arXiv:2603.01268v1 [cs.DS] 1 Mar 2026

异质超图从其图投影恢复的可达性

Alexander Morgan * Chenghao Guo †

摘要

我们提出并分析了一个异质随机超图模型，并给出了从观测到的投影图恢复超边的可达性结果。我们观测一个投影图，它合并了所有度上的随机超边，其中一条投影边出现当且仅当两个顶点至少同时出现在一条超边中。我们的目标是针对某个 j，重构度为 d_j 的原始超边集合。我们的可达性结果基于在投影图中选取大小为 d_j 的极大团这一思想，并且我们证明该算法在关于密度的一个自然条件下能够成功。该可达性条件推广了 d-均匀超图在无噪投影和含噪投影下的一个已知阈值。我们猜想该阈值对于恢复具有最大度的超边是最优的。

1 引言

超边恢复问题自然地作为社区检测这一更大类问题的一部分出现。常见的社区检测问题涉及对友谊连接和社交网络等场景中成对关系的研究，其中任意一对共享同一潜在簇的顶点之间会形成连接。受现实世界社交网络的启发，具有群组结构的图（例如随机交集图）受到了显著关注[BBN20，vdHKV21]。此外，研究者还广泛研究了超图随机块模型 [ACKZ15，DW，GD17，GJ23，GP23，GP24，SZ22]。

这一问题在现实场景中有诸多应用，例如科学合著网络 [新04]。近期的相关工作开启了这一理论研究 [BGP24、BGPY25] ，确定了在无噪情形下从图投影重构超图的精确阈值。除了理论界限之外，人们还发展了算法方法，用于解决从投影图进行重构的问题；近年来，这一问题在理论和实证方面都得到了广泛研究。这些方法包括贝叶斯方法，即根据投影图从后验分布中进行采样 [LYA23、YPP21] , 基于采样超图的评分方法 [WK24] , , 以及用于恢复加权超图的基础模型 [CFLL24]。关于这一模型及其在真实网络中的各种应用的全面概述，可参见 [BGPY25]。

然而， [BGP24、BGPY25]中讨论的情形远非现实。真实网络中存在噪声，更重要的是，还存在不同大小的相互作用。含噪情形在 [GLX25],中进行了讨论，可将其视为由度为 2 的超边和度为d超边组成的混合情形。在他们的工作中，观测图是一个含噪投影，其中真实边以 p 的概率被保留， q非边则以 p 的概率被添加。其结果确定了检测和重构的精确阈值，尤其揭示了这一问题中的检测-重构差距现象。

一般异质超图是一个更具挑战性的研究问题。在这里，我们分析如何从投影图恢复异质超图。具体而言，我们研究以下模型。

定义 1（异质随机超图）. 考虑顶点集合 $V = [n]$，并固定超边度 $2 \le d_1 < \cdots < d_k$ 及相应的 $\delta_1, \ldots, \delta_k$，其中 $0 < \delta_j < 1$。假设每个度为 d_j 的超边以相同的概率 $p_j = n^{1-d_j+\delta_j+o(1)}$ 出现，并且所有度上的所有超边独立出现。一条边出现在投影图中，当且仅当该边中的两个顶点同时出现在至少一个超边中。

*xmorgan@mit.edu
†chenghao@mit.edu

在观测到由所有度上的超边所产生的投影图后，我们试图恢复度为 d_j的$\widehat{}$超边。设 $\mathcal{H}_j$ 是度为 d_j 的超边的真实集合， $\mathcal{H}_j$ 是我们估计的集合。若这两个集合的对称差的大小相对于 $\mathcal{H}_j$的期望大小为 o (1)，则我们认为恢复成功。具体而言，我们的目标是实现

$$\mathbb{E}\left[\left|\mathcal{H}_j \triangle \widehat{\mathcal{H}}_j\right|\right] = o(1)\mathbb{E}\left[|\mathcal{H}_j|\right],$$

我们将其作为成功恢复的判据。

1.1 结果

我们分析以下算法的性能。这里，极大团指不属于更大团子集的团。

算法 1 极大团估计

将度为 d_j 的超边估计为所有大小为 d_j 的极大团构成的集合。

根据定义 1所述的异质超图模型，我们将证明，当 $d_j \geq 3$ 时，可以恢复度为 $1-o$ 的超边，即期望恢复 (1) 条超边。

$$\delta^* \triangleq \max\{\delta_1, \ldots, \delta_k\} < \frac{d_j - 2}{d_j} + \frac{2\delta_j}{d_j(d_j-1)} \tag{1}$$

通过应用算法 1。特别是，我们将证明以下结果。

定理 1.1。 考虑定义 1 中指定的模型。固定一个 j ，其中 $d_j \geq 3$，并假设该 j 满足方程 1。令 $\mathcal{H}_j$ 为度为 d_j 的真超边集合。通过将算法 $\mathcal{H}_j$ 应用于$\widehat{}$投影图 1 来估计 ψ ，得到 $\mathcal{H}_j$ 成功，即:

$$\mathbb{E}\left[\left|\mathcal{H}_j \triangle \widehat{\mathcal{H}}_j\right|\right] = o(1)\mathbb{E}\left[|\mathcal{H}_j|\right].$$

大小为 d_j 的团可能由真超边产生，但也可能由不同度的超边组合而成，从而在投影超图中覆盖该团。这种情况的一种方式是，所有 $\binom{d_j}{2}$ 条边分别都被“噪声”超边覆盖。也可能存在其他构型。不过，我们将证明，如果能够保证由全边噪声构型产生的大小为 d_j 的团的密度，小于大小为 d_j 的真超边的密度，即:

$$-\binom{d_j}{2} + \binom{d_j}{2}\delta^* < 1 - d_j + \delta_j, \tag{2}$$

其中

$$\delta^* \triangleq \max\{\delta_1, \ldots, \delta_k\},$$

则算法1成功。特别地，全边噪声构型是我们需要考虑的最不利构型。

注意，方程1 等价于方程2。所有超边的同时恢复 $d_j \geq 3$是可能的，当且仅当方程1 中的所有必要不等式在 j上同时得到满足。通过从方程1 中提取条件

$$\delta_j < \frac{d_j - 2}{d_j} + \frac{2\delta_j}{d_j(d_j-1)},$$

这等价于条件

$$\delta_j < \frac{d-1}{d+1},$$

我们看到，我们的可达性结果自然推广了 d-一致超图情形下的$\frac{d-1}{d+1}$ 阈值。但我们强调，这是一个可达性结果；例如，对于恢复度为 $d_j < d_k$的超边，可能可以实现更好的性能。

在 d一致超图设定中，导致大小为d 的团的两种主导构型是真实超边构型和全边噪声构型。这些相互竞争的构$\frac{d-1}{d+1}$ 型决定了阈值。定理1.1 指出，这一思想的广义版本仍可实现，其中噪声可能由不同度的超边引起。

2 记号

我们用 d 表示超边的度。我们用 p 表示某条特定超边出现的概率，其中 $p = n^{1-d+\delta+o(1)}$，而 $0 < \delta < 1$ 是与 d 相关的固定参数。

在异质随机超图模型中，我们考虑 k 种度，即令 $2 \leq d_1 < \cdots < d_k$，其中度为 d_j 的超边以相同的概率 $p_j = n^{1-d_j+\delta_j+o(1)}$ 独立出现，且 $0 < \delta_j < 1$。我们还假设不同度的超边也相互独立地出现。

我们用 $\mathcal{H}$ 表示一个随机且可能是异质的超图，用 $\mathcal{H}_j$ 表示异质模型下仅由度为 d_j 的超边构成的随机超图。我们观察到 ψ，它是由 $\mathcal{H}$ 产生的随机投影超图。

我们用 E 表示边集，用 V 表示顶点集。此外，为方便起见，$[d] = \{1, \ldots, d\}$。给定边集 E，我们用 $\mathcal{U}$ 表示这些边的一个覆盖。特别地，令 V 表示由 E 张成的顶点集。那么，$\mathcal{U} \subset 2^V$ 满足

- For all $u \in \mathcal{U}$, $|u| \geq 2$, and $E \not\subset \bigcup_{u' \in \mathcal{U} \setminus \{u\}} \binom{u'}{2}$
- $E \subset \bigcup_{u \in \mathcal{U}} \binom{u}{2}$.

最后，$\mathsf{Bin}(n, p)$ 表示二项分布，我们将用 A 表示一个概率事件，具体考虑的是哪个事件将由上下文说明。

3 预备知识

我们将用到以下基本概率事实。证明留在附录中。

引理3.1. 设 $X_n \sim \mathsf{Bin}(m_n, p_n)$，其中 $m_n \to +\infty$ 且 $m_n p_n \to 0$。则，

$$\mathbb{P}(X_n \geq 1) = (1 - o(1)) m_n p_n. \tag{3}$$

更具体地说，

$$\mathbb{P}(X_n \geq 1) \geq (1 - o(1)) m_n p_n \tag{4}$$

and

$$\mathbb{P}(X_n \geq 1) \leq m_n p_n. \tag{5}$$

推论 3.2. 设 $X_n \sim \mathsf{Bin}\left(n^{d-d'+o(1)}\ n^{1-d+\delta+o(1)}\right)$，其中 $\delta < 1$ 且 $2 \leq d' < d$。则，

$$\mathbb{P}(X_n \geq 1) = n^{1-d'+\delta+o(1)}.$$

虽然这里不作此假设，但之后我们将把它应用于整数 d', d。

引理 3.3（最大速率占优）。固定 $r \geq 1$，并假设 $A_1,$ $\ldots$ $=$ $, A_r$ 是满足 $\mathbb{P}(A_j)$ $n^{\alpha_j+o(1)}$ 和 $\alpha_j < 0$ 的事件。则，

$$\mathbb{P}\left(\bigcup_{j=1}^{r} A_j\right) = n^{\max_j \alpha_j + o(1)}.$$

请注意，此结果不要求独立性。

4 概述

接下来，我们正式证明定理 1.1。在第5节中，我们证明子图出现在投影图 ψ 中的概率受一个优化问题 g 控制。在第6节中，我们在所关注的核心情形下求解该优化问题。证明将在第 7节结束。

5 关于 ψ 上事件的概率

本节旨在刻画某个子图 E 出现在投影图 $\mathbb{P}(E \subset \psi)$ 中的概率。

引理 3.1 和推论3.2 的作用在于，它们使我们能够推断由一个度为 d' 的超边所导致的较小团 d出现的概率。引理 5.1 对此进行了形式化。

引理 5.1 （隐含超边概率）。假设度为 $d \geq 2$ 的超边以概率 $p = n^{1-d+\delta+o(1)}$ 独立同分布地出现，其中 $0 < \delta < 1$。考虑固定的（相对于 n）顶点集 V、V'，其中$V' \subset V$且 $2 \leq d' = |V'| \leq d$。令 $A_{V'}$ 表示存在一条包含 V' 中所有节点且不包含 $V \setminus V'$ 中任何节点的超边这一事件。那么，

$$\mathbb{P}(A_{V'}) = n^{1-d'+\delta+o(1)}. \tag{6}$$

更具体地说，

$$\begin{aligned}\mathbb{P}(A_{V'}) &= \mathbb{P}\left(\mathrm{Bin}\left(\binom{n-|V|}{d-d'}, p\right) \geq 1\right) \\ &= 1-(1-p)^{\binom{n-|V|}{d-d'}} \\ &= (1-o(1))\binom{n-|V|}{d-d'}p,\end{aligned}$$

其上界为

$$\mathbb{P}(A_{V'}) \leq \binom{n-|V|}{d-d'}p.$$

对于其他满足 $V'' \neq V'$ 和 $2 \leq |V''| \leq d$ 的候选顶点集 $V'' \subset V$， $A_{V'}$ 和 $A_{V''}$ 相互独立，因为它们分别依赖于互不相交的超边集，因此可解释为“隐含超边概率”。

证明。 情况 $d' = d$ 显然成立。当 $d' < d$时，有 $\binom{n-|V|}{d-d'} = n^{d-d'+o_{(1)}}$ 个候选超边对 $A_{V'}$有贡献，且每个都以相同的概率 $p = n^{1-d+\delta+o(1)}$独立出现。结论由引理3.1和推论3.2得出。

□

接下来，我们着手研究多个度 $d_1 < \cdots < d_k$的情况。这使我们能够进一步推导出引理 5.1的一个略为一般化的版本。

引理 5.2 (一般隐含超边概率)。 考虑定义 1所指定的模型。再次固定顶点集 V， V'，其中 $V' \subset V$和 $2 \leq d' = |V'| \leq d_k$。令 $A_{V'}$表示存在一条超边（任意度数）包含 V' 中的所有节点且不包含 $V \setminus V'$中任何节点的事件。于是，

$$\mathbb{P}(A_{V'}) = n^{1-d'+\Delta+o(1)} \tag{7}$$

其中

$$\Delta \triangleq \max_{j:d_j \geq d'} \delta_j.$$

更具体地说，

$$\begin{aligned}\mathbb{P}(A_{V'}) &= 1-\prod_{j:d_j \geq d'}(1-p_j)^{\binom{n-|V|}{d_j-d'}} \\ &= (1-o(1))\sum_{j:d_j \geq d'}\binom{n-|V|}{d_j-d'}p_j,\end{aligned}$$

其上界为

$$\mathbb{P}(A_{V'}) \leq \sum_{j:d_j \geq d'}\binom{n-|V|}{d_j-d'}p_j.$$

引理 5.1 中关于其他候选顶点集的评论 $V'' \subset V$ 仍然适用，因为所有超边都是独立出现的。

证明。 对于任意 j 满足 $d_j \geq d'$的情形，令 $\mathcal{A}_{j,V'}$ 表示存在一条度为d_j 的超边，该超边包含 V' 中的所有节点，且不包含 $V \setminus V'$中的任何节点的事件。于是，根据引理 5.1 和 3.3，

$$\begin{aligned}\mathbb{P}(A_{V'}) &= \mathbb{P}\left(\bigcup_{j:d_j \geq d'} \mathcal{A}_{j,V'}\right) \\ &= n^{1-d'+\Delta+o(1)}.\end{aligned}$$

根据并集界，

$$\mathbb{P}(A_{V'}) \leq \sum_{j:d_j \geq d'} \binom{n-|V|}{d_j - d'} p_j,$$

而博纳费罗尼不等式表明

$$\begin{aligned}\mathbb{P}(A_{V'}) &\geq \sum_{j:d_j \geq d'} (1-o(1)) \binom{n-|V|}{d_j - d'} p_j \left[1 - \sum_{r>j} (1-o(1)) \binom{n-|V|}{d_r - d'} p_r\right] \\ &\geq (1-o(1)) \sum_{j:d_j \geq d'} \binom{n-|V|}{d_j - d'} p_j.\end{aligned}$$

□

我们可以 现在分析投影图中出现特定边构型的概率 .

定理 5.3。 考虑定义1所指定的模型。固定一个边集合 E，并令 V为这些边所张成的顶点集合。于是，令 ψ 表示由所有度的超边产生的组合投影图，

$$\mathbb{P}(E \subset \psi) = n^{g(E,\boldsymbol{\Delta})+o(1)},$$

其中 $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{|V|})$ 由 $\boldsymbol{\Delta}_j = \max_{r:d_r \geq j} \delta_r$ 给出，约定当 $j > d_k$ 时 $\boldsymbol{\Delta}_j = -\infty$，而 $g(E, \boldsymbol{\Delta})$ 是我们现在定义的。对于满足 $\boldsymbol{\Delta}_j < 1$ 的 $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_{|V|})$，令

$$g(E, \boldsymbol{\Delta}) \triangleq \max_{\mathcal{U}} \sum_{u \in \mathcal{U}} 1 + \boldsymbol{\Delta}_{|u|} - |u|, \tag{8}$$

其中，最大值是在所有有效的 $\mathcal{U}$ 覆盖 E上取的，即要求 $\mathcal{U} \subset 2^V$ 满足

- 对于所有 $u \in \mathcal{U}$、 $|u| \geq 2$ 和 $E \not\subset \bigcup_{u' \in \mathcal{U} \setminus \{u\}} \binom{u'}{2}$
- $E \subset \bigcup_{u \in \mathcal{U}} \binom{u}{2}$。

给定一个固定的 E，候选覆盖 $\mathcal{U}$只有有限个。

为了更好地控制概率，我们有下界

$$\mathbb{P}(E \subset \psi) \geq (1-o(1)) \prod_{u \in \mathcal{U}} \sum_{j:d_j \geq |u|} \binom{n-|V|}{d_j - |u|} p_j$$

对于任意一个特定的覆盖 $\mathcal{U}$，以及上界

$$\mathbb{P}(E \subset \psi) \leq \sum_{\mathcal{U}} \prod_{u \in \mathcal{U}} \sum_{j:d_j \geq |u|} \binom{n-|V|}{d_j - |u|} p_j.$$

证明。 固定一个候选覆盖 $\mathcal{U}$ ，使得 $|u| \leq d_k$ 对所有 $u \in \mathcal{U}$成立。对于某个特定的 $u \in \mathcal{U}$，令 A_u表示存在一条超边，其包含 u 中的所有顶点，但不包含 $V \setminus u$中任何顶点的事件，并令 $A_{\mathcal{U}} = \bigcap_{u \in \mathcal{U}} A_u$。则引理 5.2 表明

$$\mathbb{P}(A_u) = n^{1-|u|+\boldsymbol{\Delta}_{|u|}+o(1)},$$

因此，由独立性可知，

$$\mathbb{P}(A_{\mathcal{U}}) = n^{\sum_{u \in \mathcal{U}} 1-|u|+\boldsymbol{\Delta}_{|u|}+o(1)}.$$

令 A 表示所有边 E 都出现在投影图中的事件，即 $A = \{E \subset \psi\}$。注意到$A = \cup_{\mathcal{U}} A_{\mathcal{U}}$。显然， $A_{\mathcal{U}} \subset A$ 因为 $A_{\mathcal{U}}$中的任意随机样本，其投影图都包含 E中的所有边，因此 $\cup_{\mathcal{U}} A_{\mathcal{U}} \subset A$。此外，对于 A中的任意随机样本，考虑所有至少包含 V中两个节点的超边。按任意顺序贪心地删去对覆盖 E而言冗余的超边。这样，剩余超边与 V 的交集构成一个有效覆盖，因此 $A \subset \cup_{\mathcal{U}} A_{\mathcal{U}}$。最后，由于候选覆盖 $\mathcal{U}$的数量有限，引理3.3 推出

$$\mathbb{P}(A) = n^{g(E,\boldsymbol{\Delta})+o(1)}.$$

关于 $\mathbb{P}(A)$ 的上下界也可由类似的分析和引理 5.2 得出。 □

注 1。 引入此函数的一个受限版本会很方便

$$g(E, \boldsymbol{\Delta}; M) \triangleq \max_{\mathcal{U}:|\mathcal{U}|=M} \sum_{u\in\mathcal{U}} 1 + \boldsymbol{\Delta}_{|u|} - |u| \tag{9}$$

使得

$$g(E, \boldsymbol{\Delta}) = \max_{1\le M\le \binom{|V|}{2}} g(E, \boldsymbol{\Delta}; M). \tag{10}$$

也可以方便地使用 $\delta^* = \max\{\delta_1, \ldots, \delta_k\}$ 来界定

$$g(E, \boldsymbol{\Delta}) \le g(E, \delta^*) \triangleq \max_{\mathcal{U}} \sum_{u\in\mathcal{U}} 1 + \delta^* - |u|,$$

其中，最大值取遍满足附加约束 $|u| \le d_k$ （对所有 $u \in \mathcal{U}$ 均成立）的覆盖 $\mathcal{U}$。

6 $g(E,\ \boldsymbol{\Delta})$ 的组合性质

现在，我们证明关于 $g(E,\ \boldsymbol{\Delta})$ 的两个关键引理，证明推迟至附录给出。

引理 6.1. 固定顶点 $V = \{1, \ldots, d\}$，满足 $d \ge 3$，并固定 $0 < \delta < 1$。令 $\boldsymbol{\Delta}_j = \delta$，其中 $2 \le j \le d-1$ 和 $\boldsymbol{\Delta}_d = -\infty$。因此，我们隐式地限制了 $\mathcal{U}$ 中集合的大小至多为 $d-1$。于是，

$$g\left(\binom{V}{2}, \boldsymbol{\Delta}\right) = \max\left\{\delta d - 2d + 3, -\binom{d}{2} + \binom{d}{2}\delta\right\},$$

可通过“$d-1$ 团加星形”覆盖 $\mathcal{U} = \{\{1, \ldots, d-1\}\} \cup \{\{1, d\}, \{2, d\}, \ldots, \{d-1, d\}\}$ 或单边覆盖 $\mathcal{U} = \binom{V}{2}$ 达成。

引理 6.2. 固定顶点 $V = \{1, \ldots, d\}$，满足 $d \ge 2$，并假设 $\boldsymbol{\Delta}$ 满足 $0 < \boldsymbol{\Delta}_j < 1$，其中 $2 \le j \le d$，且

$$\delta \triangleq \boldsymbol{\Delta}_2 = \max_{2\le j\le d} \boldsymbol{\Delta}_j.$$

定义边集 $E = \{(1, j)\}_{j=2}^{d}$，表示一个星形图构型。然后， $g(E; \boldsymbol{\Delta}) = -(d-1) + (d-1)\delta$ 由单条边覆盖 $\mathcal{U} = E = \{\{1, j\}\}_{j=2}^{d}$ 达成。

7 可达性结果定理 证明1.1

证明。 我们有

$$\mathbb{E}\left[|\mathcal{H}_j|\right] = \binom{n}{d_j} n^{1-d_j+\delta_j+o(1)}$$

and

$$\mathbb{E}\left[\left|\mathcal{H}_j \triangle \widehat{\mathcal{H}}_j\right|\right] = \binom{n}{d_j}\left[\mathbb{P}(A_{\mathrm{FP}}) + \mathbb{P}(A_{\mathrm{FN}})\right],$$

其中， A_{FP} 是估计 $[d_j]$时发生的假阳性事件，即 $A_{\mathrm{FP}} = \{[d_j] \notin \mathcal{H}_j\} \cap \{[d_j] \in \widehat{\mathcal{H}}_j\}$；而$A_{\mathrm{FN}}$ 是估计 $[d_j]$时发生的假阴性事件，即 $A_{\mathrm{FN}} = \{[d_j] \in \mathcal{H}_j\} \cap \{[d_j] \notin \widehat{\mathcal{H}}_j\}$。因此，只需证明

$$\mathbb{P}(A_{\mathrm{FP}}) = o(1) \cdot n^{1-d_j+\delta_j+o(1)}$$

and

$$\mathbb{P}(A_{\mathrm{FN}}) = o(1) \cdot n^{1-d_j+\delta_j+o(1)}.$$

首先，我们分析假阳性的发生率。注意，任何包含所有节点的大小为 $d_r \geq d_j$ 的超边 $[d_j]$ 都不会影响假阳性概率，因为这会导致 $[d_j]$ 不是极大的，或者意味着 $[d_j]$ 是一条真实的超边。因此，出现假阳性的唯一方式是形成一个覆盖 $\mathcal{U}$，其覆盖对象为 $E = \binom{[d_j]}{2}$，且满足 $|\mathcal{U}| \geq 2$。$1-o$ 不是超边的概率为 $1-o(1)$，对定理 5.3 的证明稍作修改可得，由大小为 $|\mathcal{U}| \geq 2$ 的覆盖导致所有边出现在 E 中的概率为 $n^{g(E;\tilde{\boldsymbol{\Delta}})+o(1)}$，其中 $\tilde{\boldsymbol{\Delta}}_\ell =$ 最大值$_{r:d_r \geq \ell}\, \delta_r$ 取遍 $\ell < \tilde{d_j}$，且 $\boldsymbol{\Delta}_{d_j} = -\infty$。这些事件相互独立，因此

$$\mathbb{P}(A_{\mathrm{FP}}) \leq n^{g(E;\tilde{\boldsymbol{\Delta}})+o(1)}.$$

现在，令 $\boldsymbol{\Delta}_\ell = \delta^*$ 对于 $\ell < d_j$ 和 $\boldsymbol{\Delta}_{d_j} = -\infty$。然后，应用引理 6.1，

$$\begin{aligned}\mathbb{P}(A_{\mathrm{FP}}) &\leq n^{g(E;\boldsymbol{\Delta})+o(1)} \\ &= n^{\max\left\{-\binom{d_j}{2}+\binom{d_j}{2}\delta^*,\, \delta^* d_j - 2d_j + 3\right\}+o(1)}.\end{aligned}$$

现在，等式 1 等价于

$$-\binom{d_j}{2} + \binom{d_j}{2}\delta^* < 1 - d_j + \delta_j.$$

此外，

$$\begin{aligned}\delta^* d_j - 2d_j + 3 &< 1 - d_j + \frac{2\delta_j}{d_j - 1} \\ &\leq 1 - d_j + \delta_j\end{aligned}$$

对于 $d_j \geq 3$。因此，

$$\mathbb{P}(A_{\mathrm{FP}}) = o(1) \cdot n^{1-d_j+\delta_j+o(1)}.$$

现在，我们分析 $[d_j]$发生假阴性的概率。要发生假阴性，超边 $[d_j]$ 必须存在，并且 $[d_j]$ 中的所有节点都必须与一个外部节点相连；也就是说，我们必须观察到连接 $[d_j]$ 与某个外部节点的星形结构。这两个事件相互独立。$[d_j]$ 是超边的概率为 $p_j = n^{1-d_j+\delta_j+o_{(1)}}$，根据定理5.3 和引理6.2，观察到连接 $[d_j]$ 与某个特定外部节点的星形结构的概率为 $n^{-d_j+d_j\delta^*+o_{(1)}}$。根据并集界，观察到连接 $[d_j]$ 与任意外部节点的星形结构的概率至多为 $n^{1-d_j+d_j\delta^*+o_{(1)}}$。因此，根据独立性，

$$\mathbb{P}(A_{\mathrm{FN}}) \leq n^{1-d_j+\delta_j+o(1)} \cdot n^{1-d_j+d_j\delta^*+o(1)}.$$

最后，根据方程1，

$$1 - d_j + d_j\delta^* < -1 + \frac{2\delta_j}{d_j - 1} \leq 0$$

对于 $d_j \geq 3$，因此

$$\mathbb{P}(A_{\mathrm{FN}}) \leq o(1) \cdot n^{1-d_j+\delta_j+o(1)},$$

证明至此完成。 □

8 结论

我们证明，在异质超图模型中，算法 1 所定义的极大团估计量能够恢复度为 d_j 且具有 $o(1)$ 对称差误差的超边，前提是大小为 d_j 的团的全边构型密度不超过度为 d_j的真实超边密度。该可实现性结果以自然的方式推广了 d-均匀超图中的既有恢复阈值，并为这一设置下的社区检测提供了一种概念上简单的方法。

这里我们还列出了一些开放问题。

1. 可达性逆命题。我们猜想，定理 1.1 中的阈值对于恢复度最大的超边而言是最优的。2. 异质超图的一般最优恢复阈值。当 $\delta_1, \cdots, \delta_k$ 任意取值且目标是恢复所有不同度的超边时，定理 1.1 中的阈值可能并非最优。3. 文献 [BGP24] 和 [BGPY25]讨论了 d-均匀超图的精确恢复。然而，异质随机超图的精确恢复尚未得到充分理解，即使是在仅存在度为 2 和度为d 的超边的含噪情形下也是如此。

参考文献

[ACKZ15] Maria C. Angelini、Francesco Caltagirone、Florent Krzakala 和 Lenka Zdeborov´a. 稀疏超图上的谱检测。载于第53届 Allerton 通信、控制与计算年会论文集（Allerton），第66–73页。IEEE，2015。[BBN20] Matthew Brennan、Guy Bresler 和 Dheeraj M. Nagaraj. 随机图中潜在几何检测的相变。《概率论及相关领域》，178(3-4)：1215–1289，2020。[BGP24] Guy Bresler、Chenghao Guo 和 Yury Polyanskiy. 从图投影重构随机超图的阈值。载于第三十七届学习理论会议论文集，《机器学习研究论文集》第247卷，第632–647页。PMLR，2024。[BGPY25] Guy Bresler、Chenghao Guo、Yury Polyanskiy 和 Andrew Yao. 从图投影部分恢复和精确恢复随机超图。arXiv 预印本 arXiv:2502.14988，2025。[CFLL24] Yang Chen、Cong Fang、Zhouchen Lin 和 Bing Liu. 预训练模型中的关系学习：从超图恢复视角出发的理论。载于第41届国际机器学习会议（ICML）论文集，第6666–6698页。PMLR，2024。[DW] Ioana Dumitriu 和 Hai-Xiao Wang. 一般非均匀超图随机块模型上的最优精确恢复。《统计年鉴》，即将发表。[GD17] Debarghya Ghoshdastidar 和 Ambedkar Dukkipati. 植入划分模型下均匀超图谱划分的一致性。《统计年鉴》，45(1)：289–315，2017。[GJ23] Julia Gaudio 和 Nirmit Joshi. 超图随机块模型中的社区检测：给定相似度矩阵时的精确恢复。载于第36届学习理论会议（COLT）论文集，第469–510页。PMLR，2023。[GLX25] Shuyang Gong、Zhangsong Li 和 Qiheng Xu. 从含噪图投影检测和重构随机超图。arXiv 预印本 arXiv:2506.17527，2025。[GP23] Yuzhou Gu 和 Yury Polyansky. 超图随机块模型的弱恢复阈值。载于第36届学习理论年会（COLT）论文集，第885–920页。PMLR，2023。[GP24] Yuzhou Gu 和 Yury Polyansky. 超图随机块模型中的社区检测及超树上的重构。载于第37届学习理论年会（COLT）论文集，第2166–2203页。PMLR，2024。[LYA23] Simon Lizotte、Jean-Gabriel Young 和 Antoine Allard. 从不确定的成对观测重构超图。《科学报告》，13(1)：21364，2023。[New04] Mark E. J. Newman. 合著网络与科学合作模式。《美国国家科学院院刊》，101(1)：5200–5205，2004。

[SZ22] Ludovic Stephan 和 Yizhe Zhu. 稀疏随机超图：非回溯谱与社区检测。载于*IEEE 第63届计算机科学基础年会（FOCS）*，第567–575页。IEEE，2022。[vdHKV21] Remco van der Hofstad、J´ulia Komj´athy 和 Vikt´oria Vadon. 带社区的随机交集图。《应用概率进展》，53(4)：1061–1089，2021。[WK24] Yanbang Wang 和 Jon Kleinberg. 从图到超图：超图投影及其重构。载于第*12届国际学习表征会议（ICLR）*，2024。[YPP21] Jean-Gabriel Young、Giovanni Petri 和 Tiago P. Peixoto. 从网络数据重构超图。《通信物理》，4(1)：135，2021。

A 延后证明

A.1 引理3.1的证明

证明。 我们有

$$\begin{aligned}\mathbb{P}(X_n \geq 1) &\geq \mathbb{P}(X_n = 1)\\ &\geq m_n p_n \left(1-p_n\right)^{m_n}\\ &= m_n p_n \left(1-\frac{m_n p_n}{m_n}\right)^{m_n}\\ &= m_n p_n (1-o(1)).\end{aligned}$$

此外，并集界或马尔可夫不等式给出式5。 □

A.2 引理3.3的证明

证明。

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^{r} A_j\right) &\geq \max_j \mathbb{P}(A_j)\\ &\geq n^{\max_j \alpha_j + o(1)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^{r} A_j\right) &\leq r\left[\max_j \mathbb{P}(A_j)\right]\\ &\leq n^{\max_j \alpha_j + o(1)}.\end{aligned}$$

□

A.3 引理6.1 的证明

证明。 这里，我们推广了 [BGP24], 中的一个引理，证明方法类似。为完整起见，我们在此给出推理过程，并对论证略作简化。

我们可以将最大化问题写为

$$g(E, \mathbf{\Delta}) = \max_{2 \leq M \leq \binom{d}{2}} \max_{\mathcal{U}:|\mathcal{U}|=M} \sum_{u \in \mathcal{U}} 1 + \delta - |u|,$$

其中我们限制 $M \geq 2$，因为可以忽略 $\mathcal{U} = \{V\}$ 这种情况。内层最大化是在满足通常条件的所有有效覆盖 $\mathcal{U}$ 上进行的。现在，令

$$\begin{aligned} g(E, \mathbf{\Delta}; M) &= \max_{\mathcal{U}:|\mathcal{U}|=M} \sum_{u \in \mathcal{U}=\{u_1,\ldots,u_M\}} 1 + \delta - |u| \\ &= \max_{\mathcal{U}:|\mathcal{U}|=M} M(1+\delta) - \sum_{u \in \mathcal{U}=\{u_1,\ldots,u_M\}} |u|. \end{aligned}$$

我们将 $|u_j| = x_j$ 放宽为实数，以进一步给出这个界。令 $y_j = {}^{x_j(x_j-1)}\,{}_2$，这意味着$x_j = {}^{1+\sqrt{1+8y_j}}_{2}$。因此，由于所有 $\binom{d}{2}$ 边都必须由 $\mathcal{U}$覆盖，

$$g(E, \mathbf{\Delta}; M) \leq \max_{\substack{y_1,\ldots,y_M \geq 1 \\ \sum_{j=1}^{M} y_j \geq \binom{d}{2}}} M(1+\delta) - \sum_{j=1}^{M} \frac{1+\sqrt{1+8y_j}}{2}.$$

我们要在凸区域上最大化一个凸函数，因此最大值将在某个顶点处或无穷远处取得。由于当任意 $y_j \to \infty$时，该函数趋近于 $-\infty$，所以最大值必定在某个顶点处取得。每个顶点对应的函数值都相同，因此选取 $y_1 = \cdots = y_{M-1} = 1$和 $y_M = \binom{d}{2} - M + 1$，得到

$$g(E, \mathbf{\Delta}; M) \leq M(\delta - 1) + 2 - \frac{1+\sqrt{1+8(\binom{d}{2} - M + 1)}}{2}. \tag{11}$$

这个关于 M的界是凸的，因此该界在 $d \leq M \leq \binom{d}{2}$ 上的最大值在$M = d$ 或 $M = \binom{d}{2}$处取得，这意味着

$$\max_{d \leq M \leq \binom{d}{2}} g(E, \mathbf{\Delta}; M) \leq \max\left\{ d\delta - 2d + 3, -\binom{d}{2} + \binom{d}{2}\delta \right\}.$$

由于“$d-1$团加星形”覆盖 $\mathcal{U} = \{\{1,\ldots,d-1\}\} \cup \{\{1,d\},\{2,d\},\ldots,\{d-1,d\}\}$ 达到值 $d\delta - 2d + 3$，而单条边覆盖 $\mathcal{U} = E$达到值 $-\binom{d}{2} + \binom{d}{2}\delta$，因此$g(E, \mathbf{\Delta}; d) = d\delta - 2d + 3$且 $g\left(E, \mathbf{\Delta}; \binom{d}{2}\right) = -\binom{d}{2} + \binom{d}{2}\delta$，所以

$$\max_{d \leq M \leq \binom{d}{2}} g(E, \mathbf{\Delta}; M) = \max\left\{ d\delta - 2d + 3, -\binom{d}{2} + \binom{d}{2}\delta \right\}. \tag{12}$$

接下来我们将证明

$$h(d) \triangleq \max_{2 \leq M \leq d} g(E, \mathbf{\Delta}; M) = d\delta - 2d + 3, \tag{13}$$

通过对 d 使用归纳法。当 $d = 3$ 时，唯一有效的覆盖是 $\mathcal{U} = \{\{1,2\},\{1,3\},\{2,3\}\}$，其达到的值为 $3\delta - 3$。现在，假设 $h(d-1) = (d-1)\delta - 2(d-1) + 3$，我们希望证明 $h(d) = d\delta - 2d + 3$。考虑由 $V_d = \{1,\ldots,d\}$ 张成的团的任意候选覆盖 $\mathcal{U}$，满足 $2 \leq |\mathcal{U}| \leq d-1$。情形 $|\mathcal{U}| = d$ 已由 $g(E, \mathbf{\Delta}; d) = d\delta - 2d + 3$ 处理。我们首先说明，可以假设存在一个顶点 $v \in V_d$ 满足

- $V_d \setminus \{v\} \notin \mathcal{U}$
- $|\{u \in \mathcal{U} : |u| = 2, \ v \in u\}| \leq 1$。

反设这样的顶点 v 不存在，即每个 $v \in V_d$ 要么有$V_d \setminus \{v\} \in \mathcal{U}$，要么有 $|\{u \in \mathcal{U} : |u| = 2, v \in u\}| \geq 2$。令 T表示满足$|\{u \in \mathcal{U}$的节点集合 $v \in V_d : |u| = 2, v \in u\}| \geq 2$，令 W表示满足 $V_d \setminus \{v\} \in \mathcal{U}$的节点集合 $v \in V_d$。我们有 $|W| + |T| \geq d$。令 $\ell = |\{u \in \mathcal{U} : |u| = 2\}$表示 $\mathcal{U}$中大小为 2 的集合的数量。所有这些边所覆盖的度总和为 2ℓ，而与 T 中的节点相关联的大小为 2 的集合的度总和至少为$2|T|$，因此 $\ell \geq |T|$。由于

$$d \leq |T| + |W| \leq \ell + |W| \leq |\mathcal{U}| \leq d - 1,$$

我们得出矛盾。

回到满足上述条件的顶点 v，考虑由 $V_d \setminus \{v\}$ 张成的团的覆盖 $\mathcal{U}' \triangleq \{u \setminus \{v\} : u \in \mathcal{U}, |u \setminus \{v\}| \geq 2\}$。注意，由于 $V_d \setminus \{v\} \notin \mathcal{U}$，因此 $2 \leq |\mathcal{U}'| \leq d-1$。于是，

$$\begin{aligned}\left[\sum_{u\in\mathcal{U}} 1+\delta-|u|\right] - \left[\sum_{u\in\mathcal{U}'} 1+\delta-|u|\right] &= \left[\sum_{u\in\mathcal{U}, v\in u, |u|>2} -1\right] + \left[\sum_{u\in\mathcal{U}, v\in u, |u|=2} -1+\delta\right] \\ &= -|\{u\in\mathcal{U} : v\in u\}| + \delta\cdot|\{u\in\mathcal{U} : |u|=2, v\in u\}| \\ &\leq -2+\delta,\end{aligned}$$

其中，我们应用了 $|\{u \in \mathcal{U} : |u| = 2, v \in u\}| \leq 1$ 和 $|\{u \in \mathcal{U} : v \in u\}| \geq 2$，从而得到最后一个不等式。（如果 $|\{u \in \mathcal{U} : v \in u\}| = 1$，那么这个唯一的元素 u 必然等于 V_d，这就会导致矛盾。）最后，由于 $2 \leq |\mathcal{U}'| \leq d-1$，归纳假设可得

$$\left[\sum_{u\in\mathcal{U}'} 1+\delta-|u|\right] \leq (d-1)\delta - 2(d-1) + 3 = d\delta - 2d + 3 + 2 - \delta,$$

so

$$\left[\sum_{u\in\mathcal{U}} 1+\delta-|u|\right] \leq d\delta - 2d + 3,$$

证明等式 13。结合等式 12 和等式 13 即可得到所需结果 lt. □

A.4 引理6.2的证明

证明。 $d = 2$ 的情形显然。当 $d \geq 3$时，我们可以通过贪心地将任意覆盖 $\mathcal{U}$ 缩减为单条边，来达到最大值。考虑任意 $u \in \mathcal{U}$，其中 $\ell = |u| \geq 3$。那么， u 覆盖了 $\ell - 1$ 条边。不失一般性，设 $u = \{1, \ldots, \ell\}$。如果我们将 u 替换为 $2 \leq j \leq \ell$对应的边集合 $\{1, j\}$，那么目标值

$$\sum_{u\in\mathcal{U}} 1 + \boldsymbol{\Delta}_{|u|} - |u|$$

增加了

$$\begin{aligned}(\ell-1)(-1+\delta) - (1+\boldsymbol{\Delta}_\ell - \ell) &\geq (\ell-1)(-1+\delta) - (1+\delta-\ell) \\ &= (\ell-2)\delta \\ &\geq 0.\end{aligned}$$

移除任何新产生的冗余边只会进一步提高目标值，因为每条边的贡献为$-1 + \delta < 0$。因此，我们可以通过转换为全边覆盖来最大化 $g(E; \boldsymbol{\Delta})$，这意味着$g(E; \boldsymbol{\Delta}) = -(d-1) + (d-1)\delta$。

□