

大图中子图的可测性质

韦立 许道云 王正才

(贵州大学计算机科学与信息学院 贵阳 550025)

摘要 对于给定的距离参数 ϵ ,性质测试算法A需以高概率正确地区分给定的对象具备预定性质 Π 与 ϵ -远离性质 Π .若存在 Π 的测试算法A满足其询问复杂性独立于规模参数 n ,则称 Π 是可测的.设 H 是一个图,性质 H -free为不含 H -子图的图所构成的集合.在有界度模型中,Goldreich与Ron证明了对任意连通图 H ,性质 H -free是可测的^[9].在邻接矩阵模型中,证明了对任意图 H ,不管其连通与否,性质 H -free是可测的.

关键词 性质测试,询问复杂性,正则引理,正则归约,可测试性

中图法分类号 TP301 **文献标识码** A

Testable Properties for Subgraphs in Large Graphs

WEI Li XU Dao-yun WANG Zheng-cai

(School of Computer Science and Information, Guizhou University, Guiyang 550025, China)

Abstract A property testing algorithm is required to highly probably distinguish the case that the input has predetermined property Π and the case that the input is ϵ -far from having the property, for any given parameter ϵ . If there is a property testing algorithm for Π such that its query complexity is independent of the input size parameter n , Π is called testable. Let H be a graph, H -free is the property of containing no copy of H . For any connected graph H , Goldreich and Ron show that H -free is testable in bounded degree model. In adjacency matrix model, it proves that H -free is testable, for any given H .

Keywords Property testing, Query complexity, Regularity lemma, Regularity reduction, Testability

1 引言

在性质测试中,给定一个对象 f 及距离参数 $\epsilon > 0$,要求设计一个概率算法A,它询问关于 f 的局部信息,以高概率正确区分 f 具有预先设定的性质 Π 与 f 是 ϵ -远离性质 Π .性质测试是标准判定任务的一种松弛:测试算法通常会拒绝那些不具备性质和“远离”该性质的对象。

一个性质即一个对象类,例如,图的二分性,即二分图所组成的类。形式上,性质 Π 是 $(\epsilon, Q(\epsilon, n))$ -可测的,若存在概率算法A,对任何输入对象 f ,A至多使用 $Q(\epsilon, n)$ 次询问,满足:

- (1) 若 f 具备性质 Π ,则A接受 f 的概率至少为 $2/3$;
- (2) 若 f 是 ϵ -远离 Π ,则A拒绝 f 的概率至少为 $2/3$ 。

测试算法A的询问复杂性 $Q(\epsilon, n)$ 与时间复杂性 $T(\epsilon, n)$ 是关于对象规模 n 及距离参数 ϵ 的函数。若测试算法A接受 Π 中成员的概率为1,则称A具备单边误差;否则,称A具备双边误差。若A的询问是相互独立的,即任何询问不依赖于此前询问的回答,也不决定之后的询问,则称 T 是非适应性的;否则,称A是适应性的。若存在 Π 的一个测试算法A满足其询问复杂性 $Q(\epsilon, n)$ 不依赖于 n ,则称 Π 是可测的。性质 Π 的询问复杂性定义为 Π 的测试器的询问复杂性的最小值。

研究性质测试与分析,主要基于如下理由:

尽管某些问题本身是多项式时间可解,但是,在某些条件下,由于无法承担扫描对象完整信息的花费,需在仅了解对象局部信息的环境下做出一个“合理”判断。

从应用角度上,可运行测试算法做预处理。若具备单边误差的测试算法拒绝,则不必运行精确判定程序。

对难解的判定问题做某种松弛的有效判定。

Rubinfeld与Sudan在研究程序验证时首次明确定义了函数的性质测试^[1]。Goldreich,Goldwasser与Ron引入组合对象的性质测试^[2],主要是图的性质。迄今为止,性质测试研究已经取得丰富的成果^[3-8],涉及到多个领域。

测试图的性质,有两个测试模型:邻接矩阵模型^[2]与有界度模型^[9]。不同模型中神谕的形式不同:在邻接矩阵模型中,使用顶点对查询,即询问两顶点是否邻接;在有界度模型中,使用邻域查询,即询问一个顶点的第 i 个邻接顶点。

在图的性质测试研究中,寻找可测图性质的结构特征是研究的难点。

在邻接矩阵模型中,已经发现,划分性质类^[2]及基于正则引理的一些性质^[3]是可测的。文献[10]得到了图可测性质的一个组合特征:性质 Π 是可测的当且仅当 Π 是可正则归约的。文献[11]将性质测试松弛为Tolerant性质测试,要求

到稿日期:2012-02-24 返修日期:2012-06-04 本文受国家自然科学基金(61262006),贵州省研究生创新基金(2010005)资助。

韦立(1981-),男,博士生,主要研究方向为可计算性与计算复杂性,E-mail: weilikb@163.com;许道云(1959-),男,教授,博士生导师,主要研究方向为可计算性与计算复杂性;王正才(1977-),男,博士生,主要研究方向为计算复杂性、信息安全。

$Tolerant(\epsilon, \delta)$ -测试算法以高概率接受 ϵ -靠近性质的对象并以高概率拒绝 δ -远离性质的对象。文献[12]定义了性质的 (ϵ, δ) -可估计概念。若对任意 $0 < \epsilon < \delta \leq 1$, Π 是 (ϵ, δ) -可估计的, 则称 Π 是可估计的。性质 Π 是可测的, 则 Π 是可估计的^[12]。可将性质 Π 的任意适应性测试算法转化为非适应性测试算法, 后者的询问次数是关于前者询问次数的多项式^[13]。进一步考察适应性测试与非适应性测试的询问复杂性之间的间隙, 文献[14]证明: 存在性质 Π , Π 的适应性测试的询问复杂性为 $O(\epsilon^{-1} \text{poly}(\log \epsilon^{-1}))$, 但使用非适应性测试算法, 仅 $O(\epsilon^{-3/2})$ 次询问是不够的; 另一方面, 存在性质 P , P 的非适应性测试的询问复杂性为 $O(\epsilon^{-1} \text{poly}(\log \epsilon^{-1}))$, 但适应性测试的询问复杂性的下界为 $\Omega(\epsilon^{-1})$ 。这表明, 某些性质的非适应性测试与适应性测试的询问复杂性的间隙可由多项式进一步缩小。

在有界度模型中, 连通性、 k -边-连通性、无环性等性质是可测的^[9]。不包含弱膨胀图的遗传性质是可测的^[15], 这是目前获取的唯一的可测试度有界图性质的特征。

子图同构问题的测试算法具有非常重要的应用。例如, 给一种化学分子一个唯一的命名是化学中一个重要的工作, 通常用图来表示化学分子, 对同一种分子, 尽管它可能有多个表示图, 但这些表示图是同构的。识别同构的表示图, 便可给出分子唯一的命名。又如, 对某些布尔公式(鸽巢公式)进行消解, 若允许对同构的子公式只进行一次消解, 则消解树的复杂性可降到多项式。

设 H 是一个图, 称 G 包含一个 H -子图, 若 G 中存在一个与 H 同构的诱导子图。令 H -free 为不含 H -子图的图所构成的集合。在有界度模型中, 对任意连通图 H , 文献[9]设计了一个基于宽度优先搜索的 H -free 测试算法, 其询问复杂性为 $O(\epsilon^{-1} d^{\text{diam}(H)})$, 其中 $\text{diam}(H)$ 为 H 的直径。由此, 对任意连通图 H , H -free 在有界度模型中是可测的, 但当 H 不连通时, 算法失效。

在邻接矩阵模型中证明了, 对任意 H , 无论 H 连通与否, 性质 H -free 是可测的。

2 基础知识

本文所考虑的图为无向简单图。令 $[n] = \{1, 2, \dots, n\}$, 不失一般性, 用 $[n]$ 表示 n 顶点图的顶点集。用 $A \Delta B$ 表示集合 A 与集合 B 的对称差。图 G 的度 $\deg(G)$ 定义为 G 中顶点的度的最大值。一个图性质 Π 是一个在同构下封闭的图类, 即满足: 若 $G \in \Pi$, $G \cong H$, 则 $H \in \Pi$ 。下文讨论中, 约定 n 足够大。我们先介绍“远离”与“靠近”两个概念。

设 $d: O \times O \rightarrow [0, 1]$ 是所考虑的对象类 O 上一个距离测度, $C \subseteq O$, 对 $w \in O$, $d(w, C) := \min\{d(w, v) : v \in C\}$, 若 $d(w, C) > \epsilon$, 则称 w 是 ϵ -远离性质 C 的; 否则, 称 w 是 ϵ -靠近性质 C 的。

在邻接矩阵模型中, 用邻接矩阵表示图, 询问图的信息, 即询问邻接矩阵中某个位置的值, 询问方式为顶点对查询。

定义 1 设图 $G = ([n], E_G)$, $H = ([n], E_H)$, Π 为一图性质, 定义距离如下:

(1) G 与 H 之间的距离 $d(G, H) := |E_G \Delta E_H| \cdot n^{-2}$;

(2) G 与 Π 的距离 $d(G, \Pi) := \min_{H \in \Pi} d(G, H)$ 。

定义 2 设 Π 是一个图性质, Π 的一个测试器是一个概率神谕机 T , 在输入规模参数 n 与距离参数 ϵ 上, 询问对象 n 顶点图 G 的局部信息, 询问形式为 $(i, j) \in E_G$?, 输出 0/1 (拒绝/接受), 满足如下两个条件:

(1) 若 $G \in \Pi$, 则 T 输出 1 的概率至少为 $2/3$;

(2) 若 $d(G, \Pi) > \epsilon$, 则 T 输出 1 的概率至多为 $1/3$ 。

在有界度模型中, 将图表示为 $f: [n] \times [d] \rightarrow [n] \cup \{\perp\}$ 的形式, 其中 f 满足: ① $f(w, i) = v$ 当且仅当存在 $j \in [d]$ 使得 $f(v, j) = w$; ② $E = \{\{u, v\} : \exists i(f(u, i) = v)\}$ 。 $f(w, i) = v$ 表明 v 为 w 的第 i 个邻接点, $f(w, i) = \perp$ 表明 w 的邻接点的个数少于 i 个。

定义 3 设图 $G = ([n], E_G)$, $H = ([n], E_H)$, 满足 $\deg(G) \leq d$, $\deg(H) \leq d$, Π 为一图性质, 定义距离如下:

(1) G 与 H 之间的距离 $d(G, H) := |E_G \Delta E_H| \cdot (dn)^{-1}$;

(2) G 与 Π 的距离 $d(G, \Pi) := \min_{H \in \Pi} d(G, H)$ 。

定义 4 设 Π 的一个图性质, Π 的一个测试器 T 是一个概率神谕机, 在输入参数 n, d 与 ϵ 上, 询问关于 n 顶点图 G 的信息, 询问形式为邻域查询, 输出 0/1, 满足如下条件:

(1) 若 $G \in \Pi$, 则 T 输出 1 的概率至少为 $2/3$;

(2) 若 $d(G, \Pi) > \epsilon$, 则 T 输出 1 的概率至多为 $1/3$ 。

一般来说, 对度有界图性质, 采用有界度模型比较合理, 但对稠密图的性质而言, 邻接矩阵模型比较合理。事实上, 存在性质 Π 使得在有界度模型中, 它是可测的, 但在邻接矩阵模型中, 它却是不可测的。例如, 性质“无孤立顶点”; 同样, 也存在性质 P , 在邻接矩阵模型中, 它是可测的, 但在有界度模型中, 它却是不可测的, 如二分性^[2,9]。

3 正则性与随机性

设 A, B 为图 G 的两个不相交的顶点子集, 用 $e(A, B)$ 表示 G 中一端点在 A 中另一端点在 B 中的边的数目, (A, B) 的边稠密度 $d(A, B) = e(A, B) / (|A| |B|)^{-1}$ 。称 (A, B) 是 λ -正则的, 若任意 $A' \subset A$ 及 $B' \subset B$ 满足 $|A'| \geq \lambda |A|$, $|B'| \geq \lambda |B|$, 均有 $|d(A, B) - d(A', B')| < \lambda$ 。

在正则对的定义中, 只需要任意 $A' \subset A$ 及 $B' \subset B$ 满足 $|A'| = \lambda |A|$, $|B'| = \lambda |B|$, 均有 $|d(A, B) - d(A', B')| < \lambda$ 即可。根据定义, 我们有如下引理。

引理 1 设 (A, B) 是稠密度为 η 的 λ -正则对, $Y \subseteq B$ 且 $|Y| \geq \lambda |B|$, 则 A 中存在多于 $(1 - \lambda) |A|$ 个在 Y 中含有多于 $(\eta - \lambda) |Y|$ 个邻接点的顶点。

证明: 令 X 为 A 中满足在 Y 中有少于 $(\eta - \lambda) |Y|$ 个邻接点的顶点集, 则 $e(X, Y) < |X| (\eta - \lambda) |Y|$, 从而

$$d(X, Y) = e(X, Y) / (|X| |Y|)^{-1} < \eta - \lambda$$

若 $|X| \geq \lambda |A|$, 由 (A, B) 的正则性有, $d(X, Y) > \eta - \lambda$, 与 $d(X, Y) < \eta - \lambda$ 矛盾, 故 $|X| < \lambda |A|$ 。故 A 中存在多于 $(1 - \lambda) |A|$ 个在 Y 中含有多于 $(\eta - \lambda) |Y|$ 个邻接点的顶点。

引理 2 设 (A, B) 是稠密度为 η 的 λ -正则对, $\delta \geq \lambda$, 对任意 $X \subset A$, $Y \subset B$ 满足 $|X| \geq \delta |A|$, $|Y| \geq \lambda |B|$, (X, Y) 是 $\max\{2\lambda, \delta^{-1}\lambda\}$ -正则的且稠密度介于 $\eta - \lambda$ 与 $\eta + \lambda$ 之间。

设 $S = \{\eta_{i,j} : 0 \leq \eta_{i,j} \leq 1, 1 \leq i < j \leq k\}$, $V_1, V_2, \dots, V_k$ 是 k 个规模为 m 的两两互不相交的顶点集, 由 V_i, V_j 支撑的二部图是一个边稠密度为 $\eta_{i,j}$ 的随机二部图。设 H 是一个 k 顶点图, $\sigma: [k] \rightarrow [k]$ 是一个置换。令

$$N(S, H, \sigma) = \prod_{(i,j) \in E_H} \eta_{\sigma(i), \sigma(j)} \prod_{(i,j) \notin E_H} (1 - \eta_{\sigma(i), \sigma(j)})$$

考虑对如下条件(c):

(c) $v_1 \in V_1, v_2 \in V_2, \dots, v_k \in V_k$, 由 $\{v_1, v_2, \dots, v_k\}$ 支撑的子图在 σ 下(σ 将 i 映到 $v_{\sigma(i)}$)与 H 同构。

显然, 任意 k 元组 $v_1 \in V_1, \dots, v_k \in V_k$ 满足条件(c)的概率为 $N(S, H, \sigma)$, 从而随机图中满足条件(c)的 k 元组的数目的期望值为 $N(S, H, \sigma)m^k$ 。利用正则性代替上述讨论中的随机性, 如下引理表明, 当正则性足够好时, $N(S, H, \sigma)m^k$ 是在正则性下满足条件(c)的 k 元组个数的一个近似值。

引理 3 对任意 $\delta > 0$ 及正整数 k , 存在 $\lambda = \lambda_3(\delta, k)$ 使得如下性质成立:

设 $V_1, V_2, \dots, V_k$ 是规模为 m 的两两不相交的顶点集, 任意 (V_i, V_j) 是 λ -正则的。令 $S = \{\eta_{i,j} : \eta_{i,j} = d(V_i, V_j), 1 \leq i < j \leq k\}$, 则对任意 k 顶点图 H 及置换 $\sigma: [k] \rightarrow [k]$, k 部图中满足条件(c)的 k 元组的个数介于 $(N(S, H, \sigma) - \delta)m^k$ 与 $(N(S, H, \sigma) + \delta)m^k$ 之间。

证明: 固定 H , 证明存在 $\lambda' = \lambda_3(\delta, k)$ 使得满足条件(c)的 k 元组个数介于 $(N(S, H, \sigma) - \delta)m^k$ 与 $(N(S, H, \sigma) + \delta)m^k$ 之间。由于 k 顶点图至多有 $2^{\binom{k}{2}}$ 个, 为了使对任意图 H 引理成立, 只需令 $\lambda = \lambda'(2^{\binom{k}{2}}, \delta, k)$ 即可。

设 H 为团, 在此情形下设 $\eta_{i,j} \geq \delta$, 用归纳法证明。当 $k = 2$ 时, 引理成立。

令 $\lambda' = \min\{\frac{\delta}{4} \binom{k-1}{2}^{-1}, \frac{\delta}{4(k-1)}, \frac{3\delta}{4} \lambda'(\frac{\delta}{4}, k-1)\}$, 考虑 $V_{\sigma(k)}$ 中的顶点。对任意 $1 \leq i < k$, 由上述引理 1, $V_{\sigma(k)}$ 中存在多于 $(1 - \frac{\delta}{4(k-1)})m$ 个满足在 $V_{\sigma(i)}$ 中有多于 $(\eta_{\sigma(i), \sigma(k)} - \frac{\delta}{4})m$ 个邻接点的顶点。因此 $V_{\sigma(k)}$ 中存在多于 $(1 - \delta/4)m$ 个满足在 $V_{\sigma(i)}$ ($1 \leq i < k$) 中有多于 $(\eta_{\sigma(i), \sigma(k)} - \frac{\delta}{4})m$ 个邻接点的顶点。对 $V_{\sigma(k)}$ 中满足此条件的任意顶点 v 及 $1 \leq i < k$, 令 $V'_{\sigma(i)}$ 为 v 在 $V_{\sigma(i)}$ 中的邻接顶点集及 $\eta'_{\sigma(i), \sigma(j)} = d(V'_{\sigma(i)}, V'_{\sigma(j)})$ 。由于 $|V'_{\sigma(i)}| \geq (\eta_{\sigma(i), \sigma(k)} - \frac{\delta}{4(k-1)})m \geq \frac{3}{4}\delta m \geq \lambda'm$, 且由引理 2, $(V'_{\sigma(i)}, V'_{\sigma(j)})$ 是 $\frac{4}{3\delta}\lambda'$ -正则的, 因此它是 $\lambda'(\frac{\delta}{4}, k-1)$ -正则的, 并且 $\eta'_{\sigma(i), \sigma(j)} \geq \eta_{\sigma(i), \sigma(j)} - \frac{\delta}{4} \binom{k-1}{2}^{-1}$ 。

由归纳假设, $V'_{\sigma(1)}, V'_{\sigma(2)}, \dots, V'_{\sigma(k-1)}$ 中至少含有 $(\prod_{1 \leq i < j < k} \eta'_{\sigma(i), \sigma(j)} - \frac{\delta}{4})m^{k-1}$ 个 $(k-1)$ -团。而

$$\begin{aligned} & (\prod_{1 \leq i < j < k} \eta'_{\sigma(i), \sigma(j)} - \frac{\delta}{4})m^{k-1} \geq (\prod_{1 \leq i < j < k} \eta_{\sigma(i), \sigma(j)} - \frac{2\delta}{4})m^{k-1} \geq \\ & (\prod_{1 \leq i < j \leq k} \eta_{\sigma(i), \sigma(j)} - \frac{3\delta}{4})m^{k-1} \end{aligned}$$

注意到 v 与 $V'_{\sigma(1)}, V'_{\sigma(2)}, \dots, V'_{\sigma(k-1)}$ 中 $(k-1)$ -团可构成一

个 k -团。由于 $V_{\sigma(k)}$ 中存在多于 $(1 - \delta/4)m$ 个这样的顶点, 故 k -团的个数至少为

$$(\prod_{1 \leq i < j \leq k} \eta_{\sigma(i), \sigma(j)} - \frac{3\delta}{4})m^{k-1}(1 - \frac{\delta}{4})m \geq (I(H, W, \sigma) - \delta)m^k$$

上界的证明与下界的证明类似。

设 $S = \{\eta_{i,j} : 0 \leq \eta_{i,j} \leq 1, 1 \leq i < j \leq k\}$, H 是一个 k 顶点图, 令 $N(S, H) = \frac{1}{\text{aut}(H)} \sum_{\sigma} N(S, H, \sigma)$, 其中 $\text{aut}(H)$ 为 H 的自同构的个数。类似地, 设 $V_1, V_2, \dots, V_k$ 是 k 个规模均为 m 的顶点集, 由 V_i, V_j 支撑的二部图是边稠密度为 $\eta_{i,j}$ 的随机二部图, 此随机图中与 H 同构的子图的个数的期望值为 $N(S, H)m^k$ 。用正则性代替上述讨论中的随机性, 可有如下结论。

引理 4 对任意 $\delta > 0$ 及正整数 k , 存在 $\lambda = \lambda_4(\delta, k)$ 使得如下性质成立:

设 $V_1, V_2, \dots, V_k$ 是 k 个规模均为 m 的两两互不相交的顶点集, 任意 (V_i, V_j) 都是 λ -正则的。令 $S = \{\eta_{i,j} : \eta_{i,j} = d(V_i, V_j), 1 \leq i < j \leq k\}$, 则对任意 k 顶点图 H , k 部图中与 H 同构的诱导子图(顶点来自不同 V_i)的个数介于 $[N(S, H) - \delta]m^k$ 与 $[N(S, H) + \delta]m^k$ 之间。

证明: 令 $\lambda_4(\delta, k) = \lambda_3(\frac{\delta}{k!}, k)$, 由引理 3, 对任意置换 σ , 满足条件(c)的 k 元组个数介于 $[N(S, H, \sigma) - \frac{\delta}{k!}]m^k$ 与 $[N(S, H, \sigma) + \frac{\delta}{k!}]m^k$ 之间。设 $v_1 \in V_1, \dots, v_k \in V_k$ 支撑的子图 H' 与 H 同构, 则存在 $\text{aut}(H)$ 个置换 σ , 使得 $v_1, \dots, v_k$ 满足条件(c), 故图中与 H 同构的子图的个数介于 $\frac{1}{\text{aut}(H)} \sum_{\sigma} [N(S, H, \sigma) - \frac{\delta}{k!}]m^k$ 与 $\frac{1}{\text{aut}(H)} \sum_{\sigma} [N(S, H, \sigma) + \frac{\delta}{k!}]m^k$ 之间, 即 $[N(S, H) - \delta]m^k$ 与 $[N(S, H) + \delta]m^k$ 之间。

4 主要结果

定理 1 在邻接矩阵模型中, 对任意 H , 性质 H -free 是可测的。

先介绍定理证明所需的相关概念及基本结论。

如下引理中, 我们利用概率方法证明, 少数边修改可将一个几乎满足稠密度为 η 的 λ -正则对转变为一个稠密度恰好为 η 的 λ -正则对。

引理 5 对任意 $0 < \delta \leq \lambda \leq 1$, 如下性质成立:

设 A, B 是两个规模均为 $m \geq m_0(\delta, \lambda)$ 的不相交顶点集, 其稠密度 $d(A, B) = \eta$ 。若对任意的规模为 λm 的子集 $A' (\subset A), B' (\subset B)$, 其稠密度 $d(A', B')$ 介于 $\eta - (\lambda + \delta)$ 与 $\eta + (\lambda + \delta)$ 之间, 则 $(2\delta\lambda^{-1})m^2$ 次边修改可将 (A, B) 变为稠密度为 η 的 λ -正则对的概率大于 0。

证明: 考虑如下两阶段操作:

阶段 1 对任意两顶点 $a \in A, b \in B$, 进行如下操作: 抛第一枚头朝上的概率为 $\delta(\delta + \lambda)^{-1}$ 的硬币 coin_1 。若 coin_1 头朝下, 则顶点对 (a, b) 不做任何修改, 跳到下一个顶点对; 若 coin_1 头朝上, 则放弃 a 与 b 之间的邻接关系, 然后抛第二枚头朝上的概率为 η 的硬币 coin_2 。若 coin_2 头朝上, 则在 a 与 b 之间添加边; 否则, 不做任何修改。

阶段2 确定性的添加或者删除至多 $m^{1.5}$ 条边, 确保 (A, B) 的稠密度为 η 。

关于阶段1的操作, 有以下几个断言。

断言1 至少以 $3/4$ 的概率, 阶段1中至多进行 $1.5\delta\gamma^{-1}m^2$ 次边修改。

证明: 注意到对任意顶点 (a, b) , 只有 coin_1 头朝上时, 才可能进行边修改。因而阶段1中边修改的次数至多为 coin_1 头朝上的次数。 coin_1 头朝上次数的期望为 $\delta(\delta+\gamma)^{-1}m^2$ 。根据 Chernoff 界, 修改次数偏离其期望值多于 $\delta m^2/2$ 的概率至多为 $2\exp\{-2(\delta/2)^2 m^2\}$ 。故当 m 足够大时, 即 $m \geq m_5(\delta, \lambda)$, 至多进行 $\delta(\delta+\gamma)^{-1}m^2 + (\delta m^2/2) \leq 1.5\delta\gamma^{-1}m^2$ 次边修改的概率大于 $3/4$ 。

断言2 阶段1修改后, $d(A, B)$ 介于 $\eta - 1/m^{0.5}$ 与 $\eta + 1/m^{0.5}$ 之间的概率大于 0 。

证明: 对任意 $a \in A, b \in B, a, b$ 由邻接变为不邻接的概率为 $\delta(\delta+\lambda)^{-1} - \delta\eta(\delta+\lambda)^{-1}$, 由不邻接变为邻接的概率为 $\delta\eta(\delta+\lambda)^{-1}$ 。故经此阶段1后, (A, B) 的边稠密度的期望值为 $\eta(1 - \delta(\delta+\lambda)^{-1} + \delta\eta(\delta+\lambda)^{-1}) + (1-\eta)\delta\eta(\delta+\lambda)^{-1} = \eta$ 。

由 Chernoff 界, $|d(A, B) - \eta| > 1/m^{0.5}$ 的概率至多为 $2\exp\{-2(1/m^{0.5})^2 m^2\}$ 。当 m 足够大时, 此概率严格小于 1 , 从而 $d(A, B)$ 介于 $\eta - 1/m^{0.5}$ 与 $\eta + 1/m^{0.5}$ 之间的概率大于 0 。

断言3 所有子集对 (A', B') 的稠密度 $d(A', B')$ 介于 $\eta - (\lambda + \delta/2)$ 与 $\eta + (\lambda + \delta/2)$ 之间的概率大于 0 , 其中, $A' \subset A, B' \subset B, |A'| = |B'| = \lambda m$ 。

证明: 固定 (A', B') , $e(A', B')$ 表示初始时 A' 与 B' 之间的边数, 经修改后, A' 与 B' 之间的边数 $e'(A', B')$ 的期望为

$$\begin{aligned} & e(A', B')(1 - \frac{\delta}{\delta+\lambda} + \frac{\delta\eta}{\delta+\lambda}) + [|A'| \parallel B'| - e(A', \\ & B')] \frac{\delta\eta}{\delta+\lambda} \\ & = e(A', B')(1 - \frac{\delta}{\delta+\lambda}) + |A'| \parallel B'| \frac{\delta\eta}{\delta+\lambda} \end{aligned}$$

初始时, $e(A', B')$ 介于 $|A'| \parallel B'|(\eta - (\lambda + \delta))$ 与 $|A'| \parallel B'|(\eta + (\lambda + \delta))$ 之间, 故 $e'(A', B')$ 的期望的下界为 $|A'| \parallel B'|[(\eta - (\lambda + \delta))(1 - \frac{\delta}{\delta+\lambda}) + \frac{\delta\eta}{\delta+\lambda}] = |A'| \parallel B'|(\eta - \lambda)$ 。

类似地, $e'(A', B')$ 的期望的上界为

$$|A'| \parallel B'|[(\eta + (\lambda + \delta))(1 - \frac{\delta}{\delta+\lambda}) + \frac{\delta\eta}{\delta+\lambda}] = |A'| \parallel B'|(\eta + \lambda)$$

由 Chernoff 界, $|e'(A', B') - E(e'(A', B'))| > \frac{\delta}{2} |A'|$ 的概率至多为 $2\exp\{-2(\delta/2)^2 (\lambda m)^2\}$ 。

当 m 足够大时, 所有 (A', B') 满足 $|e'(A', B') - E(e'(A', B'))| \leq \frac{\delta}{2} |A'|$ 的概率大于 0 , 即修改后所有 (A', B') 的稠密度 $d(A', B')$ 满足 $\eta - (\lambda + \delta/2) \leq d(A', B') \leq \eta + (\lambda + \delta/2)$ 的概率大于 0 。

由上述几个断言, 至多进行 $1.5\delta\lambda^{-1}m^2$ 次边修改, (A, B) 以非零概率具备如下性质:

- (1) $\eta - 1/m^{0.5} < d(A, B) < \eta + 1/m^{0.5}$;
- (2) 任意规模为 λm 的 $A' \subset A, B' \subset B, \eta - (\gamma + \delta/2) \leq d$

$(A', B') \leq \eta + (\gamma + \delta/2)$ 成立。

当 m 足够大时, $1/m^{0.5} \leq \delta/2$ 。在阶段2, 确定性的添加或者删除至多 $m^{1.5}$ 条边可使得 $d(A, B) = \eta$ 。对任何规模为 λm 的 $A' \subset A, B' \subset B, d(A', B')$ 至多改变 $(1/m^{0.5})\lambda^2 \leq \delta/2$, 从而有 $\eta - \lambda < d(A', B') < \eta + \lambda$, 因而 (A, B) 是 λ -正则的。

由上述讨论, 故至多进行 $1.5\delta\lambda^{-1}m^2 + m^{1.5} \leq 2\delta\lambda^{-1}m^2$ 次边修改将 (A, B) 变为稠密度为 η 的 λ -正则对的概率大于 0 。

引理6 对任意 $0 < \delta \leq \lambda \leq 1$, 如下性质成立:

设 (A, B) 是边稠密度介于 $\eta - \delta$ 与 $\eta + \delta$ 之间的 λ -正则对, 其中 $|A| = |B| = m \geq m_6(\eta, \delta)$, 则至多 $2\delta m^2$ 次边修改可将 (A, B) 变为稠密度为 η 的 λ -正则对的概率大于 0 。

证明: 设 $d(A, B) = \eta + \zeta$, 其中 $|\zeta| \leq \delta$ 。当 $\zeta > 0$ 时, 分如下两种情形进行讨论。当 $\zeta < 0$, 可类似讨论。

(1) $\zeta \leq \delta\lambda^2$ 。独立、均匀地删除 ζm^2 条边。删除边后, (A, B) 的稠密度恰好为 η 。对任意规模为 λm 的子集 $A' \subset A, B' \subset B$, 初始时, $d(A', B')$ 介于 $\eta + \zeta - \lambda$ 与 $\eta + \zeta + \lambda$ 之间, 删除边后, 其稠密度仍然介于 $\eta + \zeta - \lambda$ 与 $\eta + \zeta + \lambda$ 之间。根据引理5, 至多 $2\zeta\lambda^{-1}m^2$ 次边修改可将 (A, B) 变为稠密度为 η 的 λ -正则对。因此, $(\zeta + 2\zeta\lambda^{-1})m^2 \leq (\delta\lambda^2 + 2\delta\lambda)m^2$ 次边修改可将 (A, B) 变为稠密度为 η 的 λ -正则对。

(2) $\zeta \geq \delta\lambda^2$ 。将 (A, B) 变为稠密度恰好为 η 的 λ -正则对的方案分为两个阶段。第一个阶段, 随机删除一些边; 第二阶段, 确定性地添加或删除一定数量的边以确保其稠密度为 η 。

在第一阶段, 以 $\zeta(\eta + \zeta)^{-1}$ 的概率独立地删除每一条边。则被删除边的数目的期望为

$$(\eta + \zeta) |A| \parallel B| \zeta(\eta + \zeta)^{-1} = \zeta |A| \parallel B| \leq \delta |A| \parallel B|。$$

删除边后 (A, B) 的稠密度 $d'(A, B)$ 的期望恰好为 η 。由于 $d(A, B) \geq \delta\lambda^2$, 根据 Chernoff 界, $|d'(A, B) - \eta| > 1/m^{0.5}$ 的概率至多为 $2\exp\{-2(1/m^{0.5})^2 \delta\lambda^2 m^2\}$ 。当 $m \geq m(\delta, \lambda)$ 时, 此概率不大于 $1/4$ 。故至多进行 $1.5\delta m^2$ 次边修改的概率大于 $3/4$ 。

对任意规模为 λm 的子集 $A' \subset A, B' \subset B$, 初始时, 由于 (A, B) 是 λ -正则的, 故 $d(A', B')$ 介于 $\eta + \zeta - \lambda$ 与 $\eta + \zeta + \lambda$ 之间。故删除边后其稠密度 $d'(A', B')$ 的期望的下界为 $(\eta + \zeta - \lambda)(1 - \zeta(\eta + \zeta)^{-1}) = \eta - \lambda + \zeta(\eta + \zeta)^{-1}$, 上界为 $(\eta + \zeta + \lambda)(1 - \zeta(\eta + \zeta)^{-1}) = \eta + \lambda - \zeta(\eta + \zeta)^{-1}$ 。

由 Chernoff 界, $|d'(A', B') - E(d'(A', B'))| > \delta/2$ 的概率至多为 $2\exp\{-2(\delta/2)^2 (\eta + \zeta - \lambda)(\lambda m)^2\}$ 。

当 $m \geq m(\eta, \delta)$ 时, 所有 (A', B') 满足 $|d'(A', B') - E(d'(A', B'))| < \delta/2$ 的概率大于 0 。

当 m 足够大时, $1/m^{0.5} \leq \delta/2$ 。故在第二阶段中, 确定性地添加或者删除至多 $m^{1.5}$ 条边, 确保其稠密度为 η 。对任意满足上述条件的 (A', B') , $d(A', B')$ 至多改变 $(1/m^{0.5})\lambda^2 \leq \delta/2$, 从而有 $\eta - \lambda < d(A', B') < \eta + \lambda$, 故 (A, B) 是 λ -正则的。

综上所述, 至多进行 $2\delta m^2$ 次边修改可将稠密度介于 $\eta - \delta$ 与 $\eta + \delta$ 的 λ -正则对转换为稠密度恰好为 η 的 λ -正则对。

称图 G 的一个划分 $\mathcal{A} = \{V_1, V_2, \dots, V_k\}$ 为一个均分, 若对任意 $i \neq j, ||V_i| - |V_j|| \leq 1$ 。均分 $\mathcal{A}$ 的一个细分是一个均分 $\mathcal{B} = \{V_{i,j}, 1 \leq i \leq k, 1 \leq j \leq l\}$, 其中 $V_{i,j} \subseteq V_i (1 \leq i \leq k, 1 \leq j \leq l)$ 。均分 $\mathcal{A}$ 的阶是 $\mathcal{A}$ 的规模。

定义 5 称图 G 的一个均分 $\mathcal{A} = \{V_1, V_2, \dots, V_k\}$ 是 λ 正则的, 如果至少存在 $(1-\lambda) \binom{k}{2}$ 个对 (i, j) 使得 (V_i, V_j) 是 λ 正则的, 这里 $1 \leq i < j \leq k$.

Szemerédi 正则引理可描述如下。

引理 7^[6] 对任意 $\lambda > 0$ 及正整数 m , 存在 $T = T_7(m, \lambda)$ 使得对任意 $n (\geq T)$ 顶点图 G 的任一阶至多为 m 的均分 $\mathcal{A}$, 存在一个阶为 $k (m \leq k \leq T)$ 的细分 $\mathcal{B}$, $\mathcal{B}$ 是 λ 正则的。

定义 6 称 $R = (\lambda, k, \{\eta_{i,j}\}_{1 \leq i < j \leq k}, \bar{R})$ 为一个正则实例, 若 R 满足: $\lambda \in (0, 1]$; k 是一个正整数; 对任意 $1 \leq i < j \leq k$, $0 \leq \eta_{i,j} \leq 1$; $\bar{R}$ 是由 (i, j) 组成的规模至多为 $\lambda \binom{k}{2}$ 的集合。称图 G 满足正则实例 R , 若 G 存在一个均分 $V_1, V_2, \dots, V_k$ 使得对任意 $(i, j) \notin \bar{R}$, (V_i, V_j) 是 λ 正则的且 $d(V_i, V_j) = \eta_{i,j}$ 。正则实例 R 的复杂度 $c(R)$ 定义为 $\max\{k, 1/\lambda\}$ 。

文献[10]证明, 对任意正则实例 R , 性质“满足 R ”是可测的。

定义 7 图性质 Π 是可正则归约的, 若对任意 $\delta > 0$, 存在 $\kappa = \kappa_{\Pi}(\delta)$ 满足: 对任意 n , 存在规模不大于 κ 的正则实例簇 $\mathcal{R} = \{R: c(R) \leq \kappa\}$, 使得对任意 $\epsilon > 0$ 及 n 顶点图 G , 如下条件成立:

(i) 若 G 满足性质 Π , 则存在 $R \in \mathcal{R}$ 使得 G 是 δ 靠近满足 R ;

(ii) 若 G 是 ϵ -远离性质 Π , 则对任意 $R \in \mathcal{R}$, G 是 $(\epsilon - \delta)$ -远离满足 R 。

在邻接矩阵模型中, 文献[10]证明了性质的可测性与可正则归约性是等价的。

定理 1 的证明: 设 $H = ([k], E_H)$, 只需证明 H -free 是可正则归约的。对任意 $\delta > 0$, 令 $\lambda' = \lambda_4(\delta, k)$, $\lambda = \min\{\lambda', \delta\}$ 。定义 $\mathcal{R}$ 为满足如下条件的正则实例构成的集合:

- (1) 正则参数为 λ ;
- (2) 阶至少为 $\max\{k, 1/\lambda\}$, 至多为 $T_7(\max\{k, 1/\lambda\}, \lambda)$;
- (3) 稠密度 $\eta_{i,j} \in \{0, \lambda, 2\lambda, \dots, 1\}$;
- (4) 不存在 $j_1, j_2, \dots, j_k$ 使得对任意 $j_s, j_t, \eta_{j_s, j_t} > 0$ 当仅当 $(s, t) \in E_H$ 。

显然, $\mathcal{R}$ 的规模与任意 $R \in \mathcal{R}$ 的复杂性均受限於一个关于 δ, k 的函数 $f(\delta, k)$ 。下面证明此归约是有效的。

若 G 是 ϵ -远离 H -free, 假设存在 $R \in \mathcal{R}$, G 是 $(\epsilon - \delta)$ -靠近性质满足正则实例 R 。根据假设, 只需修改 G 中至多 $(\epsilon - \delta)n^2$ 条边可使其满足 R 。现设均分 $\{V_1, V_2, \dots, V_l\}$ 满足 R , 再删除 $V_1, V_2, \dots, V_l$ 中的边, 至多有 $l^{-1}n^2 \leq \lambda n^2 \leq \delta n^2$ 条。上述两步操作至多修改了 ϵn^2 条边, 由 (4) 知, 修改后的图不包含同构于 H 的子图, 这与 G 是 ϵ -远离 H -free 矛盾。

若 $G \in H$ -free, 由正则引理, 存在 G 的一个阶为 l 的 λ -正则均分 $\{V_1, \dots, V_l\}$, 其中, $\max\{k, 1/\lambda\} \leq l \leq T_7(\max\{k, 1/\lambda\}, \lambda)$ 。根据 λ' 的定义及 $\lambda \leq \lambda'$ 知, 不存在 $j_1, j_2, \dots, j_k$ 使得 $(V_{j_s}, V_{j_t}) (1 \leq s < t \leq k)$ 是 λ 正则的且满足 $\eta_{j_s, j_t} > 0$ 当仅当 $(s, t) \in E_H$, 否则存在诱导子图与 H 同构。利用此划分, 构造正则实例 R 如下: R 的正则参数、阶与上述划分的正则参数、阶一致, 稠密度 $\eta_{i,j}$ 的选择与 $d(V_i, V_j)$ 相关, 它们至多相差 λ 。

利用引理 6, 对划分中的每个正则对 (V_i, V_j) , 在保持正则性的同时, 至多需要 $2\lambda(n/l)^2$ 边修改使其稠密度恰好为 $\eta_{i,j}$ 。由于划分中至多含有 $l^2/2$ 个正则对, 因而至多需要 $\lambda(n/l)^2 l^2 = \lambda n^2 \leq \delta n^2$ 次边修改可使上述划分满足 R , 即修改后得到的图 G' 满足 R , 从而图 G 是 δ -靠近满足 R 。

结束语 目前, 正则引理是理解稠密图的可测试性特征的理论工具, 但对于有界度图的可测试性与一般稀疏图的可测试性的特征刻画, 一般性的理论成果较少。进一步工作中, 我们希望能去寻找此类性质的更多特征。

参考文献

- [1] Rubinfeld R, Sudan M. Robust characterizations of polynomials with applications to program testing [J]. SIAM Journal on Computing, 1996, 25(2): 252-271
- [2] Goldreich O, Goldwasser S, Ron D. Property testing and its connection to learning and approximation [J]. JACM, 1998, 45(4): 653-750
- [3] Alon N, Fischer E, Krivelevich M, et al. Efficient testing of large graphs [J]. Combinatorica, 2000, 20: 451-476
- [4] Ron D. Property testing: A learning theory perspective [J]. Foundations and Trends in Machine Learning, 2008, 1(3): 307-402
- [5] Goldreich O. Introduction to testing graph properties [R]. TR10-082. ECCC, 2010
- [6] Chen H. Testing algebraic geometric codes [J]. Science China Series A, Mathematics, 2009, 52: 2171-2176
- [7] Valiant P. Testing symmetric property of distributions [A] // Proc. of the 40th Annual ACM STOC, 2008 [C]. Victoria, British Columbia, Canada: ACM, 2008: 383-392
- [8] Ben-Sasson E, Sudan M. Robust locally testable codes and products of codes [J]. Random Structures and Algorithms, 2006, 28(4): 387-402
- [9] Goldreich O, Ron D. Property Testing in Bounded Degree Graphs [J]. Algorithmica, 2002, 32(2): 302-343
- [10] Alon N, Fischer E, Newman I, et al. A combinatorial characterization of the testable graph properties: it's all about regularity [J]. SIAM J. Computing, 2009, 39: 143-167
- [11] Parnas M, Ron D, Rubinfeld R. Tolerant property testing and distance approximation [J]. J. of Computer and System Sciences, 2006, 72(6): 1012-1042
- [12] Fischer E, Newman I. Testing versus estimation of graph properties [J]. SIAM Journal on Computing, 2007, 37(2): 482-501
- [13] Goldreich O, Trevisan L. Three theorems regarding testing graph properties [J]. Random Structures and Algorithms, 2003, 23(1): 23-57
- [14] Goldreich O, Ron D. Algorithmic aspects of property testing in the dense graphs model [J]. SIAM J. Comput., 2011, 40(2): 376-445
- [15] Czumaj A, Shapira A, Sohler C. Testing Hereditary Properties of Nonexpanding Bounded-Degree Graphs [J]. SIAM Journal on Computing, 2009, 38(6): 2499-2510
- [16] Szemerédi E. Regular partitions of graphs [A] // Proc. Colloque Inter. CNRS, 1978 [C]. Paris: CNRS, 1978: 399-401