Testing subgraphs in large graphs

Noga Alon [*]

Abstract

Let H be a fixed graph with h vertices, let G be a graph on n vertices and suppose that at least ϵn^2 edges have to be deleted from it to make it H-free. It is known that in this case G contains at least $f(\epsilon, H)n^h$ copies of H. We show that the largest possible function $f(\epsilon, H)$ is polynomial in ϵ if and only if H is bipartite. This implies that there is a one-sided error property tester for checking H-freeness, whose query complexity is polynomial in $1/\epsilon$, **if and only if** H is bipartite.

1 Introduction

1.1 Preliminaries

All graphs considered here are finite, undirected, and have neither loops nor parallel edges.

Let P be a property of graphs, that is, a family of graphs closed under graph isomorphism. A graph G with n vertices is *ϵ-far from satisfying* P if no graph $\tilde{G}$ with the same vertex set, which differs from G in no more than ϵn^2 places, (i.e., can be constructed from G by adding and removing no more than ϵn^2 edges), satisfies P. An *ϵ-tester* 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, say, $\frac{2}{3}$ between the case of G satisfying P and the case of G being ϵ-far from satisfying P. Such a tester is a *one-sided* tester if when G satisfies P the tester determines that this is the case (with probability 1). Obviously, the probability $\frac{2}{3}$ appearing above can be replaced by any constant smaller than 1, by repeating the algorithm an appropriate number of times.

The property P is called *strongly-testable*, if for every fixed $\epsilon > 0$ there exists a one-sided ϵ-tester for P whose total number of queries is bounded only by a function of ϵ, which is independent of the size of the input graph.

[*] Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: noga@math.tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. A preliminary version of this paper appeared in the Proceedings of the 42^{nd} IEEE FOCS.

1.2 The main result

For a fixed graph H (with at least one edge), let P_H denote the property of being H-free. Therefore, G satisfies P_H iff it contains no (not necessarily induced) subgraph isomorphic to H. It is known that for each fixed graph H, the property P_H is strongly-testable. This is proved (implicitly) in [1], see also [2]. The proof in [1] relies on the regularity lemma of Szemerédi [21], and thus provides a one-sided ϵ-tester for P_H whose query-complexly is bounded by a function which, though independent of the size of the input graph G, has a huge dependency on ϵ and the size of H. For some graphs H, however, there are more efficient testers; for example, if H is a single edge, it is easy to see that there is a one-sided ϵ tester for P_H, which makes only $O(1/\epsilon)$ queries. Our main result here is a precise characterization of all graphs H for which there are one-sided ϵ-testers whose query-complexity (and running time) is polynomial in $1/\epsilon$.

Theorem 1 *Let H be fixed graph on h vertices.*
(i) If H is bipartite, then for every $\epsilon > 0$ there is a one-sided ϵ-tester for P_H whose query-complexity (and running time) are bounded by

$$O(h^2(\frac{1}{2\epsilon})^{h^2/4}).$$

(ii) If H is non-bipartite, then there exists a constant $c = c(H) > 0$ such that the query-complexity (and running time) of any one-sided ϵ-tester for P_H is at least

$$(\frac{c}{\epsilon})^{c\log(c/\epsilon)}.$$

Thus, for example, for all sufficiently small $\epsilon > 0$ and all sufficiently large n, it is much easier to test if an input graph G on n vertices is $K_{100,100}$-free, than to test if it is, say, C_5-free.

1.3 Related work

The general notion of property testing was first formulated by Rubinfeld and Sudan [20], who were motivated mainly by its connection to the study of program checking. The study of the notion of testability for combinatorial objects, and mainly for labeled graphs, was introduced by Goldreich, Goldwasser and Ron [11], who showed that all graph properties describable by the existence of a partition of a certain type, and among them k-colorability, have efficient testers. The fact that k-colorability is strongly testable is, in fact, implicitly proven already in [6] for $k = 2$ and in [17] (see also [1]) for general k, using the Regularity Lemma of Szemerédi [21], but in the context of property testing it is first studied in [11], where far more efficient algorithms are described. These have been further improved in [4].

In [2] it is shown that every first order graph property without a quantifier alternation of type "$\forall\exists$" has testers whose query complexity is independent of the size of the input graph (but has a huge dependence on ϵ). These properties contain the properties P_H whose query complexity is studied here.

The notion of property testing has been investigated in other contexts as well, including the context of regular languages, [3], functions [9] , hypergraphs [8] and other contexts. See [18] for a survey on the topic.

1.4 Organization

The main result consists of two parts. The first one (Theorem 1, part (i)) is not difficult, and relies on known techniques in Extremal Graph Theory dealing with the problem of Zarankiewicz. These techniques, initiated in [16], are applied in Section 2 to show that for any bipartite H, any graph G which is ϵ-far from being H-free, contains many copies of H. Therefore, the ϵ-tester can find a copy of H in any such G with high probability, without making too many queries.

To prove the second part of Theorem 1 we have to construct, for any non-bipartite graph H and any small $\epsilon > 0$, a graph G which is ϵ-far from being H-free and yet contains relatively few copies of H. The proof of this part, described in Section 3, is more difficult, and applies some properties of graph homomorphisms as well as certain constructions in additive number theory, based on (simple variants of) the construction of Behrend [5] of dense subsets of the first n integers without three-term arithmetic progressions.

The final Section 4 contains some concluding remarks and open problems.

Throughout the paper we assume, whenever this is needed, that the number of vertices n of the graph G is sufficiently large. In order to simplify the presentation, we omit all floor and ceiling signs whenever these are not crucial.

2 Bipartite subgraphs

A *homomorphism* of a graph H into a graph G is a function from the vertex set of H to that of G, so that adjacent vertices are mapped into adjacent vertices. Note that the function does not have to be injective. Thus, for example, every bipartite graph can be mapped homomorphically into an edge. More generally, a graph is k-colorable if and only if it admits a homomorphism into a complete graph on k vertices. It is more convenient to count the number of homomorphisms of a graph H into a graph G, than to count the number of subgraphs of G isomorphic to H. The next lemma shows that every dense graph contains many copies of any bipartite graph. The proof is based on known techniques, initiated in [16].

Lemma 2.1 *For every two integers $s \geq t \geq 1$ and for every graph $G = (V, E)$ on n vertices with at least ϵn^2 edges, the number of homomorphisms from a labelled copy of the complete bipartite graph $H = K_{s,t}$ into G is at least $(2\epsilon)^{st} n^{s+t}$.*

Proof: Let $d_1 \geq d_2 \geq \ldots \geq d_n$ be the degrees of the vertices of G, and let $\overline{d} = (\sum_{i=1}^n d_i)/n$ ($\geq 2\epsilon n$) denote the average degree. The number of homomorphisms from a labelled star $K_{1,t}$ into G is

$$\sum_{i=1}^n d_i^t \geq n\overline{d}^t \geq n(2\epsilon n)^t = (2\epsilon)^t n^{t+1},$$

where the first inequality follows from the convexity of the function z^t. Put $N = n^t$, and classify the homomorphisms above into N classes, according to the ordered set of images of the t leaves of the star. Let $D_1, D_2, \ldots, D_N$ be the numbers of homomorphisms of the N possible types. Note that each ordered s-tuple of (not necessarily distinct) homomorphisms of the same type defines a homomorphism of $H = K_{s,t}$ into G by mapping the star whose apex is vertex number i of the first color class of $K_{s,t}$ according to the homomorphism number i in the s-tuple. It follows that the total number of homomorphisms of H into G is at least

$$\sum_{i=1}^N D_i^s \geq N((2\epsilon)^t n)^s = (2\epsilon)^{st} n^{s+t},$$

where the first inequality follows from the convexity of the function z^s. This completes the proof. ■

Corollary 2.1 *For every fixed $\epsilon > 0$, and every fixed two integers $s \geq t \geq 1$, and for any graph G with n vertices and at least ϵn^2 edges, the number of subgraphs of G isomorphic to $H = K_{s,t}$ is at least*

$$(1 + o(1))\binom{n}{s}\binom{n}{t}(2\epsilon)^{st}$$

if $s > t$, and at least

$$(\frac{1}{2} + o(1))\binom{n}{s}\binom{n}{t}(2\epsilon)^{st}$$

for $s = t$, where the $o(1)$ terms tend to 0 as n tends to infinity.

Proof: The number of homomorphisms of H into G which are not injective is at most $O(n^{s+t-1}) = o(n^{s+t})$, and the result thus follows from the previous lemma, after dividing by the number of automorphisms of H. ■

It is worth noting that as shown by the random graph $G(n, 2\epsilon)$ on n labelled vertices in which each pair of vertices, randomly and independently , is an edge with probability 2ϵ, the assertion of the last corollary is tight.

Proof of Theorem 1, part (i): Let H be a bipartite graph with $h = s + t$ vertices (and at least one edge), and suppose it has a bipartition with color classes of sizes s and t. If $G = (V, E)$ is ϵ-far from being H-free then it obviously has at least ϵn^2 edges. Therefore, by Corollary 2.1, it has at least

$$(\frac{1}{2} + o(1))\binom{n}{s}\binom{n}{t}(2\epsilon)^{st}$$

copies of H. Thus, if we choose, randomly and independently, say,

$$10/(2\epsilon)^{st} \leq 10(\frac{1}{2\epsilon})^{h^2/4}$$

pairs of disjoint sets of sizes s and t, and check if they form a copy of $K_{s,t}$ (and hence contain a copy of H), the probability to find a copy of H exceeds $2/3$. The ϵ-tester will thus simply decide that G is H-free iff it finds no copy of H. If G is indeed H-free, then the tester will surely report that's the case. If it is ϵ-far from being H-free, then the probability the tester reports it is not H-free exceeds $2/3$. This completes the proof of Theorem 1, part (i). ■

Remark: By the discussion above, every graph G on sufficiently many vertices with a quadratic number of edges contains a copy of every fixed bipartite graph. Therefore there is a very simple and efficient **two-sided error** algorithm for testing P_H, for every fixed bipartite graph H, based on estimating the number of edges in the input graph G by sampling. The proof above is needed as we deal here with one-sided error testers. See also Section 4 for more details.

3 Non-bipartite subgraphs

In this section we apply techniques from additive number theory, based on the construction of Behrend [5] of dense sets of integers with no three-term arithmetic progressions, together with some properties of graph homomorphisms, to prove part (ii) of Theorem 1.

A linear equation with integer coefficients

$$\sum a_i x_i = 0 \tag{1}$$

in the unknowns x_i is *homogeneous* if $\sum a_i = 0$. If $X \subseteq M = \{1, 2, \ldots, m\}$, we say that X *has no non-trivial solution to* (1), if whenever $x_i \in X$ and $\sum a_i x_i = 0$, it follows that all x_i are equal. Thus, for example, X has no nontrivial solution to the equation $x_1 - 2x_2 + x_3 = 0$ iff it contains no three-term arithmetic progression.

Lemma 3.1 *For every fixed integer $r \geq 2$ and every positive integer m, there exists a subset $X \subset M = \{1, 2, \ldots, m\}$ of size at least*

$$|X| \geq \frac{m}{e^{10\sqrt{\log m \log r}}}$$

with no non-trivial solution to the equation

$$x_1 + x_2 + \ldots + x_r = r x_{r+1}. \tag{2}$$

Proof: Let d be an integer (to be chosen later) and define

$$X = \{\sum_{i=0}^{k} x_i d^i \mid x_i < \frac{d}{r} \ (0 \leq i \leq k) \ \wedge \sum_{i=0}^{k} x_i^2 = B\},$$

where $k = \lfloor \log m / \log d \rfloor - 1$ and B is chosen to maximize the cardinality of X. If $x_1, \ldots x_{r+1} \in X$ satisfy (2) and

$$x_j = \sum_{i=0}^{k} x_{i,j} d^i, \quad \text{for} \quad 1 \leq j \leq r+1$$

then, for every i, $0 \leq i \leq k$

$$x_{i,1} + x_{i,2} + \ldots + x_{i,r} = r x_{i,r+1}.$$

By the convexity of the function $f(z) = z^2$ this implies that

$$x_{i,1}^2 + x_{i,2}^2 + \ldots + x_{i,r}^2 \geq r x_{i,r+1}^2,$$

and the inequality is strict unless all $(r+1)$ numbers $x_{i,j}$ are equal. Thus, X has no nontrivial solution to (2). The size of X satisfies

$$|X| \geq \frac{m}{d^2 r^{k+1} (k+1)^{\frac{d^2}{r^2}}}$$

Take $d = \lfloor e^{\sqrt{\log m \log r}} \rfloor$ to conclude (with room to spare) that

$$|X| \geq \frac{m}{e^{10\sqrt{\log m \log r}}}. \qquad \blacksquare$$

We next apply the construction in the last lemma to construct, for every odd integer $r+1 \geq 3$, a relatively dense graph consisting of pairwise edge disjoint copies of C_{r+1}- the cycle of length $r+1$, which does not contain too many copies of C_{r+1}. Let m be an integer, let $X \subset \{1, 2, \ldots m\}$ be a set satisfying the assertion of Lemma 3.1, and define, for each $1 \leq i \leq r+1$, $V_i = \{1, 2, \ldots im\}$ where, with a slight abuse of notation, we think on the sets V_i as being pairwise disjoint. Let $T = T(r, m)$ be the $r+1$-partite graph on the classes of vertices $V_1, V_2, \ldots, V_{r+1}$, whose edges are defined as follows. For each j, $1 \leq j \leq m$, and for each $x \in X$ the vertices $j \in V_1, j+x \in V_2, j+2x \in V_3, \ldots, j+rx \in V_{r+1}$ form a cycle of length $r+1$ in this order. Therefore, $\{j+ix, j+(i+1)x\}$ is an edge between V_{i+1} and V_{i+2} for all $1 \leq j \leq m, x \in X$ and $0 \leq i \leq r-1$, and $\{j, j+rx\}$ is an edge between V_1 and V_{r+1} for all $1 \leq j \leq m, x \in X$.

Lemma 3.2 *For every even integer $r \geq 2$, and every m, the graph $T(r,m)$ defined above has $(r+1)(r+2)m/2$ vertices, $(r+1)m|X| \geq \frac{m^2}{e^{10\sqrt{\log m \log r}}}$ edges, and precisely $m|X|$ ($< m^2$) copies of the cycle C_{r+1}.*

Proof: The number of vertices and edges of $T(r,m)$ is obviously as stated, as the $m|X|$ cycles appearing in its construction are pairwise edge-disjoint. We thus only have to show that it does not contain any additional cycles C_{r+1} besides those used in the construction. Note that the graph obtained from T by deleting all edges connecting V_1 and V_{r+1} is bipartite, and hence contains no odd cycles. It is thus easy to check that every copy of C_{r+1} in T must contain an edge between V_i

and V_{i+1} for each $1 \leq i \leq r$, and one edge between V_{r+1} and V_1. Therefore, there are $j \leq m$ and elements $x_1, x_2, \ldots, x_{r+1} \in X$, such that the vertices of the cycle are $j \in V_1, j+x_1 \in V_2, j+x_1+x_2 \in V_3, \ldots, j+x_1+x_2+\ldots+x_r \in V_{r+1}$ and $x_1+x_2+\ldots+x_r = rx_{r+1}$. However, by the definition of X this implies that $x_1 = x_2 = \ldots = x_{r+1}$, implying the desired result. ■

An s-blow-up of a graph $K = (V(K), E(K))$ is the graph obtained from K by replacing each vertex of K by an independent set of size s, and each edge of K by a complete bipartite subgraph whose vertex classes are the independent sets corresponding to the ends of the edge.

Lemma 3.3 *Let $H = (V(H), E(H))$ be a graph with h vertices, let $K = (V(K), E(K))$ be another graph on at most h vertices, and let $T = (V(T), E(T))$ be an s-blow-up of K. Suppose there is a homomorphism*

$$f : V(H) \mapsto V(K)$$

from H to K and suppose $s \geq h$. Let $R \subset E(T)$ be a subset of the set of edges of T, and suppose that each copy of H in T contains at least one edge of R. Then

$$|R| \geq \frac{|E(T)|}{|E(K)||E(H)|} > \frac{|E(T)|}{h^4}.$$

Proof: Let $g : V(H) \mapsto V(T)$ be a random injective mapping obtained by defining, for each vertex $v \in V(K)$, the images of the vertices in $f^{-1}(v) \in V(H)$ randomly, in a one-to-one fashion, among all s vertices of T in the independent set that corresponds to the vertex v. Obviously, g maps adjacent vertices of H into adjacent vertices of T, and hence the image of g contains a copy of H in T. Each edge of H is mapped to one of the corresponding s^2 edges of T according to a uniform distribution, and hence the probability it is mapped onto a member of R does not exceed $|R|/s^2$. It follows that the expected number of edges of H mapped to members of R is at most $\frac{|R||E(H)|}{s^2}$, and as, by assumption, this random variable is always at least 1, we conclude that $\frac{|R||E(H)|}{s^2} \geq 1$. The desired result follows, since $s^2 = |E(T)|/|E(K)|$. ■

Lemma 3.4 *For every fixed, non-bipartite graph $H = (V(H), E(H))$ on h vertices, there is a constant $c = c(H) > 0$, such that for every positive $\epsilon < \epsilon_0(H)$ and every integer $n > n_0(\epsilon)$, there is a graph G on n vertices which is ϵ-far from being H-free, and yet contains at most $(\epsilon/c)^{c \log(c/\epsilon)} n^h$ copies of H.*

Proof: Let $r+1$ denote the length of the shortest odd cycle of H. Let K be a subgraph of H, such that there is a homomorphism of H to K and K has the minimum possible number of edges among all subgraphs of H satisfying this property. The graph K is called the *core* of H (see [15]), a notion that has some interesting properties. Note that K must contain a cycle of length $r+1$, as a homomorphic image of any odd cycle must contain an odd cycle which is not longer, and K, which is a subgraph of H, does not contain odd cycles of length shorter than $r+1$. Let k denote the number of vertices of K, and let us number its vertices $\{v_1, v_2, \dots, v_k\}$ such that the first $r+1$ vertices $v_1, v_2, \dots v_{r+1}$ form a cycle in this order. By the minimality of K, every homomorphism of K into itself must be an automorphism, implying that in any homomorphism of H into K there is a cycle of length $r+1$ in H which is mapped onto the cycle of K on the first $r+1$ vertices.

Given a small $\epsilon > 0$, let m be the largest integer satisfying

$$\epsilon \leq \frac{1}{h^8 e^{10\sqrt{\log m \log h}}}.$$

It is easy to check that this m satisfies

$$m \geq (\frac{c}{\epsilon})^{c \log(c/\epsilon)}$$

for an appropriate $c = c(h) > 0$. Let $X \subset \{1, 2, \dots, m\}$ be as in Lemma 3.1. We next define a graph F from K in a way similar to the one described in the paragraph preceding Lemma 3.2. Let $V_1, V_2, \dots V_k$ be pairwise disjoint sets of vertices, where $|V_i| = im$ and we denote the vertices of V_i by $\{1, 2, \dots, im\}$. For each j, $1 \leq j \leq m$, for each $x \in X$ and for each edge $v_p v_q$ of K, let $j+(p-1)x \in V_p$ be adjacent to $j + (q-1)x \in V_q$. Note that the induced subgraph of F on the union of the first $(r+1)$ vertex classes, is precisely the graph $T(r, m)$ considered in Lemma 3.2. Finally, define

$$s = \lfloor \frac{n}{|V(F)|} \rfloor = \lfloor \frac{2n}{k(k+1)m} \rfloor$$

and let G be the s-blow-up of F (together with some isolated vertices, if needed, to make sure that the number of vertices is precisely n).

Since G consists of pairwise edge disjoint s-blow-ups of K it follows, by Lemma 3.3, that one has to delete at least a fraction of $1/h^4$ of its edges to destroy all copies of H in it. By the definition of m and the construction of X this implies, after taking the edge-density of G into account, that G is ϵ-far from being H-free.

We next claim that any copy of H in G must contain a cycle of length $r+1$ in the induced subgraph of G on the first $(r+1)$ vertex classes of it. To see this, note that there is a natural homomorphism of G onto K, obtained by first mapping G homomorphically onto F (by mapping each class of s vertices into the vertex of F to which it corresponds), and then by mapping all vertices of V_i to v_i. This homomorphism maps each copy of H in G homomorphically into K, and hence, using the discussion in the first paragraph of the proof, maps some cycle C of length $r+1$ in the copy of H considered onto the cycle on the first $r+1$ vertices of K. The definition of the homomorphism thus implies the assertion of the claim.

By Lemma 3.2 it follows that the number of such cycles is at most $m^2 s^{r+1} \leq n^{r+1}/m$, and this implies that the total number of copies of H in G does not exceed n^h/m, implying the desired result. ■

Proof of Theorem 1, part (ii): Let H be a non-bipartite graph on h vertices and suppose $\epsilon > 0$. Given a one-sided tester for testing H-freeness we may assume, without loss of generality, that it queries about all pairs of a randomly chosen set of vertices (otherwise, as explained in [2], 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. This may only square the number of queries. See also [13] for a more detailed proof of this statement.) As the algorithm is a one-sided-error algorithm, it can report that G is not H-free only if it finds a copy of H in it. By Lemma 3.4 there is a graph G on n vertices which is ϵ-far from being H-free and yet contains at most $(\epsilon/c)^{c \log(c/\epsilon)} n^h$ copies of H. The expected number of copies of H inside a randomly chosen set of x vertices in such a graph is at most $\binom{x}{h}(\epsilon/c)^{c \log(c/\epsilon)}$, which is far smaller than 1 unless x exceeds $(c'/\epsilon)^{c' \log(c'/\epsilon)}$ for some $c' = c'(H) > 0$, implying the desired result. ■

4 Concluding remarks and open problems

We have characterized all graphs H for which the property P_H of being H-free has a one-sided tester whose query complexity is polynomial in $(1/\epsilon)$. The situation for two-sided error algorithms is more complicated, and although the characterization for this case may be the same, this remains open. As mentioned at the end of Section 2, for every bipartite graph H there is a trivial (two-sided-error) algorithm for testing P_H which makes only $O(1/\epsilon)$ queries (and this number can be easily seen to be optimal, up to the multiplicative constant). Indeed, the algorithm only has to sample random edges of G and estimate if the total number of edges is $\Omega(\epsilon n^2)$. Since every graph with a quadratic number of edges contains every fixed bipartite graph, this indeed provides the required tester. On the other hand, it is easy to see that any one-sided tester for testing, say, $K_{1000,1000}$-freeness must ask far more than $O(1/\epsilon)$ queries, as there are graphs which are ϵ-far from being $K_{1000,1000}$-free and yet contain only $O(\epsilon^{10^6} n^{2000})$ copies of $K_{1000,1000}$. The problem of finding nontrivial lower bounds for the best possible query complexity of two-sided error testers for the property P_H for various graphs H seems interesting (and difficult).

It would be interesting to improve the upper bound for the query complexity of the best one-sided tester for P_H for non-bipartite graphs H. At the moment, the only known upper bound is a tower type function of $1/\epsilon$. Even the special case $H = K_3$ would be of considerable interest, because of its connection to the problem of the maximum possible density of a subset of $\{1, 2, \ldots, n\}$ with no three-term arithmetic progression. This problem received a considerable amount of attention over the years, see [19], [14], [22], [7]. A proof that any graph on n vertices which is ϵ-far from being triangle-free contains at least, say, $2^{-c/\epsilon^2} n^3$ triangles for some fixed $c > 0$ would suffice to improve the best known bound for the arithmetic progression problem.

Another intriguing problem is that of estimating the best possible (one-sided and two-sided) query complexity of the property P_H^* of not containing any **induced** copy of a fixed graph H. We can show that for certain fixed graphs H (like a star with two leaves) there are one-sided testers for P_H^* whose query complexity is polynomial in $1/\epsilon$, whereas for some other graphs H (like a star with three leaves) there are no such efficient testers. It would be interesting to study this problem further.

Acknowledgment. I would like to thank Oded Goldreich and Michael Krivelevich for many helpful discussions.

References

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

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Proc. 40^{th} Annual Symp. on Foundations of Computer Science (FOCS), New York, NY, IEEE (1999), 656–666. Also: Combinatorica 20 (2000), 451-476.

[3] N. Alon, M. Krivelevich, I. Newman and M. Szegedy, Regular languages are testable with a constant number of queries, Proc. 40^{th} Annual Symp. on Foundations of Computer Science (FOCS), New York, NY, IEEE (1999), 645–655. Also: SIAM J. on Computing 30 (2001), 1842-1862.

[4] N. Alon and M. Krivelevich, *Testing k-colorability*, SIAM J. Discrete Math., to appear.

[5] 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.

[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] J. Bourgain, On triples in arithmetic progressions, Geom. Funct. Anal. 9 (1999), 968-984.

[8] A. Czumaj and C. Sohler, Testing hypergraph coloring, Proc. ICALP 2001, to appear.

[9] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica*, 19, (1999), 175-220.

[10] , O. Goldreich, S. Goldwasser, E. Lehman, D. Ron and A. Samorodnitsky, Testing monotonicity, Combinatorica 20 (2000), 301-337.

[11] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *Proceedings of the* 37^{th} *Annual IEEE FOCS* (1996), 339–348. Also: Journal of the ACM 45 (1998), 653–750.

[12] O. Goldreich and D. Ron, Property testing in bounded degree graphs, Proc. 29^{th} STOC (1997), 406-415.

[13] O. Goldreich and L. Trevisan, Three theorems regarding testing graph properties, Proc. 42^{nd} IEEE FOCS, IEEE (2001), to appear.

[14] D. R. Heath-Brown, Integer sets containing no arithmetic progressions, *J. London Math. Soc.* 35 (1987), 385-394.

[15] P. Hell and J. Nesetril, The core of a graph, Discrete Math 109 (1992), 117-126.

[16] T. Kővári, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.*, 3, (1954), 50-57.

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

[18] D. Ron, Property testing, to appear in P. M. Pardalos, S. Rajasekaran, J. Reif and J. D. P. Rolim, editors, *Handbook of Randomized Algorithms*, Kluwer Academic Publishers, 2001.

[19] K. Roth, On certain sets of integers, *J. London Math. Soc.* 28 (1953), 104-109.

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

[21] 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.

[22] E. Szemerédi, Integer sets containing no arithmetic progressions, *Acta Math. Acad. Sci. Hungar.* 56 (1990), 155-158.