arXiv:2603.24545v1 [math.ST] 25 Mar 2026

随机图中局部几何的检测：信息论与计算极限

Jinho Bok*、Shuangping Li†和 SophieH.Yu‡

2026年3月25日

摘要

我们研究随机图中局部几何的检测问题。我们引入模型$\mathcal{G}(n,\ p,\ d,\ k)$，其中，一个平均大小为 k 的隐藏社区，其内部的边按照 $\mathbb{S}^{d-1}$上的随机几何图生成，而其余所有边则遵循 Erd˝os–R´enyi 模型 $\mathcal{G}(n,\ p)$。随机几何图通过对 $\mathbb{S}^{d-1}$上的潜在向量内积进行阈值化生成，其中每条边的边缘概率均等于 p。这意味着，在边缘分布层面，$\mathcal{G}(n,\ p,\ d,\ k)$与 $\mathcal{G}(n,\ p)$无法区分，而信号完全存在于局部几何所诱导的边依赖中。

我们研究检测的信息论极限和计算极限。在信息论方面，我们的上界来自三种基于带符号三角形计数的检验：全局检验、扫描检验和约束扫描检验；下界则来自两种互补方法：通过 Wishart–GOE比较得到的截断二阶矩，以及 KL 散度的张量化。这些结果共同确定了固定 p 时的检测阈值为 $d = \Theta(\tilde{k}^2 \vee k^6/n^3)$，并将当前最优界从完整模型（即 $k = n$）推广至渐消的 p。在计算方面，我们指出了计算—统计差距，并通过低次多项式框架提供证据，同时证明长度为 $\ell \geq 4$ 的带符号环计数并非最优。

*J. Bok 就职于宾夕法尼亚大学沃顿商学院统计与数据科学系。电子邮箱：jinhobok@wharton.upenn.edu。

†S. Li 任职于耶鲁大学统计与数据科学系。电子邮箱：shuangping.li@yale.edu。

‡S. H. Yu 任职于宾夕法尼亚大学沃顿商学院运营、信息与决策系。电子邮箱：hysophie@wharton.upenn.edu。

目录

1 引言 **1**
1.1 我们的贡献.31.2 相关文献 .41.3 记号 .7

2 主要结果 **7**
2.1 信息论上界 .72.2 信息论下界 .92.3 计算下界 .10

3 技术概述 **11**

4 信息论上界的证明 **13**
4.1 全局检验 .144.2 扫描检验 .154.3 约束扫描检验 .16

5 信息论下界的证明 **19**
5.1 截断二阶矩 .195.2 KL 散度的张量化 .24

6 计算下界的证明 **27**6.1 低次下界 .276.2 较长环计数的次优性 .29

7 讨论 **30**

参考文献 **31**

A 辅助引理 **37**A.1 集中不等式 .37A.2 有符号子图计数的上界 .39

B 第 4 节中的延后证明 **39**B.1 带符号三角形计数的方差（引理 4.4）.39B.2 有符号楔形计数的典型行为（引理 4.5）.41B.2.1 有符号楔形计数的方差 .44

C 延后证明　　第5节中的证明46
C.1 Wishart 与球面 Wishart 之间的全变差距离（命题 5.1）. . . . . 46C.2 似然比的分解（引理 5.3）.51C.3 三次超几何的上界（引理 5.4）.55C.4 邻域分布的典型行为（引理 5.5）.56

D 第6节中的延后证明 **61**
D.1 有符号环计数的紧界（命题 6.1）.61D.1.1 Gegenbauer 多项式与球谐函数.61D.1.2 有符号环计数的紧界.62

1 引言

跨多个领域的网络往往包含固有结构 [New10, Bar16]：社交网络中的社群 [HLL83, GN02, For 10]，生物系统中的功能模块 [HHLM99, SM03, BO04]，以及通信网络中的异常子图 [PCMP05, ATK15]。从含噪观测中检测此类结构是一项基础统计问题，也推动了概率论、组合学和算法理论的重大进展。这项任务的典型模型包括随机块模型 [HLL83, DKMZ11, Abb17]，植入团 [Jer92, Kuˇc95]，植入稠密子图 [ACV14, HWX15, VAC15]，以及隐匿匹配 [MMX21, DWXY23]，，每一种模型都可作为研究结构化随机图中统计与计算相变的基准。

这些模型中的隐藏子图通常被假定为“简单”的，即具有鲜明的组合形状，或相较于背景具有更高的边密度。虽然这类假设在分析上较为方便，但它们可能与现实世界网络的情况不符：子图的标志性特征或许不在于其密度或形状，而在于其顶点如何通过潜在特征相互关联。具体而言，边的形成受到这些特征之间相似性的影响（例如个人资料、文本表示、生物学摘要），这种相似性通常通过潜在空间来建模 [HRH02，Pen03，HRT07，NC16]。从这一视角来看，待检测结构更适合用与底层几何结构相一致的交互模式来描述。

当每个顶点都与其他顶点相似或模仿其他顶点时，这一区分尤为重要。例如，在社交网络中，目标子图可能是由机器人账号包围的一小群真实用户 [FVD+16]，也可能是一组为施加影响力而协同行动的账号 [PHT+21]；类似地，在经济网络中，子图可能由市场中合谋的企业构成[MO18，WK19]。在这类情况下，尽管子图内部在潜在空间层面存在交互，每个顶点在统计上看起来仍可能彼此相似。此外，这些交互往往本质上错综复杂，其特点是在高维空间中具有上下文或纵向特征。

受上述基于几何的信号启发，我们引入一种随机图模型，其中一个规模较小的隐藏社区呈现局部几何。我们关注基本任务检测：给定一个定义在顶点集 $[n]$ 上的观测图 G，判断它是由无信号的零模型生成，还是由我们提出的备择模型生成。

形式化地，我们考虑如下假设检验问题

$$\mathcal{P} := \mathcal{G}(n,p,d,k) \quad \text{vs.} \quad \mathcal{Q} := \mathcal{G}(n,p)\,,$$

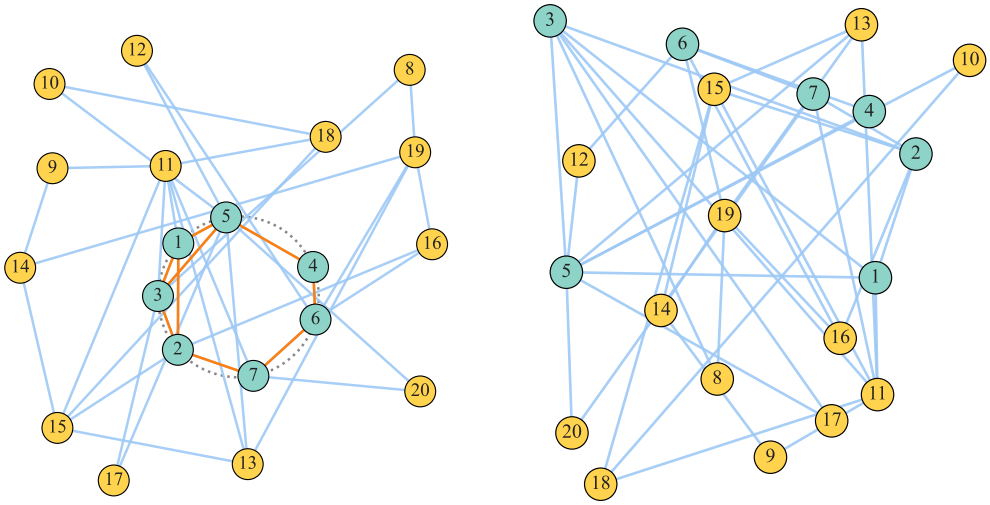
其中 $\mathcal{Q}$ 是 Erd˝os–R´enyi 模型 [ER59]，而 $\mathcal{P}$ 是高维随机几何图模型 [DGLU11]的植入版本。在 $\mathcal{P}$下，一个隐藏集合 S 的期望规模为 k（我们称之为社区），其携带潜在特征向量，而其中的边 S 按照维 d潜在空间中的几何邻近性形成；其余所有边的行为均与 $\mathcal{G}(n,p)$ 相同。现在给出 $\mathcal{P}$ 的正式定义。

定义 1.1（具有局部高维几何的随机图）。 样本 $G \sim \mathcal{P}$ 按如下方式抽取：

1。每个顶点 $i \in [n]$ 以概率 $k/$独立加入社区 S $n.$

2。每个社区顶点 $i \in S$ 都会获得一个潜在特征向量 U_i i $\overset{\text{i.d.}}{\sim} \mathcal{U}(\mathbb{S}^{d-1}).$

3. 对于任意 $i,j \in [n]$，若 $i \neq j$，当且仅当 $i,j \in S$ 时，边 ij 存在（p,d）；否则，边 ij 以概率 p 独立存在。阈值 τ（p,d） 的选取满足$\mathbb{P}$（$\langle U_i, U_j\rangle \geq \tau$（$p,d$）$) = p$。



(a) (b)

图1：从 $\mathcal{G}(n,p,d,k)$ 中抽样得到的同一张图的两种绘图方式，其中 $n=20$, $p=0.18$, $k=7$；为便于可视化，我们设定 $d=2$ 。（a）顶点的位置反映了潜在几何： $k=7$ 个社群顶点（以青绿色圆圈标出）的潜在向量取自 $\mathcal{U}(\mathbb{S}^1)$，因此它们在潜在空间中位于一个圆上。橙色的边基于几何邻近性，揭示了局部几何所诱导的富含环的结构。（b）顶点随机放置，植入社区在视觉上与 Erd˝os–R´enyi 背景无法区分。

我们只能观察到最终的图 G；社区 S 和潜在向量 $\{U_i\}$均不可观测。此外，根据构造，每条边的边缘出现概率均为 p，无论顶点属于 S与否，其边缘邻域分布都相同。因此，信号在一阶矩层面不可见，而是由局部几何诱导的依赖结构承载。图1展示了从 $\mathcal{G}(n,p,d,k)$ 中抽取的一个样本。当 $k=n$ 时，我们将分布 $\mathcal{G}(n,p,d,k)$ 记作 $\mathcal{G}(n,p,d)$；这是文献中研究的传统高维随机几何图 [DGLU11] ，我们称之为完整模型。

我们研究当 $n\to\infty$ 时何时可以进行检测，并允许参数 p,d,k 依赖于 n（因此也彼此依赖）。我们采用以下标准概念。

定义 1.2。 对于 $\mathcal{P}$ 与 $\mathcal{Q}$ 的检测问题，若检验统计量 $f(G)$ 以阈值 γ 达到

*(a)*强检测 若 $\mathbb{P}_{G\sim\mathcal{Q}}(f(G)>\gamma)+\mathbb{P}_{G\sim\mathcal{P}}(f(G)\le\gamma)=o(1)$;

*(b)*弱检测 若 $\mathbb{P}_{G\sim\mathcal{Q}}(f(G)>\gamma)+\mathbb{P}_{G\sim\mathcal{P}}(f(G)\le\gamma)=1-\Omega(1)$。

众所周知（由 Neyman–Pearson 引理可知），第一类错误与第二类错误之和的下确界等于 $1-d_{\mathrm{TV}}(\mathcal{P},\ \mathcal{Q})$。特别地，如果 $d_{\mathrm{TV}}(\mathcal{P},\ \mathcal{Q})=o(1)$，则任何检验都无法实现弱检测。直观而言，随着维度 d 增大，检测会变得更加困难，因为几何约束会导致高维情形下边之间的依赖性更弱。我们的主要目标是定量刻画在检测仍然可行的条件下， d 可以有多大（作为 n， p， k 的函数）。

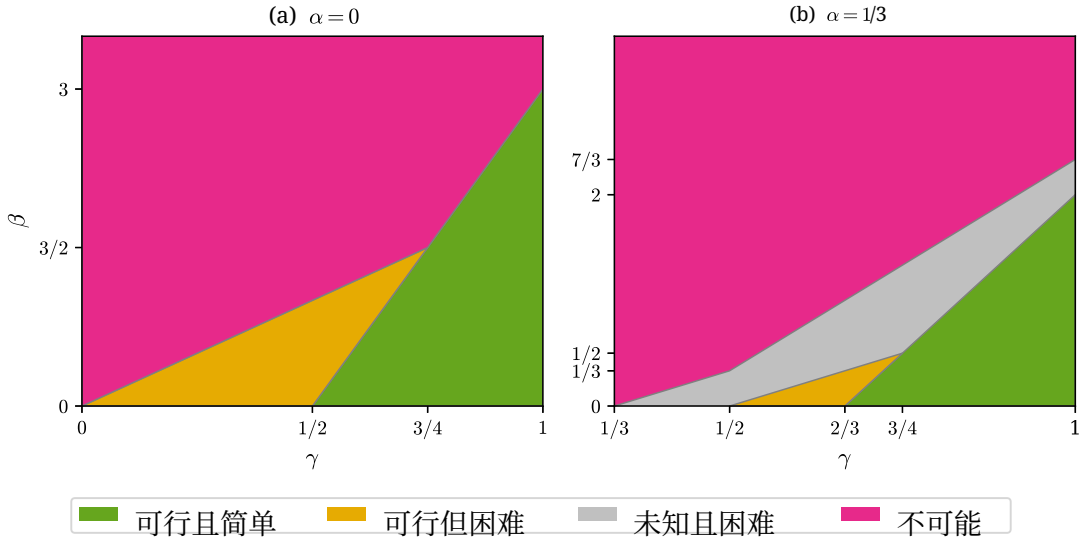


图2：检测的相图，其中 (a) $\alpha = 0$ 和 (b) $\alpha = 1/3$。请注意，两个图的刻度不同。在可能且容易相（绿色）中，可以通过高效的检验统计量进行强检测。在可能且困难相（黄色）中，可以通过低效的检验统计量进行强检测，而低次多项式算法无法实现弱检测。在未知且困难相（灰色）中，强检测是否可能尚无定论，但低次多项式算法无法实现弱检测。最后，在不可能相（洋红色）中，弱检测是不可能的。不可能相还延伸至图中所示范围之外的所有 $\beta > 0$取值。

1.1 我们的贡献

我们刻画了何时能够检测 $\mathcal{P}$ 与 $\mathcal{Q}$ 之间的差异，并分别给出信息论极限和计算极限，说明它们如何随 n, p, d, k变化。以下定理总结了对数密度设定下的主要结果 [BCC+10]；所得相图见图 2。

定理 1.3（非正式表述）。设

$$p = \Theta(n^{-\alpha}), \quad d = \Theta(n^{\beta}), \quad k = \Theta(n^{\gamma}),$$

其中 $0 \le \alpha < 1$、$\beta > 0$ 和 $0 < \gamma \le 1$。

(i) 如果 $\beta < 6\gamma - 3\alpha - 3$，可以使用高效可计算的检验统计量实现强检测 .

(*ii*) 如果 $\beta < 2\gamma - 3\alpha$，可以使用低效可计算的检验统计量实现强检测 .

(*iii*) 如果满足以下任一条件，则不可能实现弱检测：

- $\beta > 2\gamma \vee (6\gamma - 3)$;
- $\beta > (2\gamma - 2\alpha) \vee (4\gamma - 2\alpha - 1)$ *and* $\gamma > \alpha$.

(iv) 如果 $\beta > 6\gamma - 3\alpha - 3$，低次多项式算法无法实现弱检测。

信息论极限。定理1.3的 (i)–(iii) 部分共同刻画了检测的信息论阈值。在上界方面，我们提出了三种基于有符号三角形计数的检验：全局检验（在整个图上计数）、扫描检验（对所有大小为 $\approx k$的子集取有符号三角形计数的最大值），以及约束扫描检验（进一步将范围限制为楔形和受控的子集）。扫描检验适用于与全局检验互补的情形——两者结合足以处理固定的 p——而当 $p = \tilde{o}$满足 (1) 时，约束扫描检验严格扩展了扫描检验的参数范围。详情参见定理2.1、2.2 和2.3。在下界方面，我们提出了两种互补的方法：截断二阶矩，用于刻画对平均社区规模 k的依赖；以及 KL 散度的张量化，用于刻画对边密度 p的依赖。详情参见定理2.5 和2.6。

在稠密情形（$\alpha = 0$）中，结合上界和下界，可精确确定检测阈值为 $d = \widetilde{\Theta}(k^2 \vee k^6/n^3)$；见图 2(a)。对于 $0 < \alpha < 1$，我们的结果将完整模型 $\mathcal{G}(n, p, d)$（特殊情形 $k = n$，即 $\gamma = 1$）的当前最优界推广，重现了 [LMSY22] 的上界 $d = \tilde{o}(n^3p^3)$ 和下界 $d = \widetilde{\Omega}(n^3p^2)$，并将其扩展到我们的植入设定，对所有满足 $\gamma \in (\alpha, 1)$ 的 $k = \Theta(n^\gamma)$ 均成立；见图 2(b)。当 $0 < \alpha < 1$ 时，我们的结果在上界和下界之间留下了一个差距（即图 2(b) 中的未知且困难区域）；我们猜想上界是紧的，而且该区域实际上属于不可能相。

计算极限。定理1.3 的第 (iv) 部分基于低次多项式框架 [Hop18、KWB22、Wei25]提供证据表明，超出第 (i) 部分的区域在计算上是困难的（背景见第1.2 节和2.3 节）；值得注意的是，在我们提出的检验中，只有全局有符号三角形计数能在多项式时间内运行。具体而言，我们证明，当 $\lfloor (\log n/\log(\log n))^2 \rfloor$ 时，不存在次数至多为 $\beta > 6\gamma - 3\alpha - 3$的多项式能够实现弱分离，这与全局检验的阈值相符；见定理2.9。我们进一步证明，长度为 $\ell \geq 4$ 的带符号环计数均无法优于三角形计数，这进一步表明，在所有带符号环计数中，全局有符号三角形计数可能是渐近最优高效检验；见命题2.10。

计算—统计差距。第 (i) 和 (ii) 部分以及第 (iv) 部分确定了一个$6\gamma - 3\alpha - 3 < \beta < 2\gamma - 3\alpha$ 的区域（图2中的黄色区域），在该区域中，检测在信息论意义上是可能的，但目前尚无已知的高效算法。有趣的是，对于完整模型 $\mathcal{G}(n, p, d)$（即 $k = n$ 没有局部性），目前尚未发现计算—统计差距，而且推测这种差距并不存在 [LMSY22]。因此，这里的差距似乎是植入的几何结构具有局部性的结果：隐藏社区可能位于指数多个位置中的任意一个，这为高效算法造成了组合搜索障碍。

1.2 相关文献

我们回顾几条与本文工作密切相关的研究方向。

高维随机几何图。 一条经典研究路线考察了固定维度下的随机几何图 d；参见 [Pen03] 以了解综述。对具有高维（即 $d \to \infty$ 为 $n \to \infty$）几何结构的随机图的研究始于 [DGLU11],，其中作者研究了 $\mathcal{G}(n, p, d)$ 的团数，并证明当 $d \geq \exp(\widetilde{\Omega}(n^2))$时，该图与 $\mathcal{G}(n, p)$ 无法区分。此后，关于高维随机几何图的研究不断增加，探讨了各种算法和统计现象。

对于 $\mathcal{G}(n, p, d)$ 与 $\mathcal{G}(n, p)$ 之间的检测， [BDER16] 的突破性工作确定了固定 $d = \Theta(n^3)$ 时的阈值。该论文还提出了一种检验方法，用于统计

带符号三角形的数量（详情见第 3 节），达到了针对一般 p [LMSY22] 的最佳已知上界 $d = \tilde{o}(n^3p^3)$。此后，有多项工作改进了 $p = o(1)$ 情形下的下界。具体而言， [BBN20] 证明了当 $p = \tilde{\Omega}(1/n^2)$ 时，下界为 $d = \tilde{\Omega}(n^3p \vee n^{7/2}p^2 \vee n)$。随后， [LMSY22], 进一步改进了这一结果，取得了最先进的结果：当 $p = \Omega(1/n)$ 时，下界为 $d = \tilde{\Omega}(n^3p^2)$；当 $p = \Theta(1/n)$ 时，下界为 $d = \Omega((\log n)^{36})$。值得注意的是，如何缩小上界和下界之间的多项式因子 p 差距，仍是一个开放问题。

除了检测之外， $\mathcal{G}$（n、 p、 d）近年来也日益受到关注，相关研究涉及其谱性质 [LMSY23、ABI24、CZ25], 低阶矩 [BB24b], 与 $\mathcal{G}$（n、 p）[的耦合], LMSY22、BB25b [潜变量估计24], MZ [以及稀有事件25], DLW $\mathcal{G}$等。 此外，学界也积极研究与n（p、 d、 []）有所不同但密切相关的模型，区别体现在诸多因素上，例如边连接规则26], LR23a、LR23b、MWX [潜在几何的度量25], BB24a、BGPS [各向同性24], EM20、BBH [齐次性19], BKL [簇结构23], LS [以及这些因素的组合。关于高维随机几何图的进一步讨论，请参阅近期综述23]DDC。

植入子图。 我们的模型可以视为一类称为植入子图的随机图中的一个特例。一般而言，这类图可以通过以下方式生成：首先抽取一个背景 Erd˝os–R´enyi 图 $G \sim \mathcal{G}(n, p)$，再独立地从另一种分布中抽取一个子图 H，使其定义在完全图上。对于顶点集 S 属于 H的情形，最终图可通过以下方式得到：用 $G[S]$ 替换诱导子图 H，或取并集 $G \cup H$。植入子图可视为尖峰随机矩阵的图（二元）版本 [Joh01、BBAP05], ，而尖峰随机矩阵是概率论和统计学中的基本研究对象。在具有社区结构的随机图模型的广阔领域中，许多植入子图模型可通过单个社区的存在来刻画。关于具有多个社区的模型（例如随机块模型 [HLL83]）的文献十分丰富，值得单独讨论；本文在此聚焦于单社区情形。

植入子图的一个经典例子是植入团 [Jer92, Kuˇc95], ：在来自 $\mathcal{G}(n, 1/2)$ 的图中隐藏一个规模较小的团（即完全图）。这个简单的模型因展现出计算—统计差距而广为人知，并与理论计算机科学 [JP00, HK11] 和高维统计学 [BR13, BB20]有广泛联系。此后，相关文献不断丰富，研究者对植入子图进行了多种选择，例如稠密子图[ACV14, HWX15, VAC15], 、树 [MST19], 、环 [BDT+20, GSXY25], 、匹配 [MMX21, DWXY23, WM25, ABAL+26]等。除了针对特定子图的研究，近期还有一系列工作 [Hul22, EH25, LPRZ25, MNWS+25, YZZ25]致力于为一般子图建立统一理论。

超越均值的信号。 植入子图的一个常见特征是，信号体现在均值层面。例如，在植入稠密子图 [ACV14]中，社区内部的边密度平均而言高于图的其余部分。换言之，信号已经存在于最低层次（一阶矩），通常无需考虑输入之间的任何交互（例如高阶矩）。因此，通常情况下，只要操作得当，对均值进行阈值化就足够了。但我们的模型并非如此，因为每个顶点的边际分布都相同。

一些论文研究了隐藏信号在均值层面不可观测的情形。例如， [ACBL12、ACBL15、ACBLV18]研究了这样的检测问题：在均值为 0 的样本中，只有一个规模较小且未知的子集内部存在依赖关系，例如正相关或马尔可夫结构。另一篇近期值得关注的论文 [KSWY25] 深入分析了一种新的有向

随机图模型，其中一个规模较小且未知的顶点子集具有潜在排序。在该模型中，排序顶点之间，从较高排名顶点指向较低排名顶点的有向边比反方向更可能被添加。由于整个图的边密度相同，均值层面不存在信号；特别是，忽略方向时，该模型等价于 $\mathcal{G}(n,\ p)$。因此，检测是通过考察成对排序的异常一致性来完成的，而非考察边密度。[1]

具有几何结构社区的随机图。近年来，人们研究了各种同时具有潜在几何结构和社区结构的随机图模型。这些模型中的社区结构通常采用随机块模型的形式，由多个社区组成：例如，几何块模型 [GMPS18，GMPS23], 几何 SBM [ABS21，GNW24，GJ25], 几何隐藏社区模型 [GGNW24，GJ26],及其不同变体 [PP20，ABD21，AKL24]。我们的模型与这些模型的不同之处在于，它只有一个隐藏社区，而非多个。一个值得注意的例外是植入稠密环[MWZ23，MWZ25], 这是一种基于潜在结构的模型，同样使用带符号三角形计数进行检测；然而，就建模组件而言，该模型与我们的模型也有根本区别。

我们重点介绍文献中两个与我们的模型具有某些共同特征的模型。在 [BBCvdH20], 提出的模型中，一个由 Erd˝os–R´enyi 图构成的小型社区隐藏在一个更大的随机几何图中。因此，从概念层面看，我们的模型可以视为他们模型的反转版本，即将社区与非社区的角色互换。在 [BMS25], 提出的另一个模型中，隐藏社区内部的边受其潜在几何结构影响，而社区外部的边则独立生成。因此，原则上，我们以相同方式结合社区结构与潜在几何结构。尽管存在这些相似之处，这些模型的数学细节与我们的模型大不相同，因此我们的结果无法与其结果直接比较。此外，我们定量刻画了几何结构中的高维性如何影响图，而这两项工作的结果要么与 d 无关，要么针对固定的 d。

低次多项式框架。 许多高维统计模型都呈现出一种称为计算—统计差距的现象：在某些参数区域中，某个任务在信息论意义上可行，但似乎不存在任何多项式时间算法。分析这一现象的一种重要方法是低次多项式框架 [Hop18], ，该框架研究可表示为输入的低次多项式的算法。对于输入维度 n，人们考虑次数至多为 D的多项式；其基本启发式是， $D = O(\log n)$ 往往能匹配多种平均情形问题中多项式时间算法的能力。许多实例支持这一启发式，因为次数为 $O(\log n)$ 的多项式可以实现或近似各种高效算法，包括谱方法、子图计数程序和近似消息传递（见 [Wei25，第 6 节。2]）。因此，广泛认为，次数为 $D = \omega(\log n)$ 时的困难性，表明存在超越多项式时间的计算困难性。

对于假设检验，低次多项式框架考察低次多项式能否实现零分布与备择分布之间的分离（见定义2.8）。实践中，这通常通过低次似然比来分析，即将似然比投影到次数至多为 D的多项式空间后所得投影的范数；详情参见例如 [KWB22] 。理解这一准则的严格含义，近来已

[1]我们提到，尽管 n 项关键技术结果与 各个模型表面上有所不同，但有趣的是，我们的分析仍共享某些其结果（例如引理 5.4）。

引起广泛关注，并成为一个活跃的研究领域 [HW21、BHJK25、HKK+26、JV26]；进一步讨论请参阅近期的综述 [Wei25] 。

1.3 记号

我们记 $[n] := \{1, \dots, n\}$，并用 $\binom{[n]}{k}$ 表示 $[n]$的所有大小为 k 的子集构成的集合。对于图 H，$V(H)$ 和 $E(H)$ 分别表示其顶点集和边集； $v(H)$ 和 $e(H)$ 分别表示它们的基数。 K_n 表示 $[n]$上的完全图。我们用 $\mathbb{S}^{d-1}$ 表示 $\mathbb{R}^d$中的单位球面，用 $\mathcal{U}(\mathbb{S}^{d-1})$表示 $\mathbb{S}^{d-1}$ 上的均匀分布（即 Haar测度）；全文中， $U_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(\mathbb{S}^{d-1})$ 表示 $i \in [n]$。对于对称矩阵 A， $\|A\|_{\mathrm{op}}$表示其算子范数。所有对数均以 exp(1) 为底，$\log^a b$ 表示 $(\log b)^a$。

在讨论渐近行为时，我们总是假设 $n \to \infty$，其中 $k = k(n) \to \infty$，且 d 足够大。我们采用标准的大-O 记号，其中对于任意 a_n， b_n、 $a_n = O(b_n)$ 和 $a_n \lesssim b_n$ 表示对某个绝对常数 $C > 0$有 $a_n \le C b_n$； $a_n = \Omega(b_n)$ 和 $a_n \gtrsim b_n$ 表示 $b_n = O(a_n)$； $a_n = \Theta(b_n)$表示$a_n = O(b_n)$且 $a_n = \Omega(b_n)$。此外， $a_n = o(b_n)$ 和 $a_n \ll b_n$ 分别表示 $\lim_{n\to\infty}(a_n/b_n) = 0$； $a_n = \omega(b_n)$，而 $a_n \gg b_n$ 表示 $b_n = o(a_n)$。对于上述各项，我们使用 O、Ω Θ、$\widetilde{o}$、$\widetilde{\omega}$ 来省略 polylog(n) 因子。

2 主要结果

本节给出检测问题的信息论上界和下界，以及计算下界。在本节中，我们假设 $p \le 1/2$；对于任意固定的$p \in (1/2,1)$，可以轻易推得（例如，参见 [BDER16，引理 3 和 4]），我们针对固定 $p \in (0, 1/2]$ 的检测阈值也成立。

2.1 信息论上界

我们提出了三种不同的检验，用于区分 $\mathcal{P}$ 和 $\mathcal{Q}$。首先，我们考虑用于计数带符号三角形的全局检验，其检验统计量定义为

$$f_{\mathrm{tri}}(G) := \sum_{i<j<\ell \in [n]} (G_{ij} - p)(G_{j\ell} - p)(G_{i\ell} - p)\,. \tag{2.1}$$

这是一个自然的候选方案，因为它在区分$\mathcal{G}(n,p)$与完整模型 $\mathcal{G}(n,p,d)$方面达到了已知的最佳性能 [BDER16、LMSY22]。

定理 2.1 (通过全局检验进行检测). 存在一个常数 $C_{2.1} > 0$，使得 *if*

$$\frac{1}{k} \le p \le \frac{1}{2} \quad \textit{and} \quad C_{2.1} \vee (5\log(1/p))^4 \le d \ll \frac{k^6 p^3}{n^3} \log^3(1/p)\,,$$

检验误差满足

$$\mathbb{P}_{G\sim\mathcal{Q}}\left(f_{\mathrm{tri}}(G) > \gamma_{\mathrm{tri}}\right) + \mathbb{P}_{G\sim\mathcal{P}}\left(f_{\mathrm{tri}}(G) \le \gamma_{\mathrm{tri}}\right) = o(1)\,,$$

其中，阈值选为[2]

$$\gamma_{\mathrm{tri}} := \frac{1}{2}\mathbb{E}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)] = \frac{1}{2}\binom{n}{3}\left(\frac{k}{n}\right)^3 \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[(G_{12}-p)(G_{23}-p)(G_{13}-p)]\,.$$

[2].....的常数因子 1/2 前面的 γ_{tri} 是任意的，可以替换为 (0，1) 中任意一个固定常数。扫描检验和约束扫描检验中的阈值也同样如此。根据引理 $\mathbb{E}_{G\sim\mathcal{P}}$ 和 $\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}$，4.1 中的表达式相等。

在全局检验中，我们实质上是将社区内的带符号三角形计数与整个图上带符号三角形计数的波动进行比较。这种波动可能相当大，尤其是在 n 远大于 k 时。因此，自然会考虑扫描检验，在这种检验中，我们转而考察大小相近的子图各自的波动。具体而言，令

$$f_{\mathrm{scan}}(G) := \max_{A\subseteq[n],|A|=k^-} \sum_{i<j<\ell\in A} (G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)\,, \tag{2.2}$$

其中 $k^- := \lfloor 0.9k \rfloor$。我们证明，当 $d = \widetilde{O}(k^2p^6)$ 时，该方法能够成功；这是一个仅对 n 呈对数依赖的阈值。

定理 2.2（通过扫描检验进行检测）。 存在一个常数 $C_{2.2} > 0$ 使得若

$$p \le \frac{1}{2}, \quad k \ge C_{2.2}\log^2 n \quad \textit{and} \quad C_{2.2} \vee (5\log(1/p))^4 \le d \le \frac{k^2p^6\log^6(1/p)}{C_{2.2}\log n}\,,$$

检验误差满足

$$\mathbb{P}_{G\sim\mathcal{Q}}\left(f_{\mathrm{scan}}(G) > \gamma_{\mathrm{scan}}\right) + \mathbb{P}_{G\sim\mathcal{P}}\left(f_{\mathrm{scan}}(G) \le \gamma_{\mathrm{scan}}\right) = o(1)\,, \tag{2.3}$$

其中阈值选为

$$\gamma_{\mathrm{scan}} := \frac{1}{2}\binom{k^-}{3}\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[(G_{12}-p)(G_{23}-p)(G_{13}-p)]\,. \tag{2.4}$$

正如我们稍后将看到的，对于任意固定的 p ，全局检验和扫描检验都是充分的，因为存在一个匹配的下界（相差至多一个对数因子）。然而，随着 $p \to 0$，扫描检验所给出的检测阈值会迅速变差。为改进这一点，我们在扫描检验的基础上增加了一些约束，并将其称为约束扫描检验。具体而言，在大小为 $\approx k$ 的子图中，我们只考虑满足楔形计数（带符号）附加条件的子图。形式化地，令

$$\widetilde{f}_{\mathrm{scan}}(G) := \max_{A\in\mathcal{C}(G),|A|=k^-} \sum_{i<j<\ell\in A} (G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)\,, \tag{2.5}$$

其中

$$\mathcal{C}(G) := \left\{ A\subseteq[n] : \sum_{i<j\in A}\left(\sum_{\ell<i,\ell\in A}(G_{\ell i}-p)(G_{\ell j}-p)\right)^2 \le \sigma^2\,, \right. \tag{2.6}$$

$$\left. \max_{i<j\in A}\left|\sum_{\ell<i,\ell\in A}(G_{\ell i}-p)(G_{\ell j}-p)\right| \le B \right\}, \tag{2.7}$$

with

$$\sigma^2 := k^3p^2 + C_{6.1}^4\frac{k^4p^4\log^2(1/p)}{d} \quad \text{and} \quad B := (2048kp^2+8)\lceil\log k\rceil\,.$$

定理 2.3（通过约束扫描检验进行检测）。 假设存在一个常数 $\delta > 0$，使得 $d \ge n^\delta, n^{-1+\delta} \le p \le 1/2$。那么存在一个常数 $C_{2.3} = C_{2.3}(\delta) > 0$ ，使得如果

$$d \le \frac{1}{C_{2.3}}\frac{k^2p^3\log^3(1/p)}{(\log^2 n)(\log^2 k)}\,,$$

~(2.3) 以 $f_{\mathrm{scan}}(G)$ 替代 $f_{\mathrm{scan}}(G)$时成立，也就是说，

$$\mathbb{P}_{G\sim\mathcal{Q}}\left(\widetilde{f}_{\mathrm{scan}}(G) > \gamma_{\mathrm{scan}}\right) + \mathbb{P}_{G\sim\mathcal{P}}\left(\widetilde{f}_{\mathrm{scan}}(G) \le \gamma_{\mathrm{scan}}\right) = o(1)\,.$$

注 2.4 （检验的比较）。 如前所述，全局检验（定理 2.1）和扫描检验（定理 2.2结合起来，足以刻画任意固定 p的检测阈值，误差在一个对数因子以内。就定理给出的性能保证而言，当 $k \gg (n^3/\log n)^{1/4}$时，全局检验更优；而当 $k \ll (n^3/\log n)^{1/4}$时，扫描检验更优。

当 $p = o(1)$ 时，除一个非常有限的范围外，约束扫描检验（定理 2.3）都优于扫描检验。事实上，根据我们的分析可直接得出，对于 $p = O((\log n)^{-1}(\log^2 k)^{-1})$，若 $d = O(k^2 p^3 \log^3(1/p)/\log n)$，则约束扫描检验能够成功，从而严格优于扫描检验；参见 (4.8) 及其周边论证。与稠密情形一样，全局检验还是约束扫描检验能够保证更好的性能，取决于 k 与 $\widetilde{\Theta}(n^{3/4})$ 的比较结果。

2.2 信息论下界

我们基于两种不同的方法，给出了检测不可能性的不同阈值。首先，我们着重刻画对 k 的依赖关系。为此，我们考虑计算生成随机图的某些随机矩阵之间的截断二阶矩。

定理 2.5 (基于截断二阶矩的下界)。 存在一个常数 $C_{2.5} > 0$，使得以下结论成立：若

$$k \le \frac{n}{5}, \quad d \ge C_{2.5} \log(1/p) \quad \text{and} \quad d \gg k^2 \vee \frac{k^6}{n^3},$$

没有任何检验能够实现弱检测。[3]

结合这个下界与上界（定理 2.1 和2.2），可以得出，对于任意固定的 p，检测阈值为

$$d = \widetilde{\Theta}\left(k^2 \vee \frac{k^6}{n^3}\right).$$

另一方面，定理2.5 中的阈值基本不依赖于 p。近期研究中的一个共同关键特征是 [BBN20、LMSY22、LR23b] 在考虑 $p = o(1)$ 时，利用 KL 散度的张量化性质（即链式法则），从而能够对模型进行“局部”比较；更多细节请参见第 3 节的技术概述。我们的下一个结果针对我们的设定对这类方法进行了改进；此外，该设定还具有社区结构。

定理 2.6 （通过张量化得到的下界）。 存在一个常数 $C_{2.6} > 0$，使得若满足以下条件：

$$\frac{C_{2.6} \log n}{k} \le p \le \frac{1}{2} \quad \text{and} \quad d \ge C_{2.6}\left(k^2 p^2 \vee \frac{k^4 p^2}{n}\right) \log^2(k/p) \log^2(1/p) \log^3 n,$$

没有任何检验能够实现弱检测。

关于对 p 的依赖，我们得到阈值中的一个多项式因子 p^2；这与 $k = n$时的最先进结果相匹配并将其推广，具体而言，这些结果由 $d = \widetilde{\Omega}(n^3 p^2)$ [LMSY22、BGPS25] 给出。

注记 2.7 （下界的比较）。 $p \ge$ 对于下界，我们关注 $O(1/k)$ 的区域；对于稀疏区域 $p = \Theta(1/k)$，$d_{\mathrm{TV}}(\mathcal{P}, \mathcal{Q}) = o(1)$ 对 $d = \Omega(\mathrm{polylog}(k))$ 已经成立

[3]如果 $k > n/5$，则没有 即使已知社区位置，若 $d \gg k^3 = \Theta(n^3)$，检验也能实现弱检测。

[LMSY22]。我们不打算研究 $p = o(1/k)$ 的情形，因为该处社区内的平均度已经是 $o(1)$。

从对 k/n 的依赖来看，定理 *2.6* 并未严格推广定理 *2.5*。事实上可以验证，视 k 的大小而定，定理 *2.5* 适用的范围更广：具体而言，当 $n^{1/2} \le k = \widetilde{O}(n^{3/4})$ 且 $p = \widetilde{\Omega}(\sqrt{n}/k)$，或 $k = \widetilde{\Omega}(n^{3/4})$ 且 $p = \widetilde{\Omega}(k/n)$ 时皆是如此。这主要源于两者底层方法的差异；详细讨论见第 *3* 节。简言之，定理 *2.5* 的证明本质上是通过 χ^2 散度来界定全变差距离，这似乎是精确刻画对社群规模依赖关系的关键。这个方法无法直接用于定理 *2.6* 的证明：这样做会失去 *KL* 散度的链式法则结构，而这种结构对于所有现有的、能够刻画对 p 依赖关系的方法都至关重要。我们认为，若要改进 k/n 对 $p = o(1)$ 的依赖关系，需要提出全新的思路；我们将此留作一个开放问题。

2.3 计算下界

尽管全局有符号三角形计数显然可以在多项式时间内计算，但基于扫描的检验通常似乎需要通过暴力算法耗费超多项式时间。这表明我们的检测问题中存在计算—统计差距；我们主张确实如此。

我们的分析基于低次多项式框架 [Hop18、KWB22、Wei25], ，该框架考察将以下多项式作为检验统计量的判据。

定义 2.8。令 f 为一个多项式。检验统计量 $f(G)$ 达到

(a) strong separation *if* $\mathbb{E}_{G\sim\mathcal{P}}[f(G)] - \mathbb{E}_{G\sim\mathcal{Q}}[f(G)] = \omega\left(\sqrt{\mathrm{Var}_{G\sim\mathcal{Q}}[f(G)]} \vee \sqrt{\mathrm{Var}_{G\sim\mathcal{P}}[f(G)]}\right)$;

(b) weak separation *if* $\mathbb{E}_{G\sim\mathcal{P}}[f(G)] - \mathbb{E}_{G\sim\mathcal{Q}}[f(G)] = \Omega\left(\sqrt{\mathrm{Var}_{G\sim\mathcal{Q}}[f(G)]} \vee \sqrt{\mathrm{Var}_{G\sim\mathcal{P}}[f(G)]}\right)$.

在低次多项式框架中，对于次数 $\omega(\log n)$ 的该判据，负面结果可作为不存在多项式时间算法的证据（有关背景，请参见第 1.2 节）。

回顾定理 2.1 可知，对于 $d = \widetilde{o}(k^6p^3/n^3)$，检测可以高效完成。以下结果与之互补，表明即使通过弱分离，也不存在低次多项式能够显著改善该阈值。

定理 2.9（低次下界）。假设存在一个常数 $\delta > 0$，使得$d \ge n^{\delta}, n^{-1+\delta} \le p \le 1/2$。若存在任意常数 $\varepsilon > 0$，使得

$$d \ge \frac{k^6}{n^{3-\varepsilon}}p^3\,,$$

不存在任何次数为-$\lfloor (\log n/\log(\log n))^2 \rfloor$ 的多项式能够实现弱分离。

一个相关问题是，除了全局有符号三角形计数之外，是否还存在其他高效算法。有符号三角形计数的一种自然推广是带符号环计数这一类方法，它们经常出现在潜在几何检测中 [BB24a, BB25a]。在下面的命题中，我们给出否定的回答，表明任何更长环的计数都严格弱于三角形计数。

命题 2.10 (较长环计数的次优性)。设 $\ell \ge 3$ 和 d 足够大，且 $d \ge (5\log(1/p))^4$。如果长度为 ℓ 的环的全局带符号计数实现强分离，那么

$$d \ll \left(\frac{k^2p\log(1/p)}{n}\right)^{\ell/(\ell-2)}.$$

在此命题中，带符号三角形计数在最大的 d范围内有效，因为右端在 $\ell = 3$处达到最大值。这表明，全局有符号三角形计数可能是渐近最优高效检验。

3 技术概述

信息论上界。 随机几何图的一个关键特征是同质性：相邻顶点的潜在向量相似，因此它们的公共邻居彼此也更有可能相邻。因此，与边密度相同的 Erd˝os–R´enyi 图相比，几何图包含更多三角形。我们的检验统计量基于带符号三角形计数$\sum_{i<j<\ell}(G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)$，并通过对每条边的指示变量进行中心化，进一步利用同质性。具体而言， $\mathcal{P}$ 下的均值能准确反映几何信号，而其方差仍然很小，因为中心化抵消了 Erd˝os–R´enyi 背景中的冗余贡献 [BDER16,LMSY22]。此外，在我们的设定中，几何结构是局部的，仅限于一个平均大小为 k 的隐藏社区。综上，这些特点促使我们设计了一系列三个检验，每个检验都针对比前一个更宽广的参数区域。

全局统计量 f_{tri} (2.1) 对整个图中的带符号三角形求和。在 $\mathcal{P}$ 下，其均值严格为正，这是因为社区内部存在额外的三角形（引理4.1和 4.3），而方差则主要由整个图上的 Erd˝os–R´enyi 波动决定。作为一条技术性说明，要验证后者，需要界定 $\mathcal{P}$下有符号四环计数的均值；为此，我们使用本文提出的适用于任意长度环的一般界。详情见本节最后一段。通过切比雪夫不等式比较均值与方差，我们得到定理 $d \ll k^6p^3\log^3(1/p)/n^3$ 的检测阈值 2.1。

全局统计量的一个局限是， k 必须足够大（例如，是 n的多项式级别），检验才能成功；当社区规模较小时，基于几何结构的信号已经被背景 Erd˝os–R´enyi 噪声稀释。为解决这一问题，我们引入扫描统计量 f_{scan} (2.2)，它对所有大小为 $k^- \approx k$的子集中的带符号三角形计数取最大值。在 $\mathcal{P}$下，通过局部应用前面相同的切比雪夫论证，植入社区以高概率具有较大的带符号三角形计数。新的挑战在于分析 $\mathcal{Q}$下的错误：证明在 $\mathcal{Q}$ 下，没有任何子集的带符号三角形计数会异常大。为了压过并集界带来的 $\binom{n}{k^-}$ 因子，我们使用一种针对次高斯变量的多项式的强集中不等式（引理A.1）。综合以上结果，我们得到定理 $d = \widetilde{O}(k^2p^6)$ 的阈值 2.2。如前所述，对于任意固定的 p，全局检验与扫描检验共同刻画了最优检测阈值，精确到一个对数因子。

然而，当 p 趋于零时，扫描检验还不够。这是因为 Erd˝os–R´enyi 背景可能通过密集波动产生较大的带符号三角形计数，而无需存在任何几何结构。这不仅是我们分析中的一个假象，而是一个根本性障碍；参见评注 4.6。我们的方法是将那些异常子集（其行为大致类似于团）刻画为具有较大的有符号楔形和 $W_{ij} := \sum_{\ell<i}(G_{\ell i}-p)(G_{\ell j}-p)$，且这一性质对许多对都成立。具体而言，对于约束扫描统计量$\sum_{i<j} W_{ij}^2$ 和 $\max_{i<j}|W_{ij}|$ (见 (2.6) 和 (2.7)) 的约束 $\sim f_{\mathrm{scan}}$ (2.5) 发挥着不同作用：前者在伯恩斯坦型集中论证中控制方差，而后者则筛除前述会产生较大有符号楔形计数的密集类团区域。

这些论证中的一个关键技术步骤，是验证在 $\mathcal{P}$下，植入社区以高概率属于约束集 $\mathcal{C}(G)$（定理4.5 ）。这相当于证明社区内的有符号楔形计数集中于其典型值附近。对于 $\sum_{i<j} W_{ij}^2$ 上的约束（见（2.6）），这是通过对均值和

由这些楔形形成的小子图的方差进行细致控制来实现的；对于 $\max_{i<j}|W_{ij}|$ 上的约束（见（2.7）），则通过对楔形建立对数阶矩界来实现。其余论证在概念上与扫描检验的论证类似，并得到定理 $d = \widetilde{O}(k^2p^3)$ 中的阈值 2.3，当 $p = \widetilde{o}(1)$ 时，该阈值严格优于扫描检验。

信息论下界。 要区分 $\mathcal{P}$ 和 $\mathcal{Q}$并建立下界要精细得多，这是因为随机几何图中的边之间存在很强的依赖关系。即使在经典模型 $\mathcal{G}(n,p,d)$ 中，最优检测阈值仍未解决：目前已知的最佳下界与推测的最优上界 p 之间存在多项式级差距 [LMSY22]。在我们的局部化设定中——它严格更为一般——我们推导出的下界与当前在较简单模型中可得的最强结果相匹配，并将其推广。我们从两种互补的方式着手证明下界（见注 2.7），分别刻画检测阈值的不同方面。

第一种方法刻画了对平均社区规模 k 的依赖，并且在稠密区域尤其有效。对于具有隐藏结构的模型，下界通常通过似然比的二阶矩（即 Ingster–Suslina 方法 [IS03]）来获得。在我们的设定中，将 $\mathcal{P}$ 和 $\mathcal{Q}$ 视为经典随机矩阵系综——即 Wishart 分布和高斯正交系综（GOE）——的推前后，这一计算便变得可行。此前已有研究比较过这些系综 [BDER16、BG18、RR19]，但这些比较是在全变差距离或 KL 散度层面进行的，其强度不及本文采用的 χ^2-散度。此外，在我们的模型中，直接使用二阶矩论证会失效，因为稀有构型的贡献占据主导。我们的分析基于对谱进行截断的论证，只保留相关的主体行为；随后，二阶矩便可由两个独立同分布社区之间重叠的一个易处理函数来界定。

这种方法面临的另一个挑战来自社区结构：需要比较 Wishart 分布与其球面变体 [PW21]。这是因为社区边是通过对 Wishart 矩阵的归一化版本进行阈值化而生成的。这里的比较颇为微妙：只应考虑严格上三角元素，[4] 而某些朴素的比较方法严格次优。[5] 通过直接分析它们各自的密度及谱性质，我们证明这两个分布在 $d \gg k^2$（命题 5.1）下不可区分。值得注意的是，这揭示了一个区域：这两个分布渐近等价，但与高斯正交系综不同——这一结果本身或许也具有独立意义。结合前文的二阶矩分析，这种更精细的矩阵比较得出了稠密区域内最优的检测下界（定理 2.5），且仅相差一个对数因子。

第二种方法刻画了对边密度的依赖 p，而第一个下界本质上独立于 p。遵循 [LMSY22、LR23b] 针对完整模型（即$k = n$）所采用的策略，我们的出发点是将随机图视为通过一次揭示一个顶点而生成，然后利用 KL 散度的张量化性质来界定全变差距离。具体而言，我们将先前的策略扩展到当前设定，把社区成员资格的判定纳入序贯过程：对每个顶点的潜变量进行扩充，使其同时编码该顶点的特征向量 U_i 和社区成员指示变量 V_i。这一方法的关键优势来自平均情形结构：每个顶点以 k/n的概率属于该社区；在给定社区成员资格的条件下，只有 $O(k$个先前揭示的顶点与之相关。

[4]Wishart 矩阵的对角元素具有连续密度，而球形 Wishart 矩阵的对角元素则服从狄拉克测度，因此两者之间的全变差距离显然等于 1。[5]例如，逐元素比较两个矩阵并使用并集界需要 $d \gg k^4$，而将这两个矩阵分别与高斯正交系综比较则需要 $d \gg k^3$ [JL15、BDER16]。

通过精细的截断论证刻画这些特征，从而得到 p依赖于2.6 定理中的下界，该下界推广了 $k = n$ 情形。

一个重要的技术步骤是界定在 $\mathcal{P}$ 与 $\mathcal{Q}$ 之间的邻域分布的平方似然比。此前，这一步仅在 $d = \tilde{\Omega}(k^3p^3)$ 的条件下完成，这足以处理 $k = n$ 情形，但不足以应对我们更一般的设定。通过新的分析，我们将这一条件改进为 $d = \tilde{\Omega}(k^2p^2)$。主要区别在于利用似然比鞅结构的策略。借助每个鞅差的集中性，我们构造了一对递归不等式，直接刻画二阶矩；与仅依据鞅本身的集中性得到的界相比，这个界更紧 [LMSY22]。

计算下界。基于低次多项式框架 [Hop18]，我们提供证据表明，任何多项式时间算法都无法在全局检验所适用的范围之外取得成功；背景介绍请参见第 1.2 节和第 2.3 节。在我们的设定中，控制所有低次多项式可归结为对所有多对数级大小子图 H 的平方傅里叶系数 $\Phi_{\mathcal{P}}(H)^2$ 求和并加以界定。我们的分析基于两个关键思路：对于森林，其贡献由推论 4.2 得出为零；对于其余子图，通用矩界（见 [BB24b]；引理 A.7）给出一个仅依赖于 $v(H)$ 和 $e(H)$ 的上界。对所有可能的 $(v(H), e(H))$ 组合求和，即可得到该结果（定理 2.9）。

一个相关的问题是，除了全局有符号三角形计数之外，是否存在其他高效检验能够达到同样的性能。我们证明，任何更长的带符号环计数表现都严格更差（命题 2.10）。这一结论源于对任意长度的带符号环计数期望的紧确估计（命题 6.1）；这一结果本身可能也值得关注。此前，我们曾利用该结果对 4-环进行分析，以研究我们的检验。主要思路是将阈值函数 $\mathbf{1}\{\langle U_i, U_j\rangle \geq \tau(p,d)\}$ 在球谐函数与 Gegenbauer 多项式构成的标准正交基上展开。根据正交性，带符号环计数的期望可化简为无穷和 $\sum_{m=1}^{\infty} c_m^{\ell}/N_m^{\ell/2-1}$，其中 c_m 是阈值函数的 Gegenbauer 系数，而N_m 是球谐函数的重数。该无穷和由 $m = 1$ 项主导，从而得到紧确估计 $p^{\ell}\log^{\ell/2}(1/p)/d^{\ell/2-1}$，相差乘法因子 $(1/C)^{\ell}$ 以及一个常数 C^{ℓ}。利用这一结果比较带符号环计数的均值与方差，我们证明 $\ell = 3$ 情形涵盖了最宽的范围，这表明在所有带符号环计数中，全局有符号三角形计数是渐近最优检验。

4 信息论上界的证明

本节将证明第 2.1 节给出的上界。

在给出证明之前，我们先回顾模型 $\mathcal{P} = \mathcal{G}(n,\ p,\ d,\ k)$ 的一些基本事实。首先，我们可以将 $\mathcal{P}$ 写成如下混合分布

$$\mathcal{P} = \mathbb{E}_S[\mathcal{P}_S]\,, \tag{4.1}$$

其中 对于每个固定的 $S \subseteq [n]$，$\mathcal{P}_S$ 都是按照定义 1 中的步骤 3 抽取的分布。 1.

与其抽取大小可变的社区二项分布$(n,\ k/n)$，通常考虑固定大小变体会更有用。对于整数 $0 \leq s \leq n$，定义

$$\mathcal{P}'_s := \mathbb{E}_{S\big||S|=s}[\mathcal{P}_S]\,. \tag{4.2}$$

注意，在对 $|S| = s$ 进行条件化后，S 是从 $\binom{[n]}{s}$ 中均匀随机选取的。大小 s 可以

在以下意义下被认为约为 k：对于

$$k^- := \lfloor 0.9k \rfloor, \quad k^+ := \lceil 1.1k \rceil,$$

任意仅依赖于 E G 的事件都满足

$$\begin{aligned} \mathbb{P}_{G\sim\mathcal{P}}(E) &\le \mathbb{P}(|S| \notin [k^-, k^+]) + \max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}(E) \\ &\le 2\exp(-\Omega(k)) + \max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}(E)\,, \end{aligned} \tag{4.3}$$

第二个不等式由应用切尔诺夫界得到。

在我们的分析中，关键量是带符号子图计数的期望。对于 Erd˝os–R´enyi 图 $\mathcal{Q}$，由于独立性，这些期望始终为 0。对于 $\mathcal{P} = \mathcal{G}(n,\ p,\ d,\ k)$，我们首先注意到，它们与 $\mathcal{G}(n,\ p,\ d)$ 下的期望有如下直接关系。

引理 *4.1*。 *</style}*设 H 是 K_n 的一个不含孤立顶点的子图。那么，

$$\mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right] = \left(\frac{k}{n}\right)^{v(H)} \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right].$$

证明。 我们有

$$\mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right] = \mathbb{E}_S\left[\mathbb{E}_{G\sim\mathcal{P}_S}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right]\right].$$

如果 $V(H) \not\subseteq S$，则内层期望为 0，因为任意一条离开 S 的边对应的中心化因子均值为 0，且与其余部分独立。如果 $V(H) \subseteq S$，则内层期望等于 $\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[\prod_{ij\in E(H)}(G_{ij}-p)]$，因为 $\mathcal{P}_S$ 在 $V(H)$ 上与 $\mathcal{G}(n,\ p,\ d)$ 一致。由于 $\mathbb{P}(V(H) \subseteq S) = (k/n)^{v(H)}$，结论得证。 □

推论 4.2。 设 H 是 K_n 的一个子图，且该子图是森林。那么 $\mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right] = 0$，或者等价地，$\mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{ij\in E(H)} G_{ij}\right] = p^{e(H)}$。

在不同的子图中，三角形（K_3）在我们的分析中起着核心作用。虽然在 $\mathcal{Q}$ 下，带符号三角形的期望为 0[6]，但在 $\mathcal{G}(n,p,d)$ 下严格为正，因此在 $\mathcal{P}$ 下也如此。

引理 4.3 ([BDER16，引理 1]和 [LMSY22，断言 A.2])。 存在一个普适常数$C_{4.3} > 0$，使得对于所有充分大的 d 和 $0 < p \le 1/2$,

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[(G_{12}-p)(G_{23}-p)(G_{13}-p)\right] \ge \frac{p^3\log^{3/2}(1/p)}{C_{4.3}\sqrt{d}}.$$

4.1 全局检验

在本小节中，我们研究全局检验，并证明定理2.1。

回忆一下 $f_{\mathrm{tri}}(G) = \sum_{i<j<\ell\in[n]}(G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)$ 的定义。我们将计算 $f_{\mathrm{tri}}(G)$ 在 $\mathcal{P}$ 和 $\mathcal{Q}$ 下的一阶矩和二阶矩，以便使用切比雪夫不等式。直接计算可得（例如，参见 [LMSY22，引理 A.1]）

$$\begin{aligned} \mathbb{E}_{G\sim\mathcal{Q}}[f_{\mathrm{tri}}(G)] &= 0\,, \\ \mathrm{Var}_{G\sim\mathcal{Q}}[f_{\mathrm{tri}}(G)] &= \binom{n}{3}p^3(1-p)^3\,. \end{aligned}$$

[6]该 r 引理中的引用 4.3 假设 $p \in (0,0.49)$；对于 $p \in [0.49, 0.5]$，可以应用命题 6.1 .

第一类错误。我们有

$$\mathbb{P}_{G\sim\mathcal{Q}}(f_{\mathrm{tri}}(G)>\gamma_{\mathrm{tri}})\le\frac{\mathrm{Var}_{G\sim\mathcal{Q}}[f_{\mathrm{tri}}(G)]}{\gamma_{\mathrm{tri}}^2},$$

由引理4.1 和 4.3,

$$\gamma_{\mathrm{tri}}=\frac{1}{2}\mathbb{E}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]\ge\frac{1}{2}\times\binom{n}{3}\left(\frac{k}{n}\right)^3\frac{p^3\log^{3/2}(1/p)}{C_{4.3}\sqrt{d}}=\Omega\left(\frac{k^3p^3\log^{3/2}(1/p)}{\sqrt{d}}\right).$$

因此,

$$\frac{\mathrm{Var}_{G\sim\mathcal{Q}}[f_{\mathrm{tri}}(G)]}{\gamma_{\mathrm{tri}}^2}=O\left(\frac{dn^3}{k^6p^3\log^3(1/p)}\right),$$

其中最后一项为 $o(1)$，当 $d\ll\frac{k^6p^3}{n^3}\log^3(1/p)$ 时。

第二类错误。与第一类错误类似，由 $2\gamma_{\mathrm{tri}}=\mathbb{E}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]$ 可得

$$\mathbb{P}_{G\sim\mathcal{P}}(f_{\mathrm{tri}}(G)\le\gamma_{\mathrm{tri}})=\mathbb{P}_{G\sim\mathcal{P}}(f_{\mathrm{tri}}(G)-\mathbb{E}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]\le-\gamma_{\mathrm{tri}})\le\frac{\mathrm{Var}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]}{\gamma_{\mathrm{tri}}^2}.$$

因此，只需控制 $\mathcal{P}$ 下全局有符号三角形计数的方差。我们将其表述为一个引理，因为后文还会用到。

引理 4.4。若 $1/k\le p\le1/2$ 和 d 充分大，且 $(5\log(1/p))^4\le d\ll k^6p^3\log^3(1/p)/n^3$，则

$$\gamma_{\mathrm{tri}}\gg\sqrt{\mathrm{Var}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]}.$$

引理 4.4 的证明见附录B.1。根据该引理，第二类错误为 $o(1)$，至此证明完毕。

4.2 扫描检验

在此，我们分析扫描检验的性能，并证明定理2.2。

第一类错误。对于 $A\subseteq[n]$，令 $f_A(G):=\sum_{i<j<\ell\in A}(G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)$。那么,

$$\begin{aligned}
\mathbb{P}_{G\sim\mathcal{Q}}(f_{\mathrm{scan}}(G)>\gamma_{\mathrm{scan}})&=\mathbb{P}_{G\sim\mathcal{Q}}\left(\max_{A\in\binom{[n]}{k^-}}f_A(G)>\gamma_{\mathrm{scan}}\right)\\
&\le\binom{n}{k^-}\mathbb{P}_{G\sim\mathcal{Q}}\left(f_{[k^-]}(G)>\gamma_{\mathrm{scan}}\right)\\
&\le n^{k^-}\cdot2\exp\left(-\frac{1}{C_{A.1}}\left(\frac{\gamma_{\mathrm{scan}}^2\log^3(1/p)}{(k^-)^3}\wedge\frac{\gamma_{\mathrm{scan}}\log^{3/2}(1/p)}{\sqrt{k^-}}\wedge\gamma_{\mathrm{scan}}^{2/3}\log(1/p)\right)\right)\\
&\le\exp\left(k\log n-\Omega\left(\frac{k^3p^6\log^6(1/p)}{d}\wedge\frac{k^{5/2}p^3\log^3(1/p)}{\sqrt{d}}\wedge\frac{k^2p^2\log^2(1/p)}{d^{1/3}}\right)\right)\\
&\le\exp\left(k\log n-\Omega\left(\frac{k^3p^6\log^6(1/p)}{d}\wedge\frac{k^2p^2\log^2(1/p)}{d^{1/3}}\right)\right),
\end{aligned}$$

其中，第一个不等式由并集界得出，第二个不等式由引理A.1得出，第三个不等式由引理4.3得出，最后一个不等式则根据假设由 $d\le k^3p^6\log^6(1/p)$ 得出。

- 如果 $d \geq k^{3/2}p^6\log^6(1/p)$，则第一项 $k^3p^6\log^6(1/p)\,d$ 是最小值。于是，若 $d \leq \frac{k^2p^6}{\log n}\log^6(1/p)\,C_0$，通过选择足够大的 $C_0 > 0$，有 $\mathbb{P}_{G\sim\mathcal{Q}}(f_{\text{扫描}}(G) > \gamma_{\text{扫描}}) \leq \exp(-k\log n) = o(1)$。

- 如果 $d < k^{3/2}p^6\log^6(1/p)$，第二项 $k^2p^2\log^2(1/p)\,d^{1/3}$ 是最小值。那么，$\mathbb{P}_{G\sim\mathcal{Q}}(f\text{scan}(G) > \gamma\text{ scan}) \leq \exp(-k\log n) = o(1)$，只要 $d \leq \frac{k^3p^6}{\log^6(1/p)}\,C_1\log^3 n$，取足够大的 $C_1 > 0$即可。对于任意常数 $C_{2.2} > C_1/C_0$，给定 $k \geq C_{2.2}\log^2 n$，这可由前面的不等式推出。

因此，任意满足 $C_{2.2} > (C_1/C_0) \vee C_0 \vee C_1$ 的取值，均可使 $\mathbb{P}_{G\sim\mathcal{Q}}(f_{\text{scan}}(G) > \gamma_{\text{scan}}) = o(1)$ 成立。

第二类错误。 根据（4.3）及事件 $E = \{f_{\text{scan}}(G) \leq \gamma_{\text{scan}}\}$，只需证明

$$\max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}\left(\max_{A\in\binom{[n]}{k^-}} f_A(G) \leq \gamma_{\text{scan}}\right) = o(1)\,,$$

其中，$\mathcal{P}'_s$ 是由 $\mathcal{P}_S$ 构成的混合分布，S 在所有大小为s 的集合中均匀选取。对于任意这样的 S，都存在一个大小恰为 k^-的子集 $S_0 \subseteq S$。考虑任意一种确定性规则来选择这样的 S_0（例如，取 S 中最小的 k^- 个元素）。由于所有子集上的最大值至少不小于这个特定S_0上的取值，因此我们有 $\max_{A\in\binom{[n]}{k^-}} f_A(G) \geq f_{S_0}(G)$。于是，

$$\begin{aligned}\max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}(E) &\leq \max_{|S|\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}_S}(f_{S_0}(G) \leq \gamma_{\text{scan}}) \\ &= \max_{|S|\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}_S}(f_{S_0}(G) \leq \mathbb{E}_{G\sim\mathcal{P}_S}[f_{S_0}(G)] - \gamma_{\text{scan}}) \\ &\leq \frac{\text{Var}_{G\sim\mathcal{P}_{S_0}}[f_{S_0}(G)]}{\gamma^2_{\text{scan}}}\,,\end{aligned} \tag{4.4}$$

由 $\mathbb{E}_{G\sim\mathcal{P}_S}[f_{S_0}(G)] = 2\gamma_{\text{scan}}$ 以及切比雪夫不等式。对于（4.4）的最后一项为 $o(1)$ 时，我们应用引理 4.4 于 f_{S_0}，将 $G[S_0]$ 视为一个大小为 k^- 且具有完整社区的独立图。

根据定理的假设，引理 4.4 所需的条件均成立。d的下界显然成立。条件 $p \geq 1/k^-$由 $1 \leq d = O(k^2p^6\log^6(1/p)/\log n) \leq O(k^2p^2/\log n)$ 得出，这意味着 $p = \Omega(\sqrt{\log n/k})$；条件 $d \ll (k^-)^3p^3\log^3(1/p)$ 则由 $k \gg 1/\log n = \Omega(p^3\log^3(1/p)/\log n)$ 得出，这意味着 $d = O(k^2p^6\log^6(1/p)/\log n) \ll k^3p^3\log^3(1/p)$。

4.3 约束扫描检验

最后，我们考虑约束扫描检验，并证明定理 2.3。

回顾（见上一小节）的记号 $f_A(G) = \sum_{i<j<\ell\in A}(G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)$，其中 $A \subseteq [n]$。这里我们首先分析第二类错误，因为其证明在概念上更简单。

第二类错误。 我们采用与（4.4）类似的方法；此处的区别在于新的错误项甚至 t

$$E = \left\{\widetilde{f}_{\text{scan}}(G) \leq \gamma_{\text{scan}}\right\} = \left\{\max\left\{f_A(G) : A \in \binom{[n]}{k^-}, (2.6) \text{ and } (2.7) \text{ hold}\right\} \leq \gamma_{\text{scan}}\right\},$$

并且由（4.3）可知，只需证明

$$\max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}(E) = o(1)\,.$$

对于每个 S 其中 $G \sim \mathcal{P}_S$，考虑一个子集 $S_0 \subseteq S$，其大小为 k^-，并由确定性规则选取。这里的关键论断是，对于 $A = S_0$，（2.6）和（2.7以高概率成立，这意味着在该事件下 $\widetilde{}$有 $f_{\text{scan}}(G) \geq f_{S_0}(G)$。从概念上说，这是因为条件方差（（2.6）的左边）和最大差值（（2.7）的左边）在植入社区 S_0内集中于其期望附近，而参数 σ^2 和 B 的选取旨在对这些典型值加以界定。

引理 4.5。设 $S_0 \subseteq S \subseteq [n]$ 满足 $|S| \in [k^-, k^+]$ 且 $|S_0| = k^-$。此外，设 $E' = E'(S_0)$ 为满足以下条件的事件：(2.6) 和 (2.7) 对 $A = S_0$成立。那么在定理 *2.3*的假设下，

$$\mathbb{P}_{G\sim\mathcal{P}_S}(E') = 1 - o(1)\,.$$

其中 $o(1)$项对 S_0 和 S的所有取值均为通用项。

证明推迟至附录 B.2。根据引理 4.5，我们有

$$\begin{aligned}\max_{s\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}'_s}(E) &\leq \max_{|S|\in[k^-,k^+]} \left(\mathbb{P}_{G\sim\mathcal{P}_S}(f_{S_0}(G) \leq \gamma_{\text{scan}}) + \mathbb{P}_{G\sim\mathcal{P}_S}\left((E')^c\right)\right)\\ &\leq \max_{|S|\in[k^-,k^+]} \mathbb{P}_{G\sim\mathcal{P}_S}(f_{S_0}(G) \leq \mathbb{E}_{G\sim\mathcal{P}_S}[f_{S_0}(G)] - \gamma_{\text{scan}}) + o(1)\\ &\leq \mathrm{Var}_{G\sim\mathcal{P}_S}[f_{S_0}(G)]/\gamma_{\text{scan}}^2 + o(1)\,,\end{aligned}$$

沿用（4.4）中的论证思路。特别地，（4.4）之后的相同论证表明，只要 $d \ll k^3p^3\log^3(1/p)$，且 d 足够大，并满足$d \geq (5\log(1/p))^4$，最后一项就会消失；这些条件已由定理对 d的假设保证。

第一类错误。对于 $\mathcal{Q}$下的错误，我们有

$$\begin{aligned}\mathbb{P}_{G\sim\mathcal{Q}}\left(\widetilde{f}_{\text{scan}}(G) > \gamma_{\text{scan}}\right) \leq \binom{n}{k^-}\mathbb{P}_{G\sim\mathcal{Q}}\Bigg(&f_{[k^-]}(G) > \gamma_{\text{scan}},\\ &\sum_{i<j\in[k^-]}\left(\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right)^2 \leq \sigma^2,\\ &\max_{i<j\in[k^-]}\left|\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right| \leq B\Bigg)\,,\end{aligned} \tag{4.5}$$

由并集界和对称性（令 $A = [k^-]$）。

现在我们构造一个鞅。首先，令 $\{\mathcal{F}_r\}$ 为通过按以下顺序逐条揭示边而定义的滤过

$$G_{12}, G_{13}, G_{23}, G_{14}, G_{24}, G_{34}, \ldots, G_{(k^-)-1,k^-}\,.$$

此外，令 $\{M_r\}$ 为关于 $\{\mathcal{F}_r\}$的 $M_r = \mathbb{E}_{G\sim\mathcal{Q}}[f_{[k^-]}(G)|\mathcal{F}_r]$的 Doob 鞅 $f_{[k^-]}(G)$。考虑揭示具有 $i < j$ 的边 ij 的时刻 r。那么

$$U_r := \left|\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right|$$

是 $\mathcal{F}_{r-1}$-可测的，并且是鞅差的上界，即

$$M_r - M_{r-1} = (G_{ij} - p) \sum_{\ell<i, \ell\in[k^-]} (G_{\ell i} - p)(G_{\ell j} - p) \le U_r \,. \tag{4.6}$$

此外，鞅差的条件方差等于

$$\begin{aligned} \mathrm{Var}_{G\sim\mathcal{Q}}[M_r - M_{r-1}|\mathcal{F}_{r-1}] &= \mathbb{E}_{G\sim\mathcal{Q}}[(G_{ij}-p)^2] \left(\sum_{\ell<i, \ell\in[k^-]} (G_{\ell i} - p)(G_{\ell j} - p) \right)^2 \\ &= p(1-p) \left(\sum_{\ell<i, \ell\in[k^-]} (G_{\ell i} - p)(G_{\ell j} - p) \right)^2 . \end{aligned} \tag{4.7}$$

由（4.6）和（4.7），我们可以应用 Freedman 不等式（引理 A.3），参数为

$$\begin{aligned} t = \gamma &= \Theta(k^3p^3 \log^{3/2}(1/p)/\sqrt{d}) \,, \\ v &= p(1-p)\sigma^2 = \Theta(k^3p^3 + k^4p^5 \log^2(1/p)/d) \,, \\ m = B &= \Theta((kp^2+1)\log k) \,. \end{aligned}$$

回想一下， t的选取来自引理 4.3，而 v 和 m 分别来自（2.6）和（2.7）；关于这些选取的进一步讨论，请参见注4.6。由（4.5），

$$\begin{aligned} &\mathbb{P}_{G\sim\mathcal{Q}} \left(\widetilde{f}_{\mathrm{scan}}(G) > \gamma_{\mathrm{scan}} \right) \\ &\le \exp\left(k\log n - \left(\frac{\gamma_{\mathrm{scan}}^2}{4v} \wedge \frac{\gamma_{\mathrm{scan}}}{2B} \right) \right) \\ &\le \exp\left(k\log n - \Omega\left(\frac{k^3p^3\log^3(1/p)}{d} \wedge k^2p\log(1/p) \wedge \frac{k^2p\log^{3/2}(1/p)}{\sqrt{d}\log k} \wedge \frac{k^3p^3\log^{3/2}(1/p)}{\sqrt{d}\log k} \right) \right) . \end{aligned}$$

通过适当选取常数 $C_0 > 0$，使得下式成立，该项为 $o(1)$：

$$d \le \frac{1}{C_0} \left(\frac{k^2p^3\log^3(1/p)}{\log n} \wedge \frac{k^2p^2\log^3(1/p)}{(\log^2 n)(\log^2 k)} \wedge \frac{k^4p^6\log^3(1/p)}{(\log^2 n)(\log^2 k)} \right) , \tag{4.8}$$

$$kp\log(1/p) \ge C_0 \log n \,. \tag{4.9}$$

若 $d \le k^2p^3\log^3(1/p)/(C_0(\log^2 n)(\log^2 k))$，则由 $n^\delta \le d$ 和 $1/p \le n^{1-\delta}$ 可得 $k^2p^3 \ge C_0 n^\delta \log^2 k/((1-\delta)\log n)$，当 n 足够大时，该式大于 $(\log n)(\log^2 k)$。对于 (4.8)，这意味着第三项不可能是最小值，因此 $C_{2.3} > C_0$ 可任意选取。对于 (4.9)，我们有$k^3p^3\log^3(1/p) \ge k^2p^3\log^3(1/p) \ge C_0 n^\delta(\log^2 n)(\log^2 k)$，当 n 足够大时， $C_{2.3} > C_0^{1/3}$ 可任意选取。

注 *4.6*（约束的作用）。我们注意到，这类约束可能是必要的，因为无约束扫描检验（定理 *2.2*）的次优性似乎是固有的。这里我们给出一种解释；作为背景，我们参考定理 *2.2* 的证明。在那里，为了使错误在$\mathcal{Q}$下趋于零，带符号三角形计数大于 $t = \widetilde{\Theta}(k^3p^3/\sqrt{d})$ 的事件必须以极小的概率发生——至多为 $1/\binom{n}{k^-} = \exp(-\widetilde{\Omega}(k))$。然而，只要存在一个大小为 $t^{1/3}$ 的团，就能使该事件发生；而这种情况发生的概率至少为 $\exp(-\widetilde{O}(t^{2/3})) = \exp(-\widetilde{O}(k^2p^2/d^{1/3}))$（这也与该事件的大偏差型标度一致）。综合起来，这要求 $d = \widetilde{O}(k^3p^6)$，

这严格劣于我们在$p = \widetilde{o}(1/k^{1/3})$时为约束扫描检验得到的阈值 $d = \widetilde{O}(k^2 p^3)$ 。

从这个意义上说，约束可以被视为防止不利于集中性的事件。例如，可以验证，对于 $p = \widetilde{o}(1/k^{1/3})$，大小为 $t^{1/3}$ 的团意味着 (2.7) 左侧至少为 $\Omega(t^{1/3})$，当 $d = \widetilde{O}(k^2 p^3)$ 时，这违反了相应的条件。

5 信息论下界的证明

5.1 截断二阶矩

我们的第一个方法是将 $\mathcal{P}$ 和 $\mathcal{Q}$ 都视为通过对某些随机矩阵进行阈值化而生成的。这实质上将两个（二值）随机图 之间的检测问题转化为两个（实值）随机矩阵 [BDER16]之间的检测问题。

与（4.3）类似，我们从以下内容开始:

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{P}, \mathcal{Q}) &= d_{\mathrm{TV}}(\mathbb{E}_{s\sim \mathrm{Binom}(n,k/n)}[\mathcal{P}'_s], \mathcal{Q}) \\ &\le \mathbb{E}_{s\sim \mathrm{Binom}(n,k/n)}[d_{\mathrm{TV}}(\mathcal{P}'_s, \mathcal{Q})] \\ &\le \mathbb{P}_{s\sim \mathrm{Binom}(n,k/n)}(s \notin [k^-, k^+]) + \max_{s\in[k^-,k^+]} d_{\mathrm{TV}}(\mathcal{P}'_s, \mathcal{Q}) \\ &\le 2\exp(-\Omega(k)) + \max_{s\in[k^-,k^+]} d_{\mathrm{TV}}(\mathcal{P}'_s, \mathcal{Q})\,, \end{aligned}$$

其中，第一个不等式由詹森不等式得出，第二个不等式由 $d_{\mathrm{TV}} \le 1$ 得出，最后一个不等式由切尔诺夫界得出。因此，只需证明全变差距离在 $s \in [k^-, k^+]$ 上均匀趋于零。

从随机图到随机矩阵。现在我们来考虑生成 $\mathcal{P}'_s$ 和 $\mathcal{Q}$ 的随机矩阵。对于 $\mathcal{Q}$，令 $M \sim dI_n + \sqrt{d}\mathrm{GOE}(n)$，也就是说，对于任意 $i,\ j \in [n]$ 且 $i \le j$,

$$M_{ij} \overset{\text{ind.}}{\sim} d\mathbf{1}\{i=j\} + \sqrt{d}\mathcal{N}(0, 1+\mathbf{1}\{i=j\})\,,$$

并且 $M_{ji} = M_{ij}$。此外，定义映射 $\alpha:\ \mathbb{R}^{n(n+1)/2} \to \{0,1\}^{n(n-1)/2}$，它按如下方式对称矩阵 X 的非对角元素进行阈值化:

$$\alpha(X)_{ij} := \mathbf{1}\{X_{ij} \ge \sqrt{d}\Phi^{-1}(1-p)\}, \quad \text{for all } i<j\in[n]\,,$$

其中 Φ 是 $\mathcal{N}(0,\ 1)$ 的累积分布函数。于是我们有

$$\alpha(M) \sim \mathcal{Q} = \mathcal{G}(n,p)\,,$$

因为对于 $\alpha(M)$，每条边都是以概率 p独立生成的。对于 $\mathcal{P}$，随机矩阵应根据社群成员资格服从不同的分布。对于 $S \subseteq [n]$，定义一个对称随机矩阵 $W^S \in \mathbb{R}^{n\times n}$，其元素如下:

$$(W^S)_{ij} := \begin{cases} W_{ij} & i,j \in S \\ M_{ij} & \text{else}\,, \end{cases}$$

其中 $W \sim \mathrm{Wishart}(n,\ d)$，也就是说，对于任意 $i,\ j \in [n]$,

$$W_{ij} = \langle Z_i, Z_j\rangle\ ,$$

其中 $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0},\ I_d)$，适用于 $i \in [n]$。然后，通过对 $W_{ij}/\sqrt{W_{ii}W_{jj}} = \langle Z_i/\|Z_i\|,\ Z_j/\|Z_j\|\rangle$进行阈值化，得到 $\mathcal{P}$ 所属社区内的边 ij。具体而言，定义 $\beta\colon\ \mathbb{R}^{n(n+1)/2} \to \{0,1\}^{n(n-1)/2}$ 为

$$\beta(X)_{ij} := \mathbf{1}\left\{X_{ij}/\sqrt{X_{ii}X_{jj}} > \tau(p,d)\right\},\quad \text{for all } i<j\in[n]\,.$$

为了生成一个样本，社区内部的边应使用 β，其他情况则应使用 α。形式上，定义映射 $\beta^S\colon \mathbb{R}^{n(n+1)/2} \to \{0,1\}^{n(n-1)/2}$，使得

$$\beta^S(X)_{ij} = \beta(X)_{ij}\mathbf{1}\{i,j\in S\} + \alpha(X)_{ij}\mathbf{1}\{i\notin S \text{ or } j\notin S\}\,,$$

我们有

$$\beta^S(W^S) \sim \mathcal{P}_S\,.$$

从这个意义上说，稍微滥用一下记号，$\mathcal{P}'_s$ 可以写作 $\mathcal{P}'_s = \mathbb{E}_{S\big||S|=s}[\beta^S(W^S)],$，它是一个混合分布。于是

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{P}'_s,\mathcal{Q}) &= d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[\beta^S(W^S)],\alpha(M)\right)\\ &\le d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[\beta^S(W^S)],\mathbb{E}_{S\big||S|=s}[\alpha(W^S)]\right) + d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[\alpha(W^S)],\alpha(M)\right)\\ &\le \underbrace{d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[\beta^S(W^S)],\mathbb{E}_{S\big||S|=s}[\alpha(W^S)]\right)}_{\text{(I)}} + \underbrace{d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[W^S],M\right)}_{\text{(II)}}, \end{aligned} \tag{5.1}$$

其中，最后一个不等式是将数据处理不等式应用于映射 α 得到的；这是因为观察到 $\mathbb{E}_{S\big||S|=s}[\alpha(W^S)]$ 的分布（推前的混合分布）等于 $\alpha\left(\mathbb{E}_{S\big||S|=s}[W^S]\right)$（混合分布的推前）。

对于证明的其余部分，我们证明 (I) 和 (II) 在 $s\in[k^-,k^+]$ 上均匀趋于零.

第一项 (I)。利用凸性（以及詹森不等式）和对称性，(5.1) 中的 (I) 至多为

$$d_{\mathrm{TV}}\left(\beta^S(W^S),\alpha(W^S)\right) = d_{\mathrm{TV}}\left((\beta^{[s]}(W^{[s]})_{ij})_{i<j\in[s]},(\alpha(W^{[s]})_{ij})_{i<j\in[s]}\right), \tag{5.2}$$

(5.2) 的等价显式形式为

$$d_{\mathrm{TV}}\left(\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}>\tau(p,d)\right\}\right)_{i<j\in[s]},\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{d}>\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right\}\right)_{i<j\in[s]}\right). \tag{5.3}$$

根据三角不等式，(5.3) 可以进一步以上界表示为

$$\begin{aligned} &d_{\mathrm{TV}}\left(\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}>\tau(p,d)\right\}\right)_{i<j\in[s]},\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}>\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right\}\right)_{i<j\in[s]}\right)\\ &\quad + d_{\mathrm{TV}}\left(\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}>\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right\}\right)_{i<j\in[s]},\left(\mathbf{1}\left\{\frac{\langle Z_i,Z_j\rangle}{d}>\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right\}\right)_{i<j\in[s]}\right)\\ &\le s^2\mathbb{P}\left(\frac{\langle Z_1,Z_2\rangle}{\|Z_1\|\,\|Z_2\|}\text{ is between }\tau(p,d),\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right) + d_{\mathrm{TV}}\left(\left(\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}\right)_{i<j\in[s]},\left(\frac{\langle Z_i,Z_j\rangle}{d}\right)_{i<j\in[s]}\right), \end{aligned}$$

这里我们使用了耦合与并集界（对 $i<j\in[s]$），并对阈值函数应用了数据处理不等式。对于第一项，由于其概率至多为 $O(1/d)$，因此其上界为 $O(k^2/d)=o(1)$。为说明这一点，我们利用这样一个事实：经过适当缩放后，正态分布与球面边缘分布非常相似。具体而言，根据 [BDER16 引理 7]（原见 [Sod07 引理 1]），该概率至多为

$$\left|\mathbb{P}\left(\frac{\langle Z_1,Z_2\rangle}{\|Z_1\|\,\|Z_2\|}\le\tau(p,d)\right)-\mathbb{P}\left(\frac{\langle Z_1,Z_2\rangle}{\|Z_1\|\,\|Z_2\|}\le\frac{\Phi^{-1}(1-p)}{\sqrt{d}}\right)\right|$$
$$\le p(O(1/d)+1-\exp(-C_0\Phi^{-1}(1-p)^4/d))\,,$$

对于某个常数 $C_0>0$，假设 $\Phi^{-1}(1-p)/\sqrt{d}$ 足够小。最后一项至多为 $O(p\log^2(1/p)/d)$，由 $\Phi^{-1}(1-p)\lesssim\sqrt{\log(1/p)}$（由 $\mathbb{P}(\mathcal{N}(0,1)>x)\le\exp(-x^2/2)\ \text{for}\ x>0$）可得，这也意味着 $d=\Omega(\log(1/p))$ 足以满足关于 $\Phi^{-1}(1-p)/\sqrt{d}$ 的条件。

第二项是 Wishart 系综与“球面 Wishart”系综之间的全变差距离（ [PW21] 中的称谓），这两种分布都不是乘积分布，并且具有 Gram 矩阵结构。虽然可以通过分别将两个矩阵与高斯正交系综进行比较来处理，但这需要$d\gg k^3$ [JL15、BDER16]。通过直接比较两个矩阵的严格上三角部分，我们证明，当 $d\gg k^2$。

命题 5.1（Wishart 与球面 Wishart 的比较）。设 $Z_i\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,I_d)$ 对于$i\in[k]$。若 $d\gg k^2$ 且 $k\to\infty$，

$$d_{\mathrm{TV}}\left(\left(\frac{\langle Z_i,Z_j\rangle}{\|Z_i\|\,\|Z_j\|}\right)_{i<j\in[k]},\left(\frac{\langle Z_i,Z_j\rangle}{d}\right)_{i<j\in[k]}\right)=o(1)\,.$$

当 $s=\Theta(k)$ 时，命题 5.1 得出 $\mathrm{(I)}=o(1)$。命题 5.1 的证明推迟至附录 C.1。值得注意的是，该结果指出了另一个矩阵系综：即使矩阵元素之间仍存在依赖关系，它在这一条件下也收敛至 Wishart 分布。

注 5.2（文献中的相关结果）。近期有几项研究证明，对于若干种（变体）*Wishart* 分布，在低于$d=k^3$的范围内存在不可区分性结果。具体而言，已知某些变体的 *Wishart* 分布在 $d\gg k^2$时与高斯正交系综不可区分，例如 *[BBH21，推论 4。3]*，*[BB25a，定理 22]*。更相关的是，*[CW19，定理 1]* 给出了一个显式刻画：当 $k^2\ll d\ll k^3$时，有一种与 *Wishart* 不可区分的分布。该结果与我们的工作不能直接比较，因为我们考虑的是不含对角元素的 *Wishart* 分布。

第二项（II）。对于（5.1）中的（II），用似然比来表示总变差会更方便。设 $w_{\ell,d}$、$m_{\ell,d}$分别为 $\mathrm{Wishart}(\ell,d)$ 和 $dI_\ell+\sqrt{d}$ 高斯正交系综 (ℓ) 的密度。那么，$\mathbb{E}_{S\big||S|=s}[W^S]$ 和 M 的分布之间的似然比 L_s 满足

$$d_{\mathrm{TV}}\left(\mathbb{E}_{S\big||S|=s}[W^S],M\right)=\frac{1}{2}\mathbb{E}_{X\sim m_{n,d}}[|L_s(X)-1|]\,,\quad\text{where } L_s(X):=\mathbb{E}_{S\big||S|=s}\left[\frac{w_{s,d}(X_{S\times S})}{m_{s,d}(X_{S\times S})}\right]\,. \tag{5.4}$$

此处，我们用 $X_{A\times B}\in\mathbb{R}^{|A|\times|B|}$ 表示 X 中对应于行 A 和列B的子矩阵。为进行条件化论证，我们定义截断似然

$$\widetilde{L}_s(X):=\mathbb{E}_{S\big||S|=s}\left[\frac{w_{s,d}(X_{S\times S})\mathbf{1}\{X_{S\times S}\in\mathcal{E}_s\}}{m_{s,d}(X_{S\times S})}\right]\,,$$

其中，对于任意 $\ell \le s$,

$$\mathcal{E}_\ell := \left\{ X \in \mathbb{R}^{\ell\times\ell} \text{ symmetric: } \left\| X_{Q\times Q} - dI_{|Q|} \right\|_{\mathrm{op}} \le 10(1+\sqrt{\log s})\sqrt{d|Q|} \text{ for all } \emptyset \subsetneq Q \subseteq [\ell] \right\}. \tag{5.5}$$

正如我们稍后将看到的，这会防止矩阵元素过大——而这正是二阶矩类计算面临的障碍。要使 (II) 成立为 $o(1)$，只需证明

$$\min_{s\in[k^-,k^+]} \mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)] \to 1\,, \tag{5.6}$$

$$\max_{s\in[k^-,k^+]} \mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)^2] = 1 + o(1)\,. \tag{5.7}$$

这是因为

$$\begin{aligned}\mathbb{E}_{X\sim m_{n,d}}[|L_s(X)-1|] &\le \mathbb{E}_{X\sim m_{n,d}}[|\widetilde{L}_s(X)-1|] + \mathbb{E}_{X\sim m_{n,d}}[L_s(X)-\widetilde{L}_s(X)] \\ &\le \sqrt{\mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)^2] - 1 + 2(1-\mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)])} + 1 - \mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)]\,,\end{aligned}$$

其中对 $s\in[k^-,k^+]$ 取最大值，再结合 (5.6) 和 (5.7) 即可推出 (II) $= o(1)$。

第一个结论 (5.6) 很容易得到，它源于 Wishart 分布的集中性。首先，我们有

$$0 \le \mathbb{E}_{X\sim m_{n,d}}[1-\widetilde{L}_s(X)] = \mathbb{P}_{X\sim w_{n,d}}(X_{[s]\times[s]} \notin \mathcal{E}_s)\,,$$

由 $L \le L$ 以及关于 S 的对称性可得。后者还可进一步用以下上界估计

$$\begin{aligned}\mathbb{P}_{X\sim w_{n,d}}(X_{[s]\times[s]} \notin \mathcal{E}_s) &\le \sum_{q=1}^{s} \binom{s}{q} \mathbb{P}_{X\sim w_{n,d}}\left(\left\| X_{[q]\times[q]} - dI_q \right\|_{\mathrm{op}} > 10(1+\sqrt{\log s})\sqrt{dq} \right) \\ &\le \sum_{q=1}^{s} 2\exp(q\log s - (3(1+\sqrt{\log s})\sqrt{q})^2/2) \le s \times 2s^{-3}\,.\end{aligned}$$

其中，第一个不等式来自并集界和对称性，第二个不等式来自标准的 Wishart 集中性 [DS01，定理 II。13] 其中 $d + 10(1+\sqrt{\log s})\sqrt{dq} > (\sqrt{d}+\sqrt{q}+3(1+\sqrt{\log s})\sqrt{q})^2$ 对所有充分大的 d 均成立。由于 $s = \Omega(k)$ 对 $s\in[k^-,k^+]$ 成立，且 $k\to\infty$，这就证明了 (5.6)。

对于第二个陈述（5.7），令 S' 为 S 的独立同分布副本。那么，

$$\begin{aligned}&\mathbb{E}_{X\sim m_{n,d}}[\widetilde{L}_s(X)^2] \\ &= \mathbb{E}_{X\sim m_{n,d}}\left[\mathbb{E}_{S\,||S|=s}\left[\frac{w_{s,d}(X_{S\times S})\mathbf{1}\{X_{S\times S}\in\mathcal{E}_s\}}{m_{s,d}(X_{S\times S})} \right] \mathbb{E}_{S'\,||S'|=s}\left[\frac{w_{s,d}(X_{S'\times S'})\mathbf{1}\{X_{S'\times S'}\in\mathcal{E}_s\}}{m_{s,d}(X_{S'\times S'})} \right] \right] \\ &= \mathbb{E}_{S,S'\,||S|=|S'|=s}\left[\mathbb{E}_{X\sim m_{n,d}}\left[\frac{w_{s,d}(X_{S\times S})\mathbf{1}\{X_{S\times S}\in\mathcal{E}_s\} w_{s,d}(X_{S'\times S'})\mathbf{1}\{X_{S'\times S'}\in\mathcal{E}_s\}}{m_{s,d}(X_{S\times S})m_{s,d}(X_{S'\times S'})} \right] \right] \\ &\le \mathbb{E}_{S,S'\,||S|=|S'|=s}\left[\mathbb{E}_{X\sim m_{n,d}}\left[\frac{w_{s,d}(X_{S\times S})w_{s,d}(X_{S'\times S'})\mathbf{1}\{X_{(S\cap S')\times(S\cap S')}\in\mathcal{E}_{|S\cap S'|}\}}{m_{s,d}(X_{S\times S})m_{s,d}(X_{S'\times S'})} \right] \right].\end{aligned} \tag{5.8}$$

接下来，我们要证明（5.8）中的内层期望仅取决于指标在 S 和 S' 上重叠的那些条目，即 $X_{(S\cap S')\times(S\cap S')}$。为方便记号，令 $R := S\cap S'$ 和 $N := |R|$。定义子矩阵[7]的（条件）密度 g、g'、h 和 h' 如下：

$$\begin{aligned}X_{S\times S\setminus R\times R}|X_{R\times R}\sim g, \quad X_{S'\times S'\setminus R\times R}|X_{R\times R}\sim g' \quad \text{where } X\sim w_{n,d}\,,\\ X_{S\times S\setminus R\times R}|X_{R\times R}\sim h, \quad X_{S'\times S'\setminus R\times R}|X_{R\times R}\sim h' \quad \text{where } X\sim m_{n,d}\,.\end{aligned}$$

[7]下文中，$X_{S\times S\setminus R\times R}$ 表示元素 $X_{(S\setminus R)\times(S\setminus R)}\cup X_{(S\setminus R)\times R}\cup X_{R\times(S\setminus R)}$。

值得注意的是，由于$m_{s,\,d}$是乘积分布，h和h'不依赖于$X_{R\times R}$。那么

$$\begin{aligned}
&\mathbb{E}_{X\sim m_{n,d}}\left[\frac{w_{s,d}(X_{S\times S})w_{s,d}(X_{S'\times S'})\mathbf{1}\{X_{(S\cap S')\times(S\cap S')}\in\mathcal{E}_{|S\cap S'|}\}}{m_{s,d}(X_{S\times S})m_{s,d}(X_{S'\times S'})}\right]\\
&=\mathbb{E}_{X\sim m_{n,d}}\left[\frac{w_{N,d}^2(X_{R\times R})\mathbf{1}\{X_{R\times R}\in\mathcal{E}_N\}g(X_{S\times S\setminus R\times R}|X_{R\times R})g'(X_{S'\times S'\setminus R\times R}|X_{R\times R})}{m_{N,d}^2(X_{R\times R})h(X_{S\times S\setminus R\times R})h'(X_{S'\times S'\setminus R\times R})}\right]\\
&=\mathbb{E}_{X_{R\times R}\sim m_{N,d}}\Bigg[\frac{w_{N,d}^2(X_{R\times R})\mathbf{1}\{X_{R\times R}\in\mathcal{E}_N\}}{m_{N,d}^2(X_{R\times R})}\\
&\quad\times\mathbb{E}_{X_{S\times S\setminus R\times R}\sim h}\left[\frac{g(X_{S\times S\setminus R\times R}|X_{R\times R})}{h(X_{S\times S\setminus R\times R})}\right]\times\mathbb{E}_{X_{S'\times S'\setminus R\times R}\sim h'}\left[\frac{g'(X_{S'\times S'\setminus R\times R}|X_{R\times R})}{h'(X_{S'\times S'\setminus R\times R})}\right]\Bigg]\\
&=\mathbb{E}_{X_{R\times R}\sim m_{N,d}}\left[\frac{w_{N,d}^2(X_{R\times R})\mathbf{1}\{X_{R\times R}\in\mathcal{E}_N\}}{m_{N,d}^2(X_{R\times R})}\right],
\end{aligned}$$

其中，第一个等号由定义得出，第二个等号来自$m_{n、\,d}$为乘积分布，而最后一个等号来自g和g'为密度。由于最后一项只依赖于$X_{R\times R}$，将记号简化为$X\sim m_{N,\,d}$，（5.8）至多为

$$\mathbb{E}_{N\left||S|=|S'|=s\right.}\left[\mathbb{E}_{X\sim m_{N,d}}\left[\frac{w_{N,d}^2(X)\mathbf{1}\{X\in\mathcal{E}_N\}}{m_{N,d}^2(X)}\right]\right]. \tag{5.9}$$

下一步是计算（5.9）中关于X的内层期望，将其视为N和d的函数。具体做法是使用泰勒展开，将$w_{N,\,d}^2/m_{N,\,d}^2$近似为m_N的谱的简单函数，d并利用高概率事件（5.5）控制各项。由于这些计算在概念上至少与先前工作的计算相似 [BDER16、RR19]，，这里我们重点讨论其在下述引理中的含义。

引理 5.3（Wishart 与高斯正交系综的比较）。设$d\gg k^2$和$k\to\infty$。则存在常数$C_{5.3}>0$以及正随机变量$f_1(X)$ $f_2(X)$ $f_3(X)$使得

$$\frac{w_{N,d}^2(X)\mathbf{1}\{X\in\mathcal{E}_N\}}{m_{N,d}^2(X)}\le(1+o(1))f_1(X)f_2(X)f_3(X)\,, \tag{5.10}$$

其中每个f_i（X）（其依赖于N、d）均满足

$$\mathbb{E}_{X\sim m_{N,d}}[f_i(X)^3]\le\exp\left(C_{5.3}\frac{N^3}{d}\right),\quad 1\le i\le 3\,.$$

这里，$1+o$（1）因子在（5.10）对$s\in[k^-,k^+]$是通用的。

证明推迟至附录C.2。对f_i（X）的上界表明，只需计算在给定$|S|=|S'|=s=\Theta$（k）的条件下，关于$N=|S\cap S'|$的 exp(O（N^3/d）) 的期望。下面的引理表明，在所需条件$d\gg k^2\vee k^6/n^3$下，该期望为$1+o$（1）。我们指出，该结果此前已发表于 [KSWY25，命题 4。3]其目标类似，都是建立检测下界，但模型不同。

引理 5.4（三次超几何的上界）。设$C_{5.4}>0$为任意常数。若$k\le n/5$且$d\gg k^2\vee k^6/n^3$，则

$$\max_{s\in[k^-,k^+]}\mathbb{E}_{N\left||S|=|S'|=s\right.}\left[\exp\left(C_{5.4}\frac{N^3}{d}\right)\right]=1+o(1)\,.$$

证明推迟至附录C.3。由（5.9）和（5.10），可得

$$\begin{aligned}
&\max_{s\in[k^-,k^+]} \mathbb{E}_{N\left||S|=|S'|=s\right.}\left[\mathbb{E}_{X\sim m_{N,d}}\left[\frac{w_{N,d}^2(X)\mathbf{1}\{X\in\mathcal{E}_N\}}{m_{N,d}^2(X)}\right]\right]\\
&\le (1+o(1))\max_{s\in[k^-,k^+]} \mathbb{E}_{N\left||S|=|S'|=s\right.}[\mathbb{E}_{X\sim m_{N,d}}[f_1(X)f_2(X)f_3(X)]]\\
&\le (1+o(1))\prod_{i=1}^{3}\left(\max_{s\in[k^-,k^+]} \mathbb{E}_{N\left||S|=|S'|=s\right.}[\mathbb{E}_{X\sim m_{N,d}}[f_i(X)^3]]\right)^{1/3}\\
&\le (1+o(1))\prod_{i=1}^{3}\left(\max_{s\in[k^-,k^+]} \mathbb{E}_{N\left||S|=|S'|=s\right.}\left[\exp\left(C_{5.3}\frac{N^3}{d}\right)\right]\right)^{1/3}\\
&\le 1+o(1)\,,
\end{aligned}$$

第二个不等式来自 H¨older 不等式，第三个不等式由引理 5.3得出，最后一个不等式由引理5.4得出。这就证明了（5.7），并完成了定理 2.5的证明。

5.2 KL 散度的张量化

在这种方法中，我们考虑一个序贯过程：在每个时刻 $i = 1,\dots,n$，引入一个新的顶点，以及它与此前各顶点之间的边。

形式上，令 B 为来自 $\mathcal{Q}$ 的一个样本的邻接矩阵，邻域向量$B_i \in \{0,1\}^{i-1}$ 表示 $i \in [n]$ 与顶点 $j < i$之间的邻接关系。类似地，我们定义A （对于 $i \in [n]$，也类似地定义 $A_i \in \{0,1\}^{i-1}$ ）为来自 $\mathcal{P}$ 的一个样本的邻接矩阵，它依赖于联合潜在向量 $X_i := (U_i, V_i) \in \mathbb{R}^{d+1}, i \in [n]$。具体而言，对所有 $i \in [n]$，独立地有 $U_i \sim \mathcal{U}(\mathbb{S}^{d-1})$（特征向量）和 $V_i \sim \mathsf{Ber}(k/n)$（社区成员指示变量）。注意，这里我们假设每个顶点 i 都有一个特征向量 U_i，而成员指示变量 V_i 决定是否使用该特征向量。

全文中，我们将使用如下记号： $A_{[i]} := (A_1, \dots, A_i)$。于是我们有

$$\begin{aligned}
2d_{\mathrm{TV}}(\mathcal{P},\mathcal{Q})^2 = 2d_{\mathrm{TV}}(A,B)^2 &\le d_{\mathrm{KL}}(A||B)\\
&\le \sum_{i=0}^{n-1}\mathbb{E}_{A_{[i]}}[d_{\mathrm{KL}}(A_{i+1}|A_{[i]}||B_{i+1})]\\
&\le \sum_{i=0}^{n-1}\mathbb{E}_{A_{[i]},X_{[i]}}[d_{\mathrm{KL}}(A_{i+1}|A_{[i]},X_{[i]}||B_{i+1})]\\
&= \sum_{i=0}^{n-1}\mathbb{E}_{X_{[i]}}[d_{\mathrm{KL}}(A_{i+1}|X_{[i]}||B_{i+1})]\,,
\end{aligned}$$

根据 Pinsker 不等式、KL 散度的链式法则（以及 B_{i+1} 与$B_{[i]}$之间的独立性），再利用 KL 散度的凸性，以及 $A_{[i]}$ 和 A_{i+1} 在给定 $X_{[i]}$的条件下相互独立这一事实。根据 [LMSY22 中命题 8 的相同论证，2], 最后一项还可以进一步以上界估计为

$$\sum_{i=0}^{n-1}\mathbb{E}_{X_{[i]}}[d_{\mathrm{KL}}(A_{i+1}|X_{[i]}||B_{i+1})] \le n\mathbb{E}_{X_{[n-1]}}[d_{\mathrm{KL}}(A_n|X_{[n-1]}||B_n)]\,.$$

因此，只需证明

$$\mathbb{E}_{X_{[n-1]}}[d_{\mathrm{KL}}(A_n|X_{[n-1]}||B_n)] = o\left(\frac{1}{n}\right). \tag{5.11}$$

令 $\mathcal{P}_n(\cdot|X_{[n-1]})\colon \{0,1\}^{n-1} \to [0,1]$ 为 $A_n|X_{[n-1]}$ 的分布，且 $\mathcal{Q}_n(\cdot)\colon \{0,1\}^{n-1} \to [0,1]$ 为 B_n 的分布。那么，式（5.11）可改写为

$$\begin{aligned}\mathbb{E}_{X_{[n-1]}}[d_{\mathrm{KL}}(A_n|X_{[n-1]}||B_n)] &= \mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}\left[\log\frac{\mathcal{P}_n(\Gamma|X_{[n-1]})}{\mathcal{Q}_n(\Gamma)}\right]\right] \\ &= \mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}\left[\log\left(\Delta(\Gamma,X_{[n-1]})+1\right)\right]\right],\end{aligned} \tag{5.12}$$

其中 $\Gamma := (\Gamma_1,\ldots,\Gamma_{n-1}) \in \{0,1\}^{n-1}$ 且 $\Delta(\Gamma, X_{[n-1]}) := \mathcal{P}_n(\Gamma|X_{[n-1]})/\mathcal{Q}_n(\Gamma) - 1$。

我们的下一个论证是，Δ 的大小可以通过两种方式显著缩减。首先，这会使邻域分布之间的比率变小。这是因为顶点n 本身可能不属于社区（V_n 尚未揭示），在这种情况下，邻域分布就是 $\mathcal{Q}_n$。事实上，我们可以写成

$$\mathcal{P}_n(\Gamma|X_{[n-1]}) = \frac{k}{n}\overline{\mathcal{P}}_n(\Gamma|X_{[n-1]}) + \left(1-\frac{k}{n}\right)\mathcal{Q}_n(\Gamma),$$

其中，$\overline{\mathcal{P}}_n(\cdot|X_{[n-1]})$ 是在给定 $V_n = 1$ 条件下 $A_n|X_{[n-1]}$ 的分布。因此，

$$\Delta(\Gamma, X_{[n-1]}) = \frac{k}{n}\left(\frac{\overline{\mathcal{P}}_n(\Gamma|X_{[n-1]})}{\mathcal{Q}_n(\Gamma)} - 1\right). \tag{5.13}$$

其次，Δ 很小，因为它通常比较的是规模远小于 n 的邻域。如果某个先前的顶点不在社区中，则它与顶点 n 之间的边服从 $\mathrm{Bern}(p)$ 分布，从而抵消 $\mathcal{Q}_n$ 中相应的因子；注意，根据对 $X_{[n-1]}$ 的条件化，可以获得关于这些先前顶点的信息。这就得到了邻域分布之间更小的比值。由于每个先前顶点以 k/n 的概率位于社区中，平均而言，相关顶点至多有 $O(k)$ 个。

为了捕捉截至目前所解释的平均情形，我们考虑以下事件：

$$\{(\Gamma,U,V)\in\mathcal{E}\} := \left\{\sum_{i=1}^{n-1} V_i \le 2k, \sum_{i\in[n-1]:V_i=1}\Gamma_i \le 4kp, |\Delta(\Gamma,X_{[n-1]})| \le \frac{k}{n}\right\}.$$

这里，$U := (U_1、\ldots\ldots、U_{n-1})$。下面的引理表明，这确实是一个高概率事件。

引理 5.5。假设 $p\in(0,\ 1/2]$。那么存在一个常数 $C_{5.5} > 0$，使得当 n 足够大且

$$kp \ge C_{5.5}\log n \quad \textit{and} \quad d \ge C_{5.5}(kp\log(1/p)\log(d/p))^2\log n,$$

那么以下结论成立：

$$\begin{aligned} n\log(1/p)\mathbb{P}_{X_{[n-1]},\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}((\Gamma,U,V)\in\mathcal{E}^c) &= o(1/n), \\ n\log(1/p)\mathbb{P}_{X_{[n-1]},\Gamma\sim\mathcal{Q}_n}((\Gamma,U,V)\in\mathcal{E}^c) &= o(1/n).\end{aligned} \tag{5.14}$$

证明推迟至附录C.4。由（5.12）可得

$$\begin{aligned}
&\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}\left[\log\left(\Delta(\Gamma,X_{[n-1]})+1\right)\right]\right]\\
&\le o\left(\frac{1}{n}\right)+\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}\left[\log\left(\Delta(\Gamma,X_{[n-1]})+1\right)\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right]\\
&\le o\left(\frac{1}{n}\right)+\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{Q}_n}\left[\left(\Delta(\Gamma,X_{[n-1]})+1\right)\Delta(\Gamma,X_{[n-1]})\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right]\\
&\le o\left(\frac{1}{n}\right)+\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{Q}_n}\left[\Delta(\Gamma,X_{[n-1]})^2\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right],
\end{aligned}$$

其中，第一个不等式由引理 5.5 以及 $\log(\Delta+1)\le n\log(1/p)$ 得出，依据是 $\mathcal{P}_n\le 1,\mathcal{Q}_n\ge p^n$；第二个不等式由 $\log(x+1)\le x$得出；第三个不等式则由

$$\begin{aligned}
&\left|\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{Q}_n}\left[\Delta(\Gamma,X_{[n-1]})\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right]\right|\\
&=\left|\mathbb{P}_{X_{[n-1]},\Gamma\sim\mathcal{Q}_n}((\Gamma,U,V)\in\mathcal{E}^c)-\mathbb{P}_{X_{[n-1]},\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})}((\Gamma,U,V)\in\mathcal{E}^c)\right|=o(1/n),
\end{aligned}$$

同样由引理5.5得出。

为方便记号，定义 ν 为满足 $V_i=1$ 的顶点 $i\in[n-1]$ 的数量，并令 $i_1<\cdots<i_\nu$ 表示这些指标。于是

$$\Delta(\Gamma,X_{[n-1]})=\frac{k}{n}\left(\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{\nu-\sum_{j=1}^{\nu}\Gamma_{i_j}}}-1\right),\tag{5.15}$$

~其中，$\mathcal{P}_{l+1}(\cdot|U_{m_1},\ldots,U_{m_l})$ 表示顶点 n 与顶点 $m_1<$之间的边的分布 $\cdots<m_l$在 $\mathcal{G}(l+1,p,d)$ 下（此处为顶点 $m_1,\ldots,m_l,n$ 上），在给定对应的潜变量$U_{m_1},\ldots,U_{m_l}$的条件下。因此，只需证明

$$\begin{aligned}
&\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{Q}_n}\left[\Delta(\Gamma,X_{[n-1]})^2\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right]\\
&=\left(\frac{k}{n}\right)^2\mathbb{E}_{X_{[n-1]}}\left[\mathbb{E}_{\Gamma\sim\mathcal{Q}_n}\left[\left(\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{\nu-\sum_{j=1}^{\nu}\Gamma_{i_j}}}-1\right)^2\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\right],
\end{aligned}\tag{5.16}$$

是 $o(1/n)$。为此，考虑任意固定的 V 和 Γ，满足 $\nu\le 2k$ 和 $\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp$ （ $\mathcal{E}$的前两个条件）。仅利用 U_{i_1}、……、 U_{i_ν}的随机性，

$$\begin{aligned}
&\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu}}\left[\left(\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{\nu-\sum_{j=1}^{\nu}\Gamma_{i_j}}}-1\right)^2\mathbf{1}\{(\Gamma,U,V)\in\mathcal{E}\}\right]\\
&\le\int_0^1\mathbb{P}_{U_{i_1},\ldots,U_{i_\nu}}\left(\left(\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{\nu-\sum_{j=1}^{\nu}\Gamma_{i_j}}}-1\right)^2\ge t\right)dt\\
&\le\int_0^{1/(k^2(\log n)/n)}1dt\\
&\quad+\int_{1/(k^2(\log n)/n)}^1\exp\left(-\frac{1}{C_{A.4}}\frac{dt}{C_0kp(kp\log(1/p)+\log(d/p))\log(1/p)\log(d/p)}+C_{A.4}\log(3k)\right)dt\\
&\le\frac{1}{\log n}\frac{n}{k^2}+\frac{C_1(kp\log(1/p)\log(d/p))^2}{d}\exp\left(-\frac{dn}{C_1(k^2p\log(1/p)\log(d/p))^2\log n}+C_1\log(3k)\right),
\end{aligned}$$

对于某些常数 $C_0 > 0$ 和 $C_1 > 0$。第一个不等式来自 $|\Delta(\Gamma,\ X_{[n-1]})| \le k/n$ （的第三个条件；参见 (5.15)），第二个不等式来自引理A.4，其中用到 $\nu \le 2k$ 和$\sum_{j=1}^{\nu}\Gamma_{i_j} \le 4kp$；最后一个不等式来自 $kp\log(1/p) + \log(d/p) \le kp\log(1/p)\log(d/p)$。因此，(5.16)至多为

$$\frac{1}{n\log n} + \frac{C_1(k^2p\log(1/p)\log(d/p))^2}{n^2d}\exp\left(-\frac{dn}{C_1(k^2p\log(1/p)\log(d/p))^2\log n} + C_1\log(3k)\right).$$

只要.....，这就是 o({v2)。

$$d \ge C_2\frac{k^4p^2}{n}\log^2(1/p)\log^2(d/p)\log^2 n\,,$$

对某个常数 $C_2 > 0$。结合引理 5.5，我们证明了要使（5.11）成立，只需满足

$$d \ge (C_{5.5} \vee C_2)\left(k^2p^2 \vee \frac{k^4p^2}{n}\right)\log^2(d/p)\log^2(1/p)\log^2 n\,.$$

这可由 $d \ge C_{2.6}$(k^2p^2 \vee k^4p^2/n) log^2(k/p)log^2(1/p) log^3 n 推出，其中常数 $C_{2.6} > 0$ 取适当值。

6 计算下界的证明

6.1 低次下界

在本小节中，我们证明定理 2.9。

对于 K_n 的任意子图 H，定义其傅里叶系数（相对于 $\mathcal{L}^2$(Q) 中的标准正交基）为

$$\Phi_{\mathcal{P}}(H) := \mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{ij\in E(H)}\frac{G_{ij}-p}{\sqrt{p(1-p)}}\right].$$

根据 [BB24b 的命题 3，1], 为了证明不存在次数至多为 D 且能够实现弱分离的多项式，只需证明

$$\sum_{H\subseteq K_n:1\le e(H)\le D}\Phi_{\mathcal{P}}(H)^2 = o(1)\,. \tag{6.1}$$

现在令 $D = \lfloor(\log n/\log(\log n))^2\rfloor$。注意，对于每个 H，在 K_n中至多有 $n^{v(H)}$ 个与 H同构的子图。此外，这些子图的 $\Phi_{\mathcal{P}}$() 的值都相同。因此，对于（6.1），只需证明

$$\sum_{\text{non-iso. }H\subseteq K_n:1\le e(H)\le D} n^{v(H)}\Phi_{\mathcal{P}}(H)^2 = o(1)\,, \tag{6.2}$$

其中 non-iso. $H \subseteq K_n$ 表示对所有非同构子图进行枚举 H 的 K_n。对于每个H，令 $H_1,\dots,H_r$为其连通分量。于是 $\Phi_{\mathcal{P}}(H) = \prod_{i=1}^{r}\Phi_{\mathcal{P}}(H_i)$，特别地，如果 H_i中任意一个是树，则由推论 $\Phi_{\mathcal{P}}(H) = 0$可得 4.2。因此，对于满足 H的 $\Phi_{\mathcal{P}}(H) \ne 0$，可以假设

使得 $v(H_i) \geq 3$ 且 $e(H_i) \geq v(H_i)$。那么对于每个 H_i，我们有

$$\begin{aligned}
|\Phi_{\mathcal{P}}(H_i)| &= \left|(p(1-p))^{-e(H_i)/2}\mathbb{E}_{G\sim\mathcal{P}}\left[\prod_{j\ell\in E(H_i)}(G_{j\ell}-p)\right]\right| \\
&\leq (p(1-p))^{-e(H_i)/2}(8p)^{e(H_i)}\left(\frac{C_{A.7}(\log^6 n)(\log^{3/2} d)}{\sqrt{d}}\right)^{\lceil (v(H_i)-1)/2\rceil}\left(\frac{k}{n}\right)^{v(H_i)} \\
&\leq (8\sqrt{2p})^{e(H_i)}\left(\frac{C_{A.7}(\log^6 n)(\log^{3/2} d)}{\sqrt{d}}\right)^{v(H_i)/3}\left(\frac{k}{n}\right)^{v(H_i)} \\
&\leq 8^{e(H_i)}\left(\frac{4C_{A.7}k^3p^{3/2}(\log^6 n)(\log^{3/2} d)}{n^3\sqrt{d}}\right)^{v(H_i)/3}.
\end{aligned} \tag{6.3}$$

这里，第一个不等式由引理 A.7（其中 $v(H_i)$ $e(H_i) \leq \log^3 n$）和引理 4.1 得出；第二个不等式由 $1-p \geq 1/2$ and $\lceil (v(H_i)-1)/2\rceil \geq v(H_i)/3$ 得出；最后一个不等式由 $\sqrt{2p} \leq 1$and $e(H_i) \geq v(H_i)$ 得出。

由（6.3），可将（6.2）的左边界定为

$$\begin{aligned}
&\sum_{\text{non-iso. } H\subseteq K_n:1\leq e(H)\leq D} n^{v(H)}\Phi_{\mathcal{P}}(H)^2 \\
&\leq \sum_{\text{non-iso. } H\subseteq K_n:1\leq e(H)\leq D} 64^{e(H)}\left(\frac{16C_{A.7}^2k^6p^3(\log^{12} n)(\log^3 d)}{n^3 d}\right)^{v(H)/3} \\
&\leq \sum_{\text{non-iso. } H\subseteq K_n:1\leq e(H)\leq D} 64^{e(H)}n^{-((\varepsilon\wedge\delta)/6)v(H)} \leq \sum_{\text{non-iso. } H\subseteq K_n:1\leq e(H)\leq D} n^{-((\varepsilon\wedge\delta)/12)v(H)}.
\end{aligned} \tag{6.4}$$

这里，对于每个常数 $\varepsilon > 0$，只要 n 足够大，第二个不等式就成立。为证明这一点，固定一个常数 $C_0 > 0$，使得 $d \mapsto d/\log^3 d$ 对 $d \geq C_0$ 单调递增。若 $k^6p^3/n^3 < C_0$，则对于所有足够大的 n，$16C_{A.7}^2k^6p^3(\log^{12} n)(\log^3 d)/d \leq 16C_{A.7}^2C_0(\log^{12} n)(\log^3 d)/n^3d \leq n^{-\delta/2}$；否则，对于所有足够大的 n，$16C_{A.7}^2k^6p^3(\log^{12} n)(\log^3 d)/n^3d \leq 16C_{A.7}^2(\log^{12} n)(\log^3(k^6p^3/n^{3-\varepsilon}))/n^{\varepsilon} \leq n^{-\varepsilon/2}$。式 (6.4) 中的最后一个不等式由 $e(H) \leq v(H)^2 \wedge D \leq v(H)\sqrt{D} \leq v(H)\log n/\log(\log n)$ 得出，这意味着对于所有足够大的 n，$64^{e(H)} \leq n^{((\varepsilon\wedge\delta)/12)v(H)}$。

我们控制（6.4）中最后一项的论证与 [BB24b，命题 3 中的论证类似。1]。如 $v(H) \leq 2e(H) \leq 2D$，我们将非同构图的数量界定为$v(H)$ 的函数。特别地，具有 $v(H) = v$ 和 $e(H) \leq D$ 的非同构图数量至多为

$$2^{\binom{v}{2}} \wedge \left(\binom{v}{2}+1\right)^D \leq 2^{v^2} \wedge \exp(2D\log v). \tag{6.5}$$

这里，$2^{\binom{v}{2}}$ 可通过枚举边的所有可能选择 $\{0,1\}^{\binom{v}{2}}$ 得出，而 $\left(\binom{v}{2}+1\right)^D$则是通过在每个时刻决定是否在 $\binom{v}{2}$ 种可能选择中添加一条边得出的。

现在任取一个函数 $g: \mathbb{N} \to \mathbb{N}$，使得 $g(n) = \omega(1)$，且 $g(n) = o(\log(\log n))$。若 $v \leq \sqrt{Dg(n)}$，则 (6.5) 至多为 $2^{v^2} \leq 2^{v(\sqrt{Dg(n)}+1)} \leq 2^{v\log n/\log(\log n)\times g(n)+v} = 2^{o(v\log n)}$。类似地，若 $v > \sqrt{Dg(n)}$，则 (6.5) 至多为 $\exp(v\times 2D(\log v)/v) \leq \exp(v\times 2D\log(\sqrt{Dg(n)})/(\sqrt{Dg(n)})) \leq \exp(v\times 2\log n/(\log(\log n)) \times \log(\log n)/g(n)) = \exp(o(v\log n))$，对所有充分大的 n 均成立。由

这些，(6.4) 中的最后一项至多为

$$\sum_{v=1}^{2D} \exp(o(v\log n)) n^{-((\varepsilon\wedge\delta)/12)v} \le \sum_{v=1}^{2D} n^{((\varepsilon\wedge\delta)/24-(\varepsilon\wedge\delta)/12)v} \le 2Dn^{-(\varepsilon\wedge\delta)/24} = o(1)\,,$$

当 $\le n^{((\varepsilon\wedge\delta)/24)v}$ 对所有 $1 \le v \le 2D$ 均成立时，对于所有充分大的 n，有 $\exp(o(v\log n))$。

6.2 较长环计数的次优性

这里，我们证明命题 2.10。以下命题是一个关键结果，它严格刻画了在 $\mathcal{G}(n,\ p,\ d)$ 下带符号环的期望。通过引理 4.1，可将其直接转换为 $\mathcal{P} = \mathcal{G}(n,\ p,\ d,\ k)$ 下的界。令 Cyc_ℓ 为一个长度为 ℓ 的环。

命题 6.1 (带符号环计数的期望)。 假设 $p \le 1/2$ 和 d 充分大，且满足

$$d \ge (5\log(1/p))^4\,.$$

那么存在一个常数 $C_{6.1} > 0$，使得对于任意 $3 \le \ell \le n$,

$$\frac{1}{C_{6.1}^{\ell}} \frac{p^\ell \log^{\ell/2}(1/p)}{d^{\ell/2-1}} \le \mathbb{E}_{G\sim\mathcal{G}(n,p,d)} \left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} (G_{ij} - p) \right] \le C_{6.1}^{\ell} \frac{p^\ell \log^{\ell/2}(1/p)}{d^{\ell/2-1}}\,.$$

证明见附录 D.1。

注 6.2 (紧确性与推论)。 尽管先前已有研究给出了该值的上界（例如，*[LMSY23、BB24b]*），我们的上界确定了所有参数的正确渐近依赖关系（底数相差一个常数因子，且以 ℓ 为指数）。此外，我们还给出了一个可能具有独立研究价值的匹配下界。例如，我们预计它可用于证明 *[CZ25]* 中关于第二大特征值的界是紧的；另见注 *6.3*。

作为命题6.1的直接推论，我们证明较长环计数的能力严格弱于三角形计数。

命题*2.10*的证明。 令 $f_\ell(G)$ 为 Cyc_ℓ 的有符号计数，其中 G。根据命题6.1 和引理4.1。

$$0 \le \mathbb{E}_{G\sim\mathcal{P}}[f_\ell(G)] - \mathbb{E}_{G\sim\mathcal{Q}}[f_\ell(G)] \le \binom{n}{\ell} \frac{(\ell-1)!}{2} \times \left(\frac{k}{n}\right)^{\ell} \frac{C_{6.1}^{\ell} p^\ell \log^{\ell/2}(1/p)}{d^{\ell/2-1}}\,,$$

作为 K_n 中长度为 ℓ 的不同环的数量，其值为 $\binom{n}{\ell}\frac{(\ell-1)!}{2}$。此外，直接计算可得

$$\mathrm{Var}_{G\sim\mathcal{Q}}[f_\ell(G)] = \binom{n}{\ell} \frac{(\ell-1)!}{2} (p(1-p))^\ell\,.$$

然后整理 $\mathbb{E}_{G\sim\mathcal{P}}[f_\ell(G)] - \mathbb{E}_{G\sim\mathcal{Q}}[f_\ell(G)] \gg \sqrt{\mathrm{Var}_{G\sim\mathcal{Q}}[f_\ell(G)]}$，即可得到结果。 □

下面简要讨论谱算法，因为带符号环计数与迹方法密切相关 [Wig58，Vu07]。

注 6.3（谱检验）。近期有几篇论文 *[LMSY23、BB24b、CZ25]* 研究了完整模型 $\mathcal{G}(n,p,d)$ 的谱性质。为方便记号，令 A 表示 $G \sim \mathcal{G}(n,p,d)$ 的中心化邻接矩阵，即 $A_{ij} := G_{ij} - p$，若 $i \neq j$，且 $A_{ii} := 0$。*[CZ25]* 证明，若 $d = \Omega(np)$，

(i) A 的特征值经验分布收敛到与 $\mathcal{G}(n,p)$ 中相同的半圆律。*(ii)* A 的最大特征值与 $\mathcal{G}(n,p)$ 的最大特征值处于同一数量级（相差至多多对数因子）。

非正式地说，这些结果表明，“朴素”的谱检验（例如，对 ¨ 的最大特征值进行阈值化）可能不如全局有符号三角形计数，后者在达到 $A\ d = o(n^3p^3\log^3(1/p))$ 时仍然有效。 *[CZ25]* 的结果基于迹方法，其中命题 *6 1* $G \sim$。可部分 $\mathcal{G}\ n\ p\ d\ k$ $d = \tilde{}\ k^2p/n$ ¨ 应用。基于此，我们预计，若 (,,,) 满足 Ω( )，则 *(i)* 和 *(ii)* 也成立，因此朴素谱检验相应的次优性仍会持续存在。

不过，这个问题有一个简单的解决办法，即考虑 $(A)^3$，而不是 A。那么，$(A)^3$的迹等于[8]全局有符号三角形计数；我们预计，对 $(A)^3$ 的最大特征值设定阈值的检验会有相同的表现。我们不打算沿这条思路继续探索，因为它似乎并不比全局检验（定理 *2.1*）有任何特别的优势，而且在概念上与之相同。

7 讨论

我们研究了一种新型随机图模型的检测问题：在一个更大的 Erd˝os–R´enyi 图中，隐藏着一个具有潜在高维几何的小型社区。该模型经过刻意设计，力求简约：它在社区中引入了非平凡的几何结构，同时不改变边缘顶点统计量。从这个意义上说，该模型为理解结构信号（而非仅仅是密度增加等因素）如何影响检测极限提供了一个清晰的框架。我们通过基于带符号三角形的检验刻画了检测阈值，并从信息论和计算两个角度考察了困难性，揭示出计算与统计之间的差距。在此过程中，我们刻画了一个区域：在该区域中，Wishart 分布与球面 Wishart 分布渐近等价，但二者均不同于高斯正交系综；此外，我们还给出了带符号环计数的紧确界，这些结果本身也可能具有独立的研究价值。

聚焦于检测问题，对于任意固定的 p，我们的上界和下界相差至多一个对数因子。对于趋于零的 p，则仍存在差距。即使对于 p的完整模型（其中 $k = n$ [LMSY22], ），改进对 k/n的依赖仍是一个重要的开放问题，而且似乎需要全新的思路。其他值得探索的开放方向包括：更精确地刻画对 $p = O$的依赖，或针对特定情形（例如稀疏的$1/k$）发展精细化结果，在这些情形下可能适用不同的技术 [LMSY22]。

在我们的模型中，社区结构为检测之外的统计任务引入了一个新的层面。一个直接的问题是，精确恢复或近似恢复社群顶点的阈值是多少。某些基于潜变量的社群模型呈现出检测-恢复差距 [MWZ23，KSWY25，MWZ25], 而类似现象也可能出现在这里。更广泛地说，在我们的局部化设定下，重新审视完整模型的统计与计算性质 $\mathcal{G}$（n，p，d）（例如，近期已在 [BB24b、BB25b、CZ25]中确立） $\mathcal{G}$（n，p，d，k）或可揭示高维几何与社区结构之间的相互作用。

[8] 严格来说，由于 Tr((A)³) 中求和项存在重复项，因此最多相差 6 倍。

致谢

我们感谢 Kiril Bangachev 就参考文献 [LR23b]及其中的观点提供了有益讨论，也感谢 JiamingXu 在项目早期就检测问题进行了有益讨论。

参考文献

[ABAL+26] Louigi Addario-Berry、Omer Angel、G´abor Lugosi、Mikl´os Z. R´acz 和 Tselil Schramm。植入匹配和生成树的统计阈值。arXiv 预印本 arXiv:2602.07669，2026年。[Abb17] Emmanuel Abbe。社区检测与随机块模型：近期进展。J. Mach. Learn. Res., 18(177):1–86, 2017。[ABD21] Konstantin Avrachenkov、Andrei Bobu 和 Maximilien Dreveton。几何图的高阶谱聚类。J. Fourier Anal. Appl., 27(22):1–29, 2021。[ABI24] Pedro Abdalla、Afonso S. Bandeira 和 Clara Invernizzi。随机几何图上自发同步的保证。SIAM J. Appl. Dyn. Syst., 23(1):779–790, 2024。[ABS21] Emmanuel Abbe、Fran¸cois Baccelli 和 Abishek Sankararaman。欧几里得随机图上的社区检测。Inf. Inference, 10(1):109–160, 2021。[ACBL12] Ery Arias-Castro、S´ebastien Bubeck 和 G´abor Lugosi。相关性检测。Ann. Statist., 40(1):412–435, 2012。[ACBL15] Ery Arias-Castro、S´ebastien Bubeck 和 G´abor Lugosi。在多变量样本中检测正相关。Bernoulli, 21(1):209–241, 2015。[ACBLV18] Ery Arias-Castro、S´ebastien Bubeck、G´abor Lugosi 和 Nicolas Verzelen。检测隐藏在白噪声中的马尔可夫随机场。Bernoulli, 24(4B):3628–3656, 2018。[ACV14] Ery Arias-Castro 和 Nicolas Verzelen。稠密随机网络中的社区检测。Ann. Statist., 42(3):940–969, 2014。[AKL24] Konstantin Avrachenkov、B. R. Vinay Kumar 和 Lasse Leskel¨a。具有几何核的块模型上的社区检测。arXiv 预印本 arXiv:2403.02802, 2024。[AS65] Milton Abramowitz 和 Irene A. Stegun。《数学函数手册：附公式、图表和数学表》，第55卷。Courier Corporation，1965。[ATK15] Leman Akoglu、Hanghang Tong 和 Danai Koutra。基于图的异常检测与描述：综述。Data Min. Knowl. Discov., 29(3):626–688, 2015。[AW15] Rados-law Adamczak 和 Pawe-l Wolff。具有有界高阶导数的非利普希茨函数的集中不等式。Probab. Theory Related Fields, 162(3-4):531–586, 2015。[Bar16] Albert-L´aszl´o Barab´asi。《网络科学》。剑桥大学出版社，2016。[BB20] Matthew Brennan 和 Guy Bresler。秘密泄漏所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，PMLR，第648–847页，2020。[BB24a] Kiril Bangachev 和 Guy Bresler。随机几何图中的几何检测：三角形的次优性与簇展开。载于《第三十七届学习理论会议论文集》，第247卷，PMLR，第427–497页，2024。[BB24b] Kiril Bangachev 和 Guy Bresler。高维随机几何图的傅里叶系数。载于《第56届ACM计算理论研讨会论文集》，第549–560页，2024。

[BB25a] Kiril Bangachev 和 Guy Bresler。随机代数图及其向 Erd˝os-R´enyi 图的收敛。Random Structures Algorithms, 66(1):e21276, 1–43, 2025年。[BB25b] Kiril Bangachev 和 Guy Bresler。随机几何图与 Erd˝os-R´enyi 图的夹逼及其应用：尖锐阈值、稳健检验与计数。载于《第57届ACM计算理论年会论文集》，第310–321页，2025年。[BBAP05] Jinho Baik、G´erard Ben Arous 和 Sandrine P´ech´e。非零复样本协方差矩阵最大特征值的相变。Ann. Probab., 33(5):1643–1697, 2005。[BBCvdH20] Gianmarco Bet、Kay Bogerd、Rui M. Castro 和 Remco van der Hofstad。检测网络中的僵尸网络。Math. Stat. Learn., 3(3-4):315–343, 2020。[BBH21] Matthew Brennan、Guy Bresler 和 Brice Huang。Wishart矩阵的 de Finetti型结果：组合结构与相变。arXiv 预印本 arXiv:2103.14011, 2021。[BBH24] Matthew Brennan、Guy Bresler 和 Brice Huang。各向异性随机几何图中高维几何检测的阈值。Random Structures Algorithms, 64(1):125–137, 2024。[BBN20] Matthew Brennan、Guy Bresler 和 Dheeraj Nagaraj。随机图中潜在几何检测的相变。Probab. Theory Related Fields, 178(3-4):1215–1289, 2020。[BCC+10] Aditya Bhaskara、Moses Charikar、Eden Chlamtac、Uriel Feige 和 Aravindan Vijayaraghavan。检测高对数密度：稠密 k-子图的一个 $O(n^{1/4})$近似。载于《第42届ACM计算理论年会论文集》，第201–210页，2010。[BDER16] S´ebastien Bubeck、Jian Ding、Ronen Eldan 和 Mikl´os Z. R´acz。随机图中高维几何的检验。Random Structures Algorithms, 49(3):503–532, 2016。[BDT+20] 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圈。Oper. Res., 68(1):53–70, 2020。[BG18] S´ebastien Bubeck 和 Shirshendu Ganguly。高维 Wishart 矩阵的熵中心极限定理与相变。Int. Math. Res. Not. IMRN, 2018(2):588–606, 2018。[BGK+01] Andreas Brieden、Peter Gritzmann、Ravindran Kannan、Victor Klee、L´aszl´o Lov´asz 和 Mikl´os Simonovits。半径的确定性和随机化多项式时间近似。Mathematika, 48(1-2):63–105, 2001。[BGPS25] Samuel Baguley、Andreas G¨obel、Marcus Pappik 和 Leon Schiller。通过 Edgeworth型展开检验高维随机环面图的阈值与谱性质。载于《第三十八届学习理论会议论文集》，第291卷，PMLR，第200–201页，2025年。[BHJK25] Rares-Darius Buhai、Jun-Ting Hsieh、Aayush Jain 和 Pravesh K. Kothari。准多项式低次猜想是错误的。载于 IEEE 第66届计算机科学基础研讨会论文集，第2577–2590页，2025年。[BKL19] Karl Bringmann、Ralph Keusch 和 Johannes Lengler。几何非齐次随机图。Theoret. Comput. Sci., 760:35–54, 2019。[BMS25] Gianmarco Bet、Riccardo Michielan 和 Clara Stegehuis。无标度随机图中的局部几何检测。J. Appl. Probab., 第1–17页，2025年。[BO04] Albert-L´aszl´o Barab´asi 和 Zolt´an N. Oltvai。网络生物学：理解细胞的功能组织。Nat. Rev. Genet., 5(2):101–113, 2004。[BR13] Quentin Berthet 和 Philippe Rigollet。稀疏主成分检测的复杂性理论下界。载于《第26届学习理论年会论文集》，第30卷，PMLR，第1046–1066页，2013年。[CW19] Didier Ch´etelat 和 Martin T. Wells。Wishart矩阵的中尺度渐近。Ann. Statist., 47(5):2639–2670, 2019。

[CZ25] Yifan Cao 和 Yizhe Zhu。高维稀疏随机几何图的谱。arXiv 预印本 arXiv:2507.06556，2025。[DDC23] Quentin Duchemin 和 Yohann De Castro。随机几何图：近期进展与展望。载于《高维概率IX——空灵卷》，Progr. Probab. 第80卷，第347–392页。Birkh¨auser，2023。[DGLU11] Luc Devroye、Andr´as Gy¨orgy、G´abor Lugosi 和 Frederic Udina。高维随机几何图及其团数。Electron. J. Probab.，16：2481–2508，2011。[DKMZ11] Aur´elien Decelle、Florent Krzakala、Cristopher Moore 和 Lenka Zdeborov´a。模块化网络随机块模型的渐近分析及其算法应用。Phys. Rev. E，84：066106，2011。[DLW25] Prabhanka Deka、Fangzhou Luo 和 Baichuan Wu。随机几何图中的稀有事件概率。arXiv 预印本 arXiv:2510.09196，2025。[DS01] Kenneth R. Davidson 和 Stanislaw J. Szarek。局部算子理论、随机矩阵与巴拿赫空间。载于《巴拿赫空间几何手册》第 I 卷，第317–366页。North-Holland，2001。[DWXY23] Jian Ding、Yihong Wu、Jiaming Xu 和 Dana Yang。植入匹配问题：精确阈值与无穷阶相变。Probab. Theory Related Fields，187(1-2)：1–71，2023。[DX13] Feng Dai 和 Yuan Xu。《球面与球上的逼近理论和调和分析》。Springer 数学专著。施普林格，2013。[EH25] Dor Elimelech 和 Wasim Huleihel。在随机图中检测任意植入子图。载于《第三十八届学习理论会议论文集》，PMLR 第291卷，第1691–1798页，2025。[EM20] Ronen Eldan 和 Dan Mikulincer。各向异性随机几何图的信息与维数。载于《泛函分析的几何方面》第 I 卷，数学讲义第2256卷，第273–324页。施普林格，2020。[ER59] Paul Erd˝os 和 Alfr´ed R´enyi。论随机图 I。Publ. Math. Debrecen，6：290–297，1959。[For10] Santo Fortunato。图中的社区检测。Phys. Rep.，486(3-5)：75–174，2010。[FVD+16] Emilio Ferrara、Onur Varol、Clayton Davis、Filippo Menczer 和 Alessandro Flammini。社交机器人账号的兴起。Commun. ACM，59(7)：96–104，2016。[GGNW24] Julia Gaudio、Charlie Guan、Xiaochun Niu 和 Ermin Wei。欧几里得随机图中的精确标签恢复。arXiv 预印本 arXiv:2407.11163，2024。[GJ25] Julia Gaudio 和 Andrew Jin。几何随机块模型中的精确恢复。arXiv 预印本 arXiv:2512.22773，2025。[GJ26] Julia Gaudio 和 Andrew Jin。几何隐藏社区模型中的精确恢复。arXiv 预印本 arXiv:2601.17591，2026。[GMPS18] Sainyam Galhotra、Arya Mazumdar、Soumyabrata Pal 和 Barna Saha。几何块模型。AAAI 人工智能会议论文集，32(1)，2018。[GMPS23] Sainyam Galhotra、Arya Mazumdar、Soumyabrata Pal 和 Barna Saha。几何块模型中的社区恢复。J. Mach. Learn. Res.，24(338)：1–53，2023。[GN02] Michelle Girvan 和 Mark E. J. Newman。社会网络和生物网络中的社区结构。Proc. Natl. Acad. Sci. U.S.A.，99(12)：7821–7826，2002。[GNW24] Julia Gaudio、Xiaochun Niu 和 Ermin Wei。几何随机块模型中的精确社区恢复。载于《2024 年 ACM-SIAM 离散算法年度研讨会论文集》，第2158–2184页，2024。

[GSXY25] Julia Gaudio、Colin Sandon、Jiaming Xu 和 Dana Yang。在随机图中寻找植入环。arXiv 预印本 arXiv:2511.04058，2025。[HHLM99] Leland H. Hartwell、John J. Hopfield、Stanislas Leibler 和 Andrew W. Murray。从分子细胞生物学到模块化细胞生物学。《自然》，402(6761 Suppl)：C47–C52，1999。[HK11] Elad Hazan 和 Robert Krauthgamer。逼近最佳纳什均衡有多难？SIAM J. Comput.，40(1)：79–91，2011。[HKK+26] Jun-Ting Hsieh、Daniel M. Kane、Pravesh K. Kothari、Jerry Li、Sidhanth Mohanty 和 Stefan Tiegel。低次启发式的严格推论。arXiv 预印本 arXiv:2601.05850，2026。[HLL83] Paul W. Holland、Kathryn Blackmond Laskey 和 Samuel Leinhardt。随机块模型：初步探索。Soc. Netw.，5(2)：109–137，1983。[Hop18] Samuel Hopkins。统计推断与平方和方法。博士论文，康奈尔大学，2018。[HRH02] Peter D. Hoff、Adrian E. Raftery 和 Mark S. Handcock。社会网络分析中的潜在空间方法。J. Amer. Statist. Assoc.，97(460)：1090–1098，2002。[HRT07] Mark S. Handcock、Adrian E. Raftery 和 Jeremy M. Tantrum。社会网络的基于模型的聚类。J. Roy. Statist. Soc. Ser. A，170(2)：301–354，2007。[Hul22] Wasim Huleihel。在随机图中推断隐藏结构。IEEE Trans. Signal Inform. Process. Netw.，8：855–867，2022。[HW21] Justin Holmgren 和 Alexander S. Wein。低次猜想的反例。载于《第12届理论计算机科学创新会议》，LIPIcs 第185卷，第75:1–75:9页，2021。[HWX15] Bruce Hajek、Yihong Wu 和 Jiaming Xu。随机图社区检测的计算复杂度下界。载于《第28届学习理论会议论文集》，PMLR 第40卷，第899–928页，2015。[IS03] Yu I. Ingster 和 Irina A. Suslina。《高斯模型下的非参数拟合优度检验》，统计学讲义第169卷。施普林格，2003。[Jer92] Mark Jerrum。大团难以通过Metropolis过程找到。Random Structures Algorithms，3(4)：347–359，1992。[JL15] Tiefeng Jiang 和 Danning Li。矩形β-Laguerre系综的逼近与大偏差。J. Theoret. Probab.，28(3)：804–847，2015。[Joh01] Iain M. Johnstone。主成分分析中最大特征值的分布。Ann. Statist.，29(2)：295–327，2001。[JP00] Ari Juels 和 Marcus Peinado。为密码安全隐藏团。Des. Codes Cryptogr.，20(3)：269–280，2000。[JV26] He Jia 和 Aravindan Vijayaraghavan。低次方法无法预测鲁棒子空间恢复。arXiv 预印本 arXiv:2603.02594，2026。[Kra07] Ilia Krasikov。雅可比多项式的一个上界。J. Approx. Theory，149(2)：116–130，2007。[KSWY25] Dmitriy Kunisky、Daniel A. Spielman、Alexander S. Wein 和 Xifan Yu。有向图中排序社区的统计推断。载于《第57届ACM计算理论年会论文集》，第2107–2117页，2025。[Kuˇc95] Ludˇek Kuˇcera。图划分问题的期望复杂度。Discrete Appl. Math.，57(2-3)：193–212，1995。

[KWB22] Dmitriy Kunisky、Alexander S. Wein 和 Afonso S. Bandeira。假设检验计算困难性的札记：基于低次似然比的预测。载于《数学分析、应用与计算》，Springer Proc. Math. Stat. 第 385 卷，第 1–50 页。施普林格，2022年。[LKJ09] Daniel Lewandowski、Dorota Kurowicka 和 Harry Joe。基于藤结构和扩展洋葱法生成随机相关矩阵。J. Multivariate Anal., 100(9):1989–2001, 2009年。[LM00] B´eatrice Laurent 和 Pascal Massart。通过模型选择对二次泛函进行自适应估计。Ann. Statist., 28(5):1302–1338, 2000。[LMSY22] Siqi Liu、Sidhanth Mohanty、Tselil Schramm 和 Elizabeth Yang。高维稀疏随机几何图的检验阈值。载于第 54 届 ACM计算理论研讨会论文集，第 672–677 页，2022年。[LMSY23] Siqi Liu、Sidhanth Mohanty、Tselil Schramm 和 Elizabeth Yang。随机几何图中的局部与全局扩张。载于第 55 届 ACM计算理论研讨会论文集，第 817–825 页，2023年。[LPRZ25] Daniel Z. Lee、Francisco Pernice、Amit Rajaraman 和 Ilias Zadik。恢复植入子图的基本极限（扩展摘要）。载于第三十八届学习理论会议论文集，PMLR 第 291 卷，第 3578–3579 页，2025年。[LR23a] Suqi Liu 和 Mikl´os Z. R´acz。含噪高维随机几何图中的相变。Electron. J. Stat., 17(2):3512–3574, 2023年。[LR23b] Suqi Liu 和 Mikl´os Z. R´acz。潜空间图与相变的概率视角。Bernoulli, 29(3):2417–2441, 2023年。[LS23] Shuangping Li 和 Tselil Schramm。高斯混合块模型中的谱聚类。arXiv 预印本 arXiv:2305.00979，2023年。[Mat13] Ji˘r´ı Matou˘sek。《离散几何讲义》，Graduate Texts in Mathematics 第 212 卷。Springer-Verlag，2013年。[MMX21] Mehrdad Moharrami、Cristopher Moore 和 Jiaming Xu。植入匹配问题：相变与精确结果。Ann. Appl. Probab., 31(6):2663–2720, 2021。[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。Ann. Appl. Probab., 35(1):523–563, 2025年。[MO18] Carlo Morselli 和 Marie Ouellet。网络相似性与合谋。Soc. Netw., 55:21–30, 2018。[MST19] 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，PMLR 第 99 卷，第 2341–2371 页，2019年。[MWX26] Cheng Mao、Yihong Wu 和 Jiaming Xu。具有光滑核的随机几何图：尖锐检测阈值与谱猜想。arXiv 预印本 arXiv:2602.14998，2026。[MWZ23] Cheng Mao、Alexander S. Wein 和 Shenduo Zhang。植入稠密环的检测-恢复差距。载于第三十六届学习理论会议论文集，PMLR 第 195 卷，第 2440–2481 页，2023年。[MWZ25] Cheng Mao、Alexander S. Wein 和 Shenduo Zhang。植入稠密环的信息论阈值。IEEE Trans. Inform. Theory, 71(2):1266–1282, 2025年。[MZ24] Cheng Mao 和 Shenduo Zhang。基于率失真理论的潜在内积恢复不可能性。载于 2024 年第 60 届 Allerton通信、控制与计算会议，第 01–08 页，2024年。

[NC16] Mark E. J. Newman 和 Aaron Clauset。带注释网络的结构与推断。Nat. Commun., 7:11863, 2016。[New10] Mark E. J. Newman。《网络：导论》。牛津大学出版社，2010。[PCMP05] Carey E. Priebe、John M. Conroy、David J. Marchette 和 Youngser Park。安然图上的扫描统计量。Comput. Math. Organiz. Theor., 11:229–247, 2005。[Pen03] Mathew Penrose。《随机几何图》，牛津概率研究丛书第 5 卷。牛津大学出版社，2003。[PHT+21] Diogo Pacheco、Pik-Mai Hui、Christopher Torres-Lugo、Bao Tran Truong、Alessandro Flam-mini 和 Filippo Menczer。揭示社交媒体上的协同网络：方法与案例研究。载于国际万维网与社交媒体会议论文集，第 15 卷，第 455–466 页，2021。[PP20] Sandrine Peche 和 Vianney Perchet。社区检测对随机几何扰动的鲁棒性。载于《神经信息处理系统进展》第 33 卷，第 17827–17837 页，2020。[PW21] Elliot Paquette 和 Andrew Vander Werf。随机几何图与球面 Wishart 矩阵。arXiv 预印本 arXiv:2110.10785，2021。[RR19] Mikl´os Z. R´acz 和 Jacob Richey。从 Wishart 到高斯正交系综的平滑过渡。J. Theoret. Probab., 32(2):898–906, 2019年。[SM03] Victor Spirin 和 Leonid A. Mirny。分子网络中的蛋白质复合物与功能模块。Proc. Natl. Acad. Sci. U.S.A., 100(21):12123–12128, 2003。[Sod07] Sasha Sodin。高维集中测度边缘分布的尾部敏感高斯渐近。载于《泛函分析的几何方面》，Lecture Notes in Math. 第 1910 卷，第 271–295 页。施普林格，2007。[VAC15] Nicolas Verzelen 和 Ery Arias-Castro。稀疏随机网络中的社区检测。Ann. Appl. Probab., 25(6):3465–3510, 2015。[Ver18] Roman Vershynin。《高维概率》，剑桥统计与概率数学丛书第 47 卷。剑桥大学出版社，2018。[Vu07] Van H. Vu。随机矩阵的谱范数。《组合数学》，27(6):721–736, 2007。[War16] Lutz Warnke。典型有界差分方法。Combin. Probab. Comput., 25(2):269–299, 2016。[Wei25] Alexander S. Wein。统计学的计算复杂度：低次多项式带来的新见解。arXiv 预印本 arXiv:2506.10748, 2025年。[Wig58] Eugene P. Wigner。关于某些对称矩阵根的分布。Ann. of Math. (2), 67:325–327, 1958。[WK19] Johannes Wachs 和 J´anos Kert´esz。通过网络方法检测公开拍卖市场中的卡特尔。Sci. Rep., 9(1):10818, 2019年。[WM25] Timothy L. H. 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。[YZZ25] Xifan Yu、Ilias Zadik 和 Peiyuan Zhang。计数星形子图对检测任意植入子图而言是常数度最优的。Math. Stat. Learn., 8(1-2):105–164, 2025年。

A 辅助引理

A.1 集中不等式

这里，我们记录几个将在整个分析过程中用到的集中不等式。每个不等式都会在不同部分中引用，读者可在需要时查阅这些结果。

前两个结果（引理 A.1 和 A.2）针对某些行为良好的分布的多项式。这些结果可以由针对次高斯变量的多项式的一般集中不等式推导得出 [AW15，定理 1。4]。

引理 A.1（带符号三角形计数的集中性）。设 $G \sim \mathcal{G}(n,p)$。则存在常数 $C_{A.1} > 0$，使得对于 $f_{\mathrm{tri}}(G) = \sum_{i<j<\ell\in[n]}(G_{ij}-p)(G_{j\ell}-p)(G_{i\ell}-p)$，对任意 $t > 0$，

$$\mathbb{P}(|f_{\mathrm{tri}}(G)| \geq t) \leq 2\exp\left(-\frac{1}{C_{A.1}}\left(\frac{t^2\log^3(1/p)}{n^3} \wedge \frac{t\log^{3/2}(1/p)}{\sqrt{n}} \wedge t^{2/3}\log(1/p)\right)\right).$$

引理 A.2（高斯正交系综迹的集中性）。设 $Y \sim$ 高斯正交系综(n)，也就是说，$Y \in \mathbb{R}^{n\times n}$ 是一个对称矩阵，其元素独立服从 $Y_{ii} \sim \mathcal{N}(0,\ 2)$ $i \in [n]$ 和 $Y_{ij} = Y_{ji} \sim \mathcal{N}(0,\ 1)$ $i < j \in [n]$。则存在常数 $C_{A.2} > 0$，使得

$$f_2(Y) := \mathrm{Tr}(Y^2) \quad and \quad f_3(Y) := \mathrm{Tr}(Y^3),$$

对于任意 $t > 0$，

$$\mathbb{P}(|\mathrm{Tr}(Y^2) - (n^2+n)| \geq t) \leq 2\exp\left(-\frac{1}{C_{A.2}}\left(\frac{t^2}{n^2} \wedge \frac{t}{n}\right)\right),$$
$$\mathbb{P}(|\mathrm{Tr}(Y^3)| \geq t) \leq 2\exp\left(-\frac{1}{C_{A.2}}\left(\frac{t^2}{n^3} \wedge t^{2/3}\right)\right).$$

我们还使用以下鞅集中性的局部版本（即 Freedman 不等式）。其证明可见，例如，[War16，引理 2。2]。

引理 A.3（Freedman 不等式）。设 $\{M_i\}_{0\leq i\leq n}$ 是相对于滤过 $\{\mathcal{F}_i\}_{0\leq i\leq n}$ 的鞅，且 U_i、$i \in [n]$ 是 $\mathcal{F}_{i-1}$-可测随机变量，并满足对所有 $i \in [n]$ 均有 $M_i - M_{i-1} \leq U_i$。那么，对于任意 $t > 0$、$v > 0$ 和 $m > 0$，

$$\mathbb{P}\left(M_n - M_0 \geq t, \sum_{i=1}^{n}\mathrm{Var}[M_i - M_{i-1}|\mathcal{F}_{i-1}] \leq v, \max_{1\leq i\leq n} U_i \leq m\right) \leq \exp\left(-\left(\frac{t^2}{4v} \wedge \frac{t}{2m}\right)\right).$$

以下结果刻画了在 $\mathcal{G}(n,\ p)$ 和 $\mathcal{G}(n,\ p,\ d)$ 下的邻域分布有何不同。[9] 注意，对于单个顶点，这些分布的边缘分布相同；该引理比较的是在给定潜变量 $\{U_i\}$ 条件下的分布。

引理 A.4（球冠集中性；[LMSY22，引理 5.1，推论 6。1]）。假设$1/n \leq p \leq 1/2$，并考虑顶点 $\ell+1$ 在顶点 1、……、ℓ

[9]引理 A.4 中的不等式使用 d、$1/p$ 的对数项，而非 n 中的对数项，正如 [LMSY22，推论 61]所述。这只是为了避免后者中使用的形式条件 $d \leq \mathrm{poly}(n)$；当前形式可仿照 [LMSY22，推论 61] 的证明，由 [LMSY22，引理 51]推导得到。

在 $\mathcal{G}(n,p)$ 和 $\mathcal{G}(n,p,d)$ 下。也就是说，对于 $\gamma \in \{0,1\}^\ell$，令

$$\mathcal{Q}_{\ell+1}(\gamma) := p^{\sum_{i=1}^{\ell}\gamma_i}(1-p)^{\ell-\sum_{i=1}^{\ell}\gamma_i},$$

$$\widetilde{\mathcal{P}}_{\ell+1}(\gamma|U_1,\dots,U_\ell) := \mathbb{E}_{U_{\ell+1}}\left[\prod_{i\in[\ell]:\gamma_i=1}\mathbf{1}\{\langle U_{\ell+1},U_i\rangle \geq \tau\}\prod_{i\in[\ell]:\gamma_i=0}\mathbf{1}\{\langle U_{\ell+1},U_i\rangle < \tau\}\right],$$

其中 $U_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(\mathbb{S}^{d-1})$。那么存在一个常数 $C_{A.4}>0$ 使得对于所有 $t\geq 0$ 以及充分大的 n，

$$\mathbb{P}_{U_1,\dots,U_\ell}\left(\left|\frac{\widetilde{\mathcal{P}}_{\ell+1}(\gamma|U_1,\dots,U_\ell)}{\mathcal{Q}_{\ell+1}(\gamma)}-1\right|\geq t\right)\leq C_{A.4}\exp\left(-\frac{d(t^2\wedge 1)}{C_{A.4}M\log(1/p)\log(d/p)}+C_{A.4}\log n\right),$$

其中 $M=M(\ell,\ p,\ \gamma):=\left(\left(\sum_{i=1}^{\ell}\gamma_i\right)\log(1/p)+\ell p+\log(d/p)\right)\left(\sum_{i=1}^{\ell}\gamma_i+\ell p\right)$。

最后，我们记录以下标准结果。第一个结果关于 χ^2分布的集中性；第二个结果指出，球面 Wishart 的特征值集中在 1 附近。

引理 A.5 (关于 $\chi^2(d)$ 的集中性)。 设 $Z_i\sim\mathcal{N}(0,I_d)$ $i\in[k]$ 相互独立。那么其概率至少为 $1-2k\exp(-d/1000)$，

$$\frac{\|Z_i\|}{\sqrt{d}}\in[0.9,1.1]\ \textit{for all}\ i\in[k].$$

证明。 证明可由 [LM00，引理 1]以及并集界得出。 □

引理 A.6 （球面 Wishart 的谱）。 设 $d\geq k$和 $U=(U_1|\dots|U_k)\in\mathbb{R}^{d\times k}$满足 $U_i\sim\mathcal{U}(\mathbb{S}^{d-1})$ $i\in[k]$ 相互独立。那么存在一个常数 $C_{A.6}>0$，使得其概率至少为 $1-(2k+2)\exp(-k/C_{A.6})$，

$$\left\|U^TU-I_k\right\|_{\text{op}}\leq C_{A.6}\sqrt{\frac{k}{d}}.$$

证明。 记 $U_i=Z_i/\|Z_i\|$ 其中 $Z_i\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,I_d)$ $i\in[k]$ 且 $Z=(Z_1|\dots|Z_k)$。于是根据 [Ver18，定理 4.6。1]，我们有

$$\left\|\frac{Z^TZ}{d}-I_k\right\|_{\text{op}}=O\left(\sqrt{\frac{k}{d}}\right),$$

其概率至少为 $1-2\exp(-\Omega(k))$。由于 $U=ZD$ 其中 $D:=\text{diag}(1/\|Z_1\|,\ \dots\dots,\ 1/\|Z_k\|)$，我们有

$$\begin{aligned}\left\|U^TU-I_k\right\|_{\text{op}}&=\left\|DZ^TZD-I_k\right\|_{\text{op}}\\&\leq\left\|\sqrt{d}D\left(\frac{Z^TZ}{d}-I_k\right)\sqrt{d}D\right\|_{\text{op}}+\left\|dD^2-I_k\right\|_{\text{op}}\\&=O\left(\sqrt{\frac{k}{d}}\right),\end{aligned}$$

以至少 $1-(2k+2)\exp(-\Omega(k))$ 的概率，其中我们用到了 $\mathbb{P}(|\|Z_1\|^2-d|>4\sqrt{dk})\leq 2\exp(-\Omega(k))$ [LM00，引理 1] 以及并集界。

□

A.2 有符号子图计数的上界

除了一些特定情形外，在随机几何图中直接计算有符号子图计数的期望很困难。近期的一项工作 [BB24b] 为此提供了一个简洁的解决方案，给出了适用于所有中等规模子图的上界，且该上界仅取决于其顶点数和边数。

引理 A.7（有符号子图计数的上界； [BB24b，定理 1。1]）。假设存在一个常数 $\delta > 0$ 使得 $d \geq n^{\delta}$ 且 $n^{-1+\delta} \leq p \leq 1/2$。那么存在一个常数 $C_{A.7} = C_{A.7}(\delta) > 0$，使得以下结论成立：对于任意满足 H 的连通图$C_{A.7}v(H)\,e(H)\log^{3/2} d \leq \sqrt{d}$,

$$\left|\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(H)}(G_{ij}-p)\right]\right| \leq (8p)^{e(H)}\left(\frac{C_{A.7}v(H)e(H)\log^{3/2} d}{\sqrt{d}}\right)^{\lceil (v(H)-1)/2\rceil}.$$

第 B 节第 4 节中的延后证明4

B.1 有符号三角形计数的方差（引理4.4）

我们将条件

$$\gamma_{\mathrm{tri}} \gg \sqrt{\mathrm{Var}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)]}\,, \tag{B.1}$$

写成关于特定参数 $\varepsilon = \varepsilon(p,\ d) > 0$ 的不等式，并利用其下界将其转化为关于 d 的条件。具体而言，定义

$$\varepsilon = \varepsilon(p,d) := \frac{\mathbb{P}_{G\sim\mathcal{G}(n,p,d)}(G_{12}G_{13}G_{23}=1|G_{23}=1)}{p^2} - 1\,.$$

利用与某个公共顶点相连的边之间的两两独立性（即 $\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[G_{12}G_{13}] = p^2$），带符号三角形的期望可简化为：

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[(G_{12}-p)(G_{13}-p)(G_{23}-p)] = \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[G_{12}G_{13}G_{23}] - p^3 = p^3\varepsilon\,,$$

其中，最后一个等式由 $\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[G_{12}G_{13}G_{23}] = p\mathbb{P}_{G\sim\mathcal{G}(n,p,d)}(G_{12}G_{13}G_{23}=1|G_{23}=1)$ 和 ε 的定义得出。于是，我们得到

$$\gamma_{\mathrm{tri}} = \frac{1}{2}\mathbb{E}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)] = \frac{1}{2}\binom{n}{3}\left(\frac{k}{n}\right)^3\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[(G_{12}-p)(G_{13}-p)(G_{23}-p)] = \Omega(k^3p^3\varepsilon)\,, \tag{B.2}$$

其中，第一个等式由引理4.1得出，第二个等式则由上方显示的等式得出。沿用 [LMSY22 中的推导，第 21–25] 个方程结合引理 4.1，对于 $T_{ijl} := (G_{ij}-p)(G_{jl}-p)(G_{il}-p) - \mathbb{E}_{G\sim\mathcal{P}}[(G_{ij}-p)(G_{jl}-p)(G_{il}-p)]$ 我们得到

$$\mathrm{Var}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)] = \binom{n}{3}\mathbb{E}_{G\sim\mathcal{P}}[(\overline{T}_{123})^2] + 12\binom{n}{4}\mathbb{E}_{G\sim\mathcal{P}}[\overline{T}_{123}\overline{T}_{124}] + 30\binom{n}{5}\mathbb{E}_{G\sim\mathcal{P}}[\overline{T}_{123}\overline{T}_{145}]\,. \tag{B.3}$$

对于式（B.3）中的第一项，我们有

$$\begin{aligned}\mathbb{E}_{G\sim\mathcal{P}}[(\overline{T}_{123})^2] &\leq \mathbb{E}_{G\sim\mathcal{P}}[(T_{123})^2]\\ &= (1-2p)^3\mathbb{E}_{G\sim\mathcal{P}}[G_{12}G_{23}G_{13}] + 3(1-2p)^2p^4 + 3(1-2p)p^5 + p^6\\ &= (1-2p)^3p^3(1+(k/n)^3\varepsilon) + 3(1-2p)^2p^4 + 3(1-2p)p^5 + p^6\\ &= O\left(p^3(1+(k/n)^3\varepsilon)\right)\,.\end{aligned} \tag{B.4}$$

对于式（B.3）中的第二项，我们首先有

$$\begin{aligned}\mathbb{E}_{G\sim\mathcal{P}}[\overline{T}_{123}\overline{T}_{124}] &\le \mathbb{E}_{G\sim\mathcal{P}}[T_{123}T_{124}] \\ &= \left(\frac{k}{n}\right)^4 \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[T_{123}T_{124}] \\ &\le \left(\frac{k}{n}\right)^4 \mathbb{E}_{\langle U_1,U_2\rangle}\left[\left(\mathbb{E}_{U_1}[G_{13}G_{23}|\,\langle U_1,U_2\rangle] - p^2\right)^2\right],\end{aligned}$$

其中等式由引理 4.1 得出，第二个不等式则由 [BDER16 的方程 24] 得出。最后一项中的期望等于在 $\mathcal{G}(n,p,d)$ 下的带符号四环计数，因为：

$$\begin{aligned}&\mathbb{E}_{\langle U_1,U_2\rangle}\left[\left(\mathbb{E}[G_{13}G_{23}|\,\langle U_1,U_2\rangle] - p^2\right)^2\right] \\ &= \mathbb{E}_{\langle U_1,U_2\rangle}\left[\left(\mathbb{E}_{U_3}[(G_{13}-p)(G_{23}-p)|U_1,U_2]\right)^2\right] \\ &= \mathbb{E}_{\langle U_1,U_2\rangle}\left[\mathbb{E}_{U_3,U_4}[(G_{13}-p)(G_{23}-p)(G_{14}-p)(G_{24}-p)|U_1,U_2]\right] \\ &= \mathbb{E}_{\langle U_1,U_2\rangle}\left[\mathbb{E}_{U_3,U_4}[(G_{13}-p)(G_{23}-p)(G_{14}-p)(G_{24}-p)|\,\langle U_1,U_2\rangle]\right],\end{aligned}$$

其中，第一个和最后一个等式由旋转不变性得出（即，对 (U_1,U_2) 进行条件化等价于对 $\langle U_1,U_2\rangle$ 进行条件化），第二个等式则源于给定 (U_1,U_2) 时，$(G_{13}-p)(G_{23}-p)$ 与 $(G_{14}-p)(G_{24}-p)$ 的条件独立性。因此，

$$\begin{aligned}\mathbb{E}_{G\sim\mathcal{P}}[\overline{T}_{123}\overline{T}_{124}] &\le \left(\frac{k}{n}\right)^4 \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}[(G_{13}-p)(G_{23}-p)(G_{14}-p)(G_{24}-p)] \\ &= O\left(\left(\frac{k}{n}\right)^4 \frac{p^4\log^2(1/p)}{d}\right),\end{aligned} \tag{B.5}$$

最后一行由命题 6.1 得出。需要指出的是，[LMSY22] 中针对这一部分的论证$G_{13}-p$假设$G_{23}-p$与$G_{14}-pG_{24}-p$在给定G_{12}时条件独立；然而，这一假设并不成立，因为这两个量都依赖于共同的潜在向量 U_1 和U_2。因此，我们通过命题 6.1采用另一种方法，该方法只需作出一个较弱的假设：d 足够大，且 $d \ge (5\log(1/p))^4$。最后，对于（B.3）中的第三项，我们有

$$\mathbb{E}_{G\sim\mathcal{P}}[\overline{T}_{123}\overline{T}_{145}] \le \mathbb{E}_{G\sim\mathcal{P}}[T_{123}T_{145}] = \left(\frac{k}{n}\right)^5 p^6\varepsilon^2\,. \tag{B.6}$$

将 (B.4)、(B.5) 和 (B.6) 代入 (B.3，得到

$$\mathrm{Var}_{G\sim\mathcal{P}}[f_{\mathrm{tri}}(G)] = O(n^3p^3 + k^3p^3\varepsilon + k^4p^4\log^2(1/p)/d + k^5p^6\varepsilon^2)\,.$$

将其与 (B.2) 比较可知，要证明 (B.1)，只需证明

$$k^6p^6\varepsilon^2 \gg n^3p^3 + k^3p^3\varepsilon + k^4p^4\log^2(1/p)/d + k^5p^6\varepsilon^2\,. \tag{B.7}$$

显然，$k^6p^6\varepsilon^2 \gg k^5p^6\varepsilon^2$ 成立。此外，由引理 4.3，可得 $\varepsilon = \Omega(\log^{3/2}(1/p)/\sqrt{d})$。因此，(B.7) 的一个充分条件是

$$d \ll \frac{k^6p^3\log^3(1/p)}{n^3} \wedge k^6p^6\log^3(1/p) \quad \text{and} \quad k^2p^2\log(1/p) \gg 1\,.$$

如果 $p \ge 1/n$，那么只要 $d \ll k^6p^3\log^3(1/p)/n^3$，关于 d 的第一个不等式就成立。只要 $p \gg 1/(k\sqrt{\log k})$，关于 k，p 的第二个不等式就成立。

B.2 有符号楔形计数的典型行为（引理4.5）

不失一般性，令 $S = [s]$ 和 $S_0 = [k^-]$。在本小节中，我们始终将 G 的分布视为 $\mathcal{P}_{[s]}$，并为简洁起见省略记号 $G \sim \mathcal{P}_{[s]}$。

验证 (2.7)。我们先考虑范围代理变量上的事件 (2.7)，其补集的概率至多为（由并集界和马尔可夫不等式得出）

$$\begin{aligned}
&\sum_{i<j\in[k^-]} \mathbb{P}\left(\left|\sum_{\ell<i,\ell\in[k^-]} (G_{\ell i}-p)(G_{\ell j}-p)\right| > B\right) \\
&\le k^2 \max_{i<j\in[k^-]} \mathbb{P}\left(\left|\sum_{\ell<i,\ell\in[k^-]} (G_{\ell i}-p)(G_{\ell j}-p)\right| > B\right) \\
&\le k^2 \frac{\max_{i<j\in[k^-]} \mathbb{E}\left[\left(\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right)^{2m}\right]}{B^{2m}},
\end{aligned} \tag{B.8}$$

对任意 $m \ge 1$；稍后，我们将选择 $m = \lceil \log k \rceil$。为了得到楔形计数矩的上界，考虑任意固定的 $i < j \in [k^-]$。于是

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{\ell<i,\ell\in[k^-]} (G_{\ell i}-p)(G_{\ell j}-p)\right)^{2m}\right] \\
&= \sum_{\ell_1,\dots,\ell_{2m}} \mathbb{E}\left[\prod_{a=1}^{2m}(G_{\ell_a i}-p)(G_{\ell_a j}-p)\right] \le \sum_{b=1}^{2m} \sum_{|\{\ell_1,\dots,\ell_{2m}\}|=b} \left|\mathbb{E}\left[\prod_{a=1}^{2m}(G_{\ell_a i}-p)(G_{\ell_a j}-p)\right]\right|,
\end{aligned} \tag{B.9}$$

其中每个 $\ell_1, \dots, \ell_{2m}$ 都小于 i。

下一步是将 (B.9) 中的每个求和项表示为真子图计数的形式，以应用引理 A.7。考虑任意固定的 $\ell_1, \dots, \ell_{2m}$，满足 $|\{\ell_1, \dots, \ell_{2m}\}| = b$。令不同元素为 $\ell_1, \dots, \ell_b$（不失一般性），

$$\left|\mathbb{E}\left[\prod_{a=1}^{2m}(G_{\ell_a i}-p)(G_{\ell_a j}-p)\right]\right| = \left|\mathbb{E}[(G_{\ell_1 i}-p)^{m_1}(G_{\ell_1 j}-p)^{m_1}\dots(G_{\ell_b i}-p)^{m_b}(G_{\ell_b j}-p)^{m_b}]\right|, \tag{B.10}$$

其中 $m_a \ge 1$（各自取决于 ℓ_1、……、ℓ_{2m}）和 $\sum_a m_a = 2m$。由于每个 $G_{\ell_a i}$，$G_{\ell_a j} \in \{0, 1\}$，我们总能将任何高阶项表示为低阶项。特别地，我们有

$$\begin{aligned}
(x-p)^{m_a} &= ((1-p)^{m_a} - (-p)^{m_a})(x-p) + p(1-p)((1-p)^{m_a-1} - (-p)^{m_a-1}) \\
&=: C_a(x-p) + D_a,
\end{aligned}$$

对于 $x \in \{0,1\}$。由于 $C_a,\ D_a \geq 0$（由 $p \leq 1/2$得出）且 $C_a \leq 1,\ D_a \leq p,$ （B.10）至多为

$$
\begin{aligned}
&|\mathbb{E}[\sum_{U\subseteq[b]}\sum_{U'\subseteq[b]}\left(\prod_{a\in U}C_a\right)\left(\prod_{a'\in U'}C_{a'}\right)\left(\prod_{a\notin U}D_a\right)\left(\prod_{a'\notin U'}D_{a'}\right)\mathbb{E}\left[\prod_{a\in U}(G_{ai}-p)^{m_1}(G_{\ell_b i}-p)^{m_1}\cdots(G_{\ell_b i}-p)^{m_b}\prod_{a'\in U'}(G_{a'j}-p)^{m_b}\right]]| \\
&=\left|\sum_{U\subseteq[b]}\sum_{U'\subseteq[b]}\left(\prod_{a\in U}C_a\right)\left(\prod_{a'\in U'}C_{a'}\right)\left(\prod_{a\notin U}D_a\right)\left(\prod_{a'\notin U'}D_{a'}\right)\mathbb{E}\left[\prod_{a\in U}(G_{ai}-p)\prod_{a'\in U'}(G_{a'j}-p)\right]\right| \\
&\leq\left|\sum_{U\subseteq[b]}\sum_{U'\subseteq[b]}\left(\prod_{a\in U}C_a\right)\left(\prod_{a'\in U'}C_{a'}\right)\left(\prod_{a\notin U}D_a\right)\left(\prod_{a'\notin U'}D_{a'}\right)\right|\mathbb{E}\left[\prod_{a\in U}(G_{ai}-p)\prod_{a'\in U'}(G_{a'j}-p)\right] \\
&\leq\left|\sum_{U\subseteq[b]}\sum_{U'\subseteq[b]}p^{b-|U|+b-|U'|}(8p)^{|U|+|U'|}\right|\leq 2^b\times 2^b\times(8p)^{2b}=(16p)^{2b}\,.
\end{aligned}
$$

这里，第一个不等式由 $C_a, D_a \geq 0$ 得出，第二个不等式则由引理 A.7 和 $C_a \leq 1, D_a \leq p$ 得出。具体来说，与边 $\{ai\colon a \in U\} \cup \{a'j\colon a' \in U'\}$ 对应的图至多有 2 个连通分量，且含有 $|U| + |U'| \leq 4m$ 条边。若 $|U| = |U'| = 0$，则期望为 $1 = (8p)^{|U|+|U'|}$；否则，对每个连通分量应用引理 A.7，并利用 $C_{A.7}v(H)e(H)\log^{3/2} d/\sqrt{d} \leq C_{A.7}(4m)^2(\log^{3/2} d)/\sqrt{d} \leq 1$（对所有充分大的 d，且 $m = \lceil \log k \rceil$），即可得到相应的上界 $(8p)^{|U|+|U'|}$。由于只要 $|\{\ell_1, \ldots, \ell_{2m}\}| = b$，最终上界 $(16p)^{2b}$ 就成立，将其代入 (B.9) 可得

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right)^{2m}\right] &\leq \sum_{b=1}^{2m}\sum_{|\{\ell_1,\ldots,\ell_{2m}\}|=b}(16p)^{2b} \\
&\leq \sum_{b=1}^{2m}\binom{j-2}{b}b^{2m}(16p)^{2b} \\
&\leq \sum_{b=1}^{2m}k^b b^{2m}(16p)^{2b}=\sum_{b=1}^{2m}b^{2m}(256kp^2)^b\,.
\end{aligned}
$$

这里，第二个不等式是由于从 $[i-1] \subseteq [j-2]$ 中选取了 $\ell_1, \ldots, \ell_{2m}$，而下一个不等式则由 $j \leq k$ 得出。

若 $256kp^2 \leq 1$，最后一项至多为 $\sum_{b=1}^{2m} b^{2m} \leq (2m)^{2m+1}$；否则，它至多为 $\sum_{b=1}^{2m} b^{2m}(256kp^2)^{2m} \leq (2m)(512kp^2m)^{2m}$。因此，对于 (B.8)，我们得到一个上界：

$$
k^2\frac{(2m)^{2m+1}\vee(2m)(512kp^2m)^{2m}}{B^{2m}}\,.
$$

最后，我们选择 m 和 B，使得该项为 $o(1)$，如 $k \to \infty$ 所示。特别地，$m = \lceil \log k \rceil$ 和 $B = 2048kp^2m + 8m$ 是成立的。

验证(2.6)。关于方差替代量的结果可由切比雪夫不等式的简单应用得到（但计算较为繁琐）。为此，我们需要计算（2.6）中统计量的均值和方差。记作 $W_{\ell ij} := (G_{\ell i} - p)^2(G_{\ell j} - p)^2$ 以及 $Q_{\ell qij} := (G_{\ell i} - p)(G_{\ell j} - p)(G_{qi} - p)(G_{qj} - p)$，（2.6）中的统计量等于

$$
\begin{aligned}
&\sum_{i<j\in[k^-]}\sum_{\ell<i,\ell\in[k^-]}W_{\ell ij}+2\sum_{i<j\in[k^-]}\sum_{\ell<q<i,\ell\in[k^-],q\in[k^-]}Q_{\ell qij} \\
&=\sum_{\ell<i<j\in[k^-]}W_{\ell ij}+2\sum_{\ell<q<i<j\in[k^-]}Q_{\ell qij}\,.
\end{aligned}
\tag{B.11}
$$

这由展开可得：

$$\begin{aligned}
&\sum_{i<j\in[k^-]}\left(\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\right)^2\\
&=\sum_{i<j\in[k^-]}\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)\sum_{q<i,q\in[k^-]}(G_{qi}-p)(G_{qj}-p)\\
&=\sum_{i<j\in[k^-]}\sum_{\ell<i,\ell\in[k^-]}(G_{\ell i}-p)^2(G_{\ell j}-p)^2\\
&\quad+2\sum_{i<j\in[k^-]}\sum_{\ell<q<i,\ell\in[k^-],q\in[k^-]}(G_{\ell i}-p)(G_{\ell j}-p)(G_{qi}-p)(G_{qj}-p)\,,
\end{aligned}$$

其中，最后一个等式是通过根据 $\ell=q$ 或 $\ell\neq q$ 分类得出的。

我们首先计算 (B.11) 的均值。根据 $\mathbb{E}[W_{\ell ij}]=p^2(1-p)^2$ 以及关于长度为 4 的环的命题 6.1 ，(B.11) 的均值至多为

$$\binom{k^-}{3}p^2(1-p)^2+2\binom{k^-}{4}C_{6.1}^4\frac{p^4\log^2(1/p)}{d}\le\frac{\sigma^2}{2}\,, \tag{B.12}$$

其中，该不等式是由 (2.6) 中 σ^2 的取值选择得出的。

对于 (B.11) 的方差，考虑中心化变量 $W_{\ell ij}:=W_{\ell ij}-\mathbb{E}[W_{\ell ij}]$ 和 $Q_{\ell qij}:=Q_{\ell qij}-\mathbb{E}[Q_{\ell qij}]$。那么，(B.11) 的方差等于

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{\ell<i<j\in[k^-]}\overline{W}_{\ell ij}+2\sum_{\ell<q<i<j\in[k^-]}\overline{Q}_{\ell qij}\right)^2\right]\\
&\le 2\mathbb{E}\left[\left(\sum_{\ell<i<j\in[k^-]}\overline{W}_{\ell ij}\right)^2\right]+8\mathbb{E}\left[\left(\sum_{\ell<q<i<j\in[k^-]}\overline{Q}_{\ell qij}\right)^2\right].
\end{aligned} \tag{B.13}$$

不等式之后的期望可利用引理 A.7 得到上界；可以证明（详见附录 B.2.1 ）(B.13) 至多为

$$O\left(k^4p^3+\frac{k^5p^6\log^3 d}{d}+\frac{k^7p^8\log^{9/2}d}{d^{3/2}}\right). \tag{B.14}$$

最后，我们利用均值 (B.12) 和方差 (B.14) 的界应用切比雪夫不等式。令 X 为 (B.11) 中的随机变量之和。那么，(2.6) 的补事件的概率为

$$\mathbb{P}(X>\sigma^2)\le\mathbb{P}(X-\mathbb{E}[X]>\sigma^2/2)\le 4\mathrm{Var}[X]/\sigma^4\,.$$

因此，只需证明 $\sigma^4\gg\mathrm{Var}[X]$。由于 σ^2 的阶为 $\Omega(k^3p^2+k^4p^4\log^2(1/p)/d)$，只需证明

$$(k^3p^2)^2+\left(\frac{k^4p^4\log^2(1/p)}{d}\right)^2\gg k^4p^3+\frac{k^5p^6\log^3 d}{d}+\frac{k^7p^8\log^{9/2}d}{d^{3/2}}\,. \tag{B.15}$$

对于右侧的每一项，我们来评估这个渐近不等式（B.15）。对于第一项，$(k^4p^4\log^2(1/p)/d)^2 \gg k^4p^3$ 等价于

$$d \ll k^2p^{5/2}\log^2(1/p)\,.$$

这由定理的假设 $d \ll k^2p^3\log^3(1/p)$，结合 $p^{1/2}\log(1/p) = O(1)$ 得出。对于其余项，在 (B.15) 的左侧应用算术-几何平均不等式（权重分别为 $(1/2,1/2)$、$(1/4,3/4)$），可得其至少为

$$\Omega\left(\frac{k^7p^6\log^2(1/p)}{d} \vee \frac{k^{15/2}p^7\log^3(1/p)}{d^{3/2}}\right).$$

要使其大于 $\left(\frac{k^5p^6}{d}\log^3 d + \frac{k^7p^8}{d^{3/2}}\log^{9/2} d\right)$ 中的其余项B.15)，应满足

$$k \gg \frac{\log^{3/2} d}{\log(1/p)} \vee \frac{p^2\log^9 d}{\log^6(1/p)}\,.$$

根据条件 $p \geq n^{-1+\delta}$ 和 $d = \widetilde{O}(k^2p^3)$，可见右侧至多关于 n为多对数级，而左侧至少关于 n为多项式级。因此，对于每个固定的 $\delta > 0$，当 n 足够大时，该结论均成立。

B.2.1 带符号楔形计数的方差

这里记录先前关于 (B.13) 至多为 (B.14) 的断言的详细证明。

$\overline{W}_{\ell ij}$ 项。 我们有

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{\ell<i<j\in[k^-]}\overline{W}_{\ell ij}\right)^2\right] \\
&= \sum_{\ell<i<j\in[k^-]}\sum_{\ell'<i'<j'\in[k^-]}\mathbb{E}\left[\overline{W}_{\ell ij}\overline{W}_{\ell' i' j'}\right] \\
&\leq k^3\mathbb{E}\left[\left(\overline{W}_{123}\right)^2\right] + k^4\left|\mathbb{E}\left[\overline{W}_{123}\overline{W}_{124}\right]\right| + k^4\left|\mathbb{E}\left[\overline{W}_{123}\overline{W}_{234}\right]\right| + k^4\left|\mathbb{E}\left[\overline{W}_{134}\overline{W}_{234}\right]\right|,
\end{aligned}$$

因为对于 ℓ、i、j 与 ℓ'、i'、j' 的所有其他组合，根据推论 4.2，其期望为 0。由 $(x-p)^2 = (1-2p)(x-p) + p(1-p)$（对 $x \in \{0,1\}$），我们有 $W_{\ell ij} = (1-2p)^2(G_{\ell i}-p)(G_{\ell j}-p) + (1-2p)p(1-p)(G_{\ell i}-p+G_{\ell j}-p)$。因此

$$\begin{aligned}
\mathbb{E}\left[\left(\overline{W}_{123}\right)^2\right] &= \mathbb{E}\left[\left((1-2p)^2(G_{12}-p)(G_{13}-p) + (1-2p)p(1-p)(G_{12}-p+G_{13}-p)\right)^2\right] \\
&\leq 3\mathbb{E}[(G_{12}-p)^2(G_{13}-p)^2] + 3p^2\mathbb{E}[(G_{12}-p)^2] + 3p^2\mathbb{E}[(G_{13}-p)^2] \\
&= O(p^2)\,,
\end{aligned}$$

其中不等式由 $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ 得出。通过类似的展开，我们有

$$\begin{aligned}
\mathbb{E}\left[\overline{W}_{123}\overline{W}_{124}\right] =& \mathbb{E}\Big[\left((1-2p)^2(G_{12}-p)(G_{13}-p) + (1-2p)p(1-p)(G_{12}-p+G_{13}-p)\right) \\
&\times\left((1-2p)^2(G_{12}-p)(G_{14}-p) + (1-2p)p(1-p)(G_{12}-p+G_{14}-p)\right)\Big] \\
&= p^2(1-p)^2(1-2p)^2\mathbb{E}[(G_{12}-p)^2] \\
&= O(p^3)\,,
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[\overline{W}_{123}\overline{W}_{234}\right] &= \mathbb{E}\Big[\left((1-2p)^2(G_{12}-p)(G_{13}-p)+(1-2p)p(1-p)(G_{12}-p+G_{13}-p)\right)\\ &\quad\times\left((1-2p)^2(G_{23}-p)(G_{24}-p)+(1-2p)p(1-p)(G_{23}-p+G_{24}-p)\right)\Big]\\ &= p(1-p)(1-2p)^3\mathbb{E}[(G_{12}-p)(G_{13}-p)(G_{23}-p)]\\ &= O(p^4\log^{3/2}(1/p)/d^{1/2})\,,\end{aligned}$$

其中最后一行由命题6.1 得出。此外,

$$\begin{aligned}\mathbb{E}\left[\overline{W}_{134}\overline{W}_{234}\right] &= \mathbb{E}\Big[\left((1-2p)^2(G_{13}-p)(G_{14}-p)+(1-2p)p(1-p)(G_{13}-p+G_{14}-p)\right)\\ &\quad\times\left((1-2p)^2(G_{23}-p)(G_{24}-p)+(1-2p)p(1-p)(G_{23}-p+G_{24}-p)\right)\Big]\\ &= (1-2p)^4\mathbb{E}[(G_{13}-p)(G_{14}-p)(G_{23}-p)(G_{24}-p)]\\ &= O(p^4\log^2(1/p)/d)\,,\end{aligned}$$

再次根据命题 6.1。因此,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{\ell<i<j\in[k^-]}\overline{W}_{\ell ij}\right)^2\right] &= O\left(k^3p^2+k^4p^3+\frac{k^4p^4\log^{3/2}(1/p)}{d^{1/2}}+\frac{k^4p^4\log^2(1/p)}{d}\right)\\ &= O(k^4p^3)\,,\end{aligned}\tag{B.16}$$

其中最后的等 质量由 $kp\gtrsim 1$, $p\log^{3/2}(1/p)/d^{1/2}=o(1)$, 以及 $p\log^2(1/p)/d=o(1)$ 得出 .

$\overline{Q}_{\ell qij}$ 项。类似地,我们有

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{\ell<q<i<j\in[k^-]}\overline{Q}_{\ell qij}\right)^2\right] &= \sum_{\ell<q<i<j\in[k^-]}\sum_{\ell'<q'<i'<j'\in[k^-]}\mathbb{E}\left[\overline{Q}_{\ell qij}\overline{Q}_{\ell'q'i'j'}\right]\\ &= \sum_{\ell<q<i<j\in[k^-]}\sum_{\ell'<q'<i'<j'\in[k^-]}(\mathbb{E}[Q_{\ell qij}Q_{\ell'q'i'j'}]-(\mathbb{E}[Q_{1234}])^2)\,.\end{aligned}$$

我们使用引理 A.7, 根据由 H 的构型形成的并图进行分类 ℓ, q, i, j和 ℓ', q', i', j'。下文记录了基于($|\mathbb{E}[Q_{\ell qij}Q_{\ell'q'i'j'}]|$ 的不同取值(v(H) e(H))的上界,以及关于指标的不同情形数量。[10]

- $(v(H)\ e(H))=(4,\ 4)$: $O(p^4)$ 的上界成立; 共有 $O(k^4)$ 种此类情形。
- $(v(H)\ e(H))=(4,\ 6)$: $O(p^6(\log\ d)^3/d)$的上界成立; 此类情形有 $O(k^4)$种。
- $(v(H)\ e(H))=(5,\ 6)$: $O(p^6(\log\ d)^3/d)$的上界成立; 此类情形有 $O(k^5)$种。
- $(v(H)\ e(H))=(5,\ 7)$: $O(p^7(\log\ d)^3/d)$的上界成立; 有 $O(k^5)$种此类情况。

[10]在此计算中, $Q_{\ell qij}Q_{\ell'i'j'q'}$可能包含指数为 2 的因子(例如, $(G_{\ell i}-p)^2$), 在这种情况下,我们考虑基于其低阶展开的上界,即$(G_{\ell i}-p)^2=(1-2p)(G_{\ell i}-p)+p(1-p)$。未包含关于 ℓ, q, i, j 以及 ℓ', q', i', j'的环不共享公共顶点的情形,因为其中每一项的贡献都是 $\mathbb{E}[Q_{\ell qij}Q_{\ell'q'i'j'}]-(\mathbb{E}[Q_{1234}])^2=0$。

- $(v(H),\ e(H)) = (6,\ 7)$：$O(p^7(\log d)^{9/2}/d^{3/2})$的上界成立；有 $O(k^6)$种此类情况。
- $(v(H),\ e(H)) = (6,\ 8)$（从现在起，两个 4-环不再共享一条边）：$O(p^8(\log d)^{9/2}/d^{3/2})$ 的上界成立；这样的情形有 $O(k^6)$ 种。
- $(v(H),\ e(H)) = (7,\ 8)$：$O(p^8(\log d)^{9/2}/d^{3/2})$ 的上界成立；这样的情形有 $O(k^7)$ 种。

结合以上结果，我们有[11]

$$\mathbb{E}\left[\left(\sum_{\ell<q<i<j\in[k^-]}\overline{Q}_{\ell qij}\right)^2\right] = O\left(k^4p^4 + \frac{k^5p^6\log^3 d}{d} + \frac{k^7p^8\log^{9/2} d}{d^{3/2}}\right). \tag{B.17}$$

结合式（B.16）和（B.17）（再加上 $k^4p^3 \gtrsim k^4p^4$），可得（B.14）。

C 第 5 节中的延后证明5

C.1 Wishart 与球面 Wishart 之间的全变差距离（命题 5.1）

设 $K := \binom{k}{2}$, $Y \in \mathbb{R}^K$ 是一个向量，其元素为 $\frac{\langle Z_i\ Z_j\rangle}{\|Z_i\|\|Z_j\|}, i<j\in[k]$，而 $W\in\mathbb{R}^K$ 是一个向量，其元素为 $\frac{\langle Z_i\ Z_j\rangle}{d}, i<j\in[k]$。

首先，将 W 表示为 Y 的一个近似各向同性线性变换。具体而言，设 $D\in\mathbb{R}^{k\times k}$ 为对角矩阵，其中 $\|Z_i\|/\sqrt{d}$ 是其第 (i,i) 个元素；设 $T_D\colon \mathbb{R}^K\to\mathbb{R}^K$ 是一个线性映射，定义为 $(T_D(x))_{ij} := D_{ii}D_{jj}x_{ij}, i<j\in[k]$。作为关键事实，可以观察到 $W = T_D(Y)$，并且 D 与 Y 相互独立（这是因为 $Z_i/\|Z_i\|$ 与 $\|Z_i\|$ 相互独立）。这些事实共同推出

$$W \overset{\mathrm{d}}{=} \mathbb{E}_D[T_D(Y)]\,.$$

此外，映射 T_D 接近恒等映射：定义 $\mathcal{E}$ 为 $\{D\in\mathcal{E}\} = \{D_{ii}\in[0.9,1.1]$ 对所有 $i\}$，我们有 $\mathbb{P}(D\notin\mathcal{E}) = o(1)$，由引理 A.5可得。于是，

$$\begin{aligned} d_{\mathrm{TV}}(Y,W) &= d_{\mathrm{TV}}(Y,\mathbb{E}_D[T_D(Y)]) \\ &\overset{(a)}{\le} \mathbb{E}_D[d_{\mathrm{TV}}(Y,T_D(Y))] \\ &\overset{(b)}{\le} \mathbb{P}(D\notin\mathcal{E}) + \mathbb{E}_D[d_{\mathrm{TV}}(Y,T_D(Y))\mathbf{1}\{D\in\mathcal{E}\}] \\ &\le o(1) + \mathbb{E}_D[d_{\mathrm{TV}}(Y,T_D(Y))\mathbf{1}\{D\in\mathcal{E}\}] \\ &\overset{(c)}{\le} o(1) + \sqrt{\mathbb{E}_D[d_{\mathrm{TV}}(Y,T_D(Y))^2\mathbf{1}\{D\in\mathcal{E}\}]}\,, \end{aligned} \tag{C.1}$$

其中，(a)由詹森不等式和全变差距离的凸性成立；(b)由 $d_{\mathrm{TV}}\le 1$成立；(c)由柯西-施瓦茨不等式成立。

为了给 $d_{\mathrm{TV}}(Y,T_D(Y))$（出现在 (C.1) 的最后一行）求上界，我们使用它们密度的显式公式。对于 $y\in\mathbb{R}^K$，定义 $y\in\mathbb{R}^{k\times k}$ 为一个对称矩阵，其对角线元素为 0，非对角线元素为 y_{ij}。于是，Y 在 y 处的密度如下[12] (例如，参见 [PW21，引理 12])：

$$f(y) = \det(I_k+\overline{y})^{(d-k-1)/2}\mathbf{1}\{I_k+\overline{y}\succ 0\}/Z(k,d)\,, \tag{C.2}$$

[11]注意，$k^6p^7(\log d)^{9/2}/d^{3/2} \ll k^7p^8(\log d)^{9/2}/d^{3/2}$ 成立，这等价于 $kp\gg 1$。为此，由定理 2.3 的假设可得 $k^3p^3\log^3(1/p) \ge k^2p^3\log^3(1/p) = \Omega(n^\delta)$，且$\log(1/p) = O(\log n)$，两者合起来意味着 $kp = \Omega(n^{\delta/3}/\log n)$。[12]文献中也将该分布称为 LKJ 分布 [LKJ09]。

其中， $Z(k,d)$ 是一个归一化常数。作为 Y 的线性变换（其中 $D \in \mathcal{E}$），$T_D(Y)$ 的密度 g_D 在 w 处为

$$g_D(w) = f\left(T_D^{-1}(w)\right)\left|\det \frac{\partial T_D^{-1}(w)}{\partial w}\right| = f\left(T_D^{-1}(w)\right)\det(D)^{-(k-1)}\,. \tag{C.3}$$

注意，对于 $D \in \mathcal{E}$, T_D^{-1} 是定义良好的。

为了避免比值 g_D/f 无界，我们在较小的集合 $\mathcal{E}'$ 中比较密度，该集合定义为

$$\mathcal{E}' := \{y \in \mathbb{R}^K : \|\overline{y}\|_{\mathrm{op}} \le 0.1\}\,.$$

值得注意的是，对于任意 $D \in \mathcal{E}$，$Y \in \mathcal{E}' \Leftrightarrow T_D(Y) \in T_D(\mathcal{E}')$ 和 $Y \in T_D(\mathcal{E}')$ 都是高概率事件。这是因为：根据引理 A.6 及 $k/d = o(1)$，$\mathbb{P}(Y \notin \mathcal{E}') \le 2\exp(-\Omega(k))$；类似地，由 $\left\|\overline{T_D^{-1}(Y)}\right\|_{\mathrm{op}} = \left\|D^{-1}YD^{-1}\right\|_{\mathrm{op}} \le \left\|D^{-1}\right\|^2_{\mathrm{op}}\left\|\overline{Y}\right\|_{\mathrm{op}} \le (1/0.9)^2\left\|\overline{Y}\right\|_{\mathrm{op}}$ 可得 $\mathbb{P}_Y(Y \notin \overline{T_D(\mathcal{E}')}) \le 2\exp(-\Omega(k))$。令 $r_D(y) := g_D(y)/f(y)$，

$$\begin{aligned} d_{\mathrm{TV}}(Y, T_D(Y)) &= \frac{1}{2}\int_{\mathbb{R}^K} |f(y) - g_D(y)|dy \\ &\le \frac{1}{2}\mathbb{P}(Y \notin \mathcal{E}') + \frac{1}{2}\mathbb{P}_Y(Y \notin T_D(\mathcal{E}')) + \frac{1}{2}\int_{\mathbb{R}^K} |f(y) - g_D(y)|\mathbf{1}\{y \in T_D(\mathcal{E}')\}dy \\ &\le o(1) + \frac{1}{2}\mathbb{E}_Y[|r_D(Y) - 1|\mathbf{1}\{Y \in T_D(\mathcal{E}')\}]\,. \end{aligned}$$

对 $(a+b)^2 \le 2(a^2+b^2)$ 两边平方，得到

$$\begin{aligned} d_{\mathrm{TV}}(Y, T_D(Y))^2 &\le o(1) + \frac{1}{2}(\mathbb{E}_Y[|r_D(Y) - 1|\mathbf{1}\{Y \in T_D(\mathcal{E}')\}])^2 \\ &= o(1) + \frac{1}{2}(\mathbb{E}_Y[|r_D(Y) - 1|\mathbf{1}\{Y \in T_D(\mathcal{E}')\}])^2\,. \end{aligned}$$

第二项还可以进一步取上界为

$$\begin{aligned} \frac{1}{2}(\mathbb{E}_Y[|r_D(Y) - 1|\mathbf{1}\{Y \in T_D(\mathcal{E}')\}])^2 &\overset{(a)}{\le} \frac{1}{2}\frac{(\mathbb{E}_Y[|r_D(Y) - 1|\mathbf{1}\{Y \in T_D(\mathcal{E}')\}])^2}{\mathbb{E}_Y[r_D(Y)/3 + 2/3]} \\ &\overset{(b)}{\le} \frac{1}{2}\mathbb{E}_Y\left[\frac{(r_D(Y)-1)^2}{r_D(Y)/3 + 2/3}\mathbf{1}\{Y \in T_D(\mathcal{E}')\}\right] \\ &\overset{(c)}{\le} \mathbb{E}_Y[(r_D(Y)\log r_D(Y) - r_D(Y) + 1)\mathbf{1}\{Y \in T_D(\mathcal{E}')\}]\,, \end{aligned} \tag{C.4}$$

其中，(a) 成立是因为 $r_D(Y) \ge 0$ 且 $\mathbb{E}_Y[r_D(Y)/3 + 2/3] \le 1$；$(b)$ 由柯西-施瓦茨不等式得出；对于所有 $x \ge 0$，(c) 根据 $(x-1)^2/(x/3+2/3) \le 2(x\log x - x + 1)$ 成立。

下一步是计算 (C.4) 中的最后一项，它是一个 KL 型期望。一旦可以应用分部积分公式（即散度定理），计算就相当简单。不过，这需要对（不光滑的）指示函数进行可微近似。该函数的形式化定义如下；证明推迟到后文。

引理 C.1. 存在一个连续可微函数 $\phi : \mathbb{R}^K \to [0,1]$，满足以下条件：

(a) $\phi(y) = 1$ *if* $\|\overline{y}\|_{\mathrm{op}} \le 0.1$.

(b) $\phi(y) = 0$ 如果 $\|y\|_{\mathrm{op}} \geq 0.15$。

(c) $\|\nabla\phi(y)\| = O(1)$ 对所有 $0.1 \leq \|y\|_{\mathrm{op}} \leq 0.15$ 均成立。

然后，根据引理 C.1，式（C.4）中的最后一项可由以下上界估计

$$\begin{aligned}
&\mathbb{E}_Y[(r_D(Y)\log r_D(Y) - r_D(Y) + 1)\mathbf{1}\{T_D^{-1}(Y) \in \mathcal{E}'\}] \\
&\leq \mathbb{E}_Y[(r_D(Y)\log r_D(Y) - r_D(Y) + 1)\phi(T_D^{-1}(Y))] \\
&\leq \mathbb{E}_Y[(r_D(Y)\log r_D(Y))\phi(T_D^{-1}(Y))] + o(1)\,,
\end{aligned} \tag{C.5}$$

其中，最后一个不等式由

$$\begin{aligned}
\mathbb{E}_Y[\phi(T_D^{-1}(Y))] - \mathbb{E}_Y[\phi(T_D^{-1}(Y))r_D(Y)] &= \mathbb{E}_Y[\phi(T_D^{-1}(Y))] - \mathbb{E}_Y[\phi(Y)] \\
&= \mathbb{E}_Y[1 - \phi(Y)] - \mathbb{E}_Y[1 - \phi(T_D^{-1}(Y))] \\
&\leq \mathbb{P}\left(\left\|\overline{Y}\right\|_{\mathrm{op}} \geq 0.1\right)\,,
\end{aligned}$$

其中，不等式由 $0 \leq \phi \leq 1$ 和 $\phi(Y) = 1$ 对 $\left\|Y\right\|_{\mathrm{op}} \leq 0.1$成立可得，最后一项为 $o(1)$，根据引理A.6。式（C.5）的最后一项可表示为

$$\begin{aligned}
&\mathbb{E}_Y[(r_D(Y)\log r_D(Y))\phi(T_D^{-1}(Y))] \\
&= \int_{y \in T_D(2\mathcal{E}')} g_D(y)\phi(T_D^{-1}(y))\log\frac{g_D(y)}{f(y)}dy \\
&= \int_{w \in 2\mathcal{E}'} f(w)\phi(w)\log\frac{f(w)(\det D)^{-(k-1)}}{f(T_D(w))}dw \\
&= \mathbb{E}_Y\left[\left(-\frac{d-k-1}{2}(\log\det(I_k + D\overline{Y}D) - \log\det(I_k + \overline{Y})) - (k-1)\log\det D\right)\phi(Y)\right] \\
&= \mathbb{E}_Y\left[\left(-\frac{d-k-1}{2}\log\det(I_k + A) - (k-1)\log\det D\right)\phi(Y)\right]\,,
\end{aligned} \tag{C.6}$$

其中，第二个等式由变量代换 $y = T_D(w)$ 及雅可比行列式 $dy = (\det D)^{k-1}dw$ 以及式（C.3）得到；第三个等式通过代入式（C.2）中的密度形式并使用 $T_D(w) = DwD$得到；最后一个等式由恒等式

$$\det(I_k + A) = \det\left((I_k + \overline{Y})^{-1/2}(I_k + D\overline{Y}D)(I_k + \overline{Y})^{-1/2}\right) = \frac{\det(I_k + D\overline{Y}D)}{\det(I_k + \overline{Y})}\,,$$

其中 $A := (I_k + \overline{Y})^{-1/2}(D\overline{Y}D - \overline{Y})(I_k + \overline{Y})^{-1/2}$。接下来，我们近似式（C.6）中的对数行列式项。我们断言

$$|\log\det(I_k + A) - \mathrm{Tr}(A)| \leq \mathrm{Tr}\left(A^2\right)\,. \tag{C.7}$$

通过泰勒展开，并令 $\{\mu_i\}_{i=1}^k$ 表示 A的特征值，我们得到

$$\log\det(I_k + A) = \sum_{i=1}^k \log(1 + \mu_i)\,.$$

对于 $Y \in 2\mathcal{E}'$，这些特征值很小，因为

$$\begin{aligned}\|A\|_{\mathrm{op}} &\le \left\|(I_k+\overline{Y})^{-1/2}\right\|_{\mathrm{op}}^2 \left\|D\overline{Y}D-\overline{Y}\right\|_{\mathrm{op}} \\ &\le \left\|(I_k+\overline{Y})^{-1/2}\right\|_{\mathrm{op}}^2 \left(\|D-I_k\|_{\mathrm{op}}^2 \left\|\overline{Y}\right\|_{\mathrm{op}} + 2\|D-I_k\|_{\mathrm{op}} \left\|\overline{Y}\right\|_{\mathrm{op}}\right) \\ &\le (1-0.2)^{-1}(0.1^2\times 0.2 + 2\times 0.1\times 0.2) \le 1/2\,,\end{aligned}$$

其中倒数第二个不等式由 $\left\|Y\right\|_{\mathrm{op}} \le 0.2$ 和 $\overline{D} \in \mathcal{E}$ 得出。然后，由于 $|\log(1+x)-x| \le x^2$ 对于 $|x| \le 1/2$ 成立，我们有

$$\left|\sum_{i=1}^k (\log(1+\mu_i)-\mu_i)\right| \le \sum_{i=1}^k \mu_i^2\,.$$

因此，(C.7) 得证。将其应用于 (C.6)，得到

$$\begin{aligned}&\mathbb{E}_Y\left[\left(-\frac{d-k-1}{2}\log\det(I_k+A)-(k-1)\log\det D\right)\phi(Y)\right] \\ &\le \underbrace{\mathbb{E}_Y\left[\left(-\frac{d-k-1}{2}\mathrm{Tr}(A)-(k-1)\log\det D\right)\phi(Y)\right]}_{\text{(I)}} + \underbrace{\mathbb{E}_Y\left[\frac{d-k-1}{2}\mathrm{Tr}(A^2)\phi(Y)\right]}_{\text{(II)}}\,. \qquad \text{(C.8)}\end{aligned}$$

对于 (C.8) 中的 (I)，第一项有

$$\begin{aligned}-\frac{d-k-1}{2}\mathbb{E}_Y[\mathrm{Tr}(A)\phi(Y)] &= -\frac{d-k-1}{2}\mathbb{E}_Y[\mathrm{Tr}((I_k+\overline{Y})^{-1}(D\overline{Y}D-\overline{Y})\phi(Y)] \\ &\overset{(a)}{=} -\int_{2\mathcal{E}'} f(y)\left\langle\frac{\partial\log f}{\partial y}, \phi(y)(T_D(y)-y)\right\rangle_{\mathbb{R}^K} dy \\ &= -\int_{2\mathcal{E}'}\left\langle\frac{\partial f}{\partial y}, \phi(y)(T_D(y)-y)\right\rangle_{\mathbb{R}^K} dy \\ &\overset{(b)}{=} \int_{2\mathcal{E}'} \mathrm{div}_y(\phi(y)(T_D(y)-y))f(y)dy \\ &\quad -\int_{\partial(2\mathcal{E}')} f(y)\left\langle\phi(y)(T_D(y)-y), \nu(y)\right\rangle_{\mathbb{R}^K} d\Sigma(y) \\ &\overset{(c)}{\le} \mathbb{E}_Y[\phi(Y)]\sum_{i<j\in[k]}(D_{ii}D_{jj}-1)+o(1)\,. \qquad \text{(C.9)}\end{aligned}$$

这里，(a) 由下式得出

$$\begin{aligned}\left\langle\frac{\partial\log f}{\partial y}(y), T_D(y)-y\right\rangle_{\mathbb{R}^K} &= \frac{d-k-1}{2}\left\langle\frac{\partial}{\partial y}\log\det(I_k+\overline{Y}), T_D(y)-y\right\rangle_{\mathbb{R}^K} \\ &= \frac{d-k-1}{2}\left.\frac{d}{dt}\right|_{t=0}\log\det\left(I_k+\overline{Y}+t(D\overline{Y}D-\overline{Y})\right) \\ &= \frac{d-k-1}{2}\mathrm{Tr}\left((I_k+\overline{Y})^{-1}(D\overline{Y}D-\overline{Y})\right)\,,\end{aligned}$$

(b) 来自向量场 $v(y)$ 的散度定理：$= \phi(y)(T_D(y)-y)$，该向量场在 $\partial(2\mathcal{E}')$ 上为零，其中：$\mathrm{div}_y v(y) := \sum_{\ell=1}^K \frac{\partial v_\ell}{\partial y_\ell}(y)$，$\nu(y)$ 是 $\partial(2\mathcal{E}')$ 上的外单位法向量，而 $d\Sigma(y)$ 是曲面测度，即，

$$-\int_{2\mathcal{E}'}\left\langle\frac{\partial f}{\partial y}(y), v(y)\right\rangle_{\mathbb{R}^K} dy = \int_{2\mathcal{E}'}\mathrm{div}_y v(y)\, f(y)\, dy - \int_{\partial(2\mathcal{E}')} f(y)\left\langle v(y), \nu(y)\right\rangle_{\mathbb{R}^K} d\Sigma(y)\,,$$

而（c）由下式得出

$$\begin{aligned}\operatorname{div}_y(\phi(y)(T_D(y)-y)) &= \sum_{i<j\in[k]} \frac{\partial}{\partial y_{ij}}\left[\phi(y)(T_D(y)-y)_{ij}\right] \\ &= \sum_{i<j\in[k]} \phi(y)(D_{ii}D_{jj}-1) + \langle \nabla\phi(y), T_D(y)-y\rangle_{\mathbb{R}^K}\,,\end{aligned}$$

其中 $|\mathbb{E}_Y[\langle \nabla\phi(Y)\ T_D(Y)\ -Y\rangle_{\mathbb{R}^K}]| \le \sqrt{\mathbb{E}_Y[\|\nabla\phi(Y)\|^2]\mathbb{E}_Y[\|T_D(Y)\ -Y\|^2]} \le O(k/\sqrt{d})\sqrt{\mathbb{P}\left(\|Y\|_{\mathrm{op}} \ge 0.1\right)} =$ $o(1)$ 由引理 A.6。将（C.9）与 $(k-1)\log\det D = \sum_{i<j\in[k]}\log(D_{ii}D_{jj})$ 结合，得

$$\begin{aligned}\text{(I)} &\le \mathbb{E}[\phi(Y)] \sum_{i<j\in[k]} (D_{ii}D_{jj} - 1 - \log(D_{ii}D_{jj})) + o(1) \\ &\le \sum_{i<j\in[k]} (D_{ii}D_{jj}-1)^2 + o(1)\,,\end{aligned}$$

其中第二个不等式由 $D_{ii}D_{jj} \in [0.9^2, 1.1^2]$ 得出：对于所有 $i<j\in[k]$，当 $D\in\mathcal{E}$ 且 $0 \le x-1-\log x \le (x-1)^2$ 时；对于所有 $x \ge 1/2$。

对于（C.8）中的（II），我们有

$$\begin{aligned}&\frac{d-k-1}{2}\mathbb{E}_Y[\operatorname{Tr}(A^2)\phi(Y)] \\ &= \frac{d-k-1}{2}\mathbb{E}_Y[\operatorname{Tr}((I_k+\overline{Y})^{-1}(D\overline{Y}D-\overline{Y})(I_k+\overline{Y})^{-1}(D\overline{Y}D-\overline{Y}))\phi(Y)] \\ &\le \frac{d-k-1}{2}\mathbb{E}_Y\left[\left\|(I_k+\overline{Y})^{-1}(D\overline{Y}D-\overline{Y})\right\|_{\mathrm{F}}^2 \phi(Y)\right] \\ &\le \frac{d-k-1}{2}\mathbb{E}_Y\left[\left\|(I_k+\overline{Y})^{-1}\right\|_{\mathrm{op}}^2 \left\|D\overline{Y}D-\overline{Y}\right\|_{\mathrm{F}}^2 \phi(Y)\right] \\ &\le 2\sum_{i<j\in[k]} (D_{ii}D_{jj}-1)^2\,,\end{aligned}$$

其中，最后一个不等式由 $\left\|(I_k+Y)^{-1}\right\|_{\mathrm{op}} \le (1-0.15)^{-1}$ 对 $Y\in\operatorname{supp}(\phi)$ 以及 $\mathbb{E}_Y[Y_{ij}^2] = 1/d$ 得出。综合（I）和（II）的计算，对于（C.8），我们有

$$\text{(I)} + \text{(II)} \le 3\sum_{i<j\in[k]} (D_{ii}D_{jj}-1)^2 + o(1)\,. \tag{C.10}$$

现在从（C.10）回溯到（C.1），我们得到

$$\begin{aligned}d_{\mathrm{TV}}(Y,W) &\le o(1) + \sqrt{o(1) + 3\mathbb{E}_D\left[\sum_{i<j\in[k]} (D_{ii}D_{jj}-1)^2 \mathbf{1}\{D\in\mathcal{E}\}\right]} \\ &\le o(1) + \sqrt{o(1) + 3k^2\mathbb{E}_D\left[(D_{11}D_{22}-1)^2\right]} \\ &= o(1) + \sqrt{o(1) + 6k^2\left(1 - \left(\mathbb{E}\left[\frac{\|Z_1\|}{\sqrt{d}}\right]\right)^2\right)} \\ &\le o(1) + \sqrt{o(1) + O(k^2/d)}\,,\end{aligned}$$

其中等号由 D_{11} 、 D_{22} 和$\mathbb{E}[D_{ii}^2] = 1$的独立性得出，这意味着$\mathbb{E}[(D_{11}D_{22} - 1)^2] = \mathbb{E}[D_{11}^2]\mathbb{E}[D_{22}^2] - 2(\mathbb{E}[D_{11}])^2 + 1 = 2(1 - (\mathbb{E}[D_{11}])^2)$；最后一个不等式来自$\mathbb{E}\left[\|Z_1\|/\sqrt{d}\right] = \sqrt{2}\Gamma((d+1)/2)/(\Gamma(d/2)\sqrt{d}) = 1 - 1/(4d) + O(1/d^2)$ [AS65，公式 6.1。47]。最后一项为 $o(1)$，如所需。

我们通过验证先前用于近似指示函数的技术引理（引理 C.1），完成命题 5.1的证明。

引理 *C.1*的证明。 首先，很容易看出，存在一个具有这些性质的一元函数。具体而言，存在一个连续可微函数 $g : \mathbb{R} \to [0,1]$，使得 $g(x) = 1$当 $|x| \le 0.11$时， $g(x) = 0$ 当 $|x| \ge 0.15$时，并且 $|g'(x)| \le C_0$ 等于某个常数 $C_0 > 0$。

因此，就我们的目的而言，只需找到算子范数$\|y\|_{\mathrm{op}}$ 的一个近似 $h(y)$，并取这两个函数的复合，即

$$\phi := g \circ h\,.$$

我们使用标准的对数和指数函数，它是光滑的：

$$h(y) := \frac{1}{\beta}\log(\mathrm{Tr}(\exp(\beta\overline{y})) + \mathrm{Tr}(\exp(-\beta\overline{y})))\,,$$

其中 $\beta := 100\log(2k)$。作为特征值的函数，可以验证 $\|y\|_{\mathrm{op}} \le h(y) \le \|y\|_{\mathrm{op}} + 0.01$。此外，根据

$$\frac{\partial}{\partial y_{ij}}h(y) = 2\left(\frac{\exp(\beta\overline{y}) - \exp(-\beta\overline{y})}{\mathrm{Tr}(\exp(\beta\overline{y})) + \mathrm{Tr}(\exp(-\beta\overline{y}))}\right)_{ij}, i < j \in [k]\,,$$

我们有

$$\|\nabla h(y)\| \le 2\left\|\frac{\exp(\beta\overline{y})}{\mathrm{Tr}(\exp(\beta\overline{y})) + \mathrm{Tr}(\exp(-\beta\overline{y}))}\right\|_F + 2\left\|\frac{\exp(-\beta\overline{y})}{\mathrm{Tr}(\exp(\beta\overline{y})) + \mathrm{Tr}(\exp(-\beta\overline{y}))}\right\|_F \le 4\,,$$

因为每个矩阵 $\exp(\pm\beta y)/(\text{迹}(\exp(\beta y)) + \text{迹}(\exp(-\beta y)))$ 都是半正定的，且迹至多为 1。要验证引理 C.1中的性质 (a)–(c)：根据定义，(a) 和 (b) 显然成立，而 (c) 可由 $\|\nabla\phi(y)\| = \|\nabla h(y)\|\,|g'(h(y))| \le 4C_0$ 得出。

□

C.2 似然比的分解（引理 5.3）

该陈述分为两部分；第一部分是确定 $1 \le i \le 3$ 对应的函数 $f_i(X)$，第二部分是证明它们关于 N 和 d 的矩界。

刻画 $f_i(X)$。 根据密度公式$w_{N,\,d}$ 和$m_{N,\,d}$ （例如参见 [RR19，第2]节），我们有

$$w_{N,d}^2(X)/m_{N,d}^2(X) = \exp(2\alpha_{N,d}(X))\,,$$

其中， $\lambda_1, \ldots, \lambda_N$是 X的特征值，

$$\begin{aligned}\alpha_{N,d}(X) := \frac{1}{2}\sum_{i=1}^{N}\left((d-N-1)\log\lambda_i - \lambda_i + \frac{1}{2d}(\lambda_i - d)^2\right)\\ + \left(\frac{N(N+3)}{4} - \frac{dN}{2}\right)\log 2 + \frac{N}{2}\log\pi + \frac{N(N+1)}{4}\log d - \sum_{i=1}^{N}\log\Gamma\left(\frac{d+1-i}{2}\right)\,.\end{aligned}$$

根据斯特林公式，$\log\,\Gamma(z)\,=(z-1/2)\log\,z-z+(1/2)\log(2\pi)\,+O(1/z)$，最后一项可近似为

$$\begin{aligned}\alpha_{N,d}(X)=\frac{1}{2}\sum_{i=1}^{N}\left((d-N-1)\log\lambda_i-\lambda_i+\frac{1}{2d}(\lambda_i-d)^2\right)\\+\frac{N(N+1)}{4}\log d-\frac{1}{2}\sum_{i=1}^{N}(d-i)\log(d+1-i)+\frac{1}{2}\sum_{i=1}^{N}(d+1-i)+O\left(\frac{N}{d}\right).\end{aligned}\tag{C.11}$$

由于 $N\le s\le\lceil 1.1k\rceil$，$O(N/d)$ 项的阶至多为 $O(k/d)=o(1)$。由

$$\log(d+1-i)=\log d+\log\left(1-\frac{i-1}{d}\right)=\log d-\frac{i-1}{d}-\frac{(i-1)^2}{2d^2}+O\left(\frac{(i-1)^3}{d^3}\right)$$

以及 $0\le\sum_{i=1}^{N}(d-i)(i-1)^3/d^3\le\sum_{i=1}^{N}(i-1)^3/d^2=O(k^4/d^2)\,=o(1)$，展开式（C.11）可简洁地写为

$$\alpha_{N,d}(X)=\sum_{i=1}^{N}t(\lambda_i)-\frac{N^3}{12d}+o(1)\,,\tag{C.12}$$

其中

$$t(x):=\frac{1}{2}\left((d-N-1)\log(x/d)-(x-d)+\frac{1}{2d}(x-d)^2\right).$$

现在考虑 t 在 $x=d$ 处的三阶泰勒展开。于是

$$\begin{aligned}t(x)&=-\frac{N+1}{2d}(x-d)+\frac{N+1}{4d^2}(x-d)^2+\frac{d-N-1}{6d^3}(x-d)^3-\frac{d-N-1}{8\xi^4}(x-d)^4\\&\le-\frac{N+1}{2d}(x-d)+\frac{N+1}{4d^2}(x-d)^2+\frac{d-N-1}{6d^3}(x-d)^3\,,\end{aligned}$$

其中 ξ 介于 x 和 d 之间，且 $d\gg k^2$ 蕴含 $d>N+1$。

从现在起，考虑 X 的仿射变换 Y 更为方便，其定义如下：

$$Y:=(X-dI_N)/\sqrt{d}\,.$$

那么，$X\sim m_{N,\,d}$，$Y_{ii}\sim\mathcal{N}(0,\ 2)$ $i\in[N]$和 $Y_{ij}=Y_{ji}\sim\mathcal{N}(0,\ 1)$ $i<j\in[N]$彼此独立。此外，$X\in\mathcal{E}_N\Leftrightarrow Y\in\mathcal{F}_N$，其中

$$\mathcal{F}_N:=\left\{Y\in\mathbb{R}^{N\times N}:\|Y_{Q\times Q}\|_{\mathrm{op}}\le 10\sqrt{|Q|}(1+\sqrt{\log s})\text{ for all }\emptyset\subsetneq Q\subseteq[N]\right\}.$$

将 t的上界代入（C.12）可得

$$\begin{aligned}\alpha_{N,d}(X)&\le-\frac{N+1}{2d^{1/2}}\sum_{i=1}^{N}\mu_i+\frac{N+1}{4d}\sum_{i=1}^{N}\mu_i^2+\frac{d-N-1}{6d^{3/2}}\sum_{i=1}^{N}\mu_i^3+o(1)\,,\\&=-\frac{N+1}{2d^{1/2}}\mathrm{Tr}(Y)+\frac{N+1}{4d}\mathrm{Tr}(Y^2)+\frac{d-N-1}{6d^{3/2}}\mathrm{Tr}(Y^3)+o(1)\,,\end{aligned}\tag{C.13}$$

其中，μ_1，……，μ_N是 Y的特征值。由（C.13）和（5.10）可得

$$\begin{aligned}&\exp(2\alpha_{N,d}(X))\mathbf{1}\{X\in\mathcal{E}_N\}\\&=\exp(2\alpha_{N,d}(X))\mathbf{1}\{Y\in\mathcal{F}_N\}\\&\le(1+o(1))\underbrace{\exp\left(-\frac{N+1}{d^{1/2}}\mathrm{Tr}(Y)\right)}_{=:f_1(X)}\underbrace{\exp\left(\frac{N+1}{2d}\mathrm{Tr}(Y^2)\right)}_{=:f_2(X)}\underbrace{\exp\left(\frac{d-N-1}{3d^{3/2}}\mathrm{Tr}(Y^3)\right)}_{=:f_3(X)}\mathbf{1}\{Y\in\mathcal{F}_N\}\,.\end{aligned}$$

在这里，我们将 $f_i(X)$ 识别为 $1\le i\le 3$ 的 Y 的幂的迹的指数化形式。

计算期望。根据其定义，我们证明 $f_1(X)$ 通过 $f_3(X)$ 的期望（对 $X \sim m_{N,\,d}$）至多为 $\exp(O(N^3/d))$。这些计算的关键在于高斯正交系综的迹的集中性（引理A.2）。

- 对于 $f_1(X)$，我们有 $\mathrm{Tr}(Y) \sim \mathcal{N}(0,\ 2N)$。因此，根据正态分布的矩母函数，

$$\mathbb{E}_{X\sim m_{N,d}}[f_1(X)^3] = \exp\left(\frac{9(N+1)^2}{2d} \times 2N\right) = \exp\left(O\left(\frac{N^3}{d}\right)\right).$$

- 对于 $f_2(X)$，$\mathrm{Tr}(Y^2$以 $N(N+1$为中心集中，且具有次指数尾（范数为 $\|\cdot\|_{\psi_1} = O(N)$），由引理 A.2可知。根据次指数集中性到矩母函数的标准转换 [Ver18，命题 2.7。1]，

$$\mathbb{E}_{X\sim m_{N,d}}[f_2(X)^3] \le \exp\left(\frac{3N(N+1)^2}{2d} + \left(\frac{3(N+1)}{2d}\right)^2 \times O(N^2)\right) = \exp\left(O\left(\frac{N^3}{d}\right)\right).$$

注意，这里我们使用了 $N^4/d^2 = O(k^4/d^2) = o(1)$ 和 $N^4/d^2 \le (N^3/d) \times O(k/d) = o(N^3/d)$。

- 对于 $f_3(X)$，首先由 $\exp(x) \le 1 + x + (x^2/2)\exp(|x|)$ 可得

$$\begin{aligned}
&\mathbb{E}_{X\sim m_{N,d}}\left[f_3(X)^3\right] \\
&= \mathbb{E}_{X\sim m_{N,d}}\left[\exp\left(\frac{d-N-1}{d^{3/2}}\mathrm{Tr}(Y^3)\right)\mathbf{1}\{Y\in\mathcal{F}_N\}\right] \\
&\le 1 + \frac{1}{2}\mathbb{E}_{X\sim m_{N,d}}\left[\left(\frac{(d-N-1)\mathrm{Tr}(Y^3)}{d^{3/2}}\right)^2 \exp\left(\frac{(d-N-1)|\mathrm{Tr}(Y^3)|}{d^{3/2}}\right)\mathbf{1}\{Y\in\mathcal{F}_N\}\right] \\
&\le 1 + \frac{1}{2d}\sqrt{\mathbb{E}_{X\sim m_{N,d}}\left[\mathrm{Tr}(Y^3)^4\right]}\sqrt{\mathbb{E}_{X\sim m_{N,d}}\left[\exp\left(\frac{2|\mathrm{Tr}(Y^3)|}{d^{1/2}}\right)\mathbf{1}\{Y\in\mathcal{F}_N\}\right]} \\
&\le 1 + O\left(\frac{N^3}{d}\right) \times \sqrt{\mathbb{E}_{X\sim m_{N,d}}\left[\exp\left(\frac{2|\mathrm{Tr}(Y^3)|}{d^{1/2}}\right)\mathbf{1}\{Y\in\mathcal{F}_N\}\right]}.
\end{aligned} \tag{C.14}$$

这里，对于第二个不等式，我们用到 $\mathrm{Tr}(Y^3)\mathbf{1}\{Y\in\mathcal{F}_N\}$ 的均值为 0，因为每个 Y_{ij} 都关于 0 对称（因此经过对称截断 $Y\in\mathcal{F}_N$后也如此）；第三个不等式由柯西-施瓦茨不等式得到，其中 $0 \le (d-N-1)/d^{3/2} \le 1/d^{1/2}$；最后一个不等式则由高斯超收缩性得到 $\sqrt{\mathbb{E}[(\mathrm{Tr}(Y^3))^4]} \lesssim \mathbb{E}[(\mathrm{Tr}(Y^3))^2] = O(N^3)$。

注意，对于 $Y\in\mathcal{F}_N$，我们有

$$|\mathrm{Tr}(Y^3)| \le N\|Y\|_{\mathrm{op}}^3 \le N \times (10\sqrt{N}(1+\sqrt{\log s}))^3 \le C_0 N^{5/2}\log^{3/2} s\,,$$

其中第一个不等式由 $|\sum_{i=1}^N \mu_i^3| \le N\max_{1\le i\le N}|\mu_i|^3$ 得出，而最后一个不等式

对某个常数 $C_0 > 0$ 成立。因此，

$$\begin{aligned}
&\mathbb{E}_{X\sim m_{N,d}}\left[\exp((2/d^{1/2})|\mathrm{Tr}(Y^3)|)\mathbf{1}\{Y\in\mathcal{F}\}\right] \\
&\le \int_0^{\exp(2C_0N^{5/2}\log^{3/2}s/d^{1/2})} \mathbb{P}_{X\sim m_{N,d}}(\exp((2/d^{1/2})|\mathrm{Tr}(Y^3)|)\ge u)du \\
&\le 1+2\sqrt{\frac{N^3}{d}}\int_0^{C_0N\log^{3/2}s} \mathbb{P}_{X\sim m_{N,d}}(|\mathrm{Tr}(Y^3)|\ge N^{3/2}u)\exp\left(2u\sqrt{\frac{N^3}{d}}\right)du \\
&\le 1+4\sqrt{\frac{N^3}{d}}\int_0^{N^{3/4}} \exp\left(-\frac{1}{C_{A.2}}u^2+2u\sqrt{\frac{N^3}{d}}\right)du \\
&\quad +4\sqrt{\frac{N^3}{d}}\int_{N^{3/4}}^{C_0N\log^{3/2}s} \exp\left(-\frac{Nu^{2/3}}{C_{A.2}}+2u\sqrt{\frac{N^3}{d}}\right)du \\
&\le 1+O\left(\sqrt{\frac{N^3}{d}}\right)\exp\left(O\left(\frac{N^3}{d}\right)\right)+4\sqrt{\frac{N^3}{d}}\int_{N^{3/4}}^{C_0N\log^{3/2}s}\exp\left(-\frac{Nu^{2/3}}{C_{A.2}}+2u\sqrt{\frac{N^3}{d}}\right)du\,,
\end{aligned} \tag{C.15}$$

其中第二个不等式由变量代换 $v=\exp((2N^{3/2}/d^{1\ 2})u$得到，而第三个不等式由推论 A.2 得出（在 $u=N^{3/4}$ 处拆分积分，此处尾部行为发生变化）。这里请注意，函数 $u\mapsto\exp(-Nu^{2/3}/C_{A.2}+2u\sqrt{N^3/d})$ 在 $u\ge 0$ 时，随 $u\le(\sqrt{d^3/N^3})/(27C_{A.2}^3)$ 单调递减，此后单调递增。因此，该函数在 $u\in[N^{3/4},C_0N\log^{3/2}s]$ 上的最大值在边界处取得，其值为

$$\begin{aligned}
&\exp(-N^{3/2}/C_{A.2}+2N^{9/4}/\sqrt{d})\vee\exp(-(C_0^{2/3}/C_{A.2})N^{5/3}\log s+2C_0N^{5/2}\log^{3/2}s/\sqrt{d}) \\
&=\exp\left(-\frac{N^{3/2}}{C_{A.2}}\left(1-O\left(\frac{N^{3/4}}{\sqrt{d}}\right)\right)\right)\vee\exp\left(-\left(\frac{C_0^{2/3}}{C_{A.2}}\right)N^{5/3}\log s\left(1-O\left(\frac{N^{5/6}\log^{1/2}s}{\sqrt{d}}\right)\right)\right) \\
&\le\exp\left(-\frac{N^{3/2}}{2C_{A.2}}\right)\vee\exp\left(-\frac{C_0^{2/3}}{2C_{A.2}}N^{5/3}\log s\right)\,,
\end{aligned}$$

由于 $N=O(k)$ 蕴含 $N^{3/4}/\sqrt{d}=o(1)$ 和 $N^{5/6}\log^{1/2}s/\sqrt{d}=o(1)$ 对于 $d\gg k^2$ 成立，因此当 k 充分大时，该不等式对所有情况都成立。于是可得

$$\begin{aligned}
&\int_{N^{3/4}}^{C_0N\log^{3/2}s}\exp\left(-Nu^{2/3}+2u\sqrt{\frac{N^3}{d}}\right)du \\
&\le C_0N\log^{3/2}s\left(\exp\left(-\frac{N^{3/2}}{2C_{A.2}}\right)\vee\exp\left(-\frac{C_0^{2/3}}{2C_{A.2}}N^{5/3}\log s\right)\right)=O\left(\frac{\log^{3/2}s}{N}\right)\,.
\end{aligned} \tag{C.16}$$

由 (C.14)、(C.15) 和 (C.16)，我们得到

$$\begin{aligned}
\mathbb{E}_{X\sim m_{N,d}}[f_3(X)^3] &\le 1+O\left(\frac{N^3}{d}\right)\left(1+O\left(\sqrt{\frac{N^3}{d}}\right)\exp\left(O\left(\frac{N^3}{d}\right)\right)+O\left(\sqrt{\frac{N}{d}}\log^{3/2}k\right)\right) \\
&\le 1+O\left(\frac{N^3}{d}\right)\left(1+o(1)+O\left(\sqrt{\frac{N^3}{d}}\right)\exp\left(O\left(\frac{N^3}{d}\right)\right)\right) \\
&\le \exp\left(O\left(\frac{N^3}{d}\right)\right)\,,
\end{aligned}$$

其中，第二个不等式由 $O(\sqrt{N/d}\log^{3/2} k) = o(1)$ 得出，前提是 $N = O(k)$ 且 $d \gg k^2$；最后一个不等式则由 $1+x(2+\sqrt{x}e^x) \le 1+2x+xe^{2x} \le 1+2x+e^{2x}(e^x-1) \le e^{3x}$ 对所有 $x \ge 0$ 成立得出。

C.3 三次超几何的上界（引理 5.4）

令 $C_0 > 0$ 为任意值，并考虑任意固定的 $s \in [k^-, k^+]$。那么，给定 $|S| = |S'| = s$ 时，$N = |S \cap S'|$ 的条件分布为超几何分布$(n,\ s,\ s)$，它随机占优于 $M \sim$ 二项分布$(s,\ s/(n-s))$。于是

$$\begin{aligned}
\mathbb{E}_{N\left||S|=|S'|=s\right.}[\exp(C_0 N^3/d)] &\le \mathbb{E}[\exp(C_0 M^3/d)] \\
&\le \exp\left(64C_0\frac{s^6}{n^3 d}\right) + \int_{64s^6/n^3}^{s^3} \mathbb{P}(M \ge u^{1/3})\frac{C_0}{d}\exp\left(\frac{C_0 u}{d}\right) du \\
&\le \exp\left(64C_0\frac{s^6}{n^3 d}\right) + \int_{64s^6/n^3}^{s^3} \mathbb{P}(M - \mathbb{E}[M] \ge (1/2)u^{1/3})\frac{C_0}{d}\exp\left(\frac{C_0 u}{d}\right) du \\
&\le \exp\left(64C_0\frac{s^6}{n^3 d}\right) + \frac{C_0}{d}\int_{64s^6/n^3}^{s^3} \exp(-u^{1/3}/100 + C_0 u/d) du\,,
\end{aligned}$$

其中第二和第三个不等式分别由 $\mathbb{E}[M] = s^2/(n-s) \le 2s^2/n \le (1/2)u^{1/3}$（对 $u \in [64s^6/n^3, s^3]$）以及切尔诺夫界得到。对于所有这样的 u，由于 $C_0 u/d \le C_0 u^{1/3}(k^+)^2/d$，其中 $(k^+)^2/d = o(1)$，存在一个仅依赖于 C_0 的常数 $C_1 > 0$，使得对于所有充分大的 k，

$$\exp(-u^{1/3}/100 + C_0 u/d) \le \exp(-u^{1/3}/C_1)\,.$$

因此，我们有

$$\begin{aligned}
&\mathbb{E}_{N\left||S|=|S'|=s\right.}[\exp(C_0 N^3/d)] \\
&\le \exp\left(64C_0\frac{s^6}{n^3 d}\right) + \frac{C_0}{d}\int_{64s^6/n^3}^{s^3} \exp(-u^{1/3}/C_1) du \\
&\le \exp\left(64C_0\frac{s^6}{n^3 d}\right) + \frac{3C_0}{d}(C_1(64s^6/n^3)^{2/3} + 2C_1^2(64s^6/n^3)^{1/3} + 2C_1^3)\exp(-4s^2/(nC_1)) \\
&\le \exp\left(64C_0\frac{(k^+)^6}{n^3 d}\right) + 3C_0\left(C_1\left(\frac{64(k^+)^6}{n^3 d}\right)^{2/3} + 2C_1^2\left(\frac{64(k^+)^6}{n^3 d}\right)^{1/3} + \frac{2C_1^3}{d}\right) \\
&= 1 + o(1)\,,
\end{aligned}$$

可按需取值。这里，第二个不等式来自 $\int_b^\infty \exp(-u^{1/3}/a) du = 3(ab^{2/3} + 2a^2 b^{1/3} + 2a^3)\exp(-b^{1/3}/a)$，第三个不等式来自 $(1/d)x^{2/3} \le (x/d)^{2/3}$，$(1/d)x^{1/3} \le (x/d)^{1/3}$，$\exp(-4s^2/(nC_1)) \le 1$其中 $s \le k^+ = \lceil 1.1k \rceil$。最后一行由假设 $d \gg k^2 \vee k^6/n^3$ 得出。

C.4 邻域分布的典型行为（引理5.5）

回忆一下，$U=(U_1,\ldots,U_{n-1})$ $V=(V_1,\ldots,V_{n-1})$。概率 $\mathbb{P}((\Gamma,U,V)\in\mathcal{E}^c)$（无论是 $\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})$ 还是 $\mathcal{Q}_n$）的上界为

$$\underbrace{\mathbb{P}_{X_{[n-1]},\Gamma}\left(\sum_{i=1}^{n-1}V_i>2k\right)}_{\text{(I)}}+\underbrace{\mathbb{P}_{X_{[n-1]},\Gamma}\left(\sum_{i\in[n-1]:V_i=1}\Gamma_i>4kp,\sum_{i=1}^{n-1}V_i\le 2k\right)}_{\text{(II)}}$$
$$+\underbrace{\mathbb{P}_{X_{[n-1]},\Gamma}\left(|\Delta(\Gamma,U,V)|>\frac{k}{n},\sum_{i=1}^{n-1}V_i\le 2k,\sum_{i\in[n-1]:V_i=1}\Gamma_i\le 4kp\right)}_{\text{(III)}}.$$

我们证明每一项都是 $o(1/(n^2\log(1/p)))$；对前两项 (I) 和 (II) 的以下分析适用于 $\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})$ 和 $\Gamma\sim\mathcal{Q}_n$。

由于 $\sum_{i=1}^{n-1}V_i\sim\mathrm{Binom}(n-1,k/n)$，根据标准切尔诺夫界，(I) 至多为 $\exp(-k/3)$。只要 $k\gtrsim\log n$，这就是 $o(1/(n^2\log(1/p)))$。

对于 (II)，首先注意，对任意 V,Γ，都有 $\mathrm{Ber}(p)^{\otimes(n-1)}$。因此 $\sum_{i\in[n-1]:\ V_i=1}\Gamma_i\sim\mathrm{Binom}(\sum_{i=1}^{n-1}V_i,p)$ 在给定 V 的条件下，其分布在事件 $\{\sum_{i=1}^{n-1}V_i\le 2k\}$ 下随机占优于 $\mathrm{Binom}(2k,p)$。因此，再次由切尔诺夫界可知，(II) 至多为 $\exp(-2kp/3)$；只要 $kp\gtrsim\log n$，这就是 $o(1/(n^2\log(1/p)))$。

对于 (III)，首先考虑 $\Gamma\sim\mathcal{Q}_n$的情况。令 ν 表示 $i\in[n-1]$ 中满足 $V_i=1$的数量（即包含在社区中的顶点），并令 $i_1<\cdots<i_\nu$表示这些指标。由引理 A.4可知，在给定 Γ和 V 的条件下，概率至多为

$$\exp\left(-\frac{1}{C_{A.4}}\frac{d}{M\log(1/p)\log(d/p)}+C_{A.4}\log(3k)\right),$$

其中 $M=M(\nu,p,(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}))=((\sum_{j=1}^{\nu}\Gamma_{i_j})\log(1/p)+\nu p+\log(d/p))(\sum_{j=1}^{\nu}\Gamma_{i_j}+\nu p)\le C_0k^2p^2\log(1/p)\log(d/p)$，其中某个常数 $C_0>0$ 取自 $\nu\le 2k$，$\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp$ 和 $kp\ge 1$。因此，(III) 至多为

$$\exp\left(-\frac{1}{C_{A.4}}\frac{d}{C_0(kp\log(1/p)\log(d/p))^2}+C_{A.4}\log(3k)\right),\tag{C.17}$$

只要 $d\gtrsim(kp\log(1/p)\log(d/p))^2\log n$，该式就是 $o(1/(n^2\log(1/p)))$。

现在我们考虑带有 $\Gamma\sim\mathcal{P}_n(\cdot|X_{[n-1]})$ 的 (III)。首先，在给定 V（带有 $\nu\le 2k$）的条件下，(III) 等于

$$
\begin{aligned}
&\mathbb{P}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\widetilde{\mathcal{P}}_{\nu+1}(\cdot|U_{i_1},\ldots,U_{i_\nu})}\left(|\Delta(\Gamma,U,V)|>\frac{k}{n},\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right)\\
&=\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}}\left[\mathbf{1}\left\{|\Delta(\Gamma,U,V)|>\frac{k}{n},\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{\nu-\sum_{j=1}^{\nu}\Gamma_{i_j}}}\right]\\
&\le\left(\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}}\left[\mathbf{1}\left\{|\Delta(\Gamma,U,V)|>\frac{k}{n},\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}\right]\right)^{1/2}\\
&\quad\times\left(\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}}\left[\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{2\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{2(\nu-\sum_{j=1}^{\nu}\Gamma_{i_j})}}\mathbf{1}\left\{\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}\right]\right)^{1/2}\\
&\le\exp\left(-\frac{1}{2C_{A.4}}\frac{d}{C_0(kp\log(1/p)\log(d/p))^2}+\frac{C_{A.4}}{2}\log(3k)\right)\\
&\quad\times\left(\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}}\left[\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{2\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{2(\nu-\sum_{j=1}^{\nu}\Gamma_{i_j})}}\mathbf{1}\left\{\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}\right]\right)^{1/2},
\end{aligned}
\tag{C.18}
$$

其中，第一个不等式来自柯西-施瓦茨不等式，第二个不等式则来自（C.17），其中$kp\ge 1$。

由于记号 i_1、……、i_ν看起来较为复杂，不失一般性地，我们将其改为更简单的指标。注意

$$
\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{2\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{2(\nu-\sum_{j=1}^{\nu}\Gamma_{i_j})}}\mathbf{1}\left\{\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}
$$

在 U_{i_1}、……、U_{i_ν}、$(\Gamma_{i_1}$、……、$\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}$（其中 $i_1<\cdots<i_\nu$ 是 $[n-1]$ 中满足 $V_i=1$ 的指标）上的分布与

$$
\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{[\nu]}|U_{[\nu]})}{\mathcal{Q}^2_{\nu+1}(\Gamma_{[\nu]})}\mathbf{1}\left\{\sum_{i=1}^{\nu}\Gamma_i\le 4kp\right\}
$$

在 $U_{[\nu]}$、$\Gamma_{[\nu]}\sim\mathcal{Q}_{\nu+1}$（其中 $[\nu]$ 是 $[n-1]$ 中满足 $V_i=1$ 的指标）由于对称性；回顾 $\Gamma_{[\nu]}=(\Gamma_1$、……、$\Gamma_\nu)$ 和 $U_{[\nu]}=(U_1$、……、$U_\nu)$。因此，对于（C.18）最后一项中的第二个因子，连同对 V 的相应变更，我们有

$$
\begin{aligned}
&\mathbb{E}_{U_{i_1},\ldots,U_{i_\nu},(\Gamma_{i_1},\ldots,\Gamma_{i_\nu})\sim\mathcal{Q}_{\nu+1}}\left[\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{i_1},\ldots,\Gamma_{i_\nu}|U_{i_1},\ldots,U_{i_\nu})}{p^{2\sum_{j=1}^{\nu}\Gamma_{i_j}}(1-p)^{2(\nu-\sum_{j=1}^{\nu}\Gamma_{i_j})}}\mathbf{1}\left\{\sum_{j=1}^{\nu}\Gamma_{i_j}\le 4kp\right\}\right]\\
&=\mathbb{E}_{U_{[\nu]},\Gamma_{[\nu]}\sim\mathcal{Q}_{\nu+1}}\left[\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{[\nu]}|U_{[\nu]})}{\mathcal{Q}^2_{\nu+1}(\Gamma_{[\nu]})}\mathbf{1}\left\{\sum_{i=1}^{\nu}\Gamma_i\le 4kp\right\}\right].
\end{aligned}
$$

现在，主要的技术部分在于控制这一项

$$
\mathbb{E}_{U_{[\nu]},\Gamma_{[\nu]}\sim\mathcal{Q}_{\nu+1}}\left[\frac{\widetilde{\mathcal{P}}^2_{\nu+1}(\Gamma_{[\nu]}|U_{[\nu]})}{\mathcal{Q}^2_{\nu+1}(\Gamma_{[\nu]})}\mathbf{1}\left\{\sum_{i=1}^{\nu}\Gamma_i\le 4kp\right\}\right].
\tag{C.19}
$$

在证明的其余部分，我们将证明这一项为 $O(1)$。若此结论成立，那么 (C.18) 中的最后一项确实为 $o(1/(n^2\log(1/p)))$，其中 $d \gtrsim (kp\log(1/p)\log(d/p))^2 \log n$。

我们指出，直接应用引理 A.4 并不足以在我们所需的范围内控制 (C.19)。特别是，似然比的平方 $\widetilde{\mathcal{P}}^2_{\nu+1}/\mathcal{Q}^2_{\nu+1}$ 可能高达 $\exp(\widetilde{\Omega}(kp))$，而要用引理 A.4 和 (C.18) 中的界 $\exp(-d/\widetilde{\Omega}(k^2p^2))$ 控制它，需要满足 $d = \widetilde{\Omega}(k^3p^3)$。我们的策略是直接界定似然比的二阶矩，而不是通过集中性来控制 (C.19)。为此，我们利用 $\mathcal{P}_{\nu+1}/\mathcal{Q}_{\nu+1}$ 的鞅 $\widetilde{}$ 结构，递归地用前一步的矩界和鞅差的矩界来界定当前步骤的矩（例如见 (C.26)）。通过反复迭代，我们证明 (C.19) 至多为 $\exp(\widetilde{O}(k^2p^2/d))$，而当 $d = \widetilde{\Omega}(k^2p^2)$ 时，该值为 $O(1)$。

从现在起，我们进一步以 $\Gamma_{[\nu]}$ 为条件（带有 $\sum_{i=1}^{\nu}\Gamma_i \le 4kp$）。记 ρ 为 $\mathbb{S}^{d-1}$ 上的均匀概率测度。此外，对于 $\mathbb{S}^{d-1}$ 上的任意分布 μ，记 $\|\mu\|_\infty := \sup_x |(d\mu/d\rho)(x)|$。特别地，若对于某个可测集 $L \subseteq \mathbb{S}^{d-1}$，有 $\mu = \rho(\cdot \cap L)/\rho(L)$，则有 $\|\mu\|_\infty = 1/\rho(L)$。

我们先介绍下面的引理，它将用于刻画每个“鞅差”（稍后将定义）的集中性（进而控制其矩）。

引理 C.2 ([LMSY22，推论 4。9])。存在一个常数 $C_{C.2} > 0$，使得以下结论成立：设 μ 是 $\mathbb{S}^{d-1}$上的一个分布，且对于 $z \in \mathbb{S}^{d-1}$，令 $X(z) := \mathbb{P}_{x\sim\mu}(\langle x, z\rangle > \tau(p, d))$。对于任意$0 \le s \le 1$,

$$\mathbb{P}_{z\sim\rho}(|X(z) - p| > ps) \le 2\exp\left(-\frac{ds^2}{C_{C.2}\log(1/p)\log(d/p)(\log\|\mu\|_\infty + \log(d/p))}\right).$$

我们使用与 [LMSY22中观察 5 相同的鞅。3]给定序列 Γ_i，$i \in [\nu]$根据条件化，定义以下随机集合序列 E_i，$i \in [\nu]$ 由 p-帽集组成（当 $\Gamma_i = 1$时）以及 p-反帽集（当 $\Gamma_i = 0$时），定义如下：

$$E_i := \begin{cases} \{z \in \mathbb{S}^{d-1} : \langle U_i, z\rangle \ge \tau(p,d)\} & \Gamma_i = 1, \\ \{z \in \mathbb{S}^{d-1} : \langle U_i, z\rangle < \tau(p,d)\} & \Gamma_i = 0\,. \end{cases}$$

注意，在 E_i个 $\sum_{i=1}^{\nu}\Gamma_i$ 中，有 p个是 $\nu - \sum_{i=1}^{\nu}\Gamma_i$-帽集，$p$个是$0 \le t \le \nu$-反帽集。现在，对于，定义如下量

$$R_t := \frac{\rho(L_t)}{\prod_{i=1}^{t}\rho(E_i)} = \frac{\rho(L_t)}{p^{\sum_{i=1}^{t}\Gamma_i}(1-p)^{t-\sum_{i=1}^{t}\Gamma_i}},$$

其中 $R_0 := 1$ 以及 $L_t := \cap_{i=1}^{t}E_i$，其中 $L_0 := \mathbb{S}^{d-1}$；注意，等式由每个 E_i的定义推出。根据 [LMSY22，观察 5，3]，$\{R_t\}$ 是关于滤过 $\mathcal{F}_t := \sigma(\{U_1, \dots, U_t\})$ 的鞅。此外，根据定义，我们有

$$\frac{\widetilde{\mathcal{P}}_{\nu+1}(\Gamma_{[\nu]}|U_{[\nu]})}{\mathcal{Q}_{\nu+1}(\Gamma_{[\nu]})}|\Gamma_{[\nu]} \overset{\mathrm{d}}{=} R_\nu\,. \tag{C.20}$$

因此，只需控制 R_ν的二阶矩。具体做法是将 R_ν 分解为（乘法）鞅差的乘积，并利用引理C.2分析每个鞅差。请注意，这与 [LMSY22中的方法不同，见引理 5。1], 其中 R_ν 被表示为（加法）鞅差之和。形式化地说，有

$$R_\nu = R_0\prod_{i=1}^{\nu}\varphi_i\,,$$

其中对于 $i \in [\nu]$,

$$\varphi_i := R_i / R_{i-1} = \frac{\rho(E_i \cap L_{i-1})}{\rho(E_i)\rho(L_{i-1})}.$$

下一步 是构造 $\mathbb{E}[R_t^2]$ 和 $\mathbb{E}[R_{t-1}^2]$之间的递归不等式。对于任意 $t \in [\nu$],

$$\begin{aligned}\mathbb{E}[R_t^2] = \mathbb{E}[R_{t-1}^2 \mathbb{E}[\varphi_t^2 | \mathcal{F}_{t-1}]] &= \mathbb{E}[R_{t-1}^2 + R_{t-1}^2 \mathbb{E}[(\varphi_t^2 - 1) | \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[R_{t-1}^2 + R_{t-1}^2 \mathbb{E}[(\varphi_t - 1)^2 | \mathcal{F}_{t-1}]],\end{aligned} \tag{C.21}$$

其中，最后一个等号由 $\mathbb{E}[\varphi_t | \mathcal{F}_{t-1}] = 1$得出。

情形 1： $\Gamma_t = 1$。 首先，考虑 $\Gamma_t = 1$ 的情形，即 E_t 是一个 p-帽。若 $R_{t-1} \geq a$ （ a 的值稍后确定），则

$$\rho(L_{t-1}) = R_{t-1} \prod_{i=1}^{t-1} \rho(E_i) \geq a p^{\sum_{i=1}^{\nu} \Gamma_i} (1-p)^{\nu - \sum_{i=1}^{\nu} \Gamma_i},$$

这意味着

$$\|\mu\|_\infty = 1/\rho(L_{t-1}) \leq 1/(a p^{\sum_{i=1}^{\nu} \Gamma_i} (1-p)^{\nu - \sum_{i=1}^{\nu} \Gamma_i}). \tag{C.22}$$

接着，应用引理 C.2 并取 $\mu = \rho(\cdot | L_{t-1})/\rho(L_{t-1})$；对于 $\{R_{t-1} \geq a\} \in \mathcal{F}_{t-1}$，我们有$\mathbb{E}[($
$\varphi_t - 1)^2 | \mathcal{F}_{t-1}]\mathbf{1}\{R_{t-1} \geq a\}$

$$\begin{aligned}&\leq \int_0^{(1/p)^2} \mathbb{P}((\varphi_t - 1)^2 > s | \mathcal{F}_{t-1}) ds \mathbf{1}\{R_{t-1} \geq a\} \\ &= \int_0^{1/p} 2u \mathbb{P}(|\varphi_t - 1| > u | \mathcal{F}_{t-1}) du \mathbf{1}\{R_{t-1} \geq a\} \\ &\leq \left(\int_0^1 4u \exp\left(- \frac{du^2}{C_{C.2} \log(1/p) \log(d/p) \lambda_a} \right) du + \frac{2}{p^2} \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda_a} \right) \right) \mathbf{1}\{R_{t-1} \geq a\} \\ &\leq \left(\frac{C_{C.2} \log(1/p) \log(d/p) \lambda_a}{d} + \frac{2}{p^2} \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda_a} \right) \right) \mathbf{1}\{R_{t-1} \geq a\},\end{aligned} \tag{C.23}$$

其中 $\lambda_a :=$ 对数 $\|\mu\|_\infty +$ 对数(d/p)；第一个不等式由 $\varphi_t \leq 1/p$得出，第二个不等式由引理 $\mathbb{P}($
$|\varphi_t - 1| > u | \mathcal{F}_{t-1}) = \mathbb{P}(|X(z) - p| > pu)$ 中的 C.2 (for $u \in [1, 1/p]$ 得出，我们只需使用 $u = 1$的上界）。注意，由（C.22），

$$\begin{aligned}\lambda_a &\leq \log(1/a) + \left(\sum_{i=1}^{\nu} \Gamma_i \right) \log(1/p) + \left(\nu - \sum_{i=1}^{\nu} \Gamma_i \right) \log(1/(1-p)) + \log(d/p) \\ &\leq \log(1/a) + 10kp \log(1/p) + \log(d/p) =: \lambda_a',\end{aligned} \tag{C.24}$$

由 $\nu \leq 2k$ 和 $\sum_{i=1}^{\nu} \Gamma_i \leq 4kp$，以及对数$(1/(1-p)) \leq 2p$ 和对数$(1/p) \geq 2/3$。特别是，尽管 λ_a 本身取决于 t，但 (λ_a' C.24) 中的最后一项不取决于 t。

另一方面，在 $\{R_{t-1} < a\} \in \mathcal{F}_{t-1}$ 上，我们有

$$\mathbb{E}[R_t^2 \mathbf{1}\{R_{t-1} < a\}] = \mathbb{E}[R_{t-1}^2 \mathbb{E}[\varphi_t^2 | \mathcal{F}_{t-1}] \mathbf{1}\{R_{t-1} < a\}] \leq a^2/p^2, \tag{C.25}$$

由 $\varphi_t \leq 1/p$。将（C.23）和（C.25）结合起来，代入（C.21），对于 $\Gamma_t = 1$，我们有

$$\mathbb{E}[R_t^2] = \mathbb{E}[R_t^2 (\mathbf{1}\{R_{t-1} \geq a\} + \mathbf{1}\{R_{t-1} < a\})] \leq \mathbb{E}[R_{t-1}^2](1 + \eta_a) + \frac{a^2}{p^2}, \tag{C.26}$$

其中

$$\eta_a := \frac{C_{C.2} \log(1/p) \log(d/p) \lambda_a'}{d} + \frac{2}{p^2} \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda_a'} \right). \tag{C.27}$$

情况 2： $\Gamma_t = 0$。 $\Gamma_t = 0$ 情况的界可以通过与第一种情况类似的方法得到。在 $\{R_{t-1} \geq a\}$，

$$
\begin{aligned}
&\mathbb{E}[(\varphi_t - 1)^2 | \mathcal{F}_{t-1}] \mathbf{1}\{R_{t-1} \geq a\} \\
&\leq \int_0^1 2u \mathbb{P}(|\varphi_t - 1| > u | \mathcal{F}_{t-1}) du \mathbf{1}\{R_{t-1} \geq a\} \\
&\leq \int_0^{p/(1-p)} 4u \exp\left(- \frac{d(1-p)^2 u^2}{C_{C.2} p^2 \log(1/p) \log(d/p) \lambda'_a} \right) du \mathbf{1}\{R_{t-1} \geq a\} \\
&\quad + \int_{p/(1-p)}^1 4u \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda'_a} \right) du \mathbf{1}\{R_{t-1} \geq a\} \\
&\leq \left(\frac{2 C_{C.2} p^2 \log(1/p) \log(d/p) \lambda'_a}{d(1-p)^2} + 4 \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda'_a} \right) \right) \mathbf{1}\{R_{t-1} \geq a\} \\
&\leq \left(\frac{8 C_{C.2} p^2 \log(1/p) \log(d/p) \lambda'_a}{d} + 4 \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda'_a} \right) \right) \mathbf{1}\{R_{t-1} \geq a\} ,
\end{aligned}
$$

其中，第一个不等式来自 $|\varphi_t - 1| \leq 1 \vee p/(1-p) \leq 1$，第二个不等式来自引理 C.2，其中 $\mu = \rho(\cdot | L_{t-1})/\rho(L_{t-1})$，且 $\mathbb{P}(|\varphi_t - 1| > u | \mathcal{F}_{t-1}) = \mathbb{P}(|X(z) - p| > (1-p)u)$（应用于 $u < p/(1-p)$；对于 $u \in [p/(1-p), 1]$，，使用 $u = p/(1-p)$ 的上界）；最后一个不等式来自 $p \leq 1/2$。

在 $\{R_{t-1} < a\} \in \mathcal{F}_{t-1}$ 上，我们有

$$
\mathbb{E}[R_t^2 \mathbf{1}\{R_{t-1} < a\}] = \mathbb{E}[R_{t-1}^2 \mathbb{E}[\varphi_t^2 | \mathcal{F}_{t-1}] \mathbf{1}\{R_{t-1} < a\}] \leq a^2/(1-p)^2 \leq 4a^2 .
$$

将这些结果按（C.26）中的方法合并，可得

$$
\mathbb{E}[R_t^2] \leq \mathbb{E}[R_{t-1}^2](1 + \theta_a) + 4a^2 , \tag{C.28}
$$

其中

$$
\theta_a := \frac{8 C_{C.2} p^2 \log(1/p) \log(d/p) \lambda'_a}{d} + 4 \exp\left(- \frac{d}{C_{C.2} \log(1/p) \log(d/p) \lambda'_a} \right) . \tag{C.29}
$$

现在，对 $t = \nu$、$\nu - 1$、$\cdots$、1 依次应用（C.26）和（C.28），可得（回忆 $\mathbb{E}[R_0^2] = 1$）

$$
\begin{aligned}
\mathbb{E}[R_\nu^2] &\leq (1 + \eta_a)^{\sum_{i=1}^\nu \Gamma_i} (1 + \theta_a)^{\nu - \sum_{i=1}^\nu \Gamma_i} \left(1 + \left(\sum_{i=1}^\nu \Gamma_i \right) \frac{a^2}{p^2} + \left(\nu - \sum_{i=1}^\nu \Gamma_i \right) 4a^2 \right) \\
&\leq \exp\left(\left(\sum_{i=1}^\nu \Gamma_i \right) \eta_a + \left(\nu - \sum_{i=1}^\nu \Gamma_i \right) \theta_a \right) \times \left(1 + \left(\sum_{i=1}^\nu \Gamma_i \right) \frac{a^2}{p^2} + \left(\nu - \sum_{i=1}^\nu \Gamma_i \right) 4a^2 \right) \\
&\leq \exp(4kp\eta_a + 2k\theta_a) \times \left(1 + \frac{4ka^2}{p} + 8ka^2 \right) ,
\end{aligned}
$$

其中第二个不等式来自 $1 + x \leq \exp(x)$，最后一个不等式来自 $\nu \leq 2k$，$\sum_{i=1}^\nu \Gamma_i \leq 4kp$。

最后，设定 $a = p/k$。那么当 $\lambda'_a = \log(dk/p^2) + 10kp \log(1/p)$ 时，我们有

$$
\mathbb{E}[R_\nu^2] \leq \exp\left(\frac{C_1 kp \log(1/p) \log(d/p) \lambda'_a}{d} + \frac{C_1 k}{p} \exp\left(- \frac{d}{C_1 \log(1/p) \log(d/p) \lambda'_a} \right) \right) \times (1 + C_1 p/k) ,
$$

其中某个常数为 $C_1 > 0$，这分别由 (C.27) 和 (C.29) 中 η_a 和 θ_a 的定义得出。只要满足 $d \gtrsim kp\log(1/p)\log(d/p)$ $\lambda'_a = kp\log(1/p)\log(d/p)(10kp\log(1/p) + \log(dk/p^2))$ 且 $kp \gtrsim \log n$，最后一项（因此由 (C.20) 可知，(C.19)）为 $O(1)$。

综合以上下界 d 和 kp 对于项 (I)–(III)，我们有引理 5.5 成立，只要

$$kp \geq C_{5.5}\log n \quad \text{and} \quad d \geq C_{5.5}(kp\log(1/p)\log(d/p))^2\log n\,,$$

对常数 $C_{5.5} > 0$ 作适当选择即可，正如所需。

D 第 6 节中的延后证明6

D.1 关于带符号环计数的紧下界（命题 6.1）

我们对命题 6.1 的证明策略相当直接（尽管涉及技术性计算）：首先，针对潜变量 $U_1, \ldots, U_\ell$，选取适当的标准正交基，并写出有符号子图的期望。正交性大幅简化了表达式，将其化为一个由组合对象构成的易处理求和。

D.1.1 Gegenbauer 多项式与球谐函数

这里介绍计算中使用的标准正交基的若干关键事实。这些内容可见于调和分析教材，例如：[DX13]；摘要见 [LS23，第 3 节和引理 4。10]。对于相互独立的 U_1，$U_2 \sim \mathcal{U}(\mathbb{S}^{d-1})$，$\mu$ 的分布 $\langle U_1, U_2\rangle$ 可表示为

$$\mu(dx) = \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}}(1-x^2)^{(d-3)/2}dx, \quad x \in [-1,1]\,.$$

在 $\mathcal{L}^2(\mu)$ 中标准正交的多项式 $q_0, q_1, \ldots$ 称为 Gegenbauer 多项式。前几个多项式如下：

$$q_0(x) = 1, \quad q_1(x) = \sqrt{d}x, \quad q_2(x) = \frac{1}{\sqrt{2}}\sqrt{\frac{d+2}{d-1}}(dx^2-1)\,.$$

此外，这些多项式满足以下递归性质：

$$q_{m+1}(x) = \sqrt{\frac{(2m+d)(2m+d-2)}{(m+1)(m+d-2)}}xq_m(x) - \sqrt{\frac{m(m+d-3)(m+d/2)}{(m+1)(m+d-2)(m+d/2-2)}}q_{m-1}(x) \quad \text{(D.1)}$$

Gegenbauer 多项式还可以进一步分解为球谐函数，这些函数关于 U_1 的分布构成标准正交系。具体而言，存在一组函数 $\{\phi_{m,t} : m \geq 0, t \in [N_m]\}$，使得

$$\mathbb{E}_{U_1}[\phi_{m,t}(U_1)\phi_{m',t'}(U_1)] = \mathbf{1}\{m = m', t = t'\}\,,$$

and

$$q_m(\langle U_1, U_2\rangle) = \frac{1}{\sqrt{N_m}}\sum_{t\in[N_m]}\phi_{m,t}(U_1)\phi_{m,t}(U_2)\,.$$

这里，N_m 表示不同的m 阶球谐函数的数量（对应于 $\mathbb{S}^{d-1}$），它满足

$$N_0 = 1, \quad N_m = \frac{d+2m-2}{m}\binom{d+m-3}{m-1}, m \geq 1\,.$$

由于 $\mathbf{1}\{\cdot \geq \tau(p,d)\} \in \mathcal{L}^2(\mu)$，我们可以写成

$$\mathbf{1}\{\langle U_1, U_2\rangle \geq \tau(p,d)\} = \sum_{m=0}^{\infty} c_m q_m(\langle U_1, U_2\rangle) = \sum_{m=0}^{\infty} \frac{c_m}{\sqrt{N_m}} \sum_{t=1}^{N_m} \phi_{m,t}(U_1)\phi_{m,t}(U_2)\,, \tag{D.2}$$

其中 $c_m := \langle \mathbf{1}\{\cdot \geq \tau(p,\ d)\},\ q_m\rangle_{\mathcal{L}^2(\mu)}$。

D.1.2 带符号环计数的紧确界

在确立这些事实之后，我们继续证明命题 6.1。回顾我们关注的主要量：

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} (G_{ij}-p)\right].$$

首先注意，Cyc_ℓ 的任意真子图都是森林。因此，根据推论4.2，

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} G_{ij}\right] - p^\ell = \mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} (G_{ij}-p)\right]. \tag{D.3}$$

然后根据（D.2），

$$\begin{aligned}
&\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} G_{ij}\right] \\
&= \mathbb{E}_{U_1,\ldots,U_\ell}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} \mathbf{1}\{\langle U_i, U_j\rangle \geq \tau(p,d)\}\right] \\
&= \mathbb{E}_{U_1,\ldots,U_\ell}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} \left(\sum_{m=0}^{\infty} \frac{c_m}{\sqrt{N_m}} \sum_{t\in[N_m]} \phi_{m,t}(U_i)\phi_{m,t}(U_j)\right)\right] \\
&= \sum_{m_1,\ldots,m_\ell\geq 0} \frac{c_{m_1}\cdots c_{m_\ell}}{\sqrt{N_{m_1}\cdots N_{m_\ell}}} \sum_{t_1\in[N_{m_1}],\ldots,t_\ell\in[N_{m_\ell}]} \mathbb{E}_{U_1,\ldots,U_\ell}[\phi_{m_1,t_1}(U_1)\phi_{m_1,t_1}(U_2)\cdots\phi_{m_\ell,t_\ell}(U_\ell)\phi_{m_\ell,t_\ell}(U_1)]\,,
\end{aligned}$$

其中每个求和项都满足（由球谐函数的正交归一性可知）

$$\begin{aligned}
&\mathbb{E}_{U_1,\ldots,U_\ell}[\phi_{m_1,t_1}(U_1)\phi_{m_1,t_1}(U_2)\cdots\phi_{m_\ell,t_\ell}(U_\ell)\phi_{m_\ell,t_\ell}(U_1)] \\
&= \mathbb{E}_{U_1}[\phi_{m_\ell,t_\ell}(U_1)\phi_{m_1,t_1}(U_1)]\cdots\mathbb{E}_{U_\ell}[\phi_{m_{\ell-1},t_{\ell-1}}(U_\ell)\phi_{m_\ell,t_\ell}(U_\ell)] \\
&= \mathbf{1}\{m_\ell = m_1, t_\ell = t_1\}\cdots\mathbf{1}\{m_{\ell-1} = m_\ell, t_{\ell-1} = t_\ell\} \\
&= \mathbf{1}\{m_1 = \cdots = m_\ell, t_1 = \cdots = t_\ell\}\,.
\end{aligned}$$

因此，我们得到

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} G_{ij}\right] = \sum_{m=0}^{\infty} \frac{c_m^\ell}{N_m^{\ell/2}} N_m = \sum_{m=0}^{\infty} \frac{c_m^\ell}{N_m^{\ell/2-1}}\,.$$

由于 $c_0 = \langle \mathbf{1}\{\cdot \geq \tau(p,\ d)\},\ q_0\rangle_{\mathcal{L}^2(\mu)} = \mathbb{P}(\langle U_1,\ U_2\rangle \geq \tau(p,\ d)) = p$ 以及 $N_0 = 1$，由 (D.3) 可得

$$\mathbb{E}_{G\sim\mathcal{G}(n,p,d)}\left[\prod_{ij\in E(\mathrm{Cyc}_\ell)} (G_{ij}-p)\right] = \sum_{m=1}^{\infty} \frac{c_m^\ell}{N_m^{\ell/2-1}}\,. \tag{D.4}$$

因此，只需界定由 Gegenbauer 多项式产生的 c_m 和 N_m 的大小。为此，我们根据 m的取值对无穷和进行分解。特别地，我们声称对于某个常数 $C>0$,

$$\frac{1}{C^\ell}\frac{p^\ell \log^{\ell/2}(1/p)}{d^{\ell/2-1}} \le \frac{c_1^\ell}{N_1^{\ell/2-1}} \le C^\ell \frac{p^\ell \log^{\ell/2}(1/p)}{d^{\ell/2-1}}, \tag{D.5}$$

$$\left|\sum_{m=2}^{\lfloor d^{1/4}\rfloor} \frac{c_m^\ell}{N_m^{\ell/2-1}}\right| = o\left(\frac{p^\ell \log^{\ell/2}(1/p)}{C^\ell d^{\ell/2-1}}\right), \tag{D.6}$$

$$\left|\sum_{m=\lfloor d^{1/4}\rfloor+1}^{\infty} \frac{c_m^\ell}{N_m^{\ell/2-1}}\right| = o\left(\frac{p^\ell}{C^\ell d^{\ell/2-1}}\right), \tag{D.7}$$

从现在起，渐近关系均针对 $d\to\infty$，并且在 $3\le \ell \le n$ 上均匀成立；显然，命题 6.1 由此得出。对于（D.5）至（D.7），我们将用 C_0，C_1，.....表示各情形中的绝对常数。

首项（$m=1$）。对于 $m=1$，有 $N_1=d$ 以及

$$c_1 = \int_{\tau(p,d)}^1 \sqrt{d}x \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}}(1-x^2)^{(d-3)/2}dx = \frac{\sqrt{d}\Gamma(d/2)}{\Gamma((d-1)/2)(d-1)\sqrt{\pi}}(1-\tau(p,d)^2)^{(d-1)/2}.$$

我们证明 c_1（等价地，$(1-\tau(p,d)^2)^{(d-1)/2}$）的阶为 $p\sqrt{\log(1/p)}$。证明分为两种情形：当 p 较小（接近 0）或较大（远离 0）时。对于较小的 p，由 [BDER16，引理 2] 可知，存在 $C_0>0$，使得对所有 $p\in(0,\ 0.49]$,

$$\frac{1}{C_0}\sqrt{\frac{\log(1/p)}{d}} \le \tau(p,d) \le C_0\sqrt{\frac{\log(1/p)}{d}}.$$

特别地，存在 $\varepsilon_0>0$，使得对所有 $p\in(0,\ \varepsilon_0]$，$\tau(p,\ d)\ge(1/C_0)\sqrt{\log(1/p)/d \ge \sqrt{2}/d}$。然后由 [BGK 引理 2.1(b)],

$$\frac{2}{C_0}p\sqrt{\log(1/p)} \le 2p\sqrt{d}\tau(p,d) \le (1-\tau(p,d)^2)^{(d-1)/2} \le 6p\sqrt{d}\tau(p,d) \le 6C_0 p\sqrt{\log(1/p)},$$

这意味着，当 $p\in(0,\varepsilon_0]$时，c_1 的阶为 $p\sqrt{\log(1/p)}$。对于 $p\in(\varepsilon_0,0.5]$，由于 $p\sqrt{\log(1/p)}$ 的阶为常数，只需证明在相同的 p 范围内，$(1-\tau(p,d)^2)^{(d-1)/2}$ 的阶为常数（对 d 均匀成立）。[LMSY22，引理 3.6] 指出

$$\tau(p,d) \le \sqrt{\frac{3\log(1/p)}{d}} \quad \text{for all} \quad 0<p\le 0.5. \tag{D.8}$$

因此，对于所有充分大的 d,

$$p^2 \le (1-\tau(p,d)^2)^{(d-1)/2} \le 1.$$

这表明，$(1-\tau(p,d)^2)^{(d-1)/2}$ 在 $p\in(\varepsilon_0,0.5]$上是常数阶。因此，c_1 的阶为 $p\sqrt{\log(1/p)}$，从而证明了 (D.5)。

较小的 m $(2 \le m \le d^{1/4})$。简要来说，在这一 m 取值范围内，我们将利用递归定义，归纳地控制 (c_m和 N_m 在D.4) 中的大小。为此，我们首先考虑 $\widetilde{c}_m \ge 0$ 的一个上界 $|c_m|$，其定义如下：

$$\widetilde{c}_m := \int_{\tau(d,p)}^1 |q_m(x)|\mu(dx)\,.$$

我们的第一个结论是，存在一个常数 $C_0 > 0$，使得

$$\widetilde{c}_m \le C_0 p\sqrt{\log(1/p)}(0.9d)^{(m-1)/8}\,, \tag{D.9}$$

对所有 $m \ge 1$。这可以通过归纳法证明；当 $m = 1$时，由 $\widetilde{c}_1 = c_1$ 以及对 c_1 的相应分析可得（在（D.5）的证明中），而当 $m = 2$ 时，由

$$\begin{aligned}
\widetilde{c}_2 &\le \int_{\tau(p,d)}^1 (dx^2+1)\frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}}(1-x^2)^{(d-3)/2}dx \\
&= \frac{d}{d-1}\tau(p,d)\times\frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}}(1-\tau(p,d)^2)^{(d-1)/2} \\
&\quad + \frac{d}{d-1}\times\int_{\tau(p,d)}^1 \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}}(1-x^2)^{(d-1)/2}dx + p \\
&\le \sqrt{d}\tau(p,d)c_1 + 3p \le C_0 p\sqrt{\log(1/p)}\left(\sqrt{d}\tau(p,d) + \frac{3}{C_0\sqrt{\log(1/p)}}\right)\,,
\end{aligned}$$

其中，第一个等式来自分部积分，第二个不等式来自 $d/(d-1) \le 2$，$(1-x^2)^{(d-1)/2} \le (1-x^2)^{(d-3)/2}$ 以及 c_1 的定义，而最后一个不等式来自$m = 1$的(D.9)。由 (D.8) 和 $d \ge (5\log(1/p))^4$, 可知，对于所有充分大的 d，因子 $\sqrt{d}\tau(p,d) + 3/(C_0\sqrt{\log(1/p)})$ 小于 $(0.9d)^{1/8}$。

现在，我们利用 Gegenbauer 多项式的递归定义，为 $\widetilde{c}_m$构造一个递归不等式。可以证明，式（D.1）中的因子满足

$$\sqrt{\frac{(2m+d)(2m+d-2)}{(m+1)(m+d-2)}} \le \sqrt{d} \quad \text{and} \quad \sqrt{\frac{m(m+d-3)(m+d/2)}{(m+1)(m+d-2)(m+d/2-2)}} \le 1\,,$$

对所有 $d \ge 4$、$m \ge 1$ 均成立。这意味着

$$|q_{m+1}(x)| \le \sqrt{d}x|q_m(x)| + |q_{m-1}(x)|\,,$$

对所有 $x \in [\tau(p,\ d), 1]$。因此，

$$\begin{aligned}
\widetilde{c}_{m+1} &\le \sqrt{d}\int_{\tau(p,d)}^1 x|q_m(x)|\mu(dx) + \int_{\tau(p,d)}^1 |q_{m-1}(x)|\mu(dx) \\
&= \int_{\tau(p,d)}^{0.9/d^{3/8}} \sqrt{d}x|q_m(x)|\mu(dx) + \int_{0.9/d^{3/8}}^1 \sqrt{d}x|q_m(x)|\mu(dx) + \widetilde{c}_{m-1} \\
&\le 0.9d^{1/8}\int_{\tau(p,d)}^1 |q_m(x)|\mu(dx) + \left(\int_{0.9/d^{3/8}}^1 dx^2\mu(dx)\right)^{1/2}\left(\int_{0.9/d^{3/8}}^1 (q_m(x))^2\mu(dx)\right)^{1/2} + \widetilde{c}_{m-1} \\
&\le 0.9d^{1/8}\widetilde{c}_m + \left(\int_{0.9/d^{3/8}}^1 d\mu(dx)\right)^{1/2} + \widetilde{c}_{m-1} \le 0.9d^{1/8}\widetilde{c}_m + 2\sqrt{d}\exp\left(-\frac{(0.9)^2}{4}d^{1/4}\right) + \widetilde{c}_{m-1}\,,
\end{aligned}$$

其中，第二个不等式由柯西-施瓦茨不等式得出，第三个不等式由 $\|q_m\|_{\mathcal{L}^2(\mu)} = 1$ 得出，最后一个不等式则由球冠集中性 [Mat13，定理 14.1 得出。1]。

现在应用归纳假设；假设 (D.9) 对 m 和 $m-1$成立，其中$m \geq 2$。于是

$$\begin{aligned}\widetilde{c}_{m+1} &\leq 0.9d^{1/8} \times C_0 p\sqrt{\log(1/p)}(0.9d)^{(m-1)/8} + 2\sqrt{d}\exp\left(-\frac{(0.9)^2}{4}d^{1/4}\right) + C_0 p\sqrt{\log(1/p)}(0.9d)^{(m-2)/8} \\ &\leq C_0 p\sqrt{\log(1/p)}(0.9d)^{(m-2)/8}(0.9d^{1/4}+2) \leq C_0 p\sqrt{\log(1/p)}(0.9d)^{m/8}\,,\end{aligned}$$

其中，第二个不等式由 $2\sqrt{d}\exp(-(0.9)^2 d^{1/4}/4) \leq \exp(-d^{1/4}/5) \leq p$ 得出，它小于 $C_0 p\sqrt{\log(1/p)}(0.9d)^{(m-2)/8}$；最后一个不等式则由 $(0.9d)^{1/4} \geq 0.9d^{1/4} + 2$ 在所有充分大的 d时成立。这就证明了第一个结论 (D.9)。

我们的第二个结论是

$$N_m \geq \frac{d^2}{2} \times (0.9d^3)^{(m-2)/4}\,, \tag{D.10}$$

对所有 $2 \leq m \leq d^{1/4}$ 均成立。对此也可以用归纳法证明，因为 $N_2 = \binom{d+2}{2}\binom{d-1}{1} \geq \frac{d^2}{2}$ 和

$$\frac{N_m}{N_{m-1}} = \frac{(d+2m-2)(d+m-3)}{m(d+2m-4)} \geq \frac{d+2m-2}{m} \times (0.9)^{1/4} \geq (0.9d^3)^{1/4}\,.$$

结合 (D.9) 和 (D.10)，可得

$$\begin{aligned}\left|\sum_{2\leq m\leq d^{1/4}} \frac{c_m^\ell}{N_m^{\ell/2-1}}\right| &\leq \sum_{2\leq m\leq d^{1/4}} \frac{(\widetilde{c}_m)^\ell}{N_m^{\ell/2-1}} \\ &\leq \frac{2^{\ell/2-1}(C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{\ell-2}}(0.9d)^{\ell/8} \sum_{m\geq 0} \frac{(0.9d)^{\ell m/8}}{(0.9d^3)^{(\ell/2-1)(m/4)}} \\ &= \frac{2^{\ell/2-1}(C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{\ell-2}}(0.9d)^{\ell/8} \sum_{m\geq 0} \left(\frac{1}{d^{1/2}}\right)^{(\ell/2-1)m} (0.9d)^{m/4} \\ &\leq \frac{2^{\ell/2-1}(C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{\ell-2}}(0.9d)^{\ell/8} \sum_{m\geq 0} \left(\frac{1}{d^{1/2}}\right)^{m/2} (0.9d)^{m/4} \\ &\leq \frac{2^{\ell/2-1}(C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{\ell-2}}(0.9d)^{\ell/8} \sum_{m\geq 0} (0.9)^{m/4} \\ &\leq \frac{100(2C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{7\ell/8-2}}\,,\end{aligned}$$

其中，在第三个和最后一个不等式中，我们使用 $\ell \geq 3$。令 $C > 0$ 为（D.5）中的常数。于是

$$\frac{100(2C_0 p)^\ell \log^{\ell/2}(1/p)}{d^{7\ell/8-2}} = \frac{p^\ell \log^{\ell/2}(1/p)}{C^\ell d^{\ell/2-1}} \times \frac{100(2C_0 C)^\ell}{d^{3\ell/8-1}}\,,$$

其中，因子 $100(2C_0 C)^\ell/d^{3\ell/8-1}$ 对于所有 o，均匀地为 $\ell \geq 3$ 当 $d \to \infty$时。这就证明了 (D.6)。

较大的 m **(**$m > d^{1/4}$**)**。对于 m的这一范围，我们不再递归地界定 c_m的大小，而是直接调用 Gegenbauer 多项式的近似结果。这会得到一个相当复杂的伽马函数表达式。由于 N_m 也可以用伽马函数表示，我们采用斯特林近似来处理这些表达式，并仔细界定 (D.4) 中由此得到的求和项，从而证明剩余的无穷和很小。

首先，我们有

$$
\begin{aligned}
|c_m| &\le \int_{\tau(p,d)}^1 |q_m(x)| \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}} (1-x^2)^{(d-3)/2} dx \\
&= \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}} \int_{\tau(p,d)}^1 \sqrt{N_m} \frac{\Gamma(m+1)\Gamma((d-1)/2)}{\Gamma((d-1)/2+m)} |P_m^{(d-3)/2,(d-3)/2}(x)| (1-x^2)^{(d-3)/2} dx \\
&= \frac{\sqrt{N_m}\Gamma(d/2)\Gamma(m+1)}{\Gamma((d-1)/2+m)\sqrt{\pi}} \int_{\tau(p,d)}^1 |P_m^{(d-3)/2,(d-3)/2}(x)| (1-x^2)^{(d-3)/2} dx \\
&\le \frac{\sqrt{N_m}\Gamma(d/2)\Gamma(m+1)}{\Gamma((d-1)/2+m)\sqrt{\pi}} \int_{\tau(p,d)}^1 \sqrt{3} \left(\frac{d-3}{2}\right)^{1/6} \left(1+\frac{d-3}{2m}\right)^{1/12} (1-x^2)^{d/4-1} dx \\
&\le \frac{\sqrt{N_m}\Gamma(d/2)\Gamma(m+1)}{\Gamma((d-1)/2+m)} d^{1/4} \int_{\tau(p,d)}^1 (1-x^2)^{d/4-1} dx \\
&\le \frac{\sqrt{N_m}\Gamma(d/2)\Gamma(m+1)}{\Gamma((d-1)/2+m)} d^{1/4} \times \frac{C_0}{\sqrt{d}} \le \frac{C_0\sqrt{N_m}\Gamma(d/2)\Gamma(m+1)}{\Gamma((d-1)/2+m) d^{1/4}} \times p \exp\left(\frac{1}{5} d^{1/4}\right),
\end{aligned}
$$

对于某个常数 $C_0 > 0$。第一个等式来自 [DX13，公式 B.2。1],[13] 第二个不等式来自 [Kra07，定理 2],，第三个不等式来自 $\sqrt{3/\pi}((d-3)/2)^{1/6}(1+(d-3)/(2m))^{1/12} \le d^{1/4}$；接下来的两个不等式分别来自 $\int_{-1}^1 (1-x^2)^{d/4-1} dx \lesssim 1/\sqrt{d}$和 $d \ge (5\log(1/p))^4$。因此，

$$
\frac{d^{\ell/2-1}}{(C_0 p)^\ell} \times \frac{|c_m|^\ell}{N_m^{\ell/2-1}} = \left(\frac{|c_m|}{C_0 p \sqrt{N_m}}\right)^\ell d^{\ell/2-1} N_m \le \frac{\Gamma(d/2)^\ell \Gamma(m+1)^\ell}{\Gamma((d-1)/2+m)^\ell} \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4} \log d\right) \times N_m . \tag{D.11}
$$

请注意，我们关注的主要求和项 $|c_m|^\ell / N_m^{\ell/2-1}$在此处已经归一化；对于 (D.7)，只需证明 (D.11) 在 $o(1/(C_0 C)^\ell)$ 上均匀成立 d。

由于我们需要对所有 $m > d^{1/4}$ 界定该和，因此在右侧我们汇集

[13] 多项式 $P_m^{(d-3)/2\,(d-3)/2}$,称为雅可比多项式，定义如下：

$$
P_m^{(d-3)/2,(d-3)/2}(x) := \frac{(-1)^m}{2^m m!} (1-x^2)^{-(d-3)/2} \frac{d^m}{dx^m} (1-x^2)^{m+(d-3)/2} .
$$

另请参见 [LS23，命题 4。3]。

依赖于 m 的因子，即，

$$\begin{aligned}
&\frac{\Gamma(m+1)^\ell}{\Gamma((d-1)/2+m)^\ell} \times N_m \\
&= \frac{\Gamma(m+1)^\ell}{\Gamma((d-1)/2+m)^\ell} \times \frac{d+2m-2}{m} \times \frac{\Gamma(d+m-2)}{\Gamma(m)\Gamma(d-1)} \\
&\le C_1^\ell \frac{(m+1)^{(m+1/2)\ell}}{((d-1)/2+m)^{(d/2+m-1)\ell}} \exp((d-3)\ell/2) \times \frac{d+2m-2}{m} \times \frac{(d+m-2)^{d+m-5/2}}{m^{m-1/2}(d-1)^{d-3/2}} \exp(1) \\
&\le C_2^\ell \frac{\exp((d-3)\ell/2)}{(d-1)^{d-3/2}} \times \frac{(m+1)^{(m+1/2)\ell}(d+m-2)^{d+m-5/2}}{((d-1)/2+m)^{(d/2+m-1)\ell-1} m^{m+1/2}} \\
&\le C_2^\ell \frac{\exp((d-3)\ell/2)2^d}{(d-1)^{d-3/2}} \times \frac{(m+1)^{(m+1/2)\ell}(d+m-2)^{m-5/2}}{((d-1)/2+m)^{(d/2+m-1)\ell-1-d} m^{m+1/2}} ,
\end{aligned} \tag{D.12}$$

对于某些常数 $C_1, C_2 > 0$，其中第一个不等式来自斯特林近似，第二个不等式来自 $d+2m-2 \le 2((d-1)/2+m)$，第三个不等式来自 $d+m-2 \le 2((d-1)/2+m)$。结合

$$\begin{aligned}
\frac{(m+1)^{m-5/2}(m+d-2)^{m-5/2}}{(m+(d-1)/2)^{2m-5}} &\le 1 , \\
\frac{(m+1)^{2m+1}}{m^{m+1/2}(m+(d-1)/2)^{m+1/2}} &\le 1 , \\
\frac{(m+1)^{(m+1/2)\ell-(3m-3/2)}}{(m+(d-1)/2)^{(m+1/2)\ell-(3m-3/2)}} &\le 1 ,
\end{aligned}$$

(注意，最后一个不等式中我们使用了 $(m+1/2)\ell - (3m-3/2) \ge 0$，该式在 $\ell \ge 3$ 时成立）(D.12) 的最后一项有上界

$$C_2^\ell \frac{\exp((d-3)\ell/2)2^d}{(d-1)^{d-3/2}} \times \frac{1}{(m+(d-1)/2)^{(\ell/2-1)d-3\ell/2+2}} .$$

回到 (D.11) 并对 $m > d^{1/4}$ 求和，可得

$$\begin{aligned}
&\frac{d^{\ell/2-1}}{(C_0 p)^\ell} \sum_{m>d^{1/4}} \frac{|c_m|^\ell}{N_m^{\ell/2-1}} \\
&\le \Gamma(d/2)^\ell \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \times \sum_{m>d^{1/4}} \frac{\Gamma(m+1)^\ell}{\Gamma((d-1)/2+m)^\ell} N_m \\
&\le \Gamma(d/2)^\ell \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \times C_2^\ell \frac{\exp((d-3)\ell/2)2^d}{(d-1)^{d-3/2}} \times \sum_{m>d^{1/4}} \frac{1}{(m+(d-1)/2)^{(\ell/2-1)d-3\ell/2+2}} \\
&\le \Gamma(d/2)^\ell \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \times C_2^\ell \frac{\exp((d-3)\ell/2)2^d}{(d-1)^{d-3/2}} \times \frac{1}{((d+2d^{1/4}-3)/2)^{(\ell/2-1)d-3\ell/2+1}} ,
\end{aligned}$$

其中最后一个不等式是通过将求和与积分进行比较得到的（这里我们利用了 $(\ell/2-1)d > 3\ell/2$，这对所有充分大的 d 都成立，与 $\ell \ge 3$ 无关）。

由于最后一行只依赖于 d 和 ℓ，根据 $\Gamma(d/2) \lesssim (d/2)^{(d-1)/2}\exp(-d/2)$，最后一行

的上界还可以进一步收紧为

$$
\begin{aligned}
&C_3^\ell \left(\frac{d}{2}\right)^{(d-1)\ell/2} \exp(-d\ell/2) \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \frac{\exp((d-3)\ell/2) 2^d}{(d-1)^{d-3/2}} \frac{1}{((d+2d^{1/4}-3)/2)^{(\ell/2-1)d-3\ell/2+1}} \\
&\le C_3^\ell \left(\frac{d}{d+2d^{1/4}-3}\right)^{(\ell/2-1)d} \left(\frac{d}{2}\right)^{d-\ell/2} d^{3\ell/2-1} \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \exp(-3\ell/2) d^{3/2} \left(\frac{2}{d-1}\right)^d \\
&\le C_3^\ell \exp\left(-\frac{8}{5}\left(\frac{\ell}{2}-1\right) d^{1/4}\right) \left(\frac{d}{d-1}\right)^d (2d)^{\ell+1/2} \exp\left(\frac{\ell}{5} d^{1/4} + \frac{\ell}{4}\log d\right) \\
&\le \exp\left(-\left(\frac{3}{5}\ell - \frac{8}{5}\right) d^{1/4} + 2l\log d\right),
\end{aligned}
$$

对于某个常数 $C_3 > 0$；只要 d 足够大，最后一个不等式就成立。特别地，对于（D.5）中的常数 $C > 0$

$$
\exp\left(-\left(\frac{3}{5}\ell - \frac{8}{5}\right) d^{1/4} + 2l\log d\right) (C_0 C)^\ell
$$

在所有 $\ell \ge 3$ 上均匀地为 $o(1)$ 阶。这意味着

$$
\frac{d^{\ell/2-1}}{(C_0 p)^\ell} \sum_{m > d^{1/4}} \frac{|c_m|^\ell}{N_m^{\ell/2-1}} \times (C_0 C)^\ell = \frac{d^{\ell/2-1} C^\ell}{p^\ell} \sum_{m > d^{1/4}} \frac{|c_m|^\ell}{N_m^{\ell/2-1}} = o(1),
$$

在 $\ell \ge 3$上均匀成立。这就证明了（D.7）。