通过返回概率检验无连通性的扩张

Nikolas Melissaris*

IRIF、法国国家科学研究中心与巴黎西岱大学nikolas@irif.fr

arXiv:2611.32677v1 [cs.DS] 24 Nov 2026

我们研究这样的问题：在能够访问邻居列表的情况下，给定一个图，如何检验其每个连通分量是否都是边扩张图。对于任意固定的 $\alpha > 2/3$，Dinur 和 Goldreich 给出了一种使用 poly$(1/\varepsilon)n^{\alpha}$ 次查询的测试器。我们得到

$$\widetilde{O}_{d,c,\beta}\left(\frac{n^{1/2+\beta}}{\varepsilon}+\frac{\sqrt{n}}{\varepsilon^4}\right)$$

对于任意固定的 $0 < \beta \leq 1$，需要进行查询。该测试器以至少 $2/3$ 的概率接受每个连通分量的边扩张至少为 c 的图，并拒绝距离每个连通分量的边扩张均为 $\Omega(\beta c^2/d)$ 至少 ε 的图。输入不需要满足任何结构性保证。

该测试器会比较两种游走长度下的成对碰撞概率，并从采样终点重合的顶点处开始进行更多游走。这些额外样本控制了具有较大返回概率的顶点所产生的贡献。我们证明，由此得到的条件迫使较长游走在一个仅有少量边离开的诱导扩张图上近似服从均匀分布。我们还证明，对于凯莱图，在无穷多个输入规模下存在 $\Omega(\sqrt{n}/\log n)$ 的下界。

1. 引言

考虑一个由两个不相交扩张图组成的图。分隔它们的割两侧之间没有边，但在任一连通分量内，每个至多包含其一半顶点的集合，都有许多条边连接到其补集，相对于该集合的大小而言如此。Dinur 和 Goldreich [12] 提出了这样的问题：在不限制连通分量的数量或大小的情况下，需要多少次查询才能以这种意义检验扩张性。

固定一个常数度上界 d。所有图均为简单无向图。在有界度模型 [16], 中，一次查询指定一个顶点 $v \in [n]$ 和一个索引 $i \in [d]$。它返回 v 的第 i 个邻居；如果 v 的邻居少于 i 个，则返回 $\perp$。测试器知道 n 和一个邻近参数 $0 < \varepsilon \leq 1$。在 $[n]$ 上的图之间的距离，是在对两个图均有效的邻居排序中，邻居表的 nd 项所不同的最小比例。

对于一个至少含有两个顶点的图 $H = (V,\ E)$，其边扩张为

$$h(H) = \min_{\emptyset \neq S \subseteq V,\ |S| \leq |V|/2} \frac{|E(S, V \setminus S)|}{|S|},$$

*获欧洲研究理事会资助，资助协议编号为 101115790（OBELiSC）。

其中，$E(S,\ V\setminus S)$ 是跨越该割的边的集合。对于单顶点图，我们设 $h(H)=+\infty$。令$\Xi_c(n)$表示最大度至多为 d 的 n-顶点图的集合，其中每个连通分量的边扩张至少为 c，并令 $\Xi_c=\bigcup_{n\geq 1}\Xi_c(n)$。如果要到达$\Xi_c(n)$中的任意图，至少需要更改 εnd 个邻接条目，则称一个图与 $\Xi_c(n)$ 相差 ε。

随机游走揭示了检验这一性质的难点。我们使用惰性随机游走，每一步以概率 $1/(2d)$ 遍历每条关联边，否则停留在原地。一步最多只需一次邻居查询：以概率 $1/2$ 停留在原地，否则就在 $[d]$ 中均匀查询一个索引，并在该索引返回邻居时移动。即使图不是正则图，均匀分布在每个连通分量上仍是平稳分布。在连通扩张图上，经过足够长的随机游走后，分布在所有 n个顶点上近似均匀。此时，两个独立终点重合的概率接近 $1/n$。我们将这一事件称为两两碰撞。连通扩张图测试器利用这些碰撞来检验混合，查询复杂度接近 $\sqrt{n}$ [17, 10, 24, 20]。

如果起点位于大小为 m 的扩张分量中，碰撞概率则接近$1/m$，而测试器并不知道 m。稀疏割还会带来另一个难点：随机游走可能先在一侧扩散，之后才有较大可能穿过割。因此，随机游走探索的集合可能小于其所在的连通分量。即使准确知道碰撞概率，也无法判断端点分布是否接近某个集合上的均匀分布。例如，两点上的均匀分布与分布 $(2/3,1/6,1/6)$ 的碰撞概率都为 $1/2$。

Dinur 和 Goldreich 首先检验端点分布是否接近某个未指定集合上的均匀分布，这项任务称为广义均匀性测试 [4, 11]。他们估计该集合的大小，然后检验扩张性。他们的均匀性检验将两两碰撞与三重碰撞进行比较，后者指三个独立端点重合。对于 m 个点上的均匀分布，要使这些事件的期望次数分别达到常数，需要 $\Theta(\sqrt{m})$ 和 $\Theta(m^{2/3})$ 个样本。三重比较解释了其查询复杂度上界：对于每个固定的 $\alpha>2/3$ [12]，上界为 $\mathrm{poly}(1/\varepsilon)n^{\alpha}$。

一次碰撞还可以确定共同的端点。我们从这些顶点出发进行更多随机游走，并将它们的返回概率与原始起点处的返回概率进行比较。结合在两种游走长度下进行的比较，这便提供了无需估计三重碰撞即可证明扩张性所需的信息。该算法适用于任意有界度图。

1.1. 我们的结果

沿用 Dinur 和 Goldreich [12]，的方法，一个伪测试器以高概率接受属于 $\Xi_c(n)$ 的图，并拒绝与 $\Xi_{c'}(n)$ 相差 ε 的图；对于较小的扩张常数 c'，同样成立。

定理 1.1。存在一个绝对常数 $\kappa>0$，具有以下性质。固定整数 $d\geq 3$、$0<c\leq d$ 和 $0<\beta\leq 1$，并令 $c_\beta=\kappa\beta c^2/d$。对于任意 $n\geq 1$ 和 $0<\varepsilon\leq 1$，存在一个随机伪测试器，使用

$$\widetilde{O}_{d,c,\beta}\left(\frac{n^{1/2+\beta}}{\varepsilon}+\frac{\sqrt{n}}{\varepsilon^4}\right)$$

邻居查询，以至少 $2/3$ 的概率接受 $\Xi_c(n)$ 中的每个图，并以至少 $2/3$ 的概率拒绝距 $\Xi_{c_\beta}(n)$ ε-远的每个图。

~这里 O在 n 和 $1/\varepsilon$ 中省略了对数因子。对于固定的 d、c、β、ε，查询指数为 $1/2+\beta$，其中 $c_\beta>0$ 与 n 无关。这给出了在无约束问题中任意接近 $1/2$ 的指数，正如 Dinur 和 Goldreich 在定理 1.3] 之前的讨论中所提出的 [12, 。参数 c_β 随 β ~，因此该定理留下了这样一个未解决的问题：在拒绝保证中，当扩张阈值固定为正值时，是否存在一个 $O(\sqrt{n})$ 界。

Dinur 和 Goldreich 证明了，在不受限制的有界度图上存在 $\Omega(\sqrt{n})$ 的下界 [12,定理 3.1]。我们证明，即使对于凯莱图也存在下界。对于有限群 Γ 和一个不包含单位元的对称集合 $S=S^{-1}$，图 $\mathrm{Cay}(\Gamma,\ S)$ 的顶点集为 Γ，并且对于$s\in S$，存在从 g 到 gs 的边。集合 S 不一定生成 Γ，因此该图可能是非连通的。

定理 1.2（凯莱图的下界）。 存在一个偶数常数度 $d \geq 6$以及常数$c_+ > c_- > 0$和 $\varepsilon_0 > 0$，使得对于无穷多个 n，任何能够将 $\Xi_{c_+}(n)$ 中的度为 d 的凯莱图与度为 d 且与 $\Xi_{c_-}(n)$ 相差 ε_0 的凯莱图区分开的随机算法，以至少 $2/3$ 的成功概率，均需使用 $\Omega(\sqrt{n}/\log n)$ 次邻居查询。

1.2. 进一步相关工作

广义均匀性下界涉及从一个未知分布中采样的算法 [4, 11]。我们的测试器还会从以观测到的碰撞位置为起点的游走分布中请求样本。Dinur 和 Goldreich 已经从普通端点开始进一步的随机游走 [12, 第 3.2 节，步骤 2]。以概率 p 到达的顶点会对碰撞概率贡献 p^2。我们将这种加权方式与两个返回时间结合起来，以获得结构性证明所需的局部条件。附录 A还给出了一个更简单的测试器，适用于每个连通分量内返回概率均为常数的情形。

局部谱方法将限制在指定区域内的游走的特征值与该区域内部的割联系起来 [8]。Kale、Peres 和 Seshadhri 表明，从大型扩展图的大多数顶点出发的随机游走，在向其附加一个小型任意图后仍然具有实用性；而 Chakraborti 等人表明，向全局均匀分布进行足够广泛的混合，会迫使删除少量顶点后得到一个几乎覆盖全部顶点的诱导扩张图 [19, 6]。在这里，游走可能在未知连通分量的未知真子集上混合。Regev 和 Shinkar 的连续时间情形中也出现了关于返回概率的柯西–施瓦茨界 [25]。

可聚类性为具有多个内部扩张区域的图提供了另一种表述。Czumaj、Peng 和 Sohler 的测试器，以及 Chiplunkar 等人的测试器，关注的是将图分解为数量有界的簇，并控制簇的内部导通率和外部导通率 [9, 7]。谱聚类预言机可对这种分解进行局部访问 [14, 21]。性质 Ξ_c 允许存在任意多个、大小互不相关的连通分量；而远离该性质的输入不一定具有预先规定的可聚类结构。

在零阶情形下，无连通性的扩张将函数取值发生变化的边数，与该函数到那些在每个连通分量上均为常数的函数之间的距离进行比较。[1]

2. 技术与证明概述

设 $G = (V,\ E)$ 为输入图。随机游走可以在尚未探索连通分量其余部分之前，就已在某个区域内扩散开来。测试器必须判断它所观察到的区域，能否经过少量编辑后保留为一个扩张分量。它不能先识别连通分量，再运行连通扩张测试器：仅识别一个大型连通分量就可能需要过多查询。

我们固定一个起始顶点 s。第一次比较考察将游走长度加倍是否会显著改变其碰撞概率。第二次比较碰撞位置处的返回概率与 s处的返回概率，并控制返回概率较大的顶点所作的贡献。这些比较构成测试器背后的主要思想。结构性论证表明，它们迫使较长游走在一个边界边较少的诱导扩张图上近似服从均匀分布。当大多数起点满足这些条件时，我们调整 Dinur 和 Goldreich [12, 的命题 3.4.4–3.4。6] 中的重叠论证，以选出覆盖几乎所有顶点的互不相交的扩展集，从而实现图修复。这些集合仅用于证明可靠性；测试器估计返回概率，但从不构造这些集合。

[1] 在 $\mathbb{F}_2$上，这些逐分量常值函数就是零次余循环。 $\mathbf{1}_S$ 到该空间的距离为 $\sum_C \min\{|S \cap C|, |C \setminus S|\}$。采用未归一化计数时，逐分量扩张性至少为 c 等价于要求函数发生变化的边数至少为该距离的 c 倍。若改为与全局常数比较，则还会要求连通性。

2.1. 两次返回时间与随机游走发生碰撞的顶点

记 P为引言中所述惰性随机游走的转移矩阵，因此， $P^{\ell}(x,\ y)$表示从 x经过 ℓ 步后到达 y 的概率。对于游走长度 t，令 $p_s = P^{2t}(s,\ \cdot)$，并

$$A_s = P^{2t}(s,s) = \|P^t(s,\cdot)\|_2^2, \qquad B_s = P^{4t}(s,s) = \|p_s\|_2^2. \tag{1}$$

这些恒等式源于 P 的对称性。因此， A_s 和 B_s 都是返回概率和碰撞概率：我们通过统计长度分别为 t 和 $2t$ 的独立终点对来估计它们。

如果 s 位于大小为 m 的扩张分量中，那么当 t 足够大时，这两个分布在该连通分量上都接近均匀分布，并且 $A_s,\ B_s \approx 1/m$。尽管测试器不知道 m，它们的比值仍接近于一。这表明可以将 A_s 与 B_s 进行比较。为了了解这种比较能得到什么，利用对称性和柯西–施瓦茨不等式可得

$$p_s(v)^2 = \langle P^t(s,\cdot), P^t(v,\cdot)\rangle^2 \le A_s A_v. \tag{2}$$

如果 A_s 接近 B_s，那么对于满足 A_v 接近 A_s 的每个顶点，这给出了 $p_s(v)$ 约为 B_s 的上界。此外， $B_s = \sum_v p_s(v)^2$ 是当 v 从 p_s 中抽取时 $p_s(v)$ 的平均值。一个接近上界的平均值几乎不可能允许某些值显著低于该上界。这正是证明中近似均匀性的来源。

较大的顶点 A_v 仍然是个问题：式（2）在这些顶点处给出的上界较弱。抽取少量普通端点未必能发现它们的重要性。如果分布中其余的概率质量分散在足够多的点上，那么一个顶点即使只获得百分之一的概率质量，仍可能造成几乎所有的碰撞。因此，我们衡量这些顶点对碰撞概率的贡献。对于足够小的精度参数 $\eta > 0$，定义

$$D_s = \{v : A_v > (1+8\eta)A_s\}, \qquad J_s = \sum_{v\in D_s} p_s(v)^2. \tag{3}$$

称 s良好，如果

$$A_s \le (1+2\eta)B_s, \qquad J_s \le \eta B_s. \tag{4}$$

第二个条件表示，碰撞概率中来自返回概率较大的顶点的部分至多为 η。这些条件用于可靠性证明，而对 G没有任何结构性保证。

这个定义说明了如何检验第二个条件。对于任意一对固定的、相互独立的 $2t$步随机游走，它们都在 v 结束的概率为 $p_s(v)^2$。我们取一批新端点，并检查其中每个至少出现两次的不同顶点。对于每个这样的顶点 v，再用新进行的 t步随机游走估计 A_v。如果 D_s 中的顶点对碰撞概率贡献足够大，那么这批端点中应该会在其中某个顶点处出现碰撞，而进一步的返回概率估计会检测出较大的值。

T批次大小很重要。在获得 B_s 的相对误差为 O（η）的估计值后，取

$$q = \Theta\left(\frac{1}{\sqrt{\eta B_s}}\right)$$

足够多的新端点。相交路径对的期望总数为Θ（$1/\eta$）。如果 $J_s > \eta B_s$，则公共端点位于 D_s 中的路径对的期望数量下界为一个较大的常数。共享同一样本的碰撞指示变量彼此相关，因此仅凭这一期望值无法证明能够检测到碰撞。引理4.1 中的二阶矩界以高概率保证在所选批次大小下检测到碰撞。这就是重复端点为证明提供所需加权的原因。

此外，如果总碰撞计数超过某个足够大的常数乘以 $1/\eta$，我们也会拒绝。否则，只有 $O(1/\eta)$ 个顶点需要进一步进行返回估计。在扩张分量上，超出上限的概率很小，而该上限保证了每次执行的查询成本。第 4节给出了确切的比较、估计和停止规则，并证明违反公式（4）的起点会以高概率被拒绝。完备性利用了更强的事实：在扩张分量上，所有相关的返回概率都接近 $1/m$；我们不要求每个良好起点都被接受。

2.2. 边界边较少的近乎均匀集

现在假设 s 是良好的。记 $p = p_s$ 和 $B = B_s$。在 D_s之外，公式（2）给出 $p(v) \le (1 + O(\eta))B$。因此，高于 B 的值在 D_s之外只能造成很小的额外贡献，而它们在 D_s 内的贡献由 J_s界定。对于从 p中抽取的 X，随机变量 $p(X)/B$ 的均值为一。其超过一的期望超额为 $O(\eta)$，因此其低于一的期望亏缺也相同。等价地，

$$\sum_v p(v)|p(v) - B| = O(\eta B).$$

因此，大部分游走质量必然集中在端点概率接近 B的顶点上。

取满足 $p(v) \ge 3B/4$ 的顶点。上述估计将落在该集合外的概率界定为 $O(\eta)$，并表明其大小乘以 B 为 $1 \pm O(\eta)$。移除 D_s中的顶点后，相对于这些尺度，其大小和概率质量仅改变 $O(\eta)$。于是得到

$$H_s = \{v : p_s(v) \ge 3B_s/4\} \setminus D_s, \qquad |H_s|B_s = 1 \pm O(\eta),$$

并且 p_s 在全变差意义下与 H_s上的均匀分布相差 $O(\eta)$。全变差距离是概率之差绝对值总和的一半。移除 D_s 还确保每个剩余顶点都满足 $A_v \le (1 + 8\eta)A_s$。这一返回界将用于证明扩张性。

仅凭近似均匀性无法界定离开 H_s的边数。两个时刻的比较提供了缺失的图信息。对于我们的惰性随机游走，谱分解给出

$$\frac{1}{2d} \sum_{\{u,v\} \in E} (p_s(u) - p_s(v))^2 \le A_s - B_s \le 2\eta B_s.$$

因此，概率不可能沿着许多边发生剧烈变化。要在固定阈值 $3B_s/4$处使用这一估计，先考虑移除 D_s之前的集合。一条离开该集合的边，要么终点是概率低于 $B_s/2$的顶点，要么终点是概率介于 $B_s/2$ 和 $3B_s/4$之间的顶点。在第一种情况下，概率至少变化 $B_s/4$，因此平方差界限制了此类边的数量。在第二种情况下，近似均匀性估计界定了该概率范围内顶点的数量，而每个顶点至多有 d 条关联边。移除 D_s 中的少数顶点只会增加少量边界边。综上，

$$|E(H_s, V \setminus H_s)| = O(d\eta)|H_s|.$$

引理 5.1 将这些规模、分布和边界估计统一证明。

2.3. 根据返回概率研究内部扩张

我们找到了一个区域，在该区域上随机游走近似均匀分布，且只有少量边通向外部。但其中仍可能包含一个稀疏内部割。例如，两个非连通扩张图的并集没有边离开其并集，但该并集并不是扩展图。我们需要排除 H_s 内部存在这种情况的可能性。

一旦随机游走已经在 m 个顶点组成的连通分量中充分混合，每个顶点的返回概率都约为 $1/m$，因此返回概率之和约为 1。两个彼此分离的连通分量各自都会贡献约 1。这表明，对 H_s 内的返回概率求和，可以检测其中是否包含多个能够使随机游走长时间停留的区域。

将 P 限制在 H_s 中对应的行和列上，得到 $K_s = P[H_s, H_s]$。K_s^ℓ 的一个元素表示：在不离开 H_s 的情况下，经过 ℓ 步到达对应端点的概率。其迹对这些路径的返回概率求和。在长度为 $2t$ 时，每个对角元素至多为不受限的返回概率 A_v，而移除 D_s 对每个这样的 A_v 都进行了界定。因此

$$\mathrm{tr}(K_s^{2t}) \le \sum_{v \in H_s} A_v \le |H_s|(1 + 8\eta)A_s \le 1 + O(\eta).$$

记 $\lambda_1 \ge \lambda_2 \ge \cdots$为 K_s的特征值。由 P的懒惰性以及对主子矩阵的限制可知，它们位于 $[0, 1]$，中。小边界使得 $\lambda_1 \ge 1 - O(\eta)$：从中均匀分布出发的随机游走

一步后离开该集合的概率仅为 $O(\eta)$。因此 λ_1^{2t}至少为 $1-O(t\eta)$。其他所有特征值的幂均非负，所以当 $|H_s|\geq 2$时，

$$\lambda_2^{2t}\leq \operatorname{tr}(K_s^{2t})-\lambda_1^{2t}\leq O(t\eta).$$

取对数得到 $-\log\lambda_2\geq(\log(1/\eta)-\log(Ct))/(2t)$，其中 C为绝对常数，且当 $\lambda_2>0$时成立。我们的参数使得 $\log(Ct)$至多为 $\log(1/\eta)$的一半。由于 $t=\Theta(d^2c^{-2}\log(n/\eta))$，因此，只要使 $\log(1/\eta)/\log(n/\eta)$保持有界于下方，就能得到常数间隙。选择 $\eta\leq n^{-\beta/3}$ 可确保这一点，对于固定的 $\beta>0$，即使 ε 迫使 η 变得更小。分离程度取决于 β，这将决定参数 c_β。

谱界首先控制原图中 H_s 的子集的割。其中一些在此割中计数的边可能完全离开 H_s。因此，我们仍需从这些边界估计过渡到诱导子图内部的扩张性。由于离开 H_s的边很少，引理 5.3 表明，只需移除少量顶点即可。剩余集合 U_s 诱导出一个扩展图，离开它的边很少，并且较长游走在其上仍近似服从均匀分布。引理 5.4 结合了这些结论，对于以下选定的参数，其扩张性为 $\Omega(\beta c^2/d)$。

2.4. 互不相交的扩展集与查询界

这些集合 U_s 不一定是连通分量，不同的起点也可能产生重叠的集合。要将图变成扩展图的并，我们必须选取互不相交的部分，并证明它们覆盖了几乎所有顶点。若在无法再添加互不相交的集合时停止，还不足以证明覆盖：与某个未覆盖起点关联的集合，可能已经与选定的部分相交。

假设我们已经选定了若干互不相交的扩展部分，并考察另一个集合 U。如果它与选定部分的重叠很小，我们可以移除这部分重叠，同时保留另一个扩展图。困难的情况是，我们无法再添加某个集合。我们在每一步都选择最大的可用集合。在 U 首次出现过多重叠之前，每个选定部分都来自一个至少与 U 一样大的集合。

现在将这些部分分别切分为它们与 U 的交集及其补集。跨割边离开 U，因此这些边的总数很少。内部扩张迫使每个割都有一个较小的侧。当总重叠足够大时，某个选定部分必定几乎所有顶点都位于 U内。接着，大小比较便迫使 U 中几乎所有顶点也都位于该部分内。这就是为何要选择最大的可用集合，以及为何要考虑 U 重叠过多的第一个阶段。引理 6.1 遵循 Dinur 和 Goldreich [12, 的重叠方法，将这一论证表述得更为严谨。见命题 3.4.4–3.4。6]。

令 R 为未被覆盖的顶点。前面的论证对每个良好起点 s 都界定了 U_s 在 R 中所占的比例。由于 p_s 在 U_s 上近似均匀分布，这样的起点仅将 $O_{d,\,c,\,\beta}(\eta)$ 的随机游走质量送入 R。令 ρ 为非良好起点所占的比例。在图中均匀选取起点开始的随机游走，其终点也服从均匀分布，因为 P 是对称且随机的。因而对所有起点取平均可得

$$\frac{|R|}{n}=\frac{1}{n}\sum_{s\in V}p_s(R)\leq\rho+O_{d,c,\beta}(\eta).$$

删除离开选定部分的边，以及所有与 R 相 incident 的边。选定的部分仍为诱导扩张图，其余顶点则变为单点集。发生变化的邻居条目数量为 $O_{d,c,\beta}((\rho+\eta)n)$。因此，如果 η 是 ε 的足够小的倍数，则一个 ε-远图必须有 $\Omega_{d,c,\beta}(\varepsilon)$ 比例的起点违反局部条件。以均匀分布抽取 $O_{d,c,\beta}(1/\varepsilon)$ 个起点，以常数概率找到其中一个违反条件的起点。在每个起点进行独立重复，可以降低顶点测试的错误率。第6节证明这些全局保证。

现在我们可以说明准确性和查询参数。选取一个足够小的常数$a_0=a_0(d,\ c,\ \beta)>0$并令

$$\eta=a_0\min\{\varepsilon,n^{-\beta/3}\},\qquad t=\Theta\big(d^2c^{-2}\log(n/\eta)\big)\,. \tag{5}$$

游走长度确保扩张性至少为 c 的每个连通分量中都能实现混合。界 $\eta\leq a_0\varepsilon$使图修复中的误差足够小。界 $\eta\leq a_0n^{-\beta/3}$ 将迹估计

转化为 $1-\lambda_2=\Omega(\beta c^2/d^2)$，即使 ε 小于 $n^{-\beta/3}$时也保持均匀分布。谱论和组合论证给出 $c_\beta=\Omega(\beta c^2/d)$，正如定理 1.1所述。

每个返回估计使用 $\widetilde{O}(\sqrt{n}/\eta^2)$ 个独立终点。由于 $B_s\ge 1/n$，在选择新的碰撞批次时，实现会将任何低于 $1/(2n)$ 的估计替换为 $1/(2n)$。即使估计失败，这也能将批次大小限制在 $O(\sqrt{n}/\eta)$ 以内。碰撞上限仅允许再进行 $O(1/\eta)$ 次返回估计，且每个随机游走步骤至多使用一次邻居查询。因此，一次试验的查询成本为 $\widetilde{O}(\sqrt{n}\,\eta^{-3}t)$。以 $O_{d\text{ 次}c\text{ 次}}$为起点，并在每个起点重复对数次，总量为

$$\widetilde{O}_{d,c,\beta}\left(\frac{\sqrt{n}}{\varepsilon\eta^3}\right)=\widetilde{O}_{d,c,\beta}\left(\frac{n^{1/2+\beta}}{\varepsilon}+\frac{\sqrt{n}}{\varepsilon^4}\right).$$

上面的指数 $1/2$ 是为证明内部扩张性所需的精度付出的代价。相同的 η 选择也决定了这一额外的查询指数和扩张阈值 c_β。

2.5. 需要大量查询的凯莱图

对于下界，即使测试器知道这是一个凯莱图，该图仍必须难以检验。首先取一个二部 Lubotzky–Phillips–Sarnak（LPS）拉马努金凯莱图。对于这些 d 正则图，拉马努金界表明，除 $\pm d$ 以外的每个邻接矩阵特征值，其绝对值都至多为 $2\sqrt{d-1}$。YES 图由固定数量 L 个该基图的不相交副本构成。将这些副本按循环顺序编号。在 NO 图中，沿选定生成元行进也会使副本索引加一，沿其逆元行进则会使副本索引减一。由此得到的图是连通的，并且忽略副本索引后，两个图都映射到同一个 LPS 基图。

连续的一块副本会在 NO 图中形成一个平衡割，其扩张性低于目标值。仅凭这一点还不能证明其距离目标性质很远：删除一个稀疏割可能会产生扩张分量。我们还证明，NO 图仍保有一个固定的正扩张下界。因此，任何只使用少量修改的修复都必须使几乎所有顶点位于同一个连通分量中。该连通分量仍会实质性地覆盖平衡割的两侧，因此要达到目标扩张性，就需要添加许多跨割边。由此可知，尽管修复可能使图不连通，该图与目标性质之间仍存在常数距离。

接下来需要对顶点名称和邻接表顺序进行随机化，以隐藏两个图之间的差异，使其无法被自适应查询算法发现。算法可以从先前见过的顶点继续探索，也可以从任意标签开始新的探索，并比较查询结果。我们将其完整查询记录与一个辅助预言机进行比较；该预言机以森林形式回答，其中每棵树都是 d-正则树。在两个困难图中的任意一个里，只要森林中的两个不同顶点尚未被映射到同一个图顶点，查询结果就可以与该预言机的回答一致。

在将隐藏的群生成元分配给各条边之前，森林就已生成。给定这片森林及其查询记录，同一棵树中任意两个顶点之间的路径，其生成元序列均匀随机，且不含相邻的互逆生成元。因此，它在 LPS 图上的投影是一条非回溯游走，也就是说，这条游走不会立即沿着刚走过的边原路返回。若要使任一难图区中的两个顶点重合，这条投影游走就必须返回起点。

不存在短环，因而短路径不会发生这种返回。自适应测试器也可以构造更长的路径，因此我们需要一个适用于任意长度的返回界。算术计数控制中等长度的情形，而 Ramanujan 特征值界控制较长路径的情形。二者结合后，得到一个关于路径长度一致成立的返回概率 $O(\log n/n)$。不同树中的顶点以概率$O(1/n)$重合。一次 Q 次查询的探索只会暴露 $O(Q)$ 个顶点，因此对所有顶点对应用并集界，可得记录距离 $O(Q^2\log n/n)$。当 Q 是一个足够小的常数乘以 $\sqrt{n}/\log n$ 时，这两个图无法以概率 $2/3$ 区分。

第 7 节介绍构造、远离性论证和一致返回估计，然后将它们应用于自适应记录。具体而言，本节还包含导致对数损失的算术估计。

3. 预备知识

在整个上界证明中， d、 c、 β均与定理 1.1中的定义相同，而 $\eta = a_0$最小$\{\varepsilon, n^{-\beta/3}\}$与式（5）中的定义相同。我们将游走长度记为

$$t = \lceil 16 d^2 c^{-2} \log(n/\eta) \rceil . \tag{6}$$

我们选取 $a_0 = a_0(d, c, \beta) > 0$足够小，使其满足下文的结构界和修复界。我们用 $C(s)$表示包含 s的连通分量，用 $\partial W = E(W, V \setminus W)$表示顶点集 W的边界边，并在 W非空时用 u_W表示其均匀分布。对于有限集上的分布 p， q，$d_{\mathrm{TV}}(p, q) = \frac{1}{2}\sum_v |p(v) - q(v)|$。

3.1. 返回概率

引言中所述随机游走的转移矩阵为

$$P(x,y) = \begin{cases} 1/(2d), & \{x,y\} \in E, \\ 1 - \deg(x)/(2d), & x = y, \\ 0, & \text{otherwise.} \end{cases}$$

等价地， $P = I - L_G/(2d)$，其中 L_G是图拉普拉斯矩阵。其特征值 L_G位于 $[0, 2d]$中，因此 P的所有特征值均位于 $[0, 1]$中。该矩阵是对称的且为随机矩阵，并且其在每个连通分量上的限制都是双随机的：每一行和每一列的元素之和均为一。对于式（1）中的量，由柯西–施瓦茨不等式和谱分解可得

$$1/n \le B_s \le A_s \le 1. \tag{7}$$

引理 3.1（混合）。 设 C 是大小为 m 且满足 $h(C) \ge c$ 的一个连通分量。对于每个 $x \in C$ 和整数 $\ell \ge 0$,

$$\|P^\ell(x,\cdot) - u_C\|_2^2 \le \exp(-c^2\ell/(4d^2)).$$

特别地，在式 t 中的选择 （6），当 a_0 充分小时，可确保

$$1/m \le A_v, B_v \le (1 + \eta/10)/m \qquad (v \in C). \tag{8}$$

证明。 如果 $m = 1$，则断言显然成立。否则，从一个非空 $S \subseteq C$， $|S| \le m/2$ 中均匀分布地出发的随机游走在一步内离开 S 的概率为 $|\partial S|/(2d|S|) \ge c/(2d)$。切格不等式给出至少为 $c^2/(8d^2)$ 的谱间隙；例如参见 [18]。将点质量 x 在标准正交特征基下展开，即可证明第一个界。与平稳向量正交给出

$$A_v = 1/m + \|P^t(v,\cdot) - u_C\|_2^2, \qquad B_v = 1/m + \|P^{2t}(v,\cdot) - u_C\|_2^2.$$

第一个误差至多为 $(\eta/n)^4 \le \eta/(10n) \le \eta/(10m)$，第二个误差不超过此值。 □

4. 测试器

我们现在具体说明第2.1节中描述的顶点测试。我们首先记录用于选择其样本并在 D_s中检测碰撞的碰撞估计。它们对任意分布成立。

引理 4.1 （碰撞估计）。 设 r 为至多 n 个点上的一个分布， $B = \sum_v r(v)^2$，与 $q \ge 2$。由 q 个独立样本定义 N_v 为 v 的重数以及 $S = \sum_v \binom{N_v}{2}$。

(i) 给定 $0 < u < 1$ 与 $0 < \zeta \le 1/3$，可以将 B 估计到相对误差 u，失败概率至多为 ζ，使用

$$O(\sqrt{n}\, u^{-2} \log(1/\zeta))$$

样本。

（*ii*）对于任意固定集合 D，令 $J=\sum_{v\in D} r(v)^2$。对于每个 $K\succ 0$，如果 $q\sqrt{J}\ge K$，则没有重复样本落在 D 中的概率为 $O(K^{-1}+K^{-2})$。

证明。 对于 $S_D=\sum_{v\in D}\binom{N_v}{2}$，不相交的样本对给出相互独立的指示变量。每组三个样本索引对应三对彼此重叠的指示变量，其联合事件是三个样本在 D 中全部相同。因此，展开方差可得

$$\mathbb{E}S_D=\binom{q}{2}J,\qquad \operatorname{Var}S_D=\binom{q}{2}(J-J^2)+6\binom{q}{3}\left(\sum_{v\in D}r(v)^3-J^2\right).$$

因此，利用 $\sum_{v\in D} r(v)^3\le J^{3/2}$，

$$\operatorname{Var}S_D\le \tfrac{1}{2}q^2J+q^3J^{3/2}.$$

当 D 等于整个定义域时，用 $\binom{q}{2}$ 归一化可得无偏估计 $\widehat{B}$，且

$$\operatorname{Var}\widehat{B}=O(B/q^2+B^{3/2}/q).$$

由于 $B\ge 1/n$，取 $q=\lceil K\sqrt{n}/u^2\rceil$ 可使相对误差大于 u 的概率至多为$O(K^{-2}+K^{-1})$，这是由切比雪夫不等式得到的。选择足够大的绝对常数 K，使该概率至多为 $1/10$。取奇数个独立批次的中位数，并使用 $O(\log(1/\zeta))$ 个批次，即可证明 (i)。对于 (ii)，由切比雪夫不等式和 $\binom{q}{2}\ge q^2/4$ 可得

$$\Pr[S_D=0]\le\frac{\operatorname{Var}S_D}{(\mathbb{E}S_D)^2}=O\left(\frac{1}{q^2J}+\frac{1}{q\sqrt{J}}\right).$$

□

固定足够大的绝对常数 K 和 C_{cap}。现在给出在起始顶点 s处进行的一次检验试验。所有估计和随机游走在给定先前查询记录的条件下，都使用新的独立随机性。它们的起始顶点和批次大小可以依赖于该查询记录。

1. 分别将 A_s 和 B_s 估计到相对精度 $u=\eta/100$，每项的失败概率至多为 $1/100$。这些估计分别使用长度为 t步和 $2t$步的随机游走的端点。若 $A_s>(1+\eta)B_s$，则拒绝。

2. 令 $\widetilde{B}_s=\max\{B_s,1/(2n)\}$，并抽取一个新的批次，其中包含

$$q=\left\lceil K/\sqrt{\eta\widetilde{B}_s}\right\rceil$$

个来自 s的独立 $2t$ 步随机游走端点。根据它们的重数计算 $S=\sum_v\binom{N_v}{2}$。若 $S>M=\lceil C_{\mathrm{cap}}K^2/\eta\rceil$，则拒绝。

3. 对每个满足 $N_y\ge 2$的不同顶点 y，使用从 y出发的新 t 步随机游走，将 A_y 估计到相对精度 u，失败概率至多为 $1/(100M)$。若任何此类估计满足

$$\widehat{A}_y>(1+4\eta)\widehat{A}_s.$$

否则接受。

只要到达最后一步，其中的顶点数至多为 M，因为每个不同的重复端点至少会产生一次碰撞。重数可以通过排序或哈希计算；不会枚举样本对。

完整的测试器使用以下过程，其中常数 $C_{\mathrm{start}}=C_{\mathrm{start}}(d,\ c,\ \beta)$ 足够大。

1. 从 $[n]$中独立且均匀分布地采样 $N_{\mathrm{start}}=\lceil C_{\mathrm{start}}/\varepsilon\rceil$ 个顶点。

2. 在每个采样顶点处，对单顶点测试进行奇数次 $\Theta(\log(100N_{\mathrm{start}}))$ 独立试验，并采用多数决策。试验次数中的绝对常数足够大。

3. 如果任何多数决策拒绝，则拒绝；否则接受。

第 6 节证明了选择 C_{start} 的合理性，并证明了完整测试器的保证。

命题 4.2 （单顶点测试）。 每次试验在每次执行中都会使用 $\widetilde{O}(\sqrt{n}\,\eta^{-3}t)$ 次邻居查询。对于违反局部条件 (4) 的顶点，它以至少 $9/10$的概率拒绝。如果其连通分量满足式 (8)，则它以至少 $9/10$的概率接受。

证明。 除概率为 $2/100$ 的情况外，前两个估计同时准确。在此事件下，$\widehat{\ }$ $\widehat{\ }$下限截断不会改变 B_s，且 $B_s/2 \le B_s \le 2B_s$。如果 $A_s > (1+2\eta)B_s$，相对精度$u = \eta/100$会迫使其在步骤 1 中拒绝。

反之，假设 $J_s > \eta B_s$。那么 $q\sqrt{J_s} \ge \overline{K}/\sqrt{2}$。根据引理 4.1(ii)，选择足够大的 K 可确保新批次中在 D_s 出现重复顶点，除非发生概率为 $1/100$的例外。如果超出上限，则已发生拒绝。否则会检查每个碰撞位置；通过对批次进行条件化，并对至多 M 个位置应用并集界，除概率为 $1/100$的情况外，所有返回估计都准确。对于 $y \in D_s$，准确的估计意味着

$$\widehat{A}_y/\widehat{A}_s \ge \frac{1-u}{1+u}(1+8\eta) > 1+4\eta,$$

因此第 3 步拒绝。这证明了拒绝保证。

在（8）条件下，准确的初始估计会使第 1 步接受，而每个受检顶点处的准确估计会使第 3 步接受。接下来需要界定 $S > M$ 的概率。在初始估计准确的条件下，$q \le 2K/\sqrt{\eta B_s}$，因此

$$\mathbb{E}S = \binom{q}{2} B_s \le 2K^2/\eta, \qquad \Pr[S > M] \le 2/C_{\mathrm{cap}} \le 1/100.$$

结合估计失败的概率界，这证明了接受概率大于 $9/10$。对于最坏情况下的查询复杂度，下限截断保证了 $q = O(\sqrt{n/\eta})$。前两个估计需要 $\widetilde{O}(\sqrt{n}\eta^{-2}t)$ 次查询。至多还需进行 $\overline{M} = O(\eta^{-1})$ 次估计。每次估计使用

$$O(\sqrt{n}\eta^{-2}t\log(100M))$$

次查询。因此，最坏情况下的查询复杂度上界成立。 □

5. 根据返回概率扩展集合

为了获得良好起点，我们首先找出一个集合，使得较长游走在其上近似服从均匀分布，并且离开该集合的边很少。其顶点的返回界还控制着受限游走的迹。随后，我们利用这一迹界证明，在移除少量顶点后仍具有内部扩张性。

5.1. 近似均匀性、边界大小和返回概率

引理 5.1 （具有小边界的近乎均匀集）。 存在一个绝对常数 C，使得对于所有充分小的 $\eta > 0$，每个良好起点 s都有一个非空集合

$$H_s = \{v : p_s(v) \ge 3B_s/4\} \setminus D_s$$

满足

$$|H_s|B_s = 1 \pm C\eta, \qquad d_{\mathrm{TV}}(p_s, u_{H_s}) \le C\eta, \qquad |E(H_s, V\setminus H_s)| \le Cd\eta|H_s|,$$

以及，借助 $K_s = P[H_s, H_s]$，

$$\mathrm{tr}(K_s^{2t}) \le 1 + C\eta. \tag{9}$$

证明。令 $p = p_s$、$A = A_s$ 和 $B = B_s$。对于 $v \notin D_s$，式（2）和式（4）给出

$$p(v) \le \sqrt{1+8\eta}\, A \le \sqrt{1+8\eta}(1+2\eta)B = (1+O(\eta))B.$$

对于 $X \sim p$，设 $Z = p(X)/B$，并记 $x_+ = \max\{x, 0\}$。于是 $\mathbb{E}Z = 1$，且

$$\mathbb{E}[(Z-1)_+] \le O(\eta) + \frac{1}{B}\sum_{v \in D_s} p(v)^2 = O(\eta).$$

期望亏损等于期望盈余。因此

$$\sum_v p(v)|p(v) - B| = B\mathbb{E}|Z-1| = O(\eta B). \tag{10}$$

令 $H_0 = \{v : p(v) \ge 3B/4\}$。亏损界意味着 $p(H_0^c) = O(\eta)$。此外，

$$\sum_{v \in H_0} |p(v) - B| \le \frac{4}{3B}\sum_v p(v)|p(v) - B| = O(\eta).$$

因此，$B|H_0| = 1 \pm O(\eta)$ 且 $d_{\mathrm{TV}}(p, u_{H_0}) = O(\eta)$。在 H_0 上，不等式 $p(v)^2 \ge 9B^2/16$和 $p(v) \le 4p(v)^2/(3B)$，连同 $J_s \le \eta B$，给出

$$|H_0 \cap D_s| = O(\eta/B), \qquad p(H_0 \cap D_s) = O(\eta).$$

移除这些顶点后，剩下 H_s，其中 $B|H_s| = 1 \pm O(\eta)$ 且 $d_{\mathrm{TV}}(p, u_{H_s}) = O(\eta)$。特别地，对于充分小的 η，H_s 非空。

对于边界，设 $(v_j)_j$为 P 的标准正交特征基，其中 $Pv_j = \lambda_j v_j$，并令 e_s为 s处的单位向量。于是

$$\langle p, (I-P)p\rangle = \sum_j \lambda_j^{4t}(1-\lambda_j)|\langle e_s, v_j\rangle|^2 \le \sum_j \lambda_j^{2t}(1-\lambda_j^{2t})|\langle e_s, v_j\rangle|^2 = A - B \le 2\eta B.$$

对于我们的随机游走，$\sum_{\{u,v\}\in E}(p(u) - p(v))^2 = 2d\langle p, (I-P)p\rangle = O(d\eta B)$。从 H_0到{v 的每条边：$p(v) < B/2$} 至少贡献 $B^2/16$，因此共有 $O(d\eta/B)$条这样的边。中间集合 $\{B/2 \le p < 3B/4\}$ 的质量为 $O(\eta)$，大小为 $O(\eta/B)$，因此至多有 $O(d\eta/B)$条边终止于此。移除 $H_0 \cap D_s$ 至多增加 $d|H_0 \cap D_s| = O(d\eta/B)$条边界边。由于 $B|H_s| = 1 \pm O(\eta)$，这就证明了$|\partial H_s| = O(d\eta)|H_s|$。

最后，每个 $v \in H_s$都有 $A_v \le (1+8\eta)A_s$。由 K_s^{2t}计数的随机游走仍停留在 H_s中，因此（$K_s^{2t}(v, v) \le P^{2t}(v, v) = A_v$。因此

$$\mathrm{tr}(K_s^{2t}) \le |H_s|(1+8\eta)A_s \le (1+O(\eta))(1+8\eta)(1+2\eta) = 1 + O(\eta).$$

这就证明了迹界，无需界定 H_s 之外的返回概率。本证明中所有隐含的常数都是绝对常数。

□

5.2. 从谱界到诱导扩张图

引理 5.2 (G中的子集扩张性)。设 H 的大小为 $m \ge 2$，令 $K = P[H, H]$，并设

$$b = \frac{|E(H, V \setminus H)|}{2dm}.$$

*Sup*设 $g > 0$, $\lambda_2(K) \le 1 - g$，且 $b \le g/100$。那么每个满足 $|S| \le m/2$ 的 $S \subseteq H$ 都满足 *es*

$$|E(S, V \setminus S)| \ge \frac{dg}{2}|S|.$$

证明。令 v_1 为 K 的单位最大特征向量，并取其非负。设 $u = m^{-1/2}\mathbf{1}_H$。由于 $u^\top K u = 1-b$，而所有其他特征值至多为 $1-g$，我们有 $1-b \le 1-g(1-|\langle u,\ v_1\rangle|^2)$，因此 $1-|\langle u,\ v_1\rangle|^2 \le b/g$。对于满足 $r = |S|/m \le 1/2$ 的 $S \subseteq H$，令 $f = |S|^{-1/2}\mathbf{1}_S$。将 $v_1 = \alpha u + w$ 写成 $w \perp u$。由前述界可得 $\|w\| \le \sqrt{b/g}$，而 $\langle f,\ u\rangle = \sqrt{r}$。因此，由柯西–施瓦茨不等式可得

$$|\langle f, v_1\rangle| \le \sqrt{r} + \sqrt{b/g}.$$

在 $b/g \le 1/100$ 条件下，右侧平方不超过 $3/4$。由谱分解可得

$$\mathbf{1}_S^\top (I-K)\mathbf{1}_S \ge \frac{g}{4}|S|.$$

对于惰性随机游走，$\mathbf{1}_S^\top (I-K)\mathbf{1}_S = |E(S,\ V\setminus S)|/(2d)$，这就证明了该结论。 □

前面的引理统计了原图中所有离开 S 的边。其中一些边可能完全离开 H，因此该界还不能证明 $G[H]$ 内部具有扩张性。即使集合 S 在 G 中的边界很大，它也可能只有很少的边连向 H 的其余部分。下一个引理利用 H 的小边界证明，移除少量顶点后会留下一个诱导扩张图。

引理 5.3 (修剪为诱导扩张图)。设 G 的最大度不超过 d，$H \subseteq V(G)$ 的大小为 $m \ge 1$，并固定 $0 < h \le d$。假设每个满足 $|S| \le m/2$ 的 $S \subseteq H$ 都满足

$$|E(S, V\setminus S)| \ge h|S|,$$

并且

$$|E(H, V\setminus H)| \le \tau m, \qquad \tau \le \frac{h^2}{100d}.$$

那么存在 $U \subseteq H$ 使得

$$|H\setminus U| \le \frac{2\tau}{h}m, \qquad h(G[U]) \ge h/4,$$

and

$$|E(U, V\setminus U)| \le C\frac{d\tau}{h}|U|$$

其中 C是一个绝对常数。

证明。令 $\alpha = h/4$。在所有满足 $|R| \le m/2$ 的 $R \subseteq H$ 中，选取一个使 $F(R) = \alpha|R| - |E(R,\ H\setminus R)|$ 最大的。由于 $F(\emptyset) = 0$，该最大化者满足 $F(R) \ge 0$。将假设的边界下界应用于 R，得到 $h|R| \le |E(R,\ H\setminus R)| + |E(R,\ V\setminus H)|$，因此 $0 \le F(R) \le \tau m - (h-\alpha)|R|$。于是 $|R| \le (4\tau/3h)m \le (2\tau/h)m$。设 $U = H\setminus R$。

反设存在某个满足 $|S| \le |U|/2$ 的 $S \subseteq U$，使得 $|E(S,\ U\setminus S)| < \alpha|S|$。若$|R| + |S| \le m/2$，则

$$\begin{aligned} F(R\cup S) - F(R) &= \alpha|S| - |E(S, U\setminus S)| + |E(S,R)| \\ &> 0, \end{aligned}$$

这与极大性矛盾。因此 $|R| + |S| > m/2$。下界 $\tau \le h^2/(100d)$ 和 $h \le d$ 蕴含$|R| \le m/50$，所以 $|S| > m/3$。现在将相同的边界下界应用于 S 可得

$$(h-\alpha)|S| < |E(S,R)| + |E(S, V\setminus H)| \le d|R| + \tau m,$$

而右侧至多为

$$\left(\frac{2d\tau}{h} + \tau\right)m \le \frac{3h}{100}m,$$

左侧大于 $hm/4$。这一矛盾证明了 $h(G[U]) \geq h/4$。最后，

$$|E(U, V \setminus U)| \leq |E(H, V \setminus H)| + d|R| \leq \left(\tau + \frac{2d\tau}{h}\right) m.$$

由于 $|U| = (1 - O(\tau/h))m$，调整常数后即可得到最终界。 □

现在可以得到所需的扩展集。式 $\eta \leq a_0 n^{-\beta/3}$ （5）中的界将迹估计转化为在所选游走长度 t下的常数谱间隙。

引理 5.4 （从良好起点得到扩展集）。 对于充分小的绝对常数 $\kappa > 0$，令$g = 32\kappa\beta c^2/d^2$ 并令 $h_* = dg/8$。若 $a_0 = a_0(d,\ c,\ \beta) > 0$ 在式 （5）中充分小，则每个良好起点 s都有一个非空集合 $U_s \subseteq C(s)$ 满足

$$h(G[U_s]) \geq h_*, \qquad |\partial U_s| \leq C_1 \eta |U_s|, \qquad d_{\mathrm{TV}}(p_s, u_{U_s}) \leq C_1 \eta,$$

其中常数 $C_1 > 0$ 仅取决于 d、 c、 β。

证明。 应用引理5.1。若 $|H_s| = 1$，取 $U_s = H_s$ 并使用单元素约定。否则令$K = P[H_s, H_s]$ 并令 $b = |\partial H_s|/(2d|H_s|) = O(\eta)$ 。其特征值满足 $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq 0$。归一化全一向量的瑞利商为 $1 - b$，因此 $\lambda_1 \geq 1 - b$。对于充分小的 a_0，有 $b < 1$。由伯努利不等式和迹界可得，对于某个绝对常数 C_2，

$$\lambda_2^{2t} \leq 1 + O(\eta) - (1 - b)^{2t} \leq C_2 t\eta.$$

为了均匀地提取谱间隙，需要使 n 和 ε，令 $L = \log(1/\eta)$ 以及 $x = \log n$。由于 $a_0 \leq 1$ 和$\eta \leq a_0 n^{-\beta/3}$，我们有 $L \geq (\beta/3)x$。对于 $L \geq 1$，式（6）还给出 $t \leq D(x + L)$ ，其中常数$D = O(d^2/c^2)$，隐含绝对常数。因此

$$\log(C_2 t) \leq \log\big(C_2 D(1 + 3/\beta)\big) + \log L.$$

选择 a_0 足够小，使得对于每个 $L \geq \log(1/a_0)$，都有 $\log(C_2 D(1 + 3/\beta)) \leq L/4$ 以及 $\log L \leq L/4$ 。这些条件仅取决于 d、 c、 β。它们意味着 $\log(1/(C_2 t\eta)) \geq L/2$。如果 $\lambda_2 > 0$，取对数可得

$$-\log \lambda_2 \geq \frac{L}{4t} = \Omega(\beta c^2/d^2).$$

这里 $L/(x + L) \geq \beta/(3 + \beta) \geq \beta/4$。选择绝对常数 κ 足够小，使得最后一个界至少为 $2g$ 且 $g \leq 1/2$。于是 $1 - e^{-2g} \geq g$，因此 $1 - \lambda_2 \geq g$。如果 $\lambda_2 = 0$，则同样的结论可立即得到。

进一步减小 a_0 ，使得 $b \leq g/100$且

$$|\partial H_s|/|H_s| \leq h_0^2/(100d), \qquad h_0 = dg/2.$$

这是可行的，因为 $b = O(\eta)$ 且 $|\partial H_s|/|H_s| = O(d\eta)$。引理 5.2 给出 $|\partial S| \geq h_0|S|$ 对于每个$S \subseteq H_s$ 满足 $|S| \leq |H_s|/2$。引理 5.3 随后给出一个诱导的 $h_0/4$-扩展图 $U_s \subseteq H_s$ ，其满足

$$|H_s \setminus U_s| = O(\eta/g)|H_s|, \qquad |\partial U_s| = O(d\eta/g)|U_s|.$$

总变差界可由下式得到

$$d_{\mathrm{TV}}(p_s, u_{U_s}) \leq d_{\mathrm{TV}}(p_s, u_{H_s}) + |H_s \setminus U_s|/|H_s| = O(\eta + \eta/g).$$

这些界给出了一个共同常数 $C_1 = C_1(d,\ c,\ \beta)$，也涵盖单元素情形。对 a_0 的所有限制仅取决于 d, c, β，并且对每个 $n \geq 1$ 和 $0 < \varepsilon \leq 1$ 都成立。此后即使减小 a_0，这些常数也不会改变。

□

6. 集合合并并证明可靠性

从良好起点得到的集合，在移除较小的重叠后，必须覆盖几乎所有顶点。下一个引理将第 2.4 节中的选择过程形式化。它将修复成本限制为与局部条件失效的起点所占比例成线性关系，而无需在这些起点处要求存在集合。证明沿用 Dinur 和 Goldreich 的重叠方法 [12, 命题 3.4.4–3.4。6],使用最大的可用集合，以及另一个集合出现过多重叠的第一个阶段。

引理 6.1 (扩展集合合并)。 固定 d 和 $0 < h \le d$。存在 τ_0、$C > 0$，它们仅取决于d、h，使得以下结论成立。设 C_0 是一个连通分量，Q 是其上的一个双随机矩阵，且 $0 \le \rho \le 1$。假设至少有 $1-\rho$ 比例的顶点 s 具有非空集合 $U_s \subseteq C_0$，满足

$$h(G[U_s]) \ge h, \quad |\partial U_s| \le \tau |U_s|, \quad d_{\mathrm{TV}}(Q(s,\cdot), u_{U_s}) \le \tau, \quad 0 < \tau \le \tau_0.$$

通过至多 $C(\rho+\tau)|C_0|$次邻居条目变更，图 C_0 可以通过边删除变为图 $\Xi_{h/4}$。

证明。令 $a = 8/h$，并取足够小的 $\tau_0 > 0$，使其仅取决于 d、h。贪心地构造互不相交的集合 V_i。设 W为它们当前的并集，称 U_s为可容许的，当且仅当 $|U_s \cap W| \le a\tau|U_s|$。选取一个大小最大的可容许集合，记为 U_i，并对 $H_i = U_i \setminus W$应用引理 5.3。事实上，H_i 的每个大小至多为 $|H_i|/2$ 的子集在 G 中的边界边至少为其大小的 h 倍，$|H_i| \ge |U_i|/2$，且$|\partial H_i| \le (1 + da)\tau|U_i|$。当 τ_0足够小时，修剪假设成立，并由此得到一个非空的$V_i \subseteq H_i$，满足

$$h(G[V_i]) \ge h/4, \quad |U_i \setminus V_i| \le b\tau|U_i|, \quad |\partial V_i| \le K_0\tau|V_i|.$$

这里的 b，$K_0 > 0$ 是仅取决于 d，h的常数。大小估计使用了重叠损失和修剪损失 $2|\partial H_i|/h$；边界估计则由 $|\partial V_i| \le |\partial H_i| + d|H_i \setminus V_i|$ 和$|V_i| \ge |U_i|/2$推出。选择 τ_0 足够小，使得 $a\tau_0 \le 1/2$、$b\tau_0 \le 1/2$成立，并且修剪假设成立。继续进行，直到不再存在可容许集合，并令 R 表示未覆盖的顶点。每次迭代都会添加一个非空的不相交集合，因此最多可以进行 $|C_0|$次迭代。

对于给定集合中的任意一个 U，令 $W_j = \bigcup_{i \le j} V_i$ 和 $W_0 = \emptyset$。令 j 为满足 $|U \cap W_j| > a\tau|U|$的第一个指标。因此，在每次选择 $i \le j$之前，U都立即是可容许的，所以由极大性可得 $|U_i| \ge |U|$。由于这些 V_i 彼此不相交，

$$\frac{h}{4} \sum_{i \le j} \min\{|V_i \cap U|, |V_i \setminus U|\} \le |\partial U| \le \tau |U|.$$

但是 $|U \cap \bigcup_{i \le j} V_i| > 8\tau|U|/h$。因此，某个 $i \le j$具有 $|V_i \cap U| > |V_i \setminus U|$。它在 U中的补集满足

$$|U \setminus V_i| \le |U| - |V_i| + |V_i \setminus U| \le (b + 4/h)\tau|U|.$$

由此可知，对于每个使 U_s有定义的起始 s，都有 $Q(s, R) \le (1 + b + 4/h)\tau$。对均匀分布的起始点取平均，并利用双随机性，得到 $|R|/|C_0| \le \rho + (1 + b + 4/h)\tau$。删除离开选定集合 V_i 的所有边，以及所有与 R关联的边。邻居条目中发生改变的数量至多为 $2K_0\tau|C_0| + 2d|R|$。这就是 $O_{d,\,h}((\rho+\tau)|C_0|)$，而剩余的非平凡连通分量的扩张率至少为 $h/4$。这些集合和修复仅用于可靠性证明；测试器不会计算它们。 □

6.1. 完备性、可靠性和查询复杂度

定理 *1.1* 的证明。 固定一个足够小的绝对 κ，如引理 5.4 中所述，并令 $g = 32\kappa\beta c^2/d^2$ 以及 $h_* = dg/8$。则 $h_* \le$ 最小值$\{d, c\}$，并且 $h_*/4 = c_\beta$ 如所述。令 $C_{\mathrm{comb}} \ge d$ 以及 τ_0 由引理 6.1 结合 $h = h_*$给出。除对 a_0施加结构界 a_0外，再要求

$$C_1\eta \le \tau_0, \qquad C_{\mathrm{comb}} C_1 \eta \le d\varepsilon/4.$$

只需进一步减小 a_0，使得 $a_0 \le$ 最小值$\{\tau_0/C_1, d/(4C_{\mathrm{comb}}C_1)\}$；这一选择与 n、ε 无关。

令 ρ 表示所有顶点中不良顶点所占的比例。对每个连通分量分别应用引理 6.1，取 $Q = P^{2t}$ 和 $\tau = C_1\eta$。将修复成本相加可知，删除边后得到的图属于 $\Xi_{c_\beta}(n)$，且至多有

$$C_{\mathrm{comb}}(\rho + C_1\eta)n$$

个邻居条目发生改变。因此，一个 ε-远的图必须满足

$$\rho \ge d\varepsilon/(2C_{\mathrm{comb}}),$$

因为更小的值会导致修复成本低于 $3d\varepsilon n/4$。

在第 C_{start} 节的测试器中选择足够大的 4，使得 $N_{\mathrm{start}} = \lceil C_{\mathrm{start}}/\varepsilon \rceil$ 个独立的均匀起点以所示密度遗漏所有顶点的概率至多为 $1/6$。命题 4.2 和多数放大保证，违反公式（4）的起点以至少 $1 - 1/(100N_{\mathrm{start}})$ 的概率被拒绝；而其连通分量满足公式（8）的起点以至少相同的概率被接受，只要试验次数中的绝对常数足够大。

在一个远离图中，违反公式（4）的顶点被采样到的概率至少为 $5/6$。放大的检验拒绝此类顶点的概率至多为 $1/(100N_{\mathrm{start}})$，因此总体拒绝概率大于 $2/3$。

对于完备性，引理3.1 在图的每个连通分量中建立了公式（8），其中图属于 $\Xi_c(n)$。命题4.2、放大和并集界给出了至少为 $99/100$的总体接受概率。

最坏情况下的查询复杂度上界为

$$\widetilde{O}_{d,c,\beta}\left(\frac{\sqrt{n}}{\varepsilon\eta^3}\right) = \widetilde{O}_{d,c,\beta}\left(\frac{n^{1/2+\beta}}{\varepsilon} + \frac{\sqrt{n}}{\varepsilon^4}\right),$$

使用 $\eta = a_0 \min\{\varepsilon, n^{-\beta/3}\}$。所有游走长度和错误放大因子关于 n 和$1/\varepsilon$均为对数级，其常数取决于 d、c、β。所有参数上界对每个 $n \ge 1$ 均成立。 □

注 6.2（固定邻居表距离）。 第*1* 节中的距离对两张图的邻居排序取最小值。如果输入表的顺序固定不变，则删除后压缩列表可能导致额外变化。从一个列表中移除 r 个邻居，最多可以通过修改 $2r$ 个条目来实现：用被保留前缀中的存活邻居替换已删除的邻居，这些存活邻居来自被丢弃的后缀，然后清空后缀。因此，删除 q 条边，在固定表约定下至多需要修改 $4q$ 个条目，而在我们的约定下为$2q$ 。将测试器的邻近参数设为 $\varepsilon/2$，即可在固定表距离下获得 ε 的可靠性。修复证明仅使用删除操作，因此这一倍增调整已经足够；所显示的渐近查询复杂度上界和c_β 均不变。

7. Cayley 图的一个下界

我们利用两个图证明定理1.2 这两个图构造在 L 个相同二部 LPS 图的顶点集副本上。YES 实例由 L 个互不连通的基图副本组成。在 NO 实例中，每个选定的生成元还会使循环坐标前进一位，而它的逆元则使该坐标后退一位。在 NO 图中返回起始顶点的生成元序列，投影到 LPS 基图后也会返回起始顶点。

Bencs 等人研究了正则高围长扩张图上一条非回溯轨迹的自相交 [5]。预言机测试器则可以分支并比较不同的探索过程，因此我们必须对其整个自适应查询记录中的碰撞数加以限制。

7.1. LPS 基图与无环坐标

固定一个满足 $p \equiv 1 \pmod 4$ 的素数，并令 $d = p+1$ 和 $r = d/2$。LPS 构造给出了一族无限的二部 d-正则 Ramanujan 图凯莱图 $X_q = \mathrm{Cay}(G_q, S_q)$，其中 $G_q = \mathrm{PGL}_2(\mathbb{F}_q)$ 是射影一般线性群，[2] $q \equiv 1 \pmod 4$ 是素数，$(p/q) = -1$，即 p 是模 q 的二次非剩余，并且 $|G_q| = m_q = q(q^2-1)$ [22]。我们采用 Alon 和 Bachmat [2] 回顾的自由群描述。存在一个自由群 $F = \langle a_1, \ldots\ldots, a_r \rangle$[3] 以及一个满射 $\pi_q \colon F \to G_q$，满足$\pi_q(uv) = \pi_q(u)\pi_q(v)$。对称生成集为

$$S_q = \{\pi_q(a_1)^{\pm 1}, \ldots, \pi_q(a_r)^{\pm 1}\}.$$

$a_1^{\pm 1}$、……、$a_r^{\pm 1}$中的一个词，若不含相邻逆元对 $a_i a_i^{-1}$或 $a_i^{-1} a_i$，则称其为约化的。在凯莱图的解释下，长度为 k 的约化词恰好对应于从固定顶点出发的 k 步非回溯游走。

记 $N_q = \ker \pi_q$为其像在 π_q <style id='10'}下为单位元的词。图 X_q 的围长至少为$(4/3)\log_p m_q$，且除 $\pm d$ 之外的所有邻接特征值的绝对值至多为 $2\sqrt{p}$ [22, 2]。

固定一个奇素数 $L \geq 5$，满足 $L \neq p$，其大小将在下文指定。我们限制于满足以下条件的素数 q above的一个无限子序列：

$$L \nmid q(q^2-1). \tag{11}$$

这样的无限子序列如下存在。选取模 p 的一个二次非剩余 b_p，以及一个剩余 $b_L \neq 0,\ \pm 1$（模 L）。根据中国剩余定理，选取模 $4pL$ 的一个既约剩余类 q_0，使其满足 $q_0 \equiv 1$（模 4）、$q_0 \equiv b_p$（模 p）以及 $q_0 \equiv b_L$（模 L）。狄利克雷定理给出无穷多个满足 $q \equiv q_0$（模 $4pL$）的素数；由于 $p \equiv 1$（模 4），由二次互反律可得 $(p/q) = (q/p) = -1$，而模 L 的选择给出公式 (11)。定义 $\phi \colon F \to \mathbb{Z}_L$，使得 $\phi(a_i) = 1$，并令 $K = \ker \phi$。条件 (11) 蕴含 G_q 没有 L 阶的商。因此 $\phi(N_q) = \mathbb{Z}_L$。否则 $N_q \subseteq K$，而 ϕ 将通过 $F/N_q \cong G_q$ 分解。于是，组合映射 $w \mapsto (\pi_q(w)\ \phi(w))$ 从 F 到 $G_q \times \mathbb{Z}_L$ 是满射的，且其核为 $N_q \cap K$。

令 $n_q = L m_q$。在 $G_q \times \mathbb{Z}_L$ 上定义两个 d-正则凯莱图：

$$Y_q = \mathrm{Cay}\left(G_q \times \mathbb{Z}_L, \{(\pi_q(a_i), 0)^{\pm 1} : 1 \leq i \leq r\}\right), \tag{12}$$

$$Z_q = \mathrm{Cay}\left(G_q \times \mathbb{Z}_L, \{(\pi_q(a_i), 1), (\pi_q(a_i)^{-1}, -1) : 1 \leq i \leq r\}\right). \tag{13}$$

投影到 G_q 会将任一图中的每个随机游走映射为 X_q中的一个随机游走。图 Y_q 是不交并，由 L 个 X_q的副本组成。在 Z_q中，附加坐标按照方程（13）中的规定变化。由于合并映射 $F \to G_q \times \mathbb{Z}_L$ 是满射的，因此它是连通的，所显示的集合生成 $G_q \times \mathbb{Z}_L$。

为了证明远离性，我们还需要 $h(Z_q)$ 的一个与 q 无关的正下界，尽管它可以依赖于固定参数 L。这样的下界将迫使任何仅使用少量编辑的修复保留几乎所有顶点于同一个连通分量中。我们利用子群 K 和固定商 F/K 中的扩张性，通过下面的引理得到该下界。

在本小节中，归一化邻接谱间隙表示为 $1 - \lambda_2$，其中 λ_2 是归一化邻接算子的第二大特征值。

引理 7.1（子群与商群产生的谱隙）。设 $F = \langle S \rangle$ 由有限对称集 S生成，设 K 是 F 的正规子群，且 F/K 有限，[4] 并设 H_j 是包含于 K 中的 F 的正规子群，且 F/H_j 有限。假设存在有限对称生成集 $T \subseteq K$ 和常数 $\kappa_K > 0$，使得每个凯莱图的归一化邻接算子

$$\mathrm{Cay}(K/H_j, T)$$

[2]这是由可逆的 2×2 矩阵构成的群，其中非零标量倍数被视为相同。[3]其元素是由符号 $a_i^{\pm 1}$ 组成的词，并在消去相邻逆元对后得到。乘法将两个词连接起来，并执行这些消去操作。[4]陪集是一个集合 $gK = \{gk : k \in K\}$。正规性意味着 $gK = Kg$ 对每个 $g \in F$都成立；此时这些陪集构成商群 F/K。

的谱隙至少为 κ_K。令 S 表示 S 在 F/K 中的像，并保留重复元素。[5] 另假设固定商凯莱多重图

$$\mathrm{Cay}(F/K, \overline{S})$$

其归一化邻接算子的谱间隙为 $\kappa_0 > 0$。如果每个 $t \in T$ 都可以表示为至多由 ℓ 个 S 中的元素相乘得到，那么凯莱图 $\mathrm{Cay}(F/H_j, S)$ 的归一化邻接算子的谱间隙至少为

$$\min\left\{\kappa_0, \frac{\kappa_K}{|S|\ell^2}\right\}.$$

证明。 固定 j，令 $\mathcal{H}$ 为 F/H_j 上的复值函数空间，并赋予通常的内积。对于 $g \in F/H_j$，令 $(\pi(g)v)(x) = v(xg)$。这些映射会置换坐标，因此保持内积，且 $v = r\ v - v\|^2 \quad \mathcal{H}^K = \{v\ \ \pi\ k\ v = v \quad k \in K\}$ $(gh)\ (\ )\ (h)$。对于有限对称集 R，记 $\mathcal{E}_R(\ ) : (2|R|)^{-1} \sum_{r\in R} \|\pi(\ )$。令：对每个都成立$(\ )$。这个子空间由在 K/H_j 的每个陪集上取常值的函数构成。由于 K 是 F 的正规子群，该子空间及其正交补在每个 $\pi(g)$ 作用下都保持不变。在 $\mathcal{H}^K$ 上，映射 $\pi(g)$ 作为商 F/K 上的平移作用。因此，在 $\mathcal{H}^K$ 中与常值函数正交的部分上，$\mathcal{E}_S(v) \geq \kappa_0\|v\|^2$。

现在令 $v \perp \mathcal{H}^K$。T 中元素的平移分别作用于 K/H_j 的每个陪集上，并且在每个陪集上，它给出了 Cay$(K/H_j,\ T)$ 上的随机游走。根据假设的谱间隙，$\mathcal{E}_T(v) \geq \kappa_K\|v\|^2$。对于 $t = s_1 \cdots s_m$，若有 $m \leq \ell$ 和 $s_i \in S$，则通过裂项求和与柯西–施瓦茨不等式可得

$$\begin{aligned}\|\pi(t)v - v\|^2 &\leq m \sum_{i=1}^{m} \|\pi(s_i)v - v\|^2 \\ &\leq \ell^2 \sum_{s\in S} \|\pi(s)v - v\|^2.\end{aligned}$$

对 $t \in T$ 取平均得到 $\mathcal{E}_T(v) \leq |S|\ell^2\mathcal{E}_S(v)$，因此 $\mathcal{E}_S(v) \geq \kappa_K\|v\|^2/(|S|\ell^2)$。这两个不变子空间彼此正交，并且由归一化的 S-平均算子保持不变。因此，前述估计给出了作用于与常数正交的函数上的谱间隙的界，正如所需。

□

引理 7.2（与 Z_q 的扩张性质）。 存在常数 $c_+ = c_+(d) > 0$ 和 $h_0 = h_0(d, L) > 0$，使得对于选定子序列中的每个 q，

(i) $Y_q \in \Xi_{c_+}(n_q)$
</style};

(ii) $h(Z_q) \geq h_0$；并且

(iii) Z_q 有一个集合 A_q，满足 $|A_q| = \lfloor L/2 \rfloor m_q$，并且

$$\frac{|E(A_q, V \setminus A_q)|}{|A_q|} = \frac{d}{\lfloor L/2 \rfloor}.$$

证明。 由于 X_q 是 d-正则 Ramanujan 图，标准的切格不等式给出了其边扩张 [18] 的均匀正下界。固定任意 $0 < c_+ < (d - 2\sqrt{d-1})/2$。的每个连通分量 Y_q 都是 $\overline{X_q}$ 的一个副本，从而证明了 (i)。

对于 (ii)，LPS商群 F/N_q 构成一个扩张子族 [22]。Abért 和 Elek [1, 定理 8] 表明，对于任意固定的有限对称生成集，一个由正规子群得到的有限商群扩张族，在限制到一个固定的有限指数子群后仍然是扩张的。在这里，每个 N_q 都是 F 的正规子群，而 K 在 F 中的指数固定为 L。因此，对于 K 的一个固定有限对称生成集 T，凯莱图 Cay$(K/(N_q \cap K)\ T)$ 构成一个扩张子族。由于 T 是固定的，Cheeger 的

[5] 不同的 生成元在商中可能具有相同的像。每个原始生成元仍会被分别计数，因 n averagin 平均。

切格不等式给出了一个常数 $\kappa_K > 0$，与 q无关，从下方界定它们归一化邻接谱间隙。

T中的每个元素都可以写成至多 $\ell = \ell(L)$ 个 $S = \{a_1^{\pm 1}, \dots, a_r^{\pm 1}\}$中元素的乘积。商 $F/K \cong \mathbb{Z}_L$ 是固定的。在 S的像中保留重数后，每个 a_i 贡献 $+1$，每个 a_i^{-1}贡献 -1，因此归一化商图邻接算子的非平凡特征值为 $\cos(2\pi j/L)$，其中 $1 \le j \le L-1$。所以其谱间隙为 $\kappa_0 = 1 - \cos(2\pi/L) > 0$。将引理 7.1 应用于 $H_j = N_q \cap K$，可得凯莱图 Z_q具有均匀的归一化谱间隙。接着由切格不等式可得 $h(Z_q) \ge h_0(d, L) > 0$。

对于 (iii)，令 $I \subseteq \mathbb{Z}_L$ 由 $k = \lfloor L/2 \rfloor$ 个连续剩余类组成，并设 $A_q = G_q \times I$。在循环 $\mathbb{Z}_L$中，离开 I 的两条边各自所对应的 $r = d/2$ 个生成元（由 a_1、……、a_r 索引）恰好贡献rm_q 条无向边。因此 $|E(A_q, V \setminus A_q)| = 2rm_q = dm_q$，从而得到所声称的比值。 □

7.2. 与目标性质的距离为常数

对于下面选定的参数，平衡割给出 $Z_q \notin \Xi_{c_}$。这表明该图不满足所要求的扩张性，但仅凭这一点还不能证明其具有远离性。删除跨割边可能会将图分割成已经具有扩张性的连通分量。图 Z_q 还具有一个取决于 L的固定正扩张下界。因此，采用少量编辑进行修复时，必须使几乎所有顶点仍处于同一个连通分量中。该连通分量仍包含稀疏割两侧的大量部分，因此要达到目标扩张性，就需要添加许多新的跨割边。引理7.3 将这一论证精确化。

引理 7.3（由稀疏平衡割得到的远离性判据）。 固定 d 以及常数 $h_0 > 0$、$c_* > h_1 \ge 0$ 和 $a \in (0,\ 1/2]$。设 G 是一个连通的、具有 N 个顶点且最大度至多为 d 的图，并满足 $h(G) \ge h_0$。假设存在 $A \subseteq V(G)$，使得

$$aN \le |A| \le N/2, \qquad |E_G(A, V \setminus A)| \le h_1 |A|.$$

于是，对于某个 $\varepsilon_0 = \varepsilon_0(d, h_0, c_*, h_1, a) > 0$，$G$ 与 $\Xi_{c_*}(N)$ 相差至少 ε_0。

证明。 令 $H \in \Xi_{c_*}(N)$，并令

$$M = |E(G) \setminus E(H)| + |E(H) \setminus E(G)|$$

表示对称差中的边数。若 H 的每个连通分量的大小都至多为 $N/2$，那么将 G 中的 h_0-扩张不等式对 H 的各个连通分量求和可知，G 中至少有 $h_0 N/2$条边连接不同的 H-连通分量。所有这些边都不属于 H，因此$M \ge h_0 N/2$。

否则，H 有唯一的连通分量 C，满足 $|C| > N/2$。令 $s = N - |C| < N/2$。C 与 $V \setminus C$ 之间的每条 G-边都不属于 H，因此

$$M \ge h_0 s. \tag{14}$$

若 $s > |A|$，则这已经给出了 $M \ge h_0 aN$。以下设 $s \le |A|$。由 A 在 C中诱导的割的两侧大小都至少为 $|A| - s$，因为 $|A \cap C| \ge |A| - s$ 和 $|C \setminus A| \ge N - s - |A| \ge |A| - s$。由于 $H[C]$ 的扩张性至少为 c_*，它至少包含 $c_*(|A| - s)$ 条跨越割 $A, V \setminus A$的边。其中至多 $h_1|A|$ 条已经是 G 的跨割边；其余每一条都属于 $E(H) \setminus E(G)$。因此

$$M \ge (c_* - h_1)|A| - c_* s. \tag{15}$$

结合公式（14）和（15），并对 $s \ge 0$ 进行最小化，得到

$$M \ge \frac{h_0(c_* - h_1)}{h_0 + c_*}|A| \ge \frac{a h_0 (c_* - h_1)}{h_0 + c_*} N.$$

因此，每个目标图与 G 至少有 $\Omega(N)$ 条边不同。更改一个邻居表条目可以移除一个邻居并添加另一个邻居，因此它在对称差中最多对应两条边。因此，有界度距离至少为同一常数除以 $2d$。 □

选取足够大的固定素数 L，使得 $c_- := 2d/\lfloor L/2 \rfloor < c_+$。那么引理 7.2 (iii) 满足 $h_1 = c_-/2$。将引理 7.3应用于 $a = \lfloor L/2 \rfloor / L$ 和 $c_* = c_-$，可得一个常数 $\varepsilon_0 > 0$，使得每个 Z_q 都与 ε_0相距 $\Xi_{c_-}(n_q)$。由于 L 在 q 变化之前已固定，所有常数都与n_q无关。

7.3. LPS 图中的返回概率

$R_q(k)$ 表示在自由生成元 $a_1^{\pm 1}, \dots, a_r^{\pm 1}$中，长度为 k 的均匀随机约化词属于 N_q 的概率。等价地，$R_q(k)$ 是 X_q 上长度为 k 的非回溯游走的返回概率。长度为 k 的约化词共有 dp^{k-1} 个。

在中间长度范围内，LPS 游走的算术描述将返回界归结为计算一个整数 N 表示为三个平方数之和的方法。我们需要一个均匀的 $O(\sqrt{N}\log(2N))$ 界；其对数因子将解释查询下界中的对数损失。Tillich 和 Zémor 此前已对这些 Ramanujan 构造给出了环计数估计 [26]。我们采用 Alon 和 Bachmat 的计数论证 [2], 以及以下均匀三平方和界。

引理 7.4（均匀三平方和界）。设

$$r_3(N) = \big|\{(x_1, x_2, x_3) \in \mathbb{Z}^3 : x_1^2 + x_2^2 + x_3^2 = N\}\big|.$$

存在一个绝对常数 C，使得对每个整数 $N \geq 1$,

$$r_3(N) \leq C\sqrt{N}\,\log(2N).$$

证明。情形 $N = 1$显然成立，因此假设 $N \geq 2$。反复使用经典恒等式$r_3(4N) = r_3(N)$可将问题化归为情形 $4 \nmid N$。高斯三平方和公式将 $r_3(N)$表示为某个量 $H(M)$的常数倍，其中 $M \leq 4N$；[6]现代版经典公式的证明见 Mortenson [23]，便于使用的类数表述见 Filaseta、Klein 和 Sabuncu [13]。因此，只需证明

$$H(M) = O(\sqrt{M}\log(2M)). \tag{16}$$

若 $H(M) = 0$，则无需证明。否则，类数公式（见 [13]）给出整数$D < 0$和 $f \geq 1$，满足 $-M = Df^2$，以及一个函数 $\chi_D \colon \mathbb{Z} \to \{-1, 0, 1\}$，[7]使得

$$H(M) = \frac{\sqrt{|D|}}{\pi} L(1, \chi_D) \sum_{e \mid f} \mu(e)\chi_D(e)\sigma(f/e),$$

这里，$L(1, \chi_D) = \sum_{m \geq 1} \chi_D(m)/m$, μ 是莫比乌斯函数，而 σ 是因子和函数。[8]该函数满足 $\chi_D(ab) = \chi_D(a)\chi_D(b)$，模 $|D|$ 周期变化，在与 $|D|$ 不互素的整数处取零，并且在一个周期内的和为零。我们还用 φ 表示欧拉函数，用 $\mathrm{rad}(f)$ 表示 f 的不同素因子的乘积。令 $S_D(f) := \sum_{e \mid f} \mu(e)\chi_D(e)\sigma(f/e)$。对于互素的$u$, v，有 $S_D(uv) = S_D(u)S_D(v)$。若 ℓ是素数且 $\ell^j \mid f$但 $\ell^{j+1} \nmid f$，则

$$S_D(\ell^j) = \sigma(\ell^j) - \chi_D(\ell)\sigma(\ell^{j-1}).$$

若要在级数中仅保留与 f 互素的项，请设定

$$L_f(1, \chi_D) := L(1, \chi_D) \prod_{\ell \mid f} \left(1 - \frac{\chi_D(\ell)}{\ell}\right).$$

[6]量 $H(M)$ 称为赫维茨类数。对 $H(M)$ 给出的公式是估计中唯一使用的类数事实。[7]在所引用的类数公式中，χ_D 是与 D相关联的二次狄利克雷特征。[8]这里，如果素数平方整除 n，则 $\mu(n) = 0$；如果 n 是 r 个互不相同的素数的乘积，则 $\mu(n) = (-1)^r$；$\mu(1) = 1$。此外，$\sigma(n) = \sum_{a \mid n} a$ 且 $\varphi(n) = |\{1 \leq a \leq n : \gcd(a, n) = 1\}|$。

直接检查情形 $\chi_D(\ell) \in \{-1,0,1\}$ 可得

$$\frac{|S_D(\ell^j)|}{\ell^j}\left|1-\frac{\chi_D(\ell)}{\ell}\right|^{-1} \le \frac{\ell}{\ell-1}.$$

将 $\ell \mid f$ 相乘，得到

$$|L(1,\chi_D)S_D(f)| \le f\,|L_f(1,\chi_D)|\,\frac{f}{\varphi(f)}. \tag{17}$$

分别估计 $L_f(1,\chi_D)$和 $f/\varphi(f)$会引入不必要的 $\log\log f$损失。相反，我们在 $L_f(1,\chi_D)$的界中保留因子 $\varphi(f)/f$；下文的式 (18) 随后会消去式 (17) 中的因子 $f/\varphi(f)$。展开

$$L_f(1,\chi_D)=\sum_{m\ge 1}\frac{a_m}{m},\qquad a_m=\chi_D(m)\mathbf{1}_{(m,f)=1}.$$

这些系数以 $Q_{\mathrm{per}}=\mathrm{lcm}(|D|,\mathrm{rad}(f))$ 为周期，且在一个周期内的和为零。事实上，在整除 D 的素数处，互素条件是多余的；而在其余素数处，根据中国剩余定理，该条件可独立分解；χ_D 在一个完整周期内的和为零。

对于第一个周期，

$$\begin{aligned}\sum_{m\le Q_{\mathrm{per}}}\frac{|a_m|}{m} &\le \sum_{\substack{m\le Q_{\mathrm{per}}\\(m,f)=1}}\frac{1}{m}\\ &\le \prod_{\substack{\ell\le Q_{\mathrm{per}}\\ \ell\nmid f}}\left(1-\frac{1}{\ell}\right)^{-1}\\ &= \prod_{\ell\le Q_{\mathrm{per}}}\left(1-\frac{1}{\ell}\right)^{-1}\prod_{\ell\mid f}\left(1-\frac{1}{\ell}\right)\\ &= O\left(\frac{\varphi(f)}{f}\log(2Q_{\mathrm{per}})\right),\end{aligned}$$

这里用到了梅滕斯界 $\prod_{\ell\le x,\,\ell\text{ 素数}}(1-1/\ell)^{-1}=O(\log(2x))$。此外，一个周期内非零系数的个数至多为 $\varphi(Q_{\mathrm{per}})\le Q_{\mathrm{per}}\varphi(f)/f$。由于完整周期的和为零，因此 (a_m) 的任意部分和都是 $O(Q_{\mathrm{per}}\varphi(f)/f)$。对尾项使用分部求和可得$|\sum_{m>Q_{\mathrm{per}}}a_m/m|=O(\varphi(f)/f)$。因此

$$|L_f(1,\chi_D)| = O\left(\frac{\varphi(f)}{f}\log(2Q_{\mathrm{per}})\right)=O\left(\frac{\varphi(f)}{f}\log(2M)\right). \tag{18}$$

将公式（18）代入公式（17后，因子 $f/\varphi(f)$ 相消，得到 $|L(1,\chi_D)S_D(f)|=O(f\log(2M))$。由于 $\sqrt{|D|f}=\sqrt{M}$，公式（16）成立，从而引理也成立。 □

引理 7.5（一致非回溯返回界）。存在常数 $C=C(p)<\infty$ 使得，对于选定子序列中所有充分大的 q 以及每个 $k\ge 1$,

$$R_q(k)\le C\frac{\log(3m_q)}{m_q}.$$

证明。对于 $k<$围长(X_q) 返回概率为零。假设 围长$(X_q)\le k\le 2\log_p m_q$。LPS 构造的算术描述 [22, 2] 将长度为k 的闭合非回溯游走与以下方程的归一化整数解对应起来：

$$x_0^2+q^2(x_1^2+x_2^2+x_3^2)=p^k. \tag{19}$$

这里，若 $x_0 > 0$ 为奇数，x_1、x_2、x_3 为偶数，且 p 不整除全部四个坐标，则称一个解为归一化解。x_0 的可能取值落在两个公差为 $2q^2$ [22, 的等差数列的并集中。定理 3.4]；另见定理 3.1]，，其中以本文采用的非回溯游走表述记录了同一结论。由于 $|x_0| \le p^{k/2}$，在此范围内这两个等差数列包含 $O(1 + p^{k/2}/q^2)$ 个可能取值。相同的 LPS 算术结论还表明，[2, 时，只要存在这样的归一化解，就有 $p^{k/2} > q^2$ 定理 3.1]，，因此这是 $O(p^{k/2}/q^2)$。对于固定的可行 x_0，令 $M = (p^k - x_0^2)/q^2$。一个归一化解满足 $M > 0$，且 $M \le p^k/q^2$。引理 7.4 将可能的三元组 $(x_1$、x_2、$x_3)$ 的数量界定为 $O(p^{k/2}q^{-1}\log(3m_q))$，其中用到了 $k \le 2\log_p m_q$。因此，方程 (19) 的归一化解数量，也就是从固定顶点出发、长度为 k 的闭合非回溯游走数量，至多为 $O(p^k q^{-3}\log(3m_q))$。将其除以长度为 k 的 $dp^{k-1} = \Theta_p(p^k)$ 个既约字，得到

$$R_q(k) = O\left(\frac{\log(3m_q)}{q^3}\right) = O\left(\frac{\log(3m_q)}{m_q}\right).$$

对于较长游走，我们使用 Ramanujan 特征值界以及正则图上非回溯游走的标准递推关系 [3]。令 Q_k 为非回溯多项式，则$Q_k(A_{X_q})(v,v)$ 统计从 k 出发、长度为 v的闭合非回溯游走。它满足

$$Q_0 = 1, \quad Q_1 = z, \quad Q_2 = z^2 - d, \qquad Q_k = zQ_{k-1} - pQ_{k-2} \quad (k \ge 3).$$

对于 $|z| \le 2\sqrt{p}$，此递推关系给出 $|Q_k(z)| \le C_p(k+1)p^{k/2}$.[10] 图 X_q 是二部 Ramanujan 图，因此在此区间之外的特征值只有 $\pm d$。群元素的左乘保持邻接关系，并且可以将任意顶点映射到其他任意顶点。因此，$Q_k(A_{X_q})$ 的每个对角元素都等于 $m_q^{-1}\mathrm{tr}(Q_k(A_{X_q}))$。除以约化词的总数 dp^{k-1}，得到 $R_q(k) \le 2/m_q + C_p k p^{-k/2}$。对于 $k \ge 2\log_p m_q$，第二项在左端点处为 $O((\log m_q)/m_q)$，之后逐渐减小。这就证明了对所有充分大的 q，所声称的界均成立。 □

在 Z_q 中，所有表示单位元的既约字都属于 $N_q \cap K \subseteq N_q$。因此，Z_q 的非回溯返回概率逐点不超过 $R_q(k)$。

7.4. 耦合自适应查询记录

我们使用以下耦合来比较自适应查询记录。考虑凯莱图 $G = \mathrm{Cay}(\Gamma, \Sigma)$，其带索引生成元采用固定的对称生成元索引 $\Sigma = \{s_1^{\pm 1}, \dots, s_r^{\pm 1}\}$，其中 $d = 2r$。对于 $k \ge 1$，令 $R_G(k)$ 表示在这些带索引生成元中均匀随机选取一个长度为 k 的既约字，其取值为 Γ 的单位元的概率。对每个群元素随机排列邻接表顺序，会使预言机算法无法获知这一生成元索引。

下面的证明避免了对先前没有发生碰撞这一条件进行条件化。我们首先让算法在一个理想的 d-正则森林上运行，其中邻接表顺序独立随机生成；随后，针对每个凯莱图，分别为森林分配隐藏的生成元标签，并将其投影到有限群中。在这个理想森林中，沿任意固定路径得到的字都均匀分布于对应长度的既约字集合中，无需以先前没有发生碰撞为条件，即使算法是自适应的也是如此。

引理 7.6 (与理想森林的耦合)。设 U =凯莱图(Γ_U, Σ_U) 和 V =凯莱图(Γ_V, Σ_V) 是定义在 n 个顶点上、具有固定对称生成元索引的简单 d-正则凯莱图。假设对于每个 $k \ge 1$，$R_U(k)$ $R_V(k) \le R(k)$。存在一个仅依赖于 d 的常数 $C_d > 0$，使得对于每个整数 $Q \ge 1$，在独立的随机顶点重标记和独立的随机邻接表顺序下，任何进行至多

[9]最后一个条件排除了 $(p^{k/2}, 0, 0, 0)$，其中 k 为正偶数。没有这一条件时，即使长度小于图的围长，该方程也会有解。[10]写作 $z = 2\sqrt{p}\cos\vartheta$。对于 $k \ge 2$，递推关系给出 $Q_k(z) = p^{k/2}\big(\sin((k+1)\vartheta) - p^{-1}\sin((k-1)\vartheta)\big)/\sin\vartheta$。使用 $|\sin(j\vartheta)| \le j|\sin\vartheta|$，并在 $\vartheta = 0, \pi$ 处取极限；$k = 1$ 的情形是直接的。

Q 次预言机查询的确定性自适应算法的
查询记录分布满足

$$d_{\mathrm{TV}}(\mathsf{T}_U, \mathsf{T}_V) \le C_d Q^2 \left(\frac{1}{n} + \sup_{1 \le k \le 2Q} R(k) \right).$$

算法可以查询 $[n]$ 中的任意标签，包括先前的预言机回答未返回过的标签。

证明。只需将图 $G = \mathsf{Cay}(\Gamma, \Sigma)$ 的查询记录与一个独立于图的理想查询记录进行比较。如果 $Q > n/4$，则增大 C_d 后，所示界显然成立，因此不妨假设 $Q \le n/4$。

定义一个理想预言机 $\mathcal{O}_\infty$，使其维护由无限 d-正则树构成的森林副本。它的回答使用 $[n]$ 中的标签，这些标签与森林中的顶点相区分。在森林的每个顶点处，将 d 条相邻树边通过独立的均匀随机双射分配到 $[d]$ 中的邻接表位置，并且仅在查询需要时才揭示该双射。当算法查询一个此前未出现过的顶点标签时，该标签便成为一个新的根。当它查询森林中某个顶点此前未查询过的邻接表位置时，预言机就沿相应的树边前进。如果相邻的森林顶点此前未出现过，则为其分配一个均匀随机的、尚未使用的顶点标签。否则，预言机返回该顶点已有的标签。对先前查询过的位置进行查询时，返回与之前相同的回答。特别地，查询沿已揭示树边返回的那个位置时，会返回父节点标签，这不构成碰撞。令 T_∞ 表示这一理想查询记录。其分布仅取决于 d 和该确定性算法。

我们将 G 的随机预言机与 $\mathcal{O}_\infty$ 耦合。为每个新理想根分配一个尚未分配给顶点标签的均匀随机群元素。对于理想森林中的每棵树，独立地按如下方式为其关联边分配隐藏生成元。在根处，使用其 d 条关联边与 Σ 之间的均匀随机双射。如果从父节点沿生成元 s 到达某个森林顶点，则强制令子节点处的父边携带 s^{-1}，并将其余 $d-1$ 个生成元均匀地分配给其余子边。任一森林顶点的像都通过将根像与唯一的根到顶点路径上的单词相乘得到；对于所得的森林顶点v的群像，记作 $\psi(v)$。

理想森林中的不同树可以使用彼此独立的局部生成元分配，即使它们最终获得相同的群像。因此，在发生这种识别后，辅助构造不必定义一个全局邻接表排序。不过，在首次识别之前，遇到的所有有限群顶点都是不同的，而这些独立的局部分配恰好构成这些顶点处的均匀随机邻接表顺序集合。通过 G 的均匀随机顶点重标记来耦合未使用的顶点标签，实际查询记录与 T_∞ 因而完全相同，除非两个不同的森林顶点获得相同的群像。

经过 Q 次查询后，理想森林至多包含 $2Q$ 个森林顶点，因为一次查询至多引入一个新根和一个新返回顶点。对理想查询记录及其已揭示森林进行条件化。隐藏生成元赋值与该查询记录相互独立，因此仍服从其原始分布。固定已揭示森林中的两个不同顶点 a，b，以下概率均以该理想查询记录为条件。

首先假设它们位于同一棵树中，树距离为 k。那么 $1 \le k \le 2Q$。沿着从 a到 b的唯一一条路径，刚经过的边所对应的生成元的逆元标记该边。非回溯规则会排除这条边，而其他每个出向生成元都是均匀选取的。因此，这条路径上的隐藏词在长度为 k 的 $d(d-1)^{k-1}$ 个约化词中恰好均匀分布。由于 $\psi(a)$ 当且仅当 $= \psi(b)$，也就是该词求值为单位元时，

$$\Pr[\psi(a) = \psi(b)] = R_G(k) \le R(k).$$

即使不以此前未发生碰撞为条件，这个等式仍然成立。

现在假设 a 和 b 位于不同的树中。不失一般性，设 b 位于根创建时间较晚的那棵树中。对截至但不包括分配给较晚创建的根的群

像的全部耦合历史进行条件化。还对两条根到顶点路径上的隐藏生成元赋值进行条件化；这些量与较晚创建的根像相互独立。较晚创建的根生成时，其像在尚未分配给可见顶点标签的群元素中均匀分布。至少有 $n - 2Q$ 个群元素可供选择。

一旦较早的根像和两条路径的既约字确定下来，等式 $\psi(a) = \psi(b)$ 至多确定较晚根像的一个可能取值。因此

$$\Pr[\psi(a) = \psi(b)] \le \frac{1}{n - 2Q} \le \frac{2}{n},$$

其中最后一个不等式使用了 $Q \le n/4$。这同样适用于由算法自适应地选择任意顶点标签所启动的树。

森林中不同顶点的有序对至多有 $\binom{2Q}{2} = O(Q^2)$ 个。由并集界可得

$$\Pr[\psi \text{ is not injective}] \le C_d Q^2 \left(\frac{1}{n} + \sup_{1 \le k \le 2Q} R(k) \right).$$

该界对于条件理想查询记录是均匀的。对其取平均并使用耦合不等式可得

$$d_{\mathrm{TV}}(\mathsf{T}_G, \mathsf{T}_\infty) \le C_d Q^2 \left(\frac{1}{n} + \sup_{1 \le k \le 2Q} R(k) \right).$$

将该估计应用于 U 和 V，并使用三角不等式即可证明该引理；将因子 2 吸收到 C_d 中。 □

对于 Y_q，带索引生成元中的既约字当且仅当属于 N_q 时取值为单位元，因此 $R_{Y_q}(k) = R_q(k)$。对于 Z_q，既约字只有在属于 $N_q \cap K \subseteq N_q$ 时才会取值为单位元，因此 $R_{Z_q}(k) \le R_q(k)$。因此，引理 7.6 适用于这对困难实例，其中$R(k) = R_q(k)$。

7.5. 下界的证明

定理 1.2 的证明。固定 p，然后如上选择 L、c_+、c_- 和 ε_0，并令 $\mathcal{N} = \{n_q = Lq(q^2 - 1) : q$ 属于选定子序列$\}$。对于每个 n_q，令 $\mathcal{Y}_q$ 和 $\mathcal{Z}_q$ 分别表示通过对 Y_q 和 Z_q 的顶点名称及邻接表顺序进行独立随机化而得到的有界度数预言机表示的分布。$\mathcal{Y}_q$ 的支持集中的每个表示所对应的底层图都属于$\Xi_{c_+}(n_q)$，而 $\mathcal{Z}_q$ 的支持集中的每个表示所对应的底层图都是一个与$\Xi_{c_-}(n_q)$相差至少 ε_0 的凯莱图。

由于 L 是常数，$m_q = \Theta(n_q)$。令 $\mathsf{T}^{\mathcal{Y}_q}$ 和 $\mathsf{T}^{\mathcal{Z}_q}$ 表示固定的确定性算法在这两种随机化预言机表示下产生的交互记录分布。引理 7.5和 7.6表明，对于任意确定性的 Q查询算法，

$$d_{\mathrm{TV}}(\mathsf{T}^{\mathcal{Y}_q}, \mathsf{T}^{\mathcal{Z}_q}) \le CQ^2 \frac{\log(3n_q)}{n_q}$$

对于所有充分大的 q，增大 C 以吸收 $1/n_q$ 项之后，

选取一个充分小的常数 $a = a(d,\ L) > 0$。若 $Q \le a\sqrt{n_q / \log(3n_q)}$，则对于所有充分大的 $1/3$，记录距离都小于 q。这样的确定性算法无法以成功概率$\mathcal{Y}_q$ 区分 $\mathcal{Z}_q$ 和 $2/3$。因此，Yao 极小极大原理 [27] 也给出了随机化算法的相同下界。由于对于充分大的$\log(3n)$ 替换为$\log n$ 只会改变一个绝对常数，n定理得证。 □

8. 局限性与开放问题

~上界尚未解决这样一个问题：能否构造一个 $O(\sqrt{n})$测试器，使其拒绝保证中的扩张阈值有一个正的常数下界。我们的迹论证将第二特征值的 $2t$次幂界定为 $O(t\eta)$。当 $t = \Theta(d^2c^{-2}\log(n/\eta))$时，要从这一估计中提取出一个常数间隙，需要 $\eta \le n^{-\Omega(1)}$。因此，较粗略的比较需要用不同的论证来证明内部扩张。对 ε 的依赖关系也是一个定量问题：不受限制的测试器在每次起始时都要额外付出 $O(1/\eta)$个返回估计，而这种代价是否不可避免尚不清楚。

对于下界，剩下的定量问题是：凯莱图承诺是否仍允许 $\Omega(\sqrt{n})$ 下界。我们的 $\sqrt{\log n}$ 损失源于引理 7.5中的中等长度关系估计。由于关于三个平方数之和的表示数量的均匀分布 $O(\sqrt{N})$ 界 $r_3(N)$ 通常并不成立 [13], ，要消除这一损失，就需要有关 LPS 关系的额外信息、更精细的自适应记录论证，或不同的困难实例族。我们的构造也未能解决这样一个问题：对于每个足够大的输入规模，是否都存在同阶下界。

参考文献

[1] Miklós Abért 和 Gábor Elek。有限群逆极限作用的动力学性质。遍历理论与动力系统，32(6):1805–1835，2012年。[2] Noga Alon 和 Eitan Bachmat。子图趋于无环的正则图。*Random Struct.Algorithms*，29(3):324–337，2006年。[3] Noga Alon、Itai Benjamini、Eyal Lubetzky 和 Sasha Sodin。非回溯随机游走的混合速度更快。当代数学通讯，9(4):585–603，2007年。[4] Tugkan Batu 和 Clément L. Canonne。广义均匀性测试。载于 Chris Umans 编，第*58*届 *IEEE* 计算机科学基础年会，*FOCS 2017*，美国加利福尼亚州伯克利，*2017*年*10*月*15*日至*17*日，第880–889页。IEEE 计算机学会，2017年。[5] Ferenc Bencs、Leslie Ann Goldberg、Matthew Jenssen、Mark Jerrum、Gábor Pete、Guus Regts 和 Yitong Yin。关于非回溯随机游走的自交时间。arXiv:2608.09729，2026年。[6] Debsoumya Chakraborti、Jaehoon Kim、Jinha Kim、Minki Kim 和 Hong Liu。良好混合顶点与近似扩张图。美国数学会会刊，150(12):5097–5110，2022年。[7] Ashish Chiplunkar、Michael Kapralov、Sanjeev Khanna、Aida Mousavifar 和 Yuval Peres。图聚类性测试：算法与下界。载于第*59*届 *IEEE* 计算机科学基础年会，*FOCS 2018*，第497–508页。IEEE 计算机学会，2018年。[8] Fan Chung。图中的随机游走与局部割。《线性代数及其应用》，423(1):22–32，2007年。[9] Artur Czumaj、Pan Peng 和 Christian Sohler。图聚类结构测试。载于第四十七届 *ACM* 计算理论年会论文集，*STOC 2015*，第723–732页。美国计算机协会，2015年。[10] Artur Czumaj 和 Christian Sohler。有界度图中的扩张性测试。*Comb.Probab.Comput.*，19(5–6):693–709，2010年。[11] Ilias Diakonikolas、Daniel M. Kane 和 Alistair Stewart。广义均匀性测试的紧确界。载于神经信息处理系统*31*进展，第6204–6213页，2018年。[12] Irit Dinur 和 Oded Goldreich。无连通性的扩张：性质测试视角。*Electron.Colloquium Comput.Complex. TR25-129*，2025年。

[13] Michael Filaseta、Jonah Klein 和 Cihan Sabuncu。等差数列中 $r_3(n)$的极值。《*Ramanujan*期刊》，69(2)，2026年。文章42。[14] Grzegorz Gluch、Michael Kapralov、Silvio Lattanzi、Aida Mousavifar 和 Christian Sohler。次线性时间内的谱聚类预言机。载于*2021*年 *ACM-SIAM*离散算法研讨会论文集，*SODA 2021*，第1598–1617页。SIAM，2021年。[15] Chris D. Godsil 和 Brendan D. McKay。游走正则图存在的可行性条件。《线性代数及其应用》，30:51–61，1980年。[16] Oded Goldreich 和 Dana Ron。有界度图中的性质测试。《算法学》，32(2):302–343，2002年。[17] Oded Goldreich 和 Dana Ron。关于有界度图中的扩张性测试。载于 Oded Goldreich 编，《复杂性与密码学研究：随机性与计算交互关系杂谈》，《计算机科学讲义》第6650卷，第68–75页。施普林格，2011年。[18] Shlomo Hoory、Nathan Linial 和 Avi Wigderson。扩展图及其应用。《美国数学会通报》，43(4):439–561，2006年。[19] Satyen Kale、Yuval Peres 和 C. Seshadhri。扩展图的噪声容忍度与次线性扩张重构。*SIAMJ.Comput.*，42(1):305–323，2013年。[20] Satyen Kale 和 C. Seshadhri。有界度图的扩张测试器。*SIAM J.Comput.*，40(3):709–720，2011年。[21] Michael Kapralov、Ekaterina Kochetkova 和 Weronika Wrzos-Kaminska。生日悖论时间内的谱聚类。载于*2026*年 *ACM-SIAM*离散算法研讨会论文集，*SODA 2026*，第3057–3071页。SIAM，2026年。[22] Alexander Lubotzky、Ralph Phillips 和 Peter Sarnak。Ramanujan图。*Comb.*，8(3):261–277，1988年。[23] Eric T. Mortenson。Kronecker型恒等式与一个数表示为三个平方数之和的表示。《伦敦数学会通报》，49(5):770–783，2017年。[24] Asaf Nachmias 和 Asaf Shapira。图的扩张性测试。*Inf. Comput.*，208(4):309–314，2010年。[25] Oded Regev 和 Igor Shinkar。随机游走中相对质量单调性的一个反例。《概率电子通信》，21:1–8，2016年。论文第8号。[26] Jean-Pierre Tillich 和 Gilles Zémor。由 Ramanujan 图构造的最优循环码。*SIAMJ.Discret.Math.*，10(3):447–459，1997年。[27] Andrew Chi-Chih Yao。概率计算：迈向统一的复杂性度量（扩展摘要）。载于第*18*届计算机科学基础年会，*FOCS 1977*，第222–227页。IEEE 计算机学会，1977年。

A. 更简单的游走正则输入测试器

当每个连通分量内的返回概率均为常数时，仅比较这两个长度就已经满足局部条件。这样，测试器便可以省略从碰撞位置出发的后续随机游走。我们记录了由此带来的对比较精度依赖性的改进。

图 H 的邻接矩阵 A_H 是 游走正则 的条件是 $(A_H^k)_{v,\,v}$ 与 v 对每个整数 $k \geq 0$ [15]都无关。图是 分支游走正则 的条件是每个连通分量都是游走正则图；不同的

连通分量不必具有相同的返回概率。取 $k=2$ 可知，每个游走正则图都是正则图。在度为 r 的连通分量上，懒惰游走矩阵为 $(1-r/(2d))I + A_H/(2d)$。其幂是 A_H 的多项式，因此它们的对角元素也都是常数。如果整个图是游走正则的，那么对所有顶点而言，对角元素都是常数。

命题 A.1。 对于定理 *1.1*中的参数，存在一个伪测试器，对$\Xi_c(n)$ 中的每个图都接受，并拒绝每个逐分量游走正则图 ε-远离 $\Xi_{c_\beta}(n)$，且每种情况下的概率至少为 $2/3$，使用

$$\widetilde{O}_{d,c,\beta}\left(\frac{n^{1/2+\beta/3}}{\varepsilon}+\frac{\sqrt{n}}{\varepsilon^2}\right)$$

次查询。如果整个输入是游走正则的，则一个起始顶点就足够，界变为

$$\widetilde{O}_{d,c,\beta}\left(n^{1/2+\beta/3}+\frac{\sqrt{n}}{\varepsilon}\right).$$

我们首先证明该测试器所使用的碰撞比较。对于连通分量 C，令

$$M_C=\max_{x\in C}A_x,\qquad b_C=\min_{x\in C}B_x,\qquad \Delta_C=M_C/b_C-1.$$

方程（7）和（2）推出

$$b_C\le B_x\le A_x\le(1+\Delta_C)b_C\qquad(x\in C)\tag{20}$$

and

$$\|P^{2t}(x,\cdot)\|_\infty\le M_C.\tag{21}$$

对于分布 r，记作 $F_j(r)=\sum_v r(v)^j$。由 $q\ge 2$ 个独立样本，令 $\widehat{F}_2(r)$ 表示两两碰撞次数除以 $\binom{q}{2}$。引理 4.1取 D 为整个定义域时，给出

$$\mathbb{E}\widehat{F}_2(r)=F_2(r),\qquad \mathrm{Var}(\widehat{F}_2(r))\le C\left(\frac{F_2(r)}{q^2}+\frac{F_3(r)-F_2(r)^2}{q}\right)\tag{22}$$

对于绝对常数 C。差值 $F_3(r)-F_2(r)^2$ 是 $r(X)$ 的方差，其中 X 取自 r。特别地，在均匀分布上该值为零。两个游走长度之间的关系给出以下界。

引理 A.2（两个游走分布的三阶矩）。 固定一个连通分量 C，令 $b=b_C$ 和$\delta=\Delta_C$，并假设 $0\le\delta\le 1$。对于任意 $x\in C$，令 $r=P^t(x,\cdot)$ 和 $p=P^{2t}(x,\cdot)$。则

$$F_3(p)-F_2(p)^2=O(\delta b^2),\qquad F_3(r)-F_2(r)^2=O(\delta b^2+\delta^{3/2}b^{3/2}),$$

其中隐含常数为绝对常数。

证明。 令 $a=\|r\|_2^2$，$b_x=\|p\|_2^2$，以及 $D=a-b_x$。方程（20）给出

$$b\le b_x\le a\le(1+\delta)b,\qquad 0\le D\le\delta b.$$

方程（21）给出 $\|p\|_\infty\le(1+\delta)b=:M$。因此，令 $G_p=F_3(p)-b_x^2$，

$$0\le G_p\le(M-b_x)b_x\le\delta(1+\delta)b^2.$$

这就证明了第一个界。

要控制短游走，请设置 $e=r-p$。对 P 进行谱分解可得

$$\|e\|_2^2\le D,\tag{23}$$

因为 $\lambda^2 \quad {}^t(1-\lambda^t)^2 \le \lambda^{2t}(1-\lambda^{2t})$，对于 $0 \le \lambda \le 1$。此外 $\langle p, \quad e\rangle = (D - \|e\|_2^2)/2 \le D/2$。因此 y,

$$\sum_v p(v)^2 e(v) = b_x\langle p, e\rangle + \sum_v p(v)(p(v) - b_x)e(v)$$
$$\le b_x D/2 + \sqrt{MG_pD}.$$

这里，柯西–施瓦茨不等式使用 $\sum_v p(v)^2(p(v) - b_x)^2 \le M\sum_v p(v)(p(v) - b_x)^2 = MG_p$。此外，$\sum_v p(v)e(v)^2 \le MD$ 和 $\sum_v e(v)^3 \le \|e\|_2^3 \le D^{3/2}$。展开 $(p+e)^3$ 并利用 $a^2 \ge b_x^2$ 可得

$$F_3(r) - a^2 \le G_p + \tfrac{3}{2}b_x D + 3\sqrt{MG_pD} + 3MD + D^{3/2}.$$

由于 M， $b_x = O(b)$、 $D \le \delta b$和 $G_p = O(\delta b^2)$，右侧为 $O(\delta b^2 + \delta^{3/2}b^{3/2})$，从而证明了所述界。 □

引理 A.3 (单一起点处的比较)。 在连通分量 C中固定 s，并令 $0 < \theta \le 1/64$。使用两个各含 $q = \lceil K\sqrt{n/\theta}\rceil$ $t \quad 2t \quad s \quad U = \widehat{F}_2\ P^t\ s\cdot \quad V =$个端点的 $\widehat{F}_2\ P^{2t}\ s\cdot \quad K \quad U \le\ 1+4\theta\ V \quad \Delta_C \le \theta$ 独立批次，在长度和起点分别为 和 的情况下，构造 $((,))$和$((,))$，其中 为充分大的绝对常数。若 ()，则接受。若 ，接受的概率至少为 $9/10$。若 C 是游走正则的且 $\Delta_C \ge 8\theta$，则拒绝的概率至少为 $9/10$。

证明。 记 $b = b_C$， $\delta = \Delta_C$，以及 $Z = U - (1+4\theta)V$。对于 $\delta \le 1$，由独立性、式（22）和引理 A.2可得

$$\operatorname{Var} Z \le C\left(\frac{b}{q^2} + \frac{\delta b^2}{q} + \frac{\delta^{3/2}b^{3/2}}{q}\right).$$

若 $\delta \le \theta$，则 $\mathbb{E}Z \le -3\theta b$。将方差除以 $\theta^2 b^2$ 后，这三项分别至多为 K^{-2}、$(K\sqrt{n})^{-1}$和 $\sqrt{\theta}/(K\sqrt{nb})$ 的常数倍。由于 $\overline{b} \ge 1/n$，切比雪夫不等式表明，当 K 足够大时，接受的概率至少为 $9/10$ 。

对于拒绝，游走正则性意味着 $A_s = (1+\delta)b$ 和 $B_s = b$，因此 $\mathbb{E}Z = (\delta - 4\theta)b \ge \delta b/2$。若 $8\theta \le \delta \le 1$，将方差除以 $\delta^2 b^2$ 后至多为

$$C\left(\frac{1}{q^2 b\delta^2} + \frac{1}{q\delta} + \frac{1}{q\sqrt{b}\sqrt{\delta}}\right) = O(1/K).$$

若 $\delta > 1$，使用 $F_3(P^t(s, \cdot)) \le ((1+\delta)b)^{3/2}$以及 $F_3(P^{2t}(s, \cdot)) - b^2 \le \delta b^2$，后者由式（21）推出。归一化方差于是至多为

$$C\left(\frac{1+\delta}{q^2 b\delta^2} + \frac{(1+\delta)^{3/2}}{q\sqrt{b}\delta^2} + \frac{1}{q\delta}\right) = O(1/K).$$

切比雪夫不等式证明了两种情况下的拒绝保证。 □

命题 A.1 的证明。 使用 η， t 以及定理 1.1证明中的常数 a_0。对于每个抽样的起点，使用 $\theta = \eta/4$应用引理 A.3 ，通过独立重复进行放大，并在任一放大后的比较检验拒绝时予以拒绝。使用与无限制测试器中相同的 $O_{d,\ c,\ \beta}(1/\varepsilon)$个均匀分布的起点。

为完整起见，引理 3.1 在每个扩张分量中给出 $\Delta_C \le \eta/10 < \theta$ ，无论该分量是否游走正则。对于逐分量游走正则图的可靠性，由于 A_v 在每个分量上都是常数，因此 D_s 为空。因此，当且仅当 $\Delta_{C(s)} \le 2\eta$时，一个起点才是良好的。定理 1.1 证明中的修复论证表明，一个 ε-远图有 $\Omega_{d,\ c,\ \beta}(\varepsilon)$ 比例的起点满足$\Delta_{C(s)} > 2\eta = 8\theta$。放大后的比较以高概率检测出任何这样的起点。

每个比较在放大之前需要 $\overline{O}(\sqrt{n}\,t/\eta)$ 次查询。该单位元

$$1/\eta = a_0^{-1}\max\{\varepsilon^{-1}, n^{\beta/3}\}$$

给出了所述的查询复杂度上界。如果整个图是游走正则图， A_s/B_s 就与 s 无关。如果这个公共比率至多为 $1+2\eta$，那么每个起点都会是良好的，而相同的修复上界将与远离性矛盾。因此，远游走正则图中的每个起点都会不满足比较条件。只需从一个起点进行常数次放大即可。 □