arXiv:2302.03236v1 [cs.LG] 7 Feb 2023

高阶结构化预测中的精确推断

Chuyang Ke普渡大学计算机科学系cke@purdue.edu

Jean Honorio普渡大学计算机科学系
jhonorio@purdue.edu

摘要

本文研究高阶结构化预测任务中的推断问题。在马尔可夫随机场的背景下，高阶推断任务的目标是在标签空间上最大化评分函数，而该评分函数可分解为一元势与高阶势之和。我们采用生成模型方法研究高阶推断问题，并提出一种两阶段凸优化算法来实现精确标签恢复。我们还提出了一类与超边扩张相关的新型超图结构性质，这些性质推动了一般高阶推断问题的成功解决。最后，我们通过一种新颖的超图切格型不等式，将算法的性能与超边扩张性质联系起来。

1 引言

结构化预测在过去 20 年中广泛应用于机器学习的多个领域，包括社交网络分析、计算机视觉、分子生物学、自然语言处理（NLP）等。一项常见任务是分配或恢复标签：给定一些可能含噪的观测，目标是为任务中的每个实体输出一个群组标签。在社交网络分析中，这可能是根据用户档案和偏好检测社群 [Kelley et al., 2012]。在计算机视觉中，研究人员希望人工智能判断某个像素属于前景还是背景 [Nowozin et al., 2011]。在生物学中，有时需要根据结构相似性对分子进行聚类 [Nugent and Meila, 2010]。在自然语言处理中，词性标注可能是最广为人知的结构化预测任务之一 [Weiss and Taskar, 2010]。

从方法论的角度来看，上述结构化预测任务中的一种标准方法是利用许多局部结构来恢复全局结构。以社交网络为例。社交网络分析中一个广泛采用的假设是亲和性——档案和偏好相似的用户更有可能成为朋友。直观而言，如果两个用户的亲和度评分较高，结构化预测算法往往会为他们分配相同的标签。类似地，这一思想也可以在马尔可夫随机场（MRF）中得到应用。假设所有实体构成一个无向图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$，那么结构化预测可以视为求解以下推断问题 [Bello 和 Honorio, 2019]:

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

其中，$\mathcal{L}$ 是标签空间，$c_v(l)$ 是将标签 l 分配给节点 v的得分，而 $c_{v_1, v_2}(l_1, l_2)$ 是将标签 l_1 和 l_2 分配给相邻节点 v_1 和 v_2的得分。在马尔可夫随机场和推断

文献中，式（1）中的这两项通常分别称为一元势 和 成对势。上述推断表述可以通过寻找使一元和成对局部得分之和最大化的配置，来恢复全局结构。

然而，在许多现实世界问题中，实体之间的交互方式可能不止成对交互。再以社交网络为例，这次我们关注学术合著网络：许多已发表论文由两位以上作者撰写 [Liu 等人， 2005]。成对结构无法捕捉这类高阶交互。从几何角度看，合著网络不再能用图表示。因此，有必要引入超图来建模高阶结构化预测问题。

本文研究高阶结构化预测问题，其中不再使用成对势，而是考虑高阶势 。采用马尔可夫随机场表述，我们关注如下形式的推断问题：

$$\underset{y\in\mathcal{L}^{|\mathcal{V}|}}{\text{maximize}} \sum_{v\in\mathcal{V},l\in\mathcal{L}} c_v(l)\cdot\mathbb{1}[y_v=l] + \sum_{\substack{e\in\mathcal{E}\\ l_1,\ldots,l_m\in\mathcal{L}\\ e=(v_1,\ldots,v_m)}} c_e(l_1,\ldots,l_m)\cdot\mathbb{1}[y_{v_1}=l_1,\ldots,y_{v_m}=l_m]\,, \qquad (2)$$

其中， m 是推断问题以及超图的阶数（每条超边连接m 个顶点），而 $c_e(l_1, \ldots\ldots, l_m)$ 是将标签 l_1 分配给通过超边 $e \in \mathcal{E}$相连的相邻节点 v_1 至 v_m 时的得分。

1.1 推断作为恢复任务

具有形式（1）中的一元势和成对势的结构化预测和推断问题，已在先前文献中得到研究。Globerson 等人 [2015] 引入了二维网格晶格情形下的标签恢复问题，并分析了近似推断的条件。沿着相同的研究路线，Foster 等人 [2018] 通过允许树分解对该模型进行了推广。另一类问题是精确推断问题，针对该问题，Bello 和 Honorio[2019] 提出了一种凸半定规划（SDP）方法。在这些工作中，标签恢复问题的动机来自一个生成模型，该模型假设存在一个真实标签向量 y^*，并基于标签交互生成（一元和成对的）观测（可能带有噪声）。

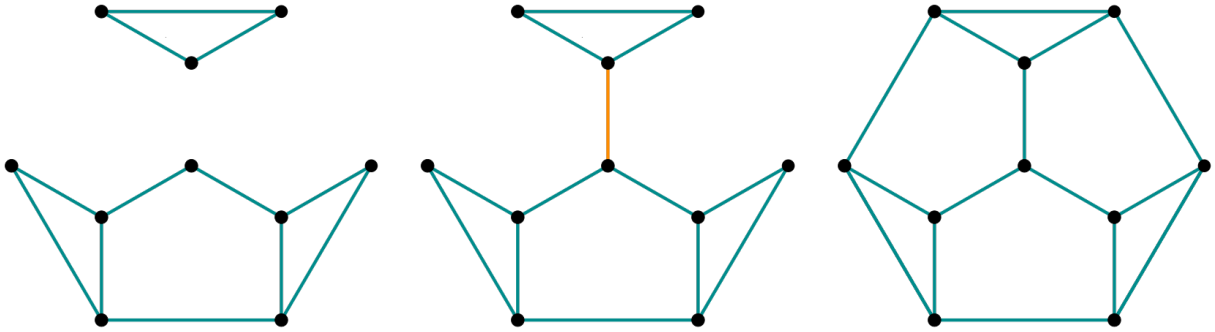
遗憾的是，对于具有高阶势和超图的结构化预测，人们所知甚少。近年来，人们尝试将一些图的性质（包括超图拉普拉斯算子、瑞利商、超边扩张、切格常数等）推广到超图 [Li 和 Milenkovic，2018，Mulas，2021，Yoshida，2019，Chan 等人，2018，Chen 等人，2017，Chang 等人， 2020]。然而，我们花了很长时间才发现，由于结构化预测任务的性质，超图定义必须满足某些属性，而上述工作中的定义无一满足这些属性。这使得设计基于超图的标签恢复算法变得十分困难。此外，上述工作均未对近似推断或精确推断提供保证。

在这项工作中，我们采用生成模型方法研究高阶推断问题。我们分析标签恢复任务，并回答以下问题：

问题 1（标签恢复）。 是否存在某种算法，能够以带噪声的一元观测和高阶局部观测为输入，并正确恢复底层的真实标签?

1.2 推断与结构性质

在马尔可夫随机场推断领域，一个长期以来的核心讨论聚焦于推断可解性与问题的某些结构性质。



(a) 非连通图（边扩展为零）　(b) 具有瓶颈的图（边扩展较小）　(c) 图　没有瓶颈（较大的边扩展）

图1： 图扩展示例。图1a显示一个非连通图。借助理想的算法，我们或许能够恢复上方子图和下方子图中的用户社区，但由于两者之间没有观测到交互，我们无法推断全局结构。图1b用一条橙色边连接了图1a中的两个分量。在这种情况下，橙色边就是“瓶颈”。移除这条橙色边后，图便不再连通。图1c在图1b的基础上又添加了两条边。在这种情况下，每个分量都彼此连通，不存在薄弱的“瓶颈”。边扩展是一种结构性质，用于刻画图中各分量的连通程度。

为说明这一点，我们首先回顾成对推断问题（即形式为 (1) 的问题）中的各类结构性质。Chandrasekaran 等人 [2012] 研究了图的树宽这一结构性质，并证明树宽较低的图是可解的。Schraudolph 和 Kamenetsky[2008] 证明了平面图可以求解。Boykov 和 Veksler [2006] 分析了具有二元标签和子模成对势函数的图。Bello 和 Honorio [2019]证明，对于“良好”的扩展图，或由“较差”的扩展图加上一个 Erdős–Rényi随机图构成的图，可以实现推断。值得强调的是，上述结构性质之间并不能直接比较，也不能相互归约。相反，它们从不同角度刻画推断问题的难度，换句话说，针对的是不同类别的图。

关于推断与结构性质的类似讨论，也见于高阶马尔可夫随机场推断的文献中。例如，Komodakis 和 Paragios [2009]研究了满足每个团对应一个子超图这一性质的超图，并证明可以通过求解线性规划（LP）实现高阶马尔可夫随机场推断。Fix 等人 [2014] 研究了具有局部完备性这一性质的超图上的推断。Gallagher 等人 [2011]分析了高阶马尔可夫随机场推断通过阶数约简实现时的性能与非子模边数量之间的关系。然而，这些工作都没有为近似推断或精确推断提供理论保证。

本文通过分析超边扩张，提出了一类新的超图结构性质。为了获得一些直观认识，我们考虑一个由两个互不连通的子网络组成的社交网络。借助理想的算法，我们可以恢复子网 1 和子网 2 中的用户社区，但由于两个子网之间完全没有互动，我们无法推断全局社群结构（例如，子网 1 和子网 2 中恢复出的社群之间的关系）。一种不那么极端的情况是网络中存在“瓶颈”，也就是说，移除这些瓶颈边会使网络断开。出于类似的原因，可以想象，如果存在噪声，那么对含有瓶颈的网络进行推断可能会很困难。图1对此作了说明。在成对图（2-图）中，这种连通性／瓶颈性质可以通过图的边

扩展（即切格常数）来刻画。要刻画高阶超图中类似的扩展性质颇具挑战，尤其是在希望将这类拓扑性质与精确推断的条件联系起来时。

问题 2（结构性质）。在什么拓扑条件下，我们的标签恢复算法能够以高概率正确运行?

我们的贡献总结。 我们的工作具有很强的理论性。本文提出了一系列新颖的定义并得出了一系列新颖的结果:

- 我们为高阶推断问题提出了一类新的超图结构性质。我们推导出一种新颖的 Cheeger 型不等式，将超图拉普拉斯算子的张量谱隙与 Cheeger 型超图扩张性质联系起来。这些超图结果不仅适用于本文模型的范畴，也有助于从事高阶推断问题研究的学者。

- 我们提出了一种两阶段方法来解决高阶结构化预测问题。我们将标签恢复问题建模为一个高阶组合优化问题，并进一步将其松弛为一种新颖的凸锥形式优化问题。

- 我们仔细分析了锥形式优化问题的 Karush–Kuhn–Tucker（KKT）条件，并推导出实现精确推断所需的充分统计条件和拓扑条件。我们的 KKT 分析保证，只要满足这些条件，所得解就能以高概率达到最优。

2 预备知识

本节将正式定义高阶精确推断问题，并介绍本文中将使用的符号。

我们用小写字体（例如 a、b、u、v）表示标量和向量，用大写字体（例如A、B、C）表示张量。我们用 $\mathbb{R}$ 表示实数集。

对于任意自然数 n，我们用 $[n]$ 表示集合$\{1,\dots,n\}$。

我们用$\mathbf{1}$ 表示全 1 向量。

为清晰起见，我们用上标 (i) 表示序列中第 i 个对象，用下标 j 表示第 j 个元素。我们用 $\circ$ 表示 Hadamard积，并用 $\otimes$ 表示外积。设 $v^{(1)},\dots,v^{(m)} \in \mathbb{R}^n$ 为一列 m 维向量，则 $v^{(1)} \otimes \dots \otimes v^{(m)}$ 是一个阶数为 m、维度为 n（或等价地，形状为 $n^{\otimes m}$）的张量，使得($v^{(1)} \otimes \dots \otimes v^{(m)})_{i_1,\dots,i_m} = v^{(1)}_{i_1} \dots v^{(m)}_{i_m}$.

2.1 张量定义

令 $A \in \mathbb{R}^{n^{\otimes m}}$ 为 m 阶、 n 维实张量。为使论述清晰，本文始终将讨论范围限定在 $m = 2$、6、10、14、.....。虽然也可以考虑其他偶数阶（$m = 4$、8、12 .....），并进行类似分析，但超图定义将涉及更多项，使论文更难阅读。相关讨论见注 3.17。

对称张量在指标的任意排列下保持不变。换言之，若对于任意排列 $\sigma\colon [m] \to [m]$，都有 $A_{\sigma(i_1,\dots,i_m)} = A_{i_1,\dots,i_m}$，则称 A 为对称张量。

我们将形状相同的两个张量 A, B 的内积定义为 $\langle A, B\rangle := \sum_{i_1,\dots,i_m=1}^{n} A_{i_1,\dots,i_m} B_{i_1,\dots,i_m}$。我们将张量 Frobenius 范数定义为 $\|A\|_{\mathrm{F}} := \sqrt{\langle A, A\rangle}$。

如果对于所有 $v \in \mathbb{R}^n$，都有 $\langle A, v^{\otimes m} \rangle \geq 0$，则称对称张量 A 为半正定（PSD）。我们用 $\mathcal{S}_+^{n\,m}$ 表示所有 m 阶、n 维半正定张量构成的凸锥。

$\mathcal{S}_+^{n\,m}$ 的对偶锥是 Caratheodory张量锥 $\mathcal{S}_+^{*\,n\,m}$，其定义为 $\mathcal{S}_+^{*\,n\,m} := \left\{ \sum_{i=1}^{\binom{m+n-1}{m}} v^{(i)\otimes m} \mid v^{(i)} \in \mathbb{R}^n \right\}$。换言之，$\mathcal{S}_+^{*\,n\,m}$ 中的每个张量都是至多 $\binom{m+n-1}{m}$ 个秩一张量之和。$\mathcal{S}_+^{n\,m}$ 和 $\mathcal{S}_+^{*\,n\,m}$ 互为对偶 [Ke and Honorio, 2022]。

对于任意张量 $A \in \mathbb{R}^{n^{\otimes m}}$，我们使用变分刻画来定义其最小张量特征值 $\lambda_{\min}(A)$（或等价地 $\lambda_1(A)$），使得 $\lambda_{\min}(A) := \min_{v\in\mathbb{R}^n, \|v\|=1} \langle A, v^{\otimes m} \rangle$。类似地，我们将其第二小张量特征值 $\lambda_2(A)$ 定义为 $\lambda_2(A) := \min_{v\in\mathbb{R}^n, \|v\|=1, v\perp v^{(1)}} \langle A, v^{\otimes m} \rangle$，其中 $v^{(1)}$ 是对应于 $\lambda_{\min}(A)$ 的特征向量。

我们将 $\sigma^{n,2^m}$ 记作形如 $\sigma(i_1, i_1, i_2, i_2, \ldots, i_{m/2}, i_{m/2})$ 的 m 元组的指标集，其中任意排列 $\sigma : [m] \to [m]$ 和 $i_j \in [n]$ 均成立。直观地说，在 $\sigma^{n,2^m}$ 的每个元组中，每个指标都重复出现偶数次。我们用 $\bar{\sigma}^{n,2^m}$ 表示集合 $\{(i_1, \ldots, i_m) \mid (i_1, \ldots, i_m) \notin \sigma^{n,2^m}$，且至少有一个指标重复两次$\}$。换句话说，$\bar{\sigma}^{n,2^m}$ 是 $\sigma^{n,2^m}$ 的补集，即排除所有指标均不重复的情况。

2.2 高阶推断任务

我们考虑这样一项任务：根据一组观测 X 和 z，预测一组 n 顶点标签 $y^* = (y_1^*, \ldots\ldots, y_n^*)$，其中 $y_i^* \in \{+1, -1\}$。X 和 z 是由某个底层的 m 均匀超图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 生成的含噪观测。具体而言，$\mathcal{V}$ 是顶点（节点）的集合，其中 $|\mathcal{V}| = n$；而 $\mathcal{E}$ 是超边的集合。

对于每个可能的 m 个顶点元组 $e = (i_1, \ldots\ldots, i_m)$，若 $e \in \mathcal{E}$，则超边观测 $X_{i_1,\ldots,i_m}$（以及所有对应的对称项 $X_{\sigma(i_1,\ldots,i_m)}$）以概率 $1-p$ 独立地采样为 $y_1^* \cdots y_m^*$，以概率 p 采样为 $-y_1^* \cdots y_m^*$。若 $e \notin \mathcal{E}$，则将 $X_{i_1,\ldots,i_m}$ 设为 0。

对于每个节点 $v_i \in \mathcal{V}$，节点观测 z_i 以概率 $1-q$ 独立地采样为 y_i^*，以概率 q 采样为 $-y_i^*$。

现在我们总结生成模型。

定义 2.1（带部分观测的高阶结构化预测）。**未知量**：真实节点标注向量 $y^* = (y_1^*, \ldots\ldots, y_n^*)$。**观测量**：部分噪声超边观测张量 $X \in \{-1, 0, +1\}^{n^{\otimes m}}$。噪声节点标签观测向量 $z \in \{-1, +1\}^n$。**任务**：根据观测 X 和 z 推断并恢复正确的节点标注向量 y^*。

3 超图结构性质与 Cheeger 型不等式

本节介绍一系列针对超图的新颖 Cheeger 型分析。这使我们能够通过图本身的拓扑超边扩张来刻画超图拉普拉斯算子的谱间隙。本节中的超图定理具有普遍性，并不限于我们推断任务中涉及的特定模型。据我们所知，以下高阶定义和结果均为原创。所有引理和定理中省略的证明均可见附录 A。

3.1 超图拓扑

我们首先介绍必要的超图拓扑定义。

定义 3.1（诱导超顶点）。给定一个 m 一致超图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$，我们用

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

表示诱导超顶点的集合。我们用 $N := |\mathcal{H}| = \binom{n}{m/2}$ 表示其基数。

定义 3.2（超顶点集的边界）。 对于任意超顶点集 $S \subset \mathcal{H}$，我们将其边界集记为

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

注意， ∂S 是由 m 阶超边和非边组成的集合。

定义 3.3（超顶点集的超边扩展）。 对于任意超顶点集 $S \subset \mathcal{H}$，我们将集合 S 的超边扩展记为

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

定义 3.4（超图的超边扩展）。 给定一个 m-均匀超图$\mathcal{G}$ =($\mathcal{V}$,$\mathcal{E}$)，其诱导超顶点为 $\mathcal{H}$，我们将超图$\mathcal{G}$的超边扩展记为

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

我们也将 $\phi_{\mathcal{G}}$ 称为该超图的切格常数。

3.2 超图拉普拉斯算子

本节将介绍与超图拉普拉斯算子相关的定义。

定义 3.5 (ζ-函数)。 $\zeta : \mathbb{R}^m \to \mathbb{R}$ 是一个定义为如下形式的函数

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

定义 3.6 (超图拉普拉斯算子)。 给定一个 m-均匀超图 $\mathcal{G}$ = ($\mathcal{V}$,$\mathcal{E}$)，我们用 $L \in \mathbb{R}^{n^{\otimes m}}$ 表示其拉普拉斯张量，该张量满足

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

定义 3.7 (瑞利商)。 对于任意超图拉普拉斯算子 L 和非零向量 $v \in \mathbb{R}^n$，瑞利商 $R_L(v)$ 定义为

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

定义 3.8（带符号拉普拉斯张量）。给定一个带有符号向量 $y \in \{-1, +1\}^n$ 的 m一致超图 $\mathcal{G}$ = ($\mathcal{V}$, $\mathcal{E}$)，我们用 $L_y \in \mathbb{R}^{n^{\otimes m}}$ 表示其带符号拉普拉斯张量，它满足

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

注 3.9。超图拉普拉斯张量 L 可以视为带符号拉普拉斯张量 L_y，其中取 $y = \mathbf{1}$。

下面介绍超图拉普拉斯算子和瑞利商的一些重要性质。

引理 3.10 (拉普拉斯特征对)。 对于任意超图拉普拉斯算子 L，$\mathbf{1}$ 是 L 的一个特征向量，其最小特征值为0。类似地，对于任意带符号超图拉普拉斯算子 L_y， y 是 L_y 的一个特征向量，其特征值为0。

引理 *3.10* 的证明。 这可直接由 ζ-函数的定义以及定义 3.6、3.8 得出。

□

引理 3.11 (缩放不变性)。 对于任意非零的 $\alpha \in \mathbb{R}$，有

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

引理 *3.11* 的证明。

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

□

引理 3.12 (缩放不变性)。 对于任意 $\delta \in \mathbb{R}$ 和 $v \in \mathbb{R}^n$， $v \perp \mathbf{1}$，有

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

引理 *3.12* 的证明。注意

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

其中，不等式成立是因为

$$\|v+\delta\mathbf{1}\|^m = \left(\sum_i (v_i+\delta)^2\right)^{m/2}$$

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

□

引理 3.13 (带符号瑞利商下界)。 对于任意拉普拉斯算子为 L 、带符号拉普拉斯矩阵为 L_y 的超图，以及任意 $\delta \in \mathbb{R}$ 和 $v \in \mathbb{R}^n$， $v \perp y$，我们有

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

引理 3.14 (度张量的存在性)。 对于任意超图拉普拉斯算子 L，令 A 表示相应的对称邻接张量，使得 $A_{i_1,\ldots,i_m} = 1$ 当且仅当 $(i_1, \ldots, i_m) \in E$。于是，存在一个高阶度张量 D，满足：*1)* 当 $i_1, \ldots, i_m$ 各不相同时，$D_{i_1,\ldots,i_m} = 0$；*2)* 对每个向量 $v \in \mathbb{R}^n$，$\langle Dv - Av^{\otimes m} i \rangle = \langle L v^{\otimes m} i \rangle$。

3.3 高阶 Cheeger 型不等式

在这里，我们将前两节联系起来，并给出我们提出的新型高阶 Cheeger 型不等式。

定理 3.15 (高阶 Cheeger 型不等式)。 对于任意 m-一致超图 $G = (V, E)$及其拉普拉斯算子 L，有

$$
\lambda_2(L) \geq \phi_G^m \, .
$$

注 3.16。 这里我们想讨论所选择的超图拉普拉斯算子以及相关的高阶定义。尽管已有多项研究尝试将一些图的性质（包括超图拉普拉斯算子、瑞利商、超边扩张、切格常数等）推广到超图 [Li 和 Milenkovic, 2018, Mulas, 2021, Yoshida, 2019, Chan 等人, 2018, Chen 等人, 2017, Chang 等人，2020], ，我们想强调的是，上述方法均不满足我们的高阶推断任务所需的要求。

为了获得一组适用于我们推断任务的超图定义，我们需要满足本节中的所有引理。具体来说，以下是必要条件：

- 为了满足引理 3.11， ζ-函数（定义 3.5）括号中的每一项的系数都必须为 1 或 -1。

- 为满足引理 3.12， ζ-函数（定义 3.5）与瑞利商分母中的范数（定义 3.7）必须具有相同的幂次。此外， ζ-函数（定义 3.5）括号中各项的系数之和必须为 0。结合最后这一条件，正项与负项的数量必须相等。

- 为满足引理 3.14，ζ-函数（定义 3.5）括号中的负项数量必须为奇数，这样在那些不含重复指标的项中，$\langle -A, v^{\otimes m}\rangle$ 就可以与$\langle L, v^{\otimes m}\rangle$ 相消。

- 为了满足引理 3.10，瑞利商必须在 $v = \mathbf{1}$ 处取得最小值。

此时应该已经很清楚，我们不能随意采用先前文献中的某个定义。我们的超图定义经过了精心构造和选择，以满足上述所有必要条件。

注 3.17。关于阶数的选择 m，请注意，若 m 为奇数，则在当前表述下，标签恢复任务将无法实现，因为瑞利商将无下界（而非取得最小值 0）。从注 3.16 可以看出，若保留当前的 ζ-函数定义，则唯一可能的 m 为 2、6、10、14、..... ζ-函数的定义可以推广到其他偶数阶：对于每个 m元组 $\mathcal{I}$，通过遍历所有可能符号的排列，对多个m次幂项取平均。为使表述清晰，我们保留当前较简单的定义，并聚焦于 $m = 2$、6、10、14、.....的情形。

4 真实标签的精确恢复

本节将介绍一种推断算法，用于恢复模型中潜在的真实标签。为此，我们采用两阶段方法。第一阶段仅利用从 X观测到的超边信息，从而将解空间缩小到两个可能的解。第二阶段借助从 z观测到的节点信息，推断出节点的正确标签赋值。

4.1 第一阶段：推断至排列

我们首先考虑以下组合问题

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

这种优化形式的问题在于它不是凸的，因此难以分析且难以求解。作为替代，我们考虑以下松弛版本

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

或者，我们将最后两个约束表示为 $-1 \le Y_{\bar{\sigma}^{n,}{}_{2}{}^{m}} \le 1$，以及 $Y_{\sigma^{n,}{}_{2}{}^{m}} = 1$。回想一下，$\mathcal{S}_+^{*\,n\,m}{}_{,}{}_{,}$ 是秩一张量的凸锥。这样做的动机是，我们使用秩一张量Y 而不是外积 $y^{\otimes m}$，从而使目标函数和约束都成为凸的。我们有 $-1 \le Y_{\bar{\sigma}^{n,}{}_{2}{}^{m}} \le 1$，因为 y 的乘积只能是 -1 或 $+1$；并且我们有 $Y_{\sigma^{n,}{}_{2}{}^{m}} = 1$，因为如果每个 y_i 都重复偶数次，我们就知道该乘积必定为+1。

我们还需证明程序（4）的正确性。我们关注的是确定这样一个范围：在该范围内，（4）能够返回精确秩一张量解 $Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$。现在我们给出关于推断的主定理。证明见附录 B。

定理 4.1 (从超边观测进行推断)。 对于一个底层超图为 $G = (V,E)$ 且超边观测为 X 的 m 阶结构化预测模型，可通过凸优化程序 (4) 以至少 $1 - \epsilon_1(\phi_{\mathcal{G}}, n, p)$ 的概率恢复秩一张量解$Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$，其中

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

注释 4.2。 一个自然的问题是，在什么样的拓扑和统计条件下，我们可以从定理 4.1 得到高概率保证？我们观察到，如果底层超图的切格常数较大，或者噪声水平较低，那么恢复更有可能成功。在第 5 节中，我们分析了几类具有良好扩张性质（切格常数较大）的有趣超图，因此可以保证以高概率恢复。

4.2 第二阶段：精确推断

在上一节中，我们为秩一张量解 $Y^* := y^{*\otimes m} = (-y^*)^{\otimes m}$ 建立了高概率推断保证。通过因式分解步骤，我们知道 y^* 或 $-y^*$ 是正确的标签向量。本节的目标是利用节点观测 z 确定正确的标签赋值。

定理 4.3。 设 $y \in \{y^*, -y^*\}$。正确的标签向量 y^* 可以通过以下程序恢复

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

以至少 $1 - \epsilon_2(n, q)$ 的概率，其中

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

定理 *4.3* 的证明。 应用霍夫丁不等式，我们得到

$$\begin{aligned}\mathbb{P}\left\{z^\top y^* \le -z^\top y^*\right\} &= \mathbb{P}\left\{z^\top y^* \le 0\right\} \\ &\le \mathrm{e}^{-(1-2q)^2 n/2}.\end{aligned}$$

□

推论 4.4。 结合定理 *4.1* 和定理 *4.3* 的结果，我们得到，正确标签向量 $y = y^*$ 的精确推断可以以至少 .....的概率实现

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

推论 *4.4* 的证明。 对定理 4.1 和定理 4.3 应用并集界。 □

注记 4.5。 给定 p，$q < 0.5$，我们观察到，只要 $n \to \infty$，就有 $\epsilon_2(n, q) \to 0$ （指数衰减）。因此，能否获得形如 $1 - O(n^{-1})$ 的高概率保证，只取决于 $\phi_{\mathcal{G}}$ 的阶数，也就是底层超图的拓扑结构。在第 5 节中，我们研究具有良好扩张性质的超图，这些超图使我们能够为精确推断获得高概率保证。

5 具有良好扩张性质的超图示例

本节中，我们考虑一些具有良好扩张性质的超图类别示例，并展示它们如何使我们的精确推断算法获得高概率保证。

我们首先分析完全超图。

定义 5.1 （完全超图）。 若一个 m-均匀超图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 对每个 m 个顶点的元组 $e = (i_1, \ldots\ldots, i_m)$ 都有 $e \in \mathcal{E}$，则称其为完全超图。

命题 5.2 （完全超图的扩张性质）。 对于任意完全超图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$，有

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

证明。 回忆一下，$\mathcal{H}$ 是诱导超顶点的集合。根据定义 3.3，对于任意 $S \subset \mathcal{H}$ with$0 < |S| \leq N/2$，我们有

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

按照定义 3.4 取最小值，得到 $\phi_{\mathcal{G}} = \frac{N}{2}$。 □

推论 5.3 (完全超图中的精确推断)。 设 $G = (V,E)$ 是一个完全超图。假设 $p, q < 0.5$。那么，当 n 足够大时，采用所提出的两阶段方法可以实现精确推断，且其概率趋于 1 。

证明。 注意，对于完全超图，我们有

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

将 $\phi_{\mathcal{G}}$ 代入 (8)，只要 n 大于某个常数 c_0，我们就有

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

□

接下来，我们聚焦于正则扩展器的情形。

定义 5.4 (d-正则扩展器)。 一个 m-均匀超图 $\mathcal{G} = (\mathcal{V},\mathcal{E})$ 若具有常数 $c > 0$，且对于任意满足 $0 < |S| \leq N/2$ 的诱导超顶点集 $S \subset \mathcal{H}$，都有以下性质，则称其为一个 d-正则扩张超图:

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

命题 5.5 (扩张超图)。 对于任意具有常数 $c > 0$ 的 d-正则扩张超图 $G = (V, E)$，我们有

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

证明。 根据定义 3.3，对于任意满足 $0 < |S| \leq N/2$ 的 $S \subset \mathcal{H}$，我们有

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

像定义 3.4 中那样取最小值，我们得到 $\phi_{\mathcal{G}} = cd$。 □

推论 5.6 (扩张超图中的精确推断)。 设 $G = (V,E)$ 是一个常数为 $c > 0$ 的 d-正则扩张超图。假设 $p, q < 0.5$。那么，当 n 足够大时，使用所提出的两阶段方法可以实现精确推断，且其概率趋近于1 ，如果

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

证明。 首先注意，(5) 中的指数项是主导因素。将 $\phi_{\mathcal{G}} = cd$代入 (5) 的第一项，我们希望确保

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

对于某个常数 $c_0 > 0$。注意， $|\mathcal{E}| \in O(n^m)$。因此， d 的一个充分条件是$d \in \Omega(n \cdot (\log n)^{1/2m})$。将 d 代入另一项也满足 $O(n^{-1})$ 界。 □

5.1 模拟结果

我们在一个阶数为 $m = 6$ 的超图上测试所提出的方法；见图 2。固定节点数为 $n = 10$。我们聚焦于第一阶段，并检查经过符号置换后可以恢复多少标签。我们实现了一个受 Ke 和 Honorio [2022], Han [2013] 启发的张量投影梯度下降求解器。对于每种设置，我们运行 20 次迭代。结果表明，如果噪声水平 p 较低且超图切格常数 $\phi_{\mathcal{G}}$ 较大，所提出的算法表现良好，并能恢复底层群结构。这与我们在定理 4.1 中的理论发现一致。

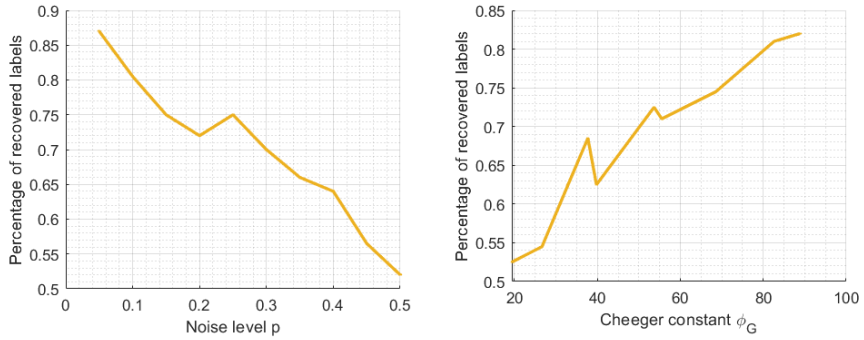


图 2： 不同噪声水平 p （上）和切格常数 ϕ_G （下）下的模拟实验。我们的结果表明，当噪声水平 p 较小且超图切格常数 $\phi_{\mathcal{G}}$ 较大时，所提出的算法表现良好，并能恢复潜在的群结构。

参考文献

Kevin Bello 和 Jean Honorio。结构化预测中的精确推断。神经信息处理系统进展，32，2019年。

Yuri Boykov 和 Olga Veksler。计算机视觉中的图割与图形学：理论与应用。载于计算机视觉数学模型手册，第 79–96 页。施普林格，2006 年。

T-H Hubert Chan、Anand Louis、Zhihao Gavin Tang 和 Chenzi Zhang。《超图拉普拉斯的谱性质与近似算法》。*ACM* 会刊（*JACM*），65(3):1–48，2018。

Venkat Chandrasekaran、Nathan Srebro 和 Prahladh Harsha。图模型中的推断复杂度。*arXiv* 预印本 *arXiv:1206.3240*，2012。Jingya Chang、Yannan Chen、Liqun Qi 和 Hong Yan。利用一种新型拉普拉斯张量进行超图聚类及其在图像处理中的应用。*SIAM*成像科学杂志，13(3):1157–1178，2020。Yannan Chen、Liqun Qi 和 Xiaoyan Zhang。用于超图划分的拉普拉斯张量费德勒向量。*SIAM*科学计算杂志，39(6):A2508–A2537，2017。Alexander Fix、Aritanan Gruber、Endre Boros 和 Ramin Zabih。一种基于超图的高阶二元马尔可夫随机场约简方法。*IEEE*模式分析与机器智能汇刊，37(7):1387–1395，2014。Dylan Foster、Karthik Sridharan 和 Daniel Reichman。具有成对测量和辅助信息的稀疏图中的推断。载于国际人工智能与统计学会议，第1810–1818页。PMLR，2018。Andrew C Gallagher、Dhruv Batra 和 Devi Parikh。马尔可夫随机场阶数约简中的推断。载于*2011*年计算机视觉与模式识别会议，第1857–1864页。IEEE，2011。Amir Globerson、Tim Roughgarden、David Sontag 和 Cafer Yildirim。结构化预测中的推断有多难？载于国际机器学习会议，第2181–2190页。PMLR，2015。Lixing Han。一种求解偶数阶对称张量实特征值的无约束优化方法。数值代数、控制与优化，3(3):583–599，2013。Chuyang Ke 和 Jean Honorio。利用一种新型凸张量锥松弛精确划分高阶模型。机器学习研究杂志，23(284):1–28，2022。Stephen Kelley、Mark Goldberg、Malik Magdon-Ismail、Konstantin Mertsalov 和 Al Wallace。社交网络中社群的定义与发现。载于复杂网络优化手册，第139–168页。施普林格，2012。Nikos Komodakis 和 Nikos Paragios。超越成对能量：高阶马尔可夫随机场的高效优化。载于*2009*年*IEEE*计算机视觉与模式识别会议，第2985–2992页。IEEE，2009。Pan Li 和 Olgica Milenkovic。次模超图：p-拉普拉斯算子、Cheeger不等式与谱聚类。载于国际机器学习会议，第3014–3023页。PMLR，2018。Xiaoming Liu、Johan Bollen、Michael L Nelson 和 Herbert Van de Sompel。数字图书馆研究社群中的合著网络。信息处理与管理，41(6):1462–1480，2005。Raffaella Mulas。均匀超图的切格割。图与组合，37(6):2265–2286，2021。Sebastian Nowozin、Christoph H Lampert 等。计算机视觉中的结构化学习与预测。基础与趋势® 计算机图形学与计算机视觉，6(3–4):185–365，2011。

Rebecca Nugent 和 Marina Meila。《分子生物学中的统计方法》中的聚类概述，第369–404页，2010年。Nicol Schraudolph 和 Dmitry Kamenetsky。平面伊辛模型中的高效精确推断。《神经信息处理系统进展》，第21卷，2008年。David Weiss 和 Benjamin Taskar。结构化预测级联。载于《第十三届人工智能与统计国际会议论文集》，第916–923页。JMLR研讨会与会议论文集，2010年。Yuichi Yoshida。子模变换的Cheeger不等式。载于《第三十届ACM-SIAM离散算法年会论文集》，第2582–2601页。SIAM，2019年。

超图结构性质与 Cheeger 型不等式的证明

为便于说明，在以下证明中，我们使用

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

用于表示拉普拉斯算子 L瑞利商的分子部分，并且

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

用于表示分母部分。

引理 *3.13* 的证明。 对于分子，我们有

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

对于分母，我们有

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

□

引理 *3.14* 的证明。 我们的目标是构造一个张量 D，使其满足 $\langle D-A, v^{\otimes m}\rangle = \langle L, v^{\otimes m}\rangle =$，对于任意向量 $v\in\mathbb{R}^n$。换言之，我们要求

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

一个关键的观察是，在左侧， D 和 A 控制着不同的条目：在没有重复指标的那些条目中，$D_{i_1,\dots,i_m}$ 等于 0；而在具有

重复指标的那些条目中。回想一下，并集 $\sigma^{n}{}_{,2}{}^{m} \cup \bar{\sigma}^{n}{}_{,2}{}^{m}$ 包含所有没有重复指标的条目索引。我们可以将上面的方程改写为

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

回顾一下

$$\begin{aligned}\langle L, v^{\otimes m}\rangle &= \frac{1}{m!\binom{m}{m/2}} \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \zeta(v_{i_1},\ldots,v_{i_m}) \\ &= \frac{1}{m!\binom{m}{m/2}} \sum_{(i_1,\ldots,i_m)\in\mathcal{E}} \left[(v_{i_1}+v_{i_2}+\cdots+v_{i_{m/2}}-v_{i_{m/2+1}}-\cdots-v_{i_m})^m+\ldots\right]. \qquad (9)\end{aligned}$$

接下来的观察是，通过展开括号中的 m次幂项，对于每个求和项（i_1、.....、 i_m），我们将得到一个由 v_{i_1}到 v_{i_m}组成的单项式序列。例如，其中包括 $C_0\cdot\, v_{i_1}v_{i_2}\ldots v_{i_{m-1}}$、 v_{i_m}、 $C_1\cdot\, v_{i_1}^2 v_{i_2}\ldots v_{i_{m-1}}$ 等，其中 C是系数。注意，所有这些单项式的阶数均为 m。

现在，我们通过将具有相同幂模式的单项式归为一组，来分析上述单项式项。形式上，我们将形如 $C\cdot\, v_{i_1}^{d_1}v_{i_2}^{d_2}\ldots v_{i_q}^{d_q}$的所有项归在一起，其中 $d_1\geq d_2\geq$ $\cdots\geq d_q\geq 1$、 $d_1+\cdots+d_q=m$。括号中列举了所有下标排列。我们用 Q 表示该组内项的数量。下面给出一些单项式组的例子：

- $C\cdot(v_{i_1}^m+v_{i_2}^m+\cdots+v_{i_m}^m)$，幂模式为 $d_1=m$。在这种情况下 $Q=m$。
- $m=6, C\cdot(v_{i_1}^5v_{i_2}^1+v_{i_1}^5v_{i_3}^1+\cdots+v_{i_6}^5v_{i_5}^1)$，幂模式为 $d_1=5, d_2=1$。在这种情况下 $Q=6\cdot 5=30$。
- $m=6, C\cdot(v_{i_1}^3v_{i_2}^2v_{i_3}^1+v_{i_1}^3v_{i_2}^2v_{i_4}^1+\cdots+v_{i_6}^3v_{i_5}^2v_{i_4}^1)$，幂模式为 $d_1=3, d_2=2, d_3=1$。在这种情况下 $Q=6\cdot 5\cdot 4=120$。

接下来我们讨论每个组中的幂模式。

如果幂模式为 $d_1=\cdots=d_m=1$（即所有分量都是一次幂分量），通过计数可知，系数C 等于 $\frac{1}{m!\binom{m}{m/2}}$。因此，这些所有分量均为一次幂分量的项，已由左侧的 $-\sum_{(i_1,\ldots,i_m)\notin\sigma^{n}{}_{,2}{}^{m}\cup\bar{\sigma}^{n,}{}_{2}{}^{m}} A_{i_1,\ldots,i_m}v_{i_1}\ldots v_{i_m}$ 平衡。这也说明了在超图拉普拉斯算子的定义中引入因子 $\frac{1}{m!\binom{m}{m/2}}$ 的必要性：它确保展开式 $\langle L, v^{\otimes m}\rangle$ 中的项 $v_{i_1}v_{i_2}\ldots v_{i_m}$ 的系数为 1。

对于具有其他幂模式（$d_1\geq\ldots d_q\geq 1, q<m$）的组，我们通过将 D个条目设为相等来平衡它们，这些条目的指标属于（i_1、.....、 i_1、 i_2、.....、 i_2、$\cdots\cdots$ i_q、.....、 i_q）的排列，其中 i_1重复 d_1 i_2 d_2 D CQ $\sum_{i_{q+1}\ldots i_m} A_{i_1\ldots i_m}$次、重复若干次，$CQ$ $\sum_{i_{q+1}\ldots i_m}\mathbb{1}[\,i_1\ldots i_m\in\,]$ D i_1 i_q 等等。我们将这些条目的值设为：$\cdot_{,,,,}\cdot_{,,}(,,)\,\mathcal{E}$。换言之，对于包含从.....到.....的指标的条目，我们将其设为与那些包含指标$\{i_1$、.....、 $i_m\}\setminus\{i_1$、.....、 $i_q\}$的条目相等。可以用上面的相同示例来说明这一点：

- 设 $D_{i_1,\ i_1}=CQ\sum_{i_2,\ i_m}A_{i_1,\ i_m}$。
- 设定 $D_{i_1,\ i_1,\ i_1,\ i_1,\ i_1,\ i_2}=CQ\sum_{i_3,\ i_4,\ i_5,\ i_6}A_{i_1,\ \ldots\ldots i_m}$，对于 $\sigma(i_1,\ i_1,\ i_1,\ i_1,\ i_1,\ i_2)$ 中的所有对称项，也作相同设定。
- 设定 $D_{i_1,\ i_1,\ i_1,\ i_2,\ i_2,\ i_3}=CQ\sum_{i_4,\ i_5,\ i_6}A_{i_1,\ \ldots\ldots i_m}$，对于 $\sigma(i_1,\ i_1,\ i_1,\ i_1,\ i_1,\ i_2)$ 中的所有对称项，也作相同设定。

完成该过程后，我们得到 $\langle L,\ v^{\otimes m}\rangle=\langle D-A,\ v^{\otimes m}\rangle$，以及 $D_{i_1,\ i_2,\ldots,i_m}=0$，如果 $i_1,\ i_2,\ldots,i_m$彼此均不相同。 □

定理 *3.15* 的证明。 设 v 是对应于第二小特征值$\lambda_2(L)$的特征向量。根据引理 3.11，不失一般性，我们假设 $\|v\| = 1$。我们还对 v 的各项进行排序，即 $v_1 \leq \cdots \leq v_n$。

我们的第一步是基于第二小特征向量 v 构造一个辅助向量 u。回顾一下，$\mathcal{H}$ 是诱导超顶点的集合（见定义 3.1）。对于任意超顶点 $h_j = \{i_1, \ldots, i_{m/2}\} \in \mathcal{H}$，基于第二特征向量 v，我们将 h_j 的 v-值定义为

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

这使我们能够按照超顶点的 v-值对所有超顶点排序，使得超顶点指标满足

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

并回顾一下，$N := |\mathcal{H}|$。我们用 M 表示满足 $M \geq \frac{1}{2}|\mathcal{H}|$ 的最小整数，并通过定义 $u := v - v(h_M)\cdot \mathbf{1}$ 引入一个平移操作。我们还找到一个常数 $c_u > 0$，并通过将 u 乘以 c_u 来缩放 u，使其满足 $(u_1 + \cdots + u_{m/2})^2 + (u_{n-m/2+1} + \cdots + u_n)^2 = 1$。注意，$u$ 满足以下性质：

- $(u_1 + \cdots + u_{m/2})^2 + (u_{n-m/2+1} + \cdots + u_n)^2 = 1$。
- $u_1 \leq \cdots \leq u_n$。
- $u_M = 0$。

根据引理 3.11 和 3.12，• $R_L(v) \geq R_L(u)$。

证明的第二步是构造一个超顶点的随机集合 S_t。为此，我们定义t 为支集 $[u_1 + \cdots + u_{m/2},\ u_{n-m/2+1} + \cdots + u_n]$ 上的随机变量，其概率密度函数为 $f(t) = 2|t|$。可以验证，t 是一个有效的随机变量，因为

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

基于 t，我们可以按如下方式构造一个超顶点的随机集合 S_t

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

这里我们考虑 S_t 在平均情况下的大小。我们有

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

and

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

将两者结合可得

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

我们还考虑边界集 ∂S_t。根据定义 3.2，如果 $h_1 \in S_t$、 $h_2 \notin S_t$ 和 $h_1 \cap h_2 = \emptyset$，则超边 $e = h_1 \cup h_2$ 属于 ∂S_t。定义简写记法 $u_h := \sum_{i \in h} u_i$，并假设 $u_{h_1} \leq u_{h_2}$。于是

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

最后，我们分析 $|\partial S_t|$的期望。由此可得

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

wh其中 (a) 由霍尔德不等式得出。重新整理上述各项可得

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

因此，存在某个 t 满足

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

或者等价地，

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

将左侧代入 $R_L(v) \geq R_L(u)$，右侧代入定义 3.4，可得

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

□

B 真实标签精确恢复的证明

定理 *4.1* 的证明。我们的目标是使用（4）恢复真实标签 $Y^* := (\pm y^*)^{\otimes m}$（符号正负可能相反）。为方便读者，我们在此重述该公式：

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

(4) 的拉格朗日对偶问题为

$$\begin{aligned} \text{最小化}_{V,V^+,V^-,A} \quad & \sum_{(i_1,\ldots,i_m)\in\sigma_2^{n,m}} V_{i_1,\ldots,i_m} + \sum_{(i_1,\ldots,i_m)\in\bar{\sigma}_2^{n,m}} (V^+_{i_1,\ldots,i_m} + V^-_{i_1,\ldots,i_m}) \\ \text{约束条件为} \quad & V_{i_1,\ldots,i_m} = 0\,，\text{若 } i_1,\ldots,i_m \text{ 均互不相同} \\ & V_{\bar{\sigma}_2^{n,m}} = V^+_{\bar{\sigma}_2^{n,m}} - V^-_{\bar{\sigma}_2^{n,m}} \\ & V^+_{\bar{\sigma}_2^{n,m}} \ge 0 \\ & V^-_{\bar{\sigma}_2^{n,m}} \ge 0 \\ & -X - A + \sum_{(i_1,\ldots,i_m)\in\sigma_2^{n,m}} V_{i_1,\ldots,i_m}\cdot E_{i_1,\ldots,i_m} + \sum_{(i_1,\ldots,i_m)\in\bar{\sigma}_2^{n,m}} (V^+_{i_1,\ldots,i_m} - V^-_{i_1,\ldots,i_m})\cdot E_{i_1,\ldots,i_m} = 0 \\ & A \in \mathcal{S}_+^{n,m}\,, \end{aligned}$$

其中，$E_{i_1,\ldots,i_m}$ 是一个张量，在条目 $(i_1,\ldots,i_m)$ 处取值为 1，其余位置均为 0。由原问题和对偶问题，我们得到以下 KKT 条件：

$$\begin{aligned} V - X - A &= 0 && \text{(Stationarity)} \\ Y_{\sigma_2^{n,m}} &= 1 \\ Y_{\bar{\sigma}_2^{n,m}} &\le 1 \\ -Y_{\bar{\sigma}_2^{n,m}} &\le 1 \\ Y &\in \mathcal{S}_+^{*,n,m} && \text{(Primal Feasibility)} \\ V_{i_1,\ldots,i_m} &= 0 \\ V_{\bar{\sigma}_2^{n,m}} &= V^+_{\bar{\sigma}_2^{n,m}} - V^-_{\bar{\sigma}_2^{n,m}} \\ V^+_{\bar{\sigma}_2^{n,m}} &\ge 0 \\ V^-_{\bar{\sigma}_2^{n,m}} &\ge 0 \\ A &\in \mathcal{S}_+^{n,m} && \text{(Dual Feasibility)} \\ V^+_{\bar{\sigma}_2^{n,m}}(Y_{\bar{\sigma}_2^{n,m}} - 1) &= 0 \\ V^-_{\bar{\sigma}_2^{n,m}}(Y_{\bar{\sigma}_2^{n,m}} + 1) &= 0 \\ \langle A, Y\rangle &= 0\,. && \text{(互补松弛)} \end{aligned}$$

在此，我们构造原始变量和对偶变量，以满足上述所有 KKT 条件。对于原始变量，我们设定 $Y = Y^*$。对于对偶变量，我们定义 $V = D_{y^*}$，其中 D_{y^*} 是由带符号拉普拉斯矩阵构造的高阶

度张量 L_{y^*}，其构造过程见引理 3.14。接着，我们定义 $V_{\sigma^{n,2}}{}^{m} = (D_{y^*})_{\sigma^{n,2}}{}^{m}$, $V^{+}_{\bar{\sigma}^{n,}{}_{2}{}^{m}} = \max((D_{y^*})_{\bar{\sigma}^{n,2}}{}^{m}, 0), V^{-}_{\bar{\sigma}^{n,}{}_{2}{}^{m}} = \min((D_{y^*})_{\bar{\sigma}^{n,2}}{}^{m}, 0)$。根据平稳性条件，我们设定 $A = L_{y^*} = D_{y^*} - X$。

至此，我们构造的原变量和对偶变量已满足上述所有 KKT 条件，唯独半正定条件 $A = L_{y^*} \in \mathcal{S}_{+}^{n}{}^{m}$, 尚未满足。根据引理 3.10，我们知道 y^* 是 L_{y^*} 的特征向量，其特征值为 0，或者等价地，$\langle L_{y^*}, Y^* \rangle = 0$。接下来需要确保，对于所有正交向量，$\min_{u \perp y^*} R_{L_{y^*}}(u) \geq 0$。除此之外，我们还希望确保解 $Y = Y^*$ 唯一。这还要求

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

不失一般性，在接下来的讨论中，我们固定 $\|u\| = 1$。我们将上述各项拆分为

$$\begin{aligned}
\min_{u \perp y^*, \|u\|=1} \langle D_{y^*} - X, u^{\otimes m} \rangle \geq & \min_{u \perp y^*, \|u\|=1} \langle D_{y^*} - \mathbb{E}\left[D_{y^*}\right], u^{\otimes m} \rangle & (10)\\
& + \min_{u \perp y^*, \|u\|=1} \langle \mathbb{E}\left[X\right] - X, u^{\otimes m} \rangle & (11)\\
& + \min_{u \perp y^*, \|u\|=1} \langle \mathbb{E}\left[D_{y^*} - X\right], u^{\otimes m} \rangle\,. & (12)
\end{aligned}$$

首先，我们对期望项 (12) 进行界定。注意，

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

其中，第一个不等式由引理 3.13 得出，第二个不等式由定理 3.15 得出。

接下来，我们使用集中不等式对 (11) 进行界定。我们有

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

设置 $t = \frac{1-2p}{2}\phi_{\mathcal{G}}^m$ 后得到

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

最后，我们对 (10) 进行界定。利用柯西-施瓦茨不等式，可得

$$-\left|\langle \mathbb{E}\left[D_{y^*}\right] - D_{y^*}, u^{\otimes m} \rangle\right| \geq -\left\|\mathrm{vec}\left(\mathbb{E}\left[D_{y^*}\right] - D_{y^*}\right)\right\| \cdot \left\|\mathrm{vec}\left(u^{\otimes m}\right)\right\|$$

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

其中，vec (·) 是向量化算子。接着，我们对每个元素应用霍夫丁不等式。根据引理 3.14 的证明中定义的构造过程， D_{y^*} 中的每个元素都是至多 $\max(|\mathcal{E}|, n^{m-1})$ 个拉德马赫随机变量之和。由此可得

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

通过并集界，我们得到

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

令 $t = \frac{(1-2p)\phi_{\mathcal{G}}^m}{2n^{m/2}}$，得到

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

结合式 (13) 和式 (14) 的结果，并应用并集界，即完成证明。 □