arXiv:2602.07669v1 [math.ST] 7 Feb 2026

植入匹配与生成树的统计阈值

Louigi Addario-Berry* Omer Angel† G´abor Lugosi‡ Mikl´os Z. R´acz§ Tselil Schramm¶

2026年2月10日

摘要

本文研究在 Erd˝os–R´enyi 随机图中检测是否存在植入式完美匹配或生成树的问题。更准确地说，我们研究这样一个假设检验问题：统计学家观察一个有 n 个顶点的图。在零假设下，该图是 Erd˝os–R´enyi 随机图 $G(n,\ q)$ 的一个实现；而在备择假设下，该图由一个 Erd˝os–R´enyi 随机图与一个随机完美匹配（或随机生成树）的并集构成。为了避免通过统计边数进行简单检测，我们调整备择假设，使两种分布下的边数期望相同。我们证明，在这两个问题中，当 $q \gg n^{-1/2}$时，没有任何检验能够优于随机猜测；而当 $q \ll n^{-1/2}$ 时，存在计算高效检验，能够以高概率正确猜测。

1 引言

植入子图问题是组合统计学中的一个基础问题，近年来受到越来越多的关注，这主要是因为它相对简单，却蕴含着丰富而有趣的复杂现象。粗略地说，植入子图问题要探究的是，能否检测或恢复嵌入大型随机图中的隐藏结构化子图。

更具体地说，令 $\mathbb{G}(n,\ p)$ 表示一个包含 n 个顶点、边概率为 p 的埃尔德什–雷尼随机图的分布；令 $\mathbb{H}$ 表示一个定义在含有 n 个顶点和 e_H 条边的图上的分布，用以表示植入的（或隐藏的）子图。考虑通过以下方式随机抽取一个含 n 个顶点的图：首先从 $\mathbb{G}(n,\ p)$ 中抽取一个图，然后抽取 $\mathbf{H} \sim \mathbb{H}$，再添加一个 $\mathbf{H}$ 的植入副本。令 $\mathbb{P}$ ($\mathbb{P}$，其中 p 表示“植入”) 表示由这一抽样过程诱导的图分布。

在检测问题中，观测到的图要么取自植入分布 $\mathbb{P}$，要么取自通常简单地视为 $\mathbb{G}(n,\ p)$ 的“零假设”分布。问题是要判断观测到的图是由这两个分布中的哪一个生成的。

在该问题的早期表述中（参考文献见下文第1.1 节），$\mathbb{H}$ 被选为所有大小为 k 的团上的均匀分布，其中 $k \ll n$。主要且尚未解决的

*麦吉尔大学数学与统计系，加拿大蒙特利尔´

†不列颠哥伦比亚大学数学系，加拿大温哥华

‡庞培法布拉大学经济与商业系，西班牙巴塞罗那；ICREA，Pg. Llu´ís Companys 23,08010 西班牙巴塞罗那；巴塞罗那经济研究生院§西北大学统计与数据科学系及计算机科学系

¶斯坦福大学统计系

问题是：当 $p = 1/2$ 时，对于哪些 k 值，存在能够以高概率识别底层分布的计算高效检验?

一个密切相关的问题是恢复问题：观测者观察到一个从植入分布 $\mathbb{P}$中抽取的图，其目标是（部分或完全地）恢复植入子图$\mathbf{H}$。

本文仅关注检测问题。具体而言，我们考虑植入一个完美匹配（假设 n 为偶数）或一棵生成树。换言之， $\mathbb{H}$ 要么是 n 个顶点的所有完美匹配上的均匀分布，要么是所有生成树上的均匀分布。在第一种情况下， $e_H = n/2$；而在第二种情况下， $e_H = n - 1$ 。由于两种情况下 e_H 都是线性的，即使 p 与零之间存在正的间隔，也可以仅通过对观测图中的边进行计数来检测是否存在植入对象。为避免通过边计数进行平凡检测，我们调整零假设 $\mathbb{Q}$ ，使得 $\mathbb{P}$ 和 $\mathbb{Q}$的期望边数相同。更准确地说，我们考虑如下假设检验问题：判断一个图是从上文定义的植入分布 $\mathbb{P}$ 中抽取，还是从零分布 $\mathbb{Q} = \mathbb{G}(n,\ q)$中抽取，其中 $q \overset{\text{定义}}{=} p + (1-p)e_H/\binom{n}{2}$。

我们考虑以下标准的假设检验问题表述。统计学家观察一个含有 n 个顶点的图 G。检验是一个函数 $f\colon \{0,1\}^{\binom{n}{2}} \to \{0,1\}$，其中$f(G) = 0$ 表示检验判定 G 来自零分布 $\mathbb{Q}$ ，而$f(G) = 1$ 表示检验判定 $\mathbb{P}$。检验 f 的风险定义为两类错误的概率之和：

$$R(f) = \Pr_{\mathbf{G}\sim\mathbb{Q}}[f(\mathbf{G}) = 1] + \Pr_{\mathbf{G}\sim\mathbb{P}}[f(\mathbf{G}) = 0] \ .$$

众所周知 ，使风险最小的最优检验是由 b 定义的似然比检验 y

$$f^*(G) = 0 \ \text{ if and only if } \ L(G) < 1 \ ,$$

其中

$$L(G) \overset{\text{def}}{=} \frac{d\mathbb{P}}{d\mathbb{Q}}(G)$$

是似然比。最优检验的风险为

$$R(f^*) = 1 - \frac{1}{2}\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}}|L(\mathbf{G}) - 1| = 1 - \mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) \ ,$$

其中 $\mathrm{d_{TV}}(\mathbb{P},\mathbb{Q})$ 表示全变差距离。因此，当且仅当 $\mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) \to 1$ 时，才能以高概率正确检测到植入结构是否存在。

本文的目标是确定 p （等价地， q）的取值，使得以高概率能够正确检测。

对 R（f^*）进行下界估计的一种标准工具，是利用如下简单观察：根据柯西–施瓦茨不等式，

$$R(f^*) \geq 1 - \frac{1}{2}\sqrt{\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}}(L(\mathbf{G})^2 - 1)} = 1 - \frac{1}{2}\sqrt{\chi^2(\mathbb{P},\mathbb{Q})} \ .$$

我们的主要结果表明，在所考虑的两个问题中， p 的高概率检测阈值都在 $n^{-1/2}$这一量级。更准确地说，在植入完美匹配问题和植入生成树问题中，若 $p = o(n^{-1/2}$，则 $R(f^*) \to 0$；若 $p = \omega(n^{-1/2}$，则 $R(f^*) \to 1$；详见下文定理3.1 和4.1 。

阈值的量级颇为引人注目。一方面，像完美匹配（或生成树）这样稀疏且分散的结构，只要隐藏其中的随机图平均度为 $o(\sqrt{n})$，就能被检测出来。另一方面，若不调整零分布的边概率，边计数对于任意 $p = o(1)$ 都能实现检测；但调整之后，当 p 的量级大于 $n^{-1/2}$时，没有任何检验能够检测出植入结构。调整 p 以隐藏边数竟会产生如此巨大的影响，这或许令人惊讶。

我们的不可能性证明依赖于对 χ^2 与 $\mathbb{P}$ 和 $\mathbb{Q}$ 之间距离的精细界定。对于可行性结果，我们设计了简单的检验，能够区分这两个分布。重要的是，这些检验在计算上是高效的，因此不存在统计/计算差距，这与许多相关的植入子图问题不同。

记号

我们使用标准的大 O 记号，并用 $o_n(1)$ 表示一个趋于 0 的函数，当$n \to \infty$。我们的约定是，随机变量以粗体表示。我们用 $H \subset G$ 表示 H 是 G 的子图。

1.1 相关工作

植入团问题 [Jer92，Kuˇc95] 催生了大量研究，包括关于检测算法的研究 [AKS98，FR10，DGGP11，DM15]；统计阈值方面的研究 [FGR+17，CLR17，RHS24，SW22]；以及复杂度下界/归约方面的研究 [BBH18，BB20，HS24，HS23，AAK+07]。

关于在边加权随机图中恢复植入式完美匹配，已有大量研究，其中植入边的权重分布不同于其他边的权重分布 [SSZ20、MMX21、DWXY23]；另请参见 [GSXY25]，，该文研究了稀疏 Erd k-因子″os–R´enyi 随机图中的植入 [EH25]。

当植入子图为树时，Massouli´e、Stephan 和 Towsley[MST19],研究了在稀疏 Erd″os–R´enyi 图中植入小型树的问题。

据我们所知，唯一一篇专门关注植入生成树问题的论文是 [MMX25]，，该文研究了如何恢复植入生成树。

近期工作 [MNWS+25] 研究了 Erd″os–R´enyi 设定下植入图检测中的全有或全无现象，并将其与 Kahn–Kalai 期望阈值联系起来。

植入子图问题还应用于秘密共享方案以及密码学中的其他问题 [BJRZ25，ABI+23]。

Erd″os-R´enyi 图中完美匹配和生成树（以及哈密顿圈）计数的分布 $\mathbb{G}(n,\ p)$ 已由 Janson 研究 [Jan94]。Janson 表明，只要 $p < 1$ 且 $p = \Omega(n^{-1/2}$，这些计数就服从对数正态分布。为了研究 $p \sim n^{-1/2}$ 的情形，Janson 研究了匹配计数的重加权的基展开，其中重加权因子随边数呈指数增长。尽管他没有对此进行讨论，这种重加权（正如我们将在第 2 节中看到的）等价于植入分布相对于 $\mathbb{G}(n,\ q')$ 的似然比，其中某个 q'（不一定对应于我们的边数校正零分布）。

在我们完成本手稿的撰写期间，我们得知了 Wee 和 Mao 最近开展的工作 [WM25]，；他们在一个相关但略有不同的设定中研究了植入匹配。在

他们的工作中，植入了一个大小为随机变量的匹配，其期望上界为 εn，其中 ε较小。在这一设定下，他们确定了检测阈值 $p = \Theta(n^{-1/2})$，并使用复杂的簇展开方法（沿用 [Jan94]）证明了阈值处的精确结果（尤其是似然比的对数正态性）。他们对匹配期望大小的限制，是簇展开得以收敛所必需的。

1.2 概述

本文其余部分的结构如下。在第2 节中，我们介绍似然比的一些性质，这些性质在证明不可能性结果时很有用。在第3 节中，我们证明关于植入完美匹配问题检测阈值的结果。在第4 节中，我们给出关于植入生成树的类似结果。

2 植入图的似然比

考虑这样一个实验：独立地从 $\mathbb{H}$ 中抽取 k 个图 $\mathbf{H}_1, \ldots, \mathbf{H}_k$，并令 $\text{collisions}(\mathbf{H}_1, \ldots, \mathbf{H}_k)$ 表示 $\mathbf{H}_1, \ldots, \mathbf{H}_k$ 共有的边数（计入重数）。本节将说明，通过研究 $\text{collisions}(\mathbf{H}_1, \ldots, \mathbf{H}_k)$ 的矩母函数，可以理解似然比 $\frac{d\mathbb{P}}{d\mathbb{Q}}$ 在 $\mathbb{Q}$ 下的 k 阶矩。

对于一个具有 n 个顶点且含有 e_G 条边的图 G，其似然比由以下表达式给出

$$L(G) = \frac{d\mathbb{P}}{d\mathbb{Q}}(G) = \frac{\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset G] \cdot p^{e_G - e_H}(1-p)^{\binom{n}{2} - e_G}}{q^{e_G}(1-q)^{\binom{n}{2} - e_G}}$$
$$= \left(\frac{1-p}{1-q}\right)^{\binom{n}{2}} \left(\frac{\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset G]}{\Pr_{\substack{\mathbf{H}\sim\mathbb{H} \\ \mathbf{G}\sim\mathbb{G}(n,p)}}[\mathbf{H} \subset \mathbf{G}]}\right) \left(\frac{p(1-q)}{q(1-p)}\right)^{e_G}. \tag{2.1}$$

写出 $\delta_n = e_H / \binom{n}{2}$ 后，通过简单的代数运算可得

$$p = \frac{q - \delta_n}{1 - \delta_n}, \qquad \frac{1-q}{1-p} = 1 - \delta_n, \qquad \frac{p(1-q)}{q(1-p)} = \frac{q - \delta_n}{q},$$

以及 上面的公式 (2.1) 可以改写为

$$L(G) = \left(\frac{1}{1-\delta_n}\right)^{\binom{n}{2}} \left(\frac{\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset G]}{\Pr_{\substack{\mathbf{H}\sim\mathbb{H} \\ \mathbf{G}\sim\mathbb{G}(n,p)}}[\mathbf{H} \subset \mathbf{G}]}\right) \left(1 - \frac{\delta_n}{q}\right)^{e_G}. \tag{2.2}$$

2.1 似然比矩

测度变换允许我们将似然比 $L(G)$ 在 $\mathbb{Q}$ 下的 k 阶矩设为与乘积测度 $\mathbb{Q}_k$ 下从图 $\mathbf{H}$ $\subset \mathbf{G}]$ 中抽取图 $\mathbf{G}$ 的 $\Pr_{\mathbf{H}\sim\mathbb{H}}[$ 矩成比例。特别地，承接表达式 式 (2.2)，定义 $r_k \overset{\text{定义}}{=} (1 - \delta_n/q)^k$，

$$\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^k = \left(\frac{1}{1-\delta_n}\right)^{k\binom{n}{2}} \frac{\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}}\left[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset \mathbf{G}]^k r_k^{e_\mathbf{G}}\right]}{\Pr_{\substack{\mathbf{H}\sim\mathbb{H} \\ \mathbf{G}\sim\mathbb{G}(n,p)}}[\mathbf{H} \subset \mathbf{G}]^k}$$
$$= \left(\frac{1}{1-\delta_n}\right)^{k\binom{n}{2}} \frac{Z_k \cdot \mathbf{E}_{\mathbf{G}\sim\mathbb{G}(n,q_k)}\left[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset \mathbf{G}]^k\right]}{\mathbf{E}_{\mathbf{G}\sim\mathbb{G}(n,p)}\left[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H} \subset \mathbf{G}]\right]^k}, \tag{2.3}$$

with

$$Z_k \stackrel{\text{def}}{=} (1-q+qr_k)^{\binom{n}{2}}, \qquad \text{and} \qquad q_k \stackrel{\text{def}}{=} \frac{qr_k}{1-q+qr_k}. \tag{2.4}$$

我们还可以进一步简化式 (2.3)的右侧。对于任意 $m \in \mathbb{N}$ 和 $a \in [0,1]$,

$$\mathbf{E}_{\mathbf{G}\sim\mathbb{G}(n,a)}\left[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H}\subset\mathbf{G}]^m\right] = \mathbf{E}_{\mathbf{H}_1,\dots,\mathbf{H}_m\sim\mathbb{H}}\left[a^{|E(\mathbf{H}_1\cup\cdots\cup\mathbf{H}_m)|}\right],$$

其中 $\mathbf{H}_1,\dots,\mathbf{H}_m \sim \mathbb{H}$ 相互独立，且 $E(\mathbf{H}_1\cup\cdots\cup\mathbf{H}_m)$ 是简单图 $\bigcup_{i=1}^m \mathbf{H}_i$的边集。在 公式 (2.3)的分子和分母中应用这一点,

$$\frac{\mathbf{E}_{\mathbf{G}\sim\mathbb{G}(n,q_k)}\left[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H}\subset\mathbf{G}]^k\right]}{\mathbf{E}_{\mathbf{G}\sim\mathbb{G}(n,p)}[\Pr_{\mathbf{H}\sim\mathbb{H}}[\mathbf{H}\subset\mathbf{G}]]^k} = \left(\frac{q_k}{p}\right)^{ke_H} \mathbf{E}_{\mathbf{H}_1,\dots,\mathbf{H}_k\sim\mathbb{H}}\left[q_k^{-\text{collisions}(\mathbf{H}_1,\dots,\mathbf{H}_k)}\right], \tag{2.5}$$

其中 $\text{collisions}(\mathbf{H}_1,\dots,\mathbf{H}_k) \stackrel{\text{定义为}}{=} ke_H - |E(\mathbf{H}_1\cup\cdots\cup\mathbf{H}_k)|$ 从多重集 $\bigcup_{i=1}^m E(\mathbf{H}_i)$ 中移除后得到一个集合所需移除的最少弧数。

结合公式 (2.3) 和 (2.5),

$$\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^k = \left(\frac{1}{1-\delta_n}\right)^{k\binom{n}{2}} \left(\frac{q_k}{p}\right)^{ke_H} Z_k \cdot \mathbf{E}_{\mathbf{H}_1,\dots,\mathbf{H}_k\sim\mathbb{H}}\left[q_k^{-\text{collisions}(\mathbf{H}_1,\dots,\mathbf{H}_k)}\right]. \tag{2.6}$$

3 植入完美匹配

本节讨论 $n \in \mathbb{N}$ 为偶数且 $\mathbb{P}$ 为 $\mathbb{G}(n,p)$、具有植入式完美匹配的情形，其中通过令 $\mathbb{H}$ 为完全图 K_n的所有完美匹配上的均匀分布来刻画这一情形。本节的主要结果如下。

定理 3.1。 若 $\mathbb{P}$ 是具有植入式完美匹配的 $\mathbb{G}(n,p)$，且 $\mathbb{Q}$ 是 $\mathbb{G}(n,q)$，其中 $q = p + \frac{1-p}{n-1}$，则

$$\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) = \begin{cases} 0 & when\ p = \omega(n^{-1/2}), \\ 1 & when\ p = o(n^{-1/2}). \end{cases}$$

我们首先考虑 $p = \omega(n^{-1/2})$ 的情形，并采用第 2 节介绍的框架，随后给出一个简单的假设检验，以证明 $p = o(n^{-1/2})$ 的情形。

引理 3.2。 若 $\mathbb{H}$ 是 K_n 上的完美匹配均匀分布，且 $p = \omega(n^{-1/2})$，则 $\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) = 0$。

证明。 只需证明 $\chi^2(\mathbb{P},\mathbb{Q}) = \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}}(L(\mathbf{G})^2 - 1) = o_n(1)$。我们先代入式 (2.6):

$$\mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^2 = \left(\frac{1}{1-\delta_n}\right)^{2\binom{n}{2}} \left(\frac{q_2}{p}\right)^{n} \cdot Z_2 \cdot \mathbf{E}_{\mathbf{H}_1,\mathbf{H}_2\sim\mathbb{H}}\left[q_2^{-\text{collisions}(\mathbf{H}_1,\mathbf{H}_2)}\right]. \tag{3.1}$$

接下来，我们估计碰撞的矩母函数$(\mathbf{H}_1,\mathbf{H}_2)$。

断言 3.3。 若 $sn \gg$ 对数 n 对数对数 n，则

$$\mathbf{E}_{\mathbf{H}_1,\mathbf{H}_2\sim\mathbb{H}}\left[s^{-\text{collisions}(\mathbf{H}_1,\mathbf{H}_2)}\right] \le (1+o_n(1))\,\mathbf{E}_{\mathbf{Y}\sim\mathrm{Pois}(\frac{1}{2})}\left[s^{-\mathbf{Y}}\right].$$

证明。 我们说明，随机变量 $\mathbf{X}$ = 碰撞$(\mathbf{H}_1,\mathbf{H}_2)$ 近似服从均值为$_1$ $_2$的泊松分布，因此上述表达式可由泊松矩母函数很好地近似。 K_n 的完美匹配集合与 n 的无序两两划分一一对应，其数量为 $(n-1)!!$。根据容斥原理，当$\mathbf{H}_1, \mathbf{H}_2$ 从该集合中独立且均匀随机选取时，对任意 $k \in [n/2]$,

$$\Pr[\mathbf{X}=k] = \frac{\binom{n/2}{k}\sum_{\ell=0}^{n/2}(-1)^\ell \cdot \binom{n/2-k}{\ell} \cdot (n-2k-2\ell-1)!!}{(n-1)!!}.$$

我们得到了该概率的两个上界。首先，对于任意 $k \in [n/2]$，我们可以从 $\ell = 0$ 项得到一个粗略的上界：

$$\Pr[\mathbf{X}=k] \le \binom{n/2}{k}\frac{(n-2k-1)!!}{(n-1)!!} = \frac{2^{n/2}}{k!\binom{n}{n/2}} \cdot \frac{1}{2^{n/2-k}}\binom{n-2k}{n/2-k} = O(\sqrt{n}) \cdot \frac{1}{2^k k!},$$

其中，最后一个不等式由以下事实得出： $\binom{2m}{m} = \frac{1}{\sqrt{\pi m}} 2^{2m}(1 + O(\frac{1}{m}))$.

现在，我们为 $k \ll n$ 得到一个更精细的上界。若 $k \ll n$，则令 k^+ 为不小于 k 的最小偶数，通过在 k^+ 项处截断容斥原理，我们得到该概率的一个上界：

$$\begin{aligned}
\Pr[\mathbf{X}=k] &\le \frac{\binom{n/2}{k}\sum_{\ell=0}^{k^+}(-1)^\ell \cdot \binom{n/2-k}{\ell} \cdot (n-2k-2\ell-1)!!}{(n-1)!!} \\
&= \frac{1}{k!}\frac{1}{\binom{n}{n/2}}\sum_{\ell=0}^{k^+}(-1)^\ell \frac{1}{\ell!} 2^{k+\ell}\binom{n-2k-2\ell}{n/2-k-\ell} \\
&= \frac{1}{k!}\frac{\sqrt{\pi n/2}}{2^n}\left(1+O(\tfrac{1}{n})\right) \cdot \sum_{\ell=0}^{k^+}(-1)^\ell \frac{1}{\ell!}\frac{2^{n-k-\ell}}{\sqrt{\pi(n/2-k-\ell)}}\left(1+O(\tfrac{1}{n})\right) \\
&= \frac{1}{2^k k!}\sum_{\ell=0}^{k^+}\left(-\frac{1}{2}\right)^\ell \frac{1}{\ell!}\left(1+O(\tfrac{k}{n})\right) \\
&\le \left(1+O(\tfrac{k}{n})\right) \cdot e^{-1/2}\frac{1}{2^k k!}.
\end{aligned}$$

令 $\mathbf{Y}$ ~服从 Pois($\frac{1}{2}$)，且 $\Pr[\mathbf{Y}= k] = e^{-1/2}\frac{1}{2^k k!}$。令 $m \ll n$ 为满足 $sm > 100\log n/\log\log n$ 的最小整数。我们对 $\mathbf{X}$ 的矩母函数进行变换：当 $k > m$ 时应用粗略上界，当 $k \le m$ 时应用更精细的上界，从而得到

$$\begin{aligned}
\sum_{k=0}^{\infty}\Pr[\mathbf{X}=k]\left(\frac{1}{s}\right)^k &\le (1+o_n(1))\sum_{k=0}^{m}\Pr[\mathbf{Y}=k]\left(\frac{1}{s}\right)^k + \sum_{k=m+1}^{n} O(\sqrt{n})\frac{1}{2^k k!}\left(\frac{1}{s}\right)^k \\
&\le (1+o_n(1))\sum_{k=0}^{m}\Pr[\mathbf{Y}=k]\left(\frac{1}{s}\right)^k + \sum_{k=m+1}^{n} O(\sqrt{nk})\left(\frac{e}{ks}\right)^k \\
&= (1+o_n(1))\mathbf{E}\left[s^{-\mathbf{Y}}\right] + o_n(1) = (1+o_n(1))\mathbf{E}\left[s^{-\mathbf{Y}}\right].\square
\end{aligned}$$

对于植入式完美匹配的情形，我们有 $\delta_n = \frac{1}{n-1}$，简单计算可知，若 $p = \omega(n^{-1/2})$，则 $q_2 = \omega(n^{-1/2})$。因此，定理 3.3 适用，其中 $s = q_2$。我们

接下来回顾一下，当 $\mathbf{Y}$ ∼服从 Pois(λ) 时，$\log(\mathbf{E}s^{-\mathbf{Y}}) = \lambda(\frac{1}{s}-1)$。将其代入 方程并取对数，可得 (3.1)

$$\log \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^2 \le o_n(1) - 2\binom{n}{2}\log(1-\delta_n) + n\log\left(\frac{q_2}{p}\right) + \log Z_2 + \frac{1}{2q_2} - \frac{1}{2}. \tag{3.2}$$

经过一些代数化简，我们得到

$$\begin{aligned} qr_2 &= (q-\delta_n)(1-\delta_n/q), \\ Z_2 &= (1-q+qr_2)^{\binom{n}{2}} = \left(1-2\delta_n+\tfrac{\delta_n^2}{q}\right)^{\binom{n}{2}}, \\ q_2 &= Z_2^{-1/\binom{n}{2}}\cdot qr_2, \\ \frac{q_2}{p} &= (1-\delta_n)\left(1-\tfrac{\delta_n}{q}\right)Z_2^{-1/\binom{n}{2}}. \end{aligned} \tag{3.3}$$

因此回到方程 (3.2),

$$\begin{aligned} \log \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^2 &\le o_n(1) + \left(n - 2\binom{n}{2}\right)\log(1-\delta_n) + n\log\left(1-\tfrac{\delta_n}{q}\right) \\ &\quad + \left(1-\tfrac{n}{\binom{n}{2}}\right)\log Z_2 + \frac{1}{2q_2} - \frac{1}{2} \\ &= o_n(1) - n(n-2)\log(1-\delta_n) + n\log(1-\tfrac{\delta_n}{q}) \\ &\quad + \tfrac{n(n-3)}{2}\log\left(1-2\delta_n+\tfrac{\delta_n^2}{q}\right) + \frac{1}{2q_2} - \frac{1}{2}. \end{aligned}$$

利用对数的泰勒展开，当 $x \to 0$ 时 $\log(1-x) = -x - \frac{x^2}{2} + o(x^2)$，并回顾 $\delta_n = \frac{1}{n-1}$ 以及 $q = \omega(n^{-1/2})$，我们有 $\log(1-\delta_n) = -\delta_n - \delta_n^2/2 + o(n^{-2})$，$\log(1-\delta_n/q) = -\delta_n/q + o(n^{-1})$，以及 $\log(1-2\delta_n+\delta_n^2/q) = -2\delta_n + \delta_n^2/q - 2\delta_n^2 + o(n^{-2})$。将这些表达式代入上面的式子，可见上面的式子等于

$$o_n(1) - n(n-2)(-\delta_n - \tfrac{1}{2}\delta_n^2) + n\left(-\tfrac{\delta_n}{q}\right) + \tfrac{n(n-3)}{2}\left(-2\delta_n + \tfrac{\delta_n^2}{q} - 2\delta_n^2\right) + \frac{1}{2q_2} - \frac{1}{2}.$$

这可化简为

$$o_n(1) + \frac{1}{2q_2} - \left(1+\Theta\left(\tfrac{1}{n}\right)\right)\frac{1}{2q}.$$

最后，利用 $q_2 = q(1+\Theta(1/(qn)))$ 以及 $q = \omega(n^{-1/2})$ 这一事实，我们得到整个表达式为 $o_n(1)$。因此 $\chi^2(\mathbb{P}, \mathbb{Q}) = \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}}(L(\mathbf{G})^2 - 1) = o_n(1)$，从而完成证明。 □

引理 3.4。 如果 $\mathbb{H}$ 是 K_n 与 $p = o$ 的完美匹配上的均匀分布，$n^{-1/2}$)那么 $\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P}, \mathbb{Q}) = 1$。

在证明引理之前，我们先从（更简单的）二部情形中给出一些直观解释。假设我们有一个定义在 $2n$ 个顶点上的二部随机图；那么，左侧顶点的度数彼此独立。在未植入情形下，每个左侧顶点的度数分布都接近$\mathbf{A} \sim \mathrm{Pois}(\lambda)$，当 $\lambda = nq$ 满足 $q, p \ll n^{-1/2}$ 时。在植入情形下，该分布则接近$\mathbf{B} \sim 1 + \mathrm{Pois}(\lambda - 1)$。$\mathbf{A}$ 和$\mathbf{B}$ 的均值相同，但方差分别为 λ 和 $\lambda - 1$。此外，$\mathbf{A}' = (\mathbf{A} - \mathbf{E}[\mathbf{A}])^2$ 和$\mathbf{B}' = (\mathbf{B} - \mathbf{E}[\mathbf{B}])^2$ 的标准差也分别为 Θ 和 λ。因此，由左侧顶点度数提供的 $n = \Omega(\lambda^2)$ 个独立同分布样本足以区分这两个分布。在非二部情形下，度数并不独立，但证明思路实际上相同。

定理 3.4 的证明。我们设计了一种假设检验，在 $p = o(n^{-1/2})$ 时能够以高概率区分 $\mathbb{P}$ 和 $\mathbb{Q}$（并注意到 $p = o(n^{-1/2})$ 蕴含 $q = o(n^{-1/2})$）。选取一个近似正则竞赛图 K_n，也就是说，将 K_n 的边划分为 P_1、……、P_n，使得 P_i 中的所有边都与顶点 i 相 incident，且每个部分包含 $\lceil (n-1)/2 \rceil$ 或 $\lfloor (n-1)/2 \rfloor$ 条边。

对于每个 $i \in [n]$，令 $\mathbf{X}_i = \sum_{e \in P_i} \mathbf{1}_{[e \in \mathbf{G}]}$，并令 $\mathbf{Y}_i = (\mathbf{X}_i - \mathbf{E}[\mathbf{X}_i])^2$（注意，在 $\mathbb{P}$ 和 $\mathbb{Q}$ 下，$\mathbf{E}\mathbf{X}_i$ 相同）。假设检验计算统计量 $\mathbf{Y} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i$，当且仅当其值大于 $\mathbf{E}_{\mathbb{Q}}\mathbf{Y} - \frac{1}{8}$ 时，才选择 $\mathbb{Q}$。我们论证，在 $\mathbb{P}$ 和 $\mathbb{Q}$ 下，均值之差至少为 $\frac{1}{4} - o_n(1)$，并且在 $\mathbb{P}$ 和 $\mathbb{Q}$ 下，$|\mathbf{Y} - \mathbf{E}\mathbf{Y}| = o_n(1)$ 均以高概率成立，由此即可得到结论。

在 $\mathbb{Q}$ 下，$\mathbf{Y}_i$ 相互独立，且每个 $\mathbf{X}_i$ 都服从 $\mathrm{Bin}(|P_i|, q)$。因此，每一项的均值为 $\mathbf{E}_{\mathbb{Q}}\mathbf{Y}_i = |P_i|q(1-q)$，而统计量的均值为 $\mathbf{E}_{\mathbb{Q}}[\mathbf{Y}] = \frac{n-1}{2} q(1-q)$。此外，由于该统计量是方差有界的独立随机变量之和，因此它会集中在均值附近：对于每个 $i \in [n]$，$\sqrt{\mathbf{E}_{\mathbb{Q}}(\mathbf{Y}_i - \mathbf{E}[\mathbf{Y}_i])^2} = O(nq+1)$。因此，应用标准的集中不等式可得，以高概率，当 $q, p \ll n^{-1/2}$ 时，$|\mathbf{Y} - \mathbf{E}_{\mathbb{Q}}\mathbf{Y}| \le O(nq+1)\sqrt{n} = o_n(1)$。

在 $\mathbb{P}$下，给定一个特定的植入匹配 H 后，$\mathbf{Y}_i$ 相互独立，其中满足 $P_i \cap E(H) = \emptyset$ 的$\mathbf{X}_i$ 的$n/2$ 服从 $\mathrm{Bin}(|P_i|, p)$ 分布，其余的服从 $1 + \mathrm{Bin}(|P_i| - 1, p)$ 分布。计算可得

$$\begin{aligned}
\mathbf{E}_{\mathbb{P}}[\mathbf{Y}_i \mid H] &= \mathbf{E}_{\mathbb{P}}\left(\sum_{e \in P_i \setminus H} (\mathbf{1}_{[e \in \mathbf{G}]} - p)\right)^2 + \left(\sum_{e \in P_i \cap E(H)} (\mathbf{1}_{[e \in \mathbf{G}]} - p) - |P_i|(q-p)\right)^2 \\
&= (|P_i \setminus E(H)|)p(1-p) + (|P_i \cap E(H)|(1-p) - (q-p)|P_i|)^2 \\
&= (|P_i| - \mathbf{1}_{[|E(H) \cap P_i| = 1]})p(1-p) + \tfrac{1}{4} + o_n(1),
\end{aligned}$$

这里用到了 $|P_i|(q-p) = 1/2 + o_n(1)$，因此

$$\mathbf{E}_{\mathbb{P}}\mathbf{Y} = \mathbf{E}_{\mathbb{H}}\mathbf{E}_{\mathbb{P}}[\mathbf{Y} \mid \mathbf{H}] = \frac{n-1}{2} p(1-p) - \tfrac{p(1-p)}{2} + \tfrac{1}{4}(1 + o_n(1)).$$

集中性同样源于这样一个事实：给定 H后，$\mathbf{Y}$ 是独立随机变量之和，且每个随机变量的方差都有界，$\sqrt{\mathbf{E}_{\mathbb{P}}(\mathbf{Y}_i - \mathbf{E}[\mathbf{Y}_i])^2} = O(np+1)$，因此以高概率有 $|\mathbf{Y} - \mathbf{E}_{\mathbb{P}}\mathbf{Y}| \le O(np+1)\sqrt{n} = o_n(1)$。

最后，我们声称均值之间相差 $\frac{1}{4} - o_n(1)$。再次回顾，$\delta_n = \frac{1}{n-1}$ 和$p = \frac{q - \delta_n}{1 - \delta_n}$，$1 - p = \frac{1-q}{1-\delta_n}$，

$$\begin{aligned}
\mathbf{E}_{\mathbb{P}}\mathbf{Y} - \mathbf{E}_{\mathbb{Q}}\mathbf{Y} &= \frac{(1-p)^2}{4}(1 + o_n(1)) - \frac{p(1-p)}{2} + \frac{n-1}{2}\left(p(1-p) - q(1-q)\right) \\
&= o_n(1) + \frac{1}{4} + \frac{n-1}{2}\left(\frac{q - \delta_n}{(1-\delta_n)}\frac{1-q}{1-\delta_n} - q(1-q)\right) \\
&= o_n(1) + \frac{1}{4} + \frac{n-1}{2}\left(-\delta_n + o(n^{-1})\right) \\
&\le -\frac{1}{4} + o_n(1). \square
\end{aligned}$$

4 Plantinga生成树

现在假设，在来自 $\mathbb{P}$ 的样本中，植入图$\mathbf{H}$ 是 K_n 的均匀随机生成树。本节将证明以下结果。

定理 4.1. 若 $\mathbb{P}$ 是带有植入生成树的 $\mathbb{G}(n,p)$，且 $\mathbb{Q}$ 是 $\mathbb{G}(n,q)$，其中 $q = p + {}^{2}{}_{(1-p)\,n}$，则

$$\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) = \begin{cases} 0 & when\ p = \omega(n^{-1/2}), \\ 1 & when\ p = o(n^{-1/2}). \end{cases}$$

为了给出全变差距离的上界，当 $p = \omega(n^{-1/2})$ 时，我们再次使用 第 2 节中建立的框架。为了控制碰撞的矩母函数，我们可以利用均匀随机生成树的边具有负关联这一事实。

事实 4.2。若 $J = (E,V)$ 是一个图，且 $\mathbf{H}$ 是 J 的均匀随机生成树，则随机变量 $\{\mathbf{1}_{[e\in\mathbf{H}]}\}_{e\in E}$ 具有负关联性。

对于子集 $E' \subset E$，令随机变量 $\mathbf{X}_{E'} = \sum_{e\in E'} \mathbf{1}_{[e\in\mathbf{H}]}$，并令 $\mathbf{Y}_{E'} = \sum_{e\in E'} \mathrm{Ber}(\Pr[e \in \mathbf{H}])$ 为一组独立伯努利随机变量之和，其边缘分布与$\mathbf{1}_{[e\in\mathbf{H}]}$相同。由于生成树的边具有负关联性，因此 [Sha00] 可知，对任意 $t \in \mathbb{R}$,

$$\mathbf{E}[\exp(t\mathbf{X}_{E'})] \le \mathbf{E}[\exp(t\mathbf{Y}_{E'})].$$

引理 4.3。若 $\mathbb{H}$ 是 K_n 的生成树上的均匀分布，且 $p = \omega(n^{-1/2})$ 则$\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P},\mathbb{Q}) = 0$。

证明。令 $\mathbf{H}_1$,$\mathbf{H}_2$ 为从 K_n 中均匀随机选取的两棵生成树。我们可以写成

$$\mathrm{collisions}(\mathbf{H}_1,\mathbf{H}_2) = \sum_{e\in\mathbf{H}_1} \mathbf{1}_{[e\in\mathbf{H}_2]}.$$

根据定理 4.2,

$$\mathbf{E}q_2^{-\mathrm{collisions}(\mathbf{H}_1,\mathbf{H}_2)} \le \mathbf{E}q_2^{-\mathbf{Y}} = \left(1 + \frac{2}{n}\left(\frac{1}{q_2} - 1\right)\right)^{n-1}$$

对于 $\mathbf{Y}$ ~二项分布(~~n~~ 1, ${}^{n-1}{}_{\binom{n}{2}}$)。将其代入公式 (2.6)，其中 $k = 2$ 并利用 $e_H = n-1$,

$$\begin{aligned} \log \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^2 \le\ & -n(n-1)\log(1-\delta_n) + 2(n-1)\log\left(\frac{q_2}{p}\right) + \log Z_2 \\ & + (n-1)\log\left(1 + \frac{2}{n}\left(\frac{1}{q_2} - 1\right)\right). \end{aligned}$$

再次应用公式 (3.3)中的简化，这进一步等于 o

$$\begin{aligned} =\ & (2(n-1) - n(n-1))\log(1-\delta_n) + 2(n-1)\log\left(1 - \tfrac{\delta_n}{q}\right) \\ & + \left(\binom{n}{2} - 2(n-1)\right)\log\left(1 - 2\delta_n + \tfrac{\delta_n^2}{q}\right) \\ & + (n-1)\log\left(1 + \frac{2}{n}\left(\frac{1}{q_2} - 1\right)\right). \end{aligned}$$

对对数进行泰勒展开，并利用 $\delta_n = \frac{2}{n}$、 p， $q = \omega(n^{-1/2})$，以及 $q_2 = q(1 + \Theta(1/(qn)))$，上面的表达式可进一步简化为

$$\begin{aligned}
&= o_n(1) - (n-2)(n-1)\left(-\delta_n - \tfrac{1}{2}\delta_n^2\right) + 2(n-1)\left(-\tfrac{\delta_n}{q}\right) \\
&\qquad + \frac{(n-4)(n-1)}{2}\left(-2\delta_n + \tfrac{\delta_n^2}{q} - 2\delta_n^2\right) \\
&\qquad + (n-1)\left(-\tfrac{2}{n} + \tfrac{2}{nq}\right) \\
&= o_n(1).
\end{aligned}$$

因此 $\chi^2(\mathbb{P},\ \mathbb{Q}) = \mathbf{E}_{\mathbf{G}\sim\mathbb{Q}} L(\mathbf{G})^2 - 1 = o_n(1)$，结论得证。 □

为了在 $p = o(n^{-1/2})$ 的情形下给全变差距离提供下界，我们使用一个与匹配所用检验几乎相同的假设检验。

引理 4.4。若 $\mathbb{H}$ 是 K_n 的生成树上的均匀分布，且 $p = o(n^{-1/2})$ 则$\lim_{n\to\infty} \mathrm{d_{TV}}(\mathbb{P}, \mathbb{Q}) = 1$。

证明。与 定理 3.4的证明类似，令 $P_1, \ldots, P_n$ 为 K_n 对应于正则或近正则竞赛图的边的一个划分。令 $\mathbf{X}_i = \sum_{e\in P_i} \mathbf{1}_{[e\in\mathbf{G}]}$，令 $\mathbf{Y}_i = (\mathbf{X}_i - \mathbf{E}\mathbf{X}_i)^2$，并令 $\mathbf{Y} = \frac{1}{n}\sum_{i=1}^n \mathbf{Y}_i$。与之前一样，我们的假设检验是计算 $\mathbf{Y}$，并且仅当 $\mathbb{P}$ 当且仅当 $\mathbf{Y} < \mathbf{E}_{\mathbb{Q}}\mathbf{Y} - \frac{1}{8}$ 时输出。

与定理 3.4的证明$\mathbf{E}_{\mathbb{Q}}[\mathbf{Y}] = \frac{n-1}{2}q(1-q)$，且以高概率$|\mathbf{Y} - \mathbf{E}_{\mathbb{Q}}\mathbf{Y}| = o_n(1)$，当 q时，$p \ll n^{-1/2}$。

在 $\mathbb{P}$ 下，为方便起见，我们记 $\mathbf{Y}_i = (\mathbf{Z}_i + \mathbf{T}_i)^2$,，其中 $\mathbf{Z}_i = \sum_{e\in P_i\setminus E(\mathbf{H})}(\mathbf{1}_{[e\in\mathbf{G}]} - p)$，且 $\mathbf{T}_i = |P_i \cap E(\mathbf{H})|(1-p) + |P_i|(p-q)$。给定 $\mathbf{H} = H$，$\mathbf{T}_i$ 均为固定常数，而每个 $\mathbf{Z}_i$ 相互独立且均值为零。因此

$$\begin{aligned}
\mathbf{E}_{\mathbb{P}}[\mathbf{Y}_i \mid H] &= \mathbf{E}_{\mathbb{P}}[\mathbf{Z}_i^2 \mid H] + T_i^2 \\
&= |P_i \setminus E(H)|p(1-p) + (|P_i \cap E(H)|(1-p) - |P_i|(q-p))^2 \\
&= |P_i \setminus E(H)|p(1-p) + (|P_i \cap E(H)|(1-p) - 1 + o_n(1))^2,
\end{aligned}$$

这里我们用到了 $|P_i|(q-p) = 1 + o_n(1)$，以及

$$\begin{aligned}
\mathbf{Var}_{\mathbb{P}}[\mathbf{Y}_i \mid H] &\le 2\mathbf{Var}_{\mathbb{P}}[\mathbf{Z}_i^2 \mid H] + 2\mathbf{Var}_{\mathbb{P}}[2\mathbf{T}_i\mathbf{Z}_i \mid H] \\
&= O(np+1)^2 + O(np+1)\cdot(|P_i \cap E(H)|(1-p) - 1 + o_n(1))^2,
\end{aligned}$$

最后，$\mathrm{Cov}_{\mathbb{P}}(\mathbf{Y}_i,\ \mathbf{Y}_j \mid H) = 0$ 对于 $i \ne j$。由此可得，给定 H，

$$\begin{aligned}
\mathbf{E}_{\mathbb{P}}[\mathbf{Y} \mid H] &= \frac{(n-2)(n-1)}{2n}p(1-p) + \mathbf{E}_{i\sim\mathrm{Unif}([n])}(|P_i \cap E(H)|(1-p) - 1 + o_n(1))^2 \\
\mathbf{Var}_{\mathbb{P}}(\mathbf{Y} \mid H), &= \frac{O(np+1)^2}{n} + \frac{O(np+1)}{n}\mathbf{E}_{i\sim\mathrm{Unif}([n])}(|P_i \cap E(H)|(1-p) - 1 + o_n(1))^2).
\end{aligned}$$

现在我们调用 $\mathbb{H}$ 的度分布性质。使用从 $\mathbb{H}$ 中抽样的 ¨ufer 编码方法， H 中顶点 i, j 的度 $\mathbf{D}_i$,$\mathbf{D}_j$ 服从平移多项分布，因此（$\mathbf{D}_i - 1$,$\mathbf{D}_j - 1$）可通过进行 $n-2$ 次试验来抽样，每次试验以概率 $1/n$ 分别使 $\mathbf{D}_i - 1, \mathbf{D}_j - 1$ 增加。因此，$\mathbf{D}_i$,$\mathbf{D}_j$ 的任意 $O(1)$ 阶联合矩，除相差 ($1 \pm o_n(1)$) 因子外，均等于独立 Pois(1) 随机变量的矩。令

$\mathbf{B}_i = |P_i \cap E(\mathbf{H})|$；由对称性， $E(H)$ 中与 i 相接的每条边，被包含在 P_i 中的概率都相同；而 $\mathbf{B}_i$, $\mathbf{B}j$ 的任意 $O(1)$ 阶联合矩，除相差 $(1 \pm o_n(1))$ 因子外，均等于独立 $\mathrm{Pois}(\frac{1}{2})$ 随机变量的联合矩。因此

$$\mathbf{E}_{\mathbb{H}}(|P_i \cap E(\mathbf{H})|(1-p) - 1 + o_n(1))^2 = \frac{3}{4} + o_n(1),$$

并使用上文给出的条件期望公式，

$$\mathbf{E}_{\mathbb{P}}\mathbf{Y} = \mathbf{E}_{\mathbb{H}}\mathbf{E}_{\mathbb{P}}[\mathbf{Y} \mid \mathbf{H}] = \frac{n}{2}p(1-p) + \tfrac{3}{4} + o_n(1).$$

我们利用全方差公式来界定$\mathbf{Y}$ 的方差：

$$\mathbf{Var}_{\mathbb{P}}(\mathbf{Y}) = \mathbf{E}_{\mathbb{H}}\mathbf{Var}_{\mathbb{P}}(\mathbf{Y} \mid \mathbf{H}) + \mathbf{Var}_{\mathbb{H}}\mathbf{E}_{\mathbb{P}}[\mathbf{Y} \mid \mathbf{H}].$$

上文已计算出条件方差，并利用$\mathbf{E}[\mathbf{B}_i^2] = O(1)$，当 $p \ll n^{-1/2}$ 时，有$\mathbf{E}_{\mathbb{H}}\mathbf{Var}_{\mathbb{P}}(\mathbf{Y} \mid \mathbf{H}) = o_n(1)$。最后，我们对条件期望的方差作上界估计。利用$\mathbf{B}_i$,$\mathbf{B}j$ 渐近独立且服从泊松分布这一点，

$$\begin{aligned}\mathbf{Var}_{\mathbb{H}}\mathbf{E}_{\mathbb{P}}[\mathbf{Y} \mid \mathbf{H}] &= \mathbf{E}_{i,j\sim[n]}\mathbf{E}_{\mathbb{H}}((\mathbf{B}_i(1-p)-1)^2 - \frac{3}{4} + o_n(1))((\mathbf{B}_j(1-p)-1)^2 - \frac{3}{4} + o_n(1)) \\ &= o_n(1).\end{aligned}$$

因此，根据切比雪夫不等式，在 $\mathbb{P}$ 下以高概率有 $|\mathbf{Y} - \mathbf{E}[\mathbf{Y}]| = o_n(1)$。最后，我们比较 $\mathbf{Y}$ 在 $\mathbb{P}$ 和 $\mathbb{Q}$下的均值，可得

$$\begin{aligned}\mathbf{E}_{\mathbb{P}}\mathbf{Y} - \mathbf{E}_{\mathbb{Q}}\mathbf{Y} &= \frac{n}{2}(p(1-p) - q(1-q)) + \frac{3}{4} + o_n(1) \\ &= \frac{n}{2}\left(\frac{q-\delta_n}{1-\delta_n}\left(1 - \frac{q-\delta_n}{1-\delta_n}\right) - q(1-q)\right) + \frac{3}{4} + o_n(1) \\ &= \frac{n}{2}\left(-\frac{2}{n} + o(\frac{1}{n})\right) + \frac{3}{4} + o_n(1) = -\frac{1}{4} + o_n(1).\end{aligned}$$

因此，在 $\mathbb{P}$ 和 $\mathbb{Q}$ 下，该假设检验均以高概率成功，由此即可得出结论。

□

5 致谢

本材料部分基于美国国家科学基金会（NSF）通过资助编号 DMS-1928930 提供的支持。作者在 2025年春季驻留于加利福尼亚州伯克利的西蒙斯·劳弗数学科学研究所（SLMath/MSRI）期间，参加了离散结构的概率与统计项目。

参考文献

[AAK+07] Noga Alon、Alexandr Andoni、Tali Kaufman、Kevin Matulef、Ronitt Rubinfeld 和 Ning Xie。《Testing k-wise and almost k-wise independence》。载于*STOC'07—第39届 ACM* 计算理论年会论文集，第496–505页。ACM，纽约，2007年。

[ABI+23] Damiano Abram、Amos Beimel、Yuval Ishai、Eyal Kushilevitz 和 Varun Narayanan。《基于植入图的密码学：对数规模消息下的安全性》。载于《密码学理论》第一部分，《计算机科学讲义》卷14369，第286–315页。Springer，沙姆， [2023] ©2023。[AKS98] Noga Alon、Michael Krivelevich 和 Benny Sudakov。《在随机图中寻找大型隐藏团》。载于第八届“随机结构与算法”国际会议论文集（波兹南，1997年），第13卷，第457–466页，1998年。[BB20] Matthew Brennan 和 Guy Bresler。《秘密泄露导致的可约性与统计—计算差距》。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。[BJRZ25] Andrej Bogdanov、Chris Jones、Alon Rosen 和 Ilias Zadik。《植入随机子图问题的低次安全性》。载于《密码学理论》第二部分，《计算机科学讲义》卷15365，第255–275页。Springer，沙姆， [2025]©2025。[CLR17] T. Tony Cai、Tengyuan Liang 和 Alexander Rakhlin。《大型含噪矩阵中子矩阵定位的计算与统计边界》。《统计年刊》，45(4):1403–1430，2017年。[DGGP11] Y. Dekel、O. Gurel-Gurevich 和 Y. Peres。《以高概率在线性时间内寻找隐藏团》。载于解析算法与组合学会议论文集，第67–75页。工业与应用数学学会，2011年。[DM15] Y. Deshpande 和 A. Montanari。《近线性时间内寻找大小为 $\sqrt{N/e}$的隐藏团》。《计算数学基础》，15(4):1069–1128，2015年。[DWXY23] Jian Ding、Yihong Wu、Jiaming Xu 和 Dana Yang。《植入匹配问题：尖锐阈值与无穷阶相变》。《概率论及相关领域》，187(1-2):1–71，2023年。[EH25] Dor Elimelech 和 Wasim Huleihel。《在随机图中检测任意植入子图》。arXiv预印本 arXiv:2503.19069，2025年。[FGR+17] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S. Vempala 和 Ying Xiao。《统计算法与检测植入团的下界》。《ACM期刊》，64(2):文章8，37页，2017年。[FR10] Uriel Feige 和 Dorit Ron。《在线性时间内寻找隐藏团》。载于第21届概率、组合与渐近方法在算法分析中的国际会议（AofA’10），《离散数学、理论计算机科学会议录》卷AM，第189–203页。离散数学与理论计算机科学协会，南锡，2010年。[GSXY25] Julia Gaudio、Colin Sandon、Jiaming Xu 和 Dana Yang。《植入 k-因子恢复中的“全有—部分有—全无”相变》。载于Nika Haghtalab和Ankur Moitra编，《第三十八届学习理论会议论文集》，《机器学习研究会议录》卷291，第2457–2459页。PMLR，2025年6月30日至7月4日。[HS23] Shuichi Hirahara 和 Nobutaka Shimizu。《困难性自增强：简化、优化与统一》。载于STOC’23—第55届ACM计算理论年度研讨会论文集，第70–83页。ACM，纽约， [2023] ©2023。

[HS24] Shuichi Hirahara 和 Nobutaka Shimizu。植入团猜想彼此等价。载于 STOC'24——第56届 ACM计算理论研讨会论文集，第358–366页。ACM，纽约，[2024] ©2024。[Jan94] Svante Janson。随机图中的生成树、哈密顿回路和完美匹配的数量。《组合学、概率与计算》，3(1):97–126，1994。[Jer92] Mark Jerrum。大型团难以被 Metropolis过程发现。《随机结构与算法》，3(4):347–359，1992。[Kuˇc95] Ludˇek Kuˇcera。图划分问题的期望复杂度。第57卷，第193–212页。1995年。组合优化1992（CO92）（牛津）。[MMX21] Mehrdad Moharrami、Cristopher Moore 和 Jiaming Xu。植入匹配问题：相变与精确结果。《应用概率年刊》，31(6):2663–2720，2021。[MMX25] Mehrdad Moharrami、Cristopher Moore 和 Jiaming Xu。植入生成树问题。arXiv预印本 arXiv:2502.08790，2025年。[MNWS+25] Elchanan Mossel、Jonathan Niles-Weed、Youngtak Sohn、Nike Sun 和 Ilias Zadik。植入子图推断中的尖锐阈值。《应用概率年刊》，35(1):523–563，2025。[MST19] Laurent Massouli´e、Ludovic Stephan 和 Don Towsley。在图中植入树并将其找回。载于《学习理论会议》，第2341–2371页。PMLR，2019年。[RHS24] Asaf Rotenberg、Wasim Huleihel 和 Ofer Shayevitz。植入二分图检测。《IEEE信息论汇刊》，70(6):4319–4334，2024。[Sha00] Qi-Man Shao。关于负相关随机变量与独立随机变量之间矩不等式的比较定理。《理论概率杂志》，13(2):343–356，2000。[SSZ20] Guilhem Semerjian、Gabriele Sicuro 和 Lenka Zdeborov´a。稀疏植入匹配问题中的恢复阈值。《物理评论E》，102(2):022304，18，2020。[SW22] Tselil Schramm 和 Alexander S. Wein。低次多项式估计中的计算障碍。《统计年刊》，50(3):1833–1858，2022。[WM25] Timothy LH Wee 和 Cheng Mao。对数似然比的簇展开：植入匹配的最优检测。arXiv预印本 arXiv:2512.14567，2025年。