Testing Linear Inequalities of Subgraph Statistics

Lior Gishboliner

School of Mathematical Sciences, Tel Aviv University, Tel Aviv, 69978, Israel
liorgis1@mail.tau.ac.il

Asaf Shapira

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
asafico@tau.ac.il

Henrique Stagni

Departamento de Ciencia da Computacao, Instituto de Matematica e Estatistica, Universidade de Sao Paulo, Brazil
stagni@gmail.com

Abstract

Property testers are fast randomized algorithms whose task is to distinguish between inputs satisfying some predetermined property $\mathcal{P}$ and those that are far from satisfying it. Since these algorithms operate by inspecting a small randomly selected portion of the input, the most natural property one would like to be able to test is whether the input does not contain certain forbidden small substructures. In the setting of graphs, such a result was obtained by Alon et al., who proved that for any finite family of graphs $\mathcal{F}$, the property of being induced $\mathcal{F}$-free (i.e. not containing an induced copy of any $F \in \mathcal{F}$) is testable.

It is natural to ask if one can go one step further and prove that more elaborate properties involving induced subgraphs are also testable. One such generalization of the result of Alon et al. was formulated by Goldreich and Shinkar who conjectured that for any finite family of graphs $\mathcal{F}$, and any linear inequality involving the densities of the graphs $F \in \mathcal{F}$ in the input graph, the property of satisfying this inequality can be tested in a certain restricted model of graph property testing. Our main result in this paper disproves this conjecture in the following strong form: some properties of this type are not testable even in the classical (i.e. unrestricted) model of graph property testing.

The proof deviates significantly from prior non-testability results in this area. The main idea is to use a linear inequality relating induced subgraph densities in order to encode the property of being a pseudo-random graph.

2012 ACM Subject Classification Mathematics of computing → Discrete mathematics

Keywords and phrases graph property testing, subgraph statistics

Digital Object Identifier 10.4230/LIPIcs.ITCS.2020.43

Funding *Asaf Shapira*: Supported in part by ISF Grant 1028/16 and ERC Starting Grant 633509.

1 Introduction

Property testers are fast randomized algorithms that distinguish between objects satisfying a certain property and objects that are "far" from the property. The systematic study of such problems originates in the seminal papers of Rubinfeld and Sudan [10] and Goldreich, Goldwasser and Ron [4], and has since become a very active area of research. We refer the reader to the book of Goldreich [3] for more background and references on the subject.

In this paper we study property testing of graph properties in the *dense graph model*. In this model, a graph is given as an $n \times n$ adjacency matrix. An n-vertex graph G is said to be ε*-far* from a graph property Π, if one has to change at least εn^2 entries in the adjacency matrix of G in order to turn it into a graph satisfying Π. A *tester* for Π is a (randomized) algorithm that, given a proximity parameter $\varepsilon \in (0, 1)$ and a graph G, accepts if G satisfies Π and rejects if G is ε-far from Π, with success probability at least $\frac{2}{3}$ in both cases. The

11th Innovations in Theoretical Computer Science Conference (ITCS 2020).
Editor: Thomas Vidick; Article No. 43; pp. 43:1–43:9
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

tester is given oracle access to the adjacency matrix of the input, to which it makes queries. We say that a tester has query complexity $q(\varepsilon, n)$ if it makes at most $q(\varepsilon, n)$ queries when invoked with proximity parameter ε on inputs with n vertices. A property Π is called *testable* if it has a tester whose query complexity is bounded by a function of ε alone, that is, it is independent of the size of the input. A tester is *canonical* if it works by sampling a random set of vertices of some size $s(\varepsilon, n)$, querying all pairs among these vertices, and making its decision based on (the isomorphism class) of the sample. The integer $s(\varepsilon, n)$ is called the *sample complexity* of the tester. Goldreich and Trevisan [7, Theorem 2] showed that every tester can be transformed to a canonical one, with the cost of possibly squaring the query complexity. A tester is *size-oblivious* if it does not know n; that is, if its function depends only on the proximity parameter ε (and not on the size of the input). The transformation of Goldreich and Trevisan [7] which turns arbitrary testers into canonical testers, preserves the property of being size-oblivious.

In this paper we study a special kind of testers, called *proximity oblivious testers*, which are defined as follows.

▶ **Definition 1.** *A* proximity oblivious tester *(POT) for a graph property* Π *is an algorithm which makes a* constant *(i.e. independent of* n and ε*) number of queries to the input and satisfies the following. There is a constant* $c \in (0, 1]$ *and a function* $f : (0, 1] \to (0, 1]$ *such that:*

1. *If the input graph satisfies* Π *then the tester accepts with probability at least* c*.*
2. *If the input graph is* ε*-far from* Π *then the tester accepts with probability at most* $c - f(\varepsilon)$*.*

Observe that a POT for Π can be used to obtain a standard tester for Π, by invoking the POT $T = \Theta(1/f(\varepsilon)^2)$ times and accepting if and only if the POT accepted in at least $(c - \frac{f(\varepsilon)}{2})T$ of the tests.

POTs were introduced by Goldreich and Ron [5], who studied *one-sided-error POTs*, namely POTs that accept every input which satisfies the property with probability 1 (this corresponds to having $c = 1$ in Definition 1). Later, Goldreich and Shinkar [6] studied general (two-sided-error) POTs in several settings, including those of general boolean functions, dense graphs and bounded degree graphs. For the dense graph model, they designed a POT for the property of being αn-regular (for a given $\alpha \in (0, 1)$), as well as for several related properties. Moreover, they considered properties of the following form: given graphs H, G, the *density* of H in G, denoted by $p(H, G)$, is the fraction of induced subgraphs of G of order $v(H)$ which are isomorphic to H. Given an integer $h \geq 2$, a rational number b and rational numbers $w_H \geq 0$, where H runs over all h-vertex graphs, the property $\Pi_{h,w,b}$ is defined as the property of all graphs G satisfying

$$\sum_H w_H \cdot p(H, G) \leq b.$$

Throughout this paper, a tuple (h, w, b) will always consist of an integer $h \geq 2$, a rational number b, and a function $w\colon \{H : v(H) = h\} \to \mathbb{Q}^+$ from the set of all h-vertex graphs to the positive rationals. The value assigned by w to a graph H is denoted by w_H.

Since property testing algorithms, and POTs in particular, work by inspecting the subgraph induced by a small sample of vertices, it is natural to ask if the property of not containing an induced copy of a fixed graph H is a testable property. Such a result was obtained by Alon, Fischer, Krivelevich and Szegedy [1] who proved that in fact for every finite family of graphs $\mathcal{F}$, the property of being induced $\mathcal{F}$-free (i.e. not containing an induced copy of F for every $F \in \mathcal{F}$) is testable. It is easy to see that the family of properties $\Pi_{h,w,b}$ forms a strict generalization of the family of properties of being induced $\mathcal{F}$-free, since the former can encode the latter. (Indeed, if all graphs in $\mathcal{F}$ have the same size h then simply set $b = 0$, $w_H = 1$ for each $H \in \mathcal{F}$, and $w_H = 0$ for each h-vertex graph H which is not in $\mathcal{F}$. If graphs in $\mathcal{F}$ have varying sizes, then take advantage of the fact that for every pair of graphs F, G and $h \geq v(F)$, it holds that $p(F,G) = \sum_H p(F,H) \cdot p(H,G)$, where the sum is over all h-vertex graphs H.)

We now arrive at an important definition.

▶ **Definition 2.** *A tuple (h, w, b) has the* removal property *if there is a function $f : (0,1] \to (0,1]$ such that for every $\varepsilon \in (0,1)$ and for every graph G, if G is ε-far from $\Pi_{h,w,b}$ then*

$$\sum_H w_H \cdot p(H,G) \geq b + f(\varepsilon).$$

As an example, the main result of [1] mentioned above is equivalent to the statement that if $b = 0$ then $\Pi_{h,w,b}$ has the removal property. Goldreich and Shinkar [6] observed that if (h, w, b) has the removal property then $\Pi_{h,w,b}$ admits a size-oblivious POT. Indeed, given an input graph G, the POT works by sampling a random induced subgraph of G of order h, and then rejecting with probability w_H if the sampled subgraph is isomorphic to H, for each H on h vertices. If G satisfies $\Pi_{h,w,b}$ then by the definition of this property, G is rejected with probability $\sum_H w_H \cdot p(H,G) \leq b$. On the other hand, if G is ε-far from $\Pi_{h,w,b}$ then by the removal property, G is rejected with probability $\sum_H w_H \cdot p(H,G) \geq b + f(\varepsilon)$. Thus, Definition 1 is satisfied with $c = 1 - b$.

Our first result in this paper is a converse of the above statement, showing that if $\Pi_{h,w,b}$ has a size-oblivious POT, then (h, w, b) has the removal property.

▶ **Theorem 3.** *For each tuple (h, w, b), if $\Pi_{h,w,b}$ has a size-oblivious POT then (h, w, b) has the removal property.*

As a corollary of the above theorem, we infer that if one "representation" of a property as $\Pi_{h,w,b}$ has the removal property, then all such representations have the removal property. This is stated in the following corollary.

▶ **Corollary 4.** *Let (h, w, b) and (h', w', b') be tuples such that $\Pi_{h,w,b} = \Pi_{h',w',b'}$. Then (h, w, b) has the removal property if and only if (h', w', b') has the removal property.*

▶ Remark 5. Theorem 3 contradicts the statement of Proposition 3.14 in [6], which states that there is a tuple (h, w, b) such that $\Pi_{h,w,b}$ has a POT but (h, w, b) does not have the removal property. We believe that the proof of this proposition is wrong. This will be explained in the full version of this paper.

Goldreich and Shinkar conjectured (see [6, Open Problem 3.11]) that every property of the form $\Pi_{h,w,b}$ has a POT . Our next theorem disproves this conjecture by showing that there are properties $\Pi_{h,w,b}$ that are not testable at all (let alone testable using a POT).

▶ **Theorem 6.** *Let K_4 denote the complete graph on 4 vertices, D_4 the diamond graph (i.e. K_4 minus an edge), P_2 the graph on 4 vertices containing a path of length 2 and an isolated vertex, C_4 the 4-cycle, P_4 the path on 4 vertices and $K_{1,3}$ the star on 4 vertices. Let w_H be the following weight function assigning a non-negative weight to each graph on 4 vertices.*

H :	K_4	$\overline{K_4}$	D_4	$\overline{D_4}$	P_2	$\overline{P_2}$	C_4	$\overline{C_4}$	$K_{1,3}$	$\overline{K_{1,3}}$	P_4
w_H :	1	$\frac{1}{2}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Then, the property

$$\Pi_{h,w,b} = \left\{ G : \sum_{H:|V(H)|=4} w_H \cdot p(H,G) \le \frac{5}{16} \right\} \tag{1}$$

is not *testable.*

Given the above theorem it is natural to ask if every property $\Pi_{h,w,b}$ can at least be tested using $o(n^2)$ edge-queries. We leave this as an open problem.

Paper overview

The rest of the paper is organized as follows. In Section 2 we prove Theorem 3. The proof of Theorem 6 appears in Section 3. The main idea behind its proof is to show that there exists a property $\Pi_{h,w,b}$ (specifically, the one defined in (1)) which encodes the property of being a pseudo-random graph. It is then not hard to show that such a property cannot be tested using a constant number of queries.

2 Proof of Theorem 3

In this section we prove Theorem 3 and Corollary 4. We will need the following auxiliary lemma.

▶ **Lemma 7.** *Let Π be a graph property, and suppose that Π has a canonical size-oblivious ε-tester $\mathcal{T}$ with sample complexity $s = s(\varepsilon)$ (and success probability $\frac{2}{3}$). Then for every $n \ge s^4$ and for every n-vertex graph G which is ε-far from Π, the following holds. For U chosen uniformly at random from $\binom{V(G)}{s^4}$, we have $\mathbb{P}[G[U] \in \Pi] \le e^{-\Omega(s)}$.*

Proof. Let $\mathcal{A}$ be the family of all s-vertex graphs A such that $\mathcal{T}$ accepts if it sees a subgraph isomorphic to A. For a graph G, we say that a sequence of subsets $S_1, \dots, S_s \in \binom{V(G)}{s}$ is *good* if $G[S_i] \in \mathcal{A}$ for at least half of the values of $1 \le i \le s$; otherwise $S_1, \dots, S_s$ is *bad*. For a sequence of vertices $W = (x_1, \dots, x_{s^2})$, we say that W is good (resp. bad) if $\{x_1, \dots, x_s\}, \{x_{s+1}, \dots, x_{2s}\}, \dots, \{x_{s^2-s+1}, \dots, x_{s^2}\}$ is good (resp. bad). Note that for a random $S \in \binom{V(G)}{s}$, if $G \in \Pi$ then $\mathbb{P}[G[S] \in \mathcal{A}] \ge \frac{2}{3}$, and if G is ε-far from Π then $\mathbb{P}[G[S] \in \mathcal{A}] \le \frac{1}{3}$. Using a standard Chernoff-type bound, one can show that the following holds for $S_1, \dots, S_s \in \binom{V(G)}{s}$ chosen uniformly at random and independently.

1. If $G \in \Pi$ then $S_1, \dots, S_s$ is good with probability at least $1 - e^{-Cs}$.
2. If G is ε-far from Π then $S_1, \dots, S_s$ is bad with probability at least $1 - e^{-Cs}$.

In both items above, C is an absolute constant.

The probability that $S_i \cap S_j \ne \emptyset$ for some $1 \le i < j \le s$ is at most $\binom{s}{2}\frac{s^2}{n} < \frac{1}{2}$, where the inequality follows from the assumption that $n \ge s^4$. So $|S_1 \cup \dots \cup S_s| = s^2$ with probability at least $\frac{1}{2}$. Conditioned on the event that $S_1, \dots, S_s$ are pairwise-disjoint, the set $S := S_1 \cup \dots \cup S_s$ has the distribution of an element of $\binom{V(G)}{s^2}$ chosen uniformly at random. Thus, a random sequence of vertices $W = (x_1, \dots, x_{s^2})$ chosen *without repetitions* satisfies the following.

1. If G satisfies Π then W is good with probability at least $1 - 2e^{-Cs}$.
2. If G is ε-far from Π then W is bad with probability at least $1 - 2e^{-Cs}$.

Now let G be a graph on $n \ge s^4$ vertices which is ε-far from Π. Consider a random pair (U, W), where U is chosen uniformly at random from $\binom{V(G)}{s^4}$, and $W = (x_1, \dots, x_{s^2})$ is a sequence of vertices sampled randomly without repetition from U. Note that W is distributed as a uniform sequence of s^2 vertices of G, sampled without repetition. Thus, $\mathbb{P}[W \text{ is good}] \leq 2e^{-Cs}$. On the other hand, if $G[U] \in \Pi$, then $\mathbb{P}[W \text{ is good} \mid U] \geq 1 - 2e^{-Cs}$. By combining these two facts, we see that

$$\mathbb{P}[G[U] \in \Pi] \leq \frac{2e^{-Cs}}{1 - 2e^{-Cs}} \leq 4e^{-Cs} = e^{-\Omega(s)}. \qquad \blacktriangleleft$$

Proof of Theorem 3. As mentioned in the introduction, a POT for $\Pi_{h,w,b}$ can be used to obtain a standard tester for $\Pi_{h,w,b}$ by invoking the POT an appropriate number of times. Moreover, it is clear that if the POT is size-oblivious, then so is the resulting tester. Next, we apply the transformation of Goldreich and Trevisan [7] to get a canonical tester $\mathcal{T}$ for $\Pi_{h,w,b}$. Since this transformation preserves the property of being size-oblivious, $\mathcal{T}$ is size-oblivious, and hence satisfies the condition of Lemma 7. Denote by $s = s(\varepsilon)$ the sample complexity of $\mathcal{T}$. We may and will assume that s is large enough as a function of the parameters h and b.

Let us denote $z(G) = \sum_H w_H \cdot p(H, G)$. By multiplying the inequality $\sum_H w_H \cdot p(H, G) \leq b$ by an appropriate integer, we can assume without loss of generality that b is an integer, and that w_H is an integer for every H.

Let G be a graph which is ε-far from $\Pi_{h,w,b}$. Our goal is to show that $z(G) \geq b + f(\varepsilon)$, for a function $f : (0, 1] \to (0, 1]$ to be chosen later. By Lemma 7, a randomly chosen $U \in \binom{V(G)}{s^4}$ satisfies $G[U] \in \Pi_{h,w,b}$ with probability at most $e^{-\Omega(s)}$. Observe that if a k-vertex graph K does not satisfy $\Pi_{h,w,b}$, then necessarily

$$z(K) = \sum_H w_H \cdot p(H, K) \geq b + \binom{k}{h}^{-1} > b + k^{-h},$$

as b and all weights w_H are integers. Thus, if $G[U] \notin \Pi_{h,w,b}$ then

$$z(G[U]) > b + |U|^{-h} = b + s^{-4h} .$$

Observe that $z(G)$ is the average of $z(G[U])$ over all $U \in \binom{V(G)}{s^4}$. Thus, using the guarantees of Lemma 7, we obtain

$$z(G) \geq (1 - e^{-\Omega(s)})(b + s^{-4h}) > b + \frac{1}{2}s^{-4h},$$

where the last inequality holds provided that s is large enough as a function of h and b. So we may take the function f in Definition 2 to be $f(\varepsilon) = \frac{1}{2}s(\varepsilon)^{-4h}$. This completes the proof. $\blacktriangleleft$

Proof of Corollary 4. We have established that (h, w, b) satisfies the removal property if and only if $\Pi_{h,w,b}$ has a size-oblivious POT. The "only if" part was explained in the introduction (see also [6]), and the "if" part is the statement of Theorem 3. Since the existence of a tester (specifically, a size-oblivious POT) does not depend on the specific representation of a given property as $\Pi_{h,w,b}$, it is now clear that the corollary holds. $\blacktriangleleft$

3 Proof of Theorem 6

Let $\Pi_{h,w,b}$ be as in the statement of Theorem 6. Denote

$$z(G) := \sum_{H:|V(H)|=4} w_H \cdot p(H, G)$$

for every graph G of order at least 4. Under this notation, $\Pi_{h,w,b} = \{G : z(G) \le b\}$. For a pair of graphs H and G, define

$$t_{\mathrm{inj}}(H,G) = \frac{1}{n^{\underline{h}}}|\{\varphi\colon V(H) \to V(G) \text{ injective s.t. } uv \in E(H) \Rightarrow \varphi(u)\varphi(v) \in E(G)\}|,$$

and

$$t_{\mathrm{ind}}(H,G) = \frac{1}{n^{\underline{h}}}|\{\varphi\colon V(H) \to V(G) \text{ injective s.t. } uv \in E(H) \Leftrightarrow \varphi(u)\varphi(v) \in E(G)\}|,$$

where $n^{\underline{h}} = n \cdot (n-1) \cdot \dots \cdot (n-h+1)$. Note that $t_{\mathrm{ind}}(H,G) = p(H,G) \cdot \mathrm{aut}(H)/h!$, where $\mathrm{aut}(H)$ is the number of automorphisms of H. The following lemma gives a simpler description of $\Pi_{h,w,b}$.

▶ **Lemma 8.** $\Pi_{h,w,b} = \{G : \phi(G) \le 0\}$, *where* $\phi(G) = 2t_{inj}(C_4,G) - t_{inj}(K_2,G) + \frac{3}{8}$.

Proof. First, note that $t_{inj}(K_2,G) = p(K_2,G)$. Next, we use the fact that

$$\begin{aligned} t_{\mathrm{inj}}(C_4,G) &= t_{\mathrm{ind}}(C_4,G) + 2t_{\mathrm{ind}}(D_4,G) + t_{\mathrm{ind}}(K_4,G) \\ &= \frac{\mathrm{aut}(C_4)}{4!} \cdot p(C_4,G) + 2\frac{\mathrm{aut}(D_4)}{4!} \cdot p(D_4,G) + \frac{\mathrm{aut}(K_4)}{4!} \cdot p(K_4,G) \\ &= \frac{1}{3}p(C_4,G) + \frac{1}{3}p(D_4,G) + p(K_4,G). \end{aligned}$$

Hence,

$$\begin{aligned} \phi(G) &= \frac{2}{3}p(C_4,G) + \frac{2}{3}p(D_4,G) + 2p(K_4,G) - p(K_2,G) + \frac{3}{8} \\ &= \frac{2}{3}p(C_4,G) + \frac{2}{3}p(D_4,G) + 2p(K_4,G) + p(\overline{K_2},G) - \frac{5}{8} \\ &= \frac{2}{3}p(C_4,G) + \frac{2}{3}p(D_4,G) + 2p(K_4,G) + \sum_{H:|V(H)|=4} p(\overline{K_2},H)p(H,G) - \frac{5}{8} \\ &= \sum_{H:|V(H)|=4} 2w_H \cdot p(H,G) - \frac{5}{8}. \end{aligned}$$

Therefore, $\phi(G) \le 0$ if and only if $\sum_{H:|V(H)|=4} w_H \cdot p(H,G) \le 5/16$ ◀

An important ingredient in the proof of Theorem 6 is the following lemma, which shows that graphs in $\Pi_{h,w,b}$ are pseudo-random. In what follows, we write $x = y \pm z$ to mean that $x \in [y-z, y+z]$.

▶ **Lemma 9.** *For every* $\delta \in (0,1)$ *there is* $n_0(\delta)$ *such that every graph* $G \in \Pi_{h,w,b}$ *on* $n \ge n_0(\delta)$ *vertices satisfies the following. For every* $U, V \subseteq V(G)$ *with* $|U|, |V| \ge \delta n$, *it holds that*

$$e(U,V) = \left(\frac{1}{2} \pm \delta\right)|U||V| .$$

Proof. We start by showing that for every $\gamma \in (0,1)$ there is $n_0(\gamma)$ such that if $G \in \Pi_{h,w,b}$ is a graph on $n \ge n_0(\gamma)$ vertices, then

$$t_{\mathrm{inj}}(K_2,G) = \frac{1}{2} \pm \gamma \text{ and } t_{\mathrm{inj}}(C_4,G) = \frac{1}{16} \pm \gamma. \tag{2}$$

It is a well-known fact (see for instance [11]) that every n-vertex graph G satisfies[1]

$$t_{\text{inj}}(C_4, G) \geq t_{\text{inj}}(K_2, G)^4 - O\left(\frac{1}{n}\right) \geq t_{\text{inj}}(K_2, G)^4 - \frac{\gamma^2}{2}, \tag{3}$$

where the last inequality holds if n is large enough. Now, observe that every $G \in \Pi_{h,w,b}$ satisfies

$$2t_{\text{inj}}(K_2, G)^4 - t_{\text{inj}}(K_2, G) + \frac{3}{8} \leq 2t_{\text{inj}}(C_4, G) + \gamma^2 - t_{\text{inj}}(K_2, G) + \frac{3}{8} = \phi(G) + \gamma^2 \leq \gamma^2, \tag{4}$$

where the last inequality follows from Lemma 8. Note that the function $x \mapsto 2x^4 - x + \frac{3}{8}$ is convex, and attains its minimum at $x = 1/2$. Therefore, if we had $t_{\text{inj}}(K_2, G) > \frac{1}{2} + \gamma$, then we would have

$$2t_{\text{inj}}(K_2, G)^4 - t_{\text{inj}}(K_2, G) + \frac{3}{8} > 2\left(\frac{1}{2} + \gamma\right)^4 - \left(\frac{1}{2} + \gamma\right) + \frac{3}{8} = 2\gamma^4 + 4\gamma^3 + 3\gamma^2 > \gamma^2.$$

Similarly, if we had $t_{\text{inj}}(K_2, G) < \frac{1}{2} - \gamma$, then we would have

$$2t_{\text{inj}}(K_2, G)^4 - t_{\text{inj}}(K_2, G) + \frac{3}{8} > 2\left(\frac{1}{2} - \gamma\right)^4 - \left(\frac{1}{2} - \gamma\right) + \frac{3}{8} = 2\gamma^4 - 4\gamma^3 + 3\gamma^2 > \gamma^2.$$

In any case, we see that $|t_{\text{inj}}(K_2, G) - \frac{1}{2}| > \gamma$ would stand in contradiction to (4). Hence, $t_{\text{inj}}(K_2, G) = \frac{1}{2} \pm \gamma$. By applying Lemma 8 again we get $t_{\text{inj}}(C_4, G) \leq \frac{1}{16} + \gamma$. By using the intermediate inequality in (3) and $t_{\text{inj}}(K_2, G) \geq \frac{1}{2} - \gamma$, we get

$$t_{\text{inj}}(C_4, G) \geq \left(\frac{1}{2} - \gamma\right)^4 - O\left(\frac{1}{n}\right) \geq \frac{1}{16} - \gamma,$$

where the last inequality can be easily verified, assuming that n is large enough, by expanding the binomial expression. We have thus established (2).

A well-known result of Chung, Graham and Wilson [2] states that for every $\delta \in (0, 1)$ there is $\gamma = \gamma(\delta)$ such that if a graph G satisfies (2), then for every $U, V \subseteq V(G)$ with $|U|, |V| \geq \delta n$, it holds that $e(U, V) = \left(\frac{1}{2} \pm \delta\right)|U||V|$. The lemma follows by combining this result with the above. ◀

For a family of graphs $\mathcal{F}$ and a graph G, we define

$$p(\mathcal{F}, G) = \sum_{F \in \mathcal{F}} p(F, G) .$$

It is well-known (see e.g. [2]) that a pseudo-random graph has approximately the same distribution of small subgraphs as a random graph with the same density. By combining this with Lemma 9, we obtain the following corollary. Note that the expected value of $p(F, G(n, \frac{1}{2}))$ is $2^{-\binom{s}{2}} \frac{s!}{\text{aut}(F)}$.

▶ **Corollary 10.** *For every $s \geq 2$ and $\delta \in (0, 1)$ there is $n_1 = n_1(s, \delta)$ such that every $G \in \Pi_{h,w,b}$ on $n \geq n_1$ vertices satisfies the following. For every family $\mathcal{F}$ of s-vertex graphs, it holds that*

$$p(\mathcal{F}, G) = \sum_{F \in \mathcal{F}} 2^{-\binom{s}{2}} \frac{s!}{aut(F)} \pm \delta.$$

[1] Usually this inequality is stated in terms of the *homomorphic density*, as $t(C_4, G) \geq t(K_2, G)^4$ (see e.g. [9]). The error-term $O(\frac{1}{n})$ accounts for the difference between the homomorphic density and the injective density.

We are now ready to prove Theorem 6.

Proof of Theorem 6. We start by showing that $\Pi_{h,w,b}$ is non-empty. More specifically, we prove that for every integer $n \geq 4$, there exists an n-vertex graph satisfying $\Pi_{h,w,b}$. Let $G \sim G(n, \frac{1}{2})$. It is easy to see that $\mathbb{E}[t_{\text{inj}}(K_2, G)] = \frac{1}{2}$ and $\mathbb{E}[t_{\text{inj}}(C_4, G)] = \frac{1}{16}$. Hence,

$$\mathbb{E}[\phi(G)] = 2\mathbb{E}[t_{\text{inj}}(C_4, G)] - \mathbb{E}[t_{\text{inj}}(K_2, G)] + \frac{3}{8} = 0.$$

It follows that there is an n-vertex graph with $\phi(G) \leq 0$, and hence $G \in \Pi_{h,w,b}$ by Lemma 8.

Now suppose by contradiction that $\Pi_{h,w,b}$ is testable. In particular, there exists a 0.1-tester for $\Pi_{h,w,b}$. This implies – by [7, Theorem 2] (see also [8]) – that there is a canonical 0.1-tester $\mathcal{T}$ for $\Pi_{h,w,b}$. Denote by s the sample complexity of $\mathcal{T}$. Then for every $n > 0$ there is a family $\mathcal{F} = \mathcal{F}(n)$ of (rejection) graphs of order s satisfying the following for every n-vertex graph G.

1. $p(\mathcal{F}, G) \leq \frac{1}{3}$ if $G \in \Pi_{h,w,b}$;
2. $p(\mathcal{F}, G) \geq \frac{2}{3}$ if G is 0.1-far from $\Pi_{h,w,b}$.

Let $n' = \max\{9s^2, n_1(s, \frac{1}{9})\}$ and $n = \max\{n_0(\frac{1}{n'}), n_1(s, \frac{1}{9})\}$, where n_1 is from Corollary 10 and n_0 is from Lemma 9. Let G' be an arbitrary n'-vertex graph which satisfies $\Pi_{h,w,b}$, and let G be the $\frac{n}{n'}$-blow-up of G'. Denote by $V_1 \sqcup \cdots \sqcup V_{n'} = V(G)$ the clusters of this blow-up.

We claim that G is 0.1-far from $\Pi_{h,w,b}$. Indeed, fix any $G^* \in \Pi_{h,w,b}$ with n vertices. By our choice of n via Lemma 9, and as $|V_i| = \frac{n}{n'}$, we must have $e_{G^*}(V_i, V_j) = (\frac{1}{2} \pm \frac{1}{n'})(n/n')^2$. But since $e_G(V_i, V_j) \in \{0, (n/n')^2\}$, we must change at least $(\frac{1}{2} - \frac{1}{n'})(n/n')^2 \geq 0.4(n/n')^2$ edges between V_i and V_j for every $1 \leq i < j \leq n'$, in order to turn G into G^*. Therefore, the edit distance between G and G^* is at least $\binom{n'}{2} \cdot 0.4(n/n')^2 \geq 0.1n^2$, as required.

Now, let $S \in \binom{V(G)}{s}$ be chosen uniformly at random, and let $\mathcal{B}$ be the event that there exists $1 \leq i \leq n'$ for which $|S \cap V_i| > 1$. Note that we have $\mathbb{P}(\mathcal{B}) \leq \binom{s}{2}/n' < \frac{1}{9}$, by the choice of n'. Observe that conditioned on $\mathcal{B}^c$, the probability that S is isomorphic to an s-vertex graph F is exactly $p(F, G')$. Hence, setting $\mathcal{F} = \mathcal{F}(n)$ and $\rho = \sum_{F \in \mathcal{F}} 2^{-\binom{s}{2}} \frac{s!}{\text{aut}(F)}$, we have

$$\begin{aligned} p(\mathcal{F}, G) &= \mathbb{P}[G[S] \in \mathcal{F}] = \mathbb{P}(G[S] \in \mathcal{F} \mid B^c) + \mathbb{P}(B) \\ &< \mathbb{P}(G[S] \in \mathcal{F} \mid B^c) + \frac{1}{9} \\ &= p(\mathcal{F}, G') + \frac{1}{9} \\ &\leq \sum_{F \in \mathcal{F}} 2^{-\binom{s}{2}} \frac{s!}{\text{aut}(F)} + \frac{1}{9} + \frac{1}{9} \\ &= \rho + \frac{2}{9}, \end{aligned} \tag{5}$$

where in the last inequality we used our choice of n' via Corollary 10. On the other hand, our choice of n via Corollary 10 implies that every n-vertex graph $G^* \in \Pi_{h,w,b}$ satisfies

$$p(\mathcal{F}, G^*) \geq \sum_{F \in \mathcal{F}} 2^{-\binom{s}{2}} \frac{s!}{\text{aut}(F)} - \frac{1}{9} = \rho - \frac{1}{9}.$$

By combining this with (5), we get $p(\mathcal{F}, G^*) > p(\mathcal{F}, G) - \frac{1}{3}$. But this stands in contradiction to $p(\mathcal{F}, G^*) \leq \frac{1}{3}$ (as $G^* \in \Pi_{h,w,b}$) and $p(\mathcal{F}, G) \geq \frac{2}{3}$ (as G is 0.1-far from $\Pi_{h,w,b}$). This completes the proof of the theorem. ◀

References

1. N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.
2. F.R.K Chung, R.L Graham, and R.M. Wilson. Quasi-random graphs. *Combinatorica*, 9:345–362, 1989.
3. O. Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.
4. O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45:653–750, 1998.
5. O. Goldreich and D. Ron. On proximity oblivious testing. *SIAM J. Comput*, 40:534–566, 2011.
6. O. Goldreich and I. Shinkar. Two-sided error proximity oblivious testing. *Random Structures Algorithms*, 48:341–383, 2016.
7. O. Goldreich and L. Trevisan. Three theorems regarding testing graph properties. *Random Structures Algorithms*, 23:23–57, 2003.
8. O. Goldreich and L. Trevisan. Errata to Three theorems regarding testing graph properties. available from `http://www.wisdom.weizmann.ac.il/~oded/p_ttt.html`, 2005.
9. L. Lovász. *Large networks and graph limits*. Providence: American Mathematical Society, 2012.
10. R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25:252–271, 1996.
11. A. Sidorenko. A correlation inequality for bipartite graphs. *Graphs and Combinatorics*, 9:201–204, 1993.