arXiv:2305.05487v1 [math.CO] 9 May 2023

图性质的测试与估计，再探讨

Lior Gishboliner* Nick Kushnir† Asaf Shapira ‡

摘要

图 G 具有 n 个顶点，若要将其转换为满足 $\mathcal{P}$的图，至少需要添加/删除 εn^2 条边，则称其与性质 $\mathcal{P}$ 相距 ε-远。$\mathcal{P}$ 的距离估计器是一种算法，给定 G 和α，$\varepsilon > 0$ 能够区分以下两种情况：G 与 $\mathcal{P}$ 相距 $(\alpha - \varepsilon)$-近，以及 G 与 $\mathcal{P}$相距 α-远。如果 $\mathcal{P}$ 具有查询复杂度仅取决于 ε的距离估计器，则称 $\mathcal{P}$ 是可估计的。

每个可估计性质显然也都是可测试的，因为测试相当于以$\alpha = \varepsilon$进行估计。性质测试领域的一个核心结果是 Fischer–Newman 定理，它指出逆命题也成立，也就是说，每个可测试性质实际上都是可估计的。Fischer 和 Newman 的证明效率很低，因为将用于 $\mathcal{P}$ 的测试算法转化为距离估计器时会产生塔型损失。这引出了一个自然的问题，Fiat–Ron 和 Hoppen–Kohayakawa–Lang–Lefmann–Stagni 最近对此进行了研究：能否找到一种只产生多项式损失的转换？我们得到以下结果。

- 我们证明，如果 $\mathcal{P}$ 是遗传的，那么可以将 $\mathcal{P}$ 的测试器转化为一个距离估计器，且损失为指数级。这比 Hoppen 等人得到的结果有指数级的改进；他们得到的转换具有双指数级损失。
- 我们证明，对于每个 $\mathcal{P}$，都可以将针对 $\mathcal{P}$ 的测试算法转化为距离估计器，且只产生双指数级损失。这优于 Fischer–Newman 的转换方法，后者会产生塔式损失。

我们这项工作的主要概念性贡献，是成功将 Fischer– Newman 的方法从本质上无效转变为有效。在技术层面，我们的主要贡献是确立 Frieze–Kannan 弱正则划分的一些性质，这些性质本身也具有独立的研究价值。

1 引言

1.1 图性质测试的背景

性质测试器是快速的随机化算法，能够区分满足某个预先指定性质的对象 $\mathcal{P}$ 与 ε距离该性质至少 $\mathcal{P}$ 的对象。在大多数情况下，ε距离至少 ε意味着，为了得到一个满足 $\mathcal{P}$的新对象，需要改变对象表示中的 $\mathcal{P}$ 比例。因此，对 $\mathcal{P}$进行测试，是经典判定问题的一种宽松形式；经典判定问题要求判断一个对象是否满足。本文研究所谓邻接矩阵模型（有时也称为稠密图模型）中的图性质。这可以说是性质测试领域研究最充分的模型之一。有关性质测试的更多背景和参考文献，请参阅。

*ETH Z¨urich。电子邮件：lior.gishboliner@math.ethz.ch。获得 SNSF 资助 200021 196965 的支持。

†特拉维夫大学数学学院，以色列特拉维夫 69978。电子邮箱：nickkushnir@mail.tau.ac.il。

‡数学学院，以色列特拉维夫特拉维夫大学，邮编 69978。电子邮箱：asafico@tau.ac.il。部分经费由 ERC 巩固者资助 863438 和 NSF-BSF 资助 20196 提供。

现在我们介绍邻接矩阵模型下图性质测试的模型。图性质 $\mathcal{P}$ 是一个对同构封闭的图族。若一个含有 n 个顶点的图 G 至少需要添加或删除 εn^2 条边，才能将 G 变成满足 $\mathcal{P}$ 的图，则称 G 与 $\mathcal{P}$ 相距 ε。若 G 与 $\mathcal{P}$ 的距离不足 ε，则称 G 与 $\mathcal{P}$ 的距离至多为 ε。针对 $\mathcal{P}$ 的测试器是一种随机化算法，给定 $\varepsilon > 0$ 后，能够以高概率（例如 $2/3$）区分满足 $\mathcal{P}$ 的图与和 $\mathcal{P}$ 相距至少 ε 的图。我们假设算法可以针对每一对 $1 \le i$、 $j \le n$ 查询输入图 G 是否包含边 (i, j)。测试器的边查询复杂度记为 $Q(\varepsilon)$，即其执行的边查询次数。如果 $\mathcal{P}$ 存在一个测试器，其边查询复杂度仅取决于 ε （而与 n 无关），则称 $\mathcal{P}$ 是可测试的。下文中我们主要讨论顶点查询复杂度，即最小的 $q = q(\varepsilon)$，使得我们可以通过检查输入图 G 的一个子图来对 $\mathcal{P}$ 进行 ε-测试；该子图由随机选取的 q 个顶点诱导而成。根据 Goldreich 和 Trevisan 的定理 [21] ，我们知道 $q(\varepsilon) \le 2Q(\varepsilon) \le q^2(\varepsilon)$。在下文的大多数（但并非所有）讨论中，我们不太关注这些二次因子。在这种情况下，我们可能会使用“查询复杂度”一词，而不说明它指的是顶点查询复杂度还是边查询复杂度。

邻接矩阵模型下的性质测试最早由 Goldreich、Goldwasser 和 Ron 引入 [20], ，他们证明了每一种划分性质（例如 k-可着色性和最大割）都是可测试的。还有若干一般性结果保证了图性质的可测试性[3, 10]。Alon 和 Shapira 得到了这类结果[5] ，他们证明每一种遗传的[1] 图性质都是可测试的。他们的证明运用了 Szemer´edi 正则性引理 [35] （另见 [33]），这是研究稠密图性质时最有用的工具之一。使用这一工具代价高昂，因为利用正则性引理得到的界是塔型的[2]。

与图性质测试相关的一个核心开放（元）问题是：何时能将一个无效的结果（例如具有塔型界的结果）转化为有效的结果，最好是多项式界？尽管这是一个定量问题，其背后实际上是以下定性问题：何时能够在避开 Szemer´edi 正则性引理的同时证明可测试性结果？可以通过给出直接的组合论证，或使用较弱的正则性引理变体（例如 Frieze–Kannan 正则性引理[18] ，下文将对此进行讨论）。例如，R¨odl 和 Duke [31] 使用正则性引理（隐式地）证明了 k-可着色性是可测试的。在 [31] 中得到的塔型界，通过一种避免使用正则性引理的直接论证，在 [20] 中改进为多项式界。Alon 和 Fox 提出的一个具体核心开放问题[4], 涉及遗传性质，询问哪些遗传性质的查询复杂度为 poly$(1/\varepsilon)$ 时是可测试的。对这一问题的系统研究见[19]。

1.2 距离估计

在稠密图模型中，如果 α 是使得 G 与 $\mathcal{P}$ 满足 α-接近的最小实数，我们就称图与 $\mathcal{P}$ 的距离为 α。换言之，这是为了获得一个满足 $\mathcal{P}$ 的图，所需添加或删除的最少边数，再除以 n^2 进行归一化。我们将这一量记为 $\mathrm{dist}_{\mathcal{P}}(G)$。对于 $\mathcal{P}$ ，距离估计器是一种随机化算法，它在给定 α、$\varepsilon > 0$ 的情况下，以高概率（例如， $2/3$）区分与 $\mathcal{P}$ 满足 $(\alpha - \varepsilon)$-接近的图和与 $\mathcal{P}$ 相距 α 的图。如果对于每个 α、ε ，都存在一个针对 $\mathcal{P}$ 的距离估计器，且其查询复杂度仅取决于 ε，那么我们就称 $\mathcal{P}$ 是可估计的。注意，对 $\mathcal{P}$ 进行测试等价于在 $\alpha = \varepsilon$ 的情况下进行距离估计，因此，这一概念至少与可测试性同样强。

距离估计最初在[30] 中得到研究，此后也在各种其他领域中得到研究

[1]如果一个图性质在进行顶点删除后仍然保持不变，则称其为遗传的。例如，具有 3-可着色性、无三角形性质，以及对于某个固定的 H 不包含诱导 H 的性质，都是遗传的。[2]塔函数 tower(x) 是一个高度为 x 的指数塔。

例如分布 [7], 字符串 [6], 稀疏图 [11, 13, 28], 布尔函数 [1, 9], 纠错码[23, 26] 以及图像处理 [8]。众所周知，在某些领域中，存在可测试但不可估计的性质 [15]。图性质测试领域中最核心、也最出人意料的结果之一是 Fischer–Newman 定理 [16], ，该定理指出，在图这一领域中，每个可测试性质也都是可估计的。与该领域的若干主要结果一样，[16] 中的证明依赖于 Szemer´edi 正则性引理 [35] ，因此将一个针对 $\mathcal{P}$ 的测试器转化为针对 $\mathcal{P}$的距离估计器时，会产生塔式损失。回到上一小节最后一段的讨论，一个自然的问题是，能否改进[16] 的转换，将针对 $\mathcal{P}$ 的测试器转化为仅有多项式损失的距离估计器。

1.3 关于遗传图性质的新结果

正如我们在上一小节中提到的，遗传图性质这一族在图性质测试领域得到了广泛研究。每个遗传性质都是可测试的，这一事实可由下面的陈述推出，其中我们用 $ind(F,G)$ 表示随机映射 $\varphi\colon V(F) \to V(G)$ 是一个单射诱导同态的概率。 [3]

引理 1.1（诱导消去引理， [5]）。对于每个 $\varepsilon > 0$ 和每个遗传的 $\mathcal{P}$，存在 $M = M_{1.1}(\varepsilon,\mathcal{P})$、 $\delta = \delta_{1.1}(\varepsilon,\mathcal{P}) > 0$ 和 $n_0 = n_{1.1}(\varepsilon,\mathcal{P})$，使得如果一个有 $n \geq n_0$ 个顶点的图 G 与 $\mathcal{P}$ 相距至少 ε，那么存在一个图 $F \notin \mathcal{P}$，满足 $|V(F)| \leq M$，且 $ind(F,G) \geq \delta$。

上述引理的第一个版本由 Alon、Fischer、Krivelevich 和 Szegedy 得出 [2]他们在.....的情况下证明了该引理 $\mathcal{P}$ 可以通过有限个禁用诱导子图来刻画。Alon 和 Shapira 在最一般的情况下证明了该引理[5]。随后，Lov´asz 和 Szegedy [27], Conlon 和 Fox [12] 以及 Borgs 等人 [10]分别给出了其他证明。此外，R¨odl 和 Schacht [32]还将其推广到了超图的情形。

注意，由引理 1.1 可立即得出，每个遗传性质都是可测试的，且其顶点查询复杂度为

$$q(\varepsilon) = \max\{n_0, M/\delta\}\ . \tag{1.1}$$

事实上，该算法从 X 个顶点中采样一个集合，查询 X 中所有顶点对的信息，当且仅当 X 上的图满足 $\mathcal{P}$ 时才接受。如果 G 满足 $\mathcal{P}$，那么该算法显然能够正确作答（概率为 1）。如果 G 与 $\mathcal{P}$ 相差至少 ε，那么根据引理 1.1，随机选取的 M 元顶点组以至少 δ 的概率张成一个图 $F \notin \mathcal{P}$ 的诱导副本。因此，一个大小为 M/δ 的样本以至少 $2/3$ 的概率包含一个 F 的诱导副本，从而保证该顶点样本不满足 $\mathcal{P}$（因为 $\mathcal{P}$ 是遗传的）。回顾一下， [21] 证明了，如果 $\mathcal{P}$ 是可测试的，那么它可以通过上述算法进行测试。因此，引理 1.1 中的界在某种程度上决定了遗传的 $\mathcal{P}$ 的查询复杂度。这引出了以下自然问题，该问题由 Hoppen 等人 [24, 25] 以及 Fiat 和 Ron [14], 提出：是否可以用测试每个遗传的 $\mathcal{P}$ 时所需的（大致）相同查询复杂度，对其进行估计，如式 (1.1) 所示?

问题 1.2. 确定每个遗传图性质 $\mathcal{P}$ 是否都可估计，其查询复杂度为

$$n_0 \cdot M/\delta\ ,$$

其中 $M = M_{1.1}(\varepsilon',\ \mathcal{P})$、 $\delta = \delta_{1.1}(\varepsilon',\ \mathcal{P})$以及 $n_0 = n_{1.1}(\varepsilon',\ \mathcal{P})$ 由引理 *1.1* 给出，其中 $\varepsilon' = \mathrm{poly}(\varepsilon)$。

注释 1.3。 对于某些遗传图性质（例如无三角形性），引理 M 和 δ 中的 *1.1* 的已知最佳界是塔型的。有人可能会认为，在这种情况下，两者之间几乎没有

[3]一个映射 $\varphi\colon V(F) \to V(G)$ 是一个诱导同态，当且仅当 $uv \in E(F)$ 当且仅当 $\varphi(u)\varphi(v) \in E(G)$.

区别，即 $\text{tower}(M/\delta)$ 界由 *[16]* 给出，而问题 *1.2*提出的界与之相同。不过，我们应强调，对于许多这类性质（例如无三角形性），尚不清楚塔式界是否紧确（事实上，目前已知的最佳下界仅略高于多项式界）。也许更重要的是，许多遗传图性质已知满足：引理 M 和 δ 中的 *1.1* 均为 ε 的多项式（例如 k-可着色性、是区间图或是线图；详见 *[19]*中的详细讨论）。对于所有这些性质，问题 *1.2* 给出的界为$\text{poly}(1/\varepsilon)$，而 $\text{tower}(1/\varepsilon)$ 给出的界为 *[16]*。

问题 1.2 由 Hoppen 等人研究。 [24, 25]他们的主要结果是，每个遗传的$\mathcal{P}$都是可估计的，且查询复杂度为 $2^{\text{poly}((1/\delta)^{M^2},\log n_0)}$。我们的第一个主要结果是对这一结果的指数级改进，使我们在解决问题 1.2.

定理 1.4. 每个遗传的 $\mathcal{P}$ 都是可估计的，查询复杂度为

$$2^{\text{poly}(M/\delta,\log n_0)}\,,$$

其中 $M = M_{1.1}(\varepsilon/2,\ \mathcal{P})$ 、 $\delta = \delta_{1.1}(\varepsilon/2,\ \mathcal{P})$ 和$n_0 = n_{1.1}(\varepsilon/2,\ \mathcal{P})$ 是引理*1.1*的参数。

注 1.5。在所有已知情形中，引理 *1.1* 中的最佳界满足 $n_0 \ll 1/\delta$，因此 *[24]* 的上界为 $2^{(1/\delta)^{O(M^2)}}$，而定理 *1.4* 中的上界为 $2^{\text{poly}(M/\delta)}$。

几乎在所有情况下，关于稠密图测试的结果都依赖于蕴含平凡算法的组合陈述。例如，测试遗传性质的算法 $\mathcal{P}$一旦我们掌握了引理 1.1 ，便是平凡的。与此形成鲜明对比的是，许多估计结果都涉及对一个顶点集合进行抽样，然后在该样本上执行一项高度非平凡的计算。本文中的情况确实如此，见引理 2.4 和 2.5的证明。不过，借助一个著名的抽样技巧 [20], ，我们可以将任何估计结果转化为组合陈述。例如，该技巧给出了定理 1.4的以下推论。

推论 1.6. 设 $q = 2^{\text{poly}(M/\delta,\log n_0)}$ 如定理 *1.4*所示。则

$$\Pr_X\left[|\text{dist}_{\mathcal{P}}(G[X]) - \text{dist}_{\mathcal{P}}(G)| \le \varepsilon\right] \ge 2/3\,,$$

其中，概率是针对从 G 中随机选取的 q 个顶点的子集 X 而言的，而 $G[X]$ 是 G 在 X 上诱导出的图。

值得注意的是，有了推论 1.6 ，现在我们可以回过头来，使用“平凡/自然”的算法重新证明定理1.4 。该算法抽取一组 q 顶点 X，计算 $\text{dist}_{\mathcal{P}}(G[X])$，然后判定 G 与 $\alpha-\varepsilon$相差不超过 $\mathcal{P}$，当且仅当 $\text{dist}_{\mathcal{P}}(G[X]) \le \alpha-\varepsilon/2$；否则，它与 α相差至少 $\mathcal{P}$。

我们对定理 1.4 的证明实际上给出了界 $2^{\text{poly}(M/\varepsilon\delta,\log n_0)}$。可以推测 $\text{poly}(M/\varepsilon\delta)$ =$\text{poly}(M/\delta)$，因为在所有已知情形中， δ 至多是 ε的多项式，而且在许多情形下要小得多。为了从形式上去除我们的界对 ε 的依赖，我们证明以下命题；若 $\mathcal{P}$ 包含所有图，或仅包含有限多个图，则 $\mathcal{P}$ 是平凡的。该命题的证明依赖于对拉姆齐定理的一种巧妙应用。

命题 1.7. 对于每个非平凡遗传性质 $\mathcal{P}$，以下结论成立。若 $q(\varepsilon)$ 表示 $\mathcal{P}$ 的顶点查询复杂度，则对于所有足够小的 ε，有

$$M/\delta \ge q(\varepsilon) \ge \Omega(1/\varepsilon)\,, \tag{1.2}$$

其中 $M = M_{1.1}(\varepsilon,\ \mathcal{P})$ 和 $\delta = \delta_{1.1}(\varepsilon,\ \mathcal{P})$ 是引理 *1.1*中的常数。

上面的左侧不等式由（1.1）得出。注意，q（ε）的下界是最优的，因为当 $\mathcal{P}$ 为不含边的性质时，该界是紧的（此时 q（ε）$=O$（$1/\varepsilon$））。

当然，也很自然地要研究问题 1.2 是否也适用于特定的遗传性质。一类自然的问题是：每个满足以下条件的遗传 $\mathcal{P}$，如果其查询复杂度为 poly($1/\varepsilon$)，是否也能以 poly($1/\varepsilon$) 的查询复杂度进行估计。Fiat 和 Ron 最近开始研究这一问题 [14]，并证明许多自然的遗传性质都满足这一结论，例如弦性以及不包含 4 个顶点的诱导路径。

1.4 关于一般图性质的新结果

鉴于上述讨论，下面的问题显得很自然。

问题 1.8。确定每个性质 $\mathcal{P}$ 若其顶点查询复杂度为 $q(\varepsilon)$，是否都能以查询复杂度 $q(\varepsilon')$ 进行估计，其中某个 ε' =poly(ε)。

在本研究之前，关于一般图性质的唯一结果 $\mathcal{P}$ 是 Fischer 和 Newman 的变换 [16]，该变换将图性质 $\mathcal{P}$ 的测试算法（其查询复杂度为 $q(\varepsilon)$）转化为查询复杂度为 tower($q(\varepsilon/2)$) 的距离估计器。利用我们为证明定理1.4所开发的工具，我们还得到了以下改进后的界。

定理 1.9. 如果 $\mathcal{P}$ 可测试，且查询复杂度为 $q(\varepsilon)$，那么它是可估计的，查询复杂度为$2^{\mathrm{poly}(1/\varepsilon)\cdot 2^{q(\varepsilon/2)}}$。

此时我们想指出，既然任何“自然”性质都满足 $q(\varepsilon)$ ≥log($1/\varepsilon$)，那么上述界可以写成 exp(exp(poly($q(\varepsilon/2)$)))。为了正式地作出这一断言，我们证明命题 1.7 的以下变体：如果存在 ε_0，使得对每个 $0<\varepsilon<\varepsilon_0$ 和 $n\geq n_0$(ε)，以下条件之一成立：每个 n 顶点图都与 $\mathcal{P}$ 相差不超过 ε，或者每个 n 顶点图都不属于 $\mathcal{P}$，那么 $\mathcal{P}$ 就是非自然的。如果 $\mathcal{P}$ 不是非自然的，那么它就是（自然地）自然的。

命题 1.10. 设 $\mathcal{P}$ 是一个自然性质，$q(\varepsilon)$ 是其顶点查询复杂度，$Q(\varepsilon)$ 是其边查询复杂度。那么

$$Q(\varepsilon)=\Omega(1/\varepsilon)\ . \tag{1.3}$$

特别地，$q(\varepsilon)$ =Ω($\sqrt{1/\varepsilon}$)。

上文的“特别地”部分直接由 Goldreich–Trevisan [21] 定理得出，该定理前文已有提及。注意，式（1.3）给出的通用下界是最优的，因为当 $\mathcal{P}$ 表示不含任何边这一性质时，该界是紧的，其中 $Q(\varepsilon)=O(1/\varepsilon)$。

1.5 Ma技术贡献及与先前方法的比较 s

先前方法总结：研究测试/估计问题时，Szemer´edi 正则性引理之所以如此有用，主要原因在于，图 G 的一个 ε-正则划分能够（近似地）确定所有较小的 F 的 $ind(F,G)$ 值。因此，从宏观层面来看，估计一个图到某个遗传性质 $\mathcal{P}$ 的距离的方法是：对 G 取一个单独的 ε-正则划分（根据正则性引理，这样的划分必然存在），然后尝试使用尽可能少的边修改来调整这个划分，使得新划分能够“预测”新图 G' 中不存在图 $F\notin\mathcal{P}$ 的诱导副本。在这一阶段需要利用的一个关键“连续性”特征是：如果 G 具有一个正则划分，且该划分中各个簇之间具有特定的边密度，而我们希望修改 G，使得在新图 G' 中存在一个正则划分，其中各个簇之间的边密度平均变化 γ，那么我们就可以

通过修改（$\gamma + o(1)$）n^2 的边来实现 G。Fischer 和 Newman [16] 关键地依赖于这样一个事实：Szemer´edi 意义下的正则划分具有这种连续性。该方法[16] 之所以无效，是因为尽管正则划分的大小是常数（即仅取决于 ε），但这个常数对 ε具有塔型依赖关系。我们应指出，[16] 的一个关键新颖思想，是一种获取输入 ´的单个 Szemer Gedi 划分的密度的方法。

Hoppen 等人 [24, 25] 之所以能够改进 [16] （对于遗传的 $\mathcal{P}$），是因为他们首先注意到，要估计 ind（F， G）对所有较小的 F而言，并不需要动用 Szemer´édi 正则性引理的全部威力。相反，可以使用 Frieze 和 Kannan 的弱正则性引理[17]，其涉及的常数仅随 ε呈指数增长。他们的证明之所以给出了双指数界，主要是因为 Frieze–Kannan 正则划分似乎不具备我们在上一段提到的 Szemer´édi 划分所具有的连续性特征。为了解决这一问题，Hoppen 等人[24, 25] 提出了一种精巧的方法，在证明的某些部分使用 Frieze–Kannan 正则划分，同时在其他部分使用完全不具备正则性[4] 特征的顶点划分（有时称为 GGR 划分，名称源自[20]）。

我们的主要技术贡献：本文的主要技术贡献在于证明，Frieze-Kannan 弱正则划分“几乎”满足我们上文提到的、相对于 Szemer´edi 划分的相同连续性特征。我们证明，如果初始划分所具有的保证略强于 Frieze–Kannan 划分的保证，并且我们只要求最终得到一个通常的 Frieze–Kannan 划分，那么实际上可以高效地修改一个 Frieze–Kannan 划分。精确表述见引理 4.10，其证明依赖于一种随机舍入类型的论证。有了上述连续性特征，我们现在可以回到 Fischer–Newman 方法，并通过充分利用 Frieze–Kannan 引理，使其成为一种有效的方法。为了确保我们的证明中只产生指数级损失，我们还需要克服一个额外的障碍，即找到一种仅使用常数次查询来寻找图的 Frieze–Kannan 划分的方法。在这里，我们引入了 Fischer–Newman 方法的一种变体，专门用于 Frieze–Kannan 划分，见引理 2.4。我们为证明定理 1.4 所开发的主要工具，也适用于证明定理 1.9。定理 1.9 中出现双指数级损失的原因在于，我们不仅需要估计单个 F 的 $ind(F, G)$（正如借助引理 1.1 在定理 1.4 中所做的那样），还需要控制所有阶为 $q(\varepsilon)$ 的图 F 的 $ind(F, G)$。我们预计，引理 2.4 和引理 4.10 将适用于未来有关图性质高效测试与估计的研究。

论文概述：第 2 节介绍本文的两个主要引理，并说明它们如何推出定理 1.4。这些引理将在第 3 节和第 4 节中证明。第 5 节证明定理 1.9。我们在第 2 节末证明命题 1.7，并在第 5 节末证明命题 1.10。我们用 $a = \text{poly}(x)$ 表示 a 被某个固定的 d从上方（当 $0 < x < 1$ 时则从下方）界定，其中该固定量与 n 或 ε无关。此外，当我们说“对于每个 a =poly(x)，都存在一个 $b = \text{poly}(x)$”时，意思是对于每个 d，都存在一个 d'，使得如果 $a \le x^d$，那么就存在一个 $b \le x^{d'}$。

[4]采用没有正则性要求的划分，其优点是它们显然具有连续性。事实上，如果我们想改变两个集合 A、B 之间的边密度 γ，只需添加/删除 $\gamma|A||B|$ 条边即可。不言而喻，使用这类划分也有各种缺点，原因是它们缺乏正则性特征。

2 关键引理及定理1.4的证明

本节的目标是陈述引理 2.4 和2.5 ，然后用它们推导定理 1.4。我们将在第 3 节和第 4 节证明这些引理。本节末尾我们还将证明命题 1.7。

为陈述引理2.4和引理2.5，我们需要一些定义。首先回顾，给定一个图 $G=(V,E)$， $V(G)$的等分划分 $A=\{V_1,\dots,V_k\}$ 是满足 $||V_i|-|V_j||\le 1$ 的划分。给定一个图 G 及其子集 X、$Y\subseteq V(G)$，我们用 $e(X,Y)$ 表示 X 与 Y 之间边的数量，并用 $d(X,Y)=e(X,Y)/|X||Y|$ 表示它们之间的密度。

定义 2.1（签名）。对于 $V(G)$ 的等分划分 $A=\{V_1,\dots,V_t\}$， A 的 (γ,ε)-签名是一个实数序列 $S=(\eta_{i,\ j})_{1\le i<j\le t}$，满足 $|d(V_i,V_j)-\eta_{i,\ j}|\le\gamma$，且对所有的 $i<j$，至多有 $\varepsilon\binom{t}{2}$ 个这样的对不满足条件。(γ,γ)-签名称为 γ-签名。

定义 2.2（划分的指标）。对于将图 $V(G)$ 划分为 t 个集合的等分划分 A，我们将 A 的指标定义为

$$ind(A)=\frac{1}{t^2}\sum_{1\le i<j\le t}d^2(V_i,V_j)\,.$$

定义 2.3（最终划分）。对于函数 $f\colon\ \mathbb{N}\to\mathbb{N}$ 和 $\gamma>0$，若由 t 个集合组成的 G 的等分划分A 是 (f,γ)-最终的，则表示不存在 $V(G)$ 的等分划分 B，其集合数至少为 t 且至多为 $f(t)$，并且满足 $ind(B)\ge ind(A)+\gamma$。

上述最终划分的概念很有用，因为（正如我们稍后将证明的）每个图都有这样的划分；此外，我们还可以设计一种算法，为输入图找到这样一个划分的签名 G。引出定理 1.4 证明的第一个关键引理正是做到了这一点。

引理 2.4。对于每个 k、$\zeta>0$，以及每个 $\gamma=\text{poly}(\zeta)$ 和 $f_\zeta(x)=x\cdot 2^{\text{poly}(1/\zeta)}$，都存在 $q=q_{2.4}(\zeta,k)$、$N=N_{2.4}(\zeta,k)$ 和 $T=T_{2.4}(\zeta,k)$，使得

$$q,N,T\le\text{poly}(k)\cdot 2^{\text{poly}(1/\zeta)}$$

并且满足以下条件。若 G 是一个至少有 N 个顶点的图，则存在一种算法，最多向 G 查询 q 次，并以至少$_{2\,3}$的概率计算出 G 的一个 (f_ζ,γ)-最终划分的 γ-签名，该划分至少有 k 个且至多有 T 个集合。

我们在第 3 节证明上述引理。以下是第二个关键引理，我们将在第 4 节证明。在其陈述中，我们使用了在引理 $ind(F,G)$ 陈述之前定义的概念1.1。粗略地说，它表明，拥有 G 的签名（参数良好）就足以估计 G 距离满足 $\mathcal{P}$ 的距离。

引理 2.5。对于每个 h、ε、$\delta>0$，存在 $\gamma=\gamma_{2.5}(h,\varepsilon,\delta)$、$s=s_{2.5}(h,\varepsilon,\delta)$ 和 $f_{2.5}^{(h,\varepsilon,\delta)}\colon\ \mathbb{N}\to\mathbb{N}$，使得

$$\gamma=\text{poly}(\varepsilon\delta/h),\quad s=\text{poly}(h/\varepsilon\delta),\quad f_{2.5}(x)=x\cdot 2^{\text{poly}(h/\varepsilon\delta)}$$

并且满足以下结论。对于每个由图组成的族，其中每个图至多有 h 个顶点，都存在一种确定性算法。该算法以图 G 的一个 $(f_{2.5},\gamma)$-最终划分 A 成 $t\ge s$ 个集合所对应的 γ-签名 S 作为输入；该图有 $n\ge N_{2.5}(h,\varepsilon,\delta,t)=\text{poly}(t)\cdot 2^{\text{poly}(h/\varepsilon\delta)}$ 个顶点。对于给定的任意 α，算法可区分以下两种情况：

(i) G 与某个图 G' 的距离为 $(\alpha-\varepsilon)$，其中对每个 $H\in\mathcal{H}$，都有 $ind(H,G')=0$.

(ii) G 与每个满足以下条件的 G' 都相距至少 α：对于每个 $H\in\mathcal{H}$， $ind(H,G')<\delta$。

证明（定理 **1.4**）。假设 $\mathcal{P}$ 是遗传图性质，并令 α，$\varepsilon > 0$。引理 1.1以 $\varepsilon/2$ 和 $\mathcal{P}$ 为输入，断言存在

$$h = M_{1.1}(\varepsilon/2), \quad \delta = \delta_{1.1}(\varepsilon/2), \quad n_0 = n_{1.1}(\varepsilon/2)\,,$$

因此，如果一个至少有 n_0 个顶点的图 G 与 $\mathcal{P}$ 相距 $\varepsilon/2$-远，那么对于某个满足$|V(H)| \le h$的 $H \notin \mathcal{P}$， $ind(H,G) \ge \delta$。我们需要描述一种算法，它对 G 进行 $2^{\mathrm{poly}(h/\delta,\log n_0)}$ 次查询，并以至少 $2/3$ 的概率区分以下两种情况： G 与 $\mathcal{P}$ 是 $(\alpha-\varepsilon)$-近的；以及 G 与 $\mathcal{P}$相距α-远。令

$$\gamma = \gamma_{2.5}(h,\varepsilon/2,\delta), \quad s = s_{2.5}(h,\varepsilon/2,\delta), \quad f = f_{2.5}^{(h,\varepsilon/2,\delta)}\,.$$

最后，令 $\zeta = \delta\varepsilon/h$，并注意到

$$\gamma = \gamma_{2.5}(h,\varepsilon/2,\delta) = \mathrm{poly}(\varepsilon\delta/2h) = \mathrm{poly}(\zeta)\,,$$

that

$$f(x) = f_{2.5}^{(h,\varepsilon/2,\delta)}(x) = x \cdot 2^{\mathrm{poly}(2h/\varepsilon\delta)} = x \cdot 2^{\mathrm{poly}(1/\zeta)}\,,$$

that

$$s = s_{2.5}(h,\varepsilon/2,\delta) = \mathrm{poly}(2h/\varepsilon\delta) = \mathrm{poly}(1/\zeta)\,.$$

此外，注意到根据命题 1.7 我们有 $\mathrm{poly}(1/\zeta) = \mathrm{poly}(h/\delta)$. 令 q, N, T 为引理 2.4 在取 $k = s$、以及 ζ, γ, f 为上述定义的参数时给出的参数。（注意， γ 和 f 满足该引理的假设。）于是，引理 2.4 保证 $q, N, T \le 2^{\mathrm{poly}(1/\zeta)} \le 2^{\mathrm{poly}(h/\delta)}$.

若 G 的顶点数少于 N，那么我们只需查询 G 的所有边，就能以概率 1 得到正确答案。此时查询次数至多为 $N^2 \le 2^{\mathrm{poly}(h/\delta)}$，符合要求。若 G 的顶点数多于 N，则可以使用引理2.4 中的算法，参数取为 k， ζ， γ， f，如上所定义。该算法至多进行 $q \le 2^{\mathrm{poly}(h/\delta)}$ 次查询，并以至少 $2/3$ 的概率返回一个 γ-签名 S，它对应于 G 的一个等分划分，划分为 $s \le t \le T$ 个集合，且该划分是（f， γ）-最终的。令

$$N' = N_{2.5}(h,\varepsilon/2,\delta,T) = \mathrm{poly}(T) \cdot 2^{\mathrm{poly}(h/\varepsilon\delta)} = 2^{\mathrm{poly}(h/\delta)}\,.$$

同样，若 G的顶点数少于 $N_1 = \max\{N', n_0\}$，那么我们只需查询 G的所有边，就能以概率 1 得到正确答案。此时查询次数至多为 $(N_1)^2 \le 2^{\mathrm{poly}(h/\delta,\log n_0)}$，符合要求。

那么，假设 G 至少有 $\max\{N, N_1\}$ 个顶点。令 $\mathcal{H}$ 为顶点数至多为 h 且不满足 $\mathcal{P}$ 的图族。现在，我们可以针对签名 S，相对于 $\mathcal{H}$，使用 $\alpha' = \alpha - \varepsilon/2$ 和 $\varepsilon/2$（而非 ε）运行引理 2.5 的算法（注意，我们选择参数时满足 $\varepsilon/2$）。如果算法判定情形（i）成立（即， G 与某个满足对每个 $H \in \mathcal{H}$ 均有 $ind(H,G') = 0$ 的 G' 是 $(\alpha'-\varepsilon/2)$-近似的），那么我们就判定 G 与 $\mathcal{P}$ 是 $(\alpha-\varepsilon)$-近似的；如果算法判定情形（ii）成立（即， G 与每个满足对每个 $H \in \mathcal{H}$ 均有 $ind(H,G') < \delta$ 的 G' 都相距至少 α'），那么我们就判定 G 与 $\mathcal{P}$ 相距至少 α。

让我们证明上述算法的正确性。如果 G 与 $\mathcal{P}$ 是 $(\alpha-\varepsilon)$-接近的，那么它与满足对每个 $H \in \mathcal{H}$ 都有 $ind(H,G') = 0$ 的图 G' 是 $(\alpha-\varepsilon)$-接近的。由于 $\alpha - \varepsilon = \alpha' - \varepsilon/2$，算法会判定情况 ($i$) 成立，因此在这种情况下算法会给出正确答案。现在假设 G与 $\mathcal{P}$ 相距 α。那么，任何与 G 是 α'-接近的 G'，都必定与 $\mathcal{P}$ 相距 $\varepsilon/2$。因此，根据引理 1.1，在任何这样的 G' 中，至少存在一个 $H \in \mathcal{H}$，使得 $ind(H,G') \ge \delta$。我们得出结论： G 与每个满足对所有 $H \in \mathcal{H}$ 都有 $ind(H,G') < \delta$ 的 G' 都相距 α'。因此，引理 2.5 中的算法会判定情况 (ii) 成立，所以我们的算法在这种情况下也会给出正确答案。 □

证明（命题 1.7）：回顾一下，图 H 在 h 个顶点上的膨胀，是通过将 H 的每个顶点 $i \in V(H)$ 替换为一个由顶点 S_i 构成的独立集，并将每条边（i，j）替换为 S_i 与 S_j 之间的完全二分图而得到的图。b-膨胀是指每个 S_i 的大小均为 b 的膨胀。假设 H_b 是 H 的一个 b-膨胀，且 f：$[h] \to \{0,1\}$ 是对 H 的顶点的一个 $0/1$ 赋值。那么，H_b^f 是通过取 H_b，然后当且仅当 $f(i) = 1$ 时将每个顶点集 S_i 变成一个团而得到的图。

我们称 H 是好的，如果存在一个 $b = b(H)$，使得对于上述任意 f，都有 $H_b^f \notin \mathcal{P}$。我们首先注意到，如果 $H \notin \mathcal{P}$，那么 H 是好的，因为我们可以取 $b = 1$。我们还注意到，单个顶点不可能是好的，因为如果一个顶点是好的，那么必然存在一个团和一个独立集，它们不满足 $\mathcal{P}$；由拉姆齐定理可知，这意味着所有足够大的图都不属于 $\mathcal{P}$，这与我们关于 $\mathcal{P}$ 非平凡的假设矛盾。

设 H 是一个不满足 $\mathcal{P}$ 的图（这样的图存在，因为 $\mathcal{P}$ 非平凡）。根据上一段，H 是好的。如果 H 在 $|V(H)| - 1$ 个顶点上的某个诱导子图也是好的，那么就用这个诱导子图替换H。设 H 是最终得到的（极小）好图。根据上一段，我们知道 $h \geq 2$。令 H' 为从H中删除顶点 h 后得到的图。那么 H' 不是好的。

固定 $0 < \varepsilon < \frac{1}{10|H|}$ 并取足够大的 n，令 G_n 为 H 的膨胀，其中替换顶点 $h \in V(H)$ 的顶点集记为 V_h，其大小为 εn；其余所有 $h-1$ 集记为 V_1、……、V_{h-1}，大小均相同 $(n - \varepsilon n)/(h-1)$。由于 H' 不是好的，我们知道，给定 $b = (n - \varepsilon n)/(h-1)$，存在一个 f'：$[h-1] \to \{0,1\}$，使得 $H_b^f \in \mathcal{P}$。对于每个 $1 \leq i \leq h-1$，当且仅当 $f(i) = 1$时，我们将 V_i 变成一个团。注意，G_n 的每个不含 V_h 中顶点的诱导子图都满足 $\mathcal{P}$。

设 $b_0 = b(H)$ 为好的图的定义中的常数，并设 $r(b_0)$ 为 b_0的拉姆齐数，即使每个有 r 个顶点的图都包含一个有 b_0 个顶点的团或独立集的最小整数。现在我们断言，G_n 与 $\mathcal{P}$相距至少 ε/C，其中 $C = 2h^4r^2$。事实上，我们断言，如果在集合 V_1、……、V_h 之间改变少于 $\varepsilon n^2/C$ 条边，那么（无论在集合 V_1、……、V_h内部做何种改变）所得图都不满足 $\mathcal{P}$。由于 $\mathcal{P}$ 是遗传的，只需证明仍存在一个不满足 $\mathcal{P}$的诱导子图。具体而言，考虑一个由顶点 v_1、……、$v_{h\cdot r}$ 组成的 $(h \cdot r)$ 元组：对每个 $1 \leq i \leq h$，从 V_i 中均匀随机地选取一个含 r个顶点的集合。固定 $i < j$。由于 $|V_i||V_j| \geq \varepsilon n^2/h^2$，我们至多修改了 V_i与 V_j之间的点对中占 $\frac{1}{2r^2h^2}$的比例。因此，样本中包含一对被修改顶点的概率至多为 $\frac{1}{2h^2}$。于是，根据并集界，样本中在某对 V_i、V_j 之间包含被修改顶点对的概率至多为 $1/2$。由此可知，存在一种选取 $h \cdot r$ 个顶点的方式，使得它们所诱导的图是 H的一个 r 膨胀。根据 r的选取方式，我们可以在该集合中找到 H 的一个 b_0膨胀，使得每个含 b_0 个顶点的集合都是一个团或独立集。由于 H 是好的，这意味着该图不满足 $\mathcal{P}$。

现在，正如我们在引言中指出的（见方程（1.1）），根据 M 和 δ的定义，M/δ 个顶点的样本以至少 $1/2$的概率包含一个不属于 $\mathcal{P}$的图。正如我们先前指出的，G_n 中不含来自 V_h 的顶点的每个子图都满足 $\mathcal{P}$。但是，要以至少 V_h 的概率命中$1/2$，必须至少采样 $1/\varepsilon$ 个顶点。这意味着我们必须有 $M/\delta \geq 1/\varepsilon$。

□

3 引理2.4的证明

该证明与 [16]中的一个证明类似。他们证明了，对于任意 f，γ，都存在一个(f，γ)-最终划分，其查询复杂度为一个常数，尽管这个常数是巨大的塔型。我们在这里证明，对于受限类型的 f，可以得到好得多的上界。为此，我们还需要依赖 [34]的一项近期结果。

3.1 预备引理

本小节将介绍一些预备引理，它们将在下一小节中用于证明引理2.4。我们需要以下 Chernoff 型大偏差不等式。

引理 3.1. 假设 $X_1,\dots,X_m$ 是 m 独立布尔随机变量，使得对于每个 $1 \le i \le m$，都有 $\Pr[X_i = 1] = p_i$。令 $E = \sum_{i=1}^m p_i$。那么，$\Pr[|\sum_{i=1}^m X_i - E| \ge \theta m] \le 2e^{-2\theta^2 m}$。

定义 3.2.（划分性质）划分性质是一个三元组 $\pi = (s, \ell, u)$，其中 s 是一个整数（划分性质的大小），ℓ 是一个包含 $\binom{s}{2}$ 个实数的向量 $0 \le \alpha_{i,j} \le 1$ 对每个 $1 \le i < j \le s$都成立，并且 u 是一个包含 $\binom{s}{2}$ 个实数的向量 $0 \le \beta_{i,j} \le 1$ 对每个 $1 \le i < j \le s$都成立。我们称图 G 满足 π，如果存在 $V(G)$ 的一个等分划分 $\{V_1, \dots, V_s\}$，使得对每个 $1 \le i < j \le s$，都有 $\alpha_{ij} \le d(V_i, V_j) \le \beta_{ij}$。

给定 s 和 μ，我们用 $\pi(s,\mu)$ 表示大小为 s 的划分性质族 π，其中每个$\alpha_{i,\ j}$和 $\beta_{i,\ j}$都是 μ 的整数倍（因此 $\pi(s,\mu)$ 包含 $\{0, \mu, 2\mu, \dots, 1\}^2\binom{s}{2}$ 个划分性质）。最后，定义 $\Pi(t,\mu) = \bigcup_{s \le t} \pi(s,\mu)$

注意，上述每个 π 都是[20], 中研究的划分性质，其中证明了它们是 μ-可测试的，查询复杂度为 $(1/\mu)^{\mathrm{poly}(s)}$。最近，这一结果被改进为 poly$(s/\mu)$，见[34]。下一个引理指出，实际上我们可以用（大致）相同的查询复杂度同时测试 $\Pi(t,\mu)$中的所有性质。

引理 3.3. 对于每个 t 和 $\mu > 0$，存在 $q = q_{3.3}(t,\mu) = \mathrm{poly}(t/\mu)$，满足如下性质：存在一个随机化算法，对于给定的图 G，向 G 进行 q 次查询，并以至少 $2/3$的概率，对于每个 $\pi \in \Pi(t,\mu)$，区分以下两种情况：G 满足 π，以及 G与 π相距至少 μ。

证明。[34] 的一项结果表明，每个 $\pi \in \Pi(s,\mu)$ 都是 μ-可测试的，且查询复杂度为 $q' = \mathrm{poly}(s/\mu)$。令 $b = |\Pi(t,\mu)| \le (1/\mu)^{t^2}$。固定 $\pi \in \Pi(t,\mu)$。如果我们执行 μ-测试算法 $10\log(b)$ 次，然后取多数结果[5], ，那么通过对引理 3.1 的标准应用，我们得到一个新算法，它进行 $\log b \cdot \mathrm{poly}(s/\mu) \le \log b \cdot \mathrm{poly}(t/\mu)$ 次查询，能够区分 G 满足 π 的情形与 G 和 π之间的距离至少为 μ 的情形，且错误概率至多为 $1/3b$。注意，为了做到这一点，我们可以抽取一个大小为 $q' \cdot 10\log(b)$ 的集合 Q，然后在 $10\log(b)$ 个互不相交、大小为 q' 的集合上运行标准的 μ测试器，以测试 π（从而保证 $10\log(b)$ 次迭代之间完全独立）。由于对每个 π，随机集合 Q 使得算法的错误概率至多为 $1/3b$，由并集界可知，算法对某个 $\pi \in \Pi(t,\mu)$ 出错的概率至多为 $1/3$。最后，该算法的查询复杂度为 $\log b \cdot \mathrm{poly}(t/\mu) = \mathrm{poly}(t/\mu)$。

□

（引理 **2.4**的证明）：给定 k，ζ，γ 以及 f_ζ（如引理陈述所述），我们定义 $T_0 = k$并对 $i \ge 1$ 定义 $T_i = f_\zeta(T_{i-1})$。现在设定以下参数。

$$N = N_{2.4}(k,\zeta) = T_{2/\gamma} = k \cdot 2^{\mathrm{poly}(1/\zeta)}, \quad T = T_{2.4}(k,\zeta) = T_{2/\gamma} = k \cdot 2^{\mathrm{poly}(1/\zeta)},$$

and

$$t = f_\zeta(T) = k \cdot 2^{\mathrm{poly}(1/\zeta)}, \quad \mu = \frac{\gamma}{48(f_\zeta(T))^2} = \frac{1}{\mathrm{poly}(k) \cdot 2^{\mathrm{poly}(1/\zeta)}}.$$

现在我们描述寻找满足引理要求的签名的算法。S 以下令 $\pi'(s,\ \mu)$ 为划分性质，其中$\beta_{i,\ j} = \alpha_{i,\ j} + \mu$ 对每个 $1 \le i < j \le s$均成立。此外，对每个 $\pi \in \pi'(s,\ \mu)$，定义 π 的指标为 $ind(\pi) = \frac{1}{t^2}\sum_{1 \le i < j \le t} \alpha_{ij}^2$。在步骤 1 中，我们运行引理3.3 的算法，使用上文定义的参数 t，μ。这是该算法中唯一的随机化部分。在算法的步骤 2 中，我们执行以下操作。

[5]也就是说，我们正在使用标准错误缩减技巧。

(*i*) 对每个 $k \le s \le t$ 集合 $M(s) = \max_\pi ind(\pi)$，其中最大值取遍步骤 1 的算法接受的所有 $\pi \in \pi'(s, \mu)$。

(*ii*) 令 $s^\star$ 为 $\{k, \dots, T\}$ 中满足以下条件的最小数：对于每个 $s' \in \{s^\star + 1, \dots, f_\zeta(s^\star)\}$，$M(s') \le M(s^\star) + \frac{3}{4}\gamma$。如果存在这样的 $s^\star$，则输出在 $s^\star$ 上取得最大值的签名 $S^\star$。否则，算法失败。

请注意，该算法的查询复杂度为 $q = q_{3.3}(t, \mu) = \mathrm{poly}(t/\mu) = \mathrm{poly}(k) \cdot 2^{\mathrm{poly}(1/\zeta)}$，这正是我们所需要的。此外，引理 3.3 保证上述算法描述中的步骤 1 至少以 $2/3$ 的概率成功。因此，接下来只需证明，在假设这一事件发生的情况下，算法的步骤 2 将返回一个 (f_ζ, γ)-最终划分。首先需要注意的是，如果算法成功，它返回的划分大小至少为 k，且至多为 T，这正符合要求。

如果步骤 1 成功，那么步骤 2 返回一个 (f_ζ, γ)-最终划分的证明，与 [16]，中断言 5.5 的证明相同，因此我们简要概述一下证明。首先，读者可能会想知道，为什么每个图都必然具有引理陈述中所说的 (f_ζ, γ)-最终划分。下面我们就用之前引入的定义，说明为什么每个 G 都有一个 $(f_\zeta, \gamma/2)$-最终划分。从 G 的任意一个等分划分 A_0（将其划分为 $T_0 = k$ 个集合）开始，并令 $ind_0 = ind(A_0)$ 表示定义 2.2 中的 A_0 的指标。如果 A_0 是 $(f_\zeta, \gamma/2)$-最终划分，那么证明完毕。否则，必定存在 G 的另一个划分 A_1，其部分数至少为 T_0 且至多为 $f(T_0)$，并且指标为 $ind(A_1) \ge ind(A_0) + \gamma/2$。由于对于每个等分划分，$0 \le ind(A) \le 1$，我们可以看出，这个过程最终会得到一个大小为 $k \le s \le T$ 的划分 A，使得 G 的所有划分——部分数至少为 s 且至多为 f({v37)——其指标都小于 $ind(A) + \gamma/2$。但这意味着 A 是 $(f_\zeta, \gamma/2)$-最终划分。因此，我们得到 G 有一个大小为 $s \le T$ 的 $(f_\zeta, \gamma/2)$-最终划分 A。

现在我们来解释如何将上述存在性证明转化为前面所述算法正确性的证明。令 $M_G(s)$ 表示大小为 s 的 G 的等分划分中最大的指标。首先，我们声称对于每个 $k \le s \le t$,

$$M(s) - \gamma/8 \le M_G(s) \le M(s) + \gamma/8. \tag{3.1}$$

对于 (3.1) 中的第二个不等式，令 A (s) 为一个有 s 个部分的等分划分，满足 $M_G(s) = ind(A)$。令 $\pi \in \pi'(s, \mu)$ 为由 A 得到的划分性质，其中将密度向下舍入到最接近的 μ 的整数倍。于是有 $|ind(A) - ind(\pi)| \le 3\mu \le \gamma/8$。因此，$M(s) \ge ind(\pi) \ge ind(A) - \gamma/8 = M_G(s) - \gamma/8$。

对于 (3.1) 中的第一个不等式，令 $\pi \in \pi'(s, \mu)$ 为算法接受的一个划分性质，并满足 $M(s) = ind(\pi)$。那么 G 必须与 π 相差不超过 μ（否则 π 本应被拒绝）。令 G' 为一个与 G 相差不超过 μ 且满足 π 的图，并令 A 为 G' 的顶点划分，作为 G' 满足 π 的见证。注意，将 G 转换为 G' 时，对于 A 的每一对部分，我们至多改变这对部分之间的密度 μs^2。因此，在 G 中，划分性质 π 是 A 的一个 $2\mu s^2$-签名（此处及下文中，我们将 π 视为一个签名）。因此，$|ind(A) - ind(\pi)| \le 6\mu s^2 \le \gamma/8$，这是根据我们对 μ 的选择得出的。现在，$M_G(s) \ge ind(A) \ge ind(\pi) - \gamma/8 = M(s) - \gamma/8$。这就证明了 (3.1)。

由上述存在性证明可知，存在 $k \le s^\star \le T$ 以及将 G 等分划分为 $s^\star$ 个部分的等分划分 A，该等分划分是 $(f_\zeta, \gamma/2)$-终止的。我们可以假设 $M_G(s^\star) = ind(A)$，因为满足这一条件的等分划分也必然是终止的。对于每个 $s^\star \le s' \le f_\zeta(s^\star)$，都有 $M_G(s') \le M_G(s^\star) + \gamma/2$。根据（3.1），这意味着对于每个 $s^\star \le s' \le f_\zeta(s^\star)$，都有 $M(s') \le M(s^\star) + 3\gamma/4$。因此，该算法将返回一个划分。

请注意，该算法不一定会返回与上述相同的签名/划分性质，即与上述划分 A 距离不超过 μ 的结果。算法选择不同划分的原因在于，可能存在另一个大小为 s 、指标更大的划分（当然，它同样是 (f_ζ, γ)-最终的），也可能存在一个具有相同性质的 $s^* < s$，或者还可能存在其他情况。

具有相同指标的划分。不过，可以反过来运用上一段中的推理，并证明如果返回了 π，那么它必定是某个 (f_ζ, γ)-最终划分的 γ-签名。 □

引理2.5 的证明

4.1 预备引理

本小节介绍一些预备引理，这些引理将在下一小节中用于证明引理 2.5。我们首先介绍 Frieze–Kannan 正则性引理 [17, 18]。我们先陈述其 γ-正则性的概念。

定义 4.1（Frieze–Kannan 正则性[18]）。设 $G = (V, E)$ 是一个图，$A = \{V_1 \dots, V_k\}$是 $V(G)$ 的一个等分划分。对于子集 $X \subseteq V$和 $1 \le i \le k$，记 $X_i = X \cap V_i$。我们称A 是 γ-Frieze–Kannan 正则的，当且仅当：

$$d^A_\square(G) := \max_{S,T\subseteq V} \frac{1}{n^2}\left|\sum_{i,j\in[k]^2}\Big(d(S_i,T_j) - d_{ij}\Big)|S_i||T_j|\right| < \gamma \tag{4.1}$$

粗略地说，如果我们能够根据交集大小 $S \cap V_i$ 和 $T \cap V_i$估计大集合 S, T之间的边数，那么划分 A 就是 γ-Frieze–Kannan 正则的，简称 γ-FK-正则。我们还需要以下稍强一些的弱正则性概念，该概念由[29]提出。

定义 4.2 (Frieze–Kannan 正则性* [29])。 在定义 *4.1*的设定下，我们称 A 是γ-Frieze–Kannan 正则的*，如果：

$$d^{\star A}_\square(G) := \max_{S,T\subseteq V} \frac{1}{n^2}\sum_{i,j\in[k]^2}\Big|d(S_i,T_j) - d_{ij}\Big||S_i||T_j| < \gamma \tag{4.2}$$

这两个概念之间的转换对于下面的引理 4.10 至关重要。设 $A = \{V_1, \dots, V_k\}$ 是 $V(G$的一个等分划分。那么，$B = \{W_1, ..., W_\ell\}$ 对 $V(G$的一个等分划分称为细化 A，如果每个 $W_i \in B$ 都包含于某个 $V_j \in A$中。下面的引理见[29]，其证明采用了 Frieze 和 Kannan 原始证明的一个简单变体[18]。

引理 4.3（Frieze–Kannan 弱正则性引理 [18],[29]）。对于每个 k_0 和 $\gamma > 0$，存在$T = T_{4.3}(k_0, \gamma) = k_0 \cdot 2^{\mathrm{poly}(1/\gamma)}$，使得对于每个至少有 T 个顶点的图 G，以下结论成立。如果 A 是 $V(G)$的一个等分划分，且划分为至多 k_0 个集合，那么存在一个将 A 细化为至多T 个集合的细化 B，使得 $d^{\star B}_\square(G) < \gamma$。

现在，我们将 $d_\square$ 的定义扩展到加权图对之间的距离，其中加权图 R 是一个完全图，因此每条边 (i, j) 都被赋予一个权重 $0 \le R(i, j) \le 1$。如果 R, R' 是具有 n 个顶点的两个加权图，那么我们定义

$$d_1(R, R') = \frac{1}{n^2}\sum_{i<j}|R(i,j) - R'(i,j)|\,, \tag{4.3}$$

and

$$d_\square(R, R') = \max_{\alpha,\beta}\frac{1}{n^2}\left|\sum_{i<j}\alpha(i)\beta(j)(R(i,j) - R'(i,j))\right|\,, \tag{4.4}$$

其中，最大值取遍所有函数 $\alpha, \beta\colon [n] \to [0, 1]$。

定义 4.4（$ind(F,R)$）。设 R 是一个定义在 $[k]$ 上的加权图，且 φ 是一个单射函数$\varphi\colon V(F)\to[k]$。我们定义

$$ind_\varphi(F,R)=\prod_{i<j\in E(F)}R(\varphi(i),\varphi(j))\prod_{i<j\notin E(F)}(1-R(\varphi(i),\varphi(j)))$$

在 φ 非单射的情况下，我们定义

$$ind_\varphi(F,R)=0$$

记 Φ 为从 $V(F)$ 到 $[k]$ 的函数集合，我们定义

$$ind(F,R)=\frac{1}{|\Phi|}\sum_{\varphi\in\Phi}ind_\varphi(F,R)\,. \tag{4.5}$$

注意，我们可以将一个签名 $S=(\eta_{i,\,j})_{1\le i<j\le t}$ 视为一个具有 t 个顶点的加权图。这意味着对于一对签名 S、S' 我们可以分别按照（$d_1(S,\,S')$）和（$d_\square(S,\,S')$）中的定义定义 $d_1(S,\,S')$ 和 $d_\square(S,\,S')$，还可以按照（4.5）定义 $ind(F,\,S)$。我们将需要以下来自[24]

引理 4.5. 假设 R、R' 是定义在 n 个顶点上的两个加权图，且 H 是定义在 h 个顶点上的一个图。那么，对于任意 $\gamma\ge d_\square(R,R')$ 和 $n\ge\frac{2}{\gamma}$，我们有 $|ind(H,R)-ind(H,R')|\le 2h^2\cdot\gamma$

证明。我们定义一个与 $ind(F,R)$ 类似的概念，但也考虑非单射函数。对于每个 $\varphi\in\Phi$（我们在此前定义了 Φ（4.5））我们令

$$ind'_\varphi(F,R)=\prod_{i<j\in E(F)}R(\varphi(i),\varphi(j))\prod_{i<j\notin E(F)}(1-R(\varphi(i),\varphi(j)))\,,$$

and

$$ind'(F,R)=\frac{1}{|\Phi|}\sum_{\varphi\in\Phi}ind'_\varphi(F,R)\,.$$

[24] 中的引理 3.2指出，若 R、R' 是定义在 n 个顶点上的两个加权图，且 H 是定义在 h个顶点上的图，则 $|ind'(H,R)-ind'(H,R')|\le h^2\cdot d_\square(R,R')$。现在取一些 $\gamma\ge d_\square(R,R')$，并假设 $n\ge\frac{2}{\gamma}$。根据伯努利不等式（$n>h$），我们有

$$|ind'(H,R)-ind(H,R)|\le 1-\frac{(n-h)^h}{n^h}=1-\left(1-\frac{h}{n}\right)^h\le\frac{h^2}{n}\,.$$

因此，

$$|ind(H,R)-ind(H,R')|\le|ind'(H,R)-ind'(H,R')|+\frac{2h^2}{n}\le h^2\cdot d_\square(R,R')+h^2\gamma\le 2h^2\cdot\gamma\,,$$

如所期望。 □

给定一个有 n 个顶点的图 G，以及一个等分划分 $A=\{V_1,\dots,V_k\}$，我们将有 V 个顶点的图 $G_A(G)$ 定义为这样一个加权图：对于每个 $u\in V_i$ 和 $v\in V_j$，其权重为 $G_A(u,v)=d(V_i,V_j)$。令 S_A 为 A 的 0-签名，也就是说，它是一个有 k 个顶点的加权图，其权重为 $S(i,j)=d(V_i,V_j)$。注意，若 k 整除 n（因此 A 中所有集合的大小相同），则 $ind(H,G_A)$ 与 $ind(H,S_A)$ 几乎相同。不难看出，对于一般的等分划分，这些量之间的差异并不大。

引理 4.6。给定一个有 n 个顶点的图 G，以及一个等分划分 $A=\{V_1,\dots,V_k\}$，设 G_A 和 S_A 如上所定义。则对于每个有 h 个顶点的图 H，都有 $|ind(H,G_A)-ind(H,S_A)|\le\frac{2h^2}{k}+\frac{2kh}{n}$。

证明。我们使用 $ind'(F,R)$ 在引理 4.5的证明中引入的定义。由 [24] 中的不等式 (5)，我们有 $|ind'(H,G_A)-ind'(H,S_A)|\le\frac{2kh}{n}$。注意，按照上述定义，与引理 4.5的证明一样，

$$|ind'(H,G_A)-ind(H,G_A)|\le\frac{h^2}{n}\text{ and }|ind'(H,S_A)-ind(H,S_A)|\le\frac{h^2}{k}\,.$$

因此，我们得到

$$|ind(H,G_A)-ind(H,S_A)|\le\frac{2h^2}{k}+\frac{2kh}{n}\,,$$

正如所需。 □

现在结合以上事实，可以得出：图的 γ-FK 划分的签名可为 $ind(H,G)$ 提供良好的近似。

引理 4.7。对于每个 h、k 和 $\delta>0$，存在

$$\gamma=\gamma_{4.7}(h,\delta)=\text{poly}(\delta/h),\quad r=r_{4.7}(h,\delta)=\text{poly}(h/\delta),\quad N=N_{4.7}(h,k,\delta)=\text{poly}(hk/\delta)\,,$$

使得若 G 是一个至少有 N 个顶点的图，且 A 是 G 的一个 γ-FK-正则划分，具有至少 r 个且至多 k 个部分，那么对于 A 的每个 γ-签名 S，我们有 $|ind(H,G)-ind(H,S)|\le\delta$，对每个具有 h 个顶点的 H 均成立。

证明。我们设 $\gamma=\frac{\delta}{8h^2}$，$r\ge2/\gamma$ 并设 $N=$最大值$\{2/\gamma,\frac{2k^2}{h},\frac{16kh}{\delta}\}$。令 G_A 和 S_A 如引理 4.6 之前所定义。如果我们将 G 视为一个带权 0/1 图（即 $G(x,y)=1$ 当且仅当 $(x,y)\in E(G)$ 时），那么我们有

$$\begin{aligned}
d_\square(G,G_A)&=\max_{\alpha,\beta}\frac{1}{n^2}\left|\sum_{x,y\in V(G)}\alpha(x)\beta(y)(G(x,y)-G_A(x,y))\right|\\
&=\max_{S,T\subseteq V}\frac{1}{n^2}\left|\sum_{i,j\in[k]^2}\sum_{x\in S_i,y\in T_j}(G(x,y)-d_{ij})\right|\\
&=\max_{S,T\subseteq V}\frac{1}{n^2}\left|\sum_{i,j\in[k]^2}e(S_i,T_j)-d_{ij}|S_i||T_j|\right|\\
&=\max_{S,T\subseteq V}\frac{1}{n^2}\left|\sum_{i,j\in[k]^2}(d(S_i,T_j)-d_{ij})|S_i||T_j|\right|\\
&=d^A_\square(G)\le\gamma
\end{aligned}$$

其中，在第二个等式中，我们用到了最大值总能由取布尔值的 α,β 达到这一事实。因此，根据引理 4.5（应用于上文定义的 γ；注意 $N\ge2/\gamma$），我们可以推得，对于每个 H，都有 $|ind(H,G)-ind(H,G_A)|\le2h^2\cdot\gamma$。根据引理 4.6，对于具有 h 个顶点的每个 H，都有 $|ind(H,G_A)-ind(H,S_A)|\le\frac{2h^2}{k}+\frac{2kh}{n}$。根据我们对 r、γ 和 N 的选取，有 $|ind(H,G_A)-$

[6]事实上，不妨假设最大值为正。那么，如果增加每个 $\alpha(x)$ 会使结果增加，我们可以将其四舍五入为 1。其余的我们都可以四舍五入为 0。然后，我们可以针对 β进行相同的舍入过程。

$ind(H,S_A)| \le \delta/4$。因此，根据三角不等式以及我们对 γ 的选取，有 $|ind(H,G) - ind(H,S_A)| < \delta/2$。最后，由于 S 是 A 的一个 γ-签名，我们有 $d_\square(S,S_A) \le d_1(S,S_A) \le \gamma$，因此再次应用引理 4.5（此处仍取 γ），还可得 $|ind(H,S) - ind(H,S_A)| \le \delta/2$。因此，再次应用三角不等式可得 $|ind(H,G) - ind(H,S)| \le \delta$，从而完成证明。 □

定义 4.8（扩展）。给定等分划分 A 的签名 $S = (\eta_{ij})_{1\le i<j\le t}$，以及 A 的细化 $B = \{W_1,\dots,W_s\}$，S 到 B 的扩展是序列 $S' = (\eta'_{ij})_{1\le i<j\le s}$，定义为 $\eta'_{i,j} = \eta_{k,l}$，如果存在 $k \ne l$ 使得 $W_i \subseteq V_k$ 且 $W_j \subseteq V_l$；并令 $\eta'_{i,j} = 0$，如果 W_i 和 W_j 都是同一个 V_k 的子集。

断言 4.9。对于每个 ε 和 s，存在 $r = r_{4.9}(\varepsilon) = \text{poly}(1/\varepsilon)$ 以及 $N = N_{4.9}(\varepsilon,s) = \text{poly}(s/\varepsilon)$，使得对定义在同一组 $n \ge N$ 个顶点上的任意一对图 G、G'，以下结论成立。若 G、G' 是 α- 接近的，且 S、S' 分别是 G、G' 关于顶点集的同一等分划分 A（划分为 $s \ge r$ 个集合）的 γ、γ'-签名，则 $d_1(S,S') \le \alpha + \varepsilon + 2(\gamma+\gamma')$。

证明。令 $r = 2/\varepsilon$。令 S_A、S'_A 分别为 A 关于 G、G' 的 0-签名。于是，假设 $n > N_{4.9}(\varepsilon,s) = \text{poly}(s/\varepsilon)$ 足够大，显然有 $d_1(S_A,S'_A) \le \alpha + \varepsilon$。由于 S 是 A 的一个 γ-签名，我们根据定义有 $d_1(S,S_A) \le 2\gamma$；同理可得 $d_1(S',S'_A) \le 2\gamma'$。因此，根据三角不等式，有 $d_1(S,S') \le \alpha + \varepsilon + 2(\gamma+\gamma')$。 □

引理 4.10。对于每个 ε 和 t，存在 $\gamma = \gamma_{4.10}(\varepsilon) = \text{poly}(\varepsilon)$ 和 $N = N_{4.10}(t,\varepsilon) = \text{poly}(t/\varepsilon)$，使得对于任意一个具有 $n \ge N$ 个顶点的图 G，若 S 是 G 的一个 γ-签名，其中 G 是由 t 个集合组成的 γ-FK-正则划分 A，那么对于任意满足以下条件的签名 S'：存在某个 δ 使得 $d_1(S,S') \le \delta$，都存在一个与 G $(\delta+\varepsilon)$-接近的图 G'，使得 A 是 G' 的一个 ε-FK-正则划分，且 S' 是 A 的一个 ε-签名。

证明。思路非常简单；我们随机修改 G，使得密度将成为 S'的密度。然而，要证明 A 将是 G' 的一个 ε-FK-正则划分，需要一个微妙的论证，该论证将利用以下事实：A 是 G的一个 γ-FK-正则* 划分。

我们设定 $\gamma = \gamma_{4.10}(\varepsilon) = \varepsilon/6$ 和 $N = N_{4.10}(t,\varepsilon) = 960t^3/\gamma^5$。给定引理陈述中所述的 G、$A = \{V_1,\dots V_t\}$、$S = (\eta_{ij})_{1\le i<j\le t}$和 $S' = (\eta'_{ij})_{1\le i<j\le t}$，我们通过以下过程从 G 得到 G'，其中所有随机选择均相互独立：

- 对于每个 i，V_i 内的边都保持不变。
- 对于满足 $\eta'_{i,j} < d(V_i,V_j)$ 的 $i<j$，V_i 与 V_j 之间的 G 的每条边都以概率 $1 - \frac{\eta'_{i,j}}{d(V_i,V_j)}$ 被移除。
- 对于满足 $\eta'_{i,j} > d(V_i,V_j)$ 的 $i<j$，V_i 与 V_j 之间的 G 中，每个不是边的顶点对都会以概率 $1 - \frac{1-\eta'_{i,j}}{1-d(V_i,V_j)}$ 变成一条边。

在下文中，我们用 $e'(X,Y)$ 表示 G' 中 X 和 Y之间的边数，并用 $d'(X,Y)$ 表示 G'中这些集合之间的密度。注意，我们生成 G' 的方式保证，对于每个 $i<j$，都有

$$\mathbb{E}[d'(V_i,V_j)] = \eta'_{i,j}\,. \tag{4.6}$$

我们首先证明，以至少 $1/2$ 的概率，对于每个 $i<j$ 和每个满足 $|X| \ge 2\gamma|V_i|$、$|Y| \ge 2\gamma|V_j|$ 的 $X \subseteq V_i$，都有 $Y \subseteq V_j$

$$|e'(X,Y) - \mathbb{E}[e'(X,Y)]| \le \frac{\gamma}{4}|X||Y|. \tag{4.7}$$

另请注意，这等价于

$$|d'(X,Y) - \mathbb{E}[d'(X,Y)]| \le \gamma/4 \,. \tag{4.8}$$

只需证明，对于固定的 $i<j$ 以及上述子集 $X \subseteq V_i$、$Y \subseteq V_j$，我们有

$$\Pr\left[|e'(X,Y) - \mathbb{E}[e'(X,Y)]| \ge \frac{\gamma}{4}|X||Y|\right] < \frac{2^{-3n/t}}{4t^2} < \frac{2^{-2(n/t+1)}}{4t^2} \,,$$

因为这样就可以对所有 $i<j$ 以及 X、Y 的取值应用并集界，从而得出结论。因此，从这里开始，我们固定 i、j、X、Y。我们先讨论 $d(V_i,V_j) > \eta'_{ij}$ 的情形。若 $d(X,Y) < \gamma/4$，则 $e'(X,Y) \ \mathbb{E}[e'(X,Y)] \le \frac{\gamma}{4}|X||Y|$ 均成立，因此 (4.7) 以概率 1 成立。现在假设 $d(X,Y) \ge \gamma/4$。由引理 3.1，取 $m = d(X,Y)|X||Y| \ge \gamma^3(n/t-1)^2 \ge \gamma^3\frac{n^2}{4t^2}$，并令 $\theta = \frac{\gamma}{4d(X,Y)}$，可得

$$\Pr[|e'(X,Y) - \mathbb{E}[e'(X,Y)]| > \frac{\gamma}{4}|X||Y|] < 2e^{-2\theta^2 m} < 2e^{-2(\frac{\gamma}{4d(X,Y)})^2\cdot\frac{\gamma^3 n^2}{4t^2}} \le 2e^{-\frac{1}{32}(\frac{\gamma^5 n^2}{t^2})} \le \frac{2^{-3n/t}}{4t^2} \,,$$

最后一个不等式中用到了关于 n 的假设。现在假设 $d(V_i,V_j) < \eta'_{ij}$。若 $d(X,Y) > 1-\gamma/4$，则同样有 (4.7) 以概率 1 成立。现在假设 $d(X,Y) \le 1-\gamma/4$。如上所述，由引理 3.1，取 $m = (1-d(X,Y))|X||Y| \ge \gamma^3(n/t-1)^2 \ge \gamma^3\frac{n^2}{4t^2}$，并令 $\theta = \frac{\gamma}{4(1-d(X,Y))}$，可得

$$\Pr[|e'(X,Y) - \mathbb{E}[e'(X,Y)]| > \frac{\gamma}{4}|X||Y|] < 2e^{-2\theta^2 m} < 2e^{-2(\frac{\gamma}{4(1-d(X,Y))})^2\cdot\frac{\gamma^3 n^2}{4t^2}} \le 2e^{-\frac{1}{32}(\frac{\gamma^5 n^2}{t^2})} \le \frac{2^{-3n/t}}{4t^2}$$

由于该陈述显然适用于 $d(V_i, V_j) = \eta'_{ij}$ 的情形（在这种情况下我们什么也不做），因此我们已经证明，对于所有 i，j 以及满足 $|X| \ge 2\gamma|V_i|$，$|Y| \ge 2\gamma|V_j|$ 的每个 $X \subseteq V_i$，$Y \subseteq V_j$，(4.8) 以至少 $1/2$ 的概率成立。现在我们将证明这一事实蕴含了该引理的所有断言。

我们首先注意到，在（$X = V_i$ 和 $Y = V_j$ 中使用（4.8）可知，对于每个 $i<j$，都有

$$|d'(V_i,V_j) - \mathbb{E}[d'(V_i,V_j)]| = |d'(V_i,V_j) - \eta'_{i,j}| \le \gamma/4 \,, \tag{4.9}$$

其中，第一个等式为（4.6）。

我们现在声称，对于每个 $i<j$ 和 $X \subseteq V_i$、$Y \subseteq V_j$，我们都有

$$|\mathbb{E}[d'(X,Y)] - \eta'_{i,j}| \le |d(X,Y) - d_{i,j}| \tag{4.10}$$

其中我们使用 $d_{i,j} = d(V_i,V_j)$。显然，当 $d_{i,j} = \eta'_{i,j}$ 时，这一点成立，因此首先假设 $d_{i,j} > \eta'_{i,j}$。令 $q_{ij} = \eta'_{i,j}/d_{i,j}$，我们看到，$V_i$ 和 V_j 之间的每条边都以 $q_{i,j}$ 的概率保留在 G' 中，因此

$$|\mathbb{E}[d'(X,Y)] - \eta'_{i,j}| = |q_{ij}\cdot d(X,Y) - \eta'_{i,j}| = q_{ij}|d(X,Y) - d_{i,j}| \le |d(X,Y) - d_{i,j}| \,.$$

类似地，如果 $d_{i,j} < \eta'_{i,j}$，那么设定 $q'_{ij} = 1 - p'_{ij} = (1-\eta'_{i,j})/(1-d_{i,j})$，我们看到，$V_i$ 与 V_j 之间缺失的每条边都会以概率 $p'_{i,j}$ 加入到 G' 中，因此

$$\begin{aligned}
|\mathbb{E}[d'(X,Y)] - \eta'_{ij}| &= |d(X,Y) + p'_{ij}(1-d(X,Y)) - \eta'_{ij}| \\
&= |q'_{ij}\cdot d(X,Y) + p'_{ij} - \eta'_{i,j}| \\
&= |q'_{ij}\cdot d(X,Y) - q'_{ij} + (1-\eta'_{i,j})|
\end{aligned}$$

$$= q'_{ij}|d(X,Y) - d_{i,j}| \le |d(X,Y) - d_{i,j}| \ .$$

现在固定一对集合 S、T，并令 L 为满足 $|S_i| \ge 2\gamma|V_i|$ 和 $|T_j| \ge 2\gamma|V_j|$ 的所有数对 i、j 组成的集合。那么

$$\begin{aligned}
\left|\sum_{i,j\in[k]^2}(d'(S_i,T_j) - d'_{ij})\cdot|S_i||T_j|\right| &\le 4\gamma n^2 + \left|\sum_{i,j\in L}(d'(S_i,T_j) - d'_{ij})\cdot|S_i||T_j|\right| \\
&\le 5\gamma n^2 + \left|\sum_{i,j\in L}(\mathbb{E}[d'(S_i,T_j)] - \eta'_{i,j})\cdot|S_i||T_j|\right| \\
&\le 5\gamma n^2 + \sum_{i,j\in L}\left|(\mathbb{E}[d'(S_i,T_j)] - \eta'_{i,j})\right|\cdot|S_i||T_j| \\
&\le 5\gamma n^2 + \sum_{i,j\in L}\left|d(S_i,T_j) - d_{ij})\right|\cdot|S_i||T_j| \\
&\le 5\gamma n^2 + \sum_{i,j\in[k]^2}\left|d(S_i,T_j) - d_{ij}\right|\cdot|S_i||T_j| \\
&\le 6\gamma n^2 = \varepsilon n^2 \ ,
\end{aligned}$$

其中第一个不等式由 L 的定义成立，第二个不等式由（4.9）和（4.8）（应用于（S_i，T_j））成立，第三个不等式是三角不等式，第四个不等式是（4.10），第六个不等式则是 $d^{\star A}_{\square}(G) \le \gamma$ 这一假设。由于上述结论对每个 S 都成立，T我们 deduce $d^{A}_{\square}(G')$，$\le \varepsilon$ 因此 A 确实是 ε-FK-正则划分 G'。

由于 $\gamma < \varepsilon$，不等式（4.9）意味着 S' 是 ε-签名，属于 A，相对于 G'，从而证明了该引理的第三个断言。由（4.9）我们还可得到

$$\frac{1}{k^2}\sum_{i,j\in[k]^2}|d'(V_i,V_j) - \eta'_{i,j}| \le \gamma \le \varepsilon/2 \ .$$

根据引理的假设，我们还有

$$\frac{1}{k^2}\sum_{i,j\in[k]^2}|\eta_{i,j} - \eta'_{i,j}| = d(S,S') \le \delta \ .$$

根据 引理的假设，S 是 A 相对于 G的 γ-签名，这意味着 t

$$\frac{1}{k^2}\sum_{i,j\in[k]^2}|d(V_i,V_j) - \eta_{i,j}| \le 2\gamma \le \varepsilon/2$$

最后，由于

$$|d'(V_i,V_j) - d(V_i,V_j)| \le |d'(V_i,V_j) - \eta'_{i,j}| + |\eta_{i,j} - \eta'_{i,j}| + |\eta_{i,j} - d(V_i,V_j)|,$$

我们推断出

$$\frac{n^2}{k^2}\sum_{i,j\in[k]^2}|d'(V_i,V_j) - d(V_i,V_j)| \le (\delta+\varepsilon)n^2 \ .$$

由于上述左侧正是我们将G 更改为 G'时所进行的边修改的准确数量，因此我们推得 G' 与 G是（$\delta+\varepsilon$）-接近的，从而证明了该引理的第一项断言。 □

我们还需要以下引理。

引理 4.11. ([2] 引理 3.7) 对于每个 ε, t ，存在 $\gamma = \gamma_{4.11}(\varepsilon) = \text{poly}(\varepsilon)$ 和 $N = N_{4.11}(t, \varepsilon) = \text{poly}(t/\varepsilon)$，满足以下条件。假设 A 是图 G 的一个等分划分，将其划分为 s 个集合，且该图有 $n \geq N$ 个顶点；并且 B 是将 A 细化为至多 t 个集合所得的细化。进一步假设 S 是 A 的任意一个 γ-签名，且 T 是其向 B 的扩展。如果 B 满足 $ind(B) \leq ind(A) + \gamma$，那么 T 是 B 的一个 ε-签名。

引理 4.12. ([16]引理 6.6) 对于每个 ε, t，存在 $N = N_{4.12}(t, \varepsilon) = \text{poly}(t/\varepsilon)$，使得对于 G 的每个等分划分 A（该图有 $n \geq N$ 个顶点，划分为 s 个集合），以及将 A 细化为至多 t 个集合所得的每个细化 B，都有 $ind(B) \geq ind(A) - \varepsilon$。

下面的观察隐含在 Frieze–Kannan 正则性引理（即引理4.3）的证明中。证明的主要步骤是证明：如果 A 是 G 的一个等分划分，将其划分为 t 个部分，并且A 不是 ε-FK-正则的*，那么 A 存在一个细化 B，将其划分为 $k \leq 16t/\varepsilon^4$ 个集合，使得 $ind(B) \geq ind(A) + \frac{\varepsilon^4}{2}$（例如，参见 [33] 中定理 1.1 的证明以及 [29]中定理 6 的证明）。

引理 4.13。对于每个 $\varepsilon > 0$，存在 $\gamma = \gamma_{4.13}(\varepsilon) = \text{poly}(\varepsilon)$ 和 $f = f(\epsilon)_{4.13}: \mathbb{N} \to \mathbb{N}$，满足 $f(x) = \text{poly}(1/\varepsilon) \cdot x$，并且图的每个 (f, γ)-最终划分也是 ε-FK-正则的*。

引理 4.14。对于每个 s 和 $\varepsilon > 0$，存在 $\gamma = \gamma_{4.14}(\varepsilon)$、$T = T_{4.14}(s, \varepsilon)$、$f = f^{(\varepsilon)}_{4.14}$ 以及 $N = N_{4.14}(\varepsilon, s)$，使得

$$\gamma = \text{poly}(\varepsilon), \quad T = s \cdot 2^{\text{poly}(1/\varepsilon)}, \quad f(x) = x \cdot 2^{\text{poly}(1/\varepsilon)}, \quad N = \text{poly}(s) \cdot 2^{\text{poly}(1/\varepsilon)}$$

并且以下结论成立。假设 G 至少有 N 个顶点，A 是将 G 划分为至多 s 个集合的 (f, γ)-最终划分，并且 S 是 A 的 γ- 签名。那么，对于定义在 G 的同一顶点集上的任意 G'，存在将 A 细化为 $t \leq T$ 个集合的细化 A'，使得

(*i*) A' 是 G' 的 ε-FK-正则* 划分。

(*ii*) A 的每个具有 $t \leq T$ 个集合的细化 A''（尤其是 A'）都是 G 的 ε-FK-正则* 划分。

(*iii*) 对于 A 的每个具有 $t \leq T$ 个集合的细化 A''，按照定义 S（4.8）中相对于 A'' 的含义，S'' 是 ε 相对于 G 的 A'' 签名（注意，A' 是这样的 A''）。

证明。给定 s 和 ε，我们定义

$$\gamma = \min\{\frac{1}{2}\gamma_{4.13}(\varepsilon), \gamma_{4.11}(\varepsilon)\} = \text{poly}(\varepsilon)\ , \quad T = T_{4.3}(s, \varepsilon) = s \cdot 2^{\text{poly}(1/\varepsilon)}\ ,$$

$$f(x) = f^{(\varepsilon)}_{4.13}(T_{4.3}(x, \varepsilon)) = 2^{\text{poly}(1/\varepsilon)} \cdot x \cdot 2^{\text{poly}(1/\varepsilon)} = x \cdot 2^{\text{poly}(1/\varepsilon)}\ ,$$

and

$$N = \max\{T, N_{4.12}(T, \gamma), N_{4.11}(T, \varepsilon)\} = \text{poly}(s) \cdot 2^{\text{poly}(1/\varepsilon)}\ .$$

给定一个 (f, γ)-最终划分 A，并假设 $N \geq N_{4.3}(s, \varepsilon)$，引理 4.3 生成了 A 的一个细化 A'，该细化将 G' 划分为至多 $T_{4.3}(s, \varepsilon)$ 个集合，并且相对于 G' 是 ε-FK- 正则*。剩下的工作是证明条目 (*ii*)-(*iii*)。引理 4.12 断言，相对于 G，任何将 A 细化为至多 T 个集合的划分 A'' 都满足 $ind(A'') \geq ind(A) - \gamma \geq ind(A) - \frac{1}{2}\gamma_{4.13}(\varepsilon)$（因为 $N > N_{4.12}(T, \gamma)$）。这意味着在 G 中，划分 A'' 是 $(f^{(\varepsilon)}_{4.13}, \gamma_{4.13}(\varepsilon))$-最终划分。事实上，如果存在

一个划分 C，其集合数至多为 $f^{(\varepsilon)}_{4.13}(T_{4.3}(x,\varepsilon))$，并且在 G 中 $ind(C) > ind(A'') + \gamma_{4.13}(\varepsilon)$，那么这将意味着 $ind(C) > ind(A) + \gamma_{4.13}(\varepsilon) - \frac{1}{2}\gamma_{4.13}(\varepsilon)$，从而与 A 的 (f,γ)-最终性相矛盾。因此，通过引理 4.13，我们可以推断 A'' 相对于 G 也是一个 ε-FK-正则* 划分，从而证明条目 (ii)。最后（双关语意），G 中 A 的 (f,γ)-最终性确保，对于将 G 划分为至多 T 个集合的每个划分 A''（特别是将 A 细化为这么多个集合的每个细化），我们都有 $ind(A'') \le ind(A) + \gamma_{4.11}(\varepsilon)$。因此，根据引理 4.11，将 S 扩展 S'' 到这样的一个 A'' 是相对于 G 的 A'' 的一个 ε- 签名，从而证明条目 (iii)。 □

4.2 引理 2.5 的证明

给定 h、ε 和 δ，我们首先选择

$$\gamma_0 = \min\{\varepsilon/10, \gamma_{4.7}(h,\delta/6), \gamma_{4.10}(\min\{\varepsilon/2, \gamma_{4.7}(h,\delta/6)\})\} = \mathrm{poly}(\varepsilon\delta/h)\ ,$$

然后定义

$$\gamma = \gamma_{4.14}(\gamma_0) = \mathrm{poly}(\varepsilon\delta/h), \quad s = \max\{r_{4.7}(h,\delta/6), r_{4.9}(\varepsilon/10), 20h^2/\delta\} = \mathrm{poly}(h/\varepsilon\delta)\ ,$$

$$f(x) = f^{(\gamma_0)}_{4.14}(x) = x \cdot 2^{\mathrm{poly}(1/\gamma_0)} = x \cdot 2^{\mathrm{poly}(h/\varepsilon\delta)}\ ,$$

作为引理2.5陈述中的常数和函数，并注意到它们满足该引理的条件。给定 t，如引理2.5中的陈述所述，我们令

$$T = T_{4.14}(t,\gamma_0)$$

并定义

$$N = \max\{N_{4.7}(h,T,\delta/6), N_{4.9}(\varepsilon/10,T), N_{4.14}(\gamma_0,s), N_{4.10}(t,\gamma_0)\} = \mathrm{poly}(t)\cdot 2^{\mathrm{poly}(h/\varepsilon\delta)}\ ,$$

为引理 2.5中的常数。

给定一个最多有 h 个顶点的图族 $\mathcal{H}$，我们如下定义一个签名族

$$\mathcal{C}_{\delta,\mathcal{H},T} = \{C : |C| \le T \text{ and } ind(H,C) \le \delta/2 \text{ for every } H \in \mathcal{H}\}\ .$$

为了使 $\mathcal{C}_{\delta,\ \mathcal{H},\ T}$ 有限，我们只将边权重为 $\eta_{i,\ j}$ 且是 $\beta = \min\{\varepsilon/10, \delta/10h^2\}$ 整数倍的签名 C 放入其中。直观而言，这是能够“证明”（因此是 C）具有该签名的图接近于诱导 $\mathcal{H}$-自由的签名集合。我们还将 $\mathcal{S}_T$ 定义为所有至多有 T 个部分且是 S 的扩展[7] 的签名集合。直观而言，这些是通过将 A 细化为至多 T 个集合所能得到的签名（回想一下，关键在于算法只能访问 S，而不能访问 G）。

现在假设给定了一个 γ-签名 S，它是某个图 G的 (f,γ)-终止划分 A（其中 f和 γ如上所定义），并且 S 有 $t \ge s$ 个部分，而 G 至少有 N 个顶点。该算法检查是否存在 $S' \in \mathcal{S}_T$和 $C \in \mathcal{C}_{\delta,\ \mathcal{H},\ T}$，使得 $d_1(S',C) \le \alpha - \frac{\varepsilon}{2}$。如果存在这样一对，算法就判定情形 (i) 成立；否则判定情形 (ii) 成立。现在我们证明该算法的正确性。

[7]注意，严格来说，根据定义 4.8，扩展必须相对于一个划分 A 及其细化 B而言，而这里我们只有签名 S。因此，这里的意思是，若取一个具有划分 A且其 0-签名为 S 的图，那么 $\mathcal{S}_T$ 就是通过将 A 细化为至多 T 个集合，并据此取 S 在这些细化上的扩展而得到的所有签名的族。当然，我们不需要任何图来构造 $\mathcal{S}_T$；只需将 S 的“部分”拆分为总计至多 T 个新的“部分”，然后按照定义 4.8，定义新顶点之间的密度 $\eta'_{i,\ j}$。

第一个方向的证明：假设存在一个图 G'，它与 G $(\alpha-\varepsilon)$-接近，并且对每个 $H\in\mathcal{H}$ 都满足 $ind(H,G')=0$。我们将证明该算法会判定情形 (i) 成立。

回顾一下，A 是将 G 划分为 $t\geq s$ 个集合的(f,γ)-最终划分，而 S 是 A的 γ-签名。根据引理 4.14，存在将 A 细化为至多 T 个集合的细化 A'，使得 A' 对 G 和 G'均为 γ_0-FK-正则*。此外，记 S' 为 S 到 A'的相应扩展，则S' 是关于 G的 A' 的 γ_0-签名。注意，$S'\in\mathcal{S}_T$。根据 γ_0的选取，这意味着 A'对 G和 G'均为$\gamma_{4.7}(h,\delta/6)$-FK-正则*，且 S'是关于 G的 A'的$\frac{1}{10}\varepsilon$-签名。令 C' 为 A' 在 G'上的0-签名。引理 4.7（使用 A' 和 G'）推出，对于所有 $H\in\mathcal{H}$，都有$|ind(H,G')-ind(H,C')|\leq\delta/6$。因此，对于所有 $H\in\mathcal{H}$，都有 $ind(H,C')\leq\delta/6$。显然，存在一个大小为 C' 的签名 C，使得 C的所有权重都是 β 的常数倍，并且 $d_1(C',C)\leq\beta$。由于 $d_\square(C',C)\leq d_1(C',C)\leq\delta/10h^2$，我们由引理 4.5（应用于$\frac{\delta}{10h^2}$，因为 $s\geq\frac{20h^2}{\delta}$）推出，对所有 $H\in\mathcal{H}$，都有$ind(H,C)\leq\delta/6+\delta/5<\delta/2$，因此$C\in\mathcal{C}_{\delta,\mathcal{H},T}$。此外，根据断言 4.9（由于 A'至少有$r_{4.9}(\varepsilon/10)$个部分，并假设 n 足够大），我们推出 $d_1(S',C)\leq\alpha-\frac{\varepsilon}{2}$（因为 G 和 G' 是$(\alpha-\varepsilon)$-接近的，且 $d_1(C,C')\leq\varepsilon/10$）。因此，$S'$ 和 C 构成一个见证，表明该算法确实会判定情形(i)成立。

第二个方向的证明：假设算法判定情形 (i) 成立。我们证明，在这种情况下，存在一个图 G'，它与 G 的距离不超过 α，并且满足 $ind(H,G')<\delta$，对所有 $H\in\mathcal{H}$ 均成立

事实上，若算法判定情形 (i) 成立，那么就存在签名 $S'\in\mathcal{S}_T$ and$C\in\mathcal{C}_{\delta,\mathcal{H},T}$，满足 $d_1(S',C)\leq\alpha-\frac{\varepsilon}{2}$。由于 $S'\in\mathcal{S}_T$，存在 A 的一个细化 A'，使得 S' 是根据 A' 对 S 进行扩展所得。引理 4.14（将 A' 视为相对于 G 对 A 的一种可能细化）断言，S' 是 A' 的一个 γ_0-签名（相对于 G）。根据 γ_0 的选取，这意味着它是 A' 的一个 $\gamma_{4.10}(\min\{\frac{\varepsilon}{2},\gamma_{4.7}(h,\delta/6)\})$-签名（相对于 G）。现在，引理 4.10（取 A' 为 G 的 γ_0-FK-正则* 划分，并取 S' 为 S、C 为 S'）推出存在一个图 G'，它与 G 的距离不超过 $(\alpha-\frac{\varepsilon}{2}+\frac{\varepsilon}{2})$，也就是与 G 的距离不超过 α；并且对于该图，C 是 A'的一个$\gamma_{4.7}(h,\delta/6)$-签名，而 A'又是 G' 上的 $\gamma_{4.7}(h,\delta/6)$-FK-正则图。引理 4.7 推出，对于所有 $H\in\mathcal{H}$，$|ind(H,G')-ind(H,C)|\leq\delta/6$。因此，对于所有 $H\in\mathcal{H}$，$ind(H,G')<\delta/2+\delta/6<\delta$，如所需。由此我们找到了所需的 G'。

5 定理1.9证明

定理1.9 的证明与定理 1.4的证明非常相似。为了帮助已经熟悉定理 1.4证明的读者，我们在几个地方指出，某些引理与我们在前文某一节中提出的引理类似。思路如下：根据 Goldreich 和 Trevisan 的一个定理 [21], ，每个可测试性质都可以由规范测试器进行测试；该测试器会抽取一组大小为 $q=q_{\mathcal{P}}(\varepsilon)$ 的顶点，并根据这些顶点诱导出的图来接受或拒绝。因此，算法的接受或拒绝只取决于 G 中顶点数为 q 的图的诱导副本数量。因此，将一个图转换为满足 $\mathcal{P}$ 的图，等价于将其转换为具有特定数量的、由 q 个顶点构成的特定图副本的图。显然，这与定理1.4 的情形非常相似，在该情形中，我们希望不属于 $\mathcal{P}$的图的副本数量非常少。之所以会多出一个指数因子，是因为我们需要控制 所有由 q 个顶点构成的图的诱导副本数量。

现在我们陈述关键引理，它们是我们在证明定理1.4时所用引理的变体。

定义 5.1. 给定组合结构有限族 $\mathcal{H}$ 上的两个分布 μ 和 ν，它们的变差距离定义为：
$|\mu-\nu|=\frac{1}{2}\sum_{H\in\mathcal{H}}|\Pr^{\mu}(H)-\Pr^{\nu}(H)|$

引理 5.2. 如果组合结构有限族 $\mathcal{H}$ 上的两个分布 μ 和 ν 满足$|\mu-\nu| \le \delta$，那么对于任意集合 $A \subset \mathcal{H}$，有 $|\Pr_\mu(A) - \Pr_\nu(A)| \le \delta$

引理 5.3. 假设 μ 和 ν 是定义在顶点集合$\{v_1,\dots,v_q\}$上的两个图的概率分布，其中每条边 v_iv_j 均独立地以概率 $\mu_{i,j}$ 和 $\nu_{i,j}$分别被选为一条边。如果对于每个 $1 \le i < j \le q$，均有$|\mu_{i,j} - \nu_{i,j}| \le \epsilon/\binom{q}{2}$，那么 μ和 ν 之间的变差距离有界于 ϵ。

定义 5.4（q-统计量）。图 G 的 q-统计量是所有具有 q 个顶点的（带标号）图上的概率分布，这些图是从 G 中随机选取 q 个不同的顶点并考察其诱导子图后得到的。对于给定的图 H，我们将按照 q-统计量抽取一个图时得到 H 的概率记为 $\Pr_G(H)$。

定义 5.5. 对于 G 的一个等分划分 $A = \{V_1,\dots,V_t\}$，以及一个 $1\le i<j\le t$ 的签名 $S = (\eta_{i,j})$，根据 S 的感知 q- 统计量是以下在具有 q 个顶点 $v_1,\dots,v_q$ 的带标记图上的分布 $\Pr_S$。首先，从 $1,\dots,t$. 中均匀随机地选取一个不重复的指标序列 $i_1,\dots,i_q$。然后，独立地，对每个 $k<l$，以概率 η_{i_k,i_l} 令 v_kv_l 为一条边（若 $i_k < i_l$），以概率 η_{i_l,i_k} 令其为一条边（若 $i_l < i_k$）。接着，将 $\Pr_S(H)$ 定义为所得带标记图等于 H 的概率。

下面的引理将在定理1.1 的证明中替代引理1.9。

引理 5.6（参见 [21]）。如果存在一个针对图性质 $\mathcal{P}$ 的 ε-测试，其进行 $Q = Q(\varepsilon)$ 次边查询，那么就存在一个由 $q = 2Q$ 个顶点上的带标记图组成的适当族 $\mathcal{H}$，使得任何满足 $\mathcal{P}$ 的图G，也满足 $\Pr_G(\mathcal{H}) \ge \frac{2}{3}$；并且任何与满足 $\mathcal{P}$ 相距 ε 的图 G，也满足 $\Pr_G(\mathcal{H}) < \frac{1}{3}$。

We now 引入引理4.7 的一个变体，该变体适用于定理 1.9的证明。

引理 5.7。对于每个 q, ε，存在 $\gamma = \gamma_{5.7}(q,\varepsilon)$ $r = r_{5.7}(q,\varepsilon)$，使得

$$\gamma = \mathrm{poly}(\varepsilon \cdot 2^{-q^2}), \quad r = \mathrm{poly}(1/\varepsilon \cdot 2^{q^2})$$

对于每个 γ-签名 S 的 γ-FK-正则等分划分 A 为 $t \ge r$ 个集合，以及一个具有$n \ge N_{5.7}(q,\varepsilon,t) = \mathrm{poly}(t/\varepsilon)$ $2^{\mathrm{poly}(q)}$ 个顶点的图 G，我们有 $|\Pr_S - \Pr_G| \le \varepsilon$，其中 $\Pr_G$是 q-统计量，而 $\Pr_S$ 是根据 S定义的感知 q-统计量。

证明。我们设

$$\gamma = \frac{\varepsilon}{56q^2 2^{\binom{q}{2}}} = \mathrm{poly}(\varepsilon \cdot 2^{-q^2}), \quad r \ge \frac{2}{\gamma} = \mathrm{poly}(1/\varepsilon \cdot 2^{q^2}), \quad N = \frac{2t}{\gamma}.$$

令 $\mathcal{H}$为任意一个由 q-顶点图组成的族。我们需要证明 $|\Pr_G(\mathcal{H}) - \Pr_S(\mathcal{H})| \le \varepsilon$。注意，对于每个具有 q个顶点的图 H，都有 $\Pr_G(H) = n^q \cdot (n-q)!\, n! \cdot ind(H,G)$。令 S_A为 A的 0-签名。因此，对于每个图 $H \in \mathcal{H}$，都有 $\Pr_{S_A}(H) = n^q \cdot (n-q)!\, n! \cdot ind(H,S_A)$。由引理 4.6 可得

$$|ind(H,G_A) - ind(H,S_A)| \le \frac{2q^2}{t} + \frac{2tq}{n} \le 2q^2\gamma$$

对每个 $H \in \mathcal{H}$。此外，由引理 4.5，我们有

$$|ind(H,G_A) - ind(H,G)| \le 2q^2\gamma$$

对每个 $H \in \mathcal{H}$。因此，由三角不等式可得

$$|\Pr_G(H) - \Pr_{S_A}(H)| \le 4q^2\gamma \cdot \frac{n^q \cdot (n-q)!}{n!} \le 8q^2\gamma \le \frac{\varepsilon}{7 \cdot 2^{\binom{q}{2}}}$$

对于每个 $H \in \mathcal{H}$。通过对 q 个顶点上 H 的所有 $2^{\binom{q}{2}}$ 种可能进行计数，我们得到 $|\Pr_G(\mathcal{H}) - \Pr_{S_A}(\mathcal{H})| < \frac{\varepsilon}{7}$。记 $S = (\eta_{i,j})_{1 \le i < j \le t}$。对于除 $\gamma\binom{t}{2}$ 以外的所有数对，我们都有 $|\eta_{i,j} - d(V_i, V_j)| < \gamma \le \frac{\varepsilon}{7q^2}$，因为 S 是 A 的一个 γ-签名。因此，以至少 $1 - \frac{4\varepsilon}{7}$ 的概率，我们会从 $1, \ldots, t$ 中抽样到互不相同的指标 $i_1, \ldots i_q$，使得每一对 $i, j \in \{i_1, \ldots, i_q\}$ 都满足 $|\eta_{i,j} - d(V_i, V_j)| < \frac{\varepsilon}{7q^2}$。因此，由引理 5.2 和 5.3，我们得到 $|\Pr_S(\mathcal{H}) - \Pr_{S_A}(\mathcal{H})| < \frac{5\varepsilon}{7}$。综上，$|\Pr_G(\mathcal{H}) - \Pr_S(\mathcal{H})| < \frac{6\varepsilon}{7} < \varepsilon$，正如所要求的。□

现在我们引入引理的一个变体2.5，它适用于定理1.9的证明。

引理 5.8. 对于每个 q 和 ε，存在 $\gamma = \gamma_{5.8}(q, \varepsilon)$、$s = s_{5.8}(q, \varepsilon)$ 和 $f_{5.8}^{(q,\varepsilon)}\colon \mathbb{N} \to \mathbb{N}$，使得

$$\gamma = \mathrm{poly}(\varepsilon \cdot 2^{-q^2}), \quad s = \mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right), \quad f_{5.8}^{(q,\varepsilon)}(x) = x \cdot 2^{\mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right)}$$

具有以下性质。对于每个拥有 q 个顶点的图族 $\mathcal{H}$，都存在一个确定性算法，它以图 G 的一个 (f, γ)-最终划分 A 的 $t \ge s$ 个集合的 γ-签名 S 为输入，其中该图有 $n \ge N_{5.8}(q, \varepsilon, t)$ 个顶点，并且对于给定的任意 α，能够区分以下两种情况：

(i) G 与某个图 G' 是 $(\alpha - \varepsilon)$-接近的，其中该图满足 $\Pr_{G'}(\mathcal{H}) \ge \frac{2}{3}$。

(ii) G 与满足 $\Pr_{G'}(\mathcal{H}) \ge \frac{1}{3}$ 的所有 G' 都相差至少 α。

证明。我们像引理的证明中那样定义以下参数 2.5。

$$\gamma_0 = \min\left\{\varepsilon/10, \gamma_{5.7}(q, \frac{1}{12}), \gamma_{4.10}(\min\{\varepsilon/2, \gamma_{5.7}(q, \frac{1}{12})\})\right\} = \mathrm{poly}(\varepsilon \cdot 2^{-q^2}),$$

并进一步定义：

$$\gamma = \gamma_{4.14}(\gamma_0) = \mathrm{poly}(\gamma_0) = \mathrm{poly}(\varepsilon \cdot 2^{-q^2})$$

$$s = \max\{r_{5.7}(q, 1/12), r_{4.9}(\varepsilon/10)\} = \mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right)$$

$$T = T_{4.14}(t, \gamma_0) = t \cdot 2^{\mathrm{poly}(1/\varepsilon) \cdot 2^{\mathrm{poly}(q)}}$$

$$f = f_{4.14}^{(\gamma_0)}(x) = x \cdot 2^{\mathrm{poly}(1/\gamma_0)} = x \cdot 2^{\mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right)}$$

$$N = \max\{N_{5.7}(q, 1/12, t)) \; N_{4.9}(\varepsilon/10, T_{4.14}(t, \gamma_0)) \; N_{4.14}(s, \gamma_0) \; N_{4.10}(\gamma_0, t)\} = t \cdot 2^{\mathrm{poly}(1/\varepsilon) \cdot 2^{\mathrm{poly}(q}}$$

给定一个至多有 q 个顶点的图族 $\mathcal{H}$，我们定义如下签名族

$$\mathcal{C}'_{\mathcal{H},T} = \{C : |C| \le T \text{ and } \Pr_C(\mathcal{H}) \ge 1/2\}.$$

为了使 $\mathcal{C}'_{\mathcal{H},\,T}$ 有限，我们只在其中加入边权重为 $\eta_{i,\,j}$ 且是 $\beta = \frac{\varepsilon}{12q^2}$ 的整数倍的签名 C。

现在假设给定了某个图 G 的某个 (f, γ)-终结（采用上述定义的 f, γ）划分 A 的一个 γ-签名 S，其中 S 有 $t \ge s$ 个部分，且 G 至少有 N 个顶点。定义 $\mathcal{S}_T$ 为所有至多有 T 个部分且是 S 的扩展的签名所组成的集合。算法检查是否存在满足 $d_1(S', C) \le \alpha - \frac{\varepsilon}{2}$ 的 $S' \in \mathcal{S}_T$ 和 $C \in \mathcal{C}'_{\mathcal{H},\,T}$。如果存在这样一对，算法就判定情形 (i) 成立；否则，它判定情形 (ii) 成立。现在我们证明该算法的正确性。

第一方向的证明：假设 G' 是某个与 G $\alpha-\varepsilon$-接近的图，并且 $\mathrm{Pr}_{G'}(\mathcal{H}) \geq 2/3$。我们将证明，该算法会判定情形 (i) 成立。回顾一下，A 是将 G 划分为 $t \geq s$ 个集合的 (f,γ)-最终划分，而 S 是 A 的一个 γ-签名。由引理 4.14，存在将 A 细化为至多 T 个集合的细化 A'，使得对于 G 和 G'，A' 都是 γ_0-FK-正则的。此外，记 S' 为 S 到 A' 的相应扩展，则 S' 是关于 G 的 A' 的一个 γ_0-签名。注意，$S' \in \mathcal{S}_T$。根据 γ_0 的选取，这意味着对于 G 和 G'，A' 都是 $\gamma_{5.7}(q,1/12)$-FK-正则的，并且 S' 是关于 G 的 A' 的一个 $\frac{1}{10}\varepsilon$-签名。令 C' 为 A' 在 G' 上的 0-签名。引理 5.7（使用 A' 和 G'）推出 $|\mathrm{Pr}_{G'}(\mathcal{H}) - \mathrm{Pr}_{C'}(\mathcal{H})| < \frac{1}{12}$。因此，$\mathrm{Pr}_{C'}(\mathcal{H}) \geq 2/3 - 1/12 \geq 7/12$。显然，存在一个签名 C，使得所有 C 的权重都是 β 的常数倍，并且 $|c_{ij} - c'_{ij}| \leq \beta$。因此，由引理 5.3，我们推得 $\mathrm{Pr}_C(\mathcal{H}) \geq 7/12 - 1/12 \geq 1/2$，于是 $C \in \mathcal{C}'_{\mathcal{H},T}$。此外，由断言 4.9（因为 A' 至少有 $r_{4.9}(\frac{\varepsilon}{10})$ 个集合，并假设 n 足够大），考虑到 G 和 G' 是 $(\alpha-\varepsilon)$- 接近的图，有 $d_1(S',C) \leq (\alpha - \frac{\varepsilon}{2})$。因此，$S'$ 和 C 提供了一个见证，表明上述过程接受 G。

第二个方向的证明：假设算法判定情形 (i) 成立。我们证明，在这种情况下，存在一个图 G'，它与 G 是 α-接近的，并且满足 $\mathrm{Pr}_{G'}(\mathcal{H}) > 1/3$。

事实上，如果算法判定情形 (i) 成立，那么就存在满足 $d_1(S',C) \leq \alpha - \frac{\varepsilon}{2}$ 的签名 $S' \in \mathcal{S}_T$ 和 $C \in \mathcal{C}'_{\mathcal{H},T}$。由于 $S' \in \mathcal{S}_T$，存在对 A 的细化 A'，使得 S' 是根据 A' 对 S 的扩展。引理 4.14（将 A' 视为相对于 G 对 A 的一种可能细化）断言，S' 是 A' 的一个 γ_0-签名（相对于 G）；根据 γ_0 的选取，这意味着它是 A' 的一个 $\gamma_{4.10}(\min\{\frac{\varepsilon}{2}, \gamma_{5.7}(q, 1\ 12\ )\})$-签名（相对于 G）。现在，引理 4.10（取 A' 作为 G 的 γ_0-FK-正则* 划分，并取 S' 为 S，C 为 S'）推出，存在一个图 G'，它与 G 是 $(\alpha - \frac{\varepsilon}{2} + \frac{\varepsilon}{2})$-接近的，即与 G 是 α-接近的，并且 C 是 A' 的一个 $\gamma_{5.7}(q, 1\ 12\ )$-签名；而后者又在 G' 上是 $\gamma_{5.7}(q, 1\ 12\ )$-FK-正则的。引理 5.7 推出 $|\mathrm{Pr}_{G'}(\mathcal{H}) - \mathrm{Pr}_C(\mathcal{H})| < \frac{1}{12}$。因此，$\mathrm{Pr}_{G'}(\mathcal{H}) \geq 1/2 - 1/12 > 1/3$，如所需。由此，我们找到了所需的 G'。 □

现在可以利用引理 1.9 和引理 2.4 以及 5.8推出定理 1.4 和引理 2.4 以及2.5。证明过程与利用引理 2.4 和 2.5 推出定理 1.4 的过程类似。

证明（定理 1.9）：设 $\mathcal{P}$ 是一个可测试图性质，并令 $\alpha, \varepsilon > 0$ 为我们希望用 (α,ε)-估计 $\mathcal{P}$ 的常数。由于 $\mathcal{P}$ 是 $\varepsilon/2$-可测试的，因此存在一个测试算法，对于给定的 $\varepsilon/2$ 和图 G，会进行 $Q(\varepsilon/2)$ 次查询。引理 5.6 告诉我们，存在一个常数 $q = q_{5.6}(\varepsilon/2) = \mathrm{poly}(Q(\varepsilon/2))$，以及一个由 $\mathcal{H}$ 个 q 顶点的图组成的图族，使得对每个$G \in \mathcal{P}$，$\mathrm{Pr}_G(\mathcal{H}) \geq 2/3$；并且对于每个与满足 $\mathcal{P}$ 相差 $\varepsilon/2$ 的图 G，$\mathrm{Pr}_G(\mathcal{H}) \leq 1/3$。因此，我们设

$$\gamma = \gamma_{5.8}(q, \varepsilon/2) = \mathrm{poly}(\varepsilon \cdot 2^{-q^2}),$$

$$f = f_{5.8}^{(q,\varepsilon/2)}(x) = x \cdot 2^{\mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right)},$$

$$k = s_{5.8}(q, \varepsilon/2) = \mathrm{poly}\left(\frac{2^{q^2}}{\varepsilon}\right)$$

现在，定义 $\zeta = \varepsilon \cdot 2^{-q^2}$ 并应用引理 2.4 中给出的算法，参数为 k、ζ、γ、f，输入图为 G。该算法最多进行 $q_{2.4}(k,\ \zeta)$ 次查询，根据引理的假设，我们有：

$$q_{2.4}(k,\zeta) = \mathrm{poly}(k)2^{\mathrm{poly}(1/\zeta)} = 2^{\mathrm{poly}(1/\varepsilon)\cdot 2^{\mathrm{poly}(q)}}.$$

以至少 $2/3$ 的概率，该算法返回一个 G 的 (f,γ)-最终等划分的 γ-签名，其中集合数量至少为 s 且至多为 $T_{2.4}(k,\gamma,f)$。现在，我们将引理 5.8 提供的算法应用于签名 S，参数为 $q,\varepsilon/2,\alpha-\varepsilon/2$。由于参数的选择，引理 5.8 保证我们能够区分以下两种情况：存在一个图 G'，它与 G 的距离为 $(\alpha-\varepsilon)$，且 $\mathrm{Pr}_{G'}(\mathcal{H}) \geq 2/3$；以及 G 与任何满足 $\mathrm{Pr}_{G'}(\mathcal{H}) > 1/3$ 的 G' 的距离均为 $(\alpha-\varepsilon/2)$。在第一种情况下，G 将被接受；在第二种情况下，则将被拒绝。为了使上述方法奏效，我们要求 $N \geq \max\{N_{2.4}(k,\gamma,f)\ N_{5.8}(q,\frac{\varepsilon}{2},T_{2.4}(k,\gamma,f))\}$。对于更小的 n，我们只需读取整个输入，并计算其与待估计性质之间的距离，此时查询复杂度为 $N^2 \leq 2^{\mathrm{poly}(1/\varepsilon)\cdot 2^{\mathrm{poly}(q)}}$。现在，我们声称，对于每个 $n \geq N$，上述算法确实是一个针对 $\mathcal{P}$ 的 (α,ε)- 估计算法。

如果 G 与 $\mathcal{P}$ 相差不超过 $(\alpha-\varepsilon)$，那么它也与某个图 G' 相差不超过 $(\alpha-\varepsilon)$，且该图满足 $\mathrm{Pr}_{G'}(\mathcal{H}) \geq 2/3$。因此，只要 S 确实是 G 的一个 (f,γ)- 最终划分的 γ- 签名，上述第一种情况就成立，而这种情况发生的概率至少为 $2/3$。因此，G 被接受的概率至少为 $2/3$。

另一方面，如果 G 与 $\mathcal{P}$相距至少 α，那么根据三角不等式，它与任何满足 $\mathrm{Pr}_{G'}(\mathcal{H}) > 1/3$ 的 G' 都相距至少 $(\alpha-\varepsilon/2)$。因此，只要 S 实际上是 G 的一个 (f,γ)-最终划分的 aγ-签名，上述第二种情况就成立，而这种情况发生的概率至少为 $2/3$。因此，G 被拒绝的概率至少为 $2/3$。 □

证明（命题 1.10）：回顾一下，我们用 $\mathrm{dist}_{\mathcal{P}}(G)$ 表示将 G 转换为满足 $\mathcal{P}$ 的图所需进行的最少边的添加/删除次数，并以 $|V(G)|^2$ 归一化。若 $\mathcal{P}$ 是自然数，则存在一个趋于零的正数序列 $\{\varepsilon_t\}_{t=1}^{\infty}$，使得对于每个 ε_t，都存在一个整数序列 $\{n_k\}_{k=1}^{\infty}$，且对于每个 k，都存在一个具有 n_k 个顶点的图 G_k，满足 $\varepsilon_t \leq \mathrm{dist}_{\mathcal{P}}(G_k) \leq 1$。现在我们将证明，对于每个 t 和每个 k，任何用于 $\mathcal{P}$ 的 ε_t-测试器在处理具有 n_k 个顶点的图时，都必须至少进行 $\frac{1}{40\varepsilon_t}$ 次边查询。

固定一个如上所述的 t。关键观察是，可以进一步假设上文定义的图 G_k 满足 $\mathrm{dist}_{\mathcal{P}}(G_k) \leq 2\varepsilon_t$。事实上，当 n_k 足够大时，我们有 $\varepsilon_t \cdot n_k^2 \geq 1$；修改 G_k 中的一条边，最多会使 $\mathrm{dist}_{\mathcal{P}}(G)$（加性地）变化 ε_t。因此，我们可以从满足 $\varepsilon_t \leq \mathrm{dist}_{\mathcal{P}}(G) \leq 1$ 的图 G_k 开始，然后逐条修改其边，直到得到一个满足 $\varepsilon_t \leq \mathrm{dist}_{\mathcal{P}}(G'_k) \leq 2\varepsilon_t$ 的新图G'_k。对于每个这样的 G_k，若其有 n_k 个顶点，由于 $\mathrm{dist}_{\mathcal{P}}(G_k) \leq 2\varepsilon_t$，存在一个有 n_k 个顶点的图 H_k，它满足 $\mathcal{P}$，且 $G_k \triangle H_k$（即两图的对称差）至多包含 $2\varepsilon_t \cdot n_k^2$ 条边。现在的关键观察是，对于 $\mathcal{P}$，一个 ε_t-测试器若进行少于 $\frac{1}{40\varepsilon_t}$ 次边查询，查询到 $G_k \triangle H_k$ 中某条边的概率至多为 $1/10$。因此，它区分 G_k 与 H_k 的概率至多为 $1/10$，所以它不可能是针对 $\mathcal{P}$ 的 ε_t-测试器。 □

参考文献

[1] N. Alon, B. Chazelle, S. Comandur, 和 D. Liue，单调函数距离的估计，《随机结构与算法》**31** (2007), 371–383. 1.2[2] N. Alon, E. Fischer, M. Krivelevich 和 M. Szegedy，大图的高效测试，《组合数学》**20** (2000), 451–476. 1.3,4.11[3] N. Alon, E. Fischer, I. Newman, 和 A. Shapira，可测试图性质的组合刻画：关键在于正则性，《SIAM计算杂志》**39** (2009), 143—167. 1.1[4] N. Alon 和 J. Fox，易于测试的图性质，《组合概率与计算》**24** (2015),646–657. 1.1

[5] N. Alon 和 A. Shapira，可通过单侧错误测试的（自然）图性质的刻画，《SIAM计算杂志》37 (2008), 1703–1727. 1.1, 1.1, 1.3[6] T. Batu, F. Ergun, J. Kilian, A. Magen, S. Raskhodnikova, R. Rubinfeld, 和 R. Sami，弱近似编辑距离的次线性算法，《ACM计算调查》35 (2003), 316–324. 1.2[7] T. Batu, L. Fortnow, R. Rubinfeld, W. Smith 和 P. White，离散分布接近性测试，《ACM杂志》60 (2013), 1-25. 1.2[8] P. Berman, M. Murzabulatov, 和 S. Raskhodnikova，图像性质的容错测试器，ICALP 2016会议录，1–14. 1.2[9] E. Blais, C. Canonne, T. Eden, A. Levi, 和 D. Ron，容错junta测试及其与次模优化和函数同构的联系，《ACM计算理论汇刊》11 (2019), 文章24. 1.2[10] C. Borgs, J. Chayes, L. Lov´asz, V.T. S´os, B. Szegedy, 和 K. Vesztergombi，图极限与参数测试，STOC 2006会议录，261—270. 1.1, 1.3[11] A. Campagna, A. Guo, 和 R. Rubinfeld，连通性和直径的局部重构器与容错测试器，APPROX 2013会议录，411–424. 1.2[12] D. Conlon 和 J. Fox，图正则性和移除引理的界，《几何与泛函分析》22 (2012), 1191–1256. 1.3[13] T. Eden, R. Levi, 和 D. Ron，有界森林度测试，SODA 2018会议录，2081–2092. 1.2[14] N. Fiat 和 D. Ron，图性质的高效距离近似，SODA 2021会议录，1618–1637. 1.3, 1.3[15] E. Fischer 和 L. Fortnow，布尔性质的容错测试与非容错测试，《计算理论》2 (2006), 173—183. 1.2[16] E. Fischer 和 I. Newman，图性质的测试与估计，《SIAM计算杂志》37 (2007), 482–501. 1.2, 1.3, 1.4, 1.5, 3, 3.1, 4.12[17] A. Frieze 和 R. Kannan，稠密问题的正则性引理与近似方案，FOCS 1996会议录，12–20. 1.5, 4.1[18] A. Frieze 和 R. Kannan，矩阵的快速近似及其应用，《组合数学》19 (1999), 175–220. 1.1, 4.1, 4.1, 4.1, 4.3[19] L. Gishboliner 和 A. Shapira，具有多项式界的移除引理，STOC 2017会议录，510–522. 1.1, 1.3[20] O. Goldreich, S. Goldwasser 和 D. Ron，性质测试及其与学习和近似的联系，《ACM杂志》45 (1998), 653–750. 1.1, 1.3, 1.5, 3.1[21] O. Goldreich 和 L. Trevisan，关于图性质测试的三个定理，《随机结构与算法》23 (2003), 23–57. 1.1, 1.3, 1.4, 5, 5.6[22] O. Goldreich，《性质测试导论》，剑桥大学出版社 (2017). 1.1

[23] V. Guruswami 和 A. Rudra，容错局部可测试码，《RANDOM 2005》论文集，306—317。1.2[24] C. Hoppen、Y. Kohayakawa、R. Lang、H. Lefmann 和 H. Stagni，关于估计到遗传图性质的距离的查询复杂度，《SIAM离散数学杂志》35（2021），1238–1251。1.3，1.3，1.5，1.5，4.1，4.1，4.1[25] C. Hoppen、Y. Kohayakawa、R. Lang、H. Lefmann 和 H. Stagni，估计与单调性质相关的参数，《组合、概率与计算》29（2020），616–632。另见《APPROX-RANDOM 2016》论文集。1.3，1.3，1.5[26] S. Kopparty 和 S. Saraf，容错线性测试与局部可测试码，《RANDOM 2009》论文集，601–614。1.2[27] L. Lov´asz 和 B. Szegedy，面向分析学家的 Szemer´edi 引理，《几何与泛函分析》17（2007），252–270。1.3[28] S. Marko 和 D. Ron，有界度图与一般稀疏图中的距离近似，《ACM算法汇刊》5（2009），22:1–22:28。1.2[29] G. Moshkovitz 和 A. Shapira，稀疏正则近似引理，《美国数学会汇刊》371（2019），第10期，6779–6814。4.1，4.2，4.1，4.3，4.1[30] M. Parnas、D. Ron 和 R. Rubinfeld，容错性质测试与距离近似，《计算机与系统科学杂志》72（2006），1012–1042。1.2[31] V. R¨odl 和 R. Duke，关于具有小子图且色数较大的图，《图与组合》1（1985），91–96。1.1[32] V. R¨odl 和 M. Schacht，消去引理的推广，《组合数学》29（2009），467–501。1.3[33] V. R¨odl 和 M. Schacht，图的正则性引理，《组合数学与计算机科学盛会》，第20卷（2010），Bolyai 数学研究丛书，287-325。1.1，4.1[34] A. Shapira 和 H. Stagni，测试划分性质的紧致界，已投稿，2023年。3，3.1，3.1[35] E. Szemer´edi，图的正则划分，载于：《CNRS国际研讨会论文集》（J. C. Bermond、J. C. Fournier、M. Las Vergnas 和 D. Sotteau 编），1978年，399–401。1.1，1.2