text{SoS} \subset \mathcal{S}_+^{n,m}$.

We believe the NP-hardness of checking whether a symmetric tensor is in the Carathéodory symmetric tensor cone is an open problem. It is known that many tensor problems are NP-hard, for example, Hillar and Lim (2013, Theorem 11.2) points out that deciding whether a symmetric 4-th order is positive semidefinite is NP-hard. To the best of our knowledge, it remains unknown if conic form programs with the Carathéodory symmetric tensor cone constraint can be solved efficiently. Furthermore, even if we relax the constraint $Y \succeq_{\mathcal{V}^{n,m}} 0$ in (3) to $Y \succeq_{\mathcal{S}_+^{n,m}} 0$, the argument in Hillar and Lim (2013) will not work and NP-hardness remains an open problem. This is because of our constraint $Y_{\sigma_2^{n,m}} = 1$ in the primal problem. More details about NP-hardness can be found in Appendix C.

We want to highlight that the utility of our novel relaxation procedure is as a proof technique. From a theoretic point of view, it allows us to apply convex optimization tools to characterize the statistical upper limits of exact partitioning for this class of tensor problems. Moreover, convexity introduces new optimization methods, including projected gradient ascent and barrier functions, and thus approximation or randomized algorithms remain possible. This is customary in many fields of machine learning, including deep learning and Bayesian networks: solving for the exact solution is NP-hard, but numerical methods (e.g., stochastic gradient descent) are feasible and efficient. To see this, we provide simulation results on both synthetic and real-world datasets in Section 6.
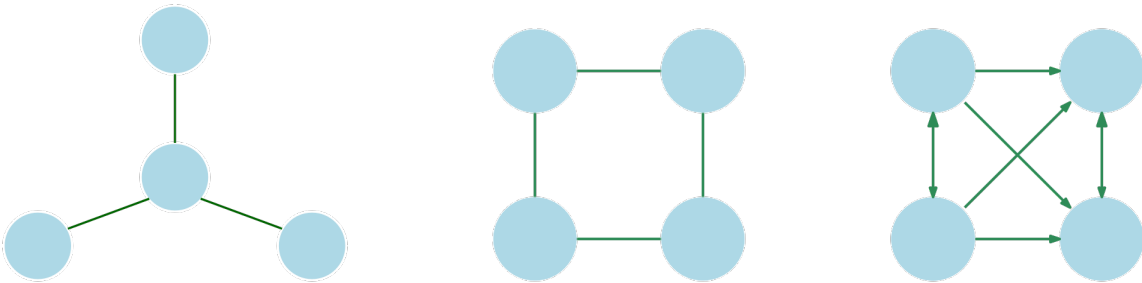
Figure 1: Examples of 4-vertex motifs. **Left:** The undirected star motif. **Middle:** The undirected cycle motif. **Right:** The directed food chain motif (Li and Milenkovic, 2017), where edges represent energy flow between species.

Recently there are some works focusing on the related problem of hypergraph / tensor clustering. To the best of our knowledge the problem setup and formulation are different. Kim et al. (2018) introduced a definition of the Hypergraph Stochastic Block Model (HSBM) as a natural extension of the Stochastic Block Model, in which every binary edge is drawn independently with a probability depending on group membership. This can be subsumed by our Example Model 1 by setting $T = 1$, $\alpha_0 = p$, and all other $\alpha_l = q$. Similarly there are results about community detection on hypergraph Degree-Corrected Block Model (hDCBM) (Ke et al., 2019), which can be subsumed by the same Example Model 1 with a weighted parameter α. Another line of work is the so-called high-order clustering (Luo and Zhang, 2022), where the task can be viewed as an extension of bi-clustering and the goal is to recover the support along each mode of a tensor. In comparison, our goal is to exactly recover the group membership vector y^* that generates the whole observed affinity tensor W.

5. High-order Example Models

In this section, we introduce several high-order models motivated from prior literature. We also show these example models belong to the class of m-HPMs by Definition 1 in Appendix D. It is worth mentioning that there is no prior theoretical work on exact partitioning in high-order models. We are providing the first results for exact partitioning in models with high-order interactions with provable theoretical guarantees. Our Example Models serve to motivate the generative models, but none of those papers contain any theoretical statistical analysis of exact partitioning.

We first consider the *high-order counting model* motivated from (Zhou et al., 2007). Suppose there exists a co-authorship network consisting of computer scientists and biologists, and every paper has m authors. On average, authors from the same discipline collaborate more than those from different backgrounds. The task is to identify the two groups of researchers, given the number of publications of each m-tuple. Naturally the co-authorship network can be modeled as a high-order counting problem. Aside from the example above, high-order counting models can be helpful in many complex application problems as pairwise models often lose high-order information, for example, categorical data (Gibson et al., 2000),

molecular biology (Zhang, 2007), and image segmentation (Agarwal et al., 2005). Next we present a generative model for high-order counting models.

Example Model 1 (High-order Counts) *Let $\mathcal{G} = (V, m, \alpha, T)$ be a high-order counting model with vertex set V and order m. $\alpha = (\alpha_0, \ldots, \alpha_{m/2}) \in [0, 1]^{m/2+1}$ is the counting parameter vector, and $T \in \mathbb{N}$ is a counting parameter. Nature generates random counts for $\mathcal{G}$ in the following way. For each m-tuple $(v_{i_1}, \ldots, v_{i_m}) \subset V$, count the group membership of the vertices. Without loss of generality assume l vertices are from the same group, where $l \in \{0, \ldots, m/2\}$. Nature then samples the corresponding count from binomial distribution: $c(v_{l_1}, \ldots, v_{l_m}) \sim \mathrm{BIN}(T, \alpha_l)$. We are interested in identifying the group membership of vertices from the observed count information $c(v_{l_1}, \ldots, v_{l_m})$.*

Proposition 12 *Example Model 1 belongs to the class of m-HPMs.*

Note that when $T = 1$, the high-order counting model can be interpreted as a random m-uniform hypergraph. As a result one can define *hypergraph cuts* as generalization of regular pairwise graph cuts (Hein et al., 2013; Benson et al., 2016). Using a similar approach one can generalize other notions from graph theory, including *clique*, *volume* and *conductance*. Hypergraph cuts have been found useful in tasks dealing with complex networks, for example, video object segmentation (Huang et al., 2009), clustering animals in a zoo dataset using categorical data (Zhou et al., 2007), among others. Next we present a generative model for hypergraph cut models.

Example Model 2 (Hypergraph Cuts) *Let $G = (V, m = 4, \alpha, T = 1)$ be a random m-uniform hypergraph generated from Model 1, and let H denote the hyperedge set. For each 4-tuple $(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}) \subset V$, we define its cut size $c(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4})$, where $c(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}) = \sum_{e \in H} \mathbf{1}[v_{i_1} \in e \vee v_{i_2} \in e \vee v_{i_3} \in e \vee v_{i_4} \in e] \cdot \mathbf{1}[(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}) \neq e]$. We are interested in inferring the group membership of vertices. Instead of observing the edge set E, we now only observe the cut sizes of every 4-tuple $(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4})$.*

Proposition 13 *Example Model 2 belongs to the class of m-HPMs.*

In the next example we are interested in the *hypergraph minimum bisection* problem. It is well-known that the problem is NP-complete on both pairwise graphs (Garey et al., 1974) and hypergraphs (Kahng, 1989).

Example Model 3 (Minimum Bisection) *Let $\mathcal{G} = (V, m, q, H)$ be a random m-uniform hypergraph with vertex set V and order m. $q \in (0, 1)$ is a activation parameter. The hyperedge set H starts empty and nature add hyperedges to H in the following way. Recall that $y^* \in \{+1, -1\}^n$ is the group assignment vector. For each m-tuple $(v_{i_1}, \ldots, v_{i_m}) \subset V$, nature first generates a temporary activation vector $b \in \{+1, -1\}^m$, such that $b_j = y^*_{i_j}$ with probability $1 - q$, and $b_j = -y^*_{i_j}$ with probability q. If $b_1 = \cdots = b_m$, nature adds a hyperedge $(v_{i_1}, \ldots, v_{i_m})$ to H. Nature then discards the value of b, and repeats the process for other m-tuples. We are interested in identifying the group membership of vertices from the hyperedge set H.*

Proposition 14 *Example Model 3 belongs to the class of m-HPMs.*

In the next example we investigate *motif models.* Motifs are simple network subgraphs and building blocks of many complex networks (Benson et al., 2016; Yaveroğlu et al., 2014; Milo et al., 2002). They can be directed or undirected. Researchers have utilized motifs to explore higher-order patterns and insights in complex systems, such as social networks (Juszczyszyn et al., 2008), air traffic patterns (Rosvall et al., 2014), and food webs (Li and Milenkovic, 2017; Benson et al., 2016). Motifs are powerful tools to represent higher-order interaction patterns of multiple entities. Figure 1 illustrates three distinct motifs of size 4.

We now present a generative model for motif models.

Example Model 4 (Motif Clustering) *Let $G = (V, \alpha, E, M, H)$ be a directed random graph, such that the vertices V are drawn from two groups S_1 and S_2. $\alpha = (\alpha_{1,1}, \alpha_{2,2}, \alpha_{1,2}, \alpha_{2,1}) \in [0, 1]^4$ is a probability parameter vector. The edge set E starts empty and nature add edges to E in the following way. For each pair $(v_1, v_2) \subset V$, if $v_1, v_2 \in S_1$, nature adds a directed edge (v_1, v_2) to E with probability $\alpha_{1,1}$; if $v_1, v_2 \in S_2$, nature adds a directed edge (v_1, v_2) to E with probability $\alpha_{2,2}$; if $v_1 \in S_1, v_2 \in S_2$, nature adds a directed edge (v_1, v_2) to with probability $\alpha_{1,2}$; otherwise nature adds a directed edge (v_1, v_2) with probability $\alpha_{2,1}$. M is a m-vertex motif of interest, and H is the set of observed motifs. H starts empty. For each m-tuple $(i_1, \ldots, i_m) \subset V$, nature adds $(i_1, \ldots, i_m)$ to H if the tuple $(i_1, \ldots, i_m)$ forms the motif M exactly. We are interested in inferring the group of vertices from the set of observed motifs H.*

Proposition 15 *Example Model 4 belongs to the class of m-HPMs.*

6. Simulation Results

In this section, we test the proposed convex optimization formulation (3) by implementing a projected gradient descent solver in MATLAB. To do so, we relax the Carathéodory symmetric tensor cone to the positive semidefinite tensor cone. Projected gradient descent is a standard method to solve constrained convex optimization problems. Our projected gradient descent solver in Algorithm 1 works as follows: starting from an initial point $Y^{(0)}$, the algorithm repeats the following assignment until a stopping condition is met:

$$Y^{(k+1)} \leftarrow P(Y^{(k)} + \zeta W),$$

where ζ is the step size of each iteration, and $P(\cdot)$ is a projection operator.

The projection operator tries to find the "closest" point to $Y^{(k)} + \zeta W$, that fulfills all the constraints in the convex problem (3). To fulfill the first two constraints $Y_{i,\ldots,i} = 1$ and $\langle Y, \mathbf{1}^{\otimes m} \rangle = 0$, the algorithm subtracts the mean of all entries in $Y^{(k)}$ from each entry, and sets the diagonal entries to be 1. To fulfill the positive semidefinite constraint $Y \succeq_{\mathcal{S}_+^{n,m}} 0$, Algorithm 1 invokes the helper function in Algorithm 2, using the method in Equation (3.1) in Han (2013), which tries to find a unit vector $v \in \mathbb{R}^n$ such that $\langle Y^{(k)}, v^{\otimes m} \rangle$ is minimized. Algorithm 1 then subtracts the tensor $\langle Y^{(k)}, v^{\otimes m} \rangle \cdot v^{\otimes m}$ from $Y^{(k)}$, making the projected tensor positive semidefinite in the direction of v. We assume the point $P(Y^{(k)} + \zeta W)$ satisfies the constraints, after repeating the projection procedure for a number of iterations. Theorem 5 in Han (2013) guarantees that the output vector v is either the zero vector $\mathbf{0}$, or v is a

critical point of the function $\left\langle Y^{(k)}, v^{\otimes m} \right\rangle$. If Algorithm 2 returns the zero vector, it is likely that the input $Y^{(k)}$ is already positive semidefinite.

Algorithm 1 Projected Gradient Descent Solver

Input: Observed affinity tensor W, step size ζ
Output: Agreement tensor $\hat{Y}$

  $Y \leftarrow \text{diag}\,(\mathbf{1})$
  for each outer iteration **do**
    $Y \leftarrow Y + \zeta W$ {gradient descent step}
    for each inner iteration **do**
      $v \leftarrow \text{Algorithm2}(Y)$
      $v \leftarrow \frac{1}{\|v\|} v$
      $V \leftarrow v^{\otimes m}$
      $c_v \leftarrow \langle V, Y \rangle$ {run projection step if Y is not PSD}
      if $c_v < 0$ **then**
        $Y \leftarrow Y - c_v Y$ $\{Y \succeq_{\mathcal{S}_+^{n,m}} 0\}$
        $Y \leftarrow Y - \left(\frac{1}{n^m} \sum_{i_1,\dots,i_m} Y_{i_1,\dots,i_m}\right)\mathbf{1}^{\otimes m}$ $\{\langle Y, \mathbf{1}^{\otimes m}\rangle = 0\}$
        $Y_{i,\dots,i} \leftarrow 1$ $\{Y_{i,\dots,i} = 1\}$
      end if
    end for
    $Y \leftarrow Y - \left(\frac{1}{n^m} \sum_{i_1,\dots,i_m} Y_{i_1,\dots,i_m}\right)\mathbf{1}^{\otimes m}$
    $Y_{i,\dots,i} \leftarrow 1$
  end for

Algorithm 2 Negative Tensor Eigenvalue Searcher (see Equation (3.1) and Theorem 5 in Han (2013))

Input: Target tensor A, step size γ
Output: Target vector x

  Initialize x
  for each iteration **do**
    if $\langle A, x^{\otimes m} \rangle < 0$ **then**
      Break
    end if
    $\nabla f_1 \leftarrow \langle I, x^{\otimes m} \rangle I x^{\otimes m-1} + A x^{\otimes m-1}$
    $x \leftarrow x - \gamma \nabla f_1$
  end for
  if $\langle A, x^{\otimes m} \rangle \geq 0$ **then**
    $x \leftarrow \mathbf{0}$
  end if

Given an input tensor A, Algorithm 2 searches for the target vector x by solving an unconstrained optimization problem as follows

$$\min f_1(x) = \frac{1}{2m} \left\langle I, x^{\otimes m} \right\rangle^2 + \frac{1}{m} \left\langle A, x^{\otimes m} \right\rangle ,$$

where I is the identity tensor, such that $I_{i,\dots,i} = 1$ and all other entries are 0 (see Han (2013, eq. (3.1)). Since both I and A are symmetric, the gradient of the objective function $f_1(x)$ is

$$\nabla f_1(x) = \langle I, x^{\otimes m} \rangle I x^{\otimes m-1} + A x^{\otimes m-1},$$

where $Ix^{\otimes m-1}$, $Ax^{\otimes m-1}$ are n-dimensional vectors, such that $(Ix^{\otimes m-1})_i = x_i^{m-1}$, and $(Ax^{\otimes m-1})_i = \sum_{i_2,\dots,i_m} A_{i,i_2,\dots,i_m} x_{i_2} \dots x_{i_m}$. Algorithm 2 repeats the gradient descent step $x \leftarrow x - \gamma \nabla f_1(x)$ and checks the objective value $f_1(x)$. If $f_1(x)$ is less than 0, the algorithm finds a negative eigenvalue and returns the corresponding eigenvector x.

Here we also present Algorithm 3, which is a modified version of Algorithm 2, and can be used to search for positive tensor eigenvectors in the same fashion.

Algorithm 3 Positive Tensor Eigenvalue Searcher
___
Input: Target tensor A, step size γ
Output: Target vector x

  Initialize x
  for each iteration **do**
    $\nabla f_1 \leftarrow \langle I, x^{\otimes m} \rangle I x^{\otimes m-1} - A x^{\otimes m-1}$
    $x \leftarrow x - \gamma \nabla f_1$
  end for
  if $\langle A, x^{\otimes m} \rangle \leq 0$ **then**
    $x \leftarrow \mathbf{0}$
  end if

Experiment 1: To validate Theorem 11, we generate synthetic high-order counting models following the definition of Example Model 1. We set the parameter $T = 1$, and $\alpha = (a, b, b)$, where a and b are signal parameters we control. In the experiment a and b iterate over the range $(0, 1)$, with an interval of 0.05. We run 20 trials for each pair of a and b. During each trial, a fourth order affinity tensor W is sampled using a entrywise Bernoulli distribution (up to symmetry), and we invoke Algorithm 1 to solve for the predicted structure tensor $\hat{Y}$. To evaluate the performance of the proposed algorithm, we run Algorithm 3 to compute the predicted label vector $\hat{y}$, which is the top tensor eigenvector of $\hat{Y}$. We then compare the corresponding sign vector $\mathrm{sign}(\hat{y})$ against the groundtruth y^*.

We test cases with n being 30, 60, and 120, respectively. We report the empirical probability of exact recovery $\mathbb{P}\{\hat{y} = y^*\}$ in Figure 2. We plot the empirical probability against $C := \log\left(\frac{(2^{1-m} F(m,p) - p_0)^2 n}{\log n}\right)$, which is equivalent to log of the left-hand side of Theorem 11 bound divided by its right-hand side. Our result suggests that as long as C is greater than a constant threshold, the proposed algorithm performs well and recovers the underlying group structure with high probability. This matches our theoretic findings in Theorem 11.

Experiment 2: We implement the proposed algorithm on a real-world dataset, *email-Eu-core* (Leskovec and Krevl, 2014). During each trial, we extract n most connected nodes from the dataset, and convert the induced subgraph into fourth order hypergraphs by counting the existence of star motifs and cycle motifs (see Figure 1). We then follow the same procedure used in the previous experiment. We also compare our results with the motif clustering algorithm (Benson et al., 2016).
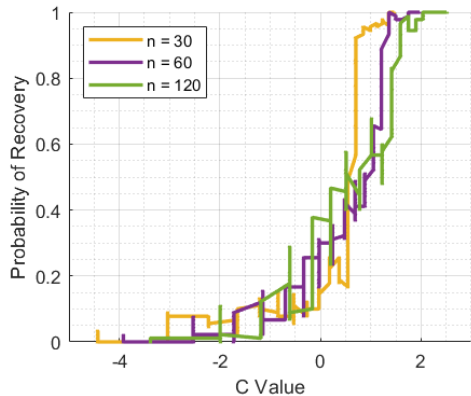
Figure 2: Simulations using synthetic fourth order planted models with different values of signal parameters p and q. In this case $m = 4$. The x-axis is set by $C := \log\left(\frac{(2^{1-m}F(m,p)-p_0)^2 n}{\log n}\right)$. The y-axis is the probability of exact partitioning $\mathbb{P}\{\hat{y} = y^*\}$. This matches our theoretic findings in Theorem 11.
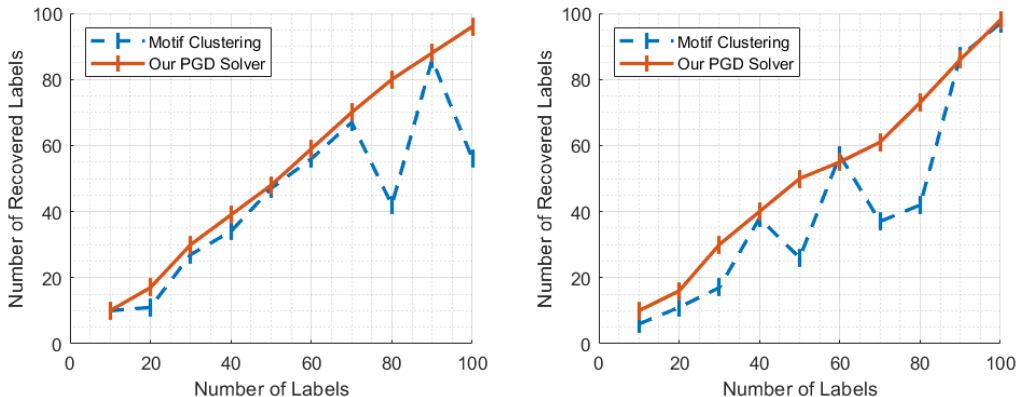

Figure 3: Simulations using the real-world dataset *email-Eu-core* (Leskovec and Krevl, 2014). We generate fourth-order hypergraphs from the dataset using the star motif and the cycle motif, respectively (Figure 1). In this case $m = 4$. The dashed curve uses the motif clustering algorithm in (Benson et al., 2016). The result suggests that the proposed method performs well on real-world datasets.

We report the number of correctly recovered labels in Figure 3. The result suggests that our method performs well compared to the motif clustering method on both hypergraphs. In particular, the figure suggests that our tensor method is more robust to noises ($n \geq 70$), since as n gets larger, newly extracted nodes are less connected, and the community structure signal becomes weaker. We also report the runtime of the proposed method in Figure 4. The runtime can be fitted almost perfectly by a third order polynomial, suggesting that our method has polynomial efficiency.
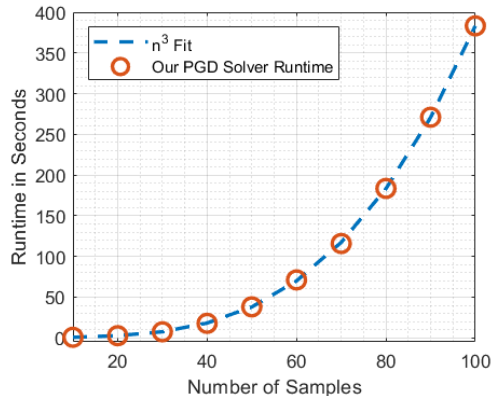
Figure 4: The runtime of Algorithm 1 versus the number of entities. The runtime can be fitted almost perfectly by a third order polynomial, suggesting that our method has polynomial efficiency.

7. Future Work

In this paper, we considered the problem of exact partitioning of high-order models. We defined a highly general class of m-HPMs, and provided novel convex relaxation analysis and provable statistical guarantees for exact partitioning.

It is still unknown, though, whether the problem of determining the membership of Carathéodory symmetric tensor cone is NP-hard or not (see discussion in Section 4). In addition, in our models it is assumed that every interaction has an order of m. It could be interesting to consider the case where interactions with order less than m are also taken into account.

We used a first-order projected gradient descent algorithm 1 to solve the high-order partitioning problem, which takes cubic time. For large-scale high-order models, finding more efficient stochastic algorithms to reduce the computational cost can be an interesting direction.

In our model, it is assumed that each entity belongs only one class. It is unclear how the partitioning problem should be tackled, if there exists entities with mixed membership. A related question is how to extend our analysis to the multiclass case. One can use a homogeneity assumption that is similar to (P1). The difference is that instead of using $m+1$ parameters (i.e., p_0 through p_m) for the two group case, now we need $\binom{m+c-1}{c-1}$ parameters, where c is the number of groups. At this point, it is unclear how to encode the homogeneity using some equations that are similar to (P1). Finally, can we have similar guarantees if some entries in the observed affinity tensor W are missing? Are some entries more important (to the task of exact partitioning) than other entries? Intuitively if a hypergraph is disconnected, there is no way to recover the membership perfectly. In the graph case (that is, $m = 2$), the connectivity can be characterized by the Cheeger constant, which provides a lower bound for the second smallest eigenvalue of a related Laplacian matrix (Ke and Honorio, 2022). It is unknown if any similar connection between tensor eigenvalues and hypergraph structures can be built. These can be interesting future directions.

Acknowledgments

Appendix A. Proofs of Tensor Lemmas

Here we highlight the connection between M in the definition of the Carathéodory symmetric tensor cone, and the *symmetric tensor rank* in the tensor (multilinear) algebra literature. For any symmetric tensor A in the complex field $\mathbb{C}$, its (complex) symmetric rank is defined as $\min\{r \mid A = \sum_{i=1}^{r} u^{(i)\otimes m}, u^{(i)} \in \mathbb{C}^n\}$ (Comon et al., 2008). If one limits the discussion to the real field, the real symmetric rank of A is defined as $\min\{r \mid A = \sum_{i=1}^{r} \lambda_i A^{(i)}, A^{(i)} \in \mathcal{U}^{n,m}, \lambda_i \in \mathbb{R}\}$. The major difference between the real symmetric rank and the Carathéodory symmetric tensor cone is that, each term in the definition of the real symmetric rank has a possibly negative coefficient λ_i. Furthermore, Comon et al. (2008) proves that the complex symmetric rank of any tensor in $\mathcal{S}^{n,m}$ is at most $\binom{m+n-1}{m}$, and Ballico (2014) shows that the real symmetric rank is at most m times the complex symmetric rank, that is, $m \cdot \binom{m+n-1}{m}$. On the other hand, every tensor in the Carathéodory symmetric tensor cone has a real symmetric rank at most $M = \binom{m+n-1}{m} + 1$.

Proof [Proof of Lemma 2] By definition of inner products, we obtain

$$
\begin{aligned}
\langle X, Y \rangle &= \sum_{i_1,\ldots,i_m=1}^{n} X_{i_1,\ldots,i_m} Y_{i_1,\ldots,i_m} = \sum_{i_1,\ldots,i_m=1}^{n} x_{i_1}^{(1)} \ldots x_{i_m}^{(m)} y_{i_1}^{(1)} \ldots y_{i_m}^{(m)} \\
&= \sum_{i_1,\ldots,i_m=1}^{n} \prod_{j=1}^{m} x_{i_j}^{(j)} y_{i_j}^{(j)} = \left(\sum_{i_1=1}^{n} x_{i_1}^{(1)} y_{i_1}^{(1)} \right) \ldots \left(\sum_{i_m=1}^{n} x_{i_m}^{(m)} y_{i_m}^{(m)} \right) = \prod_{i=1}^{m} x^{(i)\top} y^{(i)}.
\end{aligned}
$$

This completes our proof. ∎

Proof [Proof of Lemma 3] We use Cauchy-Schwarz inequality in our proof. Note that for any $A \in \mathcal{S}^{n,m}$,

$$
\begin{aligned}
\lambda_{\text{tmax}}(A)^2 &= \max_{u \in \mathbb{R}^n, \|u\|=1} \langle A, u^{\otimes m} \rangle^2 \leq \max_{u \in \mathbb{R}^n, \|u\|=1} \langle A, A \rangle \cdot \langle u^{\otimes m}, u^{\otimes m} \rangle \\
&= \|A\|^2 \max_{u \in \mathbb{R}^n, \|u\|=1} (u^\top u)^m = \|A\|^2.
\end{aligned}
$$

Since norms are nonnegative, we have $\lambda_{\text{tmax}}(A) \leq \|A\|_{\text{F}}$. ∎

Proof [Proof of Lemma 4] For any $A, B \in \mathcal{S}_+^{n,m}$ and $u \in \mathbb{R}^n$, we have

$$
\langle \theta_1 A + \theta_2 B, u^{\otimes m} \rangle = \theta_1 \langle A, u^{\otimes m} \rangle + \theta_2 \langle B, u^{\otimes m} \rangle \geq 0,
$$

if $\theta_1, \theta_2 \geq 0$. ∎

Proof [Proof of Lemma 5] First note that $\mathcal{V}^{n,m}$ is a cone. That is, for every $A \in \mathcal{V}^{n,m}$ and $\theta \geq 0$, we have $\theta A \in \mathcal{V}^{n,m}$. Next we show $\mathcal{V}^{n,m}$ is convex by considering the convex hull of the set of rank-one tensors $\mathcal{U}^{n,m}$. Note that for any tensor $A \in \text{conv}(\mathcal{U}^{n,m})$, $\dim \mathcal{S}^{n,m} = M - 1$ is the maximum number of possibly different entries in A due to symmetry. Applying Carathéodory's theorem to $\mathcal{U}^{n,m}$ leads to $\mathcal{V}^{n,m} = \left\{ \sum_{i=1}^{M} A^{(i)} \mid A^{(i)} \in \mathcal{U}^{n,m}, i \in [M] \right\} = \text{conv}(\mathcal{U}^{n,m})$. This completes our proof. ∎

Proof [Proof of Lemma 6] To prove the first part, by Lemma 2, for any $u, v \in \mathbb{R}^n$, we have $\langle u^{\otimes m}, v^{\otimes m} \rangle = (u^\top v)^m \geq 0$. Thus, by definition of $\mathcal{S}_+^{n,m}$, $u^{\otimes m} \in \mathcal{S}_+^{n,m}$. To prove the second part, by definition of dual cones, we have $(\mathcal{U}^{n,m})^* = \{A \in \mathcal{S}^{n,m} \mid \langle A, B \rangle \geq 0, \forall B \in \mathcal{U}^{n,m}\} = \mathcal{S}_+^{n,m}$. This completes our proof. ∎

Proof [Proof of Lemma 7] The statement of the lemma is proposed in Luo et al. (2015) without proofs. Here we give a proof for completeness. We use $\text{cl}(\cdot)$ to denote the *closure* of a set, and $\text{conv}(\cdot)$ to denote the *convex hull* of a set.

First we prove that $\mathcal{V}^{n,m}$ is a subset of $\mathcal{S}_+^{n,m}$. Note that for any $A \in \mathcal{V}^{n,m}$, one can write A as the summation of at most M tensors from $\mathcal{U}^{n,m}$. In other words, $A = \sum_{i=1}^{M} A^{(i)}$, where each $A^{(i)} \in \mathcal{U}^{n,m}$. Since $A^{(i)} \in \mathcal{U}^{n,m} \subset \mathcal{S}_+^{n,m}$, we have $A = \sum_{i=1}^{M} A^{(i)} \in \mathcal{S}_+^{n,m}$. Thus $\mathcal{V}^{n,m} \subset \mathcal{S}_+^{n,m}$.

Next we prove that $\mathcal{V}^{n,m}$ is closed, by showing that if a set contains all limit points, then the set is closed (Munkres, 2014). Without loss of generality assume A is a limit point of $\mathcal{V}^{n,m}$, such that $\|A\|_F < \infty$. By definition of limit points, A can be approximated by points in $\mathcal{V}^{n,m}$. Mathematically this means that one can find an infinite sequence $\{A_1, A_2, \ldots\}_{j=1}^{\infty} \subset \mathcal{V}^{n,m}$, such that $\lim_{j \to \infty} A_j = A$. Since every A_j is in set $\mathcal{V}^{n,m}$, by definition there exists a collection of n-dimensional vectors $\{x^{(ij)}\}_{i=1}^{M} \subset \mathbb{R}^n$, such that $A_j = \sum_{i=1}^{M} (x^{(ij)})^{\otimes m}$. We then consider the tensor Frobenius norm of A_j's. Note that the infinite sequence of tensor Frobenius norm $\{\|A_1\|_F, \|A_2\|_F, \ldots\}_{j=1}^{\infty}$ is bounded above with respect to A, since $\|A\|_F$ is bounded above. Now expanding the j-th term in the tensor Frobenius norm sequence, we obtain

$$\sum_{i=1}^{M} \left\| x^{(ij)} \right\|_F^{2m} = \|A_j\|_F^2 - \sum_{i \neq k} \left((x^{(ij)})^\top (x^{(kj)}) \right)^m.$$

Since m is even, every term of the summation on the right-hand side is nonnegative. From the fact that $\{\|A_j\|_F\}$ is bounded above and summation is nonnegative, one can tell that the tensor Frobenius norm $\left\| x^{(ij)} \right\|_F$ is bounded above for every i and j. Without loss of generality, we assume there exists $x^{(i)} = \lim_{j \to \infty} x^{(ij)}$ for every $i \in [M]$. It follows that

$$A = \lim_{j \to \infty} A_j = \lim_{j \to \infty} \sum_{i=1}^{M} (x^{(ij)})^{\otimes m} = \sum_{i=1}^{M} (x^{(i)})^{\otimes m}.$$

Then by definition, $A \in \mathcal{V}^{n,m}$. Since every limit point of $\mathcal{V}^{n,m}$ is contained by itself, topology tells us $\mathcal{V}^{n,m}$ is closed, and $\text{cl}(\mathcal{V}^{n,m}) = \mathcal{V}^{n,m}$.

Note that for any tensor $A \in \text{conv}(\mathcal{U}^{n,m})$, $\dim \mathcal{S}^{n,m}$ is the maximum number of possibly different entries in A due to symmetry. We define a bijective mapping $\text{vec}(\cdot) : \mathcal{S}^{n,m} \to$

$\mathbb{R}^{\dim \mathcal{S}^{n,m}}$, which takes a tensor and unfolds it to a vector. Since $M = \dim \mathcal{S}^{n,m} + 1$, applying Carathéodory's theorem to $\text{vec}(\cdot)$ with basis from $\text{vec}(\mathcal{U}^{n,m})$, we have $\mathcal{V}^{n,m} = \text{conv}(\mathcal{U}^{n,m})$. Since $\mathcal{V}^{n,m}$ is closed and $\mathcal{V}^{n,m} \subset \mathcal{S}^{n,m}_+$, we have $\mathcal{V}^{n,m} = \text{cl}(\text{conv}(\mathcal{U}^{n,m}))$. Also by Lemma 6, since $\mathcal{S}^{n,m}_+ = (\mathcal{U}^{n,m})^*$, we have $(\mathcal{S}^{n,m}_+)^* = (\mathcal{U}^{n,m})^{**} = \text{cl}(\text{conv}(\mathcal{U}^{n,m}))$ since for any set C, $(C)^{**} = \text{cl}(\text{conv}(C))$. Thus we have $(\mathcal{S}^{n,m}_+)^* = \mathcal{V}^{n,m}$. ∎

Proof [Proof of Lemma 8] Note that for any cone $\mathcal{K}$, we have $(\mathcal{K}^*)^* = \text{cl}(\text{conv}(\mathcal{K}))$. Since $\mathcal{S}^{n,m}_+$ is closed and convex, we have $((\mathcal{S}^{n,m}_+)^*)^* = \text{cl}(\text{conv}(\mathcal{S}^{n,m}_+)) = \mathcal{S}^{n,m}_+$. By Lemma 7 we obtain $(\mathcal{V}^{n,m})^* = \mathcal{S}^{n,m}_+$. ∎

Appendix B. Proof of Main Theorem

Proof [Proof of Lemma 9] To prove the lemma, we verify the true primal variable Y^* and the contructed dual variables (V^*, A^*, η^*) satisfy all KKT conditions.

Regarding the primal variable, note that $Y^* = y^{*\otimes m}$, and y^* contains equal number of $+1$'s and -1's. As a result, we have $Y^*_{\sigma_2^{n,m}} = 1$, and by Lemma 2, $\langle Y^*, \mathbf{1}^{\otimes m} \rangle = \langle y^{*\otimes m}, \mathbf{1}^{\otimes m} \rangle = (y^{*\top}\mathbf{1})^m = 0$. Also note that $y^{*\otimes m}$ by definition is in the $\mathcal{V}^{n,m}$ cone. Thus we have shown the groundtruth $Y^* = y^{*\otimes m}$ fulfills (Primal Feasibility).

Now we show the constructed dual variables satisfy the other three KKT conditions. First the (Stationarity) condition is trivially satisfied by our construction $A^* = V^* - W + \eta^*\mathbf{1}^{\otimes m}$. Next regarding the (Complementary Slackness) condition, note that

$$\langle A^*, Y^* \rangle = \langle V^* - W + \eta^*\mathbf{1}^{\otimes m}, Y^* \rangle = \langle V^*, Y^* \rangle - \langle W, Y^* \rangle \ .$$

By construction, V^* is a diagonal tensor with $V^*_{i,\ldots,i} = \sum_{i_2,\ldots,i_m} W_{i,i_2,\ldots,i_m} y^*_i y^*_{i_2} \cdots y^*_{i_m}$ for every $i \in [n]$. This leads to

$$\langle V^*, Y^* \rangle - \langle W, Y^* \rangle = \sum_i \sum_{i_2,\ldots,i_m} W_{i,i_2,\ldots,i_m} y^*_i y^*_{i_2} \cdots y^*_{i_m} - \sum_{i_1,\ldots,i_m} W_{i_1,\ldots,i_m} y^*_{i_1} \cdots y^*_{i_m} = 0 \ ,$$

satisfying the (Complementary Slackness) condition.

Finally we discuss the (Dual Feasibility) condition, which requires $A^* = V^* - W + \eta^*\mathbf{1}^{\otimes m} \succeq_{\mathcal{S}^{n,m}_+} 0$. Recall from Section 2, that $A^* \in \mathcal{S}^{n,m}_+$, if $\langle A^*, u^{\otimes m} \rangle \geq 0$ for all vector u. This is equivalent to requiring $\min_{\|u\|=1} \langle A^*, u^{\otimes m} \rangle \geq 0$. We discuss two cases of u. In the first case, for every unit vector u that is not orthogonal to $\mathbf{1}$, we have $\langle A^*, u^{\otimes m} \rangle = \langle V^* - W + \eta^*\mathbf{1}^{\otimes m}, u^{\otimes m} \rangle = \langle V^* - W, u^{\otimes m} \rangle + \eta(\mathbf{1}^{\top}u)^m$. Regarding the first term, note that each entry of W is bounded by B as in Definition 1, and each entry of V^* is bounded by Bn^{m-1}. By Lemma 3 we obtain $\langle V^* - W, u^{\otimes m} \rangle \geq -\|V^*\|_{\text{F}} - \|W\|_{\text{F}} \geq -Bn^{m-1/2} - Bn^{m/2}$, which is a finite quantity bounded below. Since m is even and η is tending to infinity, the summation $\langle V^* - W, u^{\otimes m} \rangle + \eta(\mathbf{1}^{\top}u)^m$ is always greater than 0, or equivalently, $\min_{\|u\|=1, u \not\perp \mathbf{1}} \langle A^*, u^{\otimes m} \rangle \geq 0$. In the second case, for every unit vector $u \perp \mathbf{1}$, we have $\langle A^*, u^{\otimes m} \rangle = \langle V^* - W + \eta^*\mathbf{1}^{\otimes m}, u^{\otimes m} \rangle = \langle V^* - W, u^{\otimes m} \rangle$, which can potentially be negative. Combining the two cases above, as long as $\min_{\|u\|=1, u \perp \mathbf{1}} \langle V^* - W, u^{\otimes m} \rangle \geq 0$, we obtain $\min_{\|u\|=1} \langle V^* - W + \eta^*\mathbf{1}^{\otimes m}, u^{\otimes m} \rangle \geq 0$ and equivalently, $A^* = V^* - W + \eta^*\mathbf{1}^{\otimes m} \succeq_{\mathcal{S}^{n,m}_+} 0$.

This fulfills the (Dual Feasibility) condition and completes our proof. ∎

Proof [Proof of Lemma 10] First we want to make sure $Y^* = y^{*\otimes m}$ is an optimal solution. By Lemma 9, it is sufficient to ensure $\min_{\|u\|=1, u\perp\mathbf{1}} \langle V^* - W, u^{\otimes m}\rangle \geq 0$. In particular, if u is a multiple of y^*, we have $\langle V^* - W, y^{*\otimes m}\rangle = 0$, which is exactly the (Complementary Slackness) condition. Now to ensure uniqueness of the solution, we want the equality above to hold only for multiples of $Y^* = y^{*\otimes m}$. This leads to our condition $\min_{\|u\|=1, u\perp\mathbf{1}, u\not\parallel y^*} \langle V^* - W, u^{\otimes m}\rangle = \lambda_{\mathbf{1}}(V^* - W) > 0$. Furthermore, given the constraint $Y_{\sigma_2^{n,m}} = 1$ as in (Primal Feasibility), all other multiples of Y^* are eliminated from the space of feasible solutions, and the only feasible solution that fulfills all KKT conditions is Y^* itself. This completes our proof. ∎

Proof [Proof of Theorem 11] Lemma 10 tells that as long as (6) holds, the conic form problem (3) returns the correct labeling. Our goal is to prove (6) holds with high probability. Note that (6) is a function of random variable W. By definition of $\lambda_{\mathbf{1}}$, we obtain the following decomposition

$$\lambda_{\mathbf{1}}(V^* - W) = \min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle V^* - W, u^{\otimes m}\rangle$$

$$\geq \min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle V^* - \mathbb{E}[V^*], u^{\otimes m}\rangle \tag{7}$$

$$+ \min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle -W + \mathbb{E}[W], u^{\otimes m}\rangle \tag{8}$$

$$+ \min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle \mathbb{E}[V^* - W], u^{\otimes m}\rangle, \tag{9}$$

and it is sufficient to prove the summation of (7), (8) and (9) is greater than 0. By definition of tensor eigenvalues, for (7) we obtain the following lower bound

$$\min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle V^* - \mathbb{E}[V^*], u^{\otimes m}\rangle \geq \lambda_{\text{tmin}}(V^* - \mathbb{E}[V^*])$$

$$= \min_i (V^*_{i,\dots,i} - \mathbb{E}\left[V^*_{i,\dots,i}\right]). \tag{10}$$

Similarly for (8), we have

$$\min_{u\perp\mathbf{1}, u\not\parallel y^*, \|u\|=1} \langle -W + \mathbb{E}[W], u^{\otimes m}\rangle \geq \lambda_{\text{tmin}}(-W + \mathbb{E}[W])$$

$$= -\lambda_{\text{tmax}}(W - \mathbb{E}[W]). \tag{11}$$

Regarding the expectation in (9), we first characterize the expectation of W. Consider the definition of (P1):

$$\mathbb{E}[W] = \sum_{k=0}^{m} p_k \sum_{\substack{z\in\{0,1\}^m \\ \mathbf{1}^\top z = k}} \bigotimes_{i=1}^{m} (z_i\mathbf{1} + (1 - z_i)y^*).$$

19

Instead of the whole tensor we now consider every single entry $W_{i_1,\ldots,i_m}$. By carefully expanding every single entry using combinatorics we obtain

$$
\begin{aligned}
\mathbb{E}\left[W_{i_1,\ldots,i_m}\right]\prod_{j=1}^{m} y_{i_j}^* &= \sum_{k=0}^{m} p_k \sum_{\substack{z\in\{0,1\}^m \\ \mathbf{1}^\top z=k}} \prod_{j=1}^{m}\left(z_j+(1-z_j)y_{i_j}^*\right)\prod_{j=1}^{m} y_{i_j}^* \\
&= \sum_{k=0}^{m} p_k \sum_{\substack{z\in\{0,1\}^m \\ \mathbf{1}^\top z=k}} \prod_{j=1}^{m} y_{i_j}^{*z_j} 1^{1-z_j} = \sum_{k=0}^{m} p_k \sum_{\substack{z\in\{0,1\}^m \\ \mathbf{1}^\top z=k}} \prod_{\substack{j=1 \\ z_j=1}}^{m} y_{i_j}^* \\
&= \sum_{k=0}^{m} p_k \sum_{s=\max(0,k-l)}^{\min(k,m-l)} (-1)^s \binom{l}{k-s}\binom{m-l}{s} \\
&= \sum_{k=0}^{m} p_k f(m,l,k),
\end{aligned}
$$

where $l=\sum_{j=1}^{m} \mathbf{1}[y_{i_j}^*=1]$ is the number of positive labels, bounded between 0 and m. The second equality above holds by the fact that we pick every combination of k terms out of $y_{i_1}^*$ through $y_{i_m}^*$, calculate the product of these k terms (either $+1$ or -1), and sum over all possible combinations. Thus by Lemma 2, we obtain

$$
\begin{aligned}
\max_{u\perp\mathbf{1},u\not\perp y^*,\|u\|=1}\left\langle\mathbb{E}\left[W\right],u^{\otimes m}\right\rangle &= \max_{u\perp\mathbf{1},u\not\perp y^*,\|u\|=1}\left\langle\sum_{k=0}^{m} p_k \sum_{\substack{b\in\{0,1\}^m \\ \mathbf{1}^\top b=k}} \bigotimes_{i=1}^{m}(b_i\mathbf{1}+(1-b_i)y^*),u^{\otimes m}\right\rangle \\
&= \max_{u\perp\mathbf{1},u\not\perp y^*,\|u\|=1}\sum_{k=0}^{m} p_k \sum_{\substack{b\in\{0,1\}^m \\ \mathbf{1}^\top b=k}}\left\langle\bigotimes_{i=1}^{m}(b_i\mathbf{1}+(1-b_i)y^*),u^{\otimes m}\right\rangle \\
&= \max_{u\perp\mathbf{1},u\not\perp y^*,\|u\|=1}\sum_{k=0}^{m} p_k \sum_{\substack{b\in\{0,1\}^m \\ \mathbf{1}^\top b=k}}(u^\top\mathbf{1})^k(u^\top y^*)^{m-k} \\
&\overset{(a)}{=} p_0(u^\top y^*)^m \\
&\leq p_0(\|u\|^2\|y^*\|^2)^{m/2} \\
&= p_0 n^{m/2},
\end{aligned}
\tag{12}
$$

where (a) holds because if $k\neq 0$, there exists at least one term of $u^\top\mathbf{1}$, which is 0 because of orthogonality.

Next we characterize $\mathbb{E}\left[V_{i,\ldots,i}^*\right]$. Naturally we consider two cases. If y_i^* is positive, the number of positive labels l in $y_i^*,y_{i_2}^*,\ldots,y_{i_m}^*$ is at least 1, and by definition of $V_{i,\ldots,i}^*$,

$$
\begin{aligned}
\mathbb{E}\left[V_{i,\ldots,i}^*\right] &= \mathbb{E}\left[\sum_{i_2,\ldots,i_m} W_{i,i_2,\ldots,i_m} y_i^* y_{i_2}^*,\ldots,y_{i_m}^*\right] \\
&= \left(\frac{n}{2}\right)^{m-1}\sum_{l=1}^{m}\binom{m-1}{l-1}\sum_{k=0}^{m} p_k f(m,l,k).
\end{aligned}
$$

On the other hand if y_i^* is negative, the number of positive labels l in $y_i^*, y_{i_2}^*, \ldots, y_{i_m}^*$ is at most $m - 1$. Similarly we obtain

$$
\begin{aligned}
\mathbb{E}\left[V_{i,\ldots,i}^*\right] &= \mathbb{E}\left[\sum_{i_2,\ldots,i_m} W_{i,i_2,\ldots,i_m} y_i^* y_{i_2}^* \cdots y_{i_m}^*\right] \\
&= \left(\frac{n}{2}\right)^{m-1} \sum_{l=0}^{m-1} \binom{m-1}{l} \sum_{k=0}^{m} p_k f(m,l,k) .
\end{aligned}
$$

We want to highlight that, in either case, we have the lower bound

$$
\mathbb{E}\left[V_{i,\ldots,i}^*\right] \geq \left(\frac{n}{2}\right)^{m-1} F(m,p) , \tag{13}
$$

for every $i \in [n]$. Then, combining (12) and (13), we obtain the following lower bound for (9)

$$
\begin{aligned}
\min_{u \perp \mathbf{1}, u \nparallel y^*, \|u\|=1} \langle \mathbb{E}[V^* - W], u^{\otimes m} \rangle &\geq \min_{u \perp \mathbf{1}, u \nparallel y^*, \|u\|=1} \sum_i \mathbb{E}\left[V_{i,\ldots,i}^*\right] u_i^m - p_0 n^{m/2} \\
&\geq \min_{u \perp \mathbf{1}, u \nparallel y^*, \|u\|=1} \left(\frac{n}{2}\right)^{m-1} F(m,p) \sum_i u_i^m - p_0 n^{m/2} \\
&\geq \left(\frac{n}{2}\right)^{m-1} F(m,p) n^{1-m/2} - p_0 n^{m/2} \\
&\geq n^{m/2}(2^{1-m} F(m,p) - p_0) , \tag{14}
\end{aligned}
$$

where the second to last inequality holds because $\sum_i u_i^m$ takes the minimum value when $|u_i| = \frac{1}{\sqrt{n}}$.

After deriving lower bounds for each of the three terms, we only need to show $(10) + (11) + (14)$ is greater than 0 with high probability. To do so we can simply divide (14) into two equal parts of size $\frac{1}{2} n^{m/2}(2^{1-m} F(m,p) - p_0)$, and show that $\min_i(V_{i,\ldots,i}^* - \mathbb{E}\left[V_{i,\ldots,i}^*\right]) + \frac{1}{2} n^{m/2}(2^{1-m} F(m,p) - p_0) > 0$ and $-\lambda_{\text{tmax}}(W - \mathbb{E}[W]) + \frac{1}{2} n^{m/2}(2^{1-m} F(m,p) - p_0) > 0$.

To show (10) is bounded below with high probability, we use Hoeffding's inequality in our proof. Note that $\mathbb{P}\left\{V_{i,\ldots,i}^* - \mathbb{E}\left[V_{i,\ldots,i}^*\right] \leq -t\right\} \leq \mathbb{E}\left[\exp\left(-\frac{2t^2}{(2B)^2 n^{m-1}}\right)\right] = \exp\left(-\frac{t^2}{2B^2 n^{m-1}}\right)$. By a union bound, we obtain $\mathbb{P}\left\{\min_i(V_{i,\ldots,i}^* - \mathbb{E}\left[V_{i,\ldots,i}^*\right]) \leq -t\right\} \leq n \exp\left(-\frac{t^2}{2B^2 n^{m-1}}\right)$. Setting $t = \frac{1}{2} n^{m/2}(2^{1-m} F(m,p) - p_0)$ leads to

$$
\begin{aligned}
\mathbb{P}\left\{\min_i(V_{i,\ldots,i}^* - \mathbb{E}\left[V_{i,\ldots,i}^*\right]) \leq -\frac{1}{2} n^{m/2}(2^{1-m} F(m,p) - p_0)\right\} &\leq n \exp\left(-\frac{n(2^{1-m} F(m,p) - p_0)^2}{8B^2}\right) \\
&\leq c_0 n^{-1} , \tag{15}
\end{aligned}
$$

and the last inequality holds given that $\frac{(2^{1-m} F(m,p) - p_0)^2}{B^2} \geq 16 \frac{\log n}{n} - 8 \frac{\log c_0}{n}$.

To show (11) is bounded below with high probability, we use the result of Lemma 3. Note that $\mathbb{P}\left\{-\lambda_{\text{tmax}}(W - \mathbb{E}[W]) \leq -t\right\} \leq \mathbb{P}\left\{-\|W - \mathbb{E}[W]\|_{\text{F}}^2 \leq -t^2\right\} \leq t^{-2} \mathbb{E}\left[\|W - \mathbb{E}[W]\|_{\text{F}}^2\right] \leq$

$\frac{\sigma^2}{t^2}$. Setting $t = \frac{1}{2}n^{m/2}(2^{1-m}F(m,p) - p_0)$ leads to

$$\mathbb{P}\left\{-\lambda_{\text{tmax}}(W - \mathbb{E}[W]) \leq -\frac{1}{2}n^{m/2}(2^{1-m}F(m,p) - p_0)\right\} \leq \frac{4\sigma^2}{n^m(2^{1-m}F(m,p) - p_0)^2}$$
$$\leq c_1 n^{-1}, \qquad (16)$$

and the last inequality holds given that $\frac{(2^{1-m}F(m,p) - p_0)^2}{\sigma^2} \geq \frac{4}{c_1}n^{1-m}$.

Combining the results (15) and (16) above, we can see that (10)+(11)+(14) is greater than 0 with probability at least $1 - (c_0 + c_1)n^{-1}$, which means the probability of $\lambda_\mathbf{1}(V^* - W) > 0$ is at least $1 - (c_0 + c_1)n^{-1}$. This completes our proof. ∎

Appendix C. NP-hardness of Positive Semidefiniteness of a Symmetric Tensor is Open

We want to highlight that since program (3) is convex, its correctness can be checked numerically by using a projected gradient descent solver, after relaxing the problem to the positive semidefinite cone. We implement a projected gradient descent solver in Algorithm 1, which uses an optimization algorithm to detect whether a given symmetric tensor is positive semidefinite or not (Han, 2013) as a subroutine. (Details about our algorithm in the next section.)

We believe the NP-hardness of checking whether a symmetric tensor is in the Carathéodory symmetric tensor cone is an open problem. Furthermore, we believe the problem is open even if we relax it to the positive semidefinite tensor cone, for any symmetric tensor S subject to the constraint $S_{\sigma_2^{n,m}} = 1$ as in program (3). It is claimed in Hillar and Lim (2013), Theorem 11.2, that deciding whether a symmetric 4-th order is nonnegative definite is NP-hard. However, by looking at their proof right before Theorem 11.2, their definition of tensor symmetry is very different from our definition. In Hillar and Lim (2013), the authors construct a 4-th order n-dimensional tensor S from a symmetric matrix $A = (a_{ij})$, such that $S_{i,j,i,j} = S_{j,i,j,i} = a_{ij}$ for all $i, j \in [n]$, and all other entries are assigned to be 0. In this way, checking whether tensor S is positive semidefinite (whether $\sum_{ij} a_{ij}x_i^2 x_j^2 \geq 0$ for all $x \in \mathbb{R}^n$) is equivalent to checking whether matrix A is copositive (whether $\sum_{ij} a_{ij}z_i z_j \geq 0$ for all $z \in [0, \infty)^n$), which is known to be a NP-hard problem (Burer, 2012). As a result, the author claims that checking positive semidefiniteness of tensor fulfilling such symmetry conditions is NP-hard. However in our case, we have $S_{i,i,j,j} = S_{i,j,j,i} = S_{i,j,i,j} = S_{j,j,i,i} = S_{j,i,i,j} = S_{j,i,j,i} = a_{ij} = 1$, according the constraint $S_{\sigma_2^{n,m}} = 1$ in program (3). As a result, certifying copositivity is easy because $\sum_{ij} a_{ij}x_i^2 x_j^2 = \sum_{ij} x_i^2 x_j^2 \geq 0$ is trivially true, which means that the NP-hardness reduction above does not work for our constraint $S_{\sigma_2^{n,m}} = 1$.

Appendix D. Proofs of High-order Example Models

We first provide the following lemma, which characterize the connection between the expectation weights α and the coefficient parameter p.

Lemma 16 *Let $\mathbb{E}[W_{i_1,\ldots,i_m}] = \alpha_l$, if l out of $y_{i_1}, \ldots, y_{i_m}$ are equal to $+1$. We denote $\alpha = (\alpha_0, \ldots, \alpha_m)$ as the vector of expectation weights of W. Let $L \in \mathbb{R}^{(m+1)\times(m+1)}$, where*

$L_{ij} = \sum_{s=\max(0,j-i)}^{\min(j-1,m-i+1)} (-1)^{i+s-1} \binom{i-1}{j-s-1} \binom{m-i+1}{s}$. *Then for any α, we have $Lp = \alpha$ and $L^{-1}\alpha = p$. We say L is the linear transformation between the expectation weights α and the coefficient parameter p.*

Proof This can be shown by looking at each entry in $\mathbb{E}[W]$. Note that

$$\mathbb{E}[W_{i_1,\ldots,i_m}] = \sum_{k=0}^{m} p_k \sum_{\substack{z \in \{0,1\}^m \\ \mathbf{1}^\top z = k}} \prod_{j=1}^{m} \left(z_j + (1-z_j)y_{i_j}^* \right)$$

$$= \sum_{k=0}^{m} p_k \sum_{\substack{z \in \{0,1\}^m \\ \mathbf{1}^\top z = k}} \prod_{j=1}^{m} y_{i_j}^{*(1-z_j)}$$

$$= \sum_{k=0}^{m} p_k \sum_{s=\max(0,k-l)}^{\min(k,m-l)} (-1)^{l+s} \binom{l}{k-s} \binom{m-l}{s},$$

where $l = \sum_{j=1}^{m} \mathbf{1}[y_{i_j}^* = 1]$ is the number of positive labels. Note that both k and l are bounded between 0 to m, where as i and j in the statement are between 1 and $m+1$. Reparameterizing $k = j - 1$ and $l = i - 1$, we obtain the expression of L_{ij}. ∎

As a side note, in terms of the expectation weights vector α, the closed form of $F(m, L^{-1}\alpha)$, as defined in the proof of Theorem 11, is $\min(\langle \Xi_m, (\alpha_0, \ldots, \alpha_{m-1}) \rangle, \langle \Xi_m, (\alpha_1, \ldots, \alpha_m) \rangle)$, where $\Xi_m \in \mathbb{R}^n$ is the m-th row of Pascal's triangle.

We now proceed to show the example models in Section 5 belong to the class of m-HPMs.

Proof [Proof of Proposition 12] We first show Model 1 belongs to the class of m-HPMs. For simplicity we only consider the case of $m = 4$. Let $n = |V|$, W be the affinity tensor, where $W_{i_1 i_2 i_3 i_4} = c(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4})$. Without loss of generality we label $y_i = 1$ if vertex i is from the first group, and -1 otherwise.

Model 1 needs to fulfill (P1), (P2) and (P3). The latter two are trivial because Model 1 is bounded, and it remains to prove (P1). From the assumption of binomial distribution, $\mathbb{E}[W_{i_1 i_2 i_3 i_4}] = T\alpha_l$ if l labels from $y_{i_1}^*, y_{i_2}^*, y_{i_3}^*, y_{i_4}^*$ are the same (w.l.o.g. pick the smaller group). To fulfill (P1), one needs to find a vector $p = (p_0, \ldots, p_4)$ satisfying both

$$\mathbb{E}[W] = \sum_{k=0}^{m} p_k \sum_{\substack{z \in \{0,1\}^m \\ \mathbf{1}^\top z = k}} \bigotimes_{i=1}^{m} (z_i \mathbf{1} + (1-z_i)y^*)$$

and

$$\mathbb{E}[W_{i_1 i_2 i_3 i_4}] = T\alpha_l,$$

for every $i_1, i_2, i_3, i_4 \in [n]$. By Lemma 16 we have the following linear system

$$Lp = \begin{pmatrix} 1 & -4 & 6 & -4 & 1 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & 0 & -2 & 0 & 1 \\ -1 & -2 & 0 & 2 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = T \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{pmatrix}.$$

By solving the linear equation system above we obtain

$$
\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = T \begin{pmatrix} \alpha_0/8 - \alpha_1/2 + 3\alpha_2/8 \\ 0 \\ \alpha_0/8 - \alpha_2/8 \\ 0 \\ \alpha_0/8 + \alpha_1/2 + 3\alpha_2/8 \end{pmatrix}.
$$

Thus, by setting p as above, (P1) is fulfilled and we have shown Model 1 belongs to the class of 4-HPMs. ∎

Proof [Proof of Proposition 13] The procedure is similar to the previous model. Model 2 needs to fulfill (P1), (P2) and (P3). Model 2 is bounded, thus one only needs to prove (P1) by finding the expectation of each single entry in W and solving for p. We use α_l' to denote the expected hypergraph cut size of $(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4})$ if l labels in $y_{i_1}^*, y_{i_2}^*, y_{i_3}^*, y_{i_4}^*$ are the same. Then by Lemma 16, solving p in the linear system

$$
Lp = \begin{pmatrix} 1 & -4 & 6 & -4 & 1 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & 0 & -2 & 0 & 1 \\ -1 & -2 & 0 & 2 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} \alpha_0' \\ \alpha_1' \\ \alpha_2' \\ \alpha_1' \\ \alpha_0' \end{pmatrix},
$$

(P1) is fulfilled and we have shown Model 2 belongs to the class of 4-HPMs. ∎

Proof [Proof of Proposition 14] We now show Model 3 belongs to the class of m-HPMs. For simplicity we only consider the case of $m = 4$. Let $n = |V|$, W be the affinity tensor, where $W_{i_1 i_2 i_3 i_4} = 1$ if $(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}) \in H$, and 0 otherwise.

Model 3 needs to fulfill (P1), (P2) and (P3). The latter two are trivial because Model 3 is bounded, and it remains to prove (P1). Note that the expectation $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right]$ depends on the group assignments $y_{i_1}^*, y_{i_2}^*, y_{i_3}^*$ and $y_{i_4}^*$. From the model definition one can find that

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = q^4 + (1-q)^4$, if all four vertices are from the same group;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = q(1-q)^3 + q^3(1-q)$, if three vertices are from one group, and one is from the other group;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = 2q^2(1-q)^2$, if two vertices are from one group and two are from the other group.

Thus by Lemma 16 we can transform the conditions into the following linear system

$$
Lp = \begin{pmatrix} 1 & -4 & 6 & -4 & 1 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & 0 & -2 & 0 & 1 \\ -1 & -2 & 0 & 2 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} q^4 + (1-q)^4 \\ q(1-q)^3 + q^3(1-q) \\ 2q^2(1-q)^2 \\ q(1-q)^3 + q^3(1-q) \\ q^4 + (1-q)^4 \end{pmatrix}.
$$

Thus, by solving for p in the linear system above, (P1) is fulfilled and Model 3 belongs to the class of 4-HPMs. ∎

Proof [Proof of Proposition 15] In Figure 1 we show three example motifs of size 4. For concreteness let us consider the last motif. This motif has been used to model food chains in the Florida Bay food web (Li and Milenkovic, 2017). In this motif, nodes are considered as species, and the directed edges represent carbon flow, i.e., a directed edge $i \rightarrow j$ can be interpreted as species i consumes species j. Therefore the motif can capture interaction and energy flow between multiple species in the food web.

We now show Model 4 with the last motif (food chain) from Figure 1 belongs to the class of 4-HPMs. Let $n = |V|$, W be the affinity tensor, where $W_{i_1 i_2 i_3 i_4} = 1$ if $(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}) \in H$, and 0 otherwise. Without loss of generality we label $y_i^* = 1$ if vertex i is from group S_1, and -1 otherwise.

Model 4 needs to fulfill (P1), (P2) and (P3). Again Model 4 is bounded, and one only needs to prove (P1) by finding the expectation of each single entry in W and solving for p. Note that the expectation $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right]$ depends on the group assignments $y_{i_1}^*, y_{i_2}^*, y_{i_3}^*$ and $y_{i_4}^*$. By careful analysis of combinations, one can find that

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = \beta_0 := 6\alpha_{1,1}^8 (1 - \alpha_{1,1})^4$, if all four labels are positive;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = \beta_1 := 3\alpha_{1,1}^4 \alpha_{1,2}^3 \alpha_{2,1} (1 - \alpha_{1,1})^2 (1 - \alpha_{2,1})^2 + 3\alpha_{1,1}^4 \alpha_{1,2} \alpha_{2,1}^3 (1 - \alpha_{1,1})^2 (1 - \alpha_{1,2})^2$, if three labels are positive;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = \beta_2 := \alpha_{1,1}^2 \alpha_{1,2}^4 \alpha_{2,2}^2 (1 - \alpha_{2,1})^4 + 4\alpha_{1,1} \alpha_{1,2}^2 \alpha_{2,1}^2 \alpha_{2,2} (1 - \alpha_{1,1})(1 - \alpha_{1,2})(1 - \alpha_{2,1})(1 - \alpha_{2,2}) + \alpha_{1,1}^2 \alpha_{2,1}^4 \alpha_{2,2}^2 (1 - \alpha_{1,2})^4$, if two labels are positive;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = \beta_3 := 3\alpha_{1,2}^3 \alpha_{2,1} \alpha_{2,2}^4 (1 - \alpha_{2,1})^2 (1 - \alpha_{2,2})^2 + 3\alpha_{1,2} \alpha_{2,1}^3 \alpha_{2,2}^4 (1 - \alpha_{1,2})^2 (1 - \alpha_{2,2})^2$, if one label is positive;

- $\mathbb{E}\left[W_{i_1 i_2 i_3 i_4}\right] = \beta_4 := 6\alpha_{2,2}^8 (1 - \alpha_{2,2})^4$, if all four labels are negative.

Thus by Lemma 16 we can transform the conditions into the following linear system

$$Lp = \begin{pmatrix} 1 & -4 & 6 & -4 & 1 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & 0 & -2 & 0 & 1 \\ -1 & -2 & 0 & 2 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}.$$

Thus, by solving for p in the linear system above, (P1) is fulfilled and Model 4 belongs to the class of 4-HPMs. ∎

References

Sameer Agarwal, Jongwoo Lim, Lihi Zelnik-Manor, Pietro Perona, David Kriegman, and Serge Belongie. Beyond pairwise clustering. In *2005 IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition (CVPR'05), volume 2, pages 838–845. IEEE, 2005.

Animashree Anandkumar, Rong Ge, Daniel Hsu, and Sham Kakade. A tensor spectral approach to learning mixed membership community models. In *Conference on Learning Theory*, pages 867–881. PMLR, 2013.

E Ballico. An upper bound for the real tensor rank and the real symmetric tensor rank in terms of the complex ranks. *Linear and Multilinear Algebra*, 62(11):1546–1552, 2014.

Boaz Barak and Ankur Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory*, pages 417–445. PMLR, 2016.

Kevin Bello, Chuyang Ke, and Jean Honorio. A thorough view of exact inference in graphs from the degree-4 sum-of-squares hierarchy. In *International Conference on Artificial Intelligence and Statistics*, pages 640–654. PMLR, 2022.

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

Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.

Samuel Burer. Copositive programming. In *Handbook on semidefinite, conic and polynomial optimization*, pages 201–218. Springer, 2012.

Qing Cai, Maoguo Gong, Lijia Ma, Shasha Ruan, Fuyan Yuan, and Licheng Jiao. Greedy discrete particle swarm optimization for large-scale social network clustering. *Information Sciences*, 316:503–516, 2015.

Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3): 1254–1279, 2008.

Michael R Garey, David S Johnson, and Larry Stockmeyer. Some simplified np-complete problems. In *Proceedings of the sixth annual ACM symposium on Theory of computing*, pages 47–63, 1974.

M.R. Garey, D.S. Johnson, and L. Stockmeyer. Some simplified np-complete graph problems. *Theoretical Computer Science*, 1(3):237 – 267, 1976. ISSN 0304-3975. doi: https://doi.org/10.1016/0304-3975(76)90059-1. URL `http://www.sciencedirect.com/science/article/pii/0304397576900591`.

David Gibson, Jon Kleinberg, and Prabhakar Raghavan. Clustering categorical data: An approach based on dynamical systems. *The VLDB Journal—The International Journal on Very Large Data Bases*, 8(3-4):222–236, 2000.

Lars Hagen and Andrew B Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE transactions on computer-aided design of integrated circuits and systems*, 11(9):1074–1085, 1992.

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

Matthias Hein, Simon Setzer, Leonardo Jost, and Syama Sundar Rangapuram. The total variation on hypergraphs-learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, pages 2427–2435, 2013.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):1–39, 2013.

Furong Huang, UN Niranjan, Mohammad Umar Hakeem, and Animashree Anandkumar. Online tensor methods for learning latent variable models. *Journal of Machine Learning Research*, 16:2797–2835, 2015.

Hsin-Chien Huang, Yung-Yu Chuang, and Chu-Song Chen. Affinity aggregation for spectral clustering. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 773–780. IEEE, 2012.

Yuchi Huang, Qingshan Liu, and Dimitris Metaxas.] video object segmentation by hypergraph cut. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1738–1745. IEEE, 2009.

Krzysztof Juszczyszyn, Przemysław Kazienko, and Katarzyna Musiał. Local topology of social network based on motif analysis. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 97–105. Springer, 2008.

Andrew B Kahng. Fast hypergraph partition. In *Proceedings of the 26th ACM/IEEE Design Automation Conference*, pages 762–766, 1989.

Chuyang Ke and Jean Honorio. Federated myopic community detection with one-shot communication. In *International Conference on Artificial Intelligence and Statistics*, pages 4937–4954. PMLR, 2022.

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

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

Marius Leordeanu and Cristian Sminchisescu. Efficient hypergraph clustering. In *Artificial Intelligence and Statistics*, pages 676–684, 2012.

Jure Leskovec and Andrej Krevl. Snap datasets: Stanford large network dataset collection, 2014.

Pan Li and Olgica Milenkovic. Inhomogeneous hypergraph clustering with applications. In *Advances in Neural Information Processing Systems*, pages 2308–2318, 2017.

Hairong Liu, Longin J Latecki, and Shuicheng Yan. Robust clustering as ensembles of affinity relations. In *Advances in neural information processing systems*, pages 1414–1422, 2010.

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

Yuetian Luo and Anru R Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *The Annals of Statistics*, 50(1):584–613, 2022.

ZiYan Luo, LiQun Qi, and YinYu Ye. Linear operators and positive semidefiniteness of symmetric tensor spaces. *Science China Mathematics*, 58(1):197–212, 2015.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594): 824–827, 2002.

James Munkres. *Topology*. Pearson Education, 2014.

Rebecca Nugent and Marina Meila. An overview of clustering applied to molecular biology. In *Statistical methods in molecular biology*, pages 369–404. Springer, 2010.

Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, 59(3):407–463, 2021.

David A Papa and Igor L Markov. Hypergraph partitioning and clustering., 2007.

Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications*, 5:4630, 2014.

Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

Ömer Nebil Yaveroğlu, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj. Revealing the hidden language of complex networks. *Scientific reports*, 4:4547, 2014.

Pingzhi Yuan and Lihua You. Some remarks on p, p0, b and b0 tensors. *Linear Algebra and its Applications*, 459:511–521, 2014.

Byoung-Tak Zhang. Random hypergraph models of learning and memory in biomolecular networks: shorter-term adaptability vs. longer-term persistency. In *2007 IEEE Symposium on Foundations of Computational Intelligence*, pages 344–349. IEEE, 2007.

Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems*, pages 1601–1608, 2007.