# Efficient testing of large graphs *

Noga Alon † Eldar Fischer ‡ Michael Krivelevich § Mario Szegedy ¶

**Abstract**

Let $P$ be a property of graphs. An $\epsilon$-test for $P$ is a randomized algorithm which, given the ability to make queries whether a desired pair of vertices of an input graph $G$ with $n$ vertices are adjacent or not, distinguishes, with high probability, between the case of $G$ satisfying $P$ and the case that it has to be modified by adding and removing more than $\epsilon n^2$ edges to make it satisfy $P$. The property $P$ is called testable, if for every $\epsilon$ there exists an $\epsilon$-test for $P$ whose total number of queries is independent of the size of the input graph. Goldreich, Goldwasser and Ron [8] showed that certain individual graph properties, like $k$-colorability admit an $\epsilon$-test. In this paper we make a first step towards a complete logical characterization of all testable graph properties, and show that properties describable by a very general type of coloring problem are testable. We use this theorem to prove that first order graph properties not containing a quantifier alternation of type "$\forall\exists$" are always testable, while we show that some properties containing this alternation are not.

Our results are proven using a combinatorial lemma, a special case of which, that may be of independent interest, is the following. A graph $H$ is called $\epsilon$-unavoidable in $G$ if all graphs that differ

*A preliminary version of this paper appeared in the Proceedings of the $40^{th}$ Symposium on Foundation of Computer Science (FOCS'99), IEEE Press 1999, 656–666.

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel, and AT&T Labs–Research, Florham Park, NJ 07932, USA. Email: noga@math.tau.ac.il. Research supported by a USA Israeli BSF grant, by a grant from the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

‡NEC Research Institute, 4 Independence Way, Princeton NJ, 08540, USA; and DIMACS. Research performed while at the Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel, and supported by the Fritz Brann Doctoral Fellowship in Engineering and Exact Sciences. E-mail: fischer@research.nj.nec.com

§Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@math.tau.ac.il. Research was performed while this author was with DIMACS Center, Rutgers University, Piscataway, NJ 08854, USA, and AT&T Labs–Research, Florham Park, NJ 07932, USA. Research supported in part by a DIMACS Postdoctoral Fellowship.

¶School of Mathematics, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, USA. E-mail: szegedy@math.ias.edu. Research was performed while this author was with AT&T Labs–Research, Florham Park, NJ 07932, USA.

Mathematics Subject Classification (2000): 68R10, 05C85, 05C35.

from $G$ in no more than $\epsilon|G|^2$ places contain an induced copy of $H$. A graph $H$ is called $\delta$-abundant in $G$ if $G$ contains at least $\delta|G|^{|H|}$ induced copies of $H$. If $H$ is $\epsilon$-unavoidable in $G$ then it is also $\delta(\epsilon,|H|)$-abundant.

# 1 Introduction

All graphs considered here are finite, undirected, and have neither loops nor parallel edges; let $\mathcal{G}$ denote the family of all such possible graphs with labeled sets of vertices. In what follows, we use the notation of [5] except where stated otherwise; in particular, $|G|$ denotes the number of vertices of a graph $G \in \mathcal{G}$.

Let $P$ be a property of graphs. A graph $G$ with $n$ vertices is called *$\epsilon$-far from satisfying $P$* if no graph $\tilde{G}$ with the same vertex set, which differs from $G$ in no more than $\epsilon n^2$ places (i.e. can be constructed from $G$ by adding and removing no more than $\epsilon n^2$ edges), satisfies $P$. An *$\epsilon$-test* for $P$ is a randomized algorithm which, given the quantity $n$ and the ability to make queries whether a desired pair of vertices of an input graph $G$ with $n$ vertices are adjacent or not, distinguishes with probability at least $\frac{2}{3}$ between the case of $G$ satisfying $P$ and the case of $G$ being $\epsilon$-far from satisfying $P$.

The property $P$ is called *testable*, if for every fixed $\epsilon > 0$ there exists an $\epsilon$-test for $P$ whose total number of queries is bounded only by a function of $\epsilon$, which is independent of the size of the input graph. Of course, the probability $\frac{2}{3}$ appearing above can be replaced by any constant smaller than 1, by performing (a constant number of) iterations of the algorithm and giving as output the majority vote.

As to be expected, all properties discussed are invariant with regard to graph isomorphisms. We assume without loss of generality that every given algorithm queries about all pairs of a randomly chosen set of vertices (otherwise, every time the algorithm queries about a vertex pair we make it query also about all pairs containing a vertex of the new pair and a vertex from previous queries). In particular, we assume that the number of queries a given test makes is a constant independent of the actual input graph $G$ (by making additional queries at the end if it is smaller than the bound), and, denoting it by $C$, conclude that the probability of every given edge of (a randomly permuted) input graph $G$ with $n$ vertices to be queried by the algorithm is exactly $C\binom{n}{2}^{-1}$.

The general notion of property testing was first formulated by Rubinfeld and Sudan [12], who were motivated mainly by its connection to the study of program checking. In [2] the notion of testability in the context of regular languages is investigated. The study of the notion of testability for combinatorial objects, and mainly for labeled graphs, was introduced by Goldreich, Goldwasser and Ron [8], who showed that all graph properties describable by the existence of a partition of a certain type, and among them $k$-colorability, are testable. In [8] the existence of a non-testable $NP$-graph property was shown too. The fact that $k$-colorability is testable is, in fact, implicitly proven already in [11] (see also [1]), using the Regularity Lemma of Szemerédi [13], but in the context of property testing it is first studied in [8].

In the present paper we study the testability of first order graph properties. These are properties that can be formulated by first order expressions about graphs, that is, expressions that contain quantifiers (over vertices), Boolean connectives, equality and adjacency. For example, the graph properties describable by a first order expression with one existential quantifier and no universal quantifiers are exactly those of containing a member of some fixed family of graphs as an induced subgraph (these are trivially testable because such a member can be added to any graph by altering a constant number of its edges). The graph properties describable by a first order expression with one universal quantifier and no existential ones are those of not containing a member of some fixed family of graphs as an induced subgraph; these are among the properties that are proven testable in the following.

Our main result is that all first order graph expressions containing at most one quantifier, as well as all expressions of the form "There exist $x_1, \ldots, x_t$ such that for all $y_1, \ldots, y_s$ $A(x_1, \ldots, x_t, y_1 \ldots, y_s)$", where $A$ is a quantifier-free first order expression are testable. We call these *expressions of type "$\exists\forall$"*. On the other hand, we also show that there are first order expressions of type "$\forall\exists$", namely expressions of the form "For all $x_1, \ldots, x_t$ there exist $y_1, \ldots, y_s$ satisfying $B(x_1, \ldots, x_t, y_1, \ldots, y_s)$" where $B$ is a quantifier-free first order expression about graphs, which are non-testable.

**Theorem 1.1** *All first order properties of type "$\exists\forall$" are testable. On the other hand, there exists a first order property of type "$\forall\exists$" which is not testable.*

Our main technical lemma is a variant of Szemerédi's Regularity Lemma. Szemerédi's Regularity Lemma plays a role in a wide variety of existential and algorithmic results in Extremal Graph Theory; see e.g. [10], [7], [1] and their references. The variant proven here may well have additional applications, besides the ones dealing with testing graph properties.

We note that the expressive power of first order expressions in the context of property testing is far greater than might be expected. For example, the property of being properly $k$-colorable for any fixed $k$, is not a first order property, but it is shown in the next section that it is equivalent to one, using a notion of equivalence which is defined there and proven to be appropriate for the purpose of proving testability results.

The rest of the paper is organized as follows. In Section 2 we introduce some notions about testing, including a notion of equivalence, and observe that any first order property of type "$\exists\forall$" is equivalent to a certain generalized coloring problem. After laying the required foundations in Section 3, Section 4 contains a statement and proof of a variant of Szemerédi's Regularity Lemma which is suited for proving results in the context of induced subgraphs (the context of not necessarily induced subgraphs corresponds to monotone graph properties). Section 6 contains the proof of the main result, showing that the aforementioned generalized coloring problem is testable. As a warmup to Section 6, Section 5 contains a proof of a special case of this result that may be of independent interest. In Section 7 we

describe our non-testable first order property, which is based on the Graph Isomorphism problem. The final Section 8 contains some concluding remarks and open problems. During the course of the proofs no attempt is made to optimize the constants involved.

## 2 Indistinguishability and first order properties

We begin by defining a notion that implies that two given graph properties are equivalent for the purpose of testing.

**Definition 1** *Two graph properties $P$ and $Q$ are called* indistinguishable *if for every $\epsilon > 0$ there exists $N = N(\epsilon)$ satisfying the following. For every graph $G$ with $n > N$ vertices satisfying $P$ there exists a graph $\tilde{G}$ with the same vertex set, differing from $G$ in no more than $\epsilon n^2$ places, which satisfies $Q$; and for every graph $H$ with $n > N$ vertices satisfying $Q$ there exists a graph $\tilde{H}$ with the same vertex set, differing from $H$ in no more than $\epsilon n^2$ places, which satisfies $P$.*

In other words, $P$ and $Q$ are indistinguishable if for every $\epsilon$ there are only a finite number of graphs which satisfy one property but are $\epsilon$-far from satisfying the other property.

**Lemma 2.1** *If $P$ and $Q$ are indistinguishable graph properties, then $P$ is testable if and only if $Q$ is testable.*

**Proof:** Without loss of generality we assume that $P$ is testable and show that in this case $Q$ is testable too. Given $\epsilon$, we construct an $\epsilon$-test for $Q$. According to the assumptions there exists an $\frac{\epsilon}{2}$-test for $P$, with success probability at least $\frac{2}{3}$, which makes $C = C(\frac{\epsilon}{2})$ queries on the input graph. By iterating this algorithm three times and deciding according to the majority vote, we obtain an algorithm with $3C$ queries and success probability at least $\frac{20}{27}$.

Let $N$ be such that if a graph $G$ with $n > N$ vertices satisfies $Q$ then there exists a graph $\tilde{G}$ with the same vertex set satisfying $P$ which differs from $G$ in no more than $\min\{\frac{\epsilon}{2}, \frac{2}{81C}\}\binom{n}{2}$ places. Our test is as follows. If the input graph $G$ has no more than $N$ vertices, we query all its edges and give accurate output according to whether it satisfies $Q$. If the graph $G$ has more than $N$ vertices, we use three iterations of the $\frac{\epsilon}{2}$-test for $P$ (each time using a random permutation of the vertex set of $G$) and output the majority vote. In this case, if $G$ is $\epsilon$-far from satisfying $Q$ then it is $\frac{\epsilon}{2}$-far from satisfying $P$ and with probability $\frac{20}{27} > \frac{2}{3}$ the output is correct; and if $G$ satisfies $Q$ then there exists a graph $\tilde{G}$ satisfying $P$ which differs from $G$ in no more than $\frac{2}{81C}\binom{n}{2}$ places, so with probability at least $1 - \frac{2}{27}$ the algorithm does not query an edge where $G$ and $\tilde{G}$ differ, and thus its output is correct with probability at least $\frac{20}{27} - \frac{2}{27} = \frac{2}{3}$. □

This notion of indistinguishability is used in the proofs connecting the combinatorial results to the testability results. For example, it is easily seen that the property of a graph being properly $k$-colorable is

indistinguishable from the property that after isolating $k$ vertices from the graph (by removing all edges containing them) the graph is properly $k$-colorable. This in turn is indistinguishable from the first order property that there exist $k$ vertices $v_1, \ldots, v_k$ such that every other vertex is adjacent to exactly one of $v_1, \ldots, v_k$, but no two endvertices of the same edge of $G$ are adjacent to the same $v_i$ (so the edges between $v_1, \ldots, v_k$ and the other vertices encode in fact a proper $k$-coloring).

The proof in Section 7 of the existence of a non-testable first order graph property is in fact based on the formulation of a first order property which encodes an isomorphism between two graphs, and is indistinguishable from the property that the two graphs have any isomorphism.

Let us now define a generalization of the notion of colorability.

**Definition 2** *Suppose we are given $c$, and a family (with repetitions) $\mathcal{F}$ of graphs, each of which is provided with a $c$-coloring (i.e. a function from its vertex set to $\{1, \ldots, c\}$ which is not necessarily a proper $c$-coloring in the usual sense).*

*A $c$-coloring of a graph $G$ is called an* $\mathcal{F}$-coloring *if no member of $\mathcal{F}$ appears as an induced subgraph of $G$ with an identical coloring. A graph $G$ is called* $\mathcal{F}$-colorable *if there exists an $\mathcal{F}$-coloring of it.*

For example, if $\mathcal{F}$ consists of $c$ copies $F_1, \ldots, F_c$ of $K_2$, both vertices of each $F_i$ being colored with $i$, then $\mathcal{F}$-colorability of a graph $G$ simply means the proper $c$-colorability of $G$ in the usual sense. On the other hand, if $c = 1$ and $\mathcal{F}$ is any fixed family of graphs, $\mathcal{F}$-colorability means the property of not having any member of $\mathcal{F}$ as an induced subgraph. Other properties, such as the property of having a coloring with two colors without any monochromatic triangle, are expressible as instances of $\mathcal{F}$-colorability too.

The following lemma shows the relevance of this notion to first order graph properties of type "$\exists\forall$".

**Lemma 2.2** *For every first order property $P$ of the form $\exists x_1, \ldots, x_t \forall y_1, \ldots, y_s A(x_1, \ldots, x_t, y_1 \ldots, y_s)$ there exists a family $\mathcal{F}$ of $(2^{t+\binom{t}{2}} + 1)$-colored graphs, each with at most $\max\{2, t+1, s\}$ vertices, such that the property $P$ is indistinguishable from the property of being $\mathcal{F}$-colorable.*

**Proof:** Given the property $\exists x_1, \ldots, x_t \forall y_1, \ldots, y_s A(x_1, \ldots, x_t, y_1 \ldots, y_s)$, we define $\mathcal{F}$ as follows. We assume that $A(x_1, \ldots, x_t, y_1 \ldots, y_s)$ allows us to restrict our attention to cases where $x_1, \ldots, x_t, y_1 \ldots, y_s$ are all assigned distinct values (otherwise there exists a property satisfying this which is identical to $P$ for all graphs with at least $s + t$ vertices, and the graphs with a smaller number of vertices do not matter for the purpose of proving indistinguishability). For the simplicity of the presentation we use the color set

$$\{(0,0)\} \cup \{(a,b) | 1 \leq a \leq 2^{\binom{t}{2}}, 1 \leq b \leq 2^t\}.$$

For what follows we also use an enumeration of the $2^{\binom{t}{2}}$ possible graphs with $t$ vertices $u_1, \ldots, u_t$, and an enumeration of the $2^t$ possible adjacency relations of a vertex $v$ to $t$ vertices $u_1, \ldots, u_t$. We impose upon

the coloring of $G$ the following restrictions. Note that each of them is expressible by forbidding certain induced subgraphs with given colorings.

- The color $(0,0)$ appears at most $t$ times in the coloring of $G$ (simply disallow all possible graphs with $t+1$ vertices which are all colored with this color).
- For $1 \leq a < a' \leq 2^{\binom{t}{2}}$ and $1 \leq b, b' \leq 2^t$, at most one of the colors $(a,b)$ and $(a',b')$ appears in the coloring (so $G$ is in fact colored by the set $\{(0,0)\} \cup \{(a,b) | 1 \leq b \leq 2^t\}$ for some fixed $a$).
- Suppose that $K$ is a graph with vertices $w_1, \ldots, w_s$, which are colored with $(a,b_1), \ldots, (a,b_s)$ respectively for some $a > 0$. In order to decide if such a $K$ is to be disallowed, we consider the graph $L$ with vertices $u_1, \ldots, u_t, v_1, \ldots, v_s$ and the following edges. The edges within $u_1, \ldots, u_t$ are defined in correspondence to $a$ (using the enumeration of all graphs with $t$ labeled vertices), and for $1 \leq j \leq s$ the edges between $v_j$ and $u_1, \ldots, u_t$ are defined in correspondence to $b_j$ (using the enumeration of all possible adjacencies of a vertex to $t$ other vertices). The subgraph of $L$ induced by $v_1, \ldots, v_s$ is made identical to $K$ where $v_i$ plays the role of $w_i$. Having thus defined $L$, we disallow the coloring of $K$ where $w_i$ is colored by $(a,b_i)$ for $1 \leq i \leq s$, if and only if $A(u_1, \ldots, u_t, v_1, \ldots, v_s)$ is false in relation to $L$.

We now claim that the property $P$ is indistinguishable from the property of being $\mathcal{F}$-colorable. If a graph $G$ satisfies $P$, then there exist vertices $u_1, \ldots, u_t$ of $G$ such that $\forall y_1, \ldots, y_s A(u_1, \ldots, u_t, y_1 \ldots, y_s)$ is satisfied over $G$ where $y_1, \ldots, y_s$ range over all vertices other than $u_1, \ldots, u_t$. We let $1 \leq a \leq \binom{t}{2}$ correspond to the subgraph of $G$ spanned by $u_1, \ldots, u_t$. We color $u_1, \ldots, u_t$ by $(0,0)$, and we color every other vertex $v$ of $G$ with $(a,b_v)$, where $1 \leq b_v \leq 2^t$ corresponds to the adjacency relations of $v$ to $u_1, \ldots, u_t$. This is clearly seen to be an $\mathcal{F}$-coloring of $G$.

On the other hand, we show that given an $\mathcal{F}$-coloring of $G$, we can modify $G$ by adding and removing no more than $tn$ edges to get a graph $\tilde{G}$ which satisfies $P$ (so we can choose $N(\epsilon) = 2t\epsilon^{-1} + 1$ for the definition of indistinguishability). Given an $\mathcal{F}$-coloring of $G$, we first modify it so there are exactly $t$ vertices colored with $(0,0)$ (if there are less than $t$, we just choose additional vertices arbitrarily and recolor them with $(0,0)$). Denote these vertices by $u_1, \ldots, u_t$. Remember that all colors appearing in the given coloring apart from $(0,0)$ share the same first coordinate, and denote it by $a$.

To define $\tilde{G}$ from $G$ the adjacencies between $u_1, \ldots, u_t$ and all the vertices of $G$ are redefined as follows. The subgraph of $G$ induced by $u_1, \ldots, u_t$ is made to correspond to $a$. Every vertex $v$ different from $u_1, \ldots, u_t$ is colored with $(a,b_v)$ for some $1 \leq b_v \leq 2^t$; the adjacencies between $v$ and $u_1, \ldots, u_t$ are then made to correspond to $b_v$. It is now easily seen that for every $v_1, \ldots, v_s$ different from $u_1, \ldots, u_t$ the proposition $A(u_1, \ldots, u_t, v_1, \ldots, v_s)$ holds in $\tilde{G}$, so $\tilde{G}$ satisfies $P$. $\square$

The property of a graph being $\mathcal{F}$-colorable, for any fixed finite family $\mathcal{F}$ of vertex colored graphs, is shown to be testable at the end of Section 6 as a corollary of a combinatorial result. In fact, something a little stronger than testability is proven, since the given algorithm has one-sided error probability – for $\mathcal{F}$-colorable graphs it gives the correct answer with probability 1.

# 3 Partitions and regularity

For every two nonempty disjoint vertex sets $A$ and $B$ of a graph $G$, we define $e(A,B)$ to be the number of edges of $G$ between $A$ and $B$. The *edge density* of the pair is defined by $d(A,B) = \frac{e(A,B)}{|A||B|}$; for convenience we also use the notation $d^2(A,B) = (d(A,B))^2$. We say that the pair $A,B$ is *$\gamma$-regular*, if for any two subsets $A'$ of $A$ and $B'$ of $B$, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, their edge density satisfies $|d(A',B') - d(A,B)| < \gamma$.

One simple yet useful property of regularity is that it is somewhat preserved when moving to subsets, as the following trivial lemma shows.

**Lemma 3.1** *If $A,B$ is a $\gamma$-regular pair with density $\delta$, and $A' \subset A$ and $B' \subset B$ satisfy $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$ for some $\epsilon \geq \gamma$, then $A',B'$ is a $\max\{2,\epsilon^{-1}\}\gamma$-regular pair with density at least $\delta - \gamma$ and at most $\delta + \gamma$.* □

The following lemma shows how the existence of regular pairs implies the existence of many induced subgraphs of a fixed type. Many similar lemmas have been proven in previous works; for completeness we present a self contained proof of this one.

**Lemma 3.2** *For every $0 < \eta < 1$ and $k$ there exist $\gamma = \gamma_{3.2}(\eta,k)$ and $\delta = \delta_{3.2}(\eta,k)$ with the following property.*

*Suppose that $H$ is a graph with vertices $v_1,\ldots,v_k$, and that $V_1,\ldots,V_k$ is a $k$-tuple of disjoint vertex sets of $G$ such that for every $1 \leq i < i' \leq k$ the pair $V_i,V_{i'}$ is $\gamma$-regular, with density at least $\eta$ if $v_iv_{i'}$ is an edge of $H$, and with density at most $1-\eta$ if $v_iv_{i'}$ is not an edge of $H$. Then, at least $\delta\prod_{i=1}^{k}|V_i|$ of the $k$-tuples $w_1 \in V_1,\ldots,w_k \in V_k$ span (induced) copies of $H$ where each $w_i$ plays the role of $v_i$.*

**Proof:** Without loss of generality assume that $H$ is the complete graph; otherwise, for each $i < i'$ such that $v_iv_{i'}$ is not an edge of $H$, exchange all edges and non-edges of $G$ between $V_i$ and $V_{i'}$ and regard $v_iv_{i'}$ as an edge of $H$. Assume also $\eta < 1$. The proof is by induction on $k$. The case $k = 1$ is trivial. Supposing that we know that $\gamma_{3.2}(\eta,k-1)$ and $\delta_{3.2}(\eta,k-1)$ exist for all $\eta$, we show that we can choose

$$\gamma = \gamma_{3.2}(\eta,k) = \min\{\frac{1}{2k-2}, \frac{1}{2}\eta\gamma_{3.2}(\frac{1}{2}\eta,k-1)\}$$

and

$$\delta = \delta_{3.2}(\eta, k) = \frac{1}{2}(\eta - \gamma)^{k-1}\delta_{3.2}(\frac{1}{2}\eta, k-1).$$

For each $1 < i \leq k$, the number of vertices of $V_1$ which have less than $(\eta - \gamma)|V_i|$ neighbors in $V_i$ is less than $\gamma|V_1|$, because otherwise these would constitute a counter example to the regularity of $V_1, V_i$. Therefore, at least $(1 - (k-1)\gamma)|V_1| \geq \frac{1}{2}|V_1|$ of the vertices of $V_1$ have at least $(\eta - \gamma)|V_i|$ neighbors in $V_i$ for all $i > 1$.

For each such vertex $w_1$ of $V_1$, let $V_i'$ denote the set of its neighbors in $V_i$. Since $\gamma \leq \frac{1}{2}\eta$, Lemma 3.1 ensures that for each $1 < i < i' \leq k$, the pair $V_i', V_{i'}'$ is $\frac{\gamma}{\eta - \gamma} \leq 2\eta^{-1}\gamma$-regular and with density at least $\eta - \gamma \geq \frac{1}{2}\eta$. The induction hypothesis now guarantees at least

$$\delta_{3.2}(\frac{1}{2}\eta, k-1)\prod_{i=2}^{k}|V_i'| \geq (\eta - \gamma)^{k-1}\delta_{3.2}(\frac{1}{2}\eta, k-1)\prod_{i=2}^{k}|V_i|$$

possible choices of $w_2 \in V_2, \ldots, w_k \in V_k$ such that the induced subgraph spanned by $w_1, \ldots, w_k$ is complete, so the lemma follows from the existence of at least $\frac{1}{2}|V_1|$ choices of such a $w_1$. $\Box$

A partition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of the vertex set of a graph is called an *equipartition* if $|V_i|$ and $|V_{i'}|$ differ by no more than 1 for all $1 \leq i < i' \leq k$ (so in particular each $V_i$ has one of two possible sizes). A *refinement* of such an equipartition $\mathcal{A}$ is an equipartition of the form $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ such that $V_{i,j}$ is a subset of $V_i$ for every $1 \leq i \leq k$ and $1 \leq j \leq l$.

The *order* $|\mathcal{A}|$ of an equipartition $\mathcal{A}$ is the number of sets in it ($k$ in the above notation). The *index* of the equipartition $\mathcal{A}$ above is defined by

$$\mathrm{ind}(\mathcal{A}) = k^{-2}\sum_{1 \leq i < i' \leq k} d^2(V_i, V_{i'}).$$

The Regularity Lemma of Szemerédi can be formulated as follows.

**Lemma 3.3 ([13])** *For every $m$ and $\epsilon > 0$ there exists a number $T = T_{3.3}(m, \epsilon)$ with the following property.*

*If $G$ is a graph with $n \geq T$ vertices, and $\mathcal{A}$ is an equipartition of the vertex set of $G$ with an order not exceeding $m$, then there exists a refinement $\mathcal{B}$ of $\mathcal{A}$ of order $k$, where $m \leq k \leq T$, for which all pairs of sets but at most $\epsilon\binom{k}{2}$ of them are $\epsilon$-regular.*

The original formulation of the lemma allows also for an exceptional set with up to $\epsilon n$ vertices outside of this equipartition, but one can first apply the original formulation with a somewhat smaller parameter instead of $\epsilon$ and then evenly distribute the exceptional vertices among the sets of the partition to obtain this formulation.

$T_{3.3}(m, \epsilon)$ may and is assumed to be monotone nondecreasing in $m$ and monotone nonincreasing in $\epsilon$. We also assume similar monotonicity properties for other functions appearing here, and assume that the

number of vertices $n$ of the graph $G$ is large enough (as a function of the other parameters) even when this is not mentioned explicitly. The following corollary, some versions of which appear in various papers applying the Regularity Lemma, is useful to what follows.

**Corollary 3.4** *For every $l$ and $\gamma$ there exists $\delta = \delta_{3.4}(l,\gamma)$ such that for every graph $G$ with $n \geq \delta^{-1}$ vertices there exist disjoint vertex sets $W_1, \ldots, W_l$ satisfying:*

- $|W_i| \geq \delta n$.
- *All $\binom{l}{2}$ pairs are $\gamma$-regular.*
- *Either all pairs are with densities at least $\frac{1}{2}$, or all pairs are with densities less than $\frac{1}{2}$.*

**Proof:** We set $\delta = \frac{1}{2}(T_{3.3}(r, \min\{r^{-1}, \gamma\}))^{-1}$, with $r$ to be chosen later. We obtain through Lemma 3.3 an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of the vertices of $G$ with $k \geq r$ and $|V_i| \geq \delta n$ for $1 \leq i \leq k$ (the assumption on $n$ guarantees that this holds for the sets with the smaller size as well), with all pairs of sets but at most $\min\{r^{-1}, \gamma\}\binom{k}{2} < (r-1)^{-1}\binom{k}{2}$ of them being $\min\{r^{-1}, \gamma\} \leq \gamma$-regular.

In particular, by Turán's Theorem (see [5]) there exist $i_1, \ldots, i_r$ such that all pairs taken from $V_{i_1}, \ldots, V_{i_r}$ are regular. We choose $r$ in a manner that Ramsey's Theorem (see [5]) ensures the existence of $j_1, \ldots, j_l$ such that either all pairs taken from $V_{i_{j_1}}, \ldots, V_{i_{j_l}}$ are with densities at least $\frac{1}{2}$, or all these pairs are with densities less than $\frac{1}{2}$. Setting $W_t = V_{i_{j_t}}$ for $1 \leq t \leq l$ we arrive at the required result. □

The proof of the Regularity Lemma uses the defect form of the Schwarz Inequality, which is also used in what follows.

**Lemma 3.5 (see [13])** *For all sequences of nonnegative numbers $X_1, \ldots, X_n$, if for some $m < n$*

$$\sum_{k=1}^{m} X_k = \frac{m}{n} \sum_{k=1}^{n} X_k + \alpha,$$

*then*

$$\sum_{k=1}^{n} X_k^2 \geq \frac{1}{n} \left( \sum_{k=1}^{n} X_k \right)^2 + \frac{\alpha^2 n}{m(n-m)}$$

*($\alpha$ need not be positive).*

The following is an immediate implication with regards to partitions of vertex pairs.

**Corollary 3.6** *Suppose that $A$ and $B$ are two disjoint sets of vertices of $G$, and $\{A_j | 1 \leq j \leq l\}$ and $\{B_j | 1 \leq j \leq l\}$ are their two respective partitions to sets of* **equal** *sizes, such that at least $\epsilon l^2$ of the possible $j, j'$ satisfy $|d(A,B) - d(A_j, B_{j'})| \geq \frac{1}{2}\epsilon$. Then,*

$$\sum_{1 \leq j, j' \leq l} d^2(A_j, B_{j'}) > l^2 (d^2(A,B) + \frac{1}{8}\epsilon^3).$$

**Proof:** Under the above conditions, either at least $\frac{1}{2}\epsilon l^2$ of the pairs $j, j'$ are such that $d(A_j, B_{j'}) - d(A,B) \geq \frac{1}{2}\epsilon$, or at least $\frac{1}{2}\epsilon l^2$ are such that $d(A_j, B_{j'}) - d(A,B) \leq -\frac{1}{2}\epsilon$. We use Lemma 3.5 with $n = l^2$, $m = \frac{1}{2}\epsilon l^2$, and $\alpha$ satisfying $|\alpha| \geq \frac{1}{4}\epsilon^2 l^2$ to obtain

$$\sum_{1 \leq j,j' \leq l} d^2(A_j, B_{j'}) \geq l^2 d^2(A,B) + \frac{\frac{1}{16}\epsilon^4 l^6}{\frac{1}{2}\epsilon(1-\frac{1}{2}\epsilon)l^4} > l^2(d^2(A,B) + \frac{1}{8}\epsilon^3)$$

as required. □

The following lemma shows that if the index of an equipartition $\mathcal{A}$ is not smaller by much than the index of its refinement $\mathcal{B}$, then most of the densities of the pairs of $\mathcal{B}$ are close to the densities of the corresponding pairs of $\mathcal{A}$.

**Lemma 3.7** *Suppose that $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ and its refinement $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ satisfy $\mathrm{ind}(\mathcal{B}) - \mathrm{ind}(\mathcal{A}) \leq \frac{1}{64}\epsilon^4$ for some $\epsilon$, and that the number of vertices of the graph is $n > 512\epsilon^{-4}kl$. Then, for all possible $i < i'$ but at most $\epsilon\binom{k}{2}$ of them, $|d(V_i, V_{i'}) - d(V_{i,j}, V_{i',j'})| < \epsilon$ holds for all but a maximum of $\epsilon l^2$ of the possible $j, j'$.*

**Proof:** Supposing the contrary and assuming $\epsilon < 1$ and $k > 1$, we show that the index of $\mathcal{B}$ is larger than that of $\mathcal{A}$ by more than $\frac{1}{64}\epsilon^4$. If not all of the sets of $\mathcal{B}$ are of exactly the same size, let $V'_{i,j}$ be $V_{i,j}$ for sets of the smaller size, and $V'_{i,j}$ be $V_{i,j}$ minus an arbitrarily chosen vertex for sets of the larger size. Defining also $V'_i = \bigcup_{1 \leq j \leq l} V'_{i,j}$, we define two new partitions $\mathcal{B}' = \{V'_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ and $\mathcal{A}' = \{V'_i | 1 \leq i \leq k\}$ of a large induced subgraph of $G$ (for each of these new partitions all its sets are of the same size). The assumption on $n$ implies that $|d(V_i, V_{i'}) - d(V'_i, V'_{i'})| < \frac{1}{256}\epsilon^4$ and $|d(V_{i,j}, V_{i',j'}) - d(V'_{i,j}, V'_{i',j'})| < \frac{1}{256}\epsilon^4$ hold for all $i, j, i', j'$. In particular, $|\mathrm{ind}(\mathcal{A}) - \mathrm{ind}(\mathcal{A}')| < \frac{1}{128}\epsilon^4$ and $|\mathrm{ind}(\mathcal{B}) - \mathrm{ind}(\mathcal{B}')| < \frac{1}{128}\epsilon^4$ hold, and for more than $\epsilon\binom{k}{2}$ of the possible $i < i'$ the inequality $|d(V'_i, V'_{i'}) - d(V'_{i,j}, V'_{i',j'})| > \epsilon - \frac{2}{256}\epsilon^4 > \frac{1}{2}\epsilon$ holds for at least $\epsilon l^2$ of the possible $j, j'$. Using Corollary 3.6, we obtain

$$\mathrm{ind}(\mathcal{B}') \geq k^{-2}l^{-2} \sum_{\substack{1 \leq i < i' \leq k \\ 1 \leq j,j' \leq l}} d^2(V'_{i,j}, V'_{i',j'}) > k^{-2}l^{-2}(l^2 \sum_{1 \leq i < i' \leq k} d^2(V'_i, V'_{i'}) + \epsilon\binom{k}{2} l^2 \frac{1}{8}\epsilon^3) \geq \mathrm{ind}(\mathcal{A}') + \frac{1}{32}\epsilon^4.$$

This implies $\mathrm{ind}(\mathcal{B}) - \mathrm{ind}(\mathcal{A}) \geq \mathrm{ind}(\mathcal{B}') - \mathrm{ind}(\mathcal{A}') - \frac{2}{128}\epsilon^4 > \frac{1}{64}\epsilon^4$, completing the proof. □

# 4 A lemma suitable for finding induced subgraphs

The following lemma, which can be considered a variant of the Regularity Lemma, is suited for dealing with induced subgraphs to be found in $G$.

**Lemma 4.1** *For every integer $m$ and function $0 < \mathcal{E}(r) < 1$ there exists a number $S = S_{4.1}(m, \mathcal{E})$ with the following property.*

*If $G$ is a graph with $n \geq S$ vertices, then there exists an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ and a refinement $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$ of $\mathcal{A}$ that satisfy:*

- $|\mathcal{A}| = k \geq m$ *but* $|\mathcal{B}| = kl \leq S$.
- *For all* $1 \leq i < i' \leq k$ *but at most* $\mathcal{E}(0)\binom{k}{2}$ *of them the pair* $V_i, V_{i'}$ *is* $\mathcal{E}(0)$*-regular.*
- *For all* $1 \leq i < i' \leq k$, *for all* $1 \leq j, j' \leq l$ *but at most* $\mathcal{E}(k)l^2$ *of them the pair* $V_{i,j}, V_{i',j'}$ *is* $\mathcal{E}(k)$*-regular.*
- *All* $1 \leq i < i' \leq k$ *but at most* $\mathcal{E}(0)\binom{k}{2}$ *of them are such that for all* $1 \leq j, j' \leq l$ *but at most* $\mathcal{E}(0)l^2$ *of them* $|d(V_i, V_{i'}) - d(V_{i,j}, V_{i',j'})| < \mathcal{E}(0)$ *holds.*

**Proof:** We may assume that $m > 1$ and that $\mathcal{E}(r)$ is monotone nonincreasing. For convenience, let $\epsilon = \mathcal{E}(0)$. We define

$$T^{(1)} = T_{3.3}(m, \epsilon),$$

and for $i > 1$ we define by induction

$$T^{(i)} = T_{3.3}(T^{(i-1)}, 2\mathcal{E}(T^{(i-1)})(T^{(i-1)})^{-2}).$$

We show that $S = 512\epsilon^{-4}T^{(64\epsilon^{-4}+1)}$ satisfies the required property.

Given $G$, define $\mathcal{A}_1$ to be an equipartition of order at least $m$ but not greater than $T^{(1)}$, such that all pairs but at most $\epsilon\binom{|\mathcal{A}_1|}{2}$ of them are $\epsilon$-regular. Define by induction for $i > 1$ the equipartition $\mathcal{A}_i$ to be a refinement of $\mathcal{A}_{i-1}$, of order not greater than $T^{(i)}$, such that all of the pairs but at most $2\mathcal{E}(T^{(i-1)})(T^{(i-1)})^{-2}\binom{|\mathcal{A}_i|}{2} \leq 2\mathcal{E}(T^{(i-1)})(|\mathcal{A}_{i-1}|)^{-2}\binom{|\mathcal{A}_i|}{2}$ are $2\mathcal{E}(T^{(i-1)})(T^{(i-1)})^{-2} < \mathcal{E}(T^{(i-1)})$-regular.

Let us now choose the minimum $i$ such that $\mathrm{ind}(\mathcal{A}_i) - \mathrm{ind}(\mathcal{A}_{i-1}) \leq \frac{1}{64}\epsilon^4$. There certainly exists such an $1 < i \leq 64\epsilon^{-4} + 1$ since the indices of the partition series are all between 0 and 1. We set $\mathcal{A} = \mathcal{A}_{i-1}$ and $\mathcal{B} = \mathcal{A}_i$, and appropriately $k = |\mathcal{A}_{i-1}| = |\mathcal{A}|$ and $l = k^{-1}|\mathcal{A}_i| = |\mathcal{A}|^{-1}|\mathcal{B}|$. We claim that $\mathcal{A}$ and $\mathcal{B}$ are the required partitions.

It is clear that $\mathcal{B}$ is a refinement of $\mathcal{A}$ and that they both satisfy the requirements with regards to their respective orders. It is also clear (by the assumption $\mathcal{E}(r) \leq \mathcal{E}(0) = \epsilon$) that $\mathcal{A}$ satisfies the requirement regarding the regularity of its pairs. Since all but at most $2\mathcal{E}(k)k^{-2}\binom{kl}{2} < \mathcal{E}(k)l^2$ of all the pairs of $\mathcal{B}$ are $\mathcal{E}(k)$-regular, the condition regarding the regularity of pairs of $\mathcal{B}$ in the formulation of the lemma follows. Finally, Lemma 3.7 shows that most densities of the pairs of $\mathcal{B}$ differ from the corresponding densities of the pairs of $\mathcal{A}$ by less than $\epsilon$, as in the formulation of the last condition of this lemma. □

**Three comments:**

- It is important to what follows that the function $\mathcal{E}(r)$ and not a constant is used in the formulation of this lemma.

- By using the algorithmic version of the Regularity Lemma from [1], one could obtain an algorithmic version of this variant.
- The function $S_{4.1}(m,\mathcal{E})$ is a fast growing function – even for moderate functions $\mathcal{E}(r)$, it is expressible in terms of the WOW function (a tower of towers) of a polynomial in $\mathcal{E}(0)$ and $m$.

In what follows, we need the following corollary.

**Corollary 4.2** *For every $m$ and $0 < \mathcal{E}(r) < 1$ there exist $S = S_{4.2}(m,\mathcal{E})$ and $\delta = \delta_{4.2}(m,\mathcal{E})$ with the following property.*

*If $G$ is a graph with $n \geq S$ vertices then there exist an equipartition $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ of $G$ and an induced subgraph $G'$ of $G$, with an equipartition $\mathcal{A}' = \{V_i' | 1 \leq i \leq k\}$ of the vertices of $G'$, that satisfy:*

- $S \geq k \geq m$.
- $V_i' \subset V_i$ *for all* $i \geq 1$, *and* $|V_i'| \geq \delta n$.
- *In the equipartition* $\mathcal{A}'$, **all** *pairs are* $\mathcal{E}(k)$*-regular.*
- *All but at most* $\mathcal{E}(0)\binom{k}{2}$ *of the pairs* $1 \leq i < i' \leq k$ *are such that* $|d(V_i,V_{i'}) - d(V_i',V_{i'}')| < \mathcal{E}(0)$.

**Proof:** We may assume $\mathcal{E}(r) \leq \mathcal{E}(0)$. Put $\epsilon = \mathcal{E}(0)$. Define $\mathcal{E}'$ by $\mathcal{E}'(r) = \min\{\mathcal{E}(r), \frac{1}{4}\epsilon, \frac{1}{2}\binom{r+2}{2}^{-1}\}$, set $S = S_{4.1}(m,\mathcal{E}')$, and $\delta = \frac{1}{2}(S_{4.1}(m,\mathcal{E}'))^{-1}$. Use Lemma 4.1 on $G$, finding the appropriate partitions $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$ and $\mathcal{B} = \{V_{i,j} | 1 \leq i \leq k, 1 \leq j \leq l\}$.

Now choose randomly, independently and uniformly $1 \leq j_i \leq l$ for each $1 \leq i \leq k$. Clearly, with probability more than $\frac{1}{2}$, all the pairs $V_{i,j_i}, V_{i',j_{i'}}$ are $\mathcal{E}(k)$-regular. Moreover, the expected number of pairs $1 \leq i < i' \leq k$ for which $|d(V_i,V_{i'}) - d(V_i',V_{i'}')| \geq \epsilon$ is no more than $\frac{1}{4}\epsilon\binom{k}{2} + \frac{1}{4}\epsilon\binom{k}{2} = \frac{1}{2}\epsilon\binom{k}{2}$, so with probability at least $\frac{1}{2}$ no more than $\epsilon\binom{k}{2}$ of the pairs satisfy this.

Therefore, there exists a choice of $j_1,\ldots,j_k$ such that all pairs $V_{i,j_i}, V_{i',j_{i'}}$ are $\mathcal{E}(k)$-regular, and all but at most $\epsilon\binom{k}{2}$ of them satisfy $|d(V_i,V_{i'}) - d(V_{i,j_i},V_{i',j_{i'}})| < \epsilon$. Defining $G'$ as the induced subgraph spanned by $\bigcup_{1\leq i\leq k} V_{i,j_i}$, and $\mathcal{A}'$ by setting $V_i' = V_{i,j_i}$ we arrive at the required result. □

# 5 Unavoidability and abundance of induced subgraphs

In the context of induced subgraphs, a graph $H$ is called *$\epsilon$-unavoidable* in $G$ if no adding and removing of up to $\epsilon|G|^2$ edges of $G$ results in $G$ not having an induced subgraph isomorphic to $H$. $H$ is called *$\delta$-abundant* if $G$ contains at least $\delta|G|^{|H|}$ (distinct) induced subgraphs isomorphic to $H$.

It is trivial that a certain degree of abundance implies a certain degree of unavoidability. The following application of the results of the previous section shows that in the context of induced subgraphs, a certain degree of unavoidability also implies a certain degree of abundance.

**Theorem 5.1** *For every $l$ and $\epsilon$ there exists $\delta = \delta(l,\epsilon)$, such that for any graph $H$ with $l$ vertices, if $H$ is $\epsilon$-unavoidable in a graph $G$, then it is also $\delta$-abundant in $G$.*

**Proof:** We assume $\epsilon < 1$ and let $n = |G|$. We set

$$\delta = \delta_{3.2}(\frac{1}{6}\epsilon, l)(\beta\delta_{4.2}(m, \min\{\frac{1}{6}\epsilon, \alpha\}))^l$$

with $m = 7\epsilon^{-1}$, and $\alpha$ and $\beta$ to be chosen later (for the quantities of Corollary 4.2 we use here the constant function $\mathcal{E}(r) = \min\{\frac{1}{6}\epsilon, \alpha\}$).

We apply Corollary 4.2 to $G$, to find $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$, $G'$ and $\mathcal{A}' = \{V_i' | 1 \leq i \leq k\}$, that satisfy $m \leq k \leq S_{4.2}(m, \min\{\frac{1}{6}\epsilon, \alpha\})$ and $|V_i'| \geq \delta_{4.2}(m, \min\{\frac{1}{6}\epsilon, \alpha\})n$, ensuring also that all pairs of $\mathcal{A}'$ are in particular $\alpha$-regular and the densities of no more than $\frac{1}{6}\epsilon\binom{k}{2}$ of them differ from those of the corresponding pairs of $\mathcal{A}$ by more than $\frac{1}{6}\epsilon$.

Choosing $\beta = \delta_{3.4}(l, \gamma_{3.2}(\frac{1}{6}\epsilon, l))$ and $\alpha = \beta\gamma_{3.2}(\frac{1}{6}\epsilon, l)$, we use Corollary 3.4 on the subgraph induced by $G$ on each $V_i'$ to obtain the appropriate $W_{i,1}, \ldots, W_{i,l}$, all of size at least $\beta|V_i'|$. Note by Lemma 3.1 that in particular for every $i, j, i', j'$ the pair $W_{i,j}, W_{i',j'}$ is $\gamma_{3.2}(\frac{1}{6}\epsilon, l)$-regular, and its density differs from that of $V_i', V_{i'}'$ by no more than $\frac{1}{6}\epsilon$.

Define $\tilde{G}$ to be the graph obtained from $G$ by adding and removing the following edges:

- For $1 \leq i < i' \leq k$ such that $|d(V_i, V_{i'}) - d(V_i', V_{i'}')| > \frac{1}{6}\epsilon$, for all $v \in V_i$ and $v' \in V_{i'}$ the pair $vv'$ becomes an edge if $d(V_i', V_{i'}') \geq \frac{1}{2}$, and becomes a non-edge otherwise. This changes less than $\frac{2}{6}\epsilon\binom{n}{2}$ edges (for $n$ large enough) because there are no more than $\frac{1}{6}\epsilon\binom{k}{2}$ such $1 \leq i < i' \leq k$.

- For $1 \leq i < i' \leq k$ such that $d(V_i', V_{i'}') < \frac{2}{6}\epsilon$, all edges between $V_i$ and $V_{i'}$ are removed. For all $1 \leq i < i' \leq k$ such that $d(V_i', V_{i'}') > 1 - \frac{2}{6}\epsilon$, all non-edges between $V_i$ and $V_{i'}$ become edges. This changes no more than $\frac{3}{6}\epsilon\binom{n}{2}$ edges in addition to those changed by the previous condition.

- If for a fixed $i$ all densities of pairs from $W_{i,1}, \ldots, W_{i,l}$ are less than $\frac{1}{2}$, all edges within the vertices of $V_i$ are removed. Otherwise, all the above mentioned densities are at least $\frac{1}{2}$ (by the choice of the $W_{i,j}$ through Corollary 3.4), in which case all non-edges within $V_i$ become edges. This changes less than $\frac{1}{6}\epsilon\binom{n}{2}$ edges by the choice of $m$ above.

By the $\epsilon$-unavoidability of $H$ in $G$, $\tilde{G}$ still contains an induced subgraph isomorphic to $H$, denote its vertices by $v_1, \ldots, v_l$. Choosing $i_1, \ldots, i_l$ such that $v_j \in V_{i_j}$ for all $1 \leq j \leq l$, we finally note that $W_{i_1,1}, \ldots, W_{i_l,l}$ satisfy the regularity and density conditions (over $G$, not $\tilde{G}$) required for Lemma 3.2 to ensure the existence of $\delta n^l = \delta|G|^{|H|}$ induced subgraphs isomorphic to $H$. $\square$

## 6 Graphs which are far from being colorable

This section is devoted to a generalization of the result of the previous section, using the notion of $\mathcal{F}$-colorability defined in Section 2. We fix a family (with repetitions) $\mathcal{F}$ of graphs, each of which is provided with some $c$-coloring of its vertices, i.e. a function from its vertex set to $\{1, \ldots, c\}$. Recall that a graph $G$ is called *$\mathcal{F}$-colorable* if there exists a $c$-coloring of $G$ such that no member of $\mathcal{F}$ appears as an induced subgraph of $G$ with an identical coloring.

If $c = 1$ and $\mathcal{F} = \{H\}$, being $\epsilon$-far from being $\mathcal{F}$-colorable means having $H$ as an $\epsilon$-unavoidable induced subgraph. The following is a generalization of Theorem 5.1.

**Theorem 6.1** *For every $\epsilon > 0$, $c$ and $l$ there exist $\delta = \delta(c, l, \epsilon)$ and $C = C(c, l, \epsilon)$ with the following property.*

*If $\mathcal{F}$ is a family of $c$-colored graphs, each with $l$ vertices, and $G$ is a graph which is $\epsilon$-far from being $\mathcal{F}$-colorable, then there exists a graph $H$ with no more than $C$ vertices, which is not $\mathcal{F}$-colorable, and is also $\delta$-abundant in $G$.*

For the proof of Theorem 6.1, we need the following simple lemma to help us construct the graph $H$.

**Lemma 6.2** *For every $c$ and $l$ there exists a bipartite graph $L = L_{6.2}(c, l)$ with a bipartition to the classes $U_1$ and $U_2$ of size $p = p_{6.2}(c, l)$ each, satisfying the following. Suppose that $X_1, \ldots, X_{l'}$ for some $l'$ are disjoint sets of vertices of a graph $G$ of size $p$ each, such that for any $1 \leq i < i' \leq l'$ the bipartite subgraph between $X_i$ and $X_{i'}$ is isomorphic to $L$, and suppose that $H$ is a graph with the vertex set $\{w_1, \ldots, w_l\}$ and that $i_1, \ldots, i_l$ are integers between 1 and $l'$.*

*If $X'_1, \ldots, X'_{l'}$ satisfy $X'_i \subset X_i$ and $|X'_i| \geq \frac{1}{c}p$, there exist $x_1 \in X'_{i_1}, \ldots, x_l \in X'_{i_l}$, all different, such that for all $1 \leq s < s' \leq l$ if $i_s \neq i_{s'}$ then $x_s x_{s'}$ is an edge of $G$ if and only if $w_s w_{s'}$ is an edge of $H$.*

**Proof:** For the purpose here we use the fact that regular pairs with specified density ranges are known to exist, for example by considering random bipartite graphs for an appropriate $p$ (a better bound on $p$ can be obtained by a direct proof without using regularity, but for the purpose here this is sufficient).

We take an $L$ which makes $U_1, U_2$ a $\min\{\frac{1}{5}, (2cl)^{-1}\gamma_{3.2}(\frac{1}{5}, l)\}$-regular pair with density between $\frac{2}{5}$ and $\frac{3}{5}$. Given $X'_1, \ldots, X'_{l'}$, choose arbitrarily vertex sets $\{Y_{i,j} | 1 \leq i \leq l', 1 \leq j \leq l\}$, all disjoint, such that $Y_{i,j} \subset X'_i$ and $|Y_{i,j}| \geq (2cl)^{-1}p$. Given $H$ and $i_1, \ldots, i_l$, Lemma 3.2 (together with Lemma 3.1) guarantees in particular the existence of $x_1 \in Y_{i_1,1}, \ldots, x_l \in Y_{i_l,l}$ satisfying the required properties. □

**Proof of Theorem 6.1:** We first show how to construct $H$ after finding in $G$ a certain structure in a similar manner to what was done in the proof of Theorem 5.1. Then, the two required properties of $H$ are proven.

We assume that $\epsilon < 1$ and that $n = |G|$ is large enough. We set

$$C = pS_{4.2}(7\epsilon^{-1}, \mathcal{E})$$

using $p = p_{6.2}(c, l)$, and defining

$$\mathcal{E}(r) = \min\{\frac{1}{6}\epsilon, \delta_{3.4}(p, \gamma_{3.2}(\frac{1}{6}\epsilon, pr))\gamma_{3.2}(\frac{1}{6}\epsilon, pr)\}.$$

$\delta$ in the formulation of the theorem is to be chosen later.

We apply Corollary 4.2 to $G$, to find $\mathcal{A} = \{V_i | 1 \leq i \leq k\}$, $G'$ and $\mathcal{A}' = \{V_i' | 1 \leq i \leq k\}$, that satisfy $7\epsilon^{-1} \leq k \leq \frac{1}{p}C$, $|V_i'| \geq \delta_{4.2}(7\epsilon^{-1}, \mathcal{E})n$, and the regularity and density properties guaranteed by the corollary with regards to $\mathcal{E}(r)$.

We use Corollary 3.4 on the subgraph of $G$ induced by each $V_i'$ to obtain $W_{i,1}, \ldots, W_{i,p}$, with all pairs being $\gamma_{3.2}(\frac{1}{6}\epsilon, pk)$-regular, and either all of the densities being at least $\frac{1}{2}$ or all of them being less than $\frac{1}{2}$. Note that now all pairs $W_{i,j}, W_{i',j'}$ are $\gamma_{3.2}(\frac{1}{6}\epsilon, pk)$-regular.

$\tilde{G}$ is defined to be the graph obtained from $G$ by adding and removing edges according to what follows. $H$ is a graph with a vertex set $\{u_{i,j} | 1 \leq i \leq k, 1 \leq j \leq p\}$, whose edges are also specified here.

- For $1 \leq i < i' \leq k$ such that $|d(V_i, V_{i'}) - d(V_i', V_{i'}')| > \frac{1}{6}\epsilon$, for all $v \in V_i$ and $v' \in V_{i'}$ the pair $vv'$ becomes an edge if $d(V_i', V_{i'}') \geq \frac{1}{2}$, and becomes a non-edge otherwise. In the first case, all pairs $u_{i,j}u_{i',j'}$ for $1 \leq j, j' \leq p$ are edges of $H$. In the second case, all these pairs are non-edges of $H$.

- For $1 \leq i < i' \leq k$ such that $d(V_i', V_{i'}') < \frac{2}{6}\epsilon$, all edges between $V_i$ and $V_{i'}$ are removed, and all pairs $u_{i,j}u_{i',j'}$ for $1 \leq j, j' \leq p$ are non-edges of $H$. For all $1 \leq i < i' \leq k$ such that $d(V_i', V_{i'}') > 1 - \frac{2}{6}\epsilon$, all non-edges between $V_i$ and $V_{i'}$ become edges, and all pairs $u_{i,j}u_{i',j'}$ for $1 \leq j, j' \leq p$ are edges of $H$.

- For $1 \leq i < i' \leq k$ such that none of the above holds, the edges of $\tilde{G}$ between $V_i$ and $V_{i'}$ remain exactly those of $G$. The edges of $H$ between $\{u_{i,1}, \ldots, u_{i,p}\}$ and $\{u_{i',1}, \ldots, u_{i',p}\}$ are then chosen to make the bipartite graph between these sets a copy of $L_{6.2}(c, l)$.

- If for a fixed $i$ all densities of pairs from $W_{i,1}, \ldots, W_{i,p}$ are less than $\frac{1}{2}$, all edges within the vertices of $V_i$ are removed. Otherwise, all the above mentioned densities are at least $\frac{1}{2}$, in which case all non-edges within $V_i$ become edges. In the second case, all pairs of vertices of $H$ from $\{u_{i,j} | 1 \leq j \leq p\}$ are edges of $H$. In the first case, all these pairs are non-edges of $H$.

Since $G$ is $\epsilon$-far from being $\mathcal{F}$-colorable, $\tilde{G}$ is not $\mathcal{F}$-colorable. We set

$$\delta = \delta_{3.2}(\frac{1}{6}\epsilon, C)(\delta_{3.4}(p, \gamma_{3.2}(\frac{1}{6}\epsilon, C))\delta_{4.2}(7\epsilon^{-1}, \mathcal{E}))^C.$$

We are now ready to prove that $H$ satisfies the required properties.

**Claim 1** *The constructed $H$ is not $\mathcal{F}$-colorable.*

**Proof:** Suppose that we are given a coloring $\mathcal{C}$ of $H$. We define a coloring $\mathcal{D}$ of $\tilde{G}$ as follows. For every $1 \leq i \leq k$, we choose a color $a$ which appears at least $\frac{1}{c}p$ times among the colors of $\{u_{i,1}, \ldots, u_{i,p}\}$ supplied by $\mathcal{C}$, and color all $v \in V_i$ with $a$; defining $X_i = \{u_{i,1}, \ldots, u_{i,p}\}$, we also define $X_i' \subset X_i$ to consist of those vertices of $X_i$ which are colored $a$.

Since $\tilde{G}$ is not $\mathcal{F}$-colorable, there exists in $\mathcal{F}$ a graph $F$ with a coloring $\mathcal{K}$ which appears with the same colors in the coloring $\mathcal{D}$ of $\tilde{G}$. Denote its vertices by $v_1, \ldots, v_l$, and choose $i_1, \ldots, i_l$ such that $v_s \in V_{i_s}$ for all $1 \leq s \leq l$. Since $|X_i'| \geq \frac{1}{c}p$ for all $i$, the construction of $H$ (by Lemma 6.2) ensures the existence of vertices $x_1 \in X_{i_1}', \ldots, x_l \in X_{i_l}'$ by which a copy of $F$ is spanned from $H$, and for which the colorings $\mathcal{C}$ and $\mathcal{K}$ agree. Thus $\mathcal{C}$ is not an $\mathcal{F}$-coloring of $H$. □

**Claim 2** *The constructed $H$ is $\delta$-abundant in $G$.*

**Proof:** This follows from Lemma 3.2, since the $W_{i,j}$ satisfy

$$|W_{i,j}| > ((\delta_{3.2}(\frac{1}{6}\epsilon, pk))^{-1}\delta)^{1/C}n$$

and also satisfy the required regularity and density conditions over $G$. □

Given $c$ and a fixed finite family $\mathcal{F}$ of $c$-colored graphs, it is now easy to show that the property of being $\mathcal{F}$-colorable is testable.

**Corollary 6.3** *For every $c$, a family $\mathcal{F}$ of $c$-colored graphs, and $\epsilon$, there exists an $\epsilon$-test for the property of being $\mathcal{F}$-colorable, which makes a constant number of queries.*

**Proof:** From Theorem 6.1 we know the existence of $C$ and $\delta$, such that for every graph $G$ which is $\epsilon$-far from being $\mathcal{F}$-colorable, there exists a graph with $C$ vertices which is not $\mathcal{F}$-colorable and is $\delta$-abundant in $G$. In particular, for a uniformly random choice of $C$ vertices of $G$, with probability at least $C!\delta$ they span a subgraph of $G$ which is not $\mathcal{F}$-colorable.

The $\epsilon$-test is designed as follows. Choosing a constant $D$ such that $(1 - C!\delta)^D \leq \frac{1}{3}$, we choose a uniformly random set of $DC$ vertices of $G$, query about all $\binom{DC}{2}$ pairs of the chosen set, and check whether the induced subgraph of $G$ spanned by this set is $\mathcal{F}$-colorable. If $G$ is $\mathcal{F}$-colorable, this induced subgraph of $G$ is also $\mathcal{F}$-colorable (with probability 1). If $G$ is $\epsilon$-far from being $\mathcal{F}$-colorable, the choice of $C$ and $D$ ensures that with probability at least $\frac{2}{3}$ the induced subgraph produced from the queries is not $\mathcal{F}$-colorable. □

**Proof of the first part of Theorem 1.1:** Immediate from Lemma 2.2 and Corollary 6.3. □

# 7 Non-testable first order properties

In the following, whenever a cycle of $G$ is mentioned, it means a subgraph of $G$ (not necessarily an induced subgraph) which is isomorphic to that cycle.

The vertices $u_1, u_2, u_3, u_4, u_5, u_6$ of a graph $G$ are said to *span an arrow in this order* if the subgraph of $G$ spanned by them contains exactly the edges $u_1u_2, u_2u_3, u_3u_4, u_4u_5, u_5u_6, u_1u_3, u_2u_4, u_2u_6$. The term "arrow" does not refer to the graphical representation of this subgraph but to the fact that its only automorphism is the identity, which allows us to say that the arrow "points" from $u_1$ to $u_5$. It has also some additional properties which are used below.

We now formulate a graph property in terms of a graph isomorphism, next show that it is equivalent to some first order property, and then that it is non-testable.

**Definition 3** *A graph $G$ with $n$ vertices is said to* satisfy property $I$ *if $n = 6s$ for some $s$, and $G$ consists of $4s$ isolated vertices and two vertex disjoint (isomorphic) copies of some graph $H$ with $s$ vertices which has no triangles or pentagons (i.e. cycles of size 3 or 5).*

**Lemma 7.1** *There exists a first order property of type "$\forall\exists$" which is indistinguishable from $I$.*

**Proof:** We define a first order property of graphs in terms of the following restrictions. Note that the first restriction is of type "$\forall\exists$", while all other restrictions are of type "$\forall$". Thus their conjunction is a first order property of type "$\forall\exists$".

- For every vertex $x$ there exist vertices $y_1, y_2, y_3, y_4, y_5$ such that the subgraph of $G$ induced by $\{x, y_1, \ldots, y_5\}$ is an arrow in some order.
- If $x_1, \ldots, x_6$ span an arrow in this order, then $x_2$, $x_3$, $x_4$ and $x_6$ do not have any neighbor in $G$ outside of $x_1, \ldots, x_6$ (this condition is equivalent to disallowing all induced subgraphs with vertices $u_1, \ldots, u_7$ such that $u_1, \ldots, u_6$ span an arrow in this order and $u_7$ is adjacent to a subset of them including at least one of $u_2, u_3, u_4, u_6$).
- If $x_1, \ldots, x_6$ span an arrow in this order, then neither $x_1$ nor $x_5$ are part of any triangle or pentagon (i.e. a cycle of size 3 or 5), except those which are contained in $\{x_1, \ldots, x_6\}$. Moreover, there is no cycle of size 6 or less which contains both $x_1$ and $x_5$ and is not contained in $\{x_1, \ldots, x_6\}$.
- If $x_1, \ldots, x_6$ and $x_7, \ldots, x_{12}$ span arrows in the respective orders, then $x_7$ is not a neighbor of $x_5$, $x_{11}$ is not a neighbor of $x_1$, and finally $x_7$ is a neighbor of $x_1$ if and only if $x_{11}$ is a neighbor of $x_5$.

**Claim 1** *If a graph $G$ with $n$ vertices satisfies these conditions, then $n = 6s$ for some $s$ and the vertex set of $G$ can be partitioned into $s$ sets of size 6 such that each of them spans an arrow.*

**Proof:** The first condition implies that every vertex of $G$ is contained in some (induced) arrow, so it is sufficient to prove that every two arrows are vertex disjoint. An arrow contains a cycle $u_1u_2u_6u_5u_4u_3u_1$ of length 6. Thus if $u_1, \ldots, u_6$ span an arrow (in this order), there is no other arrow which contains both $u_1$ and $u_5$ because they are not both contained in any other such cycle. Since $u_2, u_3, u_4, u_6$ have no neighbors outside of $u_1, \ldots, u_6$, no arrow can contain any of them without containing both $u_1$ and $u_5$ because of its 2-connectedness. Finally, no arrow can contain one of the vertices $u_1$ and $u_5$ without containing any of $u_2, u_3, u_4, u_6$ because this would make this vertex a part of a triangle or a pentagon which is not contained in $\{u_1, \ldots, u_6\}$. □

**Claim 2** *If a graph $G$ with $n = 6s$ vertices satisfies the above conditions, and the vertices of $G$ are relabeled $u_1, \ldots, u_n$ such that for every $1 \leq i \leq s$ the vertices $u_{6i-5}, \ldots, u_{6i}$ span an arrow in this order, then the subgraph $H$ induced by $\{u_{6i-5} | 1 \leq i \leq s\}$ and the subgraph $H'$ induced by $\{u_{6i-1} | 1 \leq i \leq s\}$ are isomorphic and have no edges between them.*

**Proof:** The last condition guarantees that there are no edges between $H$ and $H'$, and also that by mapping the vertex $u_{6i-5}$ to the vertex $u_{6i-1}$ for every $1 \leq i \leq s$ one gets an isomorphism between $H$ and $H'$. □

Returning to the proof of the lemma, let us first assume that $G$ satisfies the conditions above, and that its vertices are relabeled $u_1, \ldots, u_n$ as in Claim 2 above. We modify $G$ by removing all edges among $u_{6i-5}, \ldots, u_{6i}$ for every fixed $i$. This changes $G$ in $8s = \frac{4}{3}n$ places. The resulting graph contains (by Claim 2) two isomorphic induced subgraphs $H$ and $H'$ with $s$ vertices each and no other edges, with $H$ and $H'$ being also triangle and pentagon free. Thus $I$ is satisfied.

On the other hand, if $G$ satisfies $I$, we relabel its vertices by $u_1, \ldots, u_n$ such that the subgraph induced by $\{u_{6i-5} | 1 \leq i \leq s\}$ and the subgraph induced by $\{u_{6i-1} | 1 \leq i \leq s\}$ are isomorphic by the mapping which takes $u_{6i-5}$ to $u_{6i-1}$, and $G$ contains no edges not contained in one of these graphs. Remember also that $G$ contains no triangles or pentagons. We then modify $G$ by placing 8 edges among $u_{6i-5}, \ldots, u_{6i}$ for each fixed $i$, so that $u_{6i-5}, \ldots, u_{6i}$ span an arrow in the modified graph in that order. The modified graph satisfies the first order property defined above (for every fixed $i$ the distance between $u_{6i-5}$ and $u_{6i-1}$ is now 3 so none of these pairs would become a part of a single cycle of size 6 or less which is not contained in the corresponding arrow). □

We say that a graph $H$ and a graph $H'$ with $s$ vertices each are *$\epsilon$-apart* if no graph which differs from $H$ in no more than $\epsilon s^2$ places is isomorphic to $H'$. In the following we use the existence of many graphs which are mutually far apart. This is a consequence of the following simple lemma.

**Lemma 7.2** *There exist constants $\epsilon = \epsilon_{7.2}$ and $S = S_{7.2}$ such that every graph $H$ with $s > S$ (labeled) vertices is $\epsilon$-far from all other graphs with the same vertex set but at most $2^{s^2/5}$ of them.*

**Proof:** Choose $\epsilon < \frac{1}{2}$ so that $(\frac{e}{\epsilon})^{\epsilon} < 2^{1/10}$, and choose $S > \epsilon^{-1}$ so that for every $s > S$ the inequality $s^s < 2^{s^2/10}$ holds. Now the number of graphs which are not $\epsilon$-apart from a given graph $H$ with $s$ vertices is bounded by

$$s! \sum_{i=0}^{\epsilon s^2} \binom{\binom{s}{2}}{i} < s! \binom{s^2}{\epsilon s^2} < s^s (\frac{e}{\epsilon})^{\epsilon s^2} < 2^{s^2/5}$$

as required. □

The following simple lemma about bipartite graphs follows from standard bounds for binomial distribution (see e.g. [3], Appendix A).

**Lemma 7.3** *There exists $T = T_{7.3}$ such that for every $t > T$, at least $\frac{1}{2}2^{t^2}$ of the possible bipartite graphs with two given (labeled) classes $U_1$ and $U_2$ of size $t$ each have minimum degree more than $\frac{1}{3}t$, and are such that for every subset $X$ of $U_1 \cup U_2$ with size between $\frac{1}{3}t$ and $t$ there are more than $\frac{1}{18}t^2$ edges between $X$ and $(U_1 \cup U_2) - X$.* □

We can now prove the existence of two graphs which are far apart and satisfy some other properties, and yet have the same statistics for small induced subgraphs.

**Proposition 7.4** *For every $D$ there exist two bipartite graphs $H = H_{7.4}(D)$ and $H' = H'_{7.4}(D)$, both with a bipartition into two classes of size $t = t_{7.4}(D)$ each, and satisfying the properties appearing in the formulation of Lemma 7.3, which are $\epsilon_{7.2}$-apart, and yet every possible graph with $D$ vertices occurs exactly the same number of times as an induced subgraph in each of them.*

**Proof:** Clearly there are less than $2^{\binom{D}{2}}$ possible graphs with $D$ vertices, and each appears less than $s^D$ times in a graph with $s$ vertices. Defining the constant $E = 2^{\binom{D}{2}}$, we note that there are no more than $(s^D)^E = 2^{DE \log s}$ possible appearance counts for all such graphs. We choose $s = 2t > \max\{S_{7.2}, 2T_{7.3}\}$ such that

$$\frac{1}{2}2^{t^2} = 2^{s^2/4-1} > 2^{DE \log s} 2^{s^2/5} = 2^{s^2/5 + DE \log s}$$

holds. The existence of the required $H$ and $H'$ with $s = 2t$ vertices each follows now from the pigeonhole principle and the previous lemmas. □

**Corollary 7.5** *Property $I$ is not testable.*

**Proof:** We show that there exists no $\epsilon$-test for $I$ where $\epsilon = \frac{1}{36}\min\{\frac{1}{2}\epsilon_{7.2}, \frac{1}{72}\}$. Assuming that there exists such a test, we may assume (see Section 1) that it queries about all pairs from a uniformly random subset of size $D$ chosen from the vertex set of the input graph, and gives output according to the resulting induced subgraph.

Let $G$ be a graph with $12t = 12t_{7.4}(D)$ vertices which consists of a vertex disjoint union of $8t$ isolated vertices and two copies of $H_{7.4}(D)$. Let $G'$ be a graph with $12t$ vertices which consists of a vertex disjoint union of $8t$ isolated vertices, a copy of $H_{7.4}(D)$, and a copy of $H'_{7.4}(D)$.

The given $\epsilon$-test would have precisely the same output probabilities for both $G$ and $G'$, because both contain the same number of each possible graph with $D$ vertices as induced subgraphs. However, $G$ satisfies $I$ (remember that $H$ has in particular no triangles or pentagons) while $G'$ is $\epsilon$-far from satisfying $I$ (because $H$ and $H'$ are $\epsilon_{7.2}$-apart, while their other properties prevent modifying $G'$ by exchanging many of the vertices between the copy of $H$ and the copy of $H'$). This is a contradiction. □

**Proof of the second part of Theorem 1.1:** Immediate from Lemma 7.1 and Corollary 7.5. □

# 8 Concluding remarks and open problems

### Monotone first order properties

We call a first order property $P$ *monotone (nondecreasing)* if it is of the form

$$\forall x_{1,1}, \ldots, x_{1,t_1} \exists x_{2,1}, \ldots, x_{2,t_2} \forall x_{3,1}, \ldots, x_{3,t_3} \ldots \exists x_{2s,1}, \ldots, x_{2s,t_{2s}} A(x_{1,1}, \ldots, x_{2s,t_{2s}})$$

where $A(x_{1,1}, \ldots, x_{2s,t_{2s}})$ is a quantifier free expression about graphs which is monotone (i.e. if it holds for some assignment of its variables with respect to a graph $G$ then it also holds for the same assignment with respect to any graph containing $G$). We assume without loss of generality that it is enough to restrict the range of the variables to the case where they are all assigned distinct values. The expressive power of these properties is far less than that of general first order properties, as shown in the following.

**Proposition 8.1** *Every monotone first order property is indistinguishable from some monotone first order property of type "∀".*

**Proof:** Given the property $P$ defined by $\forall x_{1,1}, \ldots, x_{1,t_1} \ldots \exists x_{2s,1}, \ldots, x_{2s,t_{2s}} A(x_{1,1}, \ldots, x_{2s,t_{2s}})$ with $A$ being monotone, we write $A$ in terms of all the (labeled) subgraphs which $x_{1,1}, \ldots, x_{1,t_1}, \ldots, x_{2s,1}, \ldots, x_{2s,t_{2s}}$ are allowed to induce. We construct the property $Q$ defined by

$$\forall x_{1,1}, \ldots, x_{1,t_1}, x_{3,1}, \ldots, x_{3,t_3}, \ldots, x_{2s-1,1}, \ldots, x_{2s-1,t_{2s-1}} B\left(x_{1,1}, \ldots, x_{2s-1,t_{2s-1}}\right),$$

where $B$ states that the subgraph induced by $x_{1,1}, \ldots, x_{1,t_1}, x_{3,1}, \ldots, x_{3,t_3}, \ldots, x_{2s-1,1}, \ldots, x_{2s-1,t_{2s-1}}$ is a restriction of one of the subgraphs, allowable by $A$ for $x_{1,1}, \ldots, x_{1,t_1}, \ldots, x_{2s,1}, \ldots, x_{2s,t_{2s}}$, to the corresponding subset of the variables. $Q$ is clearly a monotone first order property of type "∀".

Clearly a graph that satisfies $P$ satisfies also $Q$. On the other hand, given a graph $G$ which satisfies $Q$, consider the graph $\tilde{G}$ obtained from $G$ by choosing arbitrarily $\sum_{i=1}^{2s} t_i$ vertices of $G$ and connecting

them to themselves as well as to all other vertices of $G$ (this adds less than $\sum_{i=1}^{2s} t_i n$ edges). It is now easy to see that $\tilde{G}$ satisfies $P$. $\Box$

In particular, all first order monotone properties are testable (the argument for first order monotone *nonincreasing* properties is analogous). In fact, proving the testability of these properties is much easier than proving testability of general "$\exists\forall$" properties, since the Regularity Lemma can be used directly without having to resort to Corollary 4.2.

### Allowing more queries

It may be interesting to define a notion of testability which uses more than a constant number of queries, and use it to further classify various graph properties (in which case it might be no longer legitimate to assume that the queries are about all pairs of a randomly chosen set of vertices, e.g. when the number of allowable queries is $\Theta(n)$).

It seems that a test for the graph isomorphism problem in particular needs a lot of queries (an $\Omega(\sqrt{\log n})$ bound follows in fact from analyzing the proof in the previous section); it would be interesting to find the exact order of magnitude of the number of queries required for testing it. An $\Omega(\sqrt{n})$ bound on the number of queries follows from analyzing the behavior of a tester on one input constructed using two random graphs, and another input constructed using two copies of the same random graph.

### The constants involved

The bound on the number of queries $C$ of the $\epsilon$-test described in Section 6 is huge. It is in fact a tower of towers in both $\epsilon^{-1}$ and the number of variables participating in the given first order expression.

For monotone first order expressions the bound is expressible by the tower function instead (since the Regularity Lemma can be used directly in this case), but this is still extremely large. Moreover, it cannot be avoided as long as the full version of the Regularity Lemma is required for the proofs, as follows from the main result of [9]. For properties shown testable in [8], $\epsilon$-testers whose number of queries is polynomial in $\epsilon^{-1}$ are given. However, it is worth noting that using the number theoretic construction of Behrend [4] we can show that any *one-sided error* $\epsilon$-test for some simple properties, like being triangle-free, requires a number of queries which cannot be bounded by a polynomial in $\epsilon^{-1}$ (by the existence of graphs which are $\epsilon$-far from being triangle-free and yet do not contain too many distinct triangles).

It would be interesting to find a different proof for the testability of properties of type "$\exists\forall$" (or even a subclass of them such as the monotone "$\forall$" properties), without using the Regularity Lemma, as it might give a better bound on the number of queries an $\epsilon$-test makes.

### Further characterization of testable graph properties

Finally, it would be very interesting to give a classification of all testable graph properties. At the moment this seems to be very difficult; many of the properties proven in [8] to be testable do not have one-sided tests, so the notion of $\mathcal{F}$-colorability presented here does not cover them.

## References


[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the Regularity Lemma, *Proceedings of the* $33^{rd}$ *IEEE FOCS* (1992), 473–481. Also: *Journal of Algorithms* 16 (1994), 80–109.

[2] N. Alon, M. Krivelevich, I. Newman and M. Szegedy, Regular languages are testable with a constant number of queries, *Proceedings of the* $40^{th}$ *IEEE FOCS* (1999), 645–655.

[3] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley (1992).

[4] F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression, *Proc. National Academy of Sciences USA* 32 (1946), 331–332.

[5] B. Bollobás, *Extremal Graph Theory*, Academic Press, New York (1978).

[6] B. Bollobás, P. Erdös, M. Simonovits, and E. Szemerédi, Extremal graphs without large forbidden subgraphs, *Annals of Discrete Mathematics* 3 (1978), 29–41.

[7] A. M. Frieze and R. Kannan, The Regularity Lemma and approximation schemes for dense problems, *Proceedings of the* $37^{th}$ *IEEE FOCS* (1996), 12–20.

[8] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *Proceedings of the* $37^{th}$ *IEEE FOCS* (1996), 339–348. Also: *Journal of the ACM*, to appear.

[9] W. T. Gowers, Lower bounds of tower type for Szemerédi's Uniformity Lemma, *Geometric and Functional Analysis* 7 (1997), 322–337.

[10] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory, In: *Combinatorics, Paul Erdös is Eighty*, Vol II (D. Miklós, V. T. Sós, T. Szönyi eds.), János Bolyai Math. Soc., Budapest (1996), 295–352.

[11] V. Rödl and R. Duke, On graphs with small subgraphs of large chromatic number, *Graphs and Combinatorics* 1 (1985), 91–96.

[12] R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM Journal of Computing* 25 (1996), 252–271.

[13] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau eds.), 1978, 399–401.