Combinatorics, Probability and Computing (2015) 24, 646–657. © Cambridge University Press 2015
doi:10.1017/S0963548314000765

Easily Testable Graph Properties

NOGA ALON[1†] and JACOB FOX[2‡]

[1]Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
and
School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA
(e-mail: nogaa@tau.ac.il)

[2]Department of Mathematics, MIT, Cambridge, MA 02139-4307, USA
(e-mail: fox@math.mit.edu)

Received 14 October 2013; revised 3 October 2014; first published online 2 February 2015

A graph on n vertices is ϵ-far from a property $\mathcal{P}$ if one has to add or delete from it at least ϵn^2 edges to get a graph satisfying $\mathcal{P}$. A graph property $\mathcal{P}$ is strongly testable if for every fixed $\epsilon > 0$ it is possible to distinguish, with one-sided error, between graphs satisfying $\mathcal{P}$ and ones that are ϵ-far from $\mathcal{P}$ by inspecting the induced subgraph on a random subset of at most $f(\epsilon)$ vertices. A property is easily testable if it is strongly testable and the function f is polynomial in $1/\epsilon$, otherwise it is hard. We consider the problem of characterizing the easily testable graph properties, which is wide open, and obtain several results in its study. One of our main results shows that testing perfectness is hard. The proof shows that testing perfectness is at least as hard as testing triangle-freeness, which is hard. On the other hand, we show that being a cograph, or equivalently, induced P_3-freeness where P_3 is a path with 3 edges, is easily testable. This settles one of the two exceptional graphs, the other being C_4 (and its complement), left open in the characterization by the first author and Shapira of graphs H for which induced H-freeness is easily testable. Our techniques yield a few additional related results, but the problem of characterizing all easily testable graph properties, or even that of formulating a plausible conjectured characterization, remains open.

2010 *Mathematics subject classification*: Primary 05C35
Secondary 05C85

† Research supported in part by a USA–Israeli BSF grant, by an ISF grant, by the Israeli I-Core program and by the Oswald Veblen Fund.
‡ Research supported by a Simons Fellowship, by NSF grant DMS 1069197, by an MIT NEC Corporation Award, by a USA–Israeli BSF grant and by an Alfred P. Sloan Fellowship.

1. Introduction

Property testing is an active area of computer science where one wishes to quickly distinguish between objects that satisfy a property from objects that are far from satisfying that property. The study of this notion was initiated by Rubinfield and Sudan [22], and subsequently Goldreich, Goldwasser and Ron [14] started the investigation of property testers for combinatorial objects. Graph property testing in particular has attracted a great deal of attention. A *property* $\mathcal{P}$ is a family of (undirected) graphs closed under isomorphism. A graph G with n vertices is *ϵ-far from satisfying* $\mathcal{P}$ if one must add or delete at least ϵn^2 edges in order to turn G into a graph satisfying $\mathcal{P}$.

An *ϵ-tester* for $\mathcal{P}$ is a randomized algorithm, which given n and the ability to check whether there is an edge between a given pair of vertices, distinguishes with probability at least $2/3$ between the cases G satisfies $\mathcal{P}$ and G is ϵ-far from satisfying $\mathcal{P}$. Such an ϵ-tester is one-sided if, whenever G satisfies $\mathcal{P}$, the ϵ-tester determines this with probability 1. A property $\mathcal{P}$ is *strongly testable* if for every fixed $\epsilon > 0$ there exists a one-sided ϵ-tester for $\mathcal{P}$ whose query complexity is bounded only by a function of ϵ, which is independent of the size of the input graph.

Call a property $\mathcal{P}$ *easily testable* if it is strongly testable with a one-sided ϵ-tester whose query complexity is polynomial in ϵ^{-1}, and otherwise call $\mathcal{P}$ *hard*. This is analogous to classical complexity theory, where an algorithm whose running time is polynomial in the input size is considered fast, and otherwise slow. Call a hereditary graph property *extendable* if, for all but finitely many graphs in the family, there is a larger graph in the family containing it as an induced subgraph. Most of the well-known hereditary graph properties are extendable. As mentioned briefly in [3] and proved in detail in [15], there is a universal one-sided ϵ-tester for extendable hereditary graph properties which has query complexity at most quadratic in the minimum possible query complexity of an optimal one-sided ϵ-tester. Indeed, it samples d random vertices (for some d), and if the subgraph they induce is in $\mathcal{P}$, it accepts, and otherwise it rejects. The query complexity of this tester is $\binom{d}{2}$, and it is at least as accurate as any tester with query complexity at most $d/2$. The query complexity a lower bound for the running time of an ϵ-tester, and, if there is a polynomial-time recognition algorithm for membership in $\mathcal{P}$, the running time is polynomial in the query complexity. So while query complexity and running time are different notions, they are often of comparable order.

For a graph H, let $\mathcal{P}_H$ denote the property of being H-free, *i.e.*, it is the family of graphs which do not contain H as a subgraph. The triangle removal lemma of Ruzsa and Szemerédi [23] is one of the most influential applications of Szemerédi's regularity lemma. It states that for every $\epsilon > 0$ there is $\delta > 0$ such that any graph on n vertices with at most δn^3 triangles can be made triangle-free by removing at most ϵn^2 edges. The triangle removal lemma is equivalent to the fact that $\mathcal{P}_{K_3}$ is strongly testable. Indeed, the algorithm samples $t = 2\delta^{-1}$ triples of vertices uniformly at random, where δ is picked according to the triangle removal lemma, and accepts if none of them forms a triangle, and otherwise rejects. Any triangle-free graph is clearly accepted. If a graph is ϵ-far from being triangle-free, then it contains at least δn^3 triangles, and the probability that none of the sampled triples forms a triangle is at most $(1-\delta)^t < 1/3$. Notice that the query complexity depends on the bound in the triangle removal lemma. As observed by Ruzsa and Szemerédi, the triangle removal lemma gives a simple proof of Roth's theorem [21] that every dense subset of the integers contains a 3-term arithmetic progression. From Behrend's construction [7], which gives a large subset of the first n positive integers without a 3-term arithmetic progression, it follows that $\delta \leqslant \epsilon^{c \log \epsilon}$ in the triangle removal lemma. This implies that testing triangle-freeness is hard. Indeed, in the universal algorithm described earlier, in a random sample of d vertices, the expected number of triangles is at most δd^3, and hence in the universal one-sided ϵ-tester for triangle-freeness, $1/3 \leqslant \delta d^3$, or equivalently, $d \geqslant (3\delta)^{-1/3}$. As discussed earlier, the query complexity of any one-sided ϵ-tester for triangle-freeness is at least $d/2$.

The triangle removal lemma was extended in [3] (see also [2]) to the graph removal lemma. It says that for each $\epsilon > 0$ and graph H on h vertices there is $\delta = \delta(\epsilon, H) > 0$ such that every graph on n vertices with at most δn^h copies of H can be made H-free by removing at most ϵn^2 edges. The graph removal lemma similarly implies that testing H-freeness is strongly testable. The proof, which uses Szemerédi's regularity lemma, gives a bound on the query complexity which is a tower of height a power of ϵ^{-1}. This was somewhat improved recently by the second author [12] to a tower of height logarithmic in ϵ^{-1}. The first author [1] showed that H-freeness is easily testable if and only if H is bipartite.

For a graph H, let $\mathcal{P}_H^*$ denote the property of being induced H-free, *i.e.*, it is the family of graphs which do not contain H as an induced subgraph. The graph removal lemma was extended by the first author, Fischer, Krivelevich and Szegedy [3] to the induced graph removal lemma, which states that for every $\epsilon > 0$ and graph H on h vertices there is $\delta > 0$ such that any graph on n vertices with at most δn^h induced copies of H can be made induced H-free by adding or removing at most ϵn^2 edges. The induced graph removal lemma is equivalent to the fact that, for any graph H, the property $\mathcal{P}_H^*$ is strongly testable. The proof, which uses a strengthening of Szemerédi's regularity lemma, gives a bound on the query complexity which is wowzer of height a power of ϵ^{-1}, which is one higher in the Ackermann hierarchy than the tower function. This has recently been improved by Conlon and the second author [10] to the tower function.

The length of a path is the number of edges it contains, and we let P_k denote the path of length k. The first author and Shapira [4] showed that for any graph H other than the paths of length at most 3, a cycle of length 4, and their complements, testing induced H-freeness is hard. For H a path of length at most 2 or their complements, induced H-freeness is easily testable. They left open the cases that H is a path of length 3 or a cycle of length 4 (and equivalently its complement). Here we settle one of the two remaining cases.

Theorem 1.1. *Induced P_3-freeness is easily testable.*

A well-known result of Seinsche [24] gives a simple structure theorem for induced P_3-free graphs. These graphs, also known as cographs, are generated from the single vertex graph by complementation and disjoint union. This is equivalent to the statement that every induced P_3-free graph or its complement is not connected.

A general result of the first author and Shapira [5] states that every hereditary family $\mathcal{P}$ of graphs is strongly testable. They further asked which hereditary graph properties are easily testable, and, in particular, for a few of the well-known hereditary families of graphs, including perfect graphs and comparability graphs.

Note that the chromatic number of a graph is at least its clique number as the vertices of any clique must receive different colours in a proper colouring. A graph is *perfect* if every induced subgraph of it satisfies that its clique number and chromatic number are equal. The study of perfect graphs was started by Berge, partly motivated by the study of the Shannon capacity in information theory, which lies between the clique number and chromatic number of a graph. Perfect graphs form a relatively large class of graphs for which several fundamental algorithmic problems which are known to be NP-hard for general graphs, such as the graph colouring problem, the maximum clique problem, and the maximum independent set problem, can all be solved in polynomial time (see [16]). Also, it has significant connections with the study of linear and integer programming (see, *e.g.*, [20]).

A famous conjecture of Berge, which was proved a few years ago by Chudnovsky, Robertson, Seymour and Thomas [9], states that a graph is perfect if and only if it contains no induced odd cycle of length at least five or the complement of one. The proof in fact establishes a stronger structural theorem for perfect graphs which was conjectured by Conforti, Cornuéjols and Vušković. It says that every perfect graph falls into one of a few basic classes, or admits one of a few kinds of special decompositions. Shortly afterwards, a proof that perfect graphs can be recognized in polynomial time (as a function of the number of vertices of the graph) was discovered by Chudnovsky, Cornuéjols, Liu, Seymour and Vušković [8].

Another well-studied hereditary family of graphs is that of comparability graphs. A comparability graph is a graph that connects pairs of elements that are comparable to each other in a partial order. Gallai [13] classified these graphs by forbidden induced subgraphs, and Dilworth's theorem [11] is equivalent to the statement that the complement of comparability graphs are perfect. Further, comparability graphs can be recognized in polynomial time (see McConnell and Spinrad [19]). Every cograph is a comparability graph, and every comparability graph is a perfect graph. It is natural to suspect that the structure theorem could hint at a polynomial in ϵ^{-1} tester for perfectness similar to testing cographs. However, we show that testing perfectness essentially requires as much query complexity (or time) as testing triangle-freeness, which is hard.

Theorem 1.2. *Testing perfectness is hard.*

Indeed, Theorem 3.1 shows that, from a graph on n vertices which is 14ϵ-far from being triangle-free, but a random sample of d vertices is with probability at least $1/2$ triangle-free, we can construct a graph on $5n$ vertices which is $\epsilon/25$-far from being induced C_5-free, but a random sample of d vertices in it is a comparability graph with probability at least $1/2$. Since every comparability graph is perfect, every perfect graph is induced C_5-free, and testing triangle-freeness is hard, this implies the above theorem that testing perfectness is hard, and further that testing for comparability graphs is hard.

Theorem 1.3. *Testing for comparability graphs is hard.*

In the next section, we show that induced P_3-freeness is easily testable. In Section 3, we show that testing perfectness is at least as hard as testing triangle-freeness, which is hard. We finish with some concluding remarks. Throughout the paper, we systematically omit floor and ceiling signs whenever they are not crucial for the sake of clarity of presentation. We also do not make any serious attempt to optimize absolute constants in our statements and proofs.

2. Induced P_3-freeness is easily testable

A *cut* for a graph $G = (V, E)$ is a partition $V = V_1 \cup V_2$ into non-empty subsets such that there are no edges between V_1 and V_2 or V_1 is complete to V_2. The following definition is a natural relaxation of a cut. For $\beta > 0$, define a β-*cut* for a graph $G = (V, E)$ as a partition $V = V_1 \cup V_2$ into non-empty subsets such that $e(V_1, V_2) \leqslant \beta|V_1||V_2|$ or $e(V_1, V_2) \geqslant (1-\beta)|V_1||V_2|$. For a graph G and vertex subset S, let $G[S]$ denote the induced subgraph of G with vertex set S. Let $c(\beta, n)$ be the least δ for which there is a graph $G = (V, E)$ on n vertices which has no β-cut and has δn^4 induced copies of P_3.

Theorem 2.1. *We have* $c(\beta, n) \geqslant (\beta/100)^{12}$.

Proof. Suppose for contradiction that there is a graph G on n vertices which does not have a β-cut and has less than δn^4 induced copies of P_3, where $\delta = (\beta/100)^{12}$. Since G has no β-cut, then G contains an induced P_3. Hence, $1 \leqslant \delta n^4$ and $n \geqslant \delta^{-1/4} \geqslant (100/\beta)^3$.

Since G has at most δn^4 induced copies of P_3, a random sample of $r = (8\delta)^{-1/4} \geqslant 10^5\beta^{-3}$ vertices has in expectation at most $\delta r^4 = 1/8$ induced copies of P_3. Hence, with probability at least $7/8$, a random sample of r vertices contains no induced P_3.

Randomly sample a set $R = S \cup T$ of $r = s + t$ vertices from V, where $s = t = r/2$. Let E_0 be the event that $G[R]$ is induced P_3-free, so the probability of event E_0 is at least $7/8$.

Since G does not have a β-cut, each vertex has more than $\beta(n-1)$ neighbours and less than $(1-\beta)(n-1)$ neighbours. Let $\alpha = \beta/2$. Hoeffding (see Section 6 of [18]) proved that the hypergeometric distribution is at least as concentrated as the corresponding binomial distribution. Thus, by the Azuma–Hoeffding inequality (see, *e.g.*, [6]), and the fact that each vertex $v \in S$ has more than $\beta(n-1)$ neighbours, the probability that a particular $v \in S$ has less than $\alpha(s-1)$ neighbours in S is at most

$$e^{-((\beta-\alpha)(s-1))^2/(2(s-1))} = e^{-(\beta-\alpha)^2(s-1)/2} \leqslant e^{-\beta^2 s/16} \leqslant \frac{1}{16s}.$$

Similarly, the probability that v has more than $(1-\alpha)(s-1)$ neighbours in S is at most $1/(16s)$. Let E_1 be the event that every vertex in S has at least $\alpha(s-1)$ and at most $(1-\alpha)(s-1)$ neighbours in S, that is, the induced subgraph $G[S]$ has minimum degree at least $\alpha(s-1)$ and maximum degree at most $(1-\alpha)(s-1)$. By the union bound, the probability of event E_1 is at least $1 - 2s \cdot 1/(16s) = 7/8$.

Let U be the set of vertices $v \in V \setminus S$ which are complete or empty to S. As the degree of each vertex of G is at least $\beta(n-1)$ and at most $(1-\beta)(n-1)$, the probability that,

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

for a given vertex v, a random subset of s vertices of $V \setminus \{v\}$ are all neighbours of v or all non-neighbours of v is at most $2(1-\beta)^s$. Hence, a given vertex has probability at most $2(1-\beta)^s$ of being in U. By linearity of expectation, the expected size of U is at most $2(1-\beta)^s n$. Let E_2 be the event that

$$|U| \leqslant 16(1-\beta)^s n \leqslant 16e^{-\beta s} n \leqslant \frac{\beta}{8} n.$$

By Markov's inequality, the probability of E_2 is at least $1 - 1/8 = 7/8$.

Let E_3 be the event that T contains no vertex from U. By linearity of expectation,

$$\mathbb{E}[|U \cap T|] = \mathbb{E}[|U|]t/n \leqslant 2(1-\beta)^s t \leqslant 2e^{-\beta s} t \leqslant \frac{1}{8}.$$

Therefore, event E_3 occurs with probability at least $7/8$.

The probability that events E_0 and E_1 both occur is at least $7/8 - 1/8 = 3/4$. If both of these events occur, then $G[S]$ has at least one and at most $2^{\alpha^{-1}}$ cuts. Indeed, as $S \subset R$ and $G[R]$ is induced P_3-free, then $G[S]$ is also induced P_3-free. It follows that $G[S]$ has a cut $S = S_1 \cup S_2$, and suppose S_1 is complete to S_2 (the case S_1 is empty to S_2 can be treated similarly). Further, there is a unique partition $S = S^1 \cup \cdots \cup S^a$ such that each S^i is non-empty and complete to each S^j with $j \neq i$, and none of the induced subgraphs $G[S^i]$ has a complete cut. As event E_1 occurs, each vertex has at most $(1-\alpha)(s-1)$ neighbours in S. As every vertex in each S^i is also adjacent to all vertices in $S \setminus S^i$, the smallest part S^i in the partition has size at least αs, and hence the number a of parts of this partition is at most α^{-1}. The cuts of $G[S]$ are precisely the pairs $\bigcup_{i \in A} S^i$ and $\bigcup_{i \in [a] \setminus A} S^i$, where A is a non-empty proper subset of $[a]$, and so there are at most $2^{\alpha^{-1}}$ such cuts.

For each cut $S = S_1 \cup S_2$, consider the partition $V \setminus S = U \cup V_0 \cup V_1 \cup V_2$ of vertices, where $v \in V \setminus S$ satisfies $v \in V_0$ if $v \notin U$ and it is not complete to S_1 and not complete to S_2, $v \in V_1$ if it is complete to S_2 but not complete to S_1, and $v \in V_2$ if it is complete to S_1 but not to S_2.

Note that if T contains a vertex from V_0, then the cut $S = S_1 \cup S_2$ of $G[S]$ does not extend to a cut of $G[R]$. If events E_i for $i = 0, 1, 2, 3$ occur, which happens with probability at least $1/2$, then $G[R]$ is induced P_3-free, so it has a cut, and no vertex in T is complete or empty to S. In this case one of the cuts of $G[S]$ extends to a cut of $G[R]$, and hence, for at least one cut of $G[S]$, no vertex of T is in the corresponding V_0.

We now condition on the occurrence of events E_i for $i = 0, 1, 2, 3$. Note that since the probability that this happens is at least $1/2$, for any other event E, the conditional probability that E occurs given that E_i occur for $i = 0, 1, 2, 3$ is at most twice the probability of E without any conditioning.

To complete the proof we claim that with positive probability E_0, E_1, E_2, E_3 occur and yet the induced subgraph on $S \cup T$ contains an induced P_3, contradicting E_0. To do so we apply the union bound over all cuts in $G[S]$ to show that with positive probability, for each such cut, either T contains a vertex of V_0 (and hence the cut cannot be extended to one in $G[R]$) or T contains a vertex v_1 in V_1 and a vertex v_2 in V_2, which are non-adjacent, providing an induced P_3 in $G[S \cup T]$ on the vertices v_1, v_2 together with a vertex $s_1 \in S_1$ not adjacent to v_1 and a vertex $s_2 \in S_2$ not adjacent to v_2.

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

We proceed with the proof of this claim. Conditioning on E_i for $i = 0, 1, 2, 3$, fix a cut (S_1, S_2) in $G[S]$ and let V_0, V_1, V_2 be as above. Consider two possible cases.

Case 1: $|V_0| \geqslant \frac{2}{\alpha t} n$.

In this case, the probability that T contains no vertex of V_0 (without the conditioning on E_3) is at most

$$\left(1 - \frac{2}{\alpha t}\right)^t \leqslant e^{-2/\alpha} < 2^{-\alpha^{-1}-1},$$

showing that even after our conditioning the probability of this event is smaller than $2^{-\alpha^{-1}}$.

Case 2: $|V_0| < \frac{2}{\alpha t} n \leqslant \frac{\beta}{8} n$.

Let $x = |U| + |V_0|$, $y = |S_1| + |V_1|$, and $z = |S_2| + |V_2|$, so $x + y + z = n$. Assume without loss of generality that $y \leqslant z$. Since the partition $V = (S_1 \cup V_1) \cup (S_2 \cup V_2 \cup U \cup V_0)$ is not a β-cut, there are at least $\beta y(z + x)$ missing edges between these two sets. Since, in addition, S_1 is complete to S_2, S_1 is complete to V_2, and V_1 is complete to S_2, then these missing edges go between V_1 and V_2 and between $S_1 \cup V_1$ and $U \cup V_0$. Thus

$$\frac{\beta}{2} yn \leqslant \beta y(z + x) \leqslant |V_1||V_2| - e(V_1, V_2) + yx.$$

If events E_i for $i = 0, 1, 2, 3$ occur, then $x \leqslant (\beta/4)n$, and hence there are at least $(\beta/4)yn$ missing edges between V_1 and V_2. In this case, every vertex of S_1 is complete to $S_2 \cup V_2$, and hence

$$(1 - \beta)(n - 1) \geqslant z = n - x - y \geqslant n - \frac{\beta}{4} n - y$$

and

$$y \geqslant \frac{3\beta}{4} n - 1 \geqslant \frac{\beta}{2} n.$$

Thus, the number of missing pairs between V_1 and V_2 in the case that events E_i for $i = 0, 1, 2, 3$ occur is at least

$$\frac{\beta}{4} yn \geqslant \frac{\beta^2}{8} n^2.$$

Let E_4 be the event that T contains the two vertices of at least one of the non-edges between V_1 and V_2. Given that there are at least $(\beta^2/8)n^2$ edges missing between V_1 and V_2, the probability that event E_4 occurs is at least the probability that at least one of $t/2$ random pairs of vertices of G contains one of the non-edges between V_1 and V_2. The probability that this does *not* happen (without the conditioning on E_3) is at most

$$\left(1 - \frac{\beta^2 n^2/8}{\binom{n}{2}}\right)^{t/2} \leqslant e^{-\beta^2 t/8} = e^{-\beta^2 10^5/(8 \cdot 2\beta^3)} = e^{-10^5/(32\alpha)} < 2^{-\alpha^{-1}-1},$$

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

and hence even after our conditioning the probability of this event is smaller than $2^{-\alpha^{-1}}$.

By the union bound it now follows that with positive probability E_i for $i = 0, 1, 2, 3$ occur, and yet $G[S \cup T]$ contains an induced P_3. This is a contradiction, completing the proof. □

Let $f(\epsilon, n)$ be the least δ for which there is a graph $G = (V, E)$ on n vertices which is ϵ-far from being induced P_3-free and has δn^4 induced copies of P_3.

Theorem 2.2. *There is $n_0 \geqslant \epsilon n$ such that $f(\epsilon, n) \geqslant c(\epsilon, n_0)\epsilon^4 \geqslant (\epsilon/100)^{16}$.*

Proof. Let $G = (V, E)$ be a graph on n vertices which is ϵ-far from being induced P_3-free. Partition V into two parts along an ϵ-cut, and continue refining parts along ϵ-cuts of the subgraphs induced by the parts until no part has an ϵ-cut, and let $V = V_1 \cup \cdots \cup V_k$ be the resulting partition. We modify edges along these ϵ-cuts to turn them into cuts, letting G' be the resulting graph. The total fraction of pairs of vertices changed in making G' from G is at most ϵ, so at least $\epsilon n^2 - \epsilon\binom{n}{2} \geqslant \epsilon n^2/2$ edges must be changed from the resulting graph G' to make it induced P_3-free. We can modify edges in each V_i to make it induced P_3-free, and the resulting graph on V is induced P_3-free, by the known characterization of cographs. If $|V_i| \leqslant \epsilon n$ for $1 \leqslant i \leqslant k$, then the number of edge modifications made to G' to obtain an induced P_3-free graph is at most

$$\sum_{i=1}^{k} \binom{|V_i|}{2} \leqslant \frac{n}{2} \max_{1 \leqslant i \leqslant k} (|V_i| - 1) < \frac{\epsilon n^2}{2},$$

a contradiction. Thus, one of the parts V_i, call it V_0, has $n_0 > \epsilon n$ vertices, and $G[V_0]$ has no β-cut. Therefore, the induced subgraph $G[V_0]$, and hence G, has at least

$$c(\epsilon, n_0)n_0^4 \geqslant c(\epsilon, n_0)\epsilon^4 n^4 \geqslant (\epsilon/100)^{12}\epsilon^4 n^4 \geqslant (\epsilon/100)^{16} n^4$$

induced copies of P_3, completing the proof. □

Consider the following one-sided ϵ-tester for induced P_3-freeness. Let $\delta = (\epsilon/100)^{16}$. The algorithm samples $t = 2\delta^{-1}$ quadruples of vertices uniformly at random, and accepts if none of them forms an induced P_3, and otherwise rejects. Any induced P_3-free graph is clearly accepted. If a graph is ϵ-far from being induced P_3-free, then it contains at least δn^4 induced P_3 by Theorem 2.2, and the probability that none of the sampled quadruples forms an induced P_3 is at most $(1 - \delta)^t < 1/3$. Note that the query complexity for this algorithm depends linearly on δ^{-1}, and hence polynomially on ϵ^{-1}, completing the proof of Theorem 1.1. □

3. Testing perfectness

We first observe a couple of equivalent versions of the triangle removal lemma. The *triangle edge cover number* $\nu(G)$ of a graph G is the minimum number of edges of G that cover all triangles in G, *i.e.*, it is the minimum number of edges of G whose deletion makes

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

G triangle-free. The triangle removal lemma thus says that for each $\epsilon > 0$ there is $\delta > 0$ such that every graph on n vertices with at most δn^3 triangles satisfies $\nu(G) \leqslant \epsilon n^2$.

The *triangle packing number* $\tau(G)$ of a graph G is the maximum number of edge-disjoint triangles in G. The following simple bounds hold for all graphs:

$$\tau(G) \leqslant \nu(G) \leqslant 3\tau(G).$$

Indeed, at least one edge from each of the edge-disjoint triangles is needed in any edge cover of the triangles in G, and deleting the $3\tau(G)$ edges from a maximum collection of edge-disjoint triangles leaves a triangle-free graph. We remark that a well-known conjecture of Tuza states that the upper bound can be improved to $\nu(G) \leqslant 2\tau(G)$. Haxell [17] improved the upper bound factor to $3 - 3/23$.

Thus, up to a constant factor change in ϵ, the triangle removal lemma is the same as saying that a graph G on n vertices with at least ϵn^2 edge-disjoint triangles contains at least δn^3 triangles. We can further suppose, up to a constant factor change in ϵ, that G is tripartite. Indeed, every graph has a tripartite subgraph which contains at least $2/9$ of the triangles in a maximum collection of edge-disjoint triangles. This can be seen by considering a uniform random tripartition. Each triangle has probability $2/9$ of having one vertex in each part, so the expected number of the edge-disjoint triangles in the tripartition is $2/9$ of the total, and there is a tripartition for which the number of edge-disjoint triangles is at least the expected number. We may thus assume G is tripartite.

Theorem 3.1. *Let T be a graph on n vertices which is 14ϵ-far from being triangle-free such that a random sample of d vertices of T is triangle-free, with probability at least $1/2$. Then there is a graph G on $5n$ vertices which is $\epsilon/25$-far from being induced C_5-free, such that a random sample of d vertices of G is a comparability graph, with probability at least $1/2$.*

Proof. By the remarks above, T contains a tripartite subgraph F which contains at least

$$\frac{1}{3} \cdot \frac{2}{9} \cdot 14\epsilon n^2 > \epsilon n^2 = (\epsilon/25)(5n)^2$$

edge-disjoint triangles. Let V_2, V_3, V_5 denote the three parts of F.

Let $G = (V, E)$ be the graph on $5n$ vertices with partition $V = V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5$, where V_1 and V_4 are of size $2n$ each, and V_2, V_3, V_5 are the parts of F. We next specify the edges between the various parts of G. Each part V_i, $1 \leqslant i \leqslant 5$, is an independent set. There are no edges between V_1 and V_2, between V_1 and V_3, between V_3 and V_4, and between V_4 and V_5. There is a complete bipartite graph between V_1 and V_4, between V_1 and V_5, and between V_2 and V_4. The edges of G are precisely the edges of F between V_2 and V_3, and between V_3 and V_5. Finally, between V_2 and V_5, the edges of G are precisely the non-edges of F.

Arbitrarily order $T_1, \ldots, T_t$, a maximum collection of $t = \tau(F) \geqslant \epsilon n^2$ edge-disjoint triangles in F. As F is a tripartite graph on n vertices, $t = \tau(F)$ is at most the product of the two smallest parts, which is at most $n^2/9$. For every triangle in F, the same three vertices in G with a vertex in V_1 and a vertex in V_4 form an induced C_5. We next show that this implies that there are t induced copies of C_5 in G, labelled $L_1, \ldots, L_t$, such that each pair intersects in at most one vertex. In fact, we greedily construct $L_1, \ldots, L_t$ so that

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

they further satisfy that the vertex set of each L_i consists of the vertices of T_i together with a vertex in V_1 and a vertex in V_4.

Suppose we have already constructed L_j for $j < i$ satisfying the desired properties. We next show how to construct L_i with the desired properties. Note that in a tripartite graph, the number of edge-disjoint triangles containing a given vertex v is at most the minimum order of the two parts not containing v. It follows that T_i has non-empty intersection with at most n of the t triangles $T_1, \ldots, T_t$. Hence, for $h = 1, 4$, at most n vertices in V_h are in at least one L_j with $j < i$ for which T_j and T_i share a vertex in common. For $h = 1, 4$, delete these vertices from V_h, and denote the resulting subset of V_h as V_h', so $|V_h'| \geqslant |V_h| - n = n$. As $i - 1 < t < n^2 \leqslant |V_1'||V_4'|$, there is a pair $(v_1, v_4) \in V_1' \times V_4'$ that is not in any L_j with $j < i$. We pick L_i to be the induced C_5 in G with vertices v_1, v_4 and the vertices of T_i. It is clear from this construction that L_i intersects each L_j with $j < i$ in at most one vertex. We can therefore greedily construct the desired t induced copies of C_5, and conclude that G is $\epsilon/25$-far from being induced C_5-free.

On the other hand, the only triples $a < b < c$ of vertices in a linear ordering which puts the vertices in V_i before V_j if $i < j$ with a adjacent to b, b adjacent to c, and a not adjacent to c are with $a \in V_2$, $b \in V_3$, and $c \in V_5$ the vertices of a triangle in F. Hence, an induced subgraph G' of G is a comparability graph if it contains no three vertices which make a triangle in F. Indeed, in this case we can define the corresponding partial order on the vertex set of G' given by $a \prec b$ if a and b are adjacent and $a \in V_i$ and $b \in V_j$ with $i < j$. Thus, by sampling d vertices uniformly at random from G, we sample at most d vertices uniformly at random from F. These at most d vertices are triangle-free in F with probability at least $1/2$, and hence the d random vertices in G form a comparability graph with probability at least $1/2$. This completes the proof. □

As discussed toward the end of the Introduction, Theorem 3.1 implies Theorem 1.2 that testing perfectness is hard, and Theorem 1.3 that testing for comparability graphs is hard.

A partially ordered set (poset) is a directed graph on a vertex set P which

- has no loops, *i.e.*, no pair (x, x) is an edge,
- has no antiparallel edges, *i.e.*, if (x, y) is an edge, then (y, x) is not an edge,
- is transitive, *i.e.*, if (x, y) is an edge and (y, z) is an edge, then (x, z) is also an edge.

The fact that testing for posets is hard (at least as hard as testing for triangle-freeness) follows from Theorem 3.1 by adding directions. However, we next sketch a simpler proof. Let T be a tripartite graph on n vertices with parts V_1, V_2, V_3 which is ϵ-far from being triangle-free. Consider the directed graph G on the same vertex set as T with $(v_1, v_2) \in V_1 \times V_2$ an edge of G if it is an edge of T, $(v_2, v_3) \in V_2 \times V_3$ an edge of G if it is an edge of T, $(v_1, v_3) \in V_1 \times V_3$ an edge of G if it is not an edge of T, and there are no other edges. At least one pair in every triangle of T must be modified to turn G into a poset, so G is ϵ-far from being a poset. Also, any subset of vertices which is triangle-free in T induces a poset in G. This implies that testing for posets is at least as hard as testing for triangle-freeness.

https://doi.org/10.1017/S0963548314000765 Published online by Cambridge University Press

4. Concluding remarks

We believe that comparing the number of queries needed to test various properties, as done in this paper comparing testing perfectness and triangle-freeness, could be an interesting direction for further research. This is the analogue in property testing to the powerful technique of hardness reductions in complexity theory. One general class of hard graph properties for testing for which to compare with is (not necessarily induced) H-freeness for H a fixed odd cycle.

We have shown that testing perfectness is hard. This is equivalent to showing that there is a graph which is ϵ-far from being perfect such that a random set of vertices of size polynomial in ϵ^{-1} is perfect with probability at least $1/2$. This still leaves the possibility of having a small witness if the graph is far from being perfect. That is, does every graph which is ϵ-far from being perfect contain an induced odd cycle or its complement of size at least 5 and at most a polynomial in ϵ^{-1}?

We have shown that testing induced P_3-freeness is easy, which is a step towards completing the classification of graphs H for which induced H-free testing is easy. It remains to determine whether or not induced C_4-freeness is easy.

Finally, it would be very interesting to characterize all easily testable graph properties. As all these properties have to be strongly testable, it follows from the main result of [5] that if we restrict ourselves only to natural properties, in the sense of [5], then these properties have to be essentially hereditary. Among the hereditary properties, properties that are known to be easily testable include the property of being k-colourable for any fixed k, as shown in [14], as well as a natural extension of it, as proved in [15]. As mentioned in the Introduction, additional easily testable (hereditary) properties are H-freeness for any bipartite H, and induced H-freeness for any path H on at most 4 vertices or its complement (where the case of 4 vertices is proved in Section 2).

Hereditary properties which are not easily testable are H-freeness for non-bipartite H, induced H-freeness for all graphs besides the paths on at most 4 vertices and their complements, as well as possibly the cycle of length 4 and its complement, perfectness and comparability. Our techniques here can be applied to provide several additional examples of easily testable and of non-easily testable hereditary properties, but most of these are somewhat artificial and not familiar graph properties. Does the above list of known results suggest a (conjectured) characterization of all easily testable hereditary graph properties? At the moment we are unable to formulate such a conjecture but hope that the results and ideas in the present paper may contribute to the study of this problem.

References

[1] Alon, N. (2002) Testing subgraphs in large graphs. *Random Struct. Alg.* **21** 359–370.
[2] Alon, N., Duke, R. A., Lefmann, H., Rödl, V. and Yuster, R. (1994) The algorithmic aspects of the Regularity Lemma. *J. Algorithms* **16** 80–109.
[3] Alon, N., Fischer, E., Krivelevich, M. and Szegedy, M. (2000) Efficient testing of large graphs. *Combinatorica* **20** 451–476.
[4] Alon, N. and Shapira, A. (2006) A characterization of easily testable induced subgraphs. *Combin. Probab. Comput.* **15** 791–805.

[5] Alon, N. and Shapira, A. (2008) A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.* **37** 1703–1727.
[6] Alon, N. and Spencer, J. H. (2008) *The Probabilistic Method*, third edition, Wiley.
[7] Behrend, F. A. (1946) On sets of integers which contain no three terms in arithmetic progression. *Proc. Nat. Acad. Sci.* **32** 331–332.
[8] Chudnovsky, M., Cornuéjols, G., Liu, X., Seymour, P. and Vušković, K. (2005) Recognizing Berge graphs. *Combinatorica* **25** 143–187.
[9] Chudnovsky, M., Robertson, N., Seymour, P. and Thomas, R. (2006) The strong perfect graph theorem. *Ann. of Math.* **164** 51–229.
[10] Conlon, D. and Fox, J. (2012) Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.* **22** 1191–1256.
[11] Dilworth, R. P. (1950) A decomposition theorem for partially ordered sets. *Ann. of Math.* **51** 161–166.
[12] Fox, J. (2011) A new proof of the graph removal lemma. *Ann. of Math.* **174** 561–579.
[13] Gallai, T. (1967) Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar.* **18** 25–66.
[14] Goldreich, O., Goldwasser, S. and Ron, D. (1998) Property testing and its applications to learning and approximation. *J. Assoc. Comput. Mach.* **45** 653–750.
[15] Goldreich, O. and Trevisan, L. (2003) Three theorems regarding testing graph properties. *Random Struct. Alg.* **23** 23–57.
[16] Grötschel, M., Lovász, L. and Schrijver, A. (1988) *Geometric Algorithms and Combinatorial Optimization*, Springer.
[17] Haxell, P. E. (1999) Packing and covering triangles in graphs. *Discrete Math.* **195** 251–254.
[18] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
[19] McConnell, R. M. and Spinrad, J. P. (1997) Linear-time transitive orientation. In *Proc. Eighth Annual ACM–SIAM Symposium on Discrete Algorithms*, ACM, pp. 19–25.
[20] Ramírez-Alfonsín, J. L. and Reed, B. A., eds (2001) *Perfect Graphs*, Wiley.
[21] Roth, K. F. (1953) On certain sets of integers. *J. London Math. Soc.* **28** 104–109.
[22] Rubinfield, R. and Sudan, M. (1996) Robust characterization of polynomials with applications to program testing. *SIAM J. Comput.* **25** 252–271.
[23] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In *Combinatorics II: Keszthely 1976*, Vol. 18 of *Colloq. Math. Soc. J. Bolyai*, North-Holland, pp. 939–945.
[24] Seinsche, D. (1974) On a property of the class of n-colorable graphs. *J. Combin. Theory Ser. B* **16** 191–193.