SIAM J. COMPUT.
Vol. 38, No. 6, pp. 2499–2510

TESTING HEREDITARY PROPERTIES OF NONEXPANDING BOUNDED-DEGREE GRAPHS*

ARTUR CZUMAJ†, ASAF SHAPIRA‡, AND CHRISTIAN SOHLER§

Abstract. We study graph properties that are testable for *bounded-degree graphs* in time independent of the input size. Our goal is to distinguish between graphs having a predetermined graph property and graphs that are far from every graph having that property. It is well known that in the bounded-degree graph model (where two graphs are considered "far" if they differ in εn edges for a positive constant ε), many graph properties cannot be tested even with a constant or even with a polylogarithmic number of queries. Therefore in this paper we focus our attention on testing graph properties for special classes of graphs. Specifically, we show that every hereditary graph property is testable with a constant number of queries *provided that every sufficiently large induced subgraph of the input graph has poor expansion*. This result implies that, for example, any hereditary property (e.g., k-colorability, H-freeness, etc.) is testable in the bounded-degree graph model for planar graphs, graphs with bounded genus, interval graphs, etc. No such results have been known before, and prior to our work, very few graph properties have been known to be testable with a constant number of queries for general graph classes in the bounded-degree graph model.

Key words. property testing, hereditary graph properties, bounded-degree graphs, planar graphs, nonexpanding graphs, randomized algorithms, approximation algorithms

AMS subject classifications. 68W20, 68W25, 68W40, 68Q25

DOI. 10.1137/070681831

1. Introduction.

The area of *property testing* deals with the problem of distinguishing between two cases: that an input object (for example, a graph, a function, or a point set) satisfies a certain predetermined property (for example, being bipartite, monotone, or in convex position) or is "far" from satisfying the property. Loosely speaking, an object is ε-far from having a property Π if it differs in an ε-fraction of its description from any object having the property Π. For example, when the object is a (dense) graph represented by an $n \times n$ adjacency matrix and the property is bipartiteness, then a graph is ε-far from bipartite if one has to delete more than $\varepsilon\, n^2$ edges to make it bipartite.

Many objects and properties are known to have randomized property testing algorithms whose time complexity is *sublinear* in the input description size; often, we can even achieve running time completely independent of the input size. In particular, sublinear-time property testing algorithms have been designed for graphs and hypergraphs, functions, point sets, formal languages, and many other structures (for the references, see, e.g., [9, 10, 11, 13, 17, 20]). After a series of results for specific

*Received by the editors February 5, 2007; accepted for publication (in revised form) January 12, 2009; published electronically April 10, 2009. A preliminary version of this paper, entitled "On testable properties in bounded degree graphs," authored by the first and third authors, appeared in the Proceedings of the 18th Annual Symposium on Discrete Algorithms (SODA), New Orleans, LA, 2007, pp. 494–501. This research was supported in part by NSF ITR grant CCR-0313219, Centre for Discrete Mathematics and Its Applications (DIMAP) and EPSRC grant EP/D063191/1, NSF DMS grant 0354600, and DFG grant Me 872/8-3.

http://www.siam.org/journals/sicomp/38-6/68183.html

†Centre for Discrete Mathematics and Its Applications (DIMAP) and Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom (A.Czumaj@warwick.ac.uk).

‡School of Mathematics and College of Computing, Georgia Institute of Technology, Atlanta, GA 30332 (asafico@math.gatech.edu).

§Institute of Computer Science I, University of Bonn, 53117 Bonn, Germany (sohler@informatik.uni-bonn.de).

problems, much attention has been devoted recently to a more general question: which properties can be tested in time independent of the input size? This question has been especially extensively investigated for properties of dense graphs represented by an adjacency matrix. It turned out that property testing in dense graphs is closely related to Szemerédi's regularity lemma. Very recently, this relation has been made explicit by showing that any property is testable if and only if it can be reduced to testing the property of satisfying a finite number of Szemerédi-partitions (see [1]). Furthermore, it has been shown in [4] that a (natural) graph property is testable with one-sided error if and only if it is either hereditary or it is close (in some well-defined sense) to a hereditary property (see also [8] for an alternative proof). These results, imply that in the adjacency matrix model, essentially any "natural" graph property can be tested with a number of queries independent of the size of the graph.

While property testing in dense graphs is relatively well understood, surprisingly little is known about property testing in sparse graphs. Properties of sparse graphs are traditionally studied in the model of *bounded-degree graphs* introduced by Goldreich and Ron [15]. In this model, the input graph G is represented by its *adjacency list* (or, equivalently, by its incidence list) and the vertex degrees are bounded by a constant d independent of the number of vertices of G (denoted by n). A testing algorithm has a constant-time access to any entry in the adjacency list by making a query to the ith neighbor of a given vertex v, and the number of accesses to the adjacency list is the query complexity of the tester. A property testing algorithm is an algorithm that for a given graph G determines if it satisfies a predetermined property Π or it is ε-far from property Π; a graph G is *ε-far* from property Π if one has to modify more than $\varepsilon\, d\, n$ edges in G to obtain a graph having property Π.

Unlike the adjacency matrix model discussed above, in the bounded-degree graph model only a few graph properties are known to be testable in constant time; see [15], where it is shown that k-edge-connectivity, H-freeness, and some other properties are testable with a constant number of queries. The study of testing bounded-degree graphs thus focused on designing property testers with a sublinear query complexity (like, $\widetilde{\mathcal{O}}(\sqrt{n})$ tester for bipartiteness [14]). Even more, it has been demonstrated that unlike in the adjacency matrix model, in the bounded-degree model many basic properties have a nonconstant query complexity. For example, acyclicity in directed graphs has $\Omega(n^{1/3})$ query complexity [6], the property of being bipartite has query complexity $\Omega(\sqrt{n})$ [15], and the query complexity of testing 3-colorability is $\Omega(n)$ [7].

In this paper, we take a new approach and study property testing in the bounded-degree model under the assumption that the input graph belongs to a certain (natural) family of graphs. The goal of this investigation is to identify natural families of graphs, such as planar graphs, for which many properties can be efficiently tested under the assumption that the input graph belongs to the family, even though the testing problem may be very hard in the general case.

For the rest of this paper, we say that a graph property is *testable* if it can be tested in time independent of the size of the input graph. A graph $G = (V, E)$ is said to have an *expansion* α if for every subset of vertices $U \subseteq V$ with $|U| \leq |V|/2$, the number of neighbors of U in $V \setminus U$ is at least $\alpha|U|$. A graph $G = (V, E)$ is called *nonexpanding* if it has expansion smaller than $1/\log^2 n$. (This is informally equivalent to the families of graphs with some good separator properties.) A graph $G = (V, E)$ is called *C-strongly nonexpanding* if every induced subgraph with at least C vertices is nonexpanding. A graph property is called *hereditary* if it is closed under vertex removal. We show the following result:

In the bounded-degree graph model, for every constant C, every hereditary graph property Π is testable for the family of C-strongly nonexpanding graphs.

The reader is referred to Theorem 3.6 from section 3.3 for the precise statement of our main result.

We note that there is no function of ε that uniformly bounds the number of queries that our algorithm performs. In other words, the number of queries can be an arbitrary function of ε. This is due to a certain graph functional that we define and use in section 3.2. We stress that this graph functional is used in order to allow us to handle arbitrary hereditary properties and that for most natural properties the running time can be bounded by a uniform function of ε. However, this function of ε is at least doubly exponential in $1/\varepsilon$ so there is still a lot of room for improvements.

Hereditary graph properties have been extensively investigated in combinatorics, graph theory, and theoretical computer science (see also the recent results about testability of hereditary graph properties in the dense graph model [4]). The class of hereditary graph properties also contains all *monotone graph properties* (properties closed under removal of edges and vertices). Many interesting graph properties are hereditary, for example, being acyclic, stable (independent set), planar, bipartite, k-colorable, chordal, perfect, interval, permutation, having no induced subgraph H, etc. (For the definitions, see the appendix; for more discussion on hereditary graph properties, see [16, 19].) Our result implies that these properties can be tested (in the bounded-degree graph model) when the input graph belongs to a family of graphs that is C-strongly nonexpanding for certain constant C. Examples of natural C-strongly nonexpanding families of bounded-degree graphs are planar graphs, graphs with bounded genus, graphs with forbidden minors, unit disk graphs, interval graphs, (planar) geometric intersection graphs, etc. We are not aware of any prior results showing testability of the above properties for nontrivial or bounded-degree graphs.

2. Preliminaries.

Let $G = (V, E)$ be an undirected graph with n vertices and maximum degree at most d. Without loss of generality, we assume that $V = \{1, \dots, n\}$. We write $[n] := \{1, \dots, n\}$. Given a subset $S \subseteq V$ of vertices, we use $G_{|S} = (S, E_{|S})$ to denote the subgraph induced by S, where $E_{|S} = \{(u, v) \in E \cap (S \times S)\}$. We assume that G is stored in the *adjacency list* model for bounded-degree graphs with maximum degree d. In this model, we have constant-time access to a function $f_G : [n] \times [d] \to [n] \cup \{+\}$, such that $f_G(v, i)$ denotes the ith neighbor of v or a special symbol $+$ in the case that v has less than i neighbors.

DEFINITION 2.1. *A graph G is* ε-far from a property Π *if one has to modify more than $\varepsilon\, d\, n$ entries in f_G to obtain a graph with property* Π.

2.1. Testing a property in a graph family.

In this paper, our main focus is on testing various graph properties for bounded-degree graphs from certain graph families (e.g., planar graphs or unit disk graphs).

An algorithm that is given n and has access to f_G is called an *ε-tester for a graph family $\mathcal{F}$* if it

(a) accepts with probability at least $2/3$ any graph $G \in \mathcal{F}$ that has property Π, and

(b) rejects with probability at least $2/3$ any graph $G \in \mathcal{F}$ that is ε-far from Π.

If the ε-tester always accepts any graph $G \in \mathcal{F}$ that has property Π, then it is said to have *one-sided error*. The ε-testers presented in this paper have one-sided error. They will in fact accept with probability 1 any graph that satisfies Π (even if it does not belong to $\mathcal{F}$).

A property is called *testable* for a family $\mathcal{F}$ if for any fixed $0 < \varepsilon < 1$ there is an ε-tester for $\mathcal{F}$ whose total number of queries to the function f_G is bounded from above by a function, which depends only on ε and not on the size n of the input graph. Following [3], we define a property Π to be *uniformly testable* if there is an ε-tester for Π that receives ε as part of the input. A property Π is said to be *nonuniformly testable* if for every fixed ε, $0 < \varepsilon < 1$, there is an ε-tester that can distinguish between graphs that have property Π from those ε-far from having Π (which may not work properly for other values of ε).

For a pair of disjoint vertex sets V_1, V_2 we denote by $e(V_1, V_2)$ the number of edges connecting vertices from V_1 with vertices from V_2. For each vertex $v \in V$, we denote its *neighborhood* by $\mathcal{N}(v) = \{u \in V : (v, u) \in E\}$. We generalize this notion to sets by defining $\mathcal{N}(S) = \bigcup_{v \in S} \mathcal{N}(v) \setminus S$. Furthermore, we let $D(v, r)$ denote the set of vertices which have distance at most r from v, i.e., $D(v, 0) = v$, $D(v, 1) = \{v\} \cup \mathcal{N}(v)$, etc.

A graph $G = (V, E)$ is called a λ-*expander* if for all $S \subseteq V$ with $|S| \leq n/2$, we have $|\mathcal{N}(S)| \geq \lambda |S|$.

Definition 2.2. *A graph $G = (V, E)$ is called* nonexpanding *if G is not a $(1/\log^2 n)$-expander.*[1]

*A graph $G = (V, E)$ is called C-*strongly nonexpanding *if every induced subgraph of G with at least C vertices is nonexpanding.*

In this paper we will consider C-strongly nonexpanding only for a constant value of C.

2.2. C-strongly nonexpanding graph families. There are many interesting classes of families of graphs that are C-strongly nonexpanding for some constant C. For example, the classical planar separator theorem [18] implies immediately that any *planar graph* is C-strongly nonexpanding for some constant C. Indeed, the planar separator theorem implies that every planar graph with n vertices (for a sufficiently large n) has a subset of vertices A, $\frac{1}{3} n \leq |A| \leq \frac{1}{2} n$, such that $|\mathcal{N}(A)| \leq 4\sqrt{n}$. Therefore, every planar graph with n vertices ($n \geq n_0$ for some constant n_0) is not a $\frac{6}{\sqrt{n}}$-expander, and hence the family of planar graphs is C-strongly nonexpanding for some constant C. As the example of planar graphs shows, all graphs with good separator properties (for graphs of bounded-degree) are C-strongly nonexpanding. Hence, other families of graphs (of bounded-degree) that are C-strongly nonexpanding include, among others: the class of graphs with bounded genus, graphs with forbidden minor, interval graphs, etc. For example, the result for graphs of bounded genus and graphs with forbidden minor follow directly from the separator theorem for such graphs. Gilbert, Hutchinson, and Tarjan [12] proved that any graph on n vertices with genus g has a separator of order $\mathcal{O}(\sqrt{gn})$, and Alon, Seymour, and Thomas [2] showed a similar result for graphs with forbidden minors: if G has no minor isomorphic to a given h-vertex graph H, then G has a separator of size $\mathcal{O}(h^{3/2} n^{1/2})$.

3. Proof of the main result. In this section we prove our main result by showing that the following algorithm is an ε-tester for any hereditary property Π and any family $\mathcal{F}$ of C-strongly nonexpanding graphs.

[1] The choice of the factor $1/\log^2 n$ can be relaxed. In fact, using a slightly more precise calculation in our analysis, one can replace $1/\log^2 n$ with $1/(\log n \log^2 \log n)$.

```
ε-Tester(G, n, Π)
        sample a set S of s_1 vertices uniformly at random
        for each v ∈ S do
            U_v = D(v, s_2)
        U = ⋃_{v∈S} U_v
        if G_{|U} does not satisfy property Π then reject
                                                    else accept
```

Observe that in a graph of maximum degree d, the size of $D(v, s_2)$ is at most $1 + d\sum_{i=0}^{s_2-1}(d-1)^i \le d^{s_2}$ for $s_2 \ge 3$. Therefore, since the algorithm has to query all edges incident to U, the number of queries to f_G is upper bounded by $s_1\, d^{1+s_2}$, which, for s_1 and s_2 independent of n, gives the number of queries independent of n. We will give the exact values for s_1 and s_2, which are independent of n but do depend on ε, $\mathcal{F}$, and Π, at the end of our analysis, in the proof of Theorem 3.6.

Since Π is hereditary, we know that our algorithm accepts any graph that has property Π (even if it does not belong to $\mathcal{F}$). Thus, we only have to show that any graph that is ε-far from Π and belongs to $\mathcal{F}$ is rejected with probability at least $\frac{2}{3}$.

We begin our analysis with the following lemma.

LEMMA 3.1. *Let $G = (V, E)$ be a C-strongly nonexpanding graph of maximum degree d. Let δ be an arbitrary positive parameter. If $n = |V| \ge \max\{2C, 2^{2/\delta^2}\}$, then one can partition V into two sets V_1 and V_2, such that $|V_1|, |V_2| \ge \frac{n}{4}$ and $e(V_1, V_2) \le \delta\, d\, n/\log^{1.5} n$.*

Proof. Since G is C-strongly nonexpanding, there exists a set $S \subseteq V$ of cardinality at most $\frac{n}{2}$ such that $|\mathcal{N}(S)| \le |S|/\log^2 n$. We first observe that if $|S| \ge \frac{n}{4}$, then we can take $V_1 = S$ and $V_2 = V \setminus S$. Indeed, since $|\mathcal{N}(S)| \le |S|/\log^2 n$, there are at most $dn/\log^2 n$ edges between V_1 and V_2. Therefore, if in addition $n \ge 2^{2/\delta^2}$, we can infer that

$$e(V_1, V_2) \le dn/\log^2 n \le \delta\, d\, n/\log^{1.5} n,$$

as needed.

Assume then that $|S| < \frac{n}{4}$ and consider the graph $G_{|V\setminus S}$ (the induced graph on $V \setminus S$). Since G is C-strongly nonexpanding and $|V \setminus S| > C$ (recall that $n > 2C$), we can apply the same arguments as above to conclude that there is a set $S' \subseteq (V \setminus S)$ of cardinality at most $\frac{n}{2}$ such that $|\mathcal{N}(S')| \le 2|S'|/\log^2 n$. If we have $|S \cup S'| \ge \frac{n}{4}$, then using the same arguments as above we are done by setting $V_1 = S \cup S'$ and $V_2 = V \setminus V_1$. Otherwise, we can replace S by $S \cup S'$ and continue in the same manner. Eventually, we have a set $S \cup S'$ with more than $\frac{n}{4}$ vertices and $|\mathcal{N}(S \cup S')| \le 2\,|S \cup S'|/\log^2 n$. If we set $V_1 = S \cup S'$ and $V_2 = V \setminus V_1$, then these sets will satisfy the condition in the lemma. ∎

Let us call a connected component *nontrivial* if it has more than a single vertex. The following is a corollary of Lemma 3.1.

COROLLARY 3.2. *For every C-strongly nonexpanding graph $G = (V, E)$ there exists a positive constant $c = c_C$, such that one can remove from G a set of at most $\varepsilon\, d\, n/2$ edges, such that*

(i) *their removal partitions G into connected components $C_1, C_2, \ldots$ of size at most $2^{c/\varepsilon^2}$ each,*
(ii) *each connected component C_i is an induced subgraph of G, and*
(iii) *no edge connects in G two nontrivial connected components C_i and C_j.*

Proof. Let $G = (V, E)$ be a C-strongly nonexpanding graph and let δ be a parameter to be chosen later. We apply Lemma 3.1 to obtain two sets V_1 and V_2 with at most $\delta\, d\, n/\log^{1.5} n$ edges connecting V_1 and V_2. Assume $|V_1| \le |V_2|$ and let $U^* = \mathcal{N}(V_1)$. Since the number of edges between V_1 and $V \setminus V_1$ is at most $\delta\, d\, n/\log^{1.5} n$, we also have $|U^*| \le \delta\, d\, n/\log^{1.5} n$. Remove from G all edges incident to U^*. Since $|U^*| \le \delta\, d\, n/\log^{1.5} n$ and G has maximum degree at most d, we removed at most $\delta\, d^2\, n/\log^{1.5} n$ edges from G. Next, let $U_1 = V_1$ and $U_2 = V_2 \setminus U^*$. Observe that for $\delta \le \log^{1.5} n/(4d)$ we have $\frac{n}{4} \le |U_1|, |U_2| \le \frac{3n}{4}$ and that there is no edge in G between U_1 and U_2.

Then we recursively apply Lemma 3.1 on the induced subgraphs $G_{|U_1}$ and $G_{|U_2}$; we proceed recursively until we obtain a subgraph of size at most $\max\{2C, 2^{2/\delta^2}\}$. In this way, we removed some number of edges from G and obtained a subgraph of G, denoted H, on $V(G)$ with connected components $C_1, \ldots, C_q$. Observe that the sets U^* obtained in the recursive calls will always result in trivial connected components, because we removed all edges incident to the vertices in U^*. Let $H_1, \ldots, H_k$ be nontrivial connected components in our new graph H. By definition, every C_i has size $|C_i| \le \max\{2C, 2^{2/\delta^2}\}$. Similarly, our construction ensures that no edge is removed between any pair of vertices in a single H_i and that there is no edge in G between any pair of graphs H_i and H_j. We now estimate the number of edges removed.

By Lemma 3.1, the number of edges removed from G is upper bounded by function $Q(n)$ defined by the following recurrence:

$$Q(n) = \begin{cases} 0 & \text{if } n \le \max\{2C, 2^{2/\delta^2}\}, \\ \delta d^2 n/\log^{1.5} n + \max_{\frac{1}{4} \le \tau \le \frac{3}{4}}\{Q(\tau\, n) + Q((1-\tau)\, n)\} & \text{if } n > \max\{2C, 2^{2/\delta^2}\}. \end{cases}$$

Since $Q(n) = O(\delta\, d^2\, n)$, we can conclude that the graph H is obtained from G by removal of at most $c'\,\delta\, d^2\, n$ edges, for some absolute positive constant $c' \ge 1$. This yields the proof by setting $\delta = \varepsilon/(2dc')$. Finally, recall that all the connected components of H had size $|C_i| \le \max\{2C, 2^{2/\delta^2}\} \le 2^{c/\varepsilon^2}$ if we took $c = c_G = 2\, d\, c'\, C$. ∎

Let us explain the importance of the three properties of the resulting graph stated in Corollary 3.2. Property (i) ensures that every connected component is small. Property (ii) ensures that if we have some induced subgraph of a nontrivial connected component H_i, then it is also an induced subgraph of G. Property (iii) ensures that if we have a set of induced subgraphs $\Lambda_{i_1}, \Lambda_{i_2}, \ldots, \Lambda_{i_\ell}$ of graphs $H_{i_1}, H_{i_2}, \ldots, H_{i_\ell}$, then these copies of the subgraphs do not intersect in H. Therefore, if we define a graph $\widehat{\Lambda}$ with ℓ connected components, where the jth connected of $\widehat{\Lambda}$ is isomorphic with Λ_{i_j}, then $\widehat{\Lambda}$ is also an induced subgraph of G.

3.1. Hereditary graph properties. It is well known (and easy to see) that any hereditary graph property Π can be characterized by a (possibly infinite) set of minimal forbidden induced subgraphs (see, e.g., [4, section 4]). Let us denote by $\mathcal{H}^{\Pi}_{forb}$ a *minimal* family of forbidden subgraphs for property Π. Notice that in general $\mathcal{H}^{\Pi}_{forb}$ may be an *infinite* family of forbidden graphs. For example, if Π is the property of being bipartite, then $\mathcal{H}^{\Pi}_{forb}$ can be chosen to be the set of all odd cycles, and if Π is the property of being chordal, then $\mathcal{H}^{\Pi}_{forb}$ is the set of all cycles of length at least 4.

Next, let us consider an arbitrary C-strongly nonexpanding graph G that is ε-far from Π. By Corollary 3.2, we can remove from G at most $\varepsilon\, d\, n/2$ edges to obtain a graph H on the same vertex set for which each connected component has at most $r = 2^{c/\varepsilon^2}$ vertices. Furthermore, if $H_1, \ldots, H_k$ are the nontrivial connected components of H, then there is no edge in G that connects any of these connected components and each H_i is an induced subgraph of G. Since G is ε-far from Π, H is still $\varepsilon/2$-far from Π. Since all connected components in H have size at most r (which is independent of n), H cannot contain as a subgraph any graph that has a connected component with more than r vertices. Let $\mathcal{J}_r$ denote the family of all graphs whose connected components have size at most r (notice that $\mathcal{J}_r$ is independent of G). We conclude that it suffices to consider the subgraphs in $\mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$.

COROLLARY 3.3. *If a C-strongly nonexpanding graph G is ε-far from Π, then the graph H defined above contains as an induced subgraph a graph from $\mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$. The same holds if we remove from H any set of at most $\varepsilon\, d\, n/2$ edges.*

Let us denote by $\mathfrak{c}(r)$ the number of *connected* (unlabeled) graphs on a set of at most r vertices; clearly $\mathfrak{c}(r) \le 2^{\binom{r}{2}}$. Let us enumerate all possible connected graphs with at least one and at most r vertices by $\mathfrak{G}_1, \ldots, \mathfrak{G}_{\mathfrak{c}(r)}$. Without loss of generality, let $\mathfrak{G}_1$ denote the subgraph that consists of a single vertex. Then, we can define any graph $\mathcal{G}$ in $\mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$ as an integer vector $f = \langle f_1, \ldots, f_{\mathfrak{c}(r)} \rangle$ of length $\mathfrak{c}(r)$, where f_i denotes the number of copies of graph $\mathfrak{G}_i$ occurring as a connected component in $\mathcal{G}$.

In what follows, we call f a *characteristic vector* of $\mathcal{G}$ (with respect to $\mathcal{H}^{\Pi}_{forb}$ and $\mathcal{J}_r$). Similarly, for a graph H whose all connected components are of size at most r, let us define an integer vector $\mathfrak{g}^{\langle H \rangle} = \langle \mathfrak{g}^{\langle H \rangle}_1, \ldots, \mathfrak{g}^{\langle H \rangle}_{\mathfrak{c}(r)} \rangle$ of length $\mathfrak{c}(r)$, with $\mathfrak{g}^{\langle H \rangle}_i$ being the number of *connected components* of H that are isomorphic to $\mathfrak{G}_i$.

LEMMA 3.4. *Let Π be a fixed hereditary property. Let G be a C-strongly nonexpanding graph of degree at most d that is ε-far from Π. Assume that we apply Corollary 3.2 on G and obtain a subgraph of G, denoted by H, with the property that all connected components of H are of size at most r. Then, there exists a graph $\mathcal{G} \in \mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$ with characteristic vector $f = \langle f_1, \ldots, f_{\mathfrak{c}(r)} \rangle$ such that for all $2 \le i \le \mathfrak{c}(r)$ it holds that if $f_i > 0$, then $\mathfrak{g}^{\langle H \rangle}_i \ge \gamma\, n$, where $\gamma = \varepsilon \cdot d/2^{r^2}$.*

Proof. Let $\mathfrak{G}_2, \ldots, \mathfrak{G}_{\mathfrak{c}(r)}$ be all connected graphs with at least 2 and at most r vertices. We will first construct a graph H' by removing some edges from H so that for any graph $\mathfrak{G}_i$, $i > 1$, either H' contains no connected component isomorphic to $\mathfrak{G}_i$ or it contains at least $\gamma\, n$ such components. We proceed sequentially over the graphs $\mathfrak{G}_2, \ldots, \mathfrak{G}_{\mathfrak{c}(r)}$. For each $\mathfrak{G}_i$ we do the following: if the number of connected components in the current graph obtained from H is smaller than $\gamma\, n$, we remove all the edges of any connected component that is isomorphic to $\mathfrak{G}_i$. Since we perform at most $\mathfrak{c}(r)$ iterations and in each iteration we remove at most $\binom{r}{2} \cdot \gamma\, n$ edges, the total number of edges removed is bounded by $\mathfrak{c}(r) \cdot \binom{r}{2} \cdot \gamma n < \varepsilon\, d\, n/2$ by our choice of γ. At the end of the process we obtain a graph H' with the property that for any graph $\mathfrak{G}_i$ either H' contains no connected component isomorphic $\mathfrak{G}_i$ or it contains at least $\gamma\, n$ such components. Observe that, when removing the edges of a component isomorphic to $\mathfrak{G}_i$, we do not change the number of components of the graph that are isomorphic to another $\mathfrak{G}_j$ for $j > 1$.

Since G was assumed to be ε-far from Π, and H was obtained from G by removing at most $\varepsilon\, d\, n/2$ edges, we have that H is $\frac{\varepsilon}{2}$-far from Π. Also, since H' is obtained from H by removing less than $\varepsilon\, d\, n/2$ edges, H' does not satisfy Π, and hence it contains a graph $\mathcal{G} \in \mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$. Now, by the conclusion of the previous paragraph, this means that if $\mathcal{G}$ has characteristic vector $\langle f_1, \ldots, f_{\mathfrak{c}(r)} \rangle$, then for every $i > 1$ for which $f_i > 0$ we must have that H' contains at least $\gamma\, n$ connected components that are isomorphic to $\mathfrak{G}_i$. Finally, observe that from the definition of the process of obtaining H' it follows that H must contain at least this many connected components that are isomorphic to $\mathfrak{G}_i$. Hence, for every $i > 1$ for which $f_i > 0$ we have $\mathfrak{g}_i^{\langle H \rangle} \geq \gamma n$. ◻

3.2. The function Ψ_Π. We now introduce a key notion that we will use to test a hereditary property Π. Given a family of pairwise nonisomorphic connected graphs $\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}$, let $m(\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\})$ be the least integer m with the property that the graph that contains m vertex disjoint, disconnected copies of each of the graphs $\mathfrak{G}_i^*$ does not satisfy Π. If no such integer m exists, then we set $m(\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}) = \infty$. For an integer r, let Π_r be the family of all sets of pairwise nonisomorphic connected graphs $\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}$ with the property that all the graphs $\mathfrak{G}_i^*$ are of size at most r and $m(\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}) < \infty$.

Definition 3.5. *For a fixed hereditary property* Π *we define a function* $\Psi_\Pi : \mathbb{N} \mapsto \mathbb{N}$ *as follows:*

$$\Psi_\Pi(r) = \max_{\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\} \in \Pi_r} m(\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}).$$

In case $\Pi_r = \emptyset$ *we set* $\Psi_\Pi(r) = 0$.

Note that the above is well defined as for a fixed integer r the set Π_r is finite.

3.3. Proof of the main theorem. We now formally state and prove the main result of this paper.

Theorem 3.6. *Let $\mathcal{F}$ be any family of C-strongly nonexpanding graphs, where C is an arbitrary constant. Then every hereditary graph property Π is nonuniformly testable for $\mathcal{F}$ with one-sided error. Furthermore, Π is uniformly testable with one-sided error if ψ_Π is computable (or if its approximation is computable, where the quality of the approximation must be independent of the input graph size).*

Proof. Clearly, our tester accepts every graph that has property Π. So, suppose that $G \in \mathcal{F}$ is ε-far from Π, and consider the subgraph H of G that is obtained via Corollary 3.2. By Lemma 3.4, there is a subgraph $\mathcal{G}$ of H (and so of G) that does not satisfy Π and has the property that (i) all its connected components $\mathfrak{G}_i$ are of size at most $r = 2^{c/\varepsilon^2}$ and (ii) except for isolated vertices, each of these connected components appears as a connected component of H at least γn times, where $\gamma = \varepsilon d / 2^{r^2}$.

Consider now the set of nonisomorphic connected components of $\mathcal{G}$ other than single vertices, and denote them as $\{\mathfrak{G}_1^*, \ldots, \mathfrak{G}_k^*\}$. We first show that our sample set contains with probability at least 2/3 at least $\Psi_\Pi(r)$ vertex disjoint copies of each of these components.

By the first paragraph of the proof, a randomly chosen vertex belongs to a connected component of $\mathcal{G}$ that is isomorphic to $\mathfrak{G}_i^*$ with probability at least γ. We can assume that $\Psi_\Pi(r) \leq \gamma n/2$, for otherwise the whole graph has $O(\Psi_\Pi(r)/\gamma)$ vertices and we can look at it completely. We will consider the process of sampling $|S|$ vertices independently and uniformly from V one after another. As long as our current sample set intersects less than $\Psi_\Pi(r)$ connected components isomorphic to $\mathfrak{G}_i^*$, we have a probability of at least $\gamma/2$ that the next vertex intersects a new component of this type. Let Y_j be a 0–1 random variable that is 1 with probability $\gamma/2$. Clearly, $\mathbf{E}[Y_j] = \gamma/2$ and $\mathbf{Var}[Y_j] \leq \mathbf{E}[Y_j^2] \leq \gamma/2$. Since the random variables are pairwise independent, we have $\mathbf{Var}[\sum_{j=1}^{|S|} Y_j] \leq \gamma|S|/2$. Therefore, by Chebyshev inequality we get

$$\mathbf{Pr}\left[\left|\sum_{j=1}^{|S|} Y_j - \mathbf{E}\left[\sum_{j=1}^{|S|} Y_j\right]\right| \ge \frac{1}{2}\mathbf{E}\left[\sum_{j=1}^{|S|} Y_j\right]\right] \le \frac{16 \cdot \mathbf{Var}\left[\sum_{j=1}^{|S|} Y_j\right]}{|S|^2\gamma^2} \le \frac{8}{\gamma|S|}.$$

Choosing $|S| \ge 24 \cdot \mathfrak{c}(r) \cdot \Psi_\Pi(r)/\gamma$, we obtain that with probability at most $\frac{1}{3\mathfrak{c}(r)}$ the sample set intersects fewer than $\Psi_\Pi(r)$ connected components isomorphic to $\mathfrak{G}_i^*$. Therefore, by the union bound the sample set S contains $k \cdot \Psi_\Pi(r)$ vertices $\{v_{i,j}\}_{1\le i\le k}^{1\le j\le \Psi_\Pi(r)}$ that belong to a distinct connected component of $\mathcal{G}$, with the property that for every $1 \le j \le \Psi_\Pi(r)$, the connected component of $\mathcal{G}$ to which $v_{i,j}$ belongs is an induced copy of $\mathfrak{G}_i^*$.

To finish the proof we have to consider three cases.

Case (i): *$\mathcal{G}$ does not contain isolated vertices as connected components.* Since $\mathcal{G} \notin \Pi$ and since the graph has no connected components consisting of single vertices, we have that $m(\{\mathfrak{G}_1^*, \dots, \mathfrak{G}_k^*\}) < \infty$ (cf. section 3.2). Now the definition of Ψ_Π guarantees that the graph obtained by taking $\Psi_\Pi(r)$ vertex disjoint copies of each of the graphs $\mathfrak{G}_i^*$ does not satisfy Π. By the above discussion, we have that for each $1 \le i \le k$, S contains vertices $v_{i,j}$ from $\Psi_\Pi(r)$ vertex disjoint connected components isomorphic to $\mathfrak{G}_i^*$. Choosing parameter $s_2 = r = 2^{c/\varepsilon^2}$ we know that graph $G_{|U}$ contains each connected component containing one of the vertices $v_{i,j}$ as an induced subgraph. Since G does not contain edges connecting vertices from distinct nontrivial connected components of $\mathcal{G}$ (here we rely on Corollary 3.2(iii) and the fact that $\mathcal{G}$ is a subgraph of H), we know that $G_{|U}$ also contains the union of the connected components containing the $v_{i,j}$ as an induced subgraph. Therefore, with probability at least $2/3$ the tester will reject G.

Case (ii): *$\mathcal{G}$ contains isolated vertices and other connected components.* In this case we get that $m(\{\mathfrak{G}_1^*, \dots, \mathfrak{G}_k^*\}) < \infty$, because isolated vertices are a vertex induced subgraph of every other graph. Thus, we can apply Case (i).

Case (iii): *$\mathcal{G}$ contains only isolated vertices.* In this case, we only have to show that $|S|$ contains an independent set in G of size at least $\Psi_\Pi(r)$. However, since G has maximum degree d, it contains an independent set of size $n/(d+1)$. Assuming that $\Psi_\Pi(r) \le n/(2(d+1))$, we can apply Chebyshev inequality as in the proof above and obtain that, for $|S| \ge 24(d+1) \cdot \Psi_\Pi(r)$, with probability at least $2/3$ the sample set S contains an independent set of size at least $\Psi_\Pi(r)$ and so the property tester rejects.

Therefore, choosing $s_1 = 24 \cdot \mathfrak{c}(r) \cdot \Psi_\Pi(r)/\gamma + 24(d+1) \cdot \Psi_\Pi(r)$ and $s_2 = r = 2^{c/\varepsilon^2}$ guarantees that our algorithm is a property tester. ∎

3.4. Discussion. *When do we need* Ψ_Π? Notice that the function Ψ_Π defined in section 3.2 is not necessarily computable. However, we only need this definition in order to obtain a general result on all hereditary properties. Observe, for example, that for any hereditary property Π that is closed under disjoint union[2] we have that $\Psi_\Pi(r) = 1$. Therefore, in these cases we have a trivial function Ψ. Furthermore, notice that any natural hereditary property, such as those discussed throughout the paper, is closed under disjoint union; therefore for such properties we get uniform testers (for any family of strongly nonexpanding graphs $\mathcal{F}$).

When does Π *have a uniform tester*? The proof of Theorem 3.6 shows that when the function Ψ_Π is computable then one can design a one-sided error uniform tester

[2]I.e., if $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ satisfy the property, then so does $G_3 = (V_1 \cup V_2, E_1 \cup E_2)$.

for Π. Using arguments similar to those used in [5], it can be shown that if the tester is allowed to use the size of the input in order to make its decisions, then all hereditary properties have a uniform tester with constant query complexity but with running time that depends on n. Following [5], let us define an *oblivious* tester as one that has no access to the size of the input when making its decisions. Given ε, an oblivious tester computes a number $q = Q(\varepsilon)$ and then asks an oracle for $D(v, q)$ for all the vertices $v \in S$, where S is a random subset of vertices of $V(G)$ of size q (recall that $D(v, q)$ is the neighborhood of v of radius q). Using the answers to these queries the tester should either accept or reject the input. Observe that the algorithm we design in the proof of Theorem 3.6 is oblivious. Therefore, if Ψ_Π is computable, then Π has an oblivious one-sided error uniform tester.

Let us show that for any hereditary property Π, the computability of Ψ_Π is not only sufficient but also necessary if one wants to design an oblivious one-sided error tester for Π. Here is a sketch of the proof. It is easy to see that an oblivious one-sided error tester for a hereditary property must accept the input if the graph that is spanned by $\bigcup_{v\in S} D(v, q)$ satisfies the property.[3] Suppose then that Π can be tested with query complexity $Q(\varepsilon)$. We claim that in this case $\Psi_\Pi(r) \le Q(1/2^{r^2})$, and since Q is assumed to be computable, then so does Ψ_Π. Indeed, for any $\{\mathfrak{G}_1, \dots, \mathfrak{G}_k\} \in \Pi_r$ and for any positive integer d, consider a graph consisting of d disjoint copies of each graph $\mathfrak{G}_i$. Let us think of this graph as consisting of d clusters C_j, where each cluster C_j contains one copy of each of the graphs $\mathfrak{G}_1, \dots, \mathfrak{G}_k$. This graph has degree bounded by r, and we claim that for all large enough d, it is $1/2^{r^2}$-far from Π. Let us denote by n the number of vertices of the graph and by m the number of vertices in each cluster C_i, and observe that $m \le r2^{\binom{r}{2}}$. Therefore, after adding/removing at most $\frac{n}{4m}$ edges, we will still have $\frac{n}{2m}$ clusters C_j which have not changed. Therefore, as $m(\{\mathfrak{G}_1, \dots, \mathfrak{G}_k\}) < \infty$ for large enough d, the new graph still does not satisfy Π. We thus conclude that for large enough d, the graph is at least $1/(4mr)$-far from satisfying Π (and $1/(4mr) \le 1/2^{r^2}$). However, since the algorithm must find a graph that does not satisfy Π, it must ask at least $m(\{\mathfrak{G}_1, \dots, \mathfrak{G}_k\})$ queries in order to succeed on such graphs. Therefore, $m(\{\mathfrak{G}_1, \dots, \mathfrak{G}_k\}) \le Q(1/2^{r^2})$ for any set $\{\mathfrak{G}_1, \dots, \mathfrak{G}_k\} \in \Pi_r$ and by the definition of Ψ_Π this means that $\Psi_\Pi \le Q(1/2^{r^2})$ as needed.

4. Conclusions. In this paper we made a first attempt to give general testability results for graphs belonging to restricted families of graphs. We showed that all hereditary graph properties are (nonuniformly) testable if the input graph is C-strongly nonexpanding for a constant C. Some interesting open questions include the following:

- Which properties can be tested for expander graphs? Which properties can be tested in $O(\sqrt{n})$ time for expander graphs?
- Which properties can be tested for strongly nonexpanding families of graphs when only the average degree of the graph is bounded?
- Which properties can be tested for *directed* graphs in sublinear time (in particular, when we can see a directed edge $\langle u, v\rangle$ only from vertex u)?

[3]Suppose the tester rejects an input even though $\bigcup_{v\in S} D(v, q)$ satisfies Π. In that case if we were to execute the tester on the graph that is defined as the disjoint union of $\{D(v, q) : v \in S\}$, it would have a nonzero probability of rejecting this graph even though it satisfies the property.

Appendix. Examples of hereditary graph properties.

For the convenience of the user, we will list now some of the well-known classes of graph properties that are known to be hereditary.

- *Being H-free:* For a fixed graph H, G is *H-free* if no subgraph of G is isomorphic to H.
- *Being induced H-free:* For a fixed graph H, G is *induced H-free* if no induced subgraph of G is isomorphic to H.
- *k-colorability:* G is *k-colorable* if its vertices can be partitioned into k sets such that no two adjacent vertices belong to the same set.
- *Perfect graphs:* G is *perfect* if for every induced subgraph H of G, the chromatic number of H equals the size of the largest clique in H [16].
- *Chordal graphs:* G is *chordal* if it contain no induced cycle of length at least 4.
- *Interval graphs:* $G = (V, E)$ (on n vertices) is an *interval graph* if there are closed intervals on the real line $I_1, I_2, \ldots, I_n$ such that $(i, j) \in E$ if and only if $I_i \cap I_j \neq \emptyset$.
- *Circular-arc graphs:* $G = (V, E)$ (on n vertices) is a *circular-arc graph* if there are closed intervals on an $I_1, I_2, \ldots, I_n$ such that $(i, j) \in E$ if and only if $I_i \cap I_j \neq \emptyset$.
- *Permutation graphs:* $G = (V, E)$ (on n vertices) is a *permutation graph* if there is a permutation σ of $\{1, 2, \ldots, n\}$ such that $(i, j) \in E$ if and only if (i, j) is an inversion under σ.
- *Comparability graphs:* G is a *comparability graph* if its edges can be oriented such that if there is a directed edge from i to j and from j to k, then there is one from i to k.
- *Asteroidal triple-free graphs:* G is *asteroidal triple-free* if it contains no independent set of 3 vertices such that each pair is joined by a path that avoids the neighborhood of the third.
- *Split graphs:* G is a *split graph* if its vertex set can be split into a clique and an independent set.

REFERENCES

[1] N. Alon, E. Fischer, I. Newman, and A. Shapira, *A combinatorial characterization of the testable graph properties: It's all about regularity*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 2006, pp. 251–260.

[2] N. Alon, P. Seymour, and R. Thomas, *A separator theorem for graphs with an excluded minor and its applications*, in Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 1990, pp. 293–299.

[3] N. Alon and A. Shapira, *Every monotone graph property is testable*, in Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 2005, pp. 128–137.

[4] N. Alon and A. Shapira, *A characterization of the (natural) graph properties testable with one-sided error*, SIAM J. Comput., 37 (2008), pp. 1703–1727.

[5] N. Alon and A. Shapira, *A separation theorem in property testing*, Combinatorica, 28 (2008), pp. 261–281.

[6] M. A. Bender and D. Ron, *Testing properties of directed graphs: Acyclicity and connectivity*, Random Structures Algorithms, 20 (2002), pp. 184–205.

[7] A. Bogdanov, K. Obata, and L. Trevisan, *A lower bound for testing* 3*-colorability in bounded-degree graphs*, in Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS), IEEE, Los Alamitos, CA, 2002, pp. 93–102.

[8] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi, *Graph limits and parameter testing*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 2006, pp. 261–270.

[9] A. Czumaj and C. Sohler, *Abstract combinatorial programs and efficient property testers*, SIAM J. Comput., 34 (2005), pp. 580–615.

[10] A. Czumaj and C. Sohler, *Sublinear-time algorithms*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 89 (2006), pp. 23–47.

[11] E. Fischer, *The art of uninformed decisions: A primer to property testing*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 75 (2001), pp. 97–126.

[12] J. R. Gilbert, J. P. Hutchinson, and R. E. Tarjan, *A separator theorem for graphs of bounded genus*, J. Algorithms, 5 (1984), pp. 391–407.

[13] O. Goldreich, *Property testing in massive graphs*, in Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer Academic Publishers, Dordrecht, 2002, pp. 123–147.

[14] O. Goldreich and D. Ron, *A sublinear bipartiteness tester for bounded degree graphs*, Combinatorica, 19 (1999), pp. 335–373.

[15] O. Goldreich and D. Ron, *Property testing in bounded degree graphs*, Algorithmica, 32 (2002), pp. 302–343.

[16] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, London, UK, 1980.

[17] R. Kumar and R. Rubinfeld, *Sublinear time algorithms*, SIGACT News, 34 (2003), pp. 57–67.

[18] R. J. Lipton and R. E. Tarjan, *A separator theorem for planar graphs*, SIAM J. Appl. Math., 36 (1979), pp. 177–189.

[19] J. L. Ramirez-Alfonsin and B. A. Reed, eds., *Perfect Graphs*, John Wiley and Sons, New York, 2001.

[20] D. Ron, *Property testing*, in Handbook of Randomized Computing, Vol. II, P. M. Pardalos, S. Rajasekaran, J. Reif, and J. D. P. Rolim, eds., Kluwer Academic Publishers, Dordrecht, 2001, pp. 597–649.