arXiv:2302.03658v2 [cs.DS] 6 Mar 2024

植入二分图检测

Asaf Rotenberg Wasim Huleihel Ofer Shayevitz 2024年3月7日

摘要

我们研究在给定的随机图中检测隐藏二分子图的问题。我们将其表述为一个假设检验问题：在零假设下，该图是一个定义在n 个顶点上的 Erd˝os-R´enyi 随机图的一个实现，其边密度为q。在备择假设下，存在一个边密度为$p > q$的植入$k_{\mathsf{R}} \times k_{\mathsf{L}}$ 二分图子图。我们刻画了这一问题的统计障碍和计算障碍。具体而言，我们推导出信息论下界，并设计和分析了与这些下界相匹配的最优算法，分别适用于稠密区域（其中p、$q = \Theta(1)$）和稀疏区域（其中p、$q = \Theta(n^{-\alpha})$，$\alpha \in (0,2]$）。我们还研究了多项式时间检验问题。与类似的结构化高维问题中常见的情形一样，我们的模型会经历“容易—困难—不可能”的相变，而计算约束会损害统计性能。为证明这种统计与计算之间的差距，我们基于低次猜想证明了计算下界，并表明在猜想认为困难的区域内，低次多项式算法无法奏效。

1 引言

在许多领域中，经常会遇到检测大型图中隐藏结构的问题，例如社交网络分析、计算生物学和计算机视觉等；文献中也已在许多有趣的情形下对此进行了深入研究。如今显而易见的是，这类结构的推断，尤其是在高维区间，会带来诸多统计和计算方面的挑战 [1]。在这项工作中，我们研究一个构成其中一些挑战基础的假设检验问题。为了从数学角度说明这一问题，我们先简要回顾一些经典结果。为此，考虑以下对隐藏子图检测问题的较一般表述

A. Rotenberg、W. Huleihel 和 O. Shayevitz 任职于特拉维夫大学电气工程与系统系，地址为以色列特拉维夫 6997801（电子邮箱：asaf.rotenberg@gmail.com、wasimh@tauex.tau.ac.il、ofersha@tauex.tau.ac.il）。
A. Rotenberg 的研究工作得到以色列科学基金会资助（项目编号：1734/21 和 1766/22）。W. Huleihel 的研究工作得到以色列科学基金会资助（项目编号：1734/21）。O. Shayevitz 的研究工作得到以色列科学基金会资助（项目编号：1766/22）。本文部分内容曾发表于 2023 年在台湾台北举行的 IEEE国际信息论研讨会。

问题。给定一个含有 n 个顶点的图，假设我们的目标是推断图中是否存在一个隐藏的植入结构，还是该图纯粹是随机生成的，例如一张 Erd˝os-R´enyi 随机图的样本。全文中，我们将后者记为 $\mathcal{G}(n, q)$，即一张含有 n 个顶点、边概率为 q 的随机图。如果植入结构是一个由 $k < n$ 个顶点构成的完全连通子集，也就是一个团，那么上述问题就是众所周知的植入团（PC）问题（例如，[2], [3], [4], [5], [6], [7]）。如果植入结构是一张 Erd˝os-R´enyi $\mathcal{G}(k, p)$ 随机图的样本，那么我们得到广为人知的植入稠密子图（PDS）问题（例如，[8], [9], [10]）。这些问题在文献中广受关注，（还）因为对于某些参数区域 (n, k, p, q)，尽管从统计上来说检测问题可以解决，但在计算上似乎很困难。

在本文中，我们关注这样的情形：$\mathcal{G}(n, q)$ 中的植入结构是一个边密度为 $p > q$ 的 $k_R \times k_L$二分图子图。当 $p = 1$ 时，植入结构表现为一个含有 $k_R + k_L$ 个顶点和 $k_R k_L$ 条边的完全二分图。当 $p < 1$ 时，$k_R k_L$ 条可能的边各自以概率 p 存在，且彼此独立。与前文提到的完全子图和稠密子图结构相比，二分图的“几何形态”更为丰富，具体体现在以下方面。我们的结果表明，植入二部图模型在两个极端之间进行插值：一方面是 PC 和 PDS 问题，另一方面则（本质上）是植入星问题。[1] 最重要的是，正如下文所述，从统计和计算角度审视时，这两个极端的行为有所不同。具体而言，在 $k_R = \Theta(k_L)$ 这一特殊情形下，也就是说植入结构是平衡二部图时，我们将证明，其统计和计算障碍与著名的 PC 和 PDS 问题基本相同。然而，若 $k_R \wedge k_L = O(\log n)$，则我们会得到与前一种情形完全不同的障碍；最值得注意的是，不存在困难区域！这一点也在 [11], 针对星形结构以及边概率 $p = 1$和 $q = \Theta(n^{-1})$ 的特殊情形中被观察到。

更一般地，我们猜想，对于一般的植入结构，其检测边界强烈依赖于以下量。[2]

定义 1（最大子图密度）。给定子图 $\Gamma = (V(\Gamma), E(\Gamma))$，其顶点集为 $V(\Gamma)$，边集为 $E(\Gamma)$。Γ 的最大子图密度定义为：

$$m(\Gamma) \triangleq \max_{\emptyset \neq H \subseteq \Gamma} \frac{|E(H)|}{|V(H)|}. \tag{1}$$

注意到 $\Omega(1) \leq m(\Gamma) \leq O(|V(\Gamma)|)$。上述猜想是合理的，因为众所周知，[12, 13] 最大子图密度在刻画 Erd˝os-R´enyi 随机图中的子图计数方面发挥着重要作用，而这又与检测是否可能的问题密切相关。现在，接着上述讨论，请注意，对于 PC 和 PDS 问题，其中

[1]回想一下，含k 个顶点的星图是一棵树，其中有一个内部节点和$k-1$ 个叶节点。[2]事实上，我们的部分分析可以推广到一般的植入结构，并且可以证明，最大子图密度确实发挥着作用。

	稠密区域	稀疏区域
不可能	$k_{\mathsf{R}} \wedge k_{\mathsf{L}} = O(\text{对数 } n), k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2 \ll n$	$\chi^2(p\|\|q) \ll \frac{1}{k_{\mathsf{R}} \wedge k_{\mathsf{L}}} \wedge \frac{n^2}{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2} \wedge \frac{n}{k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2}$
Hard	$k_{\mathsf{R}} \wedge k_{\mathsf{L}} = \Omega(\text{对数 } n), k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2 \ll n$	$\frac{1}{k_{\mathsf{R}} \wedge k_{\mathsf{L}}} \ll \chi^2(p\|\|q) \ll n^2 \frac{k_{\mathsf{R}}^2}{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2} \wedge \frac{n k_{\mathsf{R}}^2}{k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2}$
Easy	$k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2 \gg n$	$\chi^2(p\|\|q) \gg n^2 \frac{k_{\mathsf{R}}^2}{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2} \wedge \frac{n k_{\mathsf{R}}^2}{k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2}$

表 1： 二部子图检测阈值。

Γ 是一个具有 k 个顶点的完全图，[3] 显然，最大子图密度为$k_2^{-1} = \Theta(k)$，也就是说，它与该结构中的顶点数成正比。在我们的问题中，其中 Γ 是一个完全二分图，我们得到：

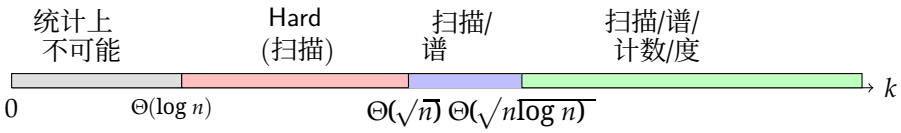
$$m(\mathsf{K}_{\mathcal{R},\mathcal{L}}) = \frac{k_{\mathsf{R}} k_{\mathsf{L}}}{k_{\mathsf{R}} + k_{\mathsf{L}}}. \tag{2}$$

因此，在平衡情形下，我们得到$m(\mathrm{K}_{\mathcal{R},\mathcal{L}}) = \Theta(|v(\mathrm{K}_{\mathcal{R},\mathcal{L}})|)$，也就是说，它与结构中的顶点数成正比，这与 PC 和 PDS 问题中的情况相同。当$k_{\mathsf{R}} \wedge k\mathrm{O}(\log n)$时，我们得到$m(\mathrm{K}_{\mathcal{R},\mathcal{L}}) = O(\log n)$。本文分析任意$k_{\mathsf{R}}$ 和k_{L}取值下的二分子图检测问题，从而涵盖最大子图密度各种可能的缩放情况；据我们所知，我们的工作首次研究了植入结构几何形状所包含的最大子图密度取遍所有可能取值的检测问题。

主要贡献。对于上述检测问题（精确定义见问题 1），我们刻画了检测的统计极限和计算极限。为此，我们推导出适用于所有算法的信息论下界形式的不可能性结果，给出了强检测不可能实现的条件。随后，我们提出若干算法，分析其性能，并证明我们的界是紧的（相差至多多对数因子）。与其他相关的高维问题类似，我们发现所推导的统计极限与所构造的高效算法的性能之间存在差距。我们猜想，在计算障碍之下不存在多项式时间算法；换言之，检测任务在平均情形下是计算困难的。为佐证这一现象，我们沿着近期一系列工作[14–19], 的思路，证明低次多项式类无法在这一推测为困难的区域内解决该检测问题。表 1 给出了我们界的一个较宽松但更简单的版本。这里回顾一下，$\chi^2(\mathrm{p}||\mathrm{q}) = (\mathrm{p}-\mathrm{q})^2\, \mathrm{q}(1-\mathrm{q})$ 是参数为 p 和 q 的两个伯努利随机变量之间的 χ^2-散度，$\mathrm{a} \vee \mathrm{b} \triangleq \max\{\mathrm{a}, \mathrm{b}\}$ 且 $\mathrm{a} \wedge \mathrm{b} \triangleq \min\{\mathrm{a}, \mathrm{b}\}$；符号“≪”隐藏了关于 n 的次多项式因子（精确定义见下文的符号说明段落）。

为进一步说明我们的结果，下面将重点讨论完全二部图这一特殊情形，其中 $p = 1$。从算法视角来看，表 1

[3]在 PC 和 PDS 中，我们都会从 $[n]$ 中均匀随机地选取k 个顶点进行“植入”；这等价于在k 个顶点上植入一个完全图，因此，无论在 PDS 中我们以p 的概率保留 Γ 中的边这一事实如何，Γ 确实都是一个大小为k 的团。



Fig图 1： 检测大小为 k 的植入团的相图，作为 k 的函数 .

中的检测阈值是通过以下三种程序的组合得到的：

- 扫描检验：扫描观测图中的最稠密二分图子图，并将其大小与某个阈值进行比较。
- 计数检验：计算观测图中的边数，并将该计数与某个阈值进行比较。
- 度检验：在观测图中寻找最大的顶点度，并将其与某个阈值进行比较。

上述过程的精确定义见第3节。需要强调的是，上述算法绝非新提出的；例如，[8,10]中就为 PC 和 PDS 问题提出了完全相同的算法。现在，对于 PC 问题，我们得到图1所示的相图。请注意，扫描检验的效果始终不逊于度检验和计数检验。此外，计数检验和度检验的性能相同。最后，存在一个区域，在该区域内，求解 PC 问题在计算上（猜想而言）是困难的。对于完全二分图情形，我们得到图2所示的相图，其中坐标轴分别对应右侧和左侧植入集合中的顶点数。与 PC 情形相比，可以看出，得到的相图复杂得多。例如，注意到存在这样一个参数区域，即 $k_R \wedge k_R = O(\log n)$，在此区域内，检测问题要么在统计上不可能，要么可以通过度检验在多项式时间内求解。 [11], 也针对星形结构这一特殊情形观察到了这一点，其中 $k_R \wedge k_R = 1$，且边概率 $p = 1$ 和 $q = \Theta(n^{-1})$。

相关工作。 过去几十年来，人们从许多不同的角度研究了PC检测问题。已有许多多项式时间算法被提出，包括简单的边计数、度检验[3]，（当 $k = \Omega\left(\sqrt{n \log n}\right)$ 时可实现检测），以及更复杂的谱方法，例如 [4]（当 $k = \Omega\left(\sqrt{n}\right)$ 时可成功）。其他子图检测问题中的统计极限以及推断算法的分析也得到了研究，例如社区检测 [20], [21]。最初在 [2] 中有人猜想，当 $k = o\left(\sqrt{n}\right)$ 时，PC问题无法在多项式时间内求解；而当 $k = \Omega(\log n)$ 时，穷举扫描则可以成功。这个PC猜想已通过归约论证应用于其他高维问题，例如稀疏主成分分析 [22] 和PDS [8], ；后者观察到了不可能—困难—容易的相变现象。关于近期综述，请读者参阅[23,24], 及其中的许多参考文献。

上述大多数高维检测问题通常涉及一种植入结构，其元素要么完全连接（PC），要么以某种概率连接（例如 PDS）。因此，现有用于确定统计和

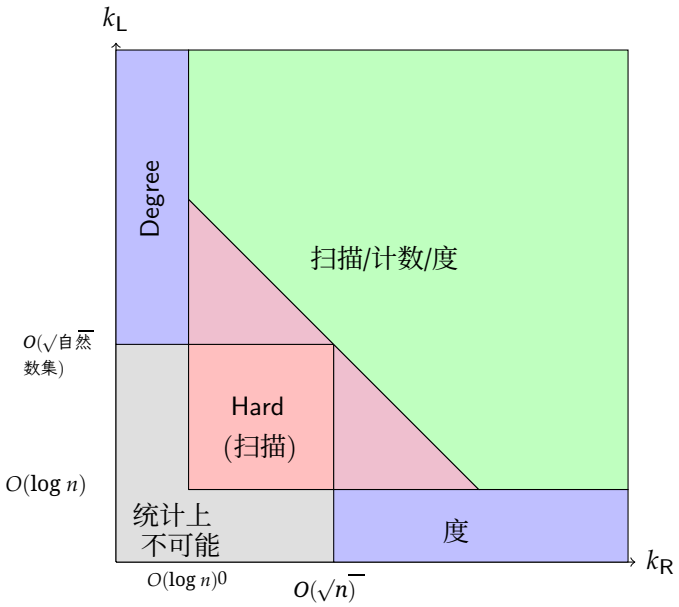


图2：相图：用于检测植入的$k_R \times k_L$ 完全二部图。

计算极限的技术专门针对这些类型的结构。不过也有一些例外。例如，[25] 研究了叠加在稀疏Erd˝os-R´enyi 随机图中的 D 叉树的检测与恢复任务，其中$p = 1$ 和$q = \Theta(n^{-1})$。如上所述，[25] 的一个特殊情形是星形子图，我们的模型也涵盖了这一情形。另一个例外是[26], ，其中讨论了植入的哈密顿圈恢复问题。近期，[19] 研究了一般植入诱导子图的检测与恢复问题；然而，由于潜在结构表现为诱导子图，[19] 中的技术和结果与我们的设定无关，因为在我们的设定中，取的是二分图与底层 Erd˝os-R´enyi 随机图的并。尽管我们的二分图框架仍有一定特殊性，但如上所述，它渐近地涵盖了可能的最大子图密度的全部范围，而且其分析在技术上颇具意义。此外，二分图在理论和实践中都有许多应用（例如，见[27], [28]），这进一步说明了本文的研究动机。

过去几年中，人们成功地严格界定了高效算法能够和不能够实现的事情。近期的研究成果，例如 [14–18,29–60] 揭示了一种引人入胜的现象，这种现象在许多具有植入结构的高维问题中普遍存在：所有计算高效算法所需的数据量，与统计最优算法所需的数据量之间存在内在差距。针对这一现象，即平均情形意义下的困难性，人们提出了各种形式的严格证据，大致可分为两类：1）特定计算模型下的失效，即证明强大的计算高效算法类别（例如低次多项式 [14–18], 平方和层级 [42–50], 统计查询算法 [51–54],消息传递算法 [55–60], 等）在该问题被猜想为计算困难的区域内会失效。2）平均情形归约，从另一个问题（例如

植入团问题，通常被猜想为计算困难，例如 [29–31,33–41]。如上所述，本文利用低次多项式的方法，为我们模型中所猜想的统计-计算差距提供证据。

记号。对于简单无向图 G，记其邻接矩阵为 A。此外，分别用 V (G) 和 E (G) 表示 G 的顶点集和边集。对于任意有限集合 $\mathcal{S}$，用 $|\mathcal{S}|$ 表示其大小。令 m(G) 表示 G 的最大子图密度，定义见定义 1。此外，令 Δ(G) 表示 G 的最大顶点度，即 $\Delta(G) \triangleq \max_i \deg_i$，其中 $\deg_i$ 表示第 i^{th} 个顶点的度，也就是以顶点 i 为一个端点的边数。令 $N^{K_{\mathcal{R},\mathcal{L}}}$，$\binom{n}{k_R}\binom{n-k_R}{k_L}$ 表示在由 [n] 个顶点构成的完全图中，所有可能的 $k_R \times k_L$ 二分图副本的数量；我们将这些副本记为 $K^{(1)}_{\mathcal{R}},\mathcal{L}, K^{(2)}_{\mathcal{R}},\mathcal{L}, \ldots, K_{(N_{K_{\mathcal{R},\mathcal{L}}})\mathcal{R},\mathcal{L}}$，并用 $\mathcal{K}_{n,k_R,k_L}$ 表示所有这些副本构成的集合。对于任意 $n \in \mathbb{N}$，令 [n] 表示整数集合 $\{1, 2, \ldots, n\}$。对于概率测度 $\mathbb{P}$ 和 $\mathbb{Q}$，令 $d_{TV}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2}\int |d\mathbb{P} - d\mathbb{Q}|$ 和 $\chi^2(\mathbb{P}||\mathbb{Q}) = \int (d\mathbb{P}-d\mathbb{Q})^2 / d\mathbb{Q}$ 分别表示全变差距离和 χ^2-散度。令 Bern(p) 和 Binomial(n, p) 分别表示参数为 p 和 n 的 Bernoulli 分布和二项分布。我们采用标准渐近记号：对于两个正数列 $\{a_n\}$ 和 $\{b_n\}$，若存在某个绝对常数 C，使得对所有 n 都有 $a_n = O(b_n)$，则记作 $a_n \leq O(b_n)$；若 $b_n = O(a_n)$，则记作 $a_n = \Omega(b_n)$；若 $a_n = O(b_n)$ 且 $a_n = \Omega(b_n)$，则记作 $a_n = \Theta(b_n)$；若 $a_n/b_n \to 0$，当 $n \to \infty$ 时，则记作 $a_n = o(b_n)$ 或 $b_n = \omega(a_n)$。记号 $\ll$ 表示关于 n 多项式级地小，即若 $\liminf_{n\to\infty} \log_n a_n < \liminf_{n\to\infty} \log_n b_n$，则 $a_n \ll b_n$；例如，$n \ll n^2$，，但 $n \not\ll n \log_2 n$。最后，对于 $a, b \in \mathbb{R}$，令 $a \vee b \triangleq \max\{a, b\}$，且 $a \wedge b \triangleq \min\{a, b\}$。

2 问题表述

我们来描述植入二分图检测的一般设定。共有 n 个顶点。我们从 n 个顶点中均匀随机选取 k_R 个，并将这个集合记为 $\mathcal{R} \subset [n]$。从剩余的 $n - k_R$ 个顶点中，我们再均匀随机选取另一个包含 k_L 个顶点的集合，并将所得集合记为 $\mathcal{L} \subseteq [n] \setminus \mathcal{R}$。我们分别将 $\mathcal{R}$和 $\mathcal{L}$ 称为右侧和左侧隐匿集合。令 $\mathcal{E}(\mathcal{R},\mathcal{L})$ 为边集合$\{(i, j)\colon i < j,\ i \in \mathcal{R},\ j \in \mathcal{L}$ 或 $i \in \mathcal{L},\ j \in \mathcal{R}\}$，并令 $K_{\mathcal{R},\mathcal{L}}$ 为一个有 n 个顶点的空图，但包含 $\mathcal{E}(\mathcal{R},\mathcal{L})$ 中的所有元素。分别用 $V(K_{\mathcal{R},\mathcal{L}})$ $\mathcal{R} \cup \mathcal{L}$ 和 $E(K_{\mathcal{R},\mathcal{L}})$ $\mathcal{E}(\mathcal{R}, \mathcal{L})$ 表示 $K_{\mathcal{R},\mathcal{L}}$ 的顶点集和边集。注意，$|V(K_{\mathcal{R},\mathcal{L}})| = k_R + k_L$，且 $|E(K_{\mathcal{R},\mathcal{L}})| = k_R k_L$。我们将 $K_{\mathcal{R},\mathcal{L}}$ 称为植入结构。最后，令 $K^{p}_{\mathcal{R},\mathcal{L}}$ 为随机图，其构造方式是以概率 p 保留 $E(K_{\mathcal{R},\mathcal{L}})$ 中的每条边。因此，我们的检测问题可以表述为以下简单的假设检验问题：在零假设 $\mathcal{H}_0$ 下，观测图 G 是边密度为 $0 < q < 1$ 的 Erd˝os-R´enyi 随机图 $\mathcal{G}(n, q)$，其中该边密度可能是 n 的函数。在备择假设 $\mathcal{H}_1$ 下，我们首先按上述方式在 n 个顶点上抽取一个随机二分图 $K^{p}_{\mathcal{R},\mathcal{L}}$，然后以概率 q 连接任意其他两个顶点，只要其中至少一个顶点位于 $\mathcal{L} \cup \mathcal{R}$ 之外；所得图即为 G。因此，植入边的概率为 p，非隐匿边的概率为 q。在不失一般性的情况下

我们始终假设 $p > q$。[4] 定义备择分布的另一种等价方式如下：从 $\mathcal{G}(n,q)$ 中抽取一个基图 $\tilde{\mathsf{G}}$，并观测该基图与 $\mathsf{K}_{\mathcal{R}}^{p'},_{\mathcal{L}}$ 的并集，即 $\mathsf{G} = \tilde{\mathsf{G}} \cup \mathsf{K}_{\mathcal{R}}^{p'},_{\mathcal{L}}$，其中 $p' = p - q\,1 - q$。[5] 我们将备择假设下的图集合记作 $\mathcal{G}(n, k_{\mathsf{R}}, k_{\mathsf{L}}, p, q)$。

问题 1（检测问题）。 植入稠密二部子图（*PDBS*）检测问题，记为 $\mathsf{PDBS}(n,k_{\mathsf{R}},k_{\mathsf{L}},p,q)$，指的是区分以下假设的问题：

$$\mathcal{H}_0 : \mathsf{G} \sim \mathcal{G}(n,q) \quad \text{vs.} \quad \mathcal{H}_1 : \mathsf{G} \sim \mathcal{G}(n, k_{\mathsf{R}}, k_{\mathsf{L}}, p, q). \tag{3}$$

我们研究上述框架在渐近区域中的情形，其中 $\mathsf{n} \to \infty$，且 $(\mathsf{k_R}, \mathsf{k_L}, \mathsf{p}, \mathsf{q})$ 也可能随 n 变化。全文中，我们假设 $\mathsf{k_R} + \mathsf{k_L} = \mathsf{o}(\mathsf{n})$。此外，“稠密区域”和“稀疏区域”分别对应 $\mathsf{p}, \mathsf{q} = \Theta(1)$ 和 $\mathsf{p}, \mathsf{q} = \Theta(\mathsf{n}^{-\alpha})$ 的情形。观察到 G 后，目标是设计一个检验 ϕ：$\{0,1\}^{\binom{n}{2}} \to \{0,1\}$，以区分 $\mathcal{H}_0$ 和 $\mathcal{H}_1$。具体而言，检验 ϕ 的第一类和第二类错误平均风险定义为 $\mathsf{R}_n(\phi) = \mathbb{P}_{\mathcal{H}_0}(\phi(\mathsf{G}) = 1) + \mathbb{P}_{\mathcal{H}_1}(\phi(\mathsf{G}) = 0)$。一列以 n 为索引的检验 ϕ_n 若满足 $\lim_{n\to\infty} \mathsf{R}_n(\phi_n) = 0$，则称其成功；若 $\lim_{n\to\infty} \mathsf{R}_n(\phi_n) > 0$，则称其失败。最后，我们将最优风险记为 $\mathsf{R}_n^\star$，即 $\mathsf{R}_n^\star$, $\inf_{\phi_n} \mathsf{R}_n(\phi_n)$。上述内容总结如下定义。

定义 2 (强检测)。 设 $\mathbb{P}_{\mathcal{H}_0}$ 和 $\mathbb{P}_{\mathcal{H}_1}$ 分别为在均匀假设和植入假设下 G 的分布。若某个可能采用随机化的算法 $\phi_n(\mathsf{G}) \in \{0,1\}$ 的第一类+<*style id='22*>第二类 风险满足 $\lim_{n\to\infty} \mathsf{R}_n(\phi_n) = 0$，则该算法实现强检测。反之，若对于所有可能采用随机化的算法序列 $\phi_n(\mathsf{G}) \in \{0,1\}$，其第一类+<*style id='47*>第二类 风险均满足 $\liminf_{n\to\infty} \mathsf{R}_n(\phi_n) > 0$，则强检测是不可能的。

3 主要结果

本节将确定植入二部子图检测问题的统计与计算极限。我们首先给出信息论下界，接着介绍算法上界。然后陈述计算下界，最后给出若干相图，以展示我们的结果如何随模型中的各项参数变化。

统计下界。 以下结果给出了强检测在统计上（或信息论上）不可能的条件。为简化记号，我们定义

$$\gamma_n(x,y) \triangleq \log\left(1 + \frac{n}{xy}\frac{\log 2}{2}\right), \tag{4}$$

[4]对于$p > q$ 情形的任何可实现性结果或$p < q$ 情形的任何反向结论，都可以通过翻转观测图中的每条边得到相应结果。[5]概率p' 的选取使得并图中植入边的概率为p，即$p = p' + q(1-p')$。

对于任意 x、$y \in \mathbb{N}$。此外，回顾一下，两个伯努利随机变量 $\mathrm{Bern}(p)$ 和 $\mathrm{Bern}(q)$ 之间的 χ^2-散度定义为 $\chi^2(p||q) = (p-q)^2\, q(1-q)$。

定理1（统计下界）。考虑 PDBS（自然数集，k_{R}，k_{L}，p，q）问题*1*中的检测问题*1*。若满足以下条件，则强检测不可能，且 $\mathsf{R}_n^\star > 1/2$：

$$\chi^2(p||q) \le \frac{n \cdot \gamma_n(k_{\mathsf{R}}, k_{\mathsf{L}})}{8 k_{\mathsf{R}} k_{\mathsf{L}}}, \tag{5a}$$

并且以下两个条件中至少有一个成立：

$$\chi^2(p||q) \le \frac{\gamma_n(k_{\mathsf{R}}, k_{\mathsf{R}})}{2k_{\mathsf{L}}} \vee \frac{\gamma_n(k_{\mathsf{L}}, k_{\mathsf{L}})}{2k_{\mathsf{R}}}, \tag{5b}$$

$$\chi^2(p||q) \le \left(\frac{1}{2k_{\mathsf{L}}} \wedge \frac{n \cdot \gamma_n(k_{\mathsf{L}}, k_{\mathsf{L}})}{8k_{\mathsf{R}}^2}\right) \vee \left(\frac{1}{2k_{\mathsf{R}}} \wedge \frac{n \cdot \gamma_n(k_{\mathsf{R}}, k_{\mathsf{R}})}{8k_{\mathsf{L}}^2}\right). \tag{5c}$$

这里有几点说明。首先需要强调的是，尽管上述定理对 p 和 q的任意取值都成立，但我们稍后会看到，针对稠密和稀疏区域时，式（5）中的某些条件会变得多余。其次，我们将看到，在多项式尺度下，式（5）中的阈值在解析上会简单得多，也更容易理解。最后需要指出，定理 1 中的各个常数并非最优；选择这些常数是为了使证明更加清晰，而我们主要关注的是检测阈值的渐近行为。

上界。现在我们来描述上界。为此，我们提出三个算法并分析其性能。定义以下统计量：

$$\mathsf{Scan}(\mathsf{A}) \triangleq \max_{(\mathcal{R}', \mathcal{L}') \in \mathcal{S}_{k_{\mathsf{R}}, k_{\mathsf{L}}}} \sum_{i \in \mathcal{R}', j \in \mathcal{L}'} \mathsf{A}_{ij}, \tag{6}$$

$$\mathsf{Count}(\mathsf{A}) \triangleq \sum_{i<j} \mathsf{A}_{ij}, \tag{7}$$

$$\mathsf{MaxDeg}(\mathsf{A}) \triangleq \max_{1 \le i \le n} \sum_{j=1}^{n} \mathsf{A}_{ij}, \tag{8}$$

其中 $\mathcal{S}_{k_{\mathsf{R}}, \mathsf{k}_{\mathsf{L}}}$，$\{(\mathcal{R}', \mathcal{L}') \subset [\mathrm{n}] \times [\mathrm{n}] : |\mathcal{R}'| = \mathrm{k}_{\mathsf{R}}, |\mathcal{L}'| = \mathrm{k}_{\mathsf{L}}, \mathcal{R}' \cap \mathcal{L}' = \emptyset\}$。接下来，我们提出的检验包括 ϕ_{Scan}，$\mathbb{I}\{\mathsf{Scan}(\mathsf{A}) \ge \tau_{\mathsf{Scan}}\}$，$\phi_{\mathsf{Count}}$，$\mathbb{I}\{\mathsf{Count}(\mathsf{A}) \ge \tau_{\mathsf{Count}}\}$，以及 ϕ_{Deg}，$\mathbb{I}\{\quad\quad \ge \tau_{\mathsf{Deg}}\}$

$\mathsf{MaxDeg}(\mathsf{A})$。如引言中所述，上述检验相当经典，已在许多其他相关问题中提出，例如[8,10,19,20,30,61–63]。此外，计数检验和最大度检验的计算复杂度为 $\mathrm{O}(\mathrm{n}^2)$ 次运算，属于多项式时间，因此效率较高。扫描检验则不然，其计算复杂度呈指数级，因此效率较低。事实上，式 (6) 中的搜索空间大小为 $|\mathcal{S}_{k_{\mathsf{R}}, \mathrm{k}_{\mathsf{L}}}| = (\,\mathrm{n}\;\mathrm{k}_{\mathsf{R}})(^{n-k_{\mathsf{R}}}\mathrm{k}_{\mathsf{L}}\,)$，当 $\mathrm{k}_{\mathsf{R}} \vee \mathrm{k}_{\mathsf{L}} = \omega(1)$ 时，搜索空间就至少已达到准多项式规模。下面的结果给出了上述各项检验风险较小的充分条件。

定理 2 (算法上界)。 固定 $\delta \in (0,1)$。 考虑 PDBS(n, k_R, k_L, p, q)问题 *1* 中的检测问题*1*，并假设 p 和 q 满足 $|p-q| = O(q)$ 且 $q < c < 1$，其中 c $0 <$ 为某个常数 < 1。那么，

1. 对于 $\tau_{\mathsf{Scan}} = k_R k_L \frac{p+q}{2}$，我们有 $R_n(\phi_{\mathsf{Scan}}) \le \delta$，条件是：

$$\chi^2(p||q) = \Omega\left(\frac{\log\left(\frac{n}{k_R \vee k_L}\right) + \frac{\log\frac{2}{\delta}}{k_R \vee k_L}}{k_R \wedge k_L}\right). \tag{9}$$

2. 对于 $\tau_{\mathsf{Count}} = \binom{n}{2} q + k_R k_L \frac{p-q}{2}$，我们有 $R_n(\phi_{\mathsf{Count}}) \le \delta$，条件是：

$$\chi^2(p||q) = \Omega\left(\frac{n^2}{k_R^2 k_L^2} \cdot \log\frac{2}{\delta}\right). \tag{10}$$

3. 对于 $\tau_{\mathsf{Deg}} = (n-1)q + k_R \vee k_L \frac{p-q}{2}$，我们有 $R_n(\phi_{\mathsf{Deg}}) \le \delta$，条件是：

$$\chi^2(p||q) = \Omega\left(\frac{n}{k_R^2 \vee k_L^2} \cdot \left(\log n + \log\frac{2}{\delta}\right)\right). \tag{11}$$

此时，我们希望将定理 1 和定理 2 中的界限缩减到稠密区域和稀疏区域，先从前者开始。在稠密情形下，我们有 $p, q = \Theta(1)$，因此 $\chi^2(p||q) = \Theta(1)$ 也成立。鉴于 k_R、$k_L = o(n)$，显然，(5a) 中的条件比 (5b) 中的条件宽松，因此两者中只有后者起作用。因此，若 (5b) 或 (5c) 成立，强检测便不可能；正如我们接下来将要展示的，(5b) 可以由几乎匹配的上界补足，因此无需考虑 (5c)。忽略部分多对数因子时，这一点最容易看出。具体而言，注意到当 $k_R^2 \vee k_L^2 \gg n$ 时，(5b) 永远不可能成立，因为 $\chi^2(p||q)$ $=$ Θ 1 5 ( )。因此，在这种情况下，必有 $k_R^2 \vee k_L^2 \ll n$，此时，(b) 可归结为条件 $k_R \wedge k_L = O(\log n)$。总而言之，在稠密区域中，若以下条件成立，则强检测在统计上不可能：

$$k_R \wedge k_L = O(\log n) \quad \text{and} \quad k_R^2 \vee k_L^2 \ll n. \tag{12}$$

这与扫描和度检验的渐近性能相仿。事实上，利用定理2，我们从（9）和（11）可知，当满足以下条件时可以实现强检测：

$$k_R \wedge k_L = \Omega(\log n) \quad \text{or} \quad k_R^2 \vee k_L^2 \gg n. \tag{13}$$

具体而言，(13) 左侧的条件保证了 $R_n(\phi_{\mathsf{Scan}}) \to 0$，当$n \to \infty$时；而 (13) 右侧的条件保证了 $R_n(\phi_{\mathsf{Deg}}) \to 0$，当 $n \to \infty$时。注意，上述结果意味着，在稠密区域，统计障碍取决于二部植入图中的量 $k_R \wedge k_L$ 和 $k_R \vee k_L$；从本质上说，它们分别等于最大子图密度和最大

植入结构的顶点度。此外，我们还想在此指出，尽管上述讨论着重于多项式尺度，尤其是关于 (12) 和 (13) 右侧的条件，但定理 1 和 2 也给出了对数因子。最后值得注意的是，在稠密区域，计数检验是多余的。稀疏区域则并非如此，下面将对此作出解释。

我们接下来讨论稀疏区域，其中 $p, q = \Theta(n^{-\alpha})$，对某个 $\alpha \in (0, 2]$,而言，这意味着 $\chi^2(p||q) = \Theta(n^{-\alpha})$。在这种情况下，我们需要定理1中的条件 (5a) 和 (5c)。与稠密情形一样，我们将这些条件简化到多项式尺度。为此，我们利用如下事实：$\gamma_n(x, y) = \Omega(1 \wedge nxy)$。于是，(5a) 可写为 $\chi^2(p||q) = O\left(n k_R k_L \wedge n^2 k_R^2 k_L^2\right)$。类似地，(5c) 可写为以下两种情形的并集：$\chi^2(p||q) = O\left(1 k_L \wedge n k_R^2 \wedge n^2 k_R^2 k_L^2\right)$，以及 $\chi^2(p||q) = O\left(1 k_R \wedge n k_L^2 \wedge n^2 k_R^2 k_L^2\right)$。将这些条件取交集可得，当 $\chi^2(p||q) = O\left(1 k_L \wedge n^2 k_R^2 k_L^2 \wedge n k_R^2\right)$，或 $\chi^2(p||q) = O\left(1 k_R \wedge n^2 k_R^2 k_L^2 \wedge n k_L^2\right)$时，强检测不可能。这里我们利用了这一事实：渐近地，$n k_R k_L$ 不可能在取最小值时占优，因为根据假设，$k_R, k_L = o(n)$，有 $1 k_R, 1 k_L = o\left(n k_R k_L\right)$。最后，不难验证，最后两个条件的并集给出：

$$\chi^2(p||q) = O\left(\frac{1}{k_R \wedge k_L} \wedge \frac{n^2}{k_R^2 k_L^2} \wedge \frac{n}{k_R^2 \vee k_L^2}\right). \tag{14}$$

这一障碍模拟了扫描、计数和度检验的渐近性能。事实上，利用定理2，由（9）、（10）和（11）可知，若满足以下条件，则可以实现强检测：

$$\chi^2(p||q) = \Omega\left(\frac{\log n}{k_R \wedge k_L} \wedge \frac{n^2 \log n}{k_R^2 k_L^2} \wedge \frac{n \log n}{k_R^2 \vee k_L^2}\right). \tag{15}$$

事实上，右侧的（(·) 表达式中从左到右的各项分别保证，当$n \to \infty$时，$R_n(\phi_{Scan}) \to 0$、$R_n(\phi_{Count}) \to 0$和$R_n(\phi_{Deg}) \to 0$。随后，我们将通过相图展示上述界限，并说明各种参数如何随n变化而权衡。一个显而易见且重要的观察是，要达到信息论下界（相差至多多对数因子），高效计数检验和度检验以及穷举扫描检验缺一不可。然而，如上所述，扫描检验并不高效。因此，从计算角度来看，人们可能会想知道，能否用一种具有相同性能保证的高效检验取代低效的扫描检验，也就是说，能否达到与扫描检验相同的统计障碍。接下来我们将依据低次多项式猜想提供证据，表明在扫描检验成功而计数和度检验失败的区域内，不存在以多项式时间运行的高效算法。

计算下界。审视上述下界和上界可以发现，统计上可实现的结果与高效可实现的结果之间存在差距。在

具体而言，在稠密区域中，参数范围为 $k_R^2 \vee k_L^2 \ll n$；在稀疏区域中，参数范围为 $1 \; k_R \wedge k$ $\ll \chi^2(p||q) \ll n^2 \; k_R^2 k_L^2 \wedge n \; k_R^2 \vee k$ {v39>，尽管可以通过穷举搜索解决检测问题，但我们没有多项式时间算法。如上所述，我们猜想，在这些参数区域内实际上不存在多项式时间算法，接下来将为这一猜想提供证据。为此，我们先简要介绍低次多项式方法。

该方法的基本前提是，所有用于解决检测问题的多项式时间算法，都可以由低次多项式刻画。迄今为止，越来越多有力证据支持这一猜想。以下思想最早形成于平方和优化领域的一系列研究中 [14,15,42,49]。我们沿用 [15,64] 中的符号和定义。任意分布 $\mathbb{P}_{\mathcal{H}_0}$ on $\Omega_n = \{0,1\}^{\binom{n}{2}}$ 都会在可测函数 $f, g: \Omega_n \to \mathbb{R}$ 上诱导出一个内积，定义为 $\langle f, g \rangle_{\mathcal{H}_0} = \mathbb{E}_{\mathcal{H}_0}[f(G)g(G)]$,; 相应的范数为 $\|f\|_{\mathcal{H}_0} = \langle f, f \rangle_{\mathcal{H}_0}^{1/2}$。我们用 $L^2(\mathbb{P}_{\mathcal{H}_0})$ 表示由满足 $\|f\|_{\mathcal{H}_0} < \infty$ 的函数构成的希尔伯特空间，并赋予上述内积和范数。核心思想是找到能够在 L^2 意义下最佳地区分 $\mathbb{P}_{\mathcal{H}_0}$ 与 $\mathbb{P}_{\mathcal{H}_1}$ 的低次多项式。设 $\mathcal{L}_{n,\le D} \subset L^2(\mathbb{P}_{\mathcal{H}_0})$ 表示由次数至多为 $D \in \mathbb{N}$ 的多项式$\Omega_n \to \mathbb{R}$ 构成的线性子空间。于是，D 次低次似然比 $L_{n,\le D}$ 是函数 L_n 在坐标次数至多为 D 的函数张成的空间上的投影，其中投影是相对于内积 $\langle \cdot,\cdot \rangle_{\mathcal{H}_0}$ 正交的。如上所述，似然比能够在 L^2 意义下最优地区分 $\mathbb{P}_{\mathcal{H}_0}$ 与 $\mathbb{P}_{\mathcal{H}_1}$。下一个引理表明，在低次多项式集合中，$D$ 次低次似然比也具有相同的性质。

引理 1(O ptimally of $L_{n,\le D}$ [14,49,64])。考虑以下优化问题 :

$$\max \mathbb{E}_{\mathcal{H}_1} f(G) \quad \text{s.t.} \quad \mathbb{E}_{\mathcal{H}_0} f^2(G) = 1, \; f \in \mathcal{L}_{n,\le D}, \tag{16}$$

那么，(16) 的唯一解 $f^\star$ 是D-低度似然比 $f^\star = L_{n,\le D} / \|L_{n,\le D}\|_{\mathcal{H}_0}$，且该优化问题的值为 $\|L_{n,\le D}\|_{\mathcal{H}_0}$。

如上所述，在计算无界区域中，似然比的一个重要性质是，若 $\|L_n\|_{\mathcal{H}_0}$ 有界，则 $\mathbb{P}_{\mathcal{H}_0}$ 和 $\mathbb{P}_{\mathcal{H}_1}$ 在统计上不可区分。以下猜想指出，这一性质的计算类比也成立，其中 $L_{n,\le D}$ 扮演似然比的角色。事实上，该猜想还提出，次数为 $\approx \log n$ 的多项式可以作为多项式时间算法的替代。下面的猜想基于 [14,15,49]，和 [15, 猜想 2.2.4]。我们给出该猜想的非正式表述，见[64, 猜想 1.16]。关于精确表述，请读者参阅例如 [15, 猜想 2.2.4] 和 [64, 第 4] 节。

猜想 1（低次猜想，非正式表述）。给定一列概率测度 $\mathbb{P}_{\mathcal{H}_0}$和$\mathbb{P}_{\mathcal{H}_1}$，如果存在 $\epsilon > 0$ 且 $D = D(n) \ge (\log n)^{1+\epsilon}$，使得 $\|L_{n,\le D}\|_{\mathcal{H}_0}$ 在 $n \to \infty$ 时仍有界，那么就不存在能够区分 $\mathbb{P}_{\mathcal{H}_0}$ 和 $\mathbb{P}_{\mathcal{H}_1}$ 的多项式时间算法。

在下文中，我们将依赖猜想 1 ，为我们的问题中观察到的统计—计算鸿沟提供依据。此处我们想提及 [15,假设 2.1。5], 该假设陈述了猜想 1 的一种更一般的形式，因为它假定，度-D多项式是 $n^{O(D)}$时间算法的替代形式。

定理 3。 考虑问题 PDBS(n,k_R,k_L, p,q) 中的检测问题，见问题 *1*。

1. 在 p,q $=\Theta(1)$ 的范围内，若 $(k_R^2 \vee k_L^2) \ll n$，则 $\|\mathbb{L}_{n,\le D}\|_{\mathcal{H}_0} \le O(1)$，对任意满足 $D \le C \log n$ 且 $C > 0$ 的情况均成立。反过来，存在一个正整数 D，使得若 $(k_R^2 \vee k_L^2) \gg n$，则 $\|\mathbb{L}_{n,\le D}\|_{\mathcal{H}_0} \ge \omega(1)$。

2. 在 $p, q = \Theta(n^{-\alpha})$ 的范围内，对某个 $\alpha \in (0,2]$，若 $^1(k_R \wedge k_L) \ll \chi^2(p\|q) \ll n^2 k_R^2 k_L^2 \wedge n(k_R^2 \vee k_L^2)$，则 $\|L_{n,\le D}\|_{\mathcal{H}_0} \le O(1)$，对任意 $D = n^{o(1)}$ 均成立。反过来，存在一个正整数 D，使得若 $\chi^2(p\|q) \gg n^2 k_R^2 k_L^2 \wedge n(k_R^2 \vee k_L^2)$，则 $\|L_{n,\le D}\|_{\mathcal{H}_0} \ge \omega(1)$。

结合猜想 1，定理 3 意味着，如果我们将度的对数级n 次多项式作为所有高效算法的替代指标，我们的计算预测，当 $n^{O(\log n)}$ 算法在稠密区域中不存在$(k_R^2 \vee k_L^2) \ll n$ 时，而在$^1(k_R \wedge k_L) \ll \chi^2(p\|q) \ll n^2 k_R^2 k_L^2 \wedge n(k_R^2 \vee k_L^2)$ 在稀疏区域中。这总结如下推论。

推论 1 (计算下界)。 考虑 PDBS(n，k_R，k_L，p，q) 问题 *1* 中的检测问题*1*，并假设猜想 *1* 成立。在以下情形中，不存在能够实现强检测的多项式时间算法：

1. 若 p，$q = \Theta(1)$ 且 $(k_R^2 \vee k_L^2) \ll n$。

2. If $p, q = \Theta(n^{-\alpha})$, *for some* $\alpha \in (0,2]$, *and* $\frac{1}{(k_R \wedge k_L)} \ll \chi^2(p\|q) \ll \frac{n^2}{k_R^2 k_L^2} \wedge \frac{n}{(k_R^2 \vee k_L^2)}$.

这些预测与前面各小节中已确立的统计—计算权衡完全一致。从定理 3的证明中可以推导出计算障碍的一个更明确公式，该公式体现了对 D 的依赖；为使论述简洁，我们选择呈现上述精化后的结果。

注 1. 我们想指出，尽管定理*3* 第二部分关注的是 $p, q = \Theta(n^{-\alpha})$的特定标度，但定理*3* 的结论及其证明，在 $p = \Theta(n^{-\alpha})$ 和 $q = \Theta(n^{-\beta})$的情形下同样成立，其中 $\alpha \le \beta \le 2\alpha$。这称为对数密度情形。

相图。定理1–3中推导出的统计和计算障碍，借助相图可以得到最佳诠释和理解。下面，我们考虑渐近情形，其中 $k = k_R + k_L$，且 (p, q) 是 n 的多项式，即 $k = \Theta(n^{\beta})$，其中 $\beta \in (0,1)$，且 $p = c \cdot q = \Theta(n^{-\alpha})$，其中 $\alpha \in (0,2]$,且 $c > 1$ 为固定值。我们不失一般性地假设 $k_R \ge k_L$。下方的相图包含三个重要区域：

1. 计算上容易的区域（蓝色和绿色区域）：存在用于检测任务的多项式时间算法（计数和度检验）。

2. 计算上困难区域（红色区域）：存在一种用于检测的低效算法（扫描检验），但从计算角度看，根据猜想该问题是困难的（不存在多项式时间算法），因为低次多项式类无法奏效。

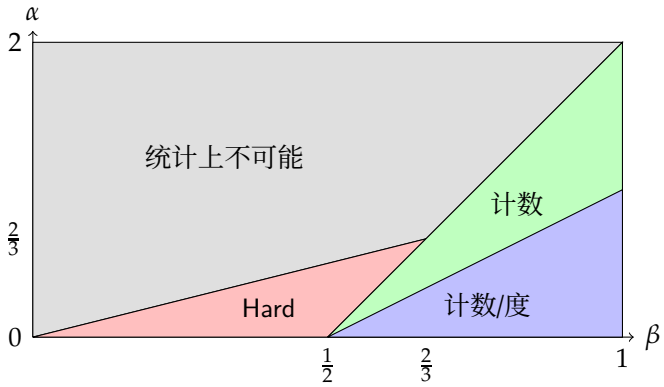


图3： 在平衡情形下，相图关于$k = \Theta(n^{\beta})$以及p、$q = \Theta(n^{-\alpha})$的变化，其中$k_R = \Theta(k_L)$。

3. 统计上不可能区域（灰色区域）：检测在统计上（或信息论意义上）是不可能的。

为得到以下相图，我们将 k_R、$k_L = \Theta(n^{\beta})$ 以及 $\chi^2(p||q) = \Theta(n^{-\alpha})$ 代入（14）和（15）以及推论 1。这里需要强调的是，以下相图反映了我们推导出的统计障碍和计算障碍的多项式尺度，这些尺度取决于各参数的多项式尺度，即 k_R、k_L、p 和 q 关于 n 的尺度。因此，这些图中没有体现我们的下界与上界之间的多对数因子差距。我们针对 k_R 和 k_L 的各种标度情形绘制了所得的相变：

- 图3 展示了“平衡情形”，其中$k_R = \Theta(k_L)$，对于该情形，我们有$m(K_{\mathcal{R}}, \mathcal{L}) = \Theta(k)$。有趣的是，我们注意到，所得相图与 PDS 问题的相图相同（例如，[8, 图 1]），其中最大子图密度的表现也相同。请注意，在这种情况下，度检验是冗余的。

- 图4展示了“轻度不平衡”的情形，此时 $k_R = \Theta(k)$，且 $\omega(\sqrt{k}) \le k_L \le o(k)$。在这种情况下，$\omega(\sqrt{k}) \le m(K_{\mathcal{R}}, \mathcal{L}) \le o(k)$，为便于说明，我们取 $k_L = \Theta(k^{\frac{2}{3}})$。

- 图5展示了“中度不平衡”的情形，此时 $k_R = \Theta(k)$，且 $\omega(1) \le k_L \le O(\sqrt{k})$。在这种情况下，$\omega(1) \le m(K_{\mathcal{R}}, \mathcal{L}) \le O(\sqrt{k})$。请注意，计数检验是冗余的；为便于说明，我们取 $k_L = \Theta(k^{\frac{1}{2}})$。

- 最后，图6展示了“极度不平衡”的情形，此时$k_R = \Theta(k)$ 且$k_L = \Theta(1)$。在这种情况下，$m(K_{\mathcal{R}}, \mathcal{L}) = \Theta(1)$。此情形在渐近意义上等价于植入星形子图，该问题已在 [11] 中研究。请注意，计数检验和扫描检验均为冗余的。

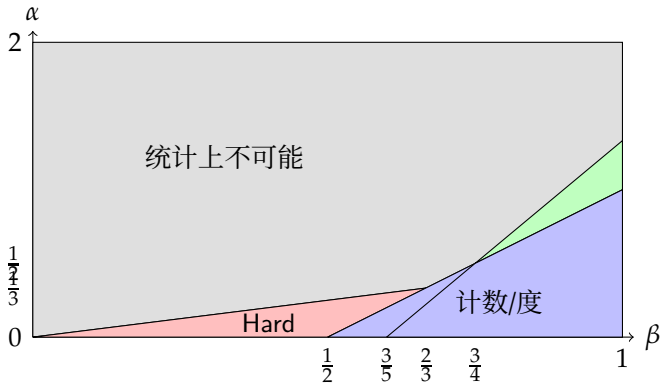


图4： 轻度不平衡情形下，关于$k = \Theta(n^{\beta})$ 以及p，$q = \Theta(n^{-\alpha})$的相图，其中$k_{\mathsf{R}} = \Theta(k)$ 且$k_{\mathsf{L}} = \Theta(k^{\frac{2}{3}})$。

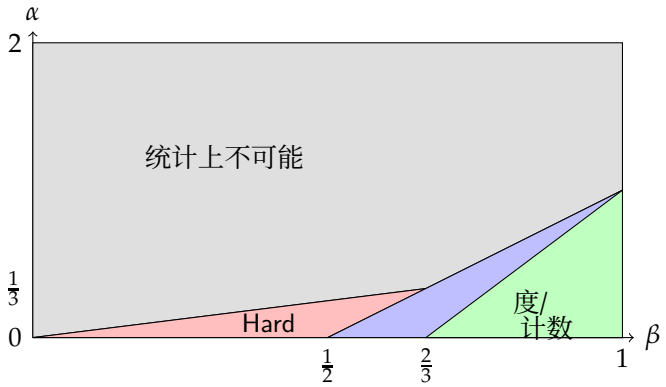


图5： 中度不平衡情形下，$k = \Theta(n^{\beta})$ 与p,$q = \Theta(n^{-\alpha})$的相图，其中$k_{\mathsf{R}} = \Theta(k)$ 且 $\omega(1) \leq k_{\mathsf{L}} \leq O(\sqrt{k})$；为便于说明，取$k_{\mathsf{L}} = \Theta(k^{\frac{1}{2}})$。

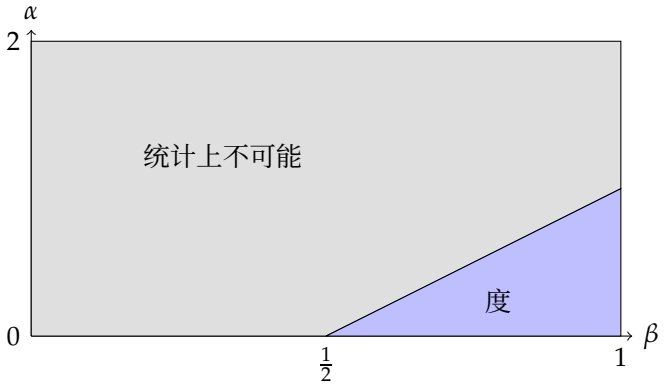


图6： 极度不平衡情形下，$k = \Theta(n^{\beta})$ 与p,$q = \Theta(n^{-\alpha})$的相图，其中$k_{\mathsf{R}} = \Theta(k)$ 且$k_{\mathsf{L}} = \Theta(1)$。

4 证明

4.1 定理1的证明

本小节中，我们通过找出使任意算法的风险严格为正的条件，来证明定理 1。为此，我们将用到以下众所周知的事实：使平均风险最小的最优检验，记为 $\phi_n^\star$，由下式给出：

$$\phi_n^\star(\mathsf{G}) \triangleq \begin{cases} 1, & \text{if } \mathsf{L}_n(\mathsf{G}) \geq 1 \\ 0, & \text{otherwise,} \end{cases} \tag{17}$$

其中 $\mathsf{L}_n(\mathsf{G}) \overset{\mathbb{P}_{\mathcal{H}_1}(\mathsf{G})}{\underline{\mathbb{P}_{\mathcal{H}_0}(\mathsf{G})}}$。与上述最优检验对应的最优风险 $\mathsf{R}_n^\star$可利用柯西–施瓦茨不等式得到如下下界：

$$\begin{aligned} \mathsf{R}_n^\star &= 1 - d_{\mathsf{TV}}\left(\mathbb{P}_{\mathcal{H}_0}, \mathbb{P}_{\mathcal{H}_1}\right) && (18)\\ &= 1 - \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}|\mathsf{L}_n(\mathsf{G}) - 1| && (19)\\ &\geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\mathcal{H}_0}[\mathsf{L}_n^2(\mathsf{G})] - 1}. && (20) \end{aligned}$$

因此，要找出最优风险为正的条件，我们可以找出使 $\mathbb{E}_{\mathcal{H}_0}[\mathsf{L}_n^2(\mathsf{G})]$ 落在区间 $(1, 5)$内的条件。

我们首先推导 $\mathcal{H}_0$ 下似然的二阶矩公式。注意，似然比由下式给出

$$\mathsf{L}_n(\mathsf{G}) = \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left[\frac{\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}_{\mathcal{R},\mathcal{L}})}{\mathbb{P}_{\mathcal{H}_0}(\mathsf{G})}\right], \tag{21}$$

其中，期望是相对于在 $\mathcal{K}_{n,k_{\mathsf{R}},k_{\mathsf{L}}}$ 上均匀分布的 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 而取的，并且$\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}_{\mathcal{R},\mathcal{L}})$ 表示给定 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 是在 $\mathcal{H}_1$ 下植入的二分图时 G 的条件分布。引入 $\mathsf{K}'_{\mathcal{R},\mathcal{L}}$ 的一个独立副本（其中 $\mathsf{K}'_{\mathcal{R},\mathcal{L}}$ 与 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 独立），平方似然比可表示为

$$\mathsf{L}_n^2(\mathsf{G}) = \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}}\left[\frac{\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}_{\mathcal{R},\mathcal{L}})\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}'_{\mathcal{R},\mathcal{L}})}{\mathbb{P}^2_{\mathcal{H}_0}(\mathsf{G})}\right], \tag{22}$$

其中，$\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}$ 表示 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 与 $\mathsf{K}'_{\mathcal{R},\mathcal{L}}$ 相互独立。交换期望的顺序可得，

$$\mathbb{E}_{\mathcal{H}_0}[\mathsf{L}_n^2(\mathsf{G})] = \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}}\left[\mathbb{E}_{\mathcal{H}_0}\left(\frac{\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}_{\mathcal{R},\mathcal{L}})\mathbb{P}_{\mathcal{H}_1}(\mathsf{G}|\mathsf{K}'_{\mathcal{R},\mathcal{L}})}{\mathbb{P}^2_{\mathcal{H}_0}(\mathsf{G})}\right)\right]. \tag{23}$$

令 A 表示 G 的邻接矩阵，对于任意一对 $(i, j) \in \binom{[n]}{2}$，我们定义 $\mathsf{f}(\mathsf{A}^{ij}) \triangleq (p/q)^{\mathsf{A}_{ij}}[(1-p)/(1-q)]^{\mathsf{A}_{ij}}$。接着注意到 $\mathbb{E}_{\mathcal{H}_0}[\mathsf{f}(\mathsf{A}^{ij})] = 1$和 $\mathbb{E}_{\mathcal{H}_0}[\mathsf{f}^2(\mathsf{A}^{ij})] = 1 + \chi^2(p||q)$。因此，

$\mathbb{E}_{\mathcal{H}_0}[\mathsf{L}_n^2(\mathsf{G})]$

$$= \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}} \left[\prod_{(i,j)\in \mathsf{K}_{\mathcal{R},\mathcal{L}} \cap \mathsf{K}'_{\mathcal{R},\mathcal{L}}} \mathbb{E}_{\mathcal{H}_0} \left[f^2 \left(\mathsf{A}_{ij} \right) \right] \cdot \prod_{(i,j)\in \mathsf{K}_{\mathcal{R},\mathcal{L}} \triangle \mathsf{K}'_{\mathcal{R},\mathcal{L}}} \mathbb{E}_{\mathcal{H}_0} \left[f \left(\mathsf{A}_{ij} \right) \right] \right] \tag{24}$$

$$= \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}} \left[\left(1 + \chi^2(p||q) \right)^{|E\left(\mathsf{K}_{\mathcal{R},\mathcal{L}} \cap \mathsf{K}'_{\mathcal{R},\mathcal{L}}\right)|} \right] \tag{25}$$

$$\leq \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}} \perp\!\!\!\perp \mathsf{K}'_{\mathcal{R},\mathcal{L}}} \left[\exp\left(\chi^2(p||q) \cdot |E\left(\mathsf{K}_{\mathcal{R},\mathcal{L}} \cap \mathsf{K}'_{\mathcal{R},\mathcal{L}} \right)| \right) \right], \tag{26}$$

其中，不等式是由以下事实得出的： $1 + x \leq \exp(x)$，其中 $x \geq 0$。注意，(26) 只是两个随机副本 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 交集的矩生成函数。我们这里的主要观察是

$$|E\left(\mathsf{K}_{\mathcal{R},\mathcal{L}} \cap \mathsf{K}'_{\mathcal{R},\mathcal{L}} \right)| = |\mathcal{R} \cap \mathcal{R}'| \cdot |\mathcal{L} \cap \mathcal{L}'| + |\mathcal{R} \cap \mathcal{L}'| \cdot |\mathcal{L} \cap \mathcal{R}'|, \tag{27}$$

这直接源于二分图的结构。然后，根据柯西–施瓦茨不等式，

$$\mathbb{E}^2 \left[e^{\chi^2(p||q)|E\left(\mathsf{K}_{\mathcal{R},\mathcal{L}} \cap \mathsf{K}'_{\mathcal{R},\mathcal{L}}\right)|} \right] = \mathbb{E}^2 \left[e^{\chi^2(p||q)(|\mathcal{R}\cap\mathcal{R}'|\cdot|\mathcal{L}\cap\mathcal{L}'| + |\mathcal{R}\cap\mathcal{L}'|\cdot|\mathcal{L}\cap\mathcal{R}'|)} \right] \tag{28}$$

$$\leq \mathbb{E}\left[e^{2\chi^2(p||q)|\mathcal{R}\cap\mathcal{R}'|\cdot|\mathcal{L}\cap\mathcal{L}'|} \right] \cdot \mathbb{E}\left[e^{2\chi^2(p||q)|\mathcal{R}\cap\mathcal{L}'|\cdot|\mathcal{L}\cap\mathcal{R}'|} \right]. \tag{29}$$

接下来，我们对（29）右侧的第一项进行界定。由于 $\mathcal{R}$ 和 $\mathcal{R}'$ 是在给定 H $|\mathcal{L} \cap \mathcal{L}'|$的条件下均匀随机抽取的，因此有：

$$|\mathcal{R} \cap \mathcal{R}'| \preceq \mathsf{Hypergeometric}\left(n - H, k_{\mathsf{R}}, k_{\mathsf{R}} \right), \tag{30}$$

其中，如果随机变量 X 被 Y 随机支配，则记作 $\mathsf{X} \preceq \mathsf{Y}$。众所周知（参见例如 [20, 引理 5]），超几何分布 $(\mathsf{n} - \mathsf{H}, \mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{R}}) \preceq$ 二项分布 $\left(\mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{R}} \mathsf{n} - \mathsf{H} \right)$。由于我们假设 $\mathsf{k}_{\mathsf{R}} + \mathsf{k}_{\mathsf{L}} = \mathsf{o}(\mathsf{n})$，因此 $\mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{L}} \leq \mathsf{n}/2$，这又意味着 $\mathsf{H} \leq \mathsf{k}_{\mathsf{L}} \leq \mathsf{n}/2$，进而有二项分布 $\left(\mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{R}} \mathsf{n} - \mathsf{H} \right) \preceq$ 二项分布 $\left(\mathsf{k}_{\mathsf{R}}, 2\mathsf{k}_{\mathsf{R}} \mathsf{n} \right)$。因此，根据全概率公式可得：

$$\mathbb{E}\left[e^{2\chi^2(p||q)|\mathcal{R}\cap\mathcal{R}'|\cdot|\mathcal{L}\cap\mathcal{L}'|} \right] = \mathbb{E}\left[\mathbb{E}\left[e^{2\chi^2(p||q)|\mathcal{R}\cap\mathcal{R}'|\cdot|\mathcal{L}\cap\mathcal{L}'|} \,\middle|\, |\mathcal{L}\cap\mathcal{L}'| = H \right] \right] \tag{31}$$

$$\leq \mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{Binomial}\left(k_{\mathsf{R}}, \frac{2k_{\mathsf{R}}}{n}\right)\cdot|\mathcal{L}\cap\mathcal{L}'|} \right] \tag{32}$$

$$\leq \mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{Binomial}\left(k_{\mathsf{R}}, \frac{2k_{\mathsf{R}}}{n}\right)\cdot\mathsf{Binomial}\left(k_{\mathsf{L}}, \frac{2k_{\mathsf{L}}}{n}\right)} \right] \tag{33}$$

$$\triangleq \mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{B}\cdot\mathsf{B}'} \right], \tag{34}$$

其中 $\mathsf{B} \triangleq$ 二项分布 $\left(k_{\mathsf{R}}, \frac{2k_{\mathsf{R}}}{n} \right)$ 和 $\mathsf{B}' \triangleq$ 二项分布 $\left(k_{\mathsf{L}}, \frac{2k_{\mathsf{L}}}{n} \right)$。同理，(29) 右侧的第二项可以如下界定：

$$\mathbb{E}\left[e^{2\chi^2(p||q)|\mathcal{R}\cap\mathcal{L}'|\cdot|\mathcal{L}\cap\mathcal{R}'|} \right] \leq \mathbb{E}\left[e^{2\chi^2(p||q)\bar{\mathsf{B}}\cdot\bar{\mathsf{B}}'} \right], \tag{35}$$

其中 $\bar{\mathsf{B}} \triangleq$ 二项分布 $\left(k_{\mathsf{R}}, {}^{2}k_{\mathsf{L}}n\right)$ 和 $\mathsf{B}' \triangleq$ 二项分布 $\left(k_{\mathsf{L}}, {}^{2}k_{\mathsf{R}}n\right)$。我们的目标现在是找出条件，使得 (29) 中的每一期望项都被上界限制，例如不超过 2。为此，我们将使用在 (34) 和 (35) 中推导出的上界。这些上界进而意味着 (26) 成立，因此二阶矩也被上界限制为 2；根据 (20)，这意味着强检测是不可能的。

我们先用两种不同的方法对（34）进行上界估计。第一种方法将得到条件（5b），第二种方法则将得到条件（5c）。因此，只要满足其中任一条件，（34）就有上界 2。先来看第一种方法。利用 $\mathsf{B}' \leq k_{\mathsf{L}}$ 以概率 1 成立这一事实，可得：

$$\mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{B}\cdot\mathsf{B}'}\right] \leq \mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{B}\cdot k_{\mathsf{L}}}\right] \tag{36}$$

$$= \left[1 + 2\frac{k_{\mathsf{R}}}{n}\left(e^{2\chi^2(p||q)k_{\mathsf{L}}} - 1\right)\right]^{k_{\mathsf{R}}}, \tag{37}$$

这里用到了二项随机变量 B 的矩生成函数。现在，由于 $2^{1k_{\mathsf{R}}} - 1 \geq \log 2\, k_{\mathsf{R}}$，若满足以下条件，则（37）右端的上界为 2：

$$2\frac{k_{\mathsf{R}}}{n}\left(e^{2\chi^2(p||q)k_{\mathsf{L}}} - 1\right) \leq \frac{\log 2}{k_{\mathsf{R}}}, \tag{38}$$

这等价于

$$\chi^2(p||q) \leq \frac{1}{2k_{\mathsf{L}}}\log\left(1 + \frac{n}{k_{\mathsf{R}}^2}\frac{\log 2}{2}\right). \tag{39}$$

如果重复上述论证，但改为在（36）中使用上界 $\mathsf{B} \leq k_{\mathsf{R}}$，那么若满足以下条件，

$$\chi^2(p||q) \leq \frac{1}{2k_{\mathsf{R}}}\log\left(1 + \frac{n}{k_{\mathsf{L}}^2}\frac{\log 2}{2}\right), \tag{40}$$

则（34）右端的上界为 2。综合这两个条件，可得：

$$\chi^2(p||q) \leq \max\left\{\frac{1}{2k_{\mathsf{L}}}\log\left(1 + \frac{n}{k_{\mathsf{R}}^2}\frac{\log 2}{2}\right), \frac{1}{2k_{\mathsf{R}}}\log\left(1 + \frac{n}{k_{\mathsf{L}}^2}\frac{\log 2}{2}\right)\right\}, \tag{41}$$

这正是条件（5b）。

接下来，我们提出另一种对（34）进行界定的方法。利用二项随机变量 B的矩生成函数，我们得到，

$$\mathbb{E}\left[e^{2\chi^2(p||q)\mathsf{B}\cdot\mathsf{B}'}\right] = \mathbb{E}\left[\left(1 + 2\frac{k_{\mathsf{R}}}{n}\left(e^{2\chi^2(p||q)\mathsf{B}'} - 1\right)\right)^{k_{\mathsf{R}}}\right]. \tag{42}$$

我们假设 $2\chi^2(p||q)k_{\mathsf{L}} \leq 1$，因此结合 $\mathsf{B}' \leq k_{\mathsf{L}}$ 以概率 1 成立这一事实，可得 $2\chi^2(p||q)\mathsf{B}' \leq 1$ 以概率 1 成立。因此，利用当 $x < 1.79$ 时，$\exp(x) - 1 \leq x + x^2$，这一事实，我们得到

$$\begin{aligned}
&\mathbb{E}\left[\left(1+2\frac{k_{\mathsf{R}}}{n}\left(e^{2\chi^2(p||q)\mathsf{B}'}-1\right)\right)^{k_{\mathsf{R}}}\right] \\
&\qquad \leq \mathbb{E}\left[\left(1+2\frac{k_{\mathsf{R}}}{n}\left(2\chi^2(p||q)\mathsf{B}'+\left(2\chi^2(p||q)\mathsf{B}'\right)^2\right)\right)^{k_{\mathsf{R}}}\right] && (43)\\
&\qquad \leq \mathbb{E}\left[\left(1+8\frac{k_{\mathsf{R}}}{n}\chi^2(p||q)\mathsf{B}'\right)^{k_{\mathsf{R}}}\right] && (44)\\
&\qquad \leq \mathbb{E}\left[\exp\left(8\frac{k_{\mathsf{R}}^2}{n}\chi^2(p||q)\mathsf{B}'\right)\right] && (45)\\
&\qquad = \left[1+2\frac{k_{\mathsf{L}}}{n}\left(e^{8\frac{k_{\mathsf{R}}^2}{n}\chi^2(p||q)}-1\right)\right]^{k_{\mathsf{L}}}, && (46)
\end{aligned}$$

其中，第二个不等式是因为 $2\chi^2(\mathrm{p}||\mathrm{q})\mathsf{B}' \leq 1$，因此 $\left(2\chi^2(\mathrm{p}||\mathrm{q})\mathsf{B}'\right)^2 \leq 2\chi^2(\mathrm{p}||\mathrm{q})\mathsf{B}'$，且该结论以概率 1 成立；第三个不等式是由于对任意 $\mathrm{x} \geq 0$，都有 $1+\mathrm{x} \leq \exp(\mathrm{x})$；最后一个等式则用到了二项随机变量 B' 的矩生成函数。与上一小节一样，利用不等式 $2\,1\mathrm{kL} - 1 \geq \underline{\log}\, 2\,\mathrm{k}_{\mathsf{L}}$，可知 (46) 以及 (34) 均以上界 2 为界，当

$$\chi^2(p||q) \leq \frac{n}{8k_{\mathsf{R}}^2}\cdot\log\left(1+\frac{n}{k_{\mathsf{L}}^2}\frac{\log 2}{2}\right). \tag{47}$$

结合我们的假设 $2\chi^2(p||q)k_{\mathsf{L}} \leq 1$，我们得到 (34) 以上界 2 为界，当

$$\chi^2(p||q) \leq \min\left\{\frac{1}{2k_{\mathsf{L}}}, \frac{n}{8k_{\mathsf{R}}^2}\cdot\log\left(1+\frac{n}{k_{\mathsf{L}}^2}\frac{\log 2}{2}\right)\right\}. \tag{48}$$

如果重复上述相同的论证，但交换B 和B'的角色，并假设 $2\chi^2(p||q)k_{\mathsf{R}} \leq 1$，那么我们将得到 (34) 以上界 2 为界，当

$$\chi^2(p||q) \leq \min\left\{\frac{1}{2k_{\mathsf{R}}}, \frac{n}{8k_{\mathsf{L}}^2}\cdot\log\left(1+\frac{n}{k_{\mathsf{R}}^2}\frac{\log 2}{2}\right)\right\}, \tag{49}$$

因此，结合起来，当34成立时，48或49中任一条件成立时，5c恰好满足式（5c）所述的条件，其上界为 2。

最后，我们对（35）给出上界。重复上述后续方法中的完全相同论证，可以证明（35）右侧有上界 2

if,

$$\chi^2(p||q) \leq \frac{n}{8k_{\mathsf{R}}k_{\mathsf{L}}} \cdot \log\left(1 + \frac{n}{k_{\mathsf{R}}k_{\mathsf{L}}}\frac{\log 2}{2}\right), \tag{50}$$

这正是我们在（5a）中陈述的条件。证明至此完成。

4.2 定理2的证明

在本小节中，我们通过对第2 节提出的三种检验各自相关的风险进行上界估计，来证明定理3，首先从扫描检验开始。

4.2.1 扫描检验

在零假设下，对于任意$(\mathcal{R}', \mathcal{L}') \in \mathcal{S}_{k_{\mathsf{R}},k_{\mathsf{L}}}$，我们有$\sum_{i\in\mathcal{R}',j\in\mathcal{L}'} \mathsf{A}^{ij} \sim$二项分布$(k_{\mathsf{R}}k_{\mathsf{L}},q)$。回顾式（6）中 Scan(A) 的定义，并注意到式（6）中的最大值是在大小为$|\mathcal{S}_{k_{\mathsf{R}},k_{\mathsf{L}}}| = \binom{n}{k_{\mathsf{R}}}\binom{n-k_{\mathsf{R}}}{k_{\mathsf{L}}}$的集合上取的，可将其上界估计如下，

$$\binom{n}{k_{\mathsf{R}}}\binom{n-k_{\mathsf{R}}}{k_{\mathsf{L}}} \leq \binom{n}{k_{\mathsf{R}}}\binom{n}{k_{\mathsf{L}}} \leq \left(\frac{en}{k_{\mathsf{R}}}\right)^{k_{\mathsf{R}}}\left(\frac{en}{k_{\mathsf{L}}}\right)^{k_{\mathsf{L}}} \leq \left(\frac{en}{k_{\mathsf{R}}\vee k_{\mathsf{L}}}\right)^{2k_{\mathsf{R}}\vee k_{\mathsf{L}}}, \tag{51}$$

其中第二个不等式由$(y/x)^x \leq \binom{y}{x} \leq (ey/x)^x$得出，最后一个不等式是因为当$a = o(n)$时，$(n/a)^a$ 作为a 的函数单调不减。因此，根据并集界和伯恩斯坦不等式，我们有，

$$\begin{aligned}
\mathbb{P}_{\mathcal{H}_0}(\phi_{\mathsf{Scan}}(\mathsf{A}) = 1) &= \mathbb{P}_{\mathcal{H}_0}(\mathsf{Scan}(\mathsf{A}) \geq \tau_{\mathsf{Scan}}) && (52)\\
&\leq \sum_{(\mathcal{R}',\mathcal{L}')\in\mathcal{S}_{k_{\mathsf{R}},k_{\mathsf{L}}}} \mathbb{P}_{\mathcal{H}_0}\left[\sum_{i\in\mathcal{R}',j\in\mathcal{L}'} \mathsf{A}_{ij} \geq \tau_{\mathsf{Scan}}\right] && (53)\\
&\leq \left(\frac{en}{k_{\mathsf{R}}\vee k_{\mathsf{L}}}\right)^{2k_{\mathsf{R}}\vee k_{\mathsf{L}}} \cdot \exp\left(-\frac{k_{\mathsf{R}}^2k_{\mathsf{L}}^2\cdot(p-q)^2/4}{2k_{\mathsf{R}}k_{\mathsf{L}}q + k_{\mathsf{R}}k_{\mathsf{L}}\cdot(p-q)/3}\right) && (54)\\
&\leq \exp\left(k_{\mathsf{R}}\vee k_{\mathsf{L}}\cdot\log\left(\frac{en}{k_{\mathsf{R}}\vee k_{\mathsf{L}}}\right) - C_1\cdot k_{\mathsf{R}}k_{\mathsf{L}}\cdot\frac{(p-q)^2}{q(1-q)}\right), && (55)
\end{aligned}$$

其中$C_1 > 0$，最后一个不等式由以下假设推出：$|p-q| = O(q)$且$q < c < 1$，其中某个常数为 $0 < c < 1$。在备择假设下，若$\mathsf{K}^{\star}_{\mathcal{R},\mathcal{L}}$表示底层的植入二分子图，则存在右侧和左侧顶点集合$(\mathcal{R}^*, \mathcal{L}^*)$，使得$\sum_{i\in\mathcal{R}^*,\ j\in\mathcal{L}^*} \mathsf{A}^{ij} \sim$二项分布$(k_{\mathsf{R}}k_{\mathsf{L}}, p)$。因此，根据乘法型切尔诺夫界，我们有，

$$\begin{aligned}
\mathbb{P}_{\mathcal{H}_1}(\phi_{\mathsf{Scan}}(\mathsf{A}) = 0) &= \mathbb{P}_{\mathcal{H}_1}(\mathsf{Scan}(\mathsf{A}) \leq \tau_{\mathsf{Scan}}) && (56)\\
&\leq \mathbb{P}_{\mathcal{H}_1}\left[\sum_{i\in\mathcal{R}^*,j\in\mathcal{L}^*} \mathsf{A}_{ij} \leq \tau_{\mathsf{Scan}}\right] && (57)
\end{aligned}$$

$$\leq \exp\left(-\frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2 \cdot (p-q)^2/4}{2k_{\mathsf{R}}k_{\mathsf{L}}p}\right) \tag{58}$$

$$\leq \exp\left(-C_2 \cdot k_{\mathsf{R}}k_{\mathsf{L}} \cdot \frac{(p-q)^2}{q\,(1-q)}\right), \tag{59}$$

对于某个常数 $C_2 > 0$，同样，最后一个不等式由以下假设推出：$|p - q| = O(q)$ 且 $q < c < 1$。因此，由式（55）和（59）可知，存在一个常数 $C_3 > 0$使得，

$$\mathsf{R}_n(\phi_{\mathsf{Scan}}) \leq 2 \cdot \exp\left(k_{\mathsf{R}} \vee k_{\mathsf{L}} \cdot \log\left(\frac{n}{k_{\mathsf{R}} \vee k_{\mathsf{L}}}\right) - C_3 \cdot k_{\mathsf{R}}k_{\mathsf{L}} \cdot \frac{(p-q)^2}{q\,(1-q)}\right), \tag{60}$$

其小于 δ，前提是，

$$\chi^2(p||q) = \frac{(p-q)^2}{q\,(1-q)} \geq C\left[\frac{k_{\mathsf{R}} \vee k_{\mathsf{L}}}{k_{\mathsf{R}}k_{\mathsf{L}}} \cdot \log\left(\frac{n}{k_{\mathsf{R}} \vee k_{\mathsf{L}}}\right) + \frac{\log\frac{2}{\delta}}{k_{\mathsf{R}}k_{\mathsf{L}}}\right] \tag{61}$$

$$= C\left[\frac{\log\left(\frac{n}{k_{\mathsf{R}} \vee k_{\mathsf{L}}}\right) + \frac{\log\frac{2}{\delta}}{k_{\mathsf{R}} \vee k_{\mathsf{L}}}}{k_{\mathsf{R}} \wedge k_{\mathsf{L}}}\right], \tag{62}$$

对于某个 $C > 0$，证明即告完成。

4.2.2 计数检验

回顾（7）中的计数统计量。在零假设下，显然 $\mathsf{Count}(\mathsf{A})$ ~服从二项分布$\left(\binom{n}{2}, q\right)$。因此，根据伯恩斯坦不等式，

$$\mathbb{P}_{\mathcal{H}_0}(\phi_{\mathsf{Count}}(\mathsf{A}) = 1) = \mathbb{P}_{\mathcal{H}_0}(\mathsf{Count}(\mathsf{A}) \geq \tau_{\mathsf{Count}}) \tag{63}$$

$$\leq \exp\left(-\frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2 \cdot (p-q)^2/4}{2\binom{n}{2}q + k_{\mathsf{R}}k_{\mathsf{L}} \cdot (p-q)/3}\right) \tag{64}$$

$$\leq \exp\left(-C_1 \cdot \frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2}{n^2} \cdot \frac{(p-q)^2}{q\,(1-q)}\right), \tag{65}$$

其中 $C_1 > 0$，最后一个不等式由以下假设推出：$|p - q| = O(q)$且 $q < c < 1$，其中某个常数为 $0 < c < 1$。在备择假设下，随机变量 $\mathsf{计数}(\mathsf{A})$ 服从二项分布$(k_{\mathsf{R}}k_{\mathsf{L}}, p)$ 与二项分布$\left(\binom{n}{2} - k_{\mathsf{R}}k_{\mathsf{L}}, q\right)$的独立和。因此，根据乘法型切尔诺夫界，

$$\mathbb{P}_{\mathcal{H}_1}(\phi_{\mathsf{Count}}(\mathsf{A}) = 0) = \mathbb{P}_{\mathcal{H}_1}(\mathsf{Count}(\mathsf{A}) \leq \tau_{\mathsf{Count}}) \tag{66}$$

$$\leq \exp\left(-\frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2 \cdot (p-q)^2/4}{2\binom{n}{2}q + 2 \cdot k_{\mathsf{R}}k_{\mathsf{L}} \cdot (p-q)}\right) \tag{67}$$

$$\leq \exp\left(-C_2 \cdot \frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2}{n^2} \cdot \frac{(p-q)^2}{q(1-q)}\right), \tag{68}$$

其中 $C_2 > 0$。因此，结合式（65）和（68），可得存在一个常数$C_3 > 0$，使得：

$$\mathsf{R}_n(\phi_{\mathsf{Count}}) \leq 2 \cdot \exp\left(-C_3 \cdot \frac{k_{\mathsf{R}}^2 k_{\mathsf{L}}^2}{n^2} \cdot \frac{(p-q)^2}{q(1-q)}\right). \tag{69}$$

小于 δ，前提是，

$$\chi^2(p||q) = \frac{(p-q)^2}{q(1-q)} \geq \frac{n^2}{C_3 \cdot k_{\mathsf{R}}^2 k_{\mathsf{L}}^2} \cdot \log\frac{2}{\delta}, \tag{70}$$

这就完成了证明。

4.2.3 度检验

回顾（8）中的度统计量。在零假设下，显然有 $\sum_j \mathsf{A}^{ij}$ ~服从二项分布$(n-1, q)$。因此，根据并集界和伯恩斯坦不等式，我们得到：

$$\begin{aligned}
\mathbb{P}_{\mathcal{H}_0}(\phi_{\mathsf{Deg}}(\mathsf{A}) = 1) &= \mathbb{P}_{\mathcal{H}_0}(\mathsf{MaxDeg}(\mathsf{A}) \geq \tau_{\mathsf{Deg}}) && (71)\\
&\leq n \cdot \exp\left(-\frac{(k_{\mathsf{R}} \vee k_{\mathsf{L}})^2 \cdot (p-q)^2/4}{2(n-1)q + k_{\mathsf{R}} \vee k_{\mathsf{L}} \cdot (p-q)/3}\right) && (72)\\
&\leq \exp\left(\log n - C_1 \cdot \frac{(k_{\mathsf{R}} \vee k_{\mathsf{L}})^2}{n} \cdot \frac{(p-q)^2}{q(1-q)}\right), && (73)
\end{aligned}$$

其中 $C_1 > 0$；最后一个不等式由以下假设推出：$|p-q| = O(q)$以及 $q < c < 1$，其中 $0 < c < 1$。在备择假设下，至少存在一行 i^*，使得 $\sum_j \mathsf{A}^{i^*j}$ 可以表示为二项分布$(k_{\mathsf{R}} \vee k_{\mathsf{L}}, p)$与二项分布$(n-1-k_{\mathsf{R}} \vee k_{\mathsf{L}}, q)$之和，且二者相互独立。因此，根据乘法型切尔诺夫界，我们得到：

$$\begin{aligned}
\mathbb{P}_{\mathcal{H}_1}(\phi_{\mathsf{Deg}}(\mathsf{A}) = 0) &= \mathbb{P}_{\mathcal{H}_1}(\mathsf{MaxDeg}(\mathsf{A}) \leq \tau_{\mathsf{Deg}}) && (74)\\
&\leq \exp\left(-\frac{(k_{\mathsf{R}} \vee k_{\mathsf{L}})^2 \cdot (p-q)^2/4}{2(n-1)q + 2 \cdot k_{\mathsf{R}} \vee k_{\mathsf{L}} \cdot (p-q)}\right) && (75)\\
&\leq \exp\left(-C_2 \cdot \frac{(k_{\mathsf{R}} \vee k_{\mathsf{L}})^2}{n} \cdot \frac{(p-q)^2}{q(1-q)}\right), && (76)
\end{aligned}$$

对于某个常数 $C_2 > 0$。因此，结合（73）和（76），我们得到：

$$\mathsf{R}_n(\phi_{\mathsf{Deg}}) \leq 2 \cdot \exp\left(\log n - C \cdot \frac{(k_{\mathsf{R}} \vee k_{\mathsf{L}})^2}{n} \cdot \frac{(p-q)^2}{q(1-q)}\right), \tag{77}$$

对于某个常数 $C > 0$。只要满足以下条件，这就小于 δ：

$$\chi^2(p||q) = \frac{(p-q)^2}{q(1-q)} \geq \frac{n}{C \cdot (k_{\mathsf{R}} \vee k_{\mathsf{L}})^2} \cdot \left(\log n + \log \frac{2}{\delta} \right), \tag{78}$$

由此完成证明。

4.3 定理3的证明

让我们将似然比 L_n 按关于 $\mathbb{P}_{\mathcal{H}_0}$ 的正交多项式基展开。具体而言，设 $f_0, f_1, \ldots, f_m$：$\Omega^n \to \mathbb{R}$ 是坐标度数为 D 的函数的一个标准正交基（关于 $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$），并且 f_0 是单位常数函数。因此，$\langle f_i, f_j \rangle_{\mathcal{H}_0} = \delta_{ij}$，其中当 $\delta_{ij} = 0$ if $i \neq j$ 时成立，且 $\delta_{ii} = 1$。接着，用这个基来计算 $\mathsf{L}_{n,\leq \mathsf{D}}$ 的范数，可得

$$\|\mathsf{L}_{n,\leq \mathsf{D}}\|^2_{\mathcal{H}_0} = \sum_{1 \leq i \leq m} \langle f_i, \mathsf{L}_{n,\leq \mathsf{D}} \rangle^2_{\mathcal{H}_0} \tag{79}$$

$$= \sum_{1 \leq i \leq m} \left[\mathbb{E}_{\mathcal{H}_0} \left[\mathsf{L}_n(\mathsf{G}) f_i(\mathsf{G}) \right] \right]^2 \tag{80}$$

$$= \sum_{1 \leq i \leq m} \left[\mathbb{E}_{\mathcal{H}_1} f_i(\mathsf{G}) \right]^2, \tag{81}$$

这里我们用到了 $(\mathsf{L}_n - \mathsf{L}_{n,\leq \mathsf{D}})$ 与 $\{f_i\}_{i=0}^m$ 正交这一事实。因此，我们需要计算 $\mathbb{E}_{\mathcal{H}_1} f_i(\mathsf{G})$，其中 f_i 是某个标准正交基函数。在我们的设定中，对于 $\alpha \subseteq \binom{[n]}{2}$，定义傅里叶特征

$$\chi_\alpha(\mathsf{G}) = \prod_{\{i,j\} \in \alpha} \frac{\mathsf{G}_{ij} - q}{\sqrt{q(1-q)}}, \tag{82}$$

对于每个 $\mathsf{G} \in \{0,1\}^{\binom{n}{2}}$。注意，任意子集 $\alpha \subseteq \binom{[n]}{2}$ 都可以与顶点集为 $[\mathrm{n}]$、由 α 中的边诱导出的图对应。因此，我们可以毫无歧义地称其为“图 α”。接着，$|\alpha|$ 表示图 α 中边的数量。此外，略微滥用记号，令 $\mathrm{V}(\alpha)$ 表示 α 的顶点集，即满足 $\mathrm{v} \in [\mathrm{n}]$ 且不被 α 的边孤立的顶点所组成的集合；最后，令 $\mathrm{v}(\alpha)$ 表示其大小，即 $\mathrm{v}(\alpha) = |\mathrm{V}(\alpha)|$。已知 [65]，$\{\chi_\alpha\}_{\alpha \subseteq \binom{[n]}{2}}$、$|\alpha| \leq \mathrm{D}$ 构成关于 $\mathbb{P}_{\mathcal{H}_0}$ 的 D 度函数的标准正交基。根据（81），下面我们计算每个此类 α 的 $\mathbb{E}_{\mathcal{H}_1} \chi_\alpha(\mathsf{G})$。我们有：

$$\mathbb{E}_{\mathcal{H}_1} \chi_\alpha(\mathsf{G}) = \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}} \mathbb{E}_{\mathcal{H}_1 | \mathsf{K}_{\mathcal{R},\mathcal{L}}} \chi_\alpha(\mathsf{G}) \tag{83}$$

$$= \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}} \mathbb{E}_{\mathcal{H}_1 | \mathsf{K}_{\mathcal{R},\mathcal{L}}} \left(\prod_{\{i,j\} \in \alpha} \frac{\mathsf{G}_{ij} - q}{\sqrt{q(1-q)}} \right) \tag{84}$$

$$= \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}} \prod_{\{i,j\} \in \alpha} \mathbb{E}_{\mathcal{H}_1 | \mathsf{K}_{\mathcal{R},\mathcal{L}}} \left(\frac{\mathsf{G}_{ij} - q}{\sqrt{q(1-q)}} \right), \tag{85}$$

这里使用了如下事实：在给定植入结构 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$ 的条件下，G 的边彼此独立。根据边 $\{i、j\}$ 及其与 α 和 G 的关系，存在两种可能的情况：

- 若 $\{i,j\} \in \alpha$ 满足 $i,j \notin V(\mathsf{K}_{\mathcal{R},\mathcal{L}})$，则

$$\mathbb{E}\left[\frac{\mathsf{G}_{ij}-q}{\sqrt{q(1-q)}}\middle| i,j \notin V(\mathsf{K}_{\mathcal{R},\mathcal{L}})\right] = 0, \tag{86}$$

  因为若 i 或 j 不属于 $\mathsf{K}_{\mathcal{R},\mathcal{L}}$，则边 $\{i,j\}$ 以概率 q 被纳入 G。

- 若 $\{i,j\} \in \alpha$ 满足 $i,j \in V(\mathsf{K}_{\mathcal{R},\mathcal{L}})$ 且 $\{i,j\} \in E(\mathsf{K}_{\mathcal{R},\mathcal{L}})$，则

$$\begin{aligned}\mathbb{E}\left[\frac{\mathsf{G}_{ij}-q}{\sqrt{q(1-q)}}\middle| i,j \in V(\mathsf{K}_{\mathcal{R},\mathcal{L}}), \{i,j\} \in E(\mathsf{K}_{\mathcal{R},\mathcal{L}})\right] &= p\sqrt{\frac{1-q}{q}} - (1-p)\sqrt{\frac{q}{1-q}} \\ &= \sqrt{\chi^2(p||q)}. \end{aligned} \tag{87}$$

- 若 $\{i,j\} \in \alpha$ 满足 $i,j \in V(\mathsf{K}_{\mathcal{R},\mathcal{L}})$ 且 $\{i,j\} \notin E(\mathsf{K}_{\mathcal{R},\mathcal{L}})$，则

$$\mathbb{E}\left[\frac{\mathsf{G}_{ij}-q}{\sqrt{q(1-q)}}\middle| i,j \in V(\mathsf{K}_{\mathcal{R},\mathcal{L}}), \{i,j\} \notin E(\mathsf{K}_{\mathcal{R},\mathcal{L}})\right] = 0. \tag{88}$$

因此，我们得到

$$\mathbb{E}_{\mathcal{H}_1|\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left(\frac{\mathsf{G}_{ij}-q}{\sqrt{q(1-q)}}\right) = \sqrt{\chi^2(p||q)} \cdot {}_{\mathbb{I}}\left\{\{i,j\} \in \alpha \cap E(\mathsf{K}_{\mathcal{R},\mathcal{L}})\right\}. \tag{89}$$

这又意味着

$$\prod_{\{i,j\}\in\alpha} \mathbb{E}_{\mathcal{H}_1|\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left(\frac{\mathsf{G}_{ij}-q}{\sqrt{q(1-q)}}\right) = (\chi^2(p||q))^{\frac{1}{2}|\alpha\cap\mathsf{K}_{\mathcal{R},\mathcal{L}}|} {}_{\mathbb{I}}\left\{\alpha \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}\right\}. \tag{90}$$

综合以上可得，

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_1}\chi_\alpha(\mathsf{G}) &= \mathbb{E}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left[(\chi^2(p||q))^{\frac{1}{2}|\alpha\cap\mathsf{K}_{\mathcal{R},\mathcal{L}}|} {}_{\mathbb{I}}\left\{\alpha \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}\right\}\right] &(91)\\ &= [\chi^2(p||q)]^{\frac{1}{2}|\alpha|}\mathbb{P}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left[\alpha \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}\right]. &(92)\end{aligned}$$

因此，由（81）可得

$$\|\mathsf{L}_{n,\le\mathsf{D}}\|^2_{\mathcal{H}_0} = 1 + \sum_{0<|\alpha|\le\mathsf{D}} [\chi^2(p||q)]^{|\alpha|}\mathbb{P}^2_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left[\alpha \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}\right], \tag{93}$$

其中，求和项 1 对应于空图 $\alpha = \emptyset$。现在，这里的主要观察是，如果 α 本身不是二分图，那么 $\mathbb{P}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}[\alpha \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}] = 0$。这是因为

二分图的任意子图本身也是二分图。下面可以很容易地看出这一点：众所周知，一个图是二分图，当且仅当它不包含奇长度环。由于二分图的子图不可能新增奇长度环，因此它也必定是二分图。因此，(93) 中对 α 的求和可以仅限于二分图。相应地，令 $\mathcal{B}_{\mathsf{D}}$ 表示完全图中所有可能的二分图子图的集合，该完全图有n 个顶点，边数至多为 D，且至少有一条边（因为$|\alpha| > 0$）。为清楚起见，我们将使用记号 α_{Bip} 代替 α，并令$v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})$ 和$v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})$ 分别表示 α_{Bip} 的左侧顶点数和右侧顶点数。因此，

$$\|\mathsf{L}_{n,\leq \mathsf{D}}\|_{\mathcal{H}_0}^2 = 1 + \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}} [\chi^2(p||q)]^{|\alpha_{\mathsf{Bip}}|} \mathbb{P}^2_{\mathsf{K}_{\mathcal{R},\mathcal{L}}} [\alpha_{\mathsf{Bip}} \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}] . \tag{94}$$

接下来，我们分别计算（94）中的求和，讨论稠密区域和稀疏区域，并从前者开始。这里需要指出，事实上，我们对稀疏区域的分析可以推广，使稠密区域成为其一个特例。尽管如此，我们仍选择分别分析这两种情形，因为我们用于稠密情形的界定技术更简单，也可能适用于其他相关问题。

4.3.1 稠密区域

考虑$p,q = \Theta(1)$的情形，此时 $\chi^2(p||q) = \Theta(1)$。我们的目标是证明，若$k_{\mathsf{R}} \vee k_{\mathsf{L}} \ll \sqrt{n}$，则 $\|\mathbb{L}_{n,\leq \mathsf{D}}\|_{\mathcal{H}_0}^2$ 有界。$\bar{}$为简化记号，我们记为$\lambda \triangleq \chi^2(p||q)$，并且不失一般性地假设$k_{\mathsf{R}} \geq k_{\mathsf{L}}$。于是，

$$\mathbb{P}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}} (\alpha_{\mathsf{Bip}} \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}) = \frac{\binom{n-v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})}{k_{\mathsf{R}}-v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})}\binom{n-k_{\mathsf{R}}-v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})}{k_{\mathsf{L}}-v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})}}{\binom{n}{k_{\mathsf{R}}}\binom{n-k_{\mathsf{R}}}{k_{\mathsf{L}}}} \tag{95}$$

$$= \frac{\binom{k_{\mathsf{R}}}{v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})}\binom{k_{\mathsf{L}}}{v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})}}{\binom{n}{v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})}\binom{n-k_{\mathsf{R}}}{v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})}} \tag{96}$$

$$\leq \left(\frac{ek_{\mathsf{R}}}{n}\right)^{v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})} \left(\frac{ek_{\mathsf{L}}}{n-k_{\mathsf{R}}}\right)^{v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})} \tag{97}$$

$$\leq \left(\frac{ek_{\mathsf{R}}}{n}\right)^{v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})} \left(\frac{ek_{\mathsf{R}}}{n-k_{\mathsf{R}}}\right)^{v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})} \tag{98}$$

$$= \left(\frac{ek_{\mathsf{R}}}{n}\right)^{v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})+v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})} \left(\frac{n}{n-k_{\mathsf{R}}}\right)^{v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})} \tag{99}$$

$$\leq \left(\frac{ek_{\mathsf{R}}}{n}\right)^{v(\alpha_{\mathsf{Bip}})} \exp\left(v_{\mathsf{L}}(\alpha_{\mathsf{Bip}}) \cdot \frac{k_{\mathsf{R}}}{n-k_{\mathsf{R}}}\right), \tag{100}$$

其中，第一个不等式由 $(y/x)^x \leq \binom{y}{x} \leq (ey/x)^x$ 得出；第二个不等式是因为 $k_{\mathsf{R}} \geq k_{\mathsf{L}}$；在最后一个不等式中，我们使用了对于 $x \geq 0$，$1 + x \leq \exp(x)$，并且 $v(\alpha_{\mathsf{Bip}}) = |V(\alpha_{\mathsf{Bip}})|$ 是 α_{Bip} 中非孤立顶点的总数。现在请注意，对于任意 $\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}$，必有 $v_{\mathsf{L}}(\alpha_{\mathsf{Bip}}) \leq \mathsf{D}$。因此，在 $k_{\mathsf{R}} \vee k_{\mathsf{L}} \ll$ 的情形下

$\sqrt{n}$，由于 $\mathsf{D} = n^{o(1)}$，我们有 $\exp\left(v_{\mathsf{L}}(\alpha_{\mathsf{Bip}}) \cdot \frac{k_{\mathsf{R}} n - k_{\mathsf{R}}}{}\right) = 1 + o(1)$，当 n 足够大时。因此，我们得到

$$\mathbb{P}_{\mathsf{K}_{\mathcal{R},\mathcal{L}}}\left(\alpha_{\mathsf{Bip}} \subseteq \mathsf{K}_{\mathcal{R},\mathcal{L}}\right) \le (1+o(1)) \cdot \left(\frac{e k_{\mathsf{R}}}{n}\right)^{v(\alpha_{\mathsf{Bip}})} \triangleq g(|V(\alpha_{\mathsf{Bip}})|). \tag{101}$$

因此，利用（94）可得

$$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 \le 1 + \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}} \lambda^{|\alpha_{\mathsf{Bip}}|} g^2(|V(\alpha_{\mathsf{Bip}})|). \tag{102}$$

分别分析 $\lambda > 1$和 $\lambda \le 1$ 这两种情况较为方便，我们先讨论前者。对于任意集合 $\mathcal{V} \subseteq [n]$，，有

$$\begin{aligned}
\sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}: V_{\alpha_{\mathsf{Bip}}} \supseteq \mathcal{V}} \lambda^{|\alpha_{\mathsf{Bip}}|} g^2(|V(\alpha_{\mathsf{Bip}})|) &= g^2(|\mathcal{V}|) \cdot \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}:\, v(\alpha_{\mathsf{Bip}}) = \mathcal{V}} \lambda^{|\alpha_{\mathsf{Bip}}|} && (103)\\
&\le g^2(|\mathcal{V}|) \cdot \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}:\, \alpha_{\mathsf{Bip}} \subseteq \binom{\mathcal{V}}{2}} \lambda^{|\alpha_{\mathsf{Bip}}|} && (104)\\
&= g^2(|\mathcal{V}|) \cdot \sum_{\ell=1}^{\min\left(\mathsf{D}, \binom{|\mathcal{V}|}{2}\right)} \binom{\binom{|\mathcal{V}|}{2}}{\ell} \lambda^{\ell} && (105)\\
&\le g^2(|\mathcal{V}|) \cdot \sum_{\ell=1}^{\min\left(\mathsf{D}, \binom{|\mathcal{V}|}{2}\right)} \left[\binom{|\mathcal{V}|}{2} \lambda\right]^{\ell} && (106)\\
&\le g^2(|\mathcal{V}|) \left(1 + \binom{|\mathcal{V}|}{2} \lambda\right)^{\min\left(\mathsf{D}, \binom{|\mathcal{V}|}{2}\right)}, && (107)
\end{aligned}$$

其中，最后一个不等式用到了如下事实：$\sum_{i=0}^{k} \binom{n}{i} \le \sum_{i=0}^{k} n^i 1^{k-i} \le (1+n)^k$。接下来，每个满足 $|\alpha_{\mathsf{Bip}}| \le \mathsf{D}$ 的 α_{Bip} 都有 $|v(\alpha_{\mathsf{Bip}})| \le 2\mathsf{D}$。此外，存在 $\binom{n}{t} \le n^t$ 个集合 $\mathcal{V} \subseteq [n]$ 个集合 $\mathcal{V}$，使得 $|\mathcal{V}| = t$。因此，

$$\begin{aligned}
\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 &\le 1 + (1+o(1)) \cdot \sum_{t=2}^{2\mathsf{D}} (1+t)^2 \left(\frac{e k_{\mathsf{R}}}{n}\right)^{2t} n^t (1 + t^2\lambda/2)^{\min\left(\mathsf{D}, t^2/2\right)} && (108)\\
&\le 1 + (1+o(1)) \cdot \sum_{t=2}^{2\mathsf{D}} (1+t)^2 \left(\frac{e^2 k_{\mathsf{R}}^2}{n}\right)^{t} \left[t\sqrt{\lambda}\right]^{2\min\left(\mathsf{D}, t^2/2\right)} && (109)\\
&= 1 + (1+o(1)) \cdot \sum_{t \le \sqrt{2\mathsf{D}}} (1+t)^2 \left(\frac{e^2 k_{\mathsf{R}}^2}{n}\right)^{t} \left[t\sqrt{\lambda}\right]^{t^2}\\
&\quad + (1+o(1)) \cdot \sum_{\sqrt{2\mathsf{D}} < t < 2\mathsf{D}} (1+t)^2 \left(\frac{e^2 k_{\mathsf{R}}^2}{n}\right)^{t} \left[t\sqrt{\lambda}\right]^{2\mathsf{D}}, && (110)
\end{aligned}$$

其中，第二个不等式由以下事实得出： $\lambda > 1$ 因此 $t^2\lambda > 2$。将右侧第一个级数中的被加数（110）记为 T_t。然后，对第一项应用比值检验可得

$$\frac{\mathsf{T}_{t+1}}{\mathsf{T}_t} = \frac{(2+t)^2}{(1+t)^2}\frac{e^2k_{\mathsf{R}}^2}{n}\frac{(t+1)^{(t+1)^2}}{t^{t^2}}\lambda^{\frac{1}{2}+t} \tag{111}$$

$$\leq \frac{(2+t)^2}{(1+t)^2}\frac{e^2k_{\mathsf{R}}^2}{n}e^t(1+t)^{1+2t}\lambda^{\frac{1}{2}+t} \tag{112}$$

$$\leq (2+t)^{1+2t}\frac{e^2k_{\mathsf{R}}^2}{n}e^t\lambda^{\frac{1}{2}+t} \tag{113}$$

$$\leq (2+\sqrt{2\mathsf{D}})^{1+2\sqrt{2\mathsf{D}}}\frac{e^2k_{\mathsf{R}}^2}{n}e^{\sqrt{2\mathsf{D}}}\lambda^{\frac{1}{2}+\sqrt{2\mathsf{D}}}, \tag{114}$$

其中，第一个不等式用到了以下事实：$(1+x) \leq \exp(x)$，当 $x \geq 0$；第二个不等式是因为 $t \geq 2$；最后一个不等式是因为 $t \leq \sqrt{2\mathsf{D}}$，并且（113）关于 t 单调递增。因此，若，

$$\frac{e^2k_{\mathsf{R}}^2}{n}e^{\sqrt{2\mathsf{D}}}(2+\sqrt{2\mathsf{D}})^{2\sqrt{2\mathsf{D}}+1}\lambda^{\frac{1}{2}+\sqrt{2\mathsf{D}}} < c < 1, \tag{115}$$

对于某个 $0 < \mathsf{c} < 1$，则（110）右侧的级数有一个常数上界。注意，对于任意 $\mathsf{D} \leq \mathsf{C}\log\mathsf{n}$ 以及任意 $\mathsf{C} > 0$，都有 $\mathsf{e}^{\sqrt{2\mathsf{D}}}(2+\sqrt{2\mathsf{D}})^{2\sqrt{2\mathsf{D}}+1}\lambda^{\frac{1}{2}+\sqrt{2}\mathsf{D}} = \mathsf{n}^{o(1)}$，因此，该因子不会影响（115）左侧项的多项式尺度。据此不难验证，对于任意 $\epsilon > 0$，只要 $\mathsf{k}_{\mathsf{R}} \leq \mathsf{n}^{\frac{1}{2}-\epsilon}$，（115）就成立。接下来，同理，我们将（110）右侧第二项中的被加数记为 T'_t。然后，对第二项应用比值检验，得到

$$\frac{\mathsf{T}'_{t+1}}{\mathsf{T}'_t} = \frac{(2+t)^2}{(1+t)^2}\frac{e^2k_{\mathsf{R}}^2}{n}\frac{(1+t)^{2\mathsf{D}}}{t^{2\mathsf{D}}} \tag{116}$$

$$\leq \frac{(2+t)^2}{(1+t)^2}\frac{e^2k_{\mathsf{R}}^2}{n}e^{\frac{2\mathsf{D}}{t}} \tag{117}$$

$$\leq e^{\sqrt{2\mathsf{D}}}\frac{(2+\sqrt{2\mathsf{D}})^2}{(1+\sqrt{2\mathsf{D}})^2}\frac{e^2k_{\mathsf{R}}^2}{n}. \tag{118}$$

同样，容易验证，由于 $\mathsf{D} = \mathrm{O}(\log\mathsf{n})$，因此 $\mathsf{e}^{\sqrt{2\mathsf{D}}\frac{(2+\sqrt{2\mathsf{D}})^2}{(1+\sqrt{2}\mathsf{D})^2}} = \mathsf{n}^{o(1)}$；于是，如果 $\mathsf{k}_{\mathsf{R}} \leq \mathsf{n}^{\frac{1}{2}-\epsilon}$，则对于任意 $\epsilon > 0$，(118) 严格小于 1，这意味着 (110) 右端的第二个级数有一个常数上界。因此，要使 (110) 在 $\mathsf{n} \to \infty$ 时有界，我们需要 $\mathsf{k}_{\mathsf{R}} \ll \sqrt{\mathsf{n}}$，正如所述。回顾我们曾假设 $\mathsf{k}_{\mathsf{R}} \geq \mathsf{k}_{\mathsf{L}}$这一事实。若情况相反，则由对称性可知，当 $\mathsf{k}_{\mathsf{L}} \ll \sqrt{\mathsf{n}}$ 时，(110) 在 $\mathsf{n} \to \infty$ 时有界；因此，一般而言，$\|\mathbb{L}_{n,\leq\mathsf{D}}\|^2_{\mathcal{H}_0}$ 在 $\mathsf{k}_{\mathsf{R}} \vee \mathsf{k}_{\mathsf{L}} \ll \sqrt{\mathsf{n}}$ 时有界。最后，我们讨论 $\lambda \leq 1$ 的情形。在这种情况下，对于任意，我们进一步估计 $\lambda^{|\alpha_{\mathsf{Bip}}|} \leq 1$。

$\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}$，于是（102）变为：

$$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 \le \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}} g^2(|V(\alpha_{\mathsf{Bip}})|) \tag{119}$$

$$\le 1 + (1+o(1)) \cdot \sum_{t=2}^{2\mathsf{D}} (1+t)^2 \left(\frac{e \cdot k_{\mathsf{R}} \vee k_{\mathsf{L}}}{n}\right)^{2t} n^t. \tag{120}$$

使用与上文相同的论证，由比值检验可知，在相同条件下，$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2$ 有界，即$k_{\mathsf{R}} \vee k_{\mathsf{L}} \ll \sqrt{n}$。

4.3.2 稀疏情形

接下来，我们分析稀疏区域，其中 $\mathsf{p}, \mathsf{q} = \Theta(\mathsf{n}^{-\alpha})$，对于 $\alpha \in (0,2]$。不失一般性，假设 $\mathsf{k}_{\mathsf{R}} \ge \mathsf{k}_{\mathsf{L}}$。此外，回想一下，$\mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{L}} = o(\mathsf{n})$，因此 $\mathsf{k}_{\mathsf{R}}, \mathsf{k}_{\mathsf{L}} \le \mathsf{n}/2$。最后，我们可以假设 $\chi^2(\mathsf{p}||\mathsf{q}) < \frac{1}{2}\mathsf{D}^2$；事实上，在稀疏区域中有$\chi^2(\mathsf{p}||\mathsf{q}) = \Theta(\mathsf{n}^{-\alpha})$，因此显然当 n 足够大且任意 $\mathsf{D} = \mathsf{n}^{o(1)}$ 时，$2\mathsf{D}^2\chi^2(\mathsf{p}||\mathsf{q}) < 1$。

使用（94）和（97），我们有

$$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 \le 1 + \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\mathsf{D}}} [\chi^2(p||q)]^{|\alpha_{\mathsf{Bip}}|} \left(\frac{ek_{\mathsf{R}}}{n}\right)^{2v_{\mathsf{R}}(\alpha_{\mathsf{Bip}})} \left(\frac{ek_{\mathsf{L}}}{n-k_{\mathsf{R}}}\right)^{2v_{\mathsf{L}}(\alpha_{\mathsf{Bip}})}. \tag{121}$$

令 $\mathcal{B}_{\ell_1,\ell_2,m}$ 表示完全图 K_n 中所有边诱导二部子图构成的集合，这些子图有n 个顶点、ℓ_1 个右侧顶点、ℓ_2 个左侧顶点以及m 条边。形式上，

$$\mathcal{B}_{\ell_1,\ell_2,m} \triangleq \{\alpha_{\mathsf{Bip}} \subseteq \mathsf{K}_n : v_{\mathsf{R}}(\alpha_{\mathsf{Bip}}) = \ell_1,\ v_{\mathsf{L}}(\alpha_{\mathsf{Bip}}) = \ell_2,\ |\alpha_{\mathsf{Bip}}| = m\}. \tag{122}$$

注意，由于 $\mathcal{B}_{\ell_1}$、ℓ_2、m 中的图均由边诱导，且我们只考虑满足$|\alpha_{\mathsf{Bip}}| \le \mathsf{D}$ 的 α_{Bip}，因此有

$$1 \le \ell_1 \le \mathsf{D},\ 1 \le \ell_2 \le \mathsf{D},\ \ell_1 \vee \ell_2 \le m \le \mathsf{D}. \tag{123}$$

现在，利用以下事实

$$\mathcal{B}_{\mathsf{D}} = \bigcup_{1 \le \ell_1,\ell_2 \le \mathsf{D}} \bigcup_{\ell_1 \vee \ell_2 \le m \le \mathsf{D}} \mathcal{B}_{\ell_1,\ell_2,m}, \tag{124}$$

我们得到

$$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 \le 1 + \sum_{\ell_1,\ell_2=1}^{\mathsf{D}} \sum_{m=\ell_1\vee\ell_2}^{\mathsf{D}} \sum_{\alpha_{\mathsf{Bip}} \in \mathcal{B}_{\ell_1,\ell_2,m}} (\chi^2(p||q))^m \left(\frac{ek_{\mathsf{R}}}{n}\right)^{2\ell_1} \left(\frac{ek_{\mathsf{L}}}{n-k_{\mathsf{R}}}\right)^{2\ell_2} \tag{125}$$

$$= 1 + \sum_{\ell_1,\ell_2=1}^{\mathsf{D}} \sum_{m=\ell_1\vee\ell_2}^{\mathsf{D}} |\mathcal{B}_{\ell_1,\ell_2,m}| \cdot (\chi^2(p||q))^m \left(\frac{ek_{\mathsf{R}}}{n}\right)^{2\ell_1} \left(\frac{ek_{\mathsf{L}}}{n-k_{\mathsf{R}}}\right)^{2\ell_2}. \tag{126}$$

我们来给$|\mathcal{B}_{\ell_1、\ \ell_2、\ m}|$作上界估计。为此，我们首先枚举从 n 个可能的顶点中选择右侧顶点和左侧顶点的所有可能方式，然后枚举从$\ell_1 \cdot \ell_2$ 条可能的边中选择 m 条边的所有可能方式。具体而言，

$$
\begin{aligned}
|\mathcal{B}_{\ell_1,\ell_2,m}| &= \binom{n}{\ell_1}\binom{n-\ell_1}{\ell_2}\binom{\ell_1\cdot\ell_2}{m} && (127)\\
&\le n^{\ell_1+\ell_2}(\ell_1\cdot\ell_2)^m && (128)\\
&\le n^{\ell_1+\ell_2}\mathsf{D}^{2m}. && (129)
\end{aligned}
$$

因此，

$$
\begin{aligned}
\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 &\le 1+\sum_{\ell_1,\ell_2=1}^{\mathsf{D}}\sum_{m=\ell_1\vee\ell_2}^{\mathsf{D}}(\mathsf{D}^2\chi^2(p||q))^m\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}\left(\frac{e^2k_{\mathsf{L}}^2 n}{(n-k_{\mathsf{R}})^2}\right)^{\ell_2} && (130)\\
&\le 1+\sum_{\ell_1,\ell_2=1}^{\mathsf{D}}\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}\left(\frac{4e^2k_{\mathsf{L}}^2}{n}\right)^{\ell_2}\sum_{m=\ell_1\vee\ell_2}^{\mathsf{D}}(\mathsf{D}^2\chi^2(p||q))^m && (131)\\
&\le 1+2\sum_{\ell_1,\ell_2=1}^{\mathsf{D}}\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}\left(\frac{4e^2k_{\mathsf{L}}^2}{n}\right)^{\ell_2}(\mathsf{D}^2\chi^2(p||q))^{\ell_1\vee\ell_2} && (132)\\
&\ 1+2\cdot(\mathcal{W}_1+\mathcal{W}_2+\mathcal{W}_3), && (133)
\end{aligned}
$$

其中，第二个不等式用到了以下事实：$k_{\mathsf{R}} \le n/2$；第三个不等式由不等式 $\sum_{i=j}^{\infty}\beta^i = \frac{\beta^j}{1-\beta} \le 2\beta^j$ 得出，适用于 $\beta < 1/2$，同时还用到了以下事实：$\chi^2(p||q) < \frac{1}{2\ \mathsf{D}^2}$，；最后，在（133）中，我们定义了：

$$
\begin{aligned}
\mathcal{W}_1 &\triangleq \sum_{\ell=1}^{\mathsf{D}}\left(\frac{4e^4k_{\mathsf{R}}^2k_{\mathsf{L}}^2}{n^2}\right)^{\ell}(\mathsf{D}^2\chi^2(p||q))^{\ell}, && (134)\\
\mathcal{W}_2 &\triangleq \sum_{\ell_1>\ell_2}^{\mathsf{D}}\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}\left(\frac{4e^2k_{\mathsf{L}}^2}{n}\right)^{\ell_2}(\mathsf{D}^2\chi^2(p||q))^{\ell_1}, && (135)\\
\mathcal{W}_3 &\triangleq \sum_{\ell_2>\ell_1}^{\mathsf{D}}\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}\left(\frac{4e^2k_{\mathsf{L}}^2}{n}\right)^{\ell_2}(\mathsf{D}^2\chi^2(p||q))^{\ell_2}. && (136)
\end{aligned}
$$

接下来，我们找出使 $\mathcal{W}_1$、 $\mathcal{W}_2$ 和 $\mathcal{W}_3$ 同时有界的条件。利用 (133)，可知在相同条件下，$\|\mathsf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2$ 也有界。从 $\mathcal{W}_1$ 开始，只要 $4\mathsf{D}^2e^4k_{\mathsf{R}}^2k_{\mathsf{L}}^2n_2 \quad \chi^2(p||q) < c < 1$，其中 c 为某个常数，(134) 中的几何级数就有界。当$\chi^2(p||q) \ll n^2\, k_{\mathsf{L}}^2k_{\mathsf{R}}^2$ 时，显然满足这一条件。接下来，我们分析 $\mathcal{W}_2$。我们有，

$$
\mathcal{W}_2 = \sum_{\ell_1=1}^{\mathsf{D}}\left(\frac{e^2k_{\mathsf{R}}^2}{n}\right)^{\ell_1}(\mathsf{D}^2\chi^2(p||q))^{\ell_1}\sum_{\ell_2=1}^{\ell_1}\left(\frac{4e^2k_{\mathsf{L}}^2}{n}\right)^{\ell_2}. \tag{137}
$$

我们考虑两种情况。如果 $4e^2k_L^2n < 1$，那么我们利用以下事实：$\sum^j{}_{i=1}\beta^i \le j\cdot\beta$，对于 $\beta < 1$，并得到，

$$\mathcal{W}_2 \le \sum_{\ell_1=1}^{\mathsf{D}} \left(\frac{e^2k_R^2}{n}\right)^{\ell_1} (\mathsf{D}^2\chi^2(p||q))^{\ell_1}\ell_1\frac{4e^2k_L^2}{n} \tag{138}$$

$$\le \mathsf{D}\sum_{\ell_1=1}^{\mathsf{D}} \left(\frac{e^2k_R^2}{n}\right)^{\ell_1} (\mathsf{D}^2\chi^2(p||q))^{\ell_1}. \tag{139}$$

因此，只要对某个常数 $139\mathsf{D}^2e^2k_R^2n\,\chi^2(p||q) < c < 1$，式（$c$）中的几何级数有界。这在 $\chi^2(p||q) \ll n\,k_R^2$ 时成立。反过来，如果 $4e^2k_L^2n \ge 1$，那么利用 $\sum^j{}_{i=1}\beta^i \le j\cdot\beta^j$ 这一事实，其中 $\beta \ge 1$，可得，

$$\mathcal{W}_2 \le \sum_{\ell_1=1}^{\mathsf{D}} \left(\frac{e^2k_R^2}{n}\right)^{\ell_1} (\mathsf{D}^2\chi^2(p||q))^{\ell_1}\ell_1 \left(\frac{4e^2k_L^2}{n}\right)^{\ell_1} \tag{140}$$

$$\le \mathsf{D}\sum_{\ell_1=1}^{\mathsf{D}} \left(\frac{4e^4k_R^2k_L^2}{n^2}\right)^{\ell_1} (\mathsf{D}^2\chi^2(p||q))^{\ell_1}. \tag{141}$$

至于 $\mathcal{W}_1$，当 $\chi^2(\mathrm{p}||\mathrm{q}) \ll \mathrm{n}^2\,\mathrm{k}_L^2\mathrm{k}_R^2$ 时，式（141）中的级数有界。最后，我们完全按照分析 $\mathcal{W}_2$ 的方式来分析 $\mathcal{W}_3$。我们得到，若 $\mathrm{e}^2\mathrm{k}_R^2\mathrm{n} < 1$，则只要对某个常数 c，满足 $4\mathrm{D}^2\mathrm{e}^2\mathrm{k}_L^2\mathrm{n}\,\chi^2(\mathrm{p}||\mathrm{q}) < \mathrm{c} < 1$，$\mathcal{W}_3$ 就有界。显然，当 $\chi^2(\mathrm{p}||\mathrm{q}) \ll \mathrm{n}\,\mathrm{k}_L^2$ 时，这一条件成立；由于我们假设 $\mathrm{k}_R \ge \mathrm{k}_L$，该条件是冗余的，因为 $\mathrm{e}^2\mathrm{k}_R^2\mathrm{n} < 1$。反过来，若 $\mathrm{e}^2\mathrm{k}_R^2\mathrm{n} \ge 1$，则只要满足 $\chi^2(\mathrm{p}||\mathrm{q}) \ll \mathrm{n}^2\,\mathrm{k}_L^2\mathrm{k}_R^2$，$\mathcal{W}_3$ 就有界。同样，由于 $\mathrm{k}_R \ge \mathrm{k}_L$，与我们针对 $\mathcal{W}_2$ 得到的条件相比，该条件的限制更弱。综上可得，$\mathcal{W}_1$、$\mathcal{W}_2$ 和 $\mathcal{W}_3$ 均有界，当：

1. 若 $k_L^2 \ll n$，则我们需要 $\chi^2(p||q) \ll n^2\,k_L^2k_R^2$，并且 $\chi^2(p||q) \ll n\,k_R^2$。由于后者占主导，交集可简化为 $\chi^2(p||q) \ll n\,k_R^2$。

2. 若 $k_L^2 \gg n$，则我们需要 $\chi^2(p||q) \ll n^2\,k_L^2k_R^2$。

注意，合并上述两种情况后，我们得到条件 $\chi^2(p||q) \ll n^2\,k_L^2k_R^2 \wedge n\,k_R^2$；若去掉 $k_R \ge k_L$ 这一假设，则得到 $\chi^2(p||q) \ll n^2\,k_L^2k_R^2 \wedge n\,k_R^2 \vee k_L^2$。证明至此完成。

4.4 反向结论

在本小节中，我们证明定理 3 的反向结论部分，即在稠密和稀疏区域中，若 $k_R^2 \vee k_L^2 \gg n$ 且 $\chi^2(p||q) \gg n^2\,k_R^2k_L^2 \wedge n\,k_R^2 \vee k_L^2$，则分别有 $\|\mathbf{L}_{n,\le\mathsf{D}}\|_{\mathcal{H}_0} \ge \omega(1)$，

分别如此。回顾一下，我们假设 k_R、$k_L = o(n)$，因此 k_R、$k_L \le n/2$。接下来，我们对 (94) 中的 $\|L_{n,\le D}\|^2_{\mathcal{H}_0}$ 给出下界。利用 (96) 和不等式 $(y/x)^x \le \binom{y}{x} \le (ey/x)^x$，可得：

$$\mathbb{P}_{K_{\mathcal{R},\mathcal{L}}}[\alpha_{Bip} \subseteq K_{\mathcal{R},\mathcal{L}}] = \frac{\binom{k_R}{v_R(\alpha_{Bip})}\binom{k_L}{v_L(\alpha_{Bip})}}{\binom{n}{v_R(\alpha_{Bip})}\binom{n-k_R}{v_L(\alpha_{Bip})}} \tag{142}$$

$$\ge \left(\frac{k_R}{en}\right)^{v_R(\alpha_{Bip})} \left(\frac{k_L}{e(n-k_R)}\right)^{v_L(\alpha_{Bip})}, \tag{143}$$

for 任意 $\alpha_{Bip} \in \mathcal{B}_D$。然后，采用导出（133）时的类似论证，我们得到

$$\|L_{n,\le D}\|^2_{\mathcal{H}_0} \ge \sum_{\ell_1,\ell_2\ge 1} \sum_{m \ge \ell_1 \vee \ell_2} |\mathcal{B}_{\ell_1,\ell_2,m}| \cdot (\chi^2(p||q))^m \left(\frac{k_R}{en}\right)^{2\ell_1} \left(\frac{k_L}{e(n-k_R)}\right)^{2\ell_2} \tag{144}$$

$$\ge \sum_{\ell_1,\ell_2\ge 1} \sum_{m \ge \ell_1 \vee \ell_2} \left(\frac{n}{e\ell_1}\right)^{\ell_1} \left(\frac{n-\ell_1}{e\ell_2}\right)^{\ell_2} (\chi^2(p||q))^m \left(\frac{k_R}{en}\right)^{2\ell_1} \left(\frac{k_L}{e(n-k_R)}\right)^{2\ell_2} \tag{145}$$

$$\ge \sum_{\ell_1,\ell_2\ge 1} (\chi^2(p||q))^{\ell_1\vee\ell_2} \left(\frac{k_R^2}{e^3 n\ell_1}\right)^{\ell_1} \left(\frac{k_L^2(n-\ell_1)}{e^3(n-k_R)^2\ell_2}\right)^{\ell_2} \tag{146}$$

$$= \tilde{\mathcal{W}}_1 + \tilde{\mathcal{W}}_2 + \tilde{\mathcal{W}}_3 \tag{147}$$

$$\ge \tilde{\mathcal{W}}_1 \vee \tilde{\mathcal{W}}_2 \vee \tilde{\mathcal{W}}_3, \tag{148}$$

其中，第二个不等式使用了以下事实：对于任意 m，$|\mathcal{B}_{\ell_1,\ell_2,m}| \ge \binom{n}{\ell_1}\binom{n-\ell_1}{\ell_2} \ge \left(\frac{ne}{\ell_1}\right)^{\ell_1}\left(\frac{n-\ell_1}{e\ell_2}\right)^{\ell_2}$，并且我们定义

$$\tilde{\mathcal{W}}_1 = \sum_{\ell=1}^{D} (\chi^2(p||q))^{\ell} \left(\frac{k_R^2}{e^3 n\ell}\right)^{\ell} \left(\frac{k_L^2(n-\ell)}{e^3(n-k_R)^2\ell}\right)^{\ell}, \tag{149}$$

$$\tilde{\mathcal{W}}_2 = \sum_{\ell_1>\ell_2}^{D} (\chi^2(p||q))^{\ell_1} \left(\frac{k_R^2}{e^3 n\ell_1}\right)^{\ell_1} \left(\frac{k_L^2(n-\ell_1)}{e^3(n-k_R)^2\ell_2}\right)^{\ell_2}, \tag{150}$$

$$\tilde{\mathcal{W}}_3 = \sum_{\ell_2>\ell_1}^{D} (\chi^2(p||q))^{\ell_2} \left(\frac{k_R^2}{e^3 n\ell_1}\right)^{\ell_1} \left(\frac{k_L^2(n-\ell_1)}{e^3(n-k_R)^2\ell_2}\right)^{\ell_2}. \tag{151}$$

为证明反向结论，只需找出条件，使得（148）右侧的$\tilde{\mathcal{W}}_1$ 或 $\tilde{\mathcal{W}}_2$ 发散；这进而意味着，在相同条件下，$\|L_{n,\le D}\|^2_{\mathcal{H}_0}$ 也发散。先从 $\mathcal{W}_1$ 入手，将 $\mathcal{W}_1$下界估计为求和中的第一项（即对应索引 $\ell = 1$ 的项），得到$\tilde{\mathcal{W}}_1 \ge \chi^2(p||q)\, k_R^2 e^3 n\, k_L^2 (n-1)\, e^3 (n-k_R)_2 \ge \chi^2(p||q)\, k_R^2 e^3 n\, k_L^2 e^3 n$，其中第二个不等式由 $n-\ell_1 (n-k_R)^2 \ge \frac{1}{n}$ 得出，因为 $\ell_1 < n$。因此，若 $\chi^2(p||q) \gg n^2 k_L^2 k_R^2$，则 $\tilde{\mathcal{W}}_1 = \omega(1)$，这

意味着 $\|\mathbf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 = \omega(1)$。接下来，我们通过分析 $\tilde{\mathcal{W}}_2$ 证明，若 $\chi^2(p||q) \gg_n k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2$，则 $\|\mathbf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 = \omega(1)$ 也成立。我们有，

$$\tilde{\mathcal{W}}_2 = \sum_{\ell_1=1}^{\mathsf{D}} (\chi^2(p||q))^{\ell_1} \left(\frac{k_{\mathsf{R}}^2}{e^3 n \ell_1}\right)^{\ell_1} \sum_{\ell_2=1}^{\ell_1} \left(\frac{k_{\mathsf{L}}^2 (n-\ell_1)}{e^3 (n-k_{\mathsf{R}})^2 \ell_2}\right)^{\ell_2} \tag{152}$$

$$\ge \sum_{\ell_1=1}^{\mathsf{D}} (\chi^2(p||q))^{\ell_1} \left(\frac{k_{\mathsf{R}}^2}{e^3 n \ell_1}\right)^{\ell_1} \left(\frac{k_{\mathsf{L}}^2 (n-\ell_1)}{e^3 (n-k_{\mathsf{R}})^2}\right) \tag{153}$$

$$\ge \frac{k_{\mathsf{L}}^2}{e^3 n} \sum_{\ell_1=1}^{\mathsf{D}} (\chi^2(p||q))^{\ell_1} \left(\frac{k_{\mathsf{R}}^2}{e^3 n \ell_1}\right)^{\ell_1} \tag{154}$$

$$= \frac{k_{\mathsf{L}}^2}{e^3 n} \frac{\zeta \cdot (\zeta^{\mathsf{D}} - 1)}{\zeta - 1}, \tag{155}$$

其中 ζ、$\chi^2(\mathrm{p}||\mathrm{q})\ k_{\mathsf{R}}^2 \mathrm{De}^3 \mathrm{n}$。我们声称存在一个整数 D，使得如果 $\zeta \gg 1$（这等价于 $\chi^2(\mathrm{p}||\mathrm{q}) \gg \mathrm{n}\ k_{\mathsf{R}}^2$），那么 $\tilde{\mathcal{W}}_2 = \omega(1)$。首先，对于 $\zeta > 2^{1/\mathsf{D}}$，注意到 (155) 进一步以下界为 $k_{\mathsf{L}}^2 e^3 n \zeta^{\mathsf{D}}$。现在，如果 $k_{\mathsf{L}}^2 \gg \mathrm{n}$，那么显然 (155) 在$\zeta \gg 1$时发散。如果 $k_{\mathsf{L}}^2 \ll \mathrm{n}$，那么由于比值 $k_{\mathsf{L}}^2 \mathrm{n}$ 随 n 的下降速度至多为多项式级，而ζ 随 n 以多项式速度增长，因此必然存在一个整数 D，使得乘积 $k_{\mathsf{L}}^2 e^3 n \zeta^{\mathsf{D}}$ 发散。令 κ_n 为满足 $\kappa_n = \omega(1)$ 的任意序列。那么，任何满足以下条件的 D

$$\mathsf{D} \ge \frac{\log \kappa_n}{\log \xi} + \frac{\log \frac{2e^3 n}{k_{\mathsf{L}}^2}}{\log \xi}, \tag{156}$$

意味着 $\tilde{\mathcal{W}}_2 = \omega(1)$。如果 $\xi \gg 1$，则对于某个 $\epsilon > 0$，有 $\xi = \Theta(n^\epsilon)$。此外，由于 $k_{\mathsf{L}}^2 \ll n$，因此对于某个 $\eta > 0$，有 $\frac{2e^3 n k_{\mathsf{L}}^2}{} = \Theta(n^\eta)$。于是，对于任意 $\kappa_n = o(\log n)$，如果我们取任意满足以下条件的整数 D，

$$\mathsf{D} \ge \frac{\eta}{\epsilon} + o(1), \tag{157}$$

然后 $\tilde{\mathcal{W}}_2 = \omega(1)$。最后，通过互换 k_{R} 和 k_{L} 的角色，我们得到条件$\chi^2(p||q) \gg_n k_{\mathsf{L}}^2$；因此，结合我们为$\tilde{\mathcal{W}}_1$得到的充分条件，可知 $\|\mathbf{L}_{n,\le \mathsf{D}}\|_{\mathcal{H}_0}^2 = \omega(1)$ 当且仅当 $\chi^2(p||q) \gg n^2 k_{\mathsf{L}}^2 k_{\mathsf{R}}^2 \wedge n\, k_{\mathsf{R}}^2 \vee k_{\mathsf{L}}^2$，如前所述。

5 结论

在这项工作中，我们研究了在大型随机图中检测植入的二部结构这一问题。我们推导出了稠密和稀疏区域中渐近紧的上下界。结果表明，这种看似简单的二部结构

本质上具有丰富的几何结构，这为检测问题带来了令人惊讶的相图，例如图2所示。最后，我们提供了证据表明，在低效的扫描检验成功而计数检验和度检验失败的参数区域内，该问题推测在计算上是困难的；我们证明了低次多项式类在该区域内无法奏效。未来研究的一个有趣方向是考虑更一般的情形，即$p = \Theta(n^{-\beta})$ 且$q = \Theta(n^{-\gamma})$，其中$\beta < \gamma$。看来，我们的一些结果和技术同样适用于这种情形，但若要获得紧的统计和计算障碍，还需要采用其他技术，尤其是在$k_{\mathsf{R}} \vee k_{\mathsf{R}} \ll \sqrt{n}$ 且 $\gamma > 2\beta$的区域，此时 $\chi^2(p|q) = \Theta\left(n^{\gamma-2\beta}\right) \gg 1$ – (n) 。最后，尽管本文分析了检测问题，恢复问题同样值得关注且十分重要。我们的检测算法可以轻松地转化为相应的恢复算法（例如，扫描检验的灵感来自最大似然估计量）。要获得下界，则需要采用其他技术。例如，对于计算下界，采用近期提出的用于恢复的低次多项式框架将会很有意思[66]。

参考文献

[1] T. Cai 和 Yihong Wu。稀疏矩阵检测的统计与计算极限。《统计年鉴》，48，2018年1月。[2] Mark Jerrum。大团逃脱 *Metropolis* 过程。《随机结构与算法》，3(4):347–359，1992年。[3] Ludek Kucera。图划分问题的期望复杂度。《离散应用数学》，57(2):193–212，1995年。[4] Noga Alon、Michael Krivelevich 和 Benny Sudakov。在随机图中寻找大型隐藏团。《随机结构与算法》，13(3-4):457–466，1998年。[5] Boaz Barak、Samuel B. Hopkins、Jonathan Kelner、Pravesh K. Kothari、Ankur Moitra 和 Aaron Potechin。植入团问题的近乎紧致平方和下界。*2016*年 *IEEE* 第 *57* 届年度计算机科学基础研讨会（*FOCS*），第 428–437 页，2016年。[6] Yael Dekel、Ori Gurel-Gurevich 和 Yuval Peres。以高概率在线性时间内寻找隐藏团。《概率与计算》，第 29–49 页，2014年。[7] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S. Vempala 和 Ying Xiao。统计算法及检测植入团的下界。《*ACM* 杂志》，64(2)，2014年。[8] Bruce Hajek、Yihong Wu 和 Jiaming Xu。随机图社区检测的计算下界。《第二十八届学习理论会议论文集》，40:899–928，2015年。

[9] 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不断增加时的子矩阵定位。机器学习研究期刊，17(27):1–57，2016年。[10] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。植入稀疏结构问题的可约性与计算下界。学习理论第三十一届会议论文集，75:48–166，2016年。[11] Laurent Massoulie、Ludovic Stephan 和 Don Towsley。在图中植入树并将其找回。学习理论第三十二届会议论文集，99:2341–2371，2019年。[12] Paul Erd´os 和 Alfr´ed R´enyi。关于随机图的演化。匈牙利科学院数学研究所出版物，5:1761，1960年。[13] B´ela Bollob´as。小子图的阈值函数。剑桥哲学学会数学会刊，90(2):197–206，1981年。[14] S. B. Hopkins 和 D. Steurer。利用少量样本进行高效贝叶斯估计：社区检测及相关问题。载于*2017*年*IEEE*第*58*届计算机科学基础年会（*FOCS*），第 379–390 页，2017年。[15] Samuel Hopkins B。统计推断与平方和方法。博士论文，康奈尔大学，2018年。[16] Afonso S. Bandeira、Dmitriy Kunisky 和 Alexander S. Wein。约束主成分分析问题中界限认证的计算困难性。载于第 *11* 届理论计算机科学创新会议（*ITCS 2020*），第 151 卷，第 78:1–78:29 页，2020年。[17] Yeshwanth Cherapanamjeri、Samuel B. Hopkins、Tarun Kathuria、Prasad Raghaven-dra 和 Nilesh Tripuraneni。重尾统计的算法：回归、协方差估计及其他。载于第 *52* 届 *ACM SIGACT*计算理论研讨会论文集，STOC 2020，第 601–609 页，2020年。[18] David Gamarnik、Aukosh Jagannath 和 Alexander S. Wein。随机优化问题的低次多项式困难性。载于*2020*年*IEEE*第*61*届计算机科学基础年会（*FOCS*），第 324–356 页，2020年。[19] Wasim Huleihel。推断随机图中的隐藏结构。*IEEE*网络信号与信息处理汇刊，8:855–867，2022年。[20] Ery Arias-Castro 和 Nicolas Verzelen。稠密随机网络中的社区检测。《统计年鉴》，42(3):940–969，2014年。[21] Nicolas Verzelen 和 Ery Arias-Castro。稀疏随机网络中的社区检测。应用概率年刊，25(6):3465–3510，2015年。[22] Quentin Berthet 和 Philippe Rigollet。稀疏主成分检测的计算复杂性理论下界。第二十六届学习理论年度会议论文集

g 理论 *y*，30:1046–1066，2013年。

[23] Emmanuel Abbe。社区检测与随机块模型：近期进展。机器学习研究期刊，18:6446–6531，2017年。[24]Yihong Wu 和 Jiaming Xu。植入结构中的统计问题：信息论与计算极限。*arXiv*，abs/1806.00118，2018年。[25] Laurent Massouli´e、Ludovic Stephan 和 Don Towsley。在图中植入树，并将其找回。载于第三十二届学习理论会议论文集，第99卷，第2341–2371页，2019年6月。[26] Vivek Bagaria、Jian Ding、David Tse、Yihong Wu 和 Jiaming Xu。通过线性规划恢复隐藏的哈密顿回路。运筹学，68(1):53–70，2020年。[27] Armen S. Asratian、Tristan M. J. Denley 和 Roland H¨aggkvist。二分图及其应用。剑桥大学出版社，1998年。[28]Georgios A Pavlopoulos、Panagiota I Kontou、Athanasia Pavlopoulou、Costas Bouyioukos、Evripides Markou 和 Pantelis G Bagos。系统生物学与医学中的二分图：方法与应用综述。*Gigascience*，7:1–31，2018年。[29] Quentin Berthet 和 Philippe Rigollet。稀疏主成分检测的计算复杂性理论下界。载于第二十六届学习理论年度会议论文集，第30卷，第1046–1066页，2013年6月12–14日。[30] Zongming Ma 和 Yihong Wu。极小极大子矩阵检测中的计算障碍。《统计年鉴》，43(3):1089–1116，2015年。[31] 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。《统计年鉴》，45(4):1403–1430，2017年8月。[32] Robert Krauthgamer、Boaz Nadler 和 Dan Vilenchik。半正定松弛能否在信息极限内解决稀疏主成分分析问题?《统计年鉴》，43(3):1300–1322，2015年。[33] Bruce Hajek、Yihong Wu 和 Jiaming Xu。随机图上社区检测的计算下界。载于第二十八届学习理论会议论文集，第40卷，第899–928页，2015年7月3–6日。[34] 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不断增加时的子矩阵定位。机器学习研究期刊，17(27):1–57，2016年。[35] Tengyao Wang、Quentin Berthet 和 Yaniv Plan。RIP认证的平均情形困难性。载于神经信息处理系统进展，第3819–3827页，2016年。[36] Tengyao Wang、Quentin Berthet 和 Richard J Samworth。稀疏主成分估计中的统计与计算权衡。《统计年鉴》，44(5):1896–1930，2016年。

[37] Chao Gao、Zongming Ma 和 Harrison H Zhou。稀疏典型相关分析：自适应估计与计算障碍。《统计年鉴》，45(5):2074–2101，2017年。[38] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。具有植入稀疏结构问题的可约性与计算下界。载于第三十一届学习理论会议论文集，第75卷，第48–166页，2018年7月6–9日。[39] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。子矩阵检测计算下界的普适性。载于第三十二届学习理论会议论文集，第99卷，第417–468页，2019年6月25–28日。[40] Yihong Wu 和 Jiaming Xu。具有植入结构的统计问题：信息论与计算极限。载于 Miguel R. D. Rodrigues 和 Yonina C. Eldar（编），数据科学中的信息论方法。剑桥大学出版社，剑桥，2020年。[41] Matthew Brennan 和 Guy Bresler。秘密泄漏导致的可约性与统计计算差距。载于第三十三届学习理论会议论文集，第125卷，第648–847页，2020年7月9–12日。[42] B. Barak、S. B. Hopkins、J. Kelner、P. Kothari、A. Moitra 和 A. Potechin。植入团问题近乎紧的平方和下界。载于*2016*年*IEEE*第*57*届计算机科学基础年会（*FOCS*），第428–437页，2016年。[43] Yash Deshpande 和 Andrea Montanari。隐藏团和隐藏子矩阵问题的改进平方和下界。载于第二十八届学习理论会议论文集，机器学习研究论文集第40卷，第523–562页，2015年7月。[44] Raghu Meka、Aaron Potechin 和 Avi Wigderson。植入团的平方和下界。载于第四十七届*ACM*计算理论研讨会年会论文集，第87–96页。ACM，2015年。[45] Tengyu Ma 和 Avi Wigderson。稀疏主成分分析的平方和下界。载于第二十八届神经信息处理系统国际会议论文集——第一卷，第1612–1620页，2015年。[46] Pravesh K. Kothari、Ryuhei Mori、Ryan O'Donnell 和 David Witmer。驳斥任意约束满足问题的平方和下界。载于第四十九届*ACM SIGACT*计算理论研讨会年会论文集，第132–145页。计算机协会，2017年。[47] Samuel B. Hopkins、Pravesh Kothari、Aaron Henry Potechin、Prasad Raghavendra 和 Tselil Schramm。关于植入团的四次平方和的整数性间隙。*ACM*算法汇刊，14(3)，2018年。[48] Prasad Raghavendra、Tselil Schramm 和 David Steurer。利用平方和证明进行高维估计。载于国际数学家大会论文集（*ICM 2018*），第4卷，第3389–3424页，2019年。

[49] Samuel B Hopkins、Pravesh K Kothari、Aaron Potechin、Prasad Raghavendra、Tselil Schramm 和 David Steurer。平方和在检测隐藏结构方面的威力。*IEEE*第五十八届计算机科学基础研讨会（*FOCS*）论文集，第720–731页，2017年。[50] Sidhanth Mohanty、Prasad Raghavendra 和 Jeff Xu。提升平方和下界：从二次到四次。载于第五十二届*ACM SIGACT*计算理论研讨会年会论文集，STOC 2020，第840–853页，2020年。[51] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S. Vempala 和 Ying Xiao。用于检测植入团的统计算法及其下界。《*ACM* 杂志》，64(2)，2017年4月。[52] Vitaly Feldman、Will Perkins 和 Santosh Vempala。具有植入解的随机可满足性问题的复杂度。载于第四十七届*ACM*计算理论研讨会年会论文集，STOC ’15，第77–86页，2015年。[53] Ilias Diakonikolas、Daniel Kane 和 Alistair Stewart。稳健估计高维高斯分布及高斯混合分布的统计查询下界。载于*2017*年*IEEE*第*58*届计算机科学基础年会（*FOCS*），第73–84页，2017年。[54] Ilias Diakonikolas、Weihao Kong 和 Alistair Stewart。稳健线性回归的高效算法与下界。载于工业与应用数学学会（*SODA’19*），第2745–2754页，2019年。[55] Lenka Zdeborov´a 和 Florent Krzakala。推断的统计物理：阈值与算法。《物理学进展》，65(5):453–552，2016年。[56] Thibault Lesieur、Florent Krzakala 和 Lenka Zdeborova。概率低秩矩阵估计的最小均方误差：关于输出信道的普适性。*2015*年第*53*届*Allerton*通信、控制与计算年会（*Allerton*），2015年9月。[57] Thibault Lesieur、Caterina de Bacco、Jess Banks、Florent Krzakala、Cris Moore 和 Lenka Zdeborova。高维高斯混合聚类中的相变与最优算法。*2016*年第*54*届*Allerton*通信、控制与计算年会（*Allerton*），2016年9月。[58] Florent Krzakala、Andrea Montanari、Federico Ricci Tersenghi、Guilhem Semerjian 和 Lenka Zdeborova。吉布斯态与随机约束满足问题的解集。《美国国家科学院院刊》，104(25):10318–10323，2007年。[59] Federico Ricci-Tersenghi、Guilhem Semerjian 和 Lenka Zdeborov´a。贝叶斯推断问题中相变的类型学。《物理评论*E*》，99(4)，2019年4月。

[60] Afonso S. Bandeira、Amelia Perry 和 Alexander S. Wein。《关于计算与统计差距的札记：基于统计物理的预测》。*Portugaliae Mathematica*，75(2):159–186，2018年。[61] Mladen Kolar、Sivaraman Balakrishnan、Alessandro Rinaldo 和 Aarti Singh。《大型噪声矩阵中结构信息的极小极大定位》。载于神经信息处理系统进展，第909–917页，2011年。[62] 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。[63] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。《子矩阵检测计算下界的普适性》。载于第三十二届学习理论会议论文集，机器学习研究论文集第99卷，第417–468页。PMLR，2019年6月25–28日。[64] Dmitriy Kunisky、Alexander S. Wein 和 Afonso S. Bandeira。《关于假设检验计算困难性的札记：基于低次数似然比的预测》。载于数学分析、应用与计算，第1–50页。Springer International Publishing，2022年。[65] Ryan O'Donnell。《布尔函数分析》。剑桥大学出版社，2014年。[66] Tselil Schramm 和 Alexander S. Wein。《低次多项式估计的计算障碍》。《统计年鉴》，50(3):1833–1858，2022年。