arXiv:2304.08135v1 [cs.DS] 17 Apr 2023

检测 低次多项式检测稠密子超图

Abhishek Dhawan*[1]、Cheng Mao†[2]和 Alexander S. Wein‡[3]

[1,2]佐治亚理工学院数学学院[3]加州大学戴维斯分校数学系

摘要

在随机图中检测植入稠密子图是一项基础性的统计与计算问题，近年来受到了广泛研究。我们研究这一问题的超图版本。令 $G^r(n,p)$ 表示具有 n 个顶点且边密度为 p 的 r 均匀 Erd″os-R´enyi 超图模型。我们考虑在 $G^r(n,n^{-\beta})$ 超图中检测是否存在一个植入的 $G^r(n^\gamma,n^{-\alpha})$ 子超图，其中 $0<\alpha<\beta<r-1$ 且 $0<\gamma<1$。我们聚焦于邻接张量各项的次数为 $n^{o(1)}$ 的多项式检验，并确定该检测问题中易解与困难区域之间的阈值。更准确地说，当 $0<\gamma<1/2$ 时，阈值为 $\alpha=\beta\gamma$；当 $1/2\le\gamma<1$ 时，阈值为 $\alpha=\beta/2+r(\gamma-1/2)$。

我们的结果在图情形下也已是新的 $r=2$，因为我们考虑的是一个微妙的对数密度区间，在该区间中，基于平均情形归约的困难性尚未得到证实。我们对低次数困难性的证明基于标准低次似然计算的一种条件变体。

目录

1 引言 2
1.1 与先前工作的关系.3

2 主要结果 6

3 检测下界 7
3.1 大 γ .8
3.2 小 γ .10
3.2.1 高概率事件.10
3.2.2 低次范数的界定.12

*电子邮箱： *abhishekdhawan@gatech.edu*。A.D. 部分研究经费由美国国家科学基金会（NSF）资助，项目编号为 DMS-2053333。

†电子邮箱： *cheng.mao@math.gatech.edu*。C.M. 部分研究经费由美国国家科学基金会（NSF）资助，项目编号为 DMS-2053333 和 DMS- 2210734。

‡电子邮箱： *aswein@ucdavis.edu*。A.S.W. 在佐治亚理工学院算法与随机性中心工作期间完成了本研究的部分内容；该中心获得了美国国家科学基金会（NSF）项目 CCF-2007443 和 CCF-2106444 的资助。

4 检测上界 **16**
4.1 大 γ .164.2 小 γ .194.2.1 平衡超图.194.2.2 强分离.20

A 检测-反驳间隙 **26**

1 引言

在给定的无向图中寻找稠密子图，是图论、计算机科学和统计学交汇处的一个经典问题，并在社会科学和生物科学中有广泛应用。过去几十年间，围绕该问题的多种变体开展了大量研究，包括植入团问题 [Kar76、Jer92], 最稠密 k-子图问题 [FS97、BCC+10]、社群检测 [For10、ACV14], 以及超图版本 [CDK+18、CPSB22、LZ22], 等。在本文中，我们关注这样一个问题：检测给定的 r-均匀超图 H （含有 n 个顶点）中是否存在植入稠密子超图。用统计假设检验的语言来说，我们考虑在 H上（在第 2 节中作形式化定义）检验以下两个分布：

- 零假设分布是边密度为 q 的 r-均匀 Erd˝os-R´enyi 超图 $G^r(n, q)$;
- 植入分布会随机选取 $\approx \rho n$ 个顶点，使其属于植入子超图；子超图内部的边以概率 p独立出现，所有其他边则以概率 q独立出现。

给定从这两种分布之一中抽取的单个超图 H ，目标是以高概率（w.h.p.）区分这两种情况，即当 $n \to \infty$ 时，概率为 $1 - o(1)$，其中 p、 q、 ρ 可以随 n变化。我们假设参数 p、 q、 ρ 已知。

我们关注的是理解计算高效的检验（即多项式时间算法）的能力与局限。目前我们缺乏能够证明此类平均情形问题（其输入为随机对象）的计算困难性的复杂性理论工具，因此当前的主流方法要么基于平均情形归约，形式化地将不同的平均情形问题联系起来（例如参见 [BB20] 及其中的参考文献），要么针对受限算法类别证明无条件下界。

我们的主要结果给出了 p、 q、 ρ 取得低阶多项式检验成功所需的精确充要条件。这是一类强大的检验，其中包括小子图计数（边、三角形等）之类的统计量。迄今为止，低阶检验已被广泛认为是计算高效检验的一种有用替代指标：对于各种高维检验问题，目前已知的最佳多项式时间算法都可由低阶类刻画；例如参见 [Hop18、KWB22]。

更具体地说，我们关注所谓的对数密度区域[BCC+10] ，其中 $p = n^{-\alpha}$， $q = n^{-\beta}$，并且 $\rho = n^{\gamma-1}$ 对于常数 $0 < \alpha < \beta < r - 1$ 和 $\gamma \in (0, 1)$。假设 $\beta < r - 1$排除了孤立顶点。我们证明，如果 $0 < \gamma < 1/2$ 且 $\alpha < \beta\gamma$，或者如果 $1/2 \le \gamma < 1$ 且$\alpha < \beta/2 + r(\gamma - 1/2)$ ，则存在一个常数次数多项式检验，能够以高概率区分零分布和植入分布。另一方面，如果 $0 < \gamma < 1/2$ 且 $\alpha > \beta\gamma$，或者如果 $1/2 \le \gamma < 1$

且 $\alpha > \beta/2 + r(\gamma - 1/2)$，那么不存在次数为 $n^{o(1)}$ 的多项式能够区分零分布和植入分布。

1.1 与先前工作的关系

植入稠密子图。我们的结果在图情形 $r = 2$ 下已经很有意思；此时我们考虑的是植入稠密子图检测问题。统计阈值已得到充分研究 [ACV14, BI13, VAC15, HWX15], ，而在各种参数区间内，通过从被认为困难的植入团问题进行归约，也已确定计算阈值 [HWX15, BBH18, BBH19]。然而，现有归约假设 p 和 q 同阶，因此，对于满足 $p = n^{-\alpha}$、$q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$ 的对数密度区间（其中常数 $0 < \alpha < \beta < 1$ 和 $\gamma \in (0, 1)$），目前还缺乏基于归约的困难性证据。在这里，该问题在 $\gamma = 1/2$ 处发生了定性变化。具体而言，当 $\gamma \geq 1/2$ 时，已知最佳的多项式时间算法只是对图中边的总数设定阈值；当 $\alpha < \beta/2+2(\gamma-1/2)$ 时，该算法即可成功。另一方面，当 $\gamma < 1/2$ 时，已知最佳的多项式时间算法是更精细的子图计数过程，在 $\alpha < \beta\gamma$ 时能够成功；参见 [BCC$^+$10] 的第 3.2 节（他们只给出了证明概要，因为其主要目标是为最坏情况图中的最稠密 k 顶点子图的近似提供保证）。

我们的主要结果表明，对于每个 $\gamma \in (0,1)$，低次检验都无法突破上述阈值，这为现有算法的最优性提供了证据。我们还给出了匹配的低次检验上界，证实上述算法均可由低次框架刻画；在 $\gamma < 1/2$ 情形下，我们的检验比 [BCC$^+$10] 的检验更简单，尽管他们的算法也能恢复植入子图。

相比之下，检测（即检验）的统计（即信息论）阈值与上述计算阈值不同，至少在某些参数范围内如此 [ACV14，VAC15]。具体而言，只要$\gamma \in (0，1/2]$，统计与计算阈值之间就存在差距。此外，当 $\gamma \in (1/2，2/3)$ 时，某些 β取值下也存在差距。我们将在下文讨论超图情形时进一步阐述。

尽管这并非我们的主要关注点，但我们注意到，关于恢复植入稠密子图（即识别植入顶点）的统计和计算方面也已有充分研究 [Ame15、CX16、Mon15、HWX17]。形式化地说，恢复问题至少与检测问题一样困难（可通过类似 [MW15第 5 节中的方法进行多项式时间归约1]）。值得注意的是，当 $\gamma > 1/2$时，恢复问题严格地更加困难，即检测与恢复的计算阈值不同（至少对于低次算法而言），恢复需要 $\alpha < \beta/2 + \gamma - 1/2$ [SW22]。当 $\gamma < 1/2$时，我们的结果表明，在与 [BCC$^+$10]的恢复算法相同的阈值处，检测具有低次数困难性，因而在此参数范围内不存在检测与恢复差距，从而解决了 [SW22第 2.4 节中遗留的一个问题1]。

低次数检验。 一个成功的 D 次数检验，是一个关于输入变量的 D 次多项式（在我们的情形中，输入变量是 $\binom{n}{r}$ 超边指示变量），其输出为实数，能够区分 （见定义 2.1）来自植入分布和零假设分布的样本。研究这类检验的想法源自一系列研究工作 [BHK$^+$19、HS17、HKP$^+$17、Hop18]；综述另见 [KWB22]。次数为 $O(\log n)$ 的检验通常被视为多项式时间检验的替代指标，因为它们涵盖了谱方法等主要算法途径。我们的上界将给出一个常数次数检验，从而得到一个用于检验的多项式时间算法；我们的下界则将排除所有次数为 $n^{o(1)}$ 的检验。

如今，证明低次测试界已有一种基于 低次似然比 的标准方法（见 [Hop18] 的第 2.3 节）：该方法归结为找出关于零假设分布的多项式标准正交基，并计算这些基多项式在植入分布下的期望。然而，我们的情形更加微妙，因为（对于$\gamma < 1/2$），低次似然比的二阶矩会因植入分布下罕见的“坏”事件而发散。因此，我们需要采用条件低次论证，即将植入分布限制在某个“好”事件上。

这种条件低次论证此前曾在少数情况下出现过 [BEH+22，CGH+22]，但我们的论证在技术层面有所不同。以往工作选择在某个事件上进行条件化，而这似乎会使直接使用正交多项式进行计算变得非常复杂；为克服这一困难，他们先将条件低次似然比与某个特定的“低重叠”二阶矩联系起来（在 Franz-Parisi 判据中也称为 [BEH+22]）。相比之下，我们的方法更为直接：我们谨慎地选择一个事件进行条件化，使得使用正交多项式的计算仍然易于处理。

稠密子超图的整数间隙 k-子图。与我们的结果互补的是一系列关于证明稠密 k-子图 [BCG+12、CM18、JPRX23]（以及 [CM18] 也研究了向超图的推广）的工作，旨在建立其凸松弛的整数间隙。这里关注的问题是在一个最坏情况的图中，寻找一个由 k 个顶点诱导的子图（给定 k），使其边密度能够保证在某个因子范围内接近此类最稠密的 k-子图。已知最佳多项式时间近似因子为 $\approx n^{1/4}$，这是由 [BCC+10]证明的。上述整数间隙结果表明，强大的凸松弛类也无法在此基础上取得改进。

事实上，这些整数间隙的证明表明，凸松弛在某些随机分布上会失效，例如我们所研究的那些分布。对于近似因子 $n^{1/4-\epsilon}$，其困难实例恰好就是参数为 $\alpha = 1/4 + \epsilon$， $\beta = 1/2$， $\gamma = 1/2$ 的检验问题（因此，我们的结果进一步表明，这一近似因子无法改进）。与我们的结果最相关的凸松弛研究，是 [JPRX23] 最近取得的结果，该结果是独立于我们工作的。我们将在下文详细比较。

与关于平方和的独立工作比较 [**JPRX23**]。在同期且相互独立的工作中，[JPRX23] 证明了 平方和（SoS）层级 （一个强大的半定规划松弛系列）在次数 $n^{\Omega(1)}$ 下无法改进上文所述的近似因子 $n^{1/4}$ 。这是关于最稠密 k子图问题整数间隙的最强已知结果。尽管人们猜测它们密切相关（参见 [HKP+17、Hop18]），但据我们所知，SoS 失效与低次检验失效之间不存在任何方向上的形式蕴含关系。

更具体地说， [JPRX23] 的作者考虑了与我们相同的假设检验问题（在图情形下为 $r = 2$），并写出了一个特定的 SoS 程序。该程序以零假设分布 $G(n, q)$ 的一个样本为输入，并尝试反驳存在一个含有 $(1 + o(1))\rho n$ 个顶点、具有 $(1 + o(1))p(\rho n)^2/2$ 条边的子图。请注意，在植入分布下，以高概率存在一个具有这种密度的子图，因此，如果 SoS 成功完成反驳任务（以高概率），那么这也就给出了一个具有相同参数 p, q, ρ 的检测（即检验）算法。 [JPRX23] 的结果表明，当 $\gamma < 1/2$ 且 $\alpha > \beta\gamma$ 时，SoS 无法解决反驳任务，这与我们的结果中的阈值相符。这为该区间内反驳问题的计算困难性提供了有力证据。然而，这未必意味着检测问题也存在困难

（这是我们工作的重点），因为检测在形式上是一个更容易的问题。

为说明上述观点，当 $\gamma > 1/2$ （这是 [JPRX23]未涵盖的一个参数区间）时，我们预计检测与反驳任务之间存在固有差距：多项式时间检测需要 $\alpha < \beta/2 + 2(\gamma - 1/2)$，而多项式时间反驳则需要 $\alpha < \beta/2 + \gamma - 1/2$；更多详情见附录A 。因此，在此参数区间内，我们预计，检查[JPRX23] 的 SoS 程序是否可行，是一种在检测问题上严格次优的算法。因此，我们的结果的一项优势在于，它直接针对检测问题，并刻画了所有 $\gamma \in (0,1)$。

上述讨论提醒我们，在将 SoS 下界作为检测（而非反驳）难度的证据时应当谨慎。不过，在本例中情况更为微妙，因为 [JPRX23] 的证明采用了伪校准方法 [BHK+19]，因此其分析中的一个关键步骤与我们证明的低次数检验下界密切相关。具体而言，他们证明低次似然比$\tilde{}$（用他们的记号表示为 $L_{\leq D}$ ） $\mathbb{E}[1]$在零假设分布下以高概率为$1 + o(1)$（和我们的结果一样，这里也涉及一些条件化论证）。我们的方法与之相关但有所不同：我们证明条件低次似然比的二阶矩（见第3节）为 $1+o(1)$，由此推出我们所需的结果（弱分离失效）。[JPRX23] 的证明还需要一些额外步骤，尤其是相当复杂且出色的分析，以证明他们的矩阵是半正定的。

植入稠密子超图。与植入稠密子图相比，针对超图的推广问题研究得较少。对于最坏情形下的最稠密子超图问题，[App12] 表明，基于伪随机性假设，要将最稠密子超图近似到某个 n^{ϵ} 因子以内（固定 $\epsilon > 0$ ）是困难的。在上述关于整数间隙的研究脉络中， [CM18] 的结果适用于 r-均匀超图，并为 Sherali-Adams 层级建立了 $\approx n^{(r-1)/4}$ 的整数间隙，与对数密度阈值相匹配。然而，与图情形不同， $r = 2$目前尚无多项式时间近似算法能在 $r \geq 4$上达到上述阈值，而 $r = 3$ 方面已由 [CDK+18]取得部分进展。

针对与我们相同的植入稠密子超图模型，[LZ22] 研究了检测问题和恢复问题。对于 $p = 1$ 和 $q = 1/2$ （即植入团）下的检测问题，已证明当 $\gamma < 1/2$时，次数为O的$(\log n)$多项式检验会失败。这与我们的结果互为补充，因为我们考虑的是 $p = n^{-\alpha}$ 和 $q = n^{-\beta}$，且 $0 < \alpha < \beta < r - 1$。对于恢复问题， [LZ22] 关注 $\gamma \geq 1/2$ 这一情形，并提出了一种在 $\alpha < \beta/2 + (r-1)(\gamma - 1/2)$ 时能够成功的高效算法。此外，若 $\alpha > \beta/2 + (r-1)(\gamma - 1/2)$，则已证明次数为 $\mathrm{polylog}(n)$ 的多项式无法恢复植入稠密子超图。请注意，这一恢复阈值不同于本文证明的检测阈值 $\alpha = \beta/2 + r(\gamma - 1/2)$；因此，这两个结果共同将[SW22] 中针对 $r = 2$ 、在 $\gamma \geq 1/2$情形下的检测与恢复差距推广到了超图情形。

另一条研究路线 [YS22、YS21a、YS21b] 关注植入稠密子超图的统计阈值。特别是，沿着 [ACV14], 的研究思路，[YS22] 证明了，检测的最优统计阈值要么由总度检验达到，要么由计算上效率低下的扫描检验达到。当前者占优时，检测阈值自然与我们的结果一致；但当扫描检验表现更好时，就会出现统计-计算间隙。更准确地说，当 $\alpha < \alpha_{\deg} := \beta/2 + r(\gamma - 1/2$时，总度检验成功；当 $\alpha < \alpha_{\mathrm{scan}} := \gamma(r - 1$时，扫描检验成功。如果 $\gamma \leq 1/2$，扫描检验严格优于我们的

条件 $\alpha < \beta\gamma$，因此存在统计-计算间隙。此外，当$1/2 < \gamma < r/(r+1)$ 且 $2r(\gamma - 1/2) < \beta < r - 2\gamma$ 时（其中，关于 γ 的第一个条件确保 β 的第二个区间非空），也存在间隙，因为此时 $\alpha_{\mathrm{scan}} > \alpha_{\mathrm{deg}}$ 且 $\alpha_{\mathrm{deg}} < \beta$ （注意，当$\alpha_{\mathrm{deg}} \geq \beta$时，总度检验对所有 $\alpha < \beta$ 都成功，因此不存在间隙）。

此外， [CPMB19、CPSB22] 通过带有加性高斯噪声的张量主成分分析模型，研究了植入稠密子超图的恢复问题。他们使用近似消息传递探究该问题的计算阈值，但其结果无法与我们的结果直接比较。最后，我们请读者参阅近期综述[LMFB23第 5 节，7] 了解更多关于稠密子超图的相关工作。

记号

令 $\mathbb{N}$ 表示正整数集。对于任意 $n \in \mathbb{N}$，令 $[n] := \{1, 2, \ldots, n\}$。在本文中，我们始终考虑 $n \to \infty$，并使用渐近记号 $O(\cdot)$、 $o(\cdot)$ 等。

对于固定整数 $r \geq 2$，令 K_n^r 表示定义在 n 个顶点上的完全 r-一致超图。对于任意超图 H，令 $V(H)$ 表示其顶点集，令 $E(H)$ 表示其边集。对于超图 H 和 $S \subseteq E(H)$，令 $H[S]$ 表示由 S 在 H 中诱导的子图。注意，任意超边集合 $S \subseteq E(K_n^r)$ 都可以视为子超图 $K_n^r[S]$，因此我们有时也将 S 视为子超图，不会产生歧义。为简便起见，我们常将超图和超边分别简称为图和边。

对于分布 $\mathcal{P}$，令 $\mathbb{E}_{\mathcal{P}}$ 表示在 $\mathcal{P}$下的期望；稍微滥用一下记号，也用 $\mathcal{P}$ 表示相应的概率。令 $\mathsf{Var}_{\mathcal{P}}(\cdot)$ 表示随机变量在 $\mathcal{P}$下的方差。令 $\mathsf{Ber}(p)$ 表示参数为 $p \in [0,1]$的伯努利分布，令 $\mathsf{Bin}(n,p)$ 表示参数为 $n \in \mathbb{N}$ 和 $p \in [0,1]$的二项分布。

2 主要结果

设 $n, r \in \mathbb{N}$满足 $n \geq r \geq 2$，并令 $0 < \alpha < \beta < r - 1$和 $\gamma \in (0, 1)$。定义 $M := \binom{n}{r}$。我们将考虑随机的 n顶点 r一致超图模型，其输出为无向超图 H，具有邻接张量 $Y \in \{0,1\}^M$。更准确地说，超边 e是 $[n]$的一个基数为 r的子集；令 $[M] := \binom{[n]}{r}$表示 K_n^r的所有超边构成的集合，并记为 $e \in [M]$。邻接张量 Y由 $e \in [M]$索引，而 $Y_e \in \{0,1\}$表示超边e的存在。

我们将稠密随机子超图的检测表述为分布 $\mathcal{P}$ 与 $\mathcal{Q}$之间的统计假设检验问题，定义如下：

- 在 $\mathcal{P}$下，令 $z_1, \ldots, z_n$为来自伯努利分布 $\mathsf{Ber}(\rho)$ 的独立同分布随机变量，其中 $\rho := n^{\gamma - 1}$，并令 $Z := \{i \in [n] : z_i = 1\}$。给定 Z，我们观测到一个具有独立超边的 r-均匀超图 H

$$Y_e \sim \begin{cases} \mathsf{Ber}(q) & \text{if } e \not\subseteq Z, \\ \mathsf{Ber}(p) & \text{if } e \subseteq Z, \end{cases}$$

  其中 $p := n^{-\alpha}$ 且 $q := n^{-\beta}$。

- 在 $\mathcal{Q}$ 下，我们观察到一个 r 一致的 Erd˝os-R´enyi 超图 H，其邻接张量为 Y，其中超边是独立同分布的 $Y_e \sim \mathsf{Ber}(q)$，且 $q := n^{-\beta}$。

为了探究检验 $\mathcal{P}$ 与 $\mathcal{Q}$之间的计算阈值，我们聚焦于低次多项式算法（例如 [Hop18、KWB22]）。令 $\mathbb{R}[Y]_{\le D}$ 表示以 Y 的各个元素为变量、次数至多为 D的多元多项式集合。稍微滥用一下记号，我们常说"一个多项式"来表示一列多项式 $f = f_n \in \mathbb{R}[Y]_{\le D}$，每个问题规模 n对应一个；此类多项式的次数 $D = D_n$ 可能随 n变化。为了研究多项式检验 $\mathcal{P}$ 与 $\mathcal{Q}$的能力，我们考虑 强分离和弱分离的概念，它们定义于[BEH+22]，前者强于后者。

定义 2.1 ([BEH+22]，定义 1.6)。根据 $n \to \infty$，若多项式 $f \in \mathbb{R}[Y]_{\le D}$ 满足以下条件

- *strongly separate* $\mathcal{P}$ *and* $\mathcal{Q}$ *if* $\sqrt{\mathrm{Var}_{\mathcal{P}}(f(Y)) \vee \mathrm{Var}_{\mathcal{Q}}(f(Y))} = o\left(|\mathbb{E}_{\mathcal{P}}[f(Y)] - \mathbb{E}_{\mathcal{Q}}[f(Y)]|\right)$;
- *weakly separate* $\mathcal{P}$ *and* $\mathcal{Q}$ *if* $\sqrt{\mathrm{Var}_{\mathcal{P}}(f(Y)) \vee \mathrm{Var}_{\mathcal{Q}}(f(Y))} = O\left(|\mathbb{E}_{\mathcal{P}}[f(Y)] - \mathbb{E}_{\mathcal{Q}}[f(Y)]|\right)$.

参见 [BEH+22]，其中详细讨论了为何这些条件对于假设检验而言是自然的。特别地，根据切比雪夫不等式，强分离意味着我们可以对 $f(Y)$ 进行阈值判定，以检验 $\mathcal{P}$ 是否不同于 $\mathcal{Q}$，且第一类和第二类错误均趋于零。我们的主要结果如下。

定理 2.2。假设我们观测到一个随机的 n 个顶点、r 均匀超图 $Y \in \{0,1\}^M$，它来自 $\mathcal{P}$ 或 $\mathcal{Q}$，其参数为 $p = n^{-\alpha}$、$q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$，其中 $r \ge 2$、$0 < \alpha < \beta < r-1$、$\gamma \in$ 属于 (0, 1)，且 $n \to \infty$ 和 $M = \binom{n}{r}$。

- （下界）假设以下条件之一成立：(1) $\gamma \ge 1/2$ 和 $\alpha > \beta/2 + r\,(\gamma - 1/2)$，或(2) $\gamma < 1/2$ 和 $\alpha > \beta\gamma$。若 $D = n^{o(1)}$ 那么，在 $\mathbb{R}[Y]_{\le D}$ 中不存在能够弱分离 $\mathcal{P}$ 和 $\mathcal{Q}$的多项式。
- （上界）假设以下条件之一成立：(1) $\gamma \ge 1/2$ 和 $\alpha < \beta/2 + r\,(\gamma - 1/2)$，或(2) $\gamma < 1/2$和 $\alpha < \beta\gamma$。存在一个正整数 D，它仅取决于（α、β、γ），以及一个 $\mathbb{R}[Y]_{\le D}$ 中的多项式，能够强分离 $\mathcal{P}$ 和 $\mathcal{Q}$。

因此，我们已经完全刻画了以下两种情形的低次数检测阈值：$\gamma \ge 1/2$和 $\gamma < 1/2$。定理2.2 中的下界证明见第3节开头。上界证明见第4节开头。特别地，在上界所述的每种情形中，都可以通过一个关于 Y各项的常数次数多项式实现强分离，因此该检验算法具有多项式时间复杂度。

3 检测下界

遵循低次多项式算法的框架 [Hop18、KWB22、BEH+22]，我们首先~介绍本节中使用的符号。令 Y_e 表示在 $\mathcal{Q}$下的标准化超边，即

$$\tilde{Y}_e := \frac{Y_e - q}{\sigma}, \tag{1}$$

其中 $\sigma := \sqrt{\mathrm{Var}_{\mathcal{Q}}(Y_e)} = \sqrt{q(1-q)}$。我们定义

$$\phi_S(Y) := \prod_{e \in S} \tilde{Y}_e, \quad \forall S \subseteq [M], \tag{2}$$

~注意，对于 $e \ne f$，在 $\mathcal{Q}$ 下，Y_e 和 Y_f 相互独立。因此，由于 Y_e 是中心化的，我们有

$$\mathbb{E}_{\mathcal{Q}}[\phi_S(Y)\phi_{S'}(Y)] = \begin{cases} 0 & \text{if } S \ne S', \\ 1 & \text{if } S = S', \end{cases}$$

因此，$\{\phi_S : S \subseteq [M]\}$ 是由 Y的函数构成的集合的一组标准正交基。

为了研究次不超过 $D > 0$ 的多项式，我们定义以下数量 y

$$\|L_{\leq D}\|^2 := \sum_{\substack{S\subseteq[M],\\ |S|\leq D}} (\mathbb{E}_{\mathcal{P}}[\phi_S(Y)])^2 . \tag{3}$$

考虑事件 $\mathcal{E}$ 以及在 $\mathcal{E}$ 上的条件分布 $\mathcal{P}'$ 时，我们类似地定义

$$\|L'_{\leq D}\|^2 := \sum_{\substack{S\subseteq[M],\\ |S|\leq D}} (\mathbb{E}_{\mathcal{P}'}[\phi_S(Y)])^2. \tag{4}$$

记号 $\|L_{\leq D}\|$ 表示以下范数：度为D 的似然比，即 $\mathcal{P}$ 与 $\mathcal{Q}$之间的似然比范数。我们注意到，$\|L_{\leq D}\|$ 与D中的量 LD([BEH+22，定义 1）相同。3], 等价性在 [Hop18第 2 节中得到了证明。3]。

定理2.2（下界）的证明。设 $\mathcal{E}$ 是一个事件，使得 $\mathcal{P}$（$\mathcal{E}$）$= 1 - o$（1）。令 $\mathcal{P}'$ 为将 $\mathcal{P}$ 在事件 $\mathcal{E}$上进行条件化后得到的分布。根据 [BEH+22命题 6，2], 为了证明 $\mathbb{R}[Y]_{\leq D}$ 中的任意多项式都不能弱分离 $\mathcal{P}$ 和 $\mathcal{Q}$，只需证明

$$\|L'_{\leq D}\|^2 = 1 + o(1). \tag{5}$$

如果 $\gamma \geq 1/2$ 和 $\alpha > \beta/2 + r\,(\gamma - 1/2)$，我们取 $\mathcal{E}$ 为完整样本空间，从而 $\mathcal{P} = \mathcal{P}'$，并在命题5中建立（3.1）。如果 $\gamma < 1/2$ 和 $\alpha > \beta\gamma$，我们定义（ $\mathcal{E}$ 11）中的高概率事件 $\mathcal{E}$，并在命题5中建立（3.5）。 □

3.1 大 γ 的下界

我们首先考虑 $\gamma \geq 1/2$ 的情形，即集合 Z 预期为“大”的情形。

命题 3.1. 设 $p = n^{-\alpha}$、 $q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$，其中 $r \geq 2$、 $0 < \alpha < \beta < r - 1$ 和 $\gamma \in$ 为固定的 (0, 1)，使得

$$\gamma \geq \frac{1}{2}, \quad \alpha > \frac{\beta}{2} + r\left(\gamma - \frac{1}{2}\right). \tag{6}$$

若 $D = n^{o(1)}$，则有

$$\|L_{\leq D}\|^2 = 1 + o(1).$$

考虑任意一个子图 $S \subseteq [M]$。我们首先计算 $\mathbb{E}_{\mathcal{P}}[\phi_S(Y)]$。

引理 3.2。对于 $S \subseteq [M]$，令 $\phi_S(Y)$ 为 (2)中定义的量，并令 $V(S) \subseteq [n]$ 表示由 S诱导的超图的顶点集。则有

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)] = \rho^{|V(S)|}\left(\frac{p-q}{\sigma}\right)^{|S|}.$$

证明。注意以下几点：

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)] = \mathbb{E}_Z\left[\mathbb{E}_{\mathcal{P}}[\phi_S(Y)|Z]\right].$$

注意到，给定 Z 后，Y_e 和 Y_f 在 $e \neq f$的条件下相互独立。特别地，我们有

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)|Z] = \prod_{e\in S} \mathbb{E}_{\mathcal{P}}[\tilde{Y}_e|Z].$$

此外，根据 $\mathcal{P}$的定义，我们有

$$\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e|Z] = \begin{cases} \frac{p-q}{\sigma} & e \subseteq Z \\ 0 & e \nsubseteq Z \end{cases}$$

有了这一点，我们便有

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)|Z] = \left(\frac{p-q}{\sigma}\right)^{|S|} \mathbb{I}\{V(S) \subseteq Z\},$$

由此可得

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)] = \rho^{|V(S)|}\left(\frac{p-q}{\sigma}\right)^{|S|},$$

正如所需。 □

令 $S_{\ell,\ m}$ 为所有具有 ℓ 个顶点和 m 条边的 K_n^r 的边诱导子超图所构成的集合，或者更正式地说，

$$S_{\ell,m} \coloneqq \{S \subseteq K_n^r \ :\ |V(S)| = \ell,\ |E(S)| = m\}. \tag{7}$$

以下不等式将在整个证明过程中使用，并且当 n 足够大时成立：

$$\frac{q}{e} \leq \sigma^2 \leq q, \quad p - q \geq \frac{p}{e}. \tag{8}$$

命题 *3.1* 的证明。注意，由于 S_{ℓ} 中的图m 是由边诱导的，且我们只考虑满足 $|S| \leq D$ 的$S \subseteq [M]$，因此有

$$r \leq \ell \leq rD, \quad \ell/r \leq m \leq D.$$

结合引理 3.2 的结果，我们可以得出

$$\|L_{\leq D}\|^2 = 1 + \sum_{\ell=r}^{rD} \sum_{m=\ell/r}^{D} \sum_{S\in S_{\ell,m}} \rho^{2\ell}\left(\frac{p-q}{\sigma}\right)^{2m},$$

其中，被加数 1 来自项 $S = \varnothing$。我们来计算$|S_{\ell,\ m}|$的一个上界。为此，我们先考虑从所有 n 个顶点中选取 ℓ 个顶点 $V(S)$ 的可能方式，然后再考虑从所有 $\binom{\ell}{r}$ 条可能的边中选取 m 条边 $E(S)$ 的可能方式。我们有

$$|S_{\ell,m}| \leq \binom{n}{\ell}\binom{\binom{\ell}{r}}{m} \leq n^{\ell}\binom{\ell}{r}^{m} \leq n^{\ell}\left(\frac{e\ell}{r}\right)^{rm} \leq n^{\ell}(eD)^{rm}, \tag{9}$$

因为 $\ell \le rD$。此外，由于 $p > q$，我们有 $(p-q)^2 \le p^2$. 现在我们有

$$\begin{aligned}\|L_{\le D}\|^2 &\le 1 + \sum_{\ell=r}^{rD} \sum_{m=\ell/r}^{D} n^{\ell} (eD)^{rm} \rho^{2\ell} \left(\frac{p-q}{\sigma}\right)^{2m} \\ &\le 1 + \sum_{\ell=r}^{rD} (n\rho^2)^{\ell} \sum_{m=\ell/r}^{D} \left(\frac{e^{r+1} D^r p^2}{q}\right)^{m} \\ &\le 1 + \sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell} \sum_{m=\ell/r}^{D} n^{(\beta-2\alpha+o(1))m},\end{aligned}$$

其中我们使用（8）以及参数所满足的假设。由（6），可得 $\beta - 2\alpha + o(1) < -r(2\gamma-1) - \delta$ 其中 $\delta = \delta(\alpha, \beta, \gamma, r) > 0$ 是一个固定常数。因此

$$\|L_{\le D}\|^2 \le 1 + \sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell} \sum_{m=\ell/r}^{D} n^{(-r(2\gamma-1)-\delta)m}.$$

由于 $\gamma \ge 1/2$，对于充分大的 n，有 $n^{-r(2\gamma-1)-\delta} < 1/2$。因此，

$$\|L_{\le D}\|^2 \le 1 + 2\sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell} n^{(-r(2\gamma-1)-\delta)\ell/r} = 1 + 2\sum_{\ell=r}^{rD} n^{-\delta\ell/r} \le 1 + 4n^{-\delta},$$

如所需。 □

3.2 小下界 γ

现在考虑 $\gamma < 1/2$ 的情形，并证明所需的阈值为 $\alpha > \beta\gamma$。在这种情况下，由于在 $\mathcal{P}$ 下存在某些稀有事件，度数为 D 的似然比 $L_{\le D}$（介于 $\mathcal{P}$ 和 $\mathcal{Q}$ 之间）的范数无界。为缓解这一问题，我们通过对这些稀有事件的补事件进行条件化来修改方法。然后，我们研究条件分布 $\mathcal{P}'$ 与 $\mathcal{Q}$ 之间的度数为 D 的似然比 $L'_{\le D}$。

3.2.1 高概率事件

回忆一下，$S_{\ell,\, m}$ 定义于（7）。取一个任意小的常数 $\delta = \delta(\alpha, \beta, \gamma) > 0$，稍后再选取。对于 $\ell \in \mathbb{N}$，定义

$$m_\ell := \left\lceil \ell \left(\frac{\gamma}{\alpha} + \delta\right) \right\rceil. \tag{10}$$

此外，我们定义一个整数对集合

$$I := \{(\ell, m) \in \mathbb{N}^2 \,:\, m_\ell \le m \le D,\, S_{\ell,m} \ne \varnothing\}.$$

令 $\mathcal{C} := H[Z]$ 表示由 Z 在观测图 H 中诱导出的子超图。换言之，$\mathcal{C}$ 是待检测的植入子超图。我们定义以下事件：

$$\begin{aligned}\mathcal{E}_{\ell,m} &:= \{\exists S \in S_{\ell,m} \text{ such that } S \subseteq \mathcal{C}\}, \\ \mathcal{E} &:= \bigcap_{(\ell,m)\in I} \mathcal{E}^c_{\ell,m},\end{aligned} \tag{11}$$

其中 $\mathcal{E}_{\ell,\ m}^{c}$ 是事件 $\mathcal{E}_{\ell,\ m}$的补事件。特别地，在 $\mathcal{E}$条件下， $\mathcal{C}$中不会出现任何诱导出稠密超图的边子集 S；这里"稠密"是指 S 的边顶点比至少为 $\frac{\gamma}{\alpha}+\delta$。我们定义一个分布 $\mathcal{P}'$ 如下：

$$\mathcal{P}'(Y) \coloneqq \frac{\mathcal{P}(Y)\mathbb{I}\{Y \in \mathcal{E}\}}{\mathcal{P}(\mathcal{E})}.$$

我们将证明 $\mathcal{P}(\mathcal{E}) = 1 - o(1)$。为辅助进行界定，我们先给出以下引理 .

引理 3.3。在模型 $\mathcal{P}$ 下，我们有

$$\frac{n^{\gamma}}{2} \le |Z| \le \frac{3\,n^{\gamma}}{2},$$

以至少为 $1 - 2\exp(-cn^{\gamma})$ 的概率成立，其中 $c > 0$ 为某个绝对常数。

证明。 注意， $|Z| = \sum_{i=1}^{n} z_i$ 是以下变量之和： n 独立同分布的$\mathsf{Ber}(\rho)$ 变量，即 $|Z| \sim \mathsf{Bin}(n,\rho)$。根据切尔诺夫界 [Ver18，练习 2.3，5] 可得

$$\mathcal{P}\left(|Z| \notin \left(1 \pm \frac{1}{2}\right)\mathbb{E}_{\mathcal{P}}[|Z|]\right) \le 2\exp\left(-C\,\frac{\mathbb{E}_{\mathcal{P}}[|Z|]}{4}\right),$$

其中 $C > 0$ 为某个绝对常数。引理由 $\mathbb{E}_{\mathcal{P}}[|Z|] = n\rho = n^{\gamma}$ 得证。 □

引理 3.4。我们有 $\mathcal{P}(\mathcal{E}) = 1 - o(1)$。

证明。 对于任意事件 A 和 B，我们有 $\mathbb{P}[A] \le \mathbb{P}[A \mid B] + \mathbb{P}[B^c]$. 令 A 为 $\mathcal{E}^c$，并令 B 为如下事件： $|Z|$ 如引理 3.3所述有界。因此，

$$\mathcal{P}(\mathcal{E}^c) \le \mathcal{P}\left(\mathcal{E}^c \,\middle|\, \frac{n^{\gamma}}{2} \le |Z| \le \frac{3\,n^{\gamma}}{2}\right) + 2\exp\left(-c\,n^{\gamma}\right). \tag{12}$$

现在，我们以 Z 的一个满足 $\frac{n^{\gamma}}{2} \le |Z| \le \frac{3\,n^{\gamma}}{2}$ 的实现为条件。为辅助证明，我们定义如下集合：

$$S_{\ell,m}(Z) \coloneqq \{S \in S_{\ell,m} : V(S) \subseteq Z\}.$$

与（9）中的论证完全相同，可得

$$|S_{\ell,m}(Z)| \le \left(\frac{3}{2}n^{\gamma}\right)^{\ell}(eD)^{rm}.$$

由此可得

$$\mathcal{P}(\mathcal{E}_{\ell,m} \mid Z) \le \sum_{S \in S_{\ell,m}(Z)} \mathcal{P}(S \subseteq E(\mathcal{C}) \mid Z) \le \left(\frac{3}{2}n^{\gamma}\right)^{\ell}(eD)^{rm} \max_{S \in S_{\ell,m}(Z)} \mathcal{P}(S \subseteq E(\mathcal{C}) \mid Z).$$

对于 $S \in S_{\ell,\ m}(Z)$，由于给定 Z 时， Y_e 和 Y_f 条件独立，因此我们有

$$\mathcal{P}(S \subseteq E(\mathcal{C}) \mid Z) = p^m = n^{-\alpha m}.$$

将此值代入上式，可得

$$\mathcal{P}(\mathcal{E}_{\ell,m} \mid Z) \le \left(\frac{3}{2}n^{\gamma}\right)^{\ell} (eD)^{rm}\, n^{-\alpha\, m} \le\ n^{\gamma\ell-(\alpha-o(1))\, m}.$$

现在我们来界定 $\mathcal{P}(\mathcal{E}^c \mid Z) \le \sum_{(\ell,\ m)\ \in I} \mathcal{P}(\mathcal{E}_{\ell,\ m} \mid Z)$。我们有

$$\mathcal{P}(\mathcal{E}^c \mid Z) \le \sum_{\ell=r}^{rD} \sum_{m=m_\ell}^{D} n^{\gamma\ell-(\alpha-o(1))\, m} = \sum_{\ell=r}^{rD} n^{\gamma\ell} \sum_{m=m_\ell}^{D} n^{-(\alpha-o(1))\, m} \le 2\sum_{\ell=r}^{rD} n^{\gamma\ell} n^{-(\alpha-o(1))\, m_\ell}.$$

由（10），有 $m_\ell \ge \ell\,(\frac{\gamma}{\alpha}+\delta)$，因此

$$\mathcal{P}(\mathcal{E}^c \mid Z) \le 2\sum_{\ell=r}^{rD} n^{-(\delta\alpha-o(1))\,\ell} \le 4n^{-\delta\alpha r/2}.$$

结合（12）即可完成证明。 □

3.2.2 界定低次范数

本小节的主要结果在于控制定义于（D 之间的次数-似然比的范数，其中 $\mathcal{P}'$ 和 $\mathcal{Q}$ 见（4）。

命题 3.5。设 $p = n^{-\alpha}$、$q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$ 具有固定的 $r \ge 2$、$0 < \alpha < \beta < r-1$ 和 $\gamma \in (0,1)$，使得

$$\gamma < \frac{1}{2}, \quad \alpha > \beta\,\gamma. \tag{13}$$

若 $D = n^{o(1)}$，则有

$$\|L'_{\le D}\|^2 = 1 + o(1).$$

为证明该命题，我们从（4）出发。注意到以下结论由 §3.2.1 得出：

$$\mathbb{E}_{\mathcal{P}'}[\phi_S(Y)] = \frac{1}{\mathcal{P}(\mathcal{E})}\,\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}] = (1+o(1))\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]. \tag{14}$$

现在只需考虑最后一项。我们首先对结果 Z 进行条件化，得到

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}] = \mathbb{E}_Z[\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z]].$$

考虑一个实现 Z。假设存在 $e \in S$ 使得 $e \not\subseteq Z$。由于 $\mathcal{E}$ 由 $\mathcal{C}$ 中的边决定，给定 Z 后，Y_e 与 $\phi_{S\setminus\{e\}}(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}$ 条件独立。特别地，有

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z] = \mathbb{E}_{\mathcal{P}}[\tilde{Y}_e \mid Z]\,\mathbb{E}_{\mathcal{P}}[\phi_{S\setminus\{e\}}(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z] = 0,$$

~ 作为 $e \not\subseteq Z$ 的 $\mathbb{E}_{\mathcal{P}}[Y_e \mid Z] = 0$。由此可得

$$\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}] = \mathbb{E}_Z[\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z]_{\mathbb{I}}\{V(S) \subseteq Z\}]. \tag{15}$$

我们将满足 $|S| \le D$ 的 $S \subseteq K_n^r$ 集合划分为“坏”子图集合和“好”子图集合，分别定义如下

$$\mathcal{B} := \bigcup_{(\ell,m)\in I} S_{\ell,m}, \quad \mathcal{G} := \{S \subseteq K_n^r \,:\, |S| \le D\} \setminus \mathcal{B}.$$

首先考虑一个良子图 S。我们证明以下引理。

引理 3.6. 设 $S \in \mathcal{G}$ 是一个含有 ℓ 个顶点和 m 条边的良子图。我们有

$$|\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]| \leq \rho^{\ell}\left(\frac{2p}{\sigma}\right)^{m}.$$

证明。 注意，由（15）和詹森不等式可得：

$$|\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]| \leq \mathbb{E}_Z[\mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z]_{\mathbb{I}}\{V(S) \subseteq Z\}]. \tag{16}$$

Let 考虑 $\mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z]$，对于 $V(S) \subseteq Z$。由于 $|\phi_S(Y)| \geq 0$，我们有 e

$$\mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z] \leq \mathbb{E}_{\mathcal{P}}[|\phi_S(Y)| \mid Z] = \left(\frac{p(1-q)+(1-p)q}{\sigma}\right)^{m} \leq \left(\frac{2p}{\sigma}\right)^{m},$$

最后一步由 $q < p$ 得出。现在我们有：

$$|\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]| \leq \left(\frac{2p}{\sigma}\right)^{m} \mathbb{E}_Z\left[{}_{\mathbb{I}}\{V(S) \subseteq Z\}\right] = \rho^{\ell}\left(\frac{2p}{\sigma}\right)^{m},$$

如所求。 □

接下来，考虑一个坏子图 S。我们将证明以下引理 a.

L引理 3.7。设 $S \in \mathcal{B}$ 是一个具有 ℓ 个顶点和 m 条边的坏子图。我们有 e

$$|\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]| \leq \rho^{\ell}\binom{m}{m_\ell - 1}\frac{q^{m-m_\ell+1}(2p)^{m_\ell-1}}{\sigma^m}.$$

证明。与引理 3.6 的证明类似，我们首先对 $\mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z]$ for $V(S) \subseteq Z$ 进行界定。根据 (11) 中 $\mathcal{E}$ 的定义， $Y \in \mathcal{E}$ 意味着 $\mathcal{C}$ 在 $S \in \mathcal{B}$ 中至多包含 $m_\ell - 1$ 条边，而 $S \in \mathcal{B}$ 有 m 条边。因此，对于 $s := m - m_\ell + 1 \geq 1$，存在 $e_1,\ldots,e_s \in S$，使得对每个 $i \in [s]$，都有 $Y_{e_i} = 0$。据此，我们有

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y \in \mathcal{E}\} \mid Z] &\leq \sum_{e_1,\ldots,e_s \in S} \mathbb{E}_{\mathcal{P}}[|\phi_S(Y)|_{\mathbb{I}}\{Y_{e_1} = \ldots = Y_{e_s} = 0\} \mid Z] \\ &\leq \sum_{e_1,\ldots,e_s \in S} \left(\frac{q}{\sigma}\right)^{s}\left(\frac{p(1-q)+(1-p)q}{\sigma}\right)^{m-s} \leq \binom{m}{s}\frac{q^s(2p)^{m-s}}{\sigma^m},\end{aligned}$$

其中最后一步成立是因为 $q < p$。再次由 (16)，可得

$$|\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]| \leq \binom{m}{s}\frac{q^s(2p)^{m-s}}{\sigma^m}\mathbb{E}_Z\left[{}_{\mathbb{I}}\{V(S) \subseteq Z\}\right] = \rho^{\ell}\binom{m}{s}\frac{q^s(2p)^{m-s}}{\sigma^m},$$

正如所需。 □

回顾我们最初要对 (4) 进行界定的目标。由 (14)，我们可以将 (4) 简化为

$$\begin{aligned}\|L'_{\leq D}\|^2 &= (1+o(1))\sum_{\substack{S \subseteq [M],\\ |S| \leq D}} \mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]^2 \\ &= (1+o(1))\left(\sum_{S \in \mathcal{G}} \mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]^2 + \sum_{S \in \mathcal{B}} \mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]^2\right).\end{aligned} \tag{17}$$

我们来考虑（17）中的第一个求和项。

引理 3.8. 我们有 $\sum_{S\in\mathcal{G}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y\in\mathcal{E}\}]^2 = 1+o(1)$。

证明。若 S是空图，那么 $\mathbb{E}_{\mathcal{P}}[\phi_S(Y)] = 1$。对于非空的 $S\in S_{\ell,\ m}\subseteq\mathcal{G}$，我们有 e

$$\frac{\ell}{r}\le m\le m_\ell-1,$$

因为 S 没有孤立顶点。由 (9) 和引理 3.6，我们得到

$$\begin{aligned}\sum_{S\in\mathcal{G}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y\in\mathcal{E}\}]^2 &\le 1+\sum_{\ell=r}^{rD}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}\sum_{S\in S_{\ell,m}}\rho^{2\ell}\left(\frac{2p}{\sigma}\right)^{2m}\\ &\le 1+\sum_{\ell=r}^{rD}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}n^{\ell}(eD)^{rm}\rho^{2\ell}\left(\frac{2p}{\sigma}\right)^{2m}\\ &\le 1+\sum_{\ell=r}^{rD}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}(n\rho^2)^{\ell}\left(\frac{e^{r+3}D^r\,p^2}{q}\right)^{m}\\ &= 1+\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}n^{(\beta-2\alpha+o(1))m},\end{aligned}$$

这里我们使用了 (8)。现在考虑两种情况。首先，假设 $\beta<2\alpha$。那么，对某个足够小的常数 $\delta_1=\delta_1(\alpha,\beta)>0$ 以及足够大的 n，我们有：

$$\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}n^{(\beta-2\alpha+o(1))m}\le\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}n^{-\delta_1 m}\le 2\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell-\delta_1\ell/r}.$$

由于 $\gamma<1/2$，我们可以进一步简化上式，得到

$$2\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell-\delta_1\ell/r}<2\sum_{\ell=r}^{rD}n^{-\delta_1\ell/r}\le 4n^{-\delta_1},$$

正如所需。现在，假设 $\beta\ge2\alpha$。我们有

$$\begin{aligned}\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell}\sum_{m=\ell/r}^{\min(m_\ell-1,D)}n^{(\beta-2\alpha+o(1))m} &\le\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell}\left(D-\frac{\ell}{r}+1\right)n^{(\beta-2\alpha+o(1))(m_\ell-1)}\\ &\le\sum_{\ell=r}^{rD}n^{(2\gamma-1)\ell+(\beta-2\alpha+o(1))(m_\ell-1)+o(1)}.\end{aligned}\tag{18}$$

现在考虑 $m_\ell-1$ 在 n的指数中的系数。根据 m_ℓ 在（10）中的定义，我们有

$$m_\ell-1<\ell\left(\frac{\gamma}{\alpha}+\delta\right).$$

有了这个结果，我们可以得到

$$(\beta-2\alpha+o(1))(m_\ell-1)<(\beta-2\alpha+o(1))\left(\frac{\gamma}{\alpha}+\delta\right)\ell<\left(\frac{\beta\gamma}{\alpha}-2\gamma+r\delta\right)\ell.$$

由此可知，式（18）中的指数可以界定为

$$(2\gamma-1)\ell+(\beta-2\alpha+o(1))(m_\ell-1)+o(1)\le\left(\frac{\beta\gamma}{\alpha}-1+2r\delta\right)\ell\le-\delta\ell, \tag{19}$$

其中最后一步由条件 $\alpha>\beta\gamma$（13）可知，只要我们选取 $\delta=\delta(\alpha,\beta,\gamma)>0$充分小。于是，（18）可进一步估计为

$$\sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell+(\beta-2\alpha+o(1))(m_\ell-1)+o(1)}\le\sum_{\ell=r}^{rD} n^{-\delta\ell}\le 2n^{-\delta r},$$

如所需。 □

现在，我们来考虑坏子图，并对（17）中的第二个求和项进行估计。

引理 3.9。我们有 $\sum_{S\in\mathcal{B}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)\,\mathbb{I}\{Y\in\mathcal{E}\}]^2=o(1)$。

证明。注意，对于 $S\in S_{\ell,m}\subseteq\mathcal{B}$，我们有

$$m_\ell\le m\le D.$$

由（9）和引理 3.7，我们有

$$\begin{aligned}
\sum_{S\in\mathcal{B}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)\,\mathbb{I}\{Y\in\mathcal{E}\}]^2 &\le \sum_{\ell=r}^{rD}\sum_{m=m_\ell}^{D}\sum_{S\in S_{\ell,m}}\rho^{2\ell}\binom{m}{m_\ell-1}^2\frac{q^{2(m-m_\ell+1)}\,(2p)^{2(m_\ell-1)}}{\sigma^{2m}}\\
&\le \sum_{\ell=r}^{rD}\sum_{m=m_\ell}^{D} n^\ell (eD)^{rm}\rho^{2\ell}\left(\frac{q}{\sigma}\right)^{2(m-m_\ell+1)}\left(\frac{2\,m\,p}{\sigma}\right)^{2(m_\ell-1)}\\
&\le \sum_{\ell=r}^{rD}\sum_{m=m_\ell}^{D}(n\rho^2)^\ell\left(\frac{e^r\,D^r\,q^2}{\sigma^2}\right)^m\left(\frac{2\,m\,p}{q}\right)^{2(m_\ell-1)}\\
&= \sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell}\,n^{2(\beta-\alpha+o(1))(m_\ell-1)}\sum_{m=m_\ell}^{D} n^{(-\beta+o(1))m},
\end{aligned}$$

其中我们使用（8）。注意，当 n 足够大时，指数中 m 的系数为负。因此，我们可以进一步将上式简化为

$$\begin{aligned}
\sum_{S\in\mathcal{B}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)\,\mathbb{I}\{Y\in\mathcal{E}\}]^2 &\le 2\sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell}\,n^{2(\beta-\alpha+o(1))(m_\ell-1)}\,n^{(-\beta+o(1))m_\ell}\\
&= 2\sum_{\ell=r}^{rD} n^{(2\gamma-1)\ell+(\beta-2\alpha+o(1))(m_\ell-1)-\beta+o(1)}.
\end{aligned}$$

我们再次考虑两种情况。首先，假设 $\beta<2\alpha$。由于 $\gamma<1/2$，当 n 足够大时，我们有：

$$\sum_{S\in\mathcal{B}}\mathbb{E}_{\mathcal{P}}[\phi_S(Y)\,\mathbb{I}\{Y\in\mathcal{E}\}]^2\le 2\sum_{\ell=r}^{rD} n^{-\beta+o(1)}\le 2rDn^{-\beta+o(1)}\le n^{-\beta/2},$$

如所需。现在，假设 $\beta \geq 2\alpha$。与证明（19）中的论证相同，可得

$$(2\gamma - 1)\ell + (\beta - 2\alpha + o(1))(m_\ell - 1) - \beta + o(1) \leq -\delta\ell - \beta.$$

特别地，我们有

$$\sum_{S \in \mathcal{B}} \mathbb{E}_{\mathcal{P}}[\phi_S(Y)_{\mathbb{I}}\{Y \in \mathcal{E}\}]^2 \leq 2\sum_{\ell=r}^{rD} n^{-\delta\ell-\beta} \leq 4n^{-\beta-\delta r},$$

正如所需。 □

命题 *3.5* 的证明。 只需将（17）与引理 3.8 和 3.9结合起来即可。 □

4 检测上界

与上一节一样，我们根据 γ 的取值将分析分为两种情况。对于 $\gamma \geq 1/2$，我们考虑 H 中超边总数的带符号计数，定义见（20）。特别地，它是邻接张量 Y 各项的线性函数。对于 $\gamma < 1/2$，我们的统计量是 平衡 子超图在 H 中出现次数的（无符号）计数，定义见（25）。子超图的大小仅取决于命题 α、β、γ 中给出的固定参数，如命题 4.4 所述。因此，（25）中的统计量是 Y 各项的常数次数多项式，并且可以在多项式时间内计算。

定理 *2.2*（上界）的证明。 若 $\gamma \geq 1/2$ 且 $\alpha < \beta/2 + r(\gamma - 1/2)$，命题 4.1 表明~，由（$T$ 20）定义的统计量 $\mathcal{P}$ 能将 $\mathcal{Q}$ 和 $\gamma < 1/2$ 强有力地区分开。若 $\alpha < \beta\gamma$ 且，命题 T 4.6 表明，由（25）定义的统计量 $\mathcal{Q}$ 能将 $\mathcal{P}$ 与 $\mathcal{Q}$ 强有力地区分开。 □

4.1 大 γ 的上界

在 $\gamma \geq 1/2$的情形下，我们将使用 H 中的超边有符号计数作为我们的检验统计量。回顾 $M = \binom{n}{r}$，其定义见 §2； $[M]$ 表示所有可能超边的集合。我们定义

$$\tilde{T} := \sum_{e \in [M]} \tilde{Y}_e, \tag{20}$$

~ ~其中 Y_e 定义于（1）中。显然， T 是 Y各个条目的一个一次多项式。

命题 4.1. 假设 $p = n^{-\alpha}$、 $q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$ 具有固定的 $r \geq 2$、 $0 < \alpha < \beta < r - 1$ 和 $\gamma \in$（0, 1），使得

$$\gamma \geq \frac{1}{2}, \quad \alpha < \frac{\beta}{2} + r\left(\gamma - \frac{1}{2}\right). \tag{21}$$

对于 $\tilde{T}$中定义的统计量 （20），有 $\sqrt{\text{方差}_{\mathcal{P}}(\tilde{T}) \vee \text{方差}_{\mathcal{Q}}(\tilde{T})} = o\left(\left|\mathbb{E}_{\mathcal{P}}[\tilde{T}] - \mathbb{E}_{\mathcal{Q}}[\tilde{T}]\right|\right)$.

首先，我们来计算每种分布下的期望和方差。

引理 4.2. 我们有

$$\mathbb{E}_{\mathcal{Q}}[\tilde{T}] = 0, \quad \mathrm{Var}_{\mathcal{Q}}(\tilde{T}) = M, \quad \mathbb{E}_{\mathcal{P}}[\tilde{T}] = M\,\rho^r \left(\frac{p-q}{\sigma}\right),$$
$$\mathrm{Var}_{\mathcal{P}}(\tilde{T}) \le M + \frac{2M\rho^r\,p}{\sigma^2} + \frac{2\,M\,rn^{r-1}\rho^{2r-1}\,p^2}{\sigma^2}.$$

˜证明。注意， Y_e 是通过在分布 $\mathcal{Q}$下对 Y_e 标准化而定义的，并且对于 $e \neq f$， Y_e 和 Y_f 在 $\mathcal{Q}$ 下相互独立。关于 $\mathbb{E}_{\mathcal{Q}}[T]$ 和 $\mathrm{Var}_{\mathcal{Q}}(t)$ 的结果可立即得出。

现在考虑一条边 $e \in [M]$。我们有

$$\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e] = \rho^r \left(\frac{p-q}{\sigma}\right) + (1-\rho^r)\cdot 0 = \rho^r \left(\frac{p-q}{\sigma}\right),$$

˜由此得到关于 $\mathbb{E}_{\mathcal{P}}[T]$的所需公式。

现在，关于 $\mathcal{P}$下的方差，我们有:

$$\mathrm{Var}_{\mathcal{P}}(\tilde{T}) = \sum_{e,f\in[M]} \mathrm{Cov}_{\mathcal{P}}(\tilde{Y}_e, \tilde{Y}_f).$$

˜ ˜注意，对于 $e \cap f = \varnothing$， Y_e 和 Y_f 相互独立。因此

$$\begin{aligned}
\mathrm{Var}_{\mathcal{P}}(\tilde{T}) &= \sum_{\substack{e,f\in[M],\\ e\cap f\neq\varnothing}} \mathrm{Cov}_{\mathcal{P}}(\tilde{Y}_e, \tilde{Y}_f) \\
&= \sum_{\substack{e,f\in[M],\\ e\cap f\neq\varnothing}} \left(\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f] - \rho^{2r}\left(\frac{p-q}{\sigma}\right)^2\right) \\
&\le \sum_{e\in[M]} \mathbb{E}_{\mathcal{P}}[\tilde{Y}_e^2] + \sum_{e\in[M]} \sum_{\substack{f\in[M]\setminus e,\\ e\cap f\neq\varnothing}} \mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f]. \qquad (22)
\end{aligned}$$

考虑第一项。我们有

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e^2] &= \rho^r \left(p\left(\frac{1-q}{\sigma}\right)^2 + (1-p)\left(\frac{q}{\sigma}\right)^2\right) + (1-\rho^r)\left(q\left(\frac{1-q}{\sigma}\right)^2 + (1-q)\left(\frac{q}{\sigma}\right)^2\right) \\
&\le \rho^r \left(\frac{p(1-q)^2 + (1-p)q^2}{\sigma^2}\right) + 1 \\
&\le \frac{2\rho^r\,p}{\sigma^2} + 1,
\end{aligned}$$

为 $q^2 < q < p$。对于式（22）中的第二项，我们先计算 $\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f]$。由于 $\tilde{Y}_e$和 $\tilde{Y}_f$在给定 Z的条件下相互独立，因此

$$\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f] = \mathbb{E}_Z[\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f \mid Z]] = \mathbb{E}_Z[\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e \mid Z]\mathbb{E}_{\mathcal{P}}[\tilde{Y}_f \mid Z]].$$

我们注意到以下几点:

$$\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e \mid Z] = \left(\frac{p-q}{\sigma}\right) \mathbb{I}\{e \subseteq Z\} + 0 \cdot \mathbb{I}\{e \not\subseteq Z\} = \left(\frac{p-q}{\sigma}\right) \mathbb{I}\{e \subseteq Z\}.$$

特别地，我们有

$$\mathbb{E}_{\mathcal{P}}[\tilde{Y}_e\tilde{Y}_f] = \mathbb{E}_Z\left[\left(\frac{p-q}{\sigma}\right)^2 \mathbb{I}\{e\cup f\subseteq Z\}\right] = \rho^{2r-|e\cap f|}\left(\frac{p-q}{\sigma}\right)^2 \le \frac{\rho^{2r-|e\cap f|}\,p^2}{\sigma^2},$$

其中最后一步 成立是因为 $p>q$。将这些界与（22）结合起来，我们现在得到

$$\mathrm{Var}_{\mathcal{P}}(\tilde{T}) \le M + \frac{2M\rho^r\,p}{\sigma^2} + \frac{\rho^{2r}\,p^2}{\sigma^2}\sum_{e\in[M]}\sum_{\substack{f\in[M]\setminus e,\\ e\cap f\neq\varnothing}}\rho^{-|e\cap f|}.$$

令 e固定。我们将考虑给出.....的上界

$$\sum_{\substack{f\in[M]\setminus e,\\ e\cap f\neq\varnothing}}\rho^{-|e\cap f|} \le \sum_{s=1}^{r-1}\sum_{v_1,\ldots,v_s\in e}\sum_{\substack{f\in[M],\\ v_1,\ldots,v_s\in f}}\rho^{-s}.$$

对于固定的 v_1、.....、 v_s， f至多有 $\binom{n}{r-s}\le n^{r-s}$种选择。类似地， v_1、.....、 v_s至多有$\binom{r}{s}\le r^s$种选择。因此

$$\sum_{\substack{f\in[M]\setminus e,\\ e\cap f\neq\varnothing}}\rho^{-|e\cap f|} \le \sum_{s=1}^{r-1} r^s\,n^{r-s}\,\rho^{-s} = n^r\sum_{s=1}^{r-1}\left(\frac{r}{n\rho}\right)^s.$$

注意到 $n\rho = n^\gamma \gg r$ ，因此我们有

$$\sum_{\substack{f\in[M]\setminus e,\\ e\cap f\neq\varnothing}}\rho^{-|e\cap f|} \le 2rn^{r-1}\rho^{-1}.$$

将此值代入方差计算，得到

$$\mathrm{Var}_{\mathcal{P}}(\tilde{T}) \le M + \frac{2M\rho^r\,p}{\sigma^2} + \frac{2\,M\,rn^{r-1}\rho^{2r-1}\,p^2}{\sigma^2},$$

如所愿。 □

命题的证明 *4.1*。 引理 4.2 意味着只需证明

$$\begin{aligned} &M^2\rho^{2r}\left(\frac{p-q}{\sigma}\right)^2 \gg M + \frac{2M\rho^r\,p}{\sigma^2} + \frac{2\,M\,rn^{r-1}\rho^{2r-1}\,p^2}{\sigma^2}\\ \iff\ &\rho^{2r}\left(\frac{p-q}{\sigma}\right)^2 \gg \frac{1}{M} + \frac{2\rho^r\,p}{M\sigma^2} + \frac{2\,rn^{r-1}\rho^{2r-1}\,p^2}{M\sigma^2}. \end{aligned}$$

首先，我们对左侧作下界估计。由（8），可得

$$LHS \ge \frac{\rho^{2r}\,p^2}{e^2q} = e^{-2}\,n^{2r(\gamma-1)+\beta-2\alpha}.$$

接下来，我们对右侧作上界估计。注意到 $M = \binom{n}{r} \geq \left(\frac{n}{r}\right)^r$，因此 $\frac{1}{M} = n^{-r+o(1)}$。由（8），可得

$$RHS \leq n^{-r+o(1)} + n^{r(\gamma-2)+\beta-\alpha+o(1)} + n^{-1+(2r-1)(\gamma-1)+\beta-2\alpha+o(1)}.$$

剩下需要验证的是

$$\begin{aligned}
&2r(\gamma-1)+\beta-2\alpha > -r+o(1),\\
&2r(\gamma-1)+\beta-2\alpha > r(\gamma-2)+\beta-\alpha+o(1),\\
&2r(\gamma-1)+\beta-2\alpha > -1+(2r-1)(\gamma-1)+\beta-2\alpha+o(1),
\end{aligned}$$

利用假设（21），这显然成立。 □

4.2 小 γ 时的上界

在本小节中，我们假设

$$\gamma < \frac{1}{2}, \quad \alpha < \beta\gamma. \tag{23}$$

与前一小节一样，我们将定义一个检验统计量 T，用于区分 $\mathcal{P}$ 和 $\mathcal{Q}$。

4.2.1 平衡超图

˜我们选择的统计量 T 将统计特定超图 H 在 H中的出现次数。我们首先定义平衡超图。

定义 4.3. 若超图 H 对每个非空的 $H' \subseteq H$都满足以下条件，则称其为平衡超图：

$$\frac{|E(H')|}{|V(H')|} \leq \frac{|E(H)|}{|V(H)|}.$$

˜现在定义我们关注的超图 H，其存在性由[RV86] 在$r = 2$ 以及 [MP22] 在一般情形下保证。

˜命题 4.4. 假设(23)，则存在一个平衡超图 H，它有 ℓ个顶点和 m 条边，并满足

$$\frac{1}{\beta} < \frac{m}{\ell} < \frac{\gamma}{\alpha}. \tag{24}$$

证明。 由于 $\alpha < \beta\gamma$ 由假设可知，存在一个有理数 λ 使得 $\frac{1}{\beta} < \lambda < \frac{\gamma}{\alpha}$。此外，注意到 $\lambda > \frac{1}{\beta} \geq \frac{1}{r-1}$. 根据 [MP22定理 2], ，存在一个平衡超图，它有 ℓ个顶点和 m 条边，使得 $\frac{m}{\ell} = \lambda$，证明完毕。 □

与（2）类似，我们定义

$$\psi_S(Y) \coloneqq \prod_{e\in S} Y_e, \quad \forall S \subseteq [M].$$

有了˜ H，我们便可以定义 T：

$$T \coloneqq \sum_{S\in\mathcal{S}} \psi_S(Y), \qquad \text{where } \mathcal{S} \coloneqq \{S \subseteq [M] \,:\, K_n^r[S] \cong \tilde{H}\}. \tag{25}$$

具体而言， T 统计 H 中与 H同构的边诱导子图数量。由于 H 中的边数 m 是仅依赖于（α、β、γ）的常数，因此统计量 T是关于 Y各项的常数次数多项式。我们指出，采用 T 的无符号计数而非有符号计数，主要是出于技术上的便利。 $\psi_S(Y)$ 的取值仅在$\{0,1\}$ 中这一事实，简化了 T方差的计算。

在转向本节的主要结果之前，我们先指出一个关于平衡超图的简单事实，这将在后面的证明中派上用场。

断言 4.5. 设 H 是一个平衡超图， H' 是 H 的真子超图，且$V(H') \subsetneq V(H)$。那么，

$$\frac{|E(H)|-|E(H')|}{|V(H)|-|V(H')|} \geq \frac{|E(H)|}{|V(H)|}.$$

证明。 若 $H'=\varnothing$，则该不等式显然成立。因此，我们可以假设 H' 非空。由于 H是平衡的，且 H' 非空，根据定义 4.3 可得

$$\frac{|E(H)|-|E(H')|}{|V(H)|-|V(H')|} \geq \frac{|E(H)|-|V(H')|\frac{|E(H)|}{|V(H)|}}{|V(H)|-|V(H')|} = \frac{|E(H)|(|V(H)|-|V(H')|)}{|V(H)|(|V(H)|-|V(H')|)} = \frac{|E(H)|}{|V(H)|},$$

如所求。 □

4.2.2 强分离

现在我们 证明，在所关注的情形下，统计量 T 可将 $\mathcal{P}$ 与 $\mathcal{Q}$ 明确区分开来 .

命题 4.6. 假设 $p=n^{-\alpha}$、 $q=n^{-\beta}$ 和 $\rho=n^{\gamma-1}$ 满足固定的 $r\geq 2$、 $0<\alpha<\beta<r-1$ 和 $\gamma\in(0,1)$，且 (23) 成立。由 (25) 定义的统计量 T满足 $\sqrt{\mathrm{Var}_{\mathcal{P}}(T)} \vee \mathrm{Var}_{\mathcal{Q}}(T) = o(|\mathbb{E}_{\mathcal{P}}[T]-\mathbb{E}_{\mathcal{Q}}[T]|)$。

我们先考虑 T 在 $\mathcal{Q}$下的情形。下面给出一个引理。

引理 4.7. 对于 T 如 (25)所定义，我们有

$$\mathbb{E}_{\mathcal{Q}}[T] = N q^m, \quad \mathrm{Var}_{\mathcal{Q}}(T) \leq \max\left\{ m^2 N n^{\ell(1-1/m)} \ell^{r(m-1)} q^{2m-1}, m^{m+1}\mathbb{E}_{\mathcal{Q}}[T] \right\},$$

其中 $N=|\mathcal{S}|$。

证明。由于在 $\mathcal{Q}$下， Y_e 与 Y_f 在 $e\neq f$ 时相互独立，因此注意到对于任意 $S\subseteq[M]$，我们有

$$\psi_S(Y) \sim \mathsf{Ber}(q^m),$$

由此可得关于 $\mathbb{E}_{\mathcal{Q}}[T]$ 的结果。对于方差，我们有

$$\mathrm{Var}_{\mathcal{Q}}(T) = \sum_{S_1,S_2\in\mathcal{S}} \mathrm{Cov}_{\mathcal{Q}}\left(\psi_{S_1}(Y),\psi_{S_2}(Y)\right).$$

Note 当 $S_1\cap S_2=\varnothing$ 时， $\psi_{S_1}(Y)$ 和 $\psi_{S_2}(Y)$ 相互独立。此外，我们有 e

$$\psi_{S_1}(Y)\psi_{S_2}(Y) = \prod_{e\in S_1} Y_e \prod_{e\in S_2} Y_e = \prod_{e\in S_1\cap S_2} Y_e^2 \prod_{e\in S_1\Delta S_2} Y_e = \prod_{e\in S_1\cap S_2} Y_e \prod_{e\in S_1\Delta S_2} Y_e = \prod_{e\in S_1\cup S_2} Y_e,$$

由于 $Y_e^2 = Y_e$。特别地，我们可以将方差界定为

$$\begin{aligned}
\mathrm{Var}_{\mathcal{Q}}(T) &= \sum_{\substack{S_1,S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} \mathrm{Cov}_{\mathcal{Q}}\left(\psi_{S_1}(Y),\psi_{S_2}(Y)\right) \\
&= \sum_{\substack{S_1,S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} \left(\mathbb{E}_{\mathcal{Q}}\left[\psi_{S_1}(Y)\psi_{S_2}(Y)\right] - \mathbb{E}_{\mathcal{Q}}\left[\psi_{S_1}(Y)\right]\,\mathbb{E}_{\mathcal{Q}}\left[\psi_{S_2}(Y)\right]\right) \\
&\le \sum_{\substack{S_1,S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} \mathbb{E}_{\mathcal{Q}}\left[\psi_{S_1\cup S_2}(Y)\right] \qquad (26)\\
&= \sum_{\substack{S_1,S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} q^{|S_1\cup S_2|} = q^{2m} \sum_{\substack{S_1,S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} q^{-|S_1\cap S_2|}.
\end{aligned}$$

固定 S_1。我们将考虑对.....进行界定

$$\sum_{\substack{S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} q^{-|S_1\cap S_2|} \le \sum_{m'=1}^{m} \sum_{\substack{S'\subseteq S_1,\\ |S'|=m'}} \sum_{\substack{S_2\in\mathcal{S},\\ S'\subseteq S_2}} q^{-m'}. \qquad (27)$$

我们先固定 S' ，并对最内层求和中 S_2 的选择数进行界定。剩余顶点 S_2 至多有 $n^{\ell-|V(S')|}$ 种选择，剩余边 $S_2\setminus S'$至多有 $\ell^{r(m-m')}$ 种选择。注意，由于 $S'\subseteq S_1$ 和$K_n^r[S_1]$ 同构于一个平衡的˜超图 H，因此我们有

$$\frac{|S'|}{|V(S')|} \le \frac{m}{\ell} \implies |V(S')| \ge \frac{\ell|S'|}{m}.$$

特别地，我们可以将(27)界定为

$$\sum_{\substack{S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} q^{-|S_1\cap S_2|} \le \sum_{m'=1}^{m} \sum_{\substack{S'\subseteq S_1,\\ |S'|=m'}} n^{\ell-|V(S')|}\,\ell^{r(m-m')}\,q^{-m'} \le n^{\ell}\,\ell^{rm} \sum_{m'=1}^{m} \left(\frac{m}{\ell^r\,n^{\ell/m}\,q}\right)^{m'},$$

这里我们利用了如下事实：给定 S_1时， S'至多有 $m^{m'}$ 种选择。注意，上述和由首项或末项主导。因此，我们得到

$$\sum_{\substack{S_2\in\mathcal{S},\\ S_1\cap S_2\neq\varnothing}} q^{-|S_1\cap S_2|} \le m n^{\ell}\,\ell^{rm} \max\left\{\frac{m}{\ell^r\,n^{\ell/m}\,q}, \left(\frac{m}{\ell^r\,n^{\ell/m}\,q}\right)^{m}\right\}.$$

有了这一结果，我们回到最初的方差计算，得到

$$\begin{aligned}
\mathrm{Var}_{\mathcal{Q}}(T) &\le m\,N\,q^{2m}\,n^{\ell}\,\ell^{rm} \max\left\{\frac{m}{\ell^r\,n^{\ell/m}\,q}, \left(\frac{m}{\ell^r\,n^{\ell/m}\,q}\right)^{m}\right\} \\
&= \max\left\{m^2\,N\,n^{\ell(1-1/m)}\,\ell^{r(m-1)}\,q^{2m-1}, m^{m+1}N\,q^m\right\},
\end{aligned}$$

正如所期望的。 □

在研究 T under $\mathcal{P}$ 的期望和方差之前，我们先证明以下引理。

引理 **4.8.** 设 S_1, $S_2 \in \mathcal{S}$ 如 (25)中所定义。假设 (23) 和 (24)成立。则有

$$\mathbb{E}_{\mathcal{P}}\left[\psi_{S_1\cup S_2}(Y)\right] \le 2^{2\ell}\,\rho^{|V(S_1\cup S_2)|}\,p^{|S_1\cup S_2|}.$$

证明。考虑 $V(S_1 \cup S_2) \cap Z$ 的所有可能实现，可得

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}\left[\psi_{S_1\cup S_2}(Y)\right] &= \sum_{V'\subseteq V(S_1\cup S_2)} \mathcal{P}(V(S_1\cup S_2)\cap Z = V')\,\mathcal{P}(\psi_{S_1\cup S_2}(Y)=1 \mid V(S_1\cup S_2)\cap Z = V') \\ &\le \sum_{V'\subseteq V(S_1\cup S_2)} \rho^{|V'|}p^{|E(V')|}\,q^{|S_1\cup S_2|-|E(V')|},\end{aligned}$$

其中， $E(V')$ 由 $S_1 \cup S_2$ 中所有顶点完全包含在 V' 中的边组成。我们断言，当 $V' = V(S_1 \cup S_2)$ 时，上述被加数达到最大值。由此断言可知，每个被加数都不超过 $\rho^{|V(S_1\cup S_2)|}\,p^{|S_1\cup S_2|}$，而 V' 的选择至多有 $2^{2\ell}$ 种。因此，结果得证。

为证明该断言，我们首先注意到，如果 V' 使 $\rho^{|V'|}p^{|E(V')|}\,q^{|S_1\cup S_2|-|E(V')|}$ 达到最大值，那么子图 $(V', E(V'))$ 不含孤立顶点。这是因为从 V' 中移除孤立顶点只会降低 ρ 中的指数，从而增大 $\rho < 1$ 下的整体值。因此，我们可以转而考虑问题

$$\max_{S'\subseteq S_1\cup S_2} \rho^{|V(S')|}\,p^{|S'|}\,q^{|S_1\cup S_2|-|S'|},$$

其中， $V(S')$ 是由 S' 诱导的顶点集合。我们将证明最大化者是 $S' = S_1 \cup S_2$。我们作如下定义：

$$S \coloneqq S_1 \cup S_2, \quad S_1' \coloneqq S' \cap S_1, \quad S_2' \coloneqq S' \setminus S_1'.$$

特 别是，$S_i' \subseteq S_i$ 和 $S_2' \cap S_1' = S_2' \cap S_1 = \varnothing$。目标是证明如下结论：

$$\rho^{|V(S)|}p^{|S|} \ge \rho^{|V(S')|}p^{|S'|}\,q^{|S|-|S'|},$$

这等价于证明

$$\rho^{|V(S)|-|V(S')|}\left(\frac{p}{q}\right)^{|S|-|S'|} \ge 1. \tag{28}$$

首先考虑 p/q 的指数。我们有

$$|S| - |S'| = |S_1| + |S_2| - |S_1 \cap S_2| - |S_1'| - |S_2'| = |S_1| - |S_1'| + |S_2| - |(S_1 \cap S_2) \cup S_2'|,$$

当 $S_2' \cap S_1 \cap S_2 = \varnothing$ 时。接下来，考虑 ρ 的指数。我们有

$$\begin{aligned}|V(S)| - |V(S')| &= |V(S_1)| + |V(S_2)| - |V(S_1)\cap V(S_2)| - |V(S_1')| - |V(S_2')| + |V(S_1')\cap V(S_2')| \\ &= |V(S_1)| - |V(S_1')| + |V(S_2)| - |V(S_1)\cap V(S_2)| - |V(S_2')| + |V(S_1')\cap V(S_2')|.\end{aligned}$$

我们注意到以下事实：

$$V(S_1')\cap V(S_2') \subseteq V(S_1)\cap V(S_2)\cap V(S_2'),$$

即 $S'_i \subseteq S_i$。因此，

$$\begin{aligned}
&|V(S)| - |V(S')| \\
&\le |V(S_1)| - |V(S'_1)| + |V(S_2)| - |V(S_1) \cap V(S_2)| - |V(S'_2)| + |V(S_1) \cap V(S_2) \cap V(S'_2)| \\
&\le |V(S_1)| - |V(S'_1)| + |V(S_2)| - |(V(S_1) \cap V(S_2)) \cup V(S'_2)| \\
&\le |V(S_1)| - |V(S'_1)| + |V(S_2)| - |V((S_1 \cap S_2) \cup S'_2)|,
\end{aligned}$$

其中，最后一个不等式是由以下事实得到的

$$V((S_1 \cap S_2) \cup S'_2) = V(S_1 \cap S_2) \cup V(S'_2) \subseteq (V(S_1) \cap V(S_2)) \cup V(S'_2).$$

现在可以化简（28）的左端，得到：

$$\begin{aligned}
&\rho^{|V(S)|-|V(S')|} \left(\frac{p}{q}\right)^{|S|-|S'|} \\
&\ge \rho^{|V(S_1)|-|V(S'_1)|+|V(S_2)|-|V((S_1\cap S_2)\cup S'_2)|} \left(\frac{p}{q}\right)^{|S_1|-|S'_1|+|S_2|-|(S_1\cap S_2)\cup S'_2|} \\
&= \rho^{|V(S_1)|-|V(S'_1)|} \left(\frac{p}{q}\right)^{|S_1|-|S'_1|} \rho^{|V(S_2)|-|V((S_1\cap S_2)\cup S'_2)|} \left(\frac{p}{q}\right)^{|S_2|-|(S_1\cap S_2)\cup S'_2|}.
\end{aligned}$$

因此，只需证明

$$\rho^{|V(S_1)|-|V(S'_1)|} \left(\frac{p}{q}\right)^{|S_1|-|S'_1|} \ge 1, \tag{29a}$$

$$\rho^{|V(S_2)|-|V((S_1\cap S_2)\cup S'_2)|} \left(\frac{p}{q}\right)^{|S_2|-|(S_1\cap S_2)\cup S'_2|} \ge 1. \tag{29b}$$

现在证明（29a）。（29b）的证明过程相同，只需将 S_1 替换为 S_2，并将 S'_1 替换为 $(S_1 \cap S_2) \cup S'_2$。首先，如果 $V(S'_1) = V(S_1)$，那么（29a）的左端可化简为 $(\frac{p}{q})^{|S_1|-|S'_1|}$，根据 $p > q$，其值至少为 1。因此，我们可以假设 $V(S'_1) \subsetneq V(S_1)$。回顾一下，$K_n^r[S_1]$ 与 $\tilde{}$ 同构于 H，这是一个具有 ℓ 个顶点和 m 条边的平衡超图。根据断言 4.5，我们有

$$\rho^{|V(S_1)|-|V(S'_1)|} \left(\frac{p}{q}\right)^{|S_1|-|S'_1|} = \left(\rho \left(\frac{p}{q}\right)^{\frac{|S_1|-|S'_1|}{|V(S_1)|-|V(S'_1)|}}\right)^{|V(S_1)|-|V(S'_1)|} \ge \left(\rho \left(\frac{p}{q}\right)^{m/\ell}\right)^{|V(S_1)|-|V(S'_1)|}.$$

注意

$$\rho \left(\frac{p}{q}\right)^{m/\ell} = n^{\gamma-1+(\beta-\alpha)m/\ell}.$$

只需证明上面的指数非负；若 $(\beta-\alpha)\, m \ge (1-\gamma)\, \ell$，则该条件成立。由（23）和（24），可得

$$(\beta-\alpha)m > (\beta-\beta\gamma)m = (1-\gamma)\beta m > (1-\gamma)\ell,$$

如所需。 □

现在我们来界定 T 在 $\mathcal{P}$下的期望和方差。

引理 4.9. 令 T 如 (25)中所定义，则有

$$\mathbb{E}_{\mathcal{P}}[T] \geq N\,\rho^{\ell}p^{m}, \quad \mathrm{Var}_{\mathcal{P}}(T) \leq 8^{\ell}\,N\,n^{\ell-1}\,\ell^{1+rm}\,\rho^{2\ell-1}p^{2m-m/\ell},$$

其中 $N = |\mathcal{S}|$。

证明。由于在给定 Z 的条件下， Y_e 和 Y_f 对于 $e \neq f$ 条件独立，我们注意到，对于特定的 $S \in \mathcal{S}$，有

$$\mathcal{P}\left(\psi_S(Y) = 1 \mid V(S) \subseteq Z\right) = p^{m}.$$

此外， $\psi_S(Y) \in \{0,1\}$，因此有

$$\mathbb{E}_{\mathcal{P}}[T] \geq \sum_{S\in\mathcal{S}} \mathcal{P}(V(S) \subseteq Z)\,\mathcal{P}\left(\psi_S(Y) = 1 \mid V(S) \subseteq Z\right) = N\rho^{\ell}p^{m}.$$

对于方差，我们注意到 z_i和 z_j对于 $i \neq j$相互独立，因此 $\psi_{S_1}(Y)$ 和 $\psi_{S_2}(Y)$ 在 $V(S_1) \cap V(S_2) = \varnothing$时相互独立。再次使用推导出 (26) 的论证，可得

$$\mathrm{Var}_{\mathcal{P}}(T) = \sum_{\substack{S_1,S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \mathrm{Cov}_{\mathcal{P}}\left(\psi_{S_1}(Y), \psi_{S_2}(Y)\right) \leq \sum_{\substack{S_1,S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \mathbb{E}_{\mathcal{P}}\left[\psi_{S_1\cup S_2}(Y)\right].$$

由引理4.8，于是得到

$$\begin{aligned}\mathrm{Var}_{\mathcal{P}}(T) &\leq \sum_{\substack{S_1,S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} 2^{2\ell}\rho^{|V(S_1\cup S_2)|}p^{|S_1\cup S_2|} \\ &= 2^{2\ell}\,\rho^{2\ell}\,p^{2m} \sum_{\substack{S_1,S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \rho^{-|V(S_1)\cap V(S_2)|}p^{-|S_1\cap S_2|}, \end{aligned} \tag{30}$$

其中我们用到了以下事实

$$|S_1 \cup S_2| = 2m - |S_1 \cap S_2|, \quad |V(S_1 \cup S_2)| = |V(S_1) \cup V(S_2)| = 2\ell - |V(S_1) \cap V(S_2)|.$$

与引理的证明4.7类似，我们固定 S_1 并对以下内容进行界定：

$$\sum_{\substack{S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \rho^{-|V(S_1)\cap V(S_2)|}p^{-|S_1\cap S_2|} \leq \sum_{\ell'=1}^{\ell} \sum_{\substack{V'\subseteq V(S_1),\\ |V'|=\ell'}} \sum_{\substack{S_2\in\mathcal{S},\\ V'\subseteq V(S_2)}} \rho^{-\ell'}p^{-|S_1\cap S_2|}. \tag{31}$$

我们首先固定一个 V' ，并对最内层求和中 S_2 的选择数量进行界定。在 S_2 中，其余顶点至多有 $n^{\ell-\ell'}$ 种选择，而 S_2 中的边至多有 ℓ^{rm} 种可能的选择。注意， $V(S_1 \cap S_2) \subseteq V(S_1) \cap V(S_2)$。此外，由于 $S_1 \cap S_2 \subseteq S_2$ 和 $K_n^r[S_2]$ 是一个平衡超图，我们有

$$\frac{|S_1 \cap S_2|}{|V(S_1) \cap V(S_2)|} \leq \frac{|S_1 \cap S_2|}{|V(S_1 \cap S_2)|} \leq \frac{m}{\ell}.$$

特别地，我们可以将31界定为

$$\sum_{\substack{S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \rho^{-|V(S_1)\cap V(S_2)|}p^{-|S_1\cap S_2|} \le \sum_{\ell'=1}^{\ell}\sum_{\substack{V'\subseteq V(S_1),\\ |V'|=\ell'}} n^{\ell-\ell'}\,\ell^{rm}\,\rho^{-\ell'}p^{-m\ell'/\ell}$$
$$\le n^{\ell}\,\ell^{rm}\sum_{\ell'=1}^{\ell}\left(\frac{\ell}{n\rho\,p^{m/\ell}}\right)^{\ell'},$$

其中我们利用了这样一个事实：给定 S_1， V' 至多有 $\ell^{\ell'}$ 种选择。现在考虑求和中的这一项。我们有

$$\frac{\ell}{n\rho\,p^{m/\ell}} = \ell\,n^{\alpha\,m/\ell-\gamma}.$$

注意，由（24）可得

$$\alpha\frac{m}{\ell}-\gamma<0.$$

特别地，对某个小常数 $\delta=\delta(\alpha,\beta,\gamma)>0$，我们有：

$$\frac{\ell}{n\rho\,p^{m/\ell}} \le \ell\,n^{-\delta} < \frac{1}{2}.$$

最后，我们可以如下界定（31）：

$$\sum_{\substack{S_2\in\mathcal{S},\\ V(S_1)\cap V(S_2)\neq\varnothing}} \rho^{-|V(S_1)\cap V(S_2)|}p^{-|S_1\cap S_2|} \le 2\,n^{\ell}\,\ell^{rm}\left(\frac{\ell}{n\rho\,p^{m/\ell}}\right).$$

有了这些结论，我们回到（30）中关于方差的原始计算，得到

$$\mathrm{Var}_{\mathcal{P}}(T') \le 2^{2\ell+1}\,N\,n^{\ell}\,\ell^{rm}\rho^{2\ell}\,p^{2m}\left(\frac{\ell}{n\rho\,p^{m/\ell}}\right) \le 8^{\ell}\,N\,n^{\ell-1}\,\ell^{1+rm}\,\rho^{2\ell-1}p^{2m-m/\ell},$$

正如所需。 □

命题 *4.6* 的证明。 设 $\lambda := N\,\rho^{\ell}\,p^{m}$。根据引理 4.9，我们有 $\mathbb{E}_{\mathcal{Q}}[T]\ge\lambda$。因此，只需证明

1. $\mathbb{E}_{\mathcal{Q}}[T]=o(\lambda)$,
2. $\mathrm{Var}_{\mathcal{Q}}(T)=o(\lambda^2)$,
3. $\mathrm{Var}_{\mathcal{P}}(T)=o(\lambda^2)$。

我们先考虑 $\mathbb{E}_{\mathcal{Q}}[T]$。根据引理 4.7，我们有

$$\frac{\mathbb{E}_{\mathcal{Q}}[T]}{\lambda}=\frac{Nq^m}{N\rho^{\ell}p^m}=n^{(1-\gamma)\ell-(\beta-\alpha)m}<n^{(1-\gamma)\ell-\beta(1-\gamma)m}=n^{(1-\gamma)(\ell-\beta m)}=o(1),$$

其中的界由条件 $\alpha<\beta\gamma$ （23）中的条件、 $\ell<m\beta$ （24）中的条件以及 $1-\gamma>0$得出。这就完成了条件1的证明。

现在，让我们考虑引理4.7中 $\mathrm{Var}_{\mathcal{Q}}(T)$ 的界。我们注意到以下内容

$$N \geq \binom{n}{\ell} \geq \left(\frac{n}{\ell}\right)^{\ell}.$$

结合这一点和上面对 $\mathbb{E}_{\mathcal{Q}}[T]$ 的界，我们得到

$$\frac{m^{m+1}\mathbb{E}_{\mathcal{Q}}[T]}{\lambda^2} \ll \frac{m^{m+1}}{\lambda} \leq \frac{m^{m+1}\ell^{\ell}}{n^{\ell}\rho^{\ell}p^{m}} = n^{\alpha m - \gamma\ell + o(1)} = o(1),$$

因为 $\alpha m < \gamma\ell$ 由（24）。现在，为了证明条件 2，还需对以下量给出界 ：

$$\begin{aligned}\frac{m^2 N n^{\ell(1-1/m)} \ell^{r(m-1)} q^{2m-1}}{\lambda^2} &= \frac{m^2 n^{\ell(1-1/m)} \ell^{r(m-1)} q^{2m-1}}{N\rho^{2\ell}p^{2m}} \\ &\leq \frac{m^2 n^{-\ell/m} \ell^{\ell+r(m-1)} q^{2m-1}}{\rho^{2\ell}p^{2m}} \\ &= n^{-\ell/m+\beta+2\ell(1-\gamma)-2m(\beta-\alpha)+o(1)}. \end{aligned} \tag{32}$$

考虑（32）中的指数。由（23）可得

$$\begin{aligned}-\ell/m + \beta + 2\ell(1-\gamma) - 2m(\beta-\alpha) &= \left(1-\gamma-\frac{1}{2m}\right)2\ell - \left(\beta\left(1-\frac{1}{2m}\right)-\alpha\right)2m \\ &< \left(1-\gamma-\frac{1}{2m}\right)2\ell - 2m\beta\left(1-\gamma-\frac{1}{2m}\right) \\ &= 2\left(1-\gamma-\frac{1}{2m}\right)(\ell - m\beta).\end{aligned}$$

由于 $\gamma < 1/2$ 且 $m \geq 1$，我们有 $1-\gamma-1/(2m) > 0$。此外， $\ell < m\beta$ 由（24）可得。因此，上述量为负，所以（32）是 $o(1)$。综上，我们已经证明了 $\mathrm{Var}_{\mathcal{Q}}(T)/\lambda^2 = o(1)$，这就是条件 2。

最后，考虑引理 4.9中 $\mathrm{Var}_{\mathcal{P}}(T)$ 的界。我们有

$$\frac{\mathrm{Var}_{\mathcal{P}}(T)}{\lambda^2} \leq \frac{8^{\ell} N n^{\ell-1} \ell^{1+rm} \rho^{2\ell-1} p^{2m-m/\ell}}{N^2\rho^{2\ell}p^{2m}} \leq \frac{8^{\ell}\ell^{1+rm+\ell}}{n\rho\, p^{m/\ell}} = n^{\alpha m/\ell - \gamma + o(1)} = o(1),$$

其中我们再次使用 $N \geq \left(\frac{n}{\ell}\right)^{\ell}$ 和 $\alpha m/\ell - \gamma < 0$ 由（24）。这证明了条件 3。 □

检测-反驳间隙

这里我们论证这一主张：对于 $\gamma > 1/2$，我们预期存在固有的检测-反驳间隙，这意味着基于检验凸松弛（例如 [JPRX23] 的平方和规划）可行性的检验，在检测方面严格次优。我们始终聚焦于图情形 $r = 2$。

我们先正式定义反驳任务。如果一个图中存在一个包含 $1 \pm o(1)\rho n$ 个顶点、含有 $1 \pm o(1)p(\rho n)^2/2$ 条边的子图，我们就称该图含有一个“稠密子图”。其中每个 $o(1)$ 都表示某个特定的 $o(1)$ 量，其选取方式保证植入分布 $\mathcal{P}$ (定义见第 2 节）以高概率（即概率为 $1 - o(1)$）含有稠密子图。反驳是如下算法任务：给定一个图 G，目标是输出 NO 或 MAYBE，并满足以下

两项保证：（1）如果 G 含有稠密子图，则输出必须为 MAYBE；（2）如果 G 是从零假设分布 $\mathcal{Q} = G(n,q)$ 中抽取的，则以概率 $1-o(1)$ 输出必须为 NO。直观来说，关键在于算法只有在绝对确定不存在稠密子图时，才允许输出 NO。以往研究中曾考虑过这类反驳（或认证）任务，例如 [WBP16、BKW20、BMR21、BBK+21、BKW22、KVWX23]。

解决反驳问题的一种自然方法，是检验用于判断是否存在稠密子图的凸松弛是否可行。如果该松弛在 $\mathcal{Q}$ 上以高概率不可行（对于某些参数 p、q、ρ），我们就得到一个成功的反驳算法：若松弛不可行，则输出 NO；否则输出 MAYBE。（注意，由于这是一个松弛，如果松弛不可行，就能证明不存在稠密子图。）这也给出了一个成功的检测算法：若松弛不可行，则输出“$\mathcal{Q}$”；否则输出“$\mathcal{P}$”。（注意，根据设计，该松弛在 $\mathcal{P}$ 上以高概率可行，因为稠密子图确实存在。）

本节的目标是论证，对于 $\gamma > 1/2$，只要 $\alpha > \beta/2 + \gamma - 1/2$，反驳问题就是计算困难的，因此它严格难于检测（因为只要 $\alpha < \beta/2 + 2$（$\gamma - 1/2$），检测就很容易）。因此，我们预计，任何基于检验凸松弛可行性的测试（按上述含义）都不可能是最优检测方法；否则，就意味着存在一个好得令人难以置信的反驳算法。事实上，如果该松弛能够成功检测，那么它在 $\mathcal{Q}$ 上必定以高概率不可行，这意味着它也能成功反驳。

我们将以类似于 [BKW20] 的方式论证反驳的困难性：我们将构造一种~不同的植入分布 $\mathcal{P}$，它与 $\mathcal{P}$ 一样，以高概率包含一个稠密子图；随后我们将证明~区分 $\mathcal{P}$ 与 $\mathcal{Q}$的低次数困难性，从而推测不存在多项式时间~算法能够区分 $\mathcal{P}$ 和 $\mathcal{Q}$。若这一猜想成立，它便形式化地蕴含了反驳的困难性~，因为成功的反驳算法可用于区分 $\mathcal{P}$ 和 $\mathcal{Q}$，如上文所述。

总而言之，本节剩余部分的目标是构造一个明若 $\gamma > 1/2$ 且 $\alpha > \beta/2 + \gamma - 1/2$，~则不存在次数为 $n^{o(1)}$ 的多项式能够多项式弱分离 $\widetilde{\mathcal{P}}$ 和 $\mathcal{Q}$。

~ ~构造 $\mathcal{P}$。现在我们构造一个辅助植入分布 $\mathcal{P}$，它具有一个植入的稠密子图，但与 $\mathcal{Q} = G(n,q)$ 相比，更难与 $\mathcal{P}$ 区分。与正文一样，我们固定参数 $0 < \alpha < \beta < 1$ 和 $\gamma \in (0,1)$，并考虑尺度关系 $p = n^{-\alpha}$、$q = n^{-\beta}$ 和 $\rho = n^{\gamma-1}$。以一种不同寻常的方式对观测图进行参数化会很方便：对于 $i < j$，令 $Y_{ij} = a := \sqrt{(1-q)/q}$，若边 (i,j) 存在；否则令 $Y_{ij} = b := -\sqrt{q}/(1-q)$。(这确保了 $\mathbb{E}_{\mathcal{Q}}[Y_{ij}] = 0$ 和 $\mathbb{E}_{\mathcal{Q}}[Y_{ij}^2] = 1$。) 从 $\widetilde{\mathcal{P}}$ 中按如下方式采样一张图。与之前一样，采样一个植入顶点集合 $Z \subseteq [n]$，其中每个顶点都以概率 ρ 独立纳入。通过 $u_i = \sqrt{(1-\rho)/\rho}$（若 $i \in Z$）以及 $u_i = -\sqrt{\rho/(1-\rho)}$（若 $i \notin Z$）定义 $u \in \mathbb{R}^n$。(这确保了 $\mathbb{E}[u_i] = 0$ 和 $\mathbb{E}[u_i^2] = 1$。) 现在，在给定 u 的条件下，令 $Y = (Y_{ij})_{i<j}$ 的各项独立取值 $Y_{ij} \in \{a,b\}$，使得 $\mathbb{E}_{\widetilde{\mathcal{P}}}[Y_{ij}|u] = \lambda u_i u_j$。选择标量 λ，使植入边具有所需的概率；也就是说，当 $i,j \in Z$ 时，边 (i,j) 以概率 p 出现；由此得到 $\lambda = (1+o(1))p\rho/\sqrt{q}$。

低次数困难性的证明。现在我们将证明，如果 $\gamma > 1/2$，$\alpha > \beta/2+\gamma-1/2$，且 $D = n^{o(1)}$，~ 那么 $\|L_{\le D}\| = 1 + o$（1），这排除了如第 3 节所述的弱分离。根据 [BBK+21，命题 B。1] 再结合揭示额外观测值（Y_{ij}）$_{i\ge j}$ 只能增加

$\|\widetilde{L_{\le D}}\|$,

$$\|\widetilde{L}_{\le D}\|^2 \le \sum_{d=0}^{D} \frac{1}{d!}\mathbb{E}\langle \lambda uu^\top, \lambda vv^\top\rangle^d = \sum_{d=0}^{D} \frac{\lambda^{2d}}{d!}\mathbb{E}\langle u, v\rangle^{2d},$$

其中 u 如上所定义，且 v 是 u的一个独立副本。注意， $\langle u,\ v\rangle$ 是由 n 个独立同分布的有界随机变量之和。按照[SW22 的引理 B3], ，我们可以得出

$$\mathbb{E}\langle u, v\rangle^{2d} \le \sqrt{2\pi}\rho^{-2d}\left[(4d\rho^2 n)^d + (8d/3)^{2d}\right].$$

结合以上结果， $\|\widetilde{L}_{\le D}\| = 1 + o(1)$成立，条件是 $\lambda \ll \rho$ 和 $\lambda^2 n \ll 1$。由于 $\gamma > 1/2$，这就归结为所需条件 $\alpha > \beta/2 + \gamma - 1/2$ 。

致谢

感谢 Guy Bresler 和 Aaron Potechin 提供了有益的讨论。

参考文献

[ACV14] Ery Arias-Castro 和 Nicolas Verzelen。稠密随机网络中的社群检测。《统计年刊》，42(3):940 – 969，2014年。[Ame15] Brendan PW Ames。通过凸松弛保证恢复植入团和稠密子图。《优化理论与应用杂志》，167:653–675，2015年。[App12] Benny Applebaum。基于随机局部单向函数、具有长扩展和低局部性的伪随机生成器。载于第四十四届美国计算机协会计算理论年会论文集，第805–816页，2012年。[BB20] Matthew Brennan 和 Guy Bresler。秘密泄露带来的可约性与统计-计算间隙。载于学习理论会议，第648–847页。PMLR，2020年。[BBH18] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。植入稀疏结构问题的可约性与计算下界。载于学习理论会议，第48–166页。PMLR，2018年。[BBH19] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。子矩阵检测计算下界的普适性。载于学习理论会议，第417–468页。PMLR，2019年。[BBK+21] Afonso S Bandeira、Jess Banks、Dmitriy Kunisky、Christopher Moore 和 Alexander S Wein。随机图中割、可着色性和社群的谱植入及其否证难度。载于学习理论会议，第410–473页。PMLR，2021年。[BCC+10] Aditya Bhaskara、Moses Charikar、Eden Chlamtac、Uriel Feige 和 Aravindan Vijayaraghavan。检测高对数密度：最稠密 k-子图问题的一个 $O(n^{1/4})$ 近似算法。载于第四十二届美国计算机协会计算理论研讨会论文集。美国计算机协会，2010年。

[BCG+12] Aditya Bhaskara、Moses Charikar、Venkatesan Guruswami、Aravindan Vijayaraghavan 和 Yuan Zhou。强半定规划松弛的最稠密 k-子图问题的多项式整数间隙。载于第二十三届 ACM-SIAM 离散算法研讨会论文集，第388–405页。SIAM，2012年。[BEH+22] Afonso S Bandeira、Ahmed El Alaoui、Samuel Hopkins、Tselil Schramm、Alexander S Wein 和 Ilias Zadik。Franz-Parisi 判据与高维统计中的计算权衡。载于《神经信息处理系统进展》，35:33831–33844，2022年。[BHK+19] Boaz Barak、Samuel Hopkins、Jonathan Kelner、Pravesh K Kothari、Ankur Moitra 和 Aaron Potechin。植入团问题的近乎紧致的平方和下界。《SIAM 计算杂志》，48(2):687–735，2019年。[BI13] Cristina Butucea 和 Yuri I Ingster。高维含噪矩阵中稀疏子矩阵的检测。Bernoulli，19(5B):2652–2688，2013年。[BKW20] Afonso S Bandeira、Dmitriy Kunisky 和 Alexander S Wein。受约束主成分分析问题中认证界限的计算困难性。载于第十一届理论计算机科学创新会议（ITCS 2020），第151卷，第78页。Schloss Dagstuhl-Leibniz 信息学中心，2020年。[BKW22] Afonso Bandeira、Dmitriy Kunisky 和 Alexander Wein。非负主成分分析的平均情形整数间隙。载于《数学与科学机器学习》，第153–171页。PMLR，2022年。[BMR21] Jess Banks、Sidhanth Mohanty 和 Prasad Raghavendra。局部统计、半定规划与社群检测。载于2021年 ACM-SIAM 离散算法研讨会（SODA）论文集，第1298–1316页。SIAM，2021年。[CDK+18] Eden Chlamtac、Michael Dinitz、Christian Konrad、Guy Kortsarz 和 George Rabanca。最稠密 k-子超图问题。《SIAM 离散数学杂志》，32(2):1458–1477，2018年。[CGH+22] Amin Coja-Oghlan、Oliver Gebhard、Max Hahn-Klimroth、Alexander S Wein 和 Ilias Zadik。群体测试中的统计与计算相变。载于学习理论会议，第4764–4781页。PMLR，2022年。[CM18] Eden Chlamtac 和 Pasin Manurangsi。与对数密度阈值相匹配的 Sherali-Adams 整数间隙。arXiv 预印本 arXiv:1804.07842，2018年。[CPMB19] Luca Corinzia、Paolo Penna、Luca Mondada 和 Joachim M Buhmann。一类社群检测生成模型的精确恢复。载于2019年 IEEE 国际信息论研讨会（ISIT），第415–419页。IEEE，2019年。[CPSB22] Luca Corinzia、Paolo Penna、Wojciech Szpankowski 和 Joachim Buhmann。植入最稠密 k-子超图问题的统计与计算阈值。载于人工智能与统计国际会议，第11615–11640页。PMLR，2022年。

[CX16] Yudong Chen 和 Jiaming Xu。《植入问题中的统计-计算权衡，以及簇和子矩阵数量增长时的子矩阵定位》。《机器学习研究杂志》，17(1):882–938，2016年。[For10] Santo Fortunato。《图中的社群检测》。《物理学报告》，486(3-5):75–174，2010年。[FS97] Uriel Feige 和 Michael Seltser。《关于最稠密 k-子图问题》。1997年。[HKP+17] Samuel B Hopkins、Pravesh K Kothari、Aaron Potechin、Prasad Raghavendra、Tselil Schramm 和 David Steurer。《平方和方法在检测隐藏结构方面的威力》。载于第 58 届计算机科学基础年会（FOCS），第720–731页。IEEE，2017年。[Hop18] Samuel Hopkins。《统计推断与平方和方法》。博士论文，康奈尔大学，2018年。[HS17] Samuel B Hopkins 和 David Steurer。《利用少量样本进行高效贝叶斯估计：社群检测及相关问题》。载于第58届计算机科学基础年会（FOCS），第379–390页。IEEE，2017年。[HWX15] Bruce Hajek、Yihong Wu 和 Jiaming Xu。《随机图上社群检测的计算下界》。载于学习理论会议，第899–928页。PMLR，2015年。[HWX17] Bruce Hajek、Yihong Wu 和 Jiaming Xu。《恢复隐藏社群的信息极限》。《IEEE信息论汇刊》，63(8):4729–4745，2017年。[Jer92] Mark Jerrum。《大型团难以被 Metropolis 过程捕获》。《随机结构与算法》，3(4):347–359，1992年。[JPRX23] Chris Jones、Aaron Potechin、Goutham Rajendran 和 Jeff Xu。《最稠密 k 子图的平方和下界》。arXiv 预印本 arXiv:2303.17506，2023年。[Kar76] R Karp。《若干组合搜索问题的概率分析》。《算法与复杂性：新方向与近期成果》，1976年。[KVWX23] Pravesh K Kothari、Santosh S Vempala、Alexander S Wein 和 Jeff Xu。《植入着色是否比植入团更容易？》arXiv 预印本 arXiv:2303.00252，2023年。[KWB22] Dmitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira。《关于假设检验计算困难性的札记：利用低次似然比进行预测》。载于《数学分析、应用与计算：ISAAC 2019》，第1–50页。Springer，2022年。[LMFB23] Tommaso Lanciano、Atsushi Miyauchi、Adriano Fazzone 和 Francesco Bonchi。《最稠密子图问题及其变体综述》。arXiv 预印本 arXiv:2303.14467，2023年。[LZ22] Yuetian Luo 和 Anru R Zhang。《具有植入结构的张量聚类：统计最优性与计算极限》。《统计年刊》，50(1):584–613，2022年。

[Mon15] Andrea Montanari。《在稀疏图中寻找一个社群》。《统计物理学杂志》，161:273–299，2015年。[MP22] AD Matushkin 和 SN Popova。《严格平衡均匀超图与零一律的推广》。《离散数学》，345(6):112835，2022年。[MW15] Zongming Ma 和 Yihong Wu。《极小极大子矩阵检测中的计算障碍》。《统计年刊》，43(3):1089–1116，2015年。[RV86] Andrzej Ruci´nski 和 Andrew Vince。《强平衡图与随机图》。《图论杂志》，10(2):251–264，1986年。[SW22] Tselil Schramm 和 Alexander S Wein。《低次多项式估计的计算障碍》。《统计年刊》，50(3):1833–1858，2022年。[VAC15] Nicolas Verzelen 和 Ery Arias-Castro。《稀疏随机网络中的社群检测》。《应用概率年鉴》，25(6):3465–3510，2015年。[Ver18] Roman Vershynin。《高维概率：数据科学中的应用导论》，第47卷。剑桥大学出版社，2018年。[WBP16] Tengyao Wang、Quentin Berthet 和 Yaniv Plan。《RIP 认证的平均情形困难性》。《神经信息处理系统进展》，29，2016年。[YS21a] Mingao Yuan 和 Zuofeng Shang。《异质稠密子超图检测》。arXiv 预印本 arXiv:2104.04047，2021年。[YS21b] Mingao Yuan 和 Zuofeng Shang。《检测子超图的信息极限》。《Stat》，10(1):e407，2021年。[YS22] Mingao Yuan 和 Zuofeng Shang。《稠密子超图检验的精确检测边界》。《Bernoulli》，28(4):2459–2491，2022年。