

On Testable Properties in Bounded Degree Graphs ^{*}

Artur Czumaj[†]

Christian Sohler^{‡ §}

Abstract

We study graph properties which 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 believed that almost all, even very simple graph properties require a large complexity to be tested for arbitrary (bounded degree) graphs. Therefore in this paper we focus our attention on testing graph properties for special classes of graphs. We call a graph family *non-expanding* if every graph in this family is not a weak expander (its expansion is $\mathcal{O}(1/\log^2 n)$, where n is the graph size). A *graph family is hereditary* if it is closed under vertex removal. Similarly, a *graph property is hereditary* if it is closed under vertex removal. Next, we call a graph property Π to be *testable* for a graph family $\mathcal{F}$ if for every graph $G \in \mathcal{F}$, in time independent of the size of G we can distinguish between the case when G satisfies property Π and when it is far from every graph satisfying property Π . In this paper we prove that

In the bounded degree graph model, any hereditary property is testable if the input graph belongs to a hereditary and non-expanding family of graphs.

As an application, our 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, in the bounded degree graph model very few graph properties have been known to be testable for any graph classes.

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, bipartiteness, monotonicity, or convex position) or is far from 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 adjacency matrix and the property is bipartiteness, then a graph is ϵ -far from bipartite if one has to delete more than ϵn^2 edges to make it bipartite.

Given oracle access to the object, many objects and properties are known to have randomized property testing algorithms whose time 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 considered for graphs and hypergraphs, functions, point sets, formal languages, and many other structures (for the references, see the excellent surveys [13, 15, 16, 22, 28]). After a series of results for specific problems, recently much attention has been devoted to study 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 [2]). Similarly, recently it has been shown that a property is testable if and only if it is either hereditary or it is close (in some well-defined sense) to a hereditary property (see [6] and [10, 25]).

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 [20]. In this model, the input graph G is represented by its *adjacency 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 i th

^{*}Research supported in part by NSF ITR grant CCR-0313219, NSF DMS grant 0354600, and DFG grant Me 872/8-3.

[†]Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom. czumaj@dcs.warwick.ac.uk.

[‡]Department of Computer Science, University of Paderborn, 33102 Paderborn, Germany. cschler@uni-paderborn.de.

[§]Work was done while the author was visiting Rutgers University.

neighbor of a given vertex v , and the number of accesses to the adjacency list corresponds to the 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 remove more than ϵdn edges from G to obtain a graph having property Π .

Unlike in the adjacency matrix model, in the bounded degree graph model only a few, very simple graph properties (like connectivity) are known to be testable in constant time [20]; instead, main research focused on designing property testers with sublinear query complexity (like, $\mathcal{O}(\sqrt{n})$ tester for bipartiteness [20]). Even more, it has been demonstrated that unlike in the adjacency matrix model, many basic properties are having non-constant complexity: testing acyclicity in directed graphs has $\Omega(n^{1/3})$ query complexity [8], the query complexity of the very basic property of being bipartite is $\Omega(\sqrt{n})$ [20], and the query complexity of testing 3-colorability is $\Omega(n)$ [9]. In fact, it is believed that very few properties can be tested in the bounded degree model with $o(\sqrt{n})$ or even with $o(n)$ query complexity.

In this paper, we address the question to characterize which properties are testable in constant time in the bounded degree model. Our approach to the problem differs from previous approaches in the following way. Instead of considering any input graph, we assume that the input graph comes from a certain family of graphs, e.g., from the family of planar graphs. We call a graph property *testable* if it can be tested in time independent of the size of the input graph. A family of graphs is called *non-expanding* if it does not contain graphs with expansion $\Omega(1/\log^2 n)$; (this is informally equivalent to the families of graphs with some separator properties). A family of graphs is called *hereditary* if it is closed under vertex removal. Similarly, a graph property is called *hereditary* if it is closed under vertex removal. We show (informally) the following result:

In the bounded degree graph model, any hereditary property is testable if the input graph belongs to a hereditary and non-expanding family of graphs.

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 adjacency matrix model [6]). The class of hereditary graph properties contains also trivially all *monotone graphs properties* (properties closed under edge removal). Many interesting graph properties are hereditary, for example, being acyclic, stable (independent set), planar, perfect, bipartite,

k -colorable, chordal, perfect, interval graph, permutation graph, comparability graph, circular-arc graph, having no induced subgraph H , etc. (see also [15, 27]).

Our result implies that these properties can be tested (in the bounded degree graph model) for the following classes of graphs: 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 these properties for non-trivial classes of 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 $f_G : [n] \times [d] \rightarrow [n] \cup \{+\}$, such that $f_G(v, i)$ denotes the i -th neighbor of v or a special symbol $+$ in 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 ϵdn 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 any graph $G \in \mathcal{F}$ that has property Π with probability at least $\frac{2}{3}$, and (b) rejects any graph $G \in \mathcal{F}$ that is ϵ -far from Π with probability at least $\frac{2}{3}$. 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 -tester presented in this paper has one-sided error.

A property is called *testable* for family $\mathcal{F}$ if for any fixed $0 < \epsilon < 1$ 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 [5], 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 *non-uniformly testable* if for every fixed ϵ , $0 < \epsilon < 1$, 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 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 $|\mathcal{N}(S)| \geq \lambda|S|$. With this, we can now define non-expanding graph families.

DEFINITION 2.2. A family $\mathcal{F}$ of graphs is called **non-expanding** if for every constant $0 < \lambda \leq 1$ there exists a constant $n_0 = n_0(\lambda)$ such that every graph in $\mathcal{F}$ with $n \geq n_0(\lambda)$ vertices is not a $(\lambda/\log^2 n)$ -expander.¹

2.2 Hereditary and non-expanding graph families.

A family $\mathcal{F}$ of graphs is called *hereditary* if it is closed under vertex removal. Similarly, a graph property is called *hereditary* if it is closed under vertex removal.

There are many interesting classes of families of graphs that are hereditary and non-expanding. For example, the *family of planar graphs* is trivially hereditary, and also the classical planar separator theorem [24] implies immediately that it is non-expanding. 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 $|\mathcal{N}(A)| \leq \mathcal{O}(\sqrt{n})$. Therefore, every planar graph with n vertices ($n \geq n_0$ for some constant n_0) is not an $\mathcal{O}(1/\sqrt{n})$ -expander, and hence the family of planar graphs is non-expanding. As the example of planar graphs shows, if a family of graphs has a good separator then it is non-expanding. Therefore, all graph families with good separator properties (for graphs of bounded degree) are non-expanding. Hence, other families of graphs (of bounded degree) that are hereditary and non-expanding 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 such graphs. And so, Gilbert et al. [14] proved that any graph on n vertices with genus g has a separator of order $\mathcal{O}(\sqrt{gn})$, and Alon et al. [4] showed a similar results 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 $\mathcal{O}(h^{3/2}n^{1/2})$.

LEMMA 2.1. *The following families of bounded degree graphs are hereditary and non-expanding:*

- *planar graphs,*

- *graphs of bounded genus,*
- *graphs with forbidden minor,*
- *interval graphs,*
- *unit disk graphs,*
- *planar intersection graphs.*

□

3 Testing hereditary properties in hereditary and non-expanding graph families

In this section we will prove our main result and we will show that the following algorithm is an ϵ -tester for any hereditary property Π and any hereditary non-expanding family of graphs $\mathcal{F}$.

ϵ -TESTER(G, n, Π)
 sample a set S of s vertices uniformly at random
for each $v \in S$ **do**
 $U_v = D(v, \lceil \frac{4}{3} n_0(\mathcal{O}(\epsilon/d)) \rceil)$
 $U = \bigcup_{v \in S} U_v$
if $G|_U$ does not satisfy property Π **then reject**
else accept

Clearly, the number of queries to f_G is upper bounded by $2s d^{\lceil \frac{4}{3} n_0(\mathcal{O}(\epsilon/d)) \rceil + 1}$, which for s being a constant independent of n , means that the number of queries is independent of n . (We will give the exact value for s (which is independent of n but does depend on ϵ and Π) at the end of our analysis, in the proof of Theorem 3.1.)

Since Π is hereditary, we know that our algorithm accepts any graph that has property Π . Thus, we only have to show that any graph that is ϵ -far from Π 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 graph from a hereditary and non-expanding graph family $\mathcal{F}$. Let $0 < \kappa < 1$. If $|V| \geq \frac{4}{3} n_0(\kappa)$, then we can partition the vertex set of G into two sets V_1 and V_2 , $|V_1|, |V_2| \geq \frac{n}{4}$, such that there are at most $\kappa d n / \log^2(\frac{3n}{4})$ edges between V_1 and V_2 .*

Proof : Let $\mathcal{F}$ be a non-expanding, hereditary family of graphs and let $|V| \geq \frac{4}{3} n_0(\kappa)$. Since $\mathcal{F}$ is non-expanding, $G \in \mathcal{F}$, and $n \geq n_0(\kappa)$, there exists a set $S \subseteq V$ of cardinality at most $\frac{n}{2}$ such that $|\mathcal{N}(S)| \leq \kappa |S| / \log^2 n$. If $|S| \geq \frac{n}{4}$, then we can take $V_1 = S$ and $V_2 = V \setminus S$. Then, since $|\mathcal{N}(S)| \leq \kappa |S| / \log^2 n$, there are at most $d \kappa |S| / \log^2 n$ edges between V_1 and V_2 , what would conclude the proof.

Otherwise, we have to consider the case when $|S| < \frac{n}{4}$. Take the graph $G|_{V \setminus S}$. Since $\mathcal{F}$ is hereditary, $G|_{V \setminus S} \in$

¹The choice of the factor $\log^2 n$ can be relaxed. In fact, all what is needed is that every graph in $\mathcal{F}$ with $n \geq n_0(\lambda)$ vertices is not a $(\lambda/\varrho(n))$ -expander, where $\varrho(n)$ is a non-decreasing function satisfying $\sum_{i=0}^{\log n} 1/\varrho(2^i) = \mathcal{O}(1)$.

$\mathcal{F}$, and $|V \setminus S| > \frac{3}{4}n \geq n_0(\kappa)$, 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 $\mathcal{N}(S') \leq \kappa |S'| / \log^2(\frac{3n}{4})$. If we have $|S \cup S'| \geq \frac{n}{4}$ 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 $\mathcal{N}(S \cup S') \leq \kappa |S \cup S'| / \log^2(\frac{3n}{4})$. 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. $\square$

Let us call a connected component *non-trivial* if it has more than a single vertex. The following is a corollary from Lemma 3.1.

COROLLARY 3.1. *Let $\mathcal{F}$ be a hereditary and non-expanding family of graphs. There exists a positive constant c such that for any graph $G \in \mathcal{F}$ there is a set of at most $\epsilon d n/2$ edges, such that (i) their removal partitions G into connected components $C_1, C_2, \dots$ of size at most $\frac{4}{3} n_0(\epsilon/(dc))$ each, (ii) each connected component C_i is an induced subgraph of G , and (iii) no edge connects in G two non-trivial connected components C_i and C_j .*

Proof : Let $G \in \mathcal{F}$ and let κ be a parameter to be set at the end of the proof. We apply Lemma 3.1 to obtain two sets V_1 and V_2 with at most $\kappa d n / \log^2(\frac{3n}{4})$ edges connecting V_1 and V_2 . Let $|V_1| \leq |V_2|$ and let $U^* = \mathcal{N}(V_1)$. Since the number of edges between V_1 and $V \setminus V_1$ is at most $\kappa d n / \log^2(\frac{3n}{4})$, we also have $|U^*| \leq \kappa d n / \log^2(\frac{3n}{4})$. Remove from G all edges incident to U^* . Since $|U^*| \leq \kappa d n / \log^2(\frac{3n}{4})$ and G has maximum degree at most d , we removed at most $\kappa d^2 n / \log^2(\frac{3n}{4})$ edges from G . Next, let $U_1 = V_1$ and $U_2 = V_2 \setminus U^*$. Observe that for $\kappa \leq \log^2(\frac{3n}{4}) / (4d)$ we have $\frac{n}{4} \leq |U_1|, |U_2| \leq \frac{3n}{4}$ 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 smaller than or equal to $\frac{4}{3} n_0(\kappa)$. In this way, we removed some number of edges in G and obtained a graph H on V with connected components $C_1, \dots, C_q$. 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 $H_1, \dots, H_k$ be non-trivial connected components in our new graph. It is easy to see that for every C_i we have $|C_i| \leq \frac{4}{3} n_0(\kappa)$. 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 . Therefore, the three promised properties of $\hat{G}$ are satisfied. Now, we only have to 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:

$$Q(n) = \begin{cases} 0 & \text{if } n \leq \frac{4}{3} n_0(\kappa), \\ \kappa d^2 n / \log^2(\frac{3n}{4}) + \max_{\frac{1}{4} \leq \tau \leq \frac{3}{4}} \{Q(\tau n) + Q((1-\tau)n)\} & \text{if } n > \frac{4}{3} n_0(\kappa). \end{cases}$$

Since $Q(n) = \Theta(\kappa d^2 n)$, we can conclude that the graph H is obtained from G by removal of at most $c \kappa d^2 n/2$ edges, for some positive constant c . This yields the proof by setting $\kappa = \epsilon/(dc)$. $\square$

Let us explain the importance of the three properties of the resulting graph stated in Corollary 3.1. Property (i) ensures that every connected component is small. Property (ii) ensures that if we have some induced subgraph of a non-trivial 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 $Q_{i_1}, Q_{i_2}, \dots, Q_{i_\ell}$ of graphs $H_{i_1}, H_{i_2}, \dots, H_{i_\ell}$, then these copies of the subgraphs do not intersect in H . Therefore, if we define a graph $\hat{Q}$ with ℓ connected components, where the j th connected of $\hat{Q}$ is isomorphic with Q_{i_j} , then $\hat{Q}$ 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., [6, Section 4]). Let us denote by $\mathcal{H}_{\text{forb}}^\Pi$ a *minimal* family of forbidden subgraphs for property Π . Notice that in general, $\mathcal{H}_{\text{forb}}^\Pi$ may be an *infinite* family of forbidden graphs.

(Observe that, for example, if Π is the property of being bipartite, then $\mathcal{H}_{\text{forb}}^\Pi$ can be chosen to be the set of all odd cycles, and if Π is the property of being chordal, then $\mathcal{H}_{\text{forb}}^\Pi$ is the set of all cycles of length at least 4.)

For simplicity of presentation (but without loss of generality) we will assume that the graphs in $\mathcal{H}_{\text{forb}}^\Pi$ contain no isolated vertices. The reason why we can make such an assumption is that every large enough bounded degree graph G will always have an arbitrary large induced subgraph that consists of isolated vertices only.

Next, let us consider a graph $G \in \mathcal{F}$ that is ϵ -far from Π . By Corollary 3.1, we can remove from G at most $\epsilon d n/2$ edges to obtain a graph H on the same vertex set for which each connected component has at most $\zeta = \frac{4}{3} \cdot n_0(\epsilon/(dc))$ vertices. Furthermore, if $H_1, \dots, H_k$ are the non-trivial 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 $\epsilon/2$ -far from Π . Since

all connected components in H have size at most ζ (which is independent of n), H cannot contain as a subgraph any graph that has a connected component with more than ζ vertices. Let $\mathcal{J}$ denote the family of all graphs whose connected components have size at most ζ (notice that $\mathcal{J}$ is independent of G). We conclude that it suffices to consider the subgraphs in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$.

LEMMA 3.2. *If $G \in \mathcal{F}$ is ϵ -far from Π , then H contains as an induced subgraph a graph from $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$. The same holds also if we remove from H any set of at most $\epsilon d n/2$ edges.* $\square$

Let us denote by c the number of *connected* (unlabeled) graphs on a set of at most ζ vertices; clearly $c \leq 2^{\zeta^2}$. Let us enumerate all possible connected graphs with at most ζ vertices by $\mathfrak{G}_1, \dots, \mathfrak{G}_c$. Then, we can define any graph $\mathcal{G}$ in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ as a c -ary integer vector $f = \langle f_1, \dots, f_c \rangle$, 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 $\mathcal{H}_{\text{forb}}^\Pi$ and $\mathcal{J}$).

Similarly, let us define a c -ary integer vector $g = \langle g_1, \dots, g_c \rangle$ with g_i being the number of *induced copies* of graph $\mathfrak{G}_i$ in H . Notice that there is a fundamental difference between the ways of counting copies of $\mathfrak{G}_i$ in $\mathcal{G}$ and in H : all copies of graphs $\mathfrak{G}_1, \dots, \mathfrak{G}_c$ counted in the characteristic vector of $\mathcal{G}$ are disjoint while the induced copies of these graphs counted in g can intersect.

3.1.1 Finiteness of $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$. As we noted earlier, the set $\mathcal{H}_{\text{forb}}^\Pi$ does not need to be finite; similarly, $\mathcal{J}$ is infinite (because it may have an arbitrary number of connected components). For our analysis, however, we need to show that their intersection is finite.

LEMMA 3.3. *$\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ is finite.*

Proof : Let us first observe that the minimality of $\mathcal{H}_{\text{forb}}^\Pi$ ensures that for any two distinct graphs $\mathcal{G}', \mathcal{G}'' \in \mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with characteristic vectors $f' = \langle f'_1, \dots, f'_c \rangle$ and $f'' = \langle f''_1, \dots, f''_c \rangle$, respectively, it is impossible that $f'_i \geq f''_i$ for every i ; that is, if $f'_i \geq f''_i$ then there is j such that $f'_j < f''_j$.

For any graph $\mathcal{G} \in \mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with characteristic vector $f = \langle f_1, \dots, f_c \rangle$, define the number of different types of its connected components to be $|\{i : f_i > 0\}|$. We will prove the lemma by induction on the number ϖ of different types of connected components of graphs in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$; notice that $1 \leq \varpi \leq c$.

If $\varpi = 1$ for a graph $\mathcal{G}$ in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$, then every connected component of $\mathcal{G}$ is identical; let it be $\mathfrak{G}_i$. Thus, any graph in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ consists of a number of disjoint

copies of $\mathfrak{G}_i$. We claim that the minimality of $\mathcal{H}_{\text{forb}}^\Pi$ ensures that there is no other graph $\mathcal{G}^*$ in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ that has only one type of connected components and that consists of a number of disjoint copies of $\mathfrak{G}_i$. This follows from the fact that the one of $\mathcal{G}$ and $\mathcal{G}^*$ that has fewer connected components would be a subgraph of the other graph. Hence, the other graph would not be a minimal forbidden subgraph. This contradicts to the definition of $\mathcal{H}_{\text{forb}}^\Pi$. Thus, for any i , there can be only one graph in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ that consists of a number of disjoint copies of $\mathfrak{G}_i$. Since there are c different graphs $\mathfrak{G}_i$, the number of graphs in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with a single type of connected components is at most c , and so it is finite.

Now assume that we know that the number of graphs in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with ϖ different types of connected components is finite. We want to show that this is also the case for $\varpi + 1$ types. Fix an arbitrary graph $\mathcal{G}$ in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ that consists of $\varpi + 1$ different types of connected components. Without loss of generality, suppose that $\mathcal{G}$ has $f_i > 0$ connected components of type i , for $1 \leq i \leq \varpi + 1$. The minimality of $\mathcal{H}_{\text{forb}}^\Pi$ implies that we can partition all graphs in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with $\varpi + 1$ different types of connected components into $\varpi + 1$ classes, $C_1, \dots, C_{\varpi+1}$, such that if a graph $\mathcal{G}'$ from $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ (with $\varpi + 1$ different types of connected components) has a characteristic vector $f' = \langle f'_1, \dots, f'_c \rangle$, then $\mathcal{G}' \in C_j$ if $f_1 \geq f'_1, \dots, f_{j-1} \geq f'_{j-1}, f_j < f'_j$. If every class is finite then we have nothing to prove, and so, let us suppose that at least one class, call it C_i , is infinite. Since f_i is finite and integral, for each graph $\mathcal{G}' \in C_i$ there are at most f_i choices for the number of copies of the connected component $\mathfrak{G}_i$. Once the number of connected components $\mathfrak{G}_i$ is fixed, we only have ϖ types of connected components left. But we know that the number of constraint graphs for ϖ types is finite and so it is for $\varpi + 1$. This completes the induction. $\square$

For a given hereditary property Π and for a given ϵ , let $\psi_\Pi(\epsilon)$ denote the maximum number of vertices of a graph in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$. Lemma 3.3 implies that $\psi_\Pi(\epsilon)$ is a constant independent of n . On the other hand, notice that g_i (an entry in the characteristic vector of H) may depend on n .

With the notation above and Lemma 3.3, we can now show the following lemma.

LEMMA 3.4. *There exists a graph $\mathcal{G}$ in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ such that for the characteristic vector $f = \langle f_1, \dots, f_c \rangle$ for $\mathcal{G}$, for all $1 \leq i \leq c$ it holds that either $f_i = 0$ or $f_i \leq g_i - \gamma n$, where $\gamma = \frac{\epsilon \cdot d}{2 \cdot c \cdot \zeta^2}$.*

Proof : For the purpose of contradiction, let us assume that such a graph $\mathcal{G}$ does not exist. Then, for any subgraph

in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ with characteristic vector $f = \langle f_1, \dots, f_c \rangle$ we have a graph $\mathfrak{G}_i$ with $f_i > 0$ and $f_i > g_i - \gamma n$. We will show that one can delete less than $\epsilon d n/2$ edges to obtain a graph such that for every i , it contains at most $\max\{g_i - \gamma n, 0\}$ induced copies of $\mathfrak{G}_i$, what by the property above implies that it contains no subgraph in $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$. This would contradict to the fact that H is $\epsilon/2$ -far from Π .

For every $1 \leq i \leq c$, delete one by one induced copies of $\mathfrak{G}_i$ from H until at most $\max\{g_i - \gamma n, 0\}$ induced copies of $\mathfrak{G}_i$ are left. Since a deletion from H of a single induced copy of $\mathfrak{G}_i$ removes at most ζ^2 edges from H , this will remove at most $c \cdot \zeta^2 \cdot \gamma n \leq \epsilon d n/2$ edges from H . Clearly, the resulting graph does not contain any subgraph from $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$, and thus, from $\mathcal{H}_{\text{forb}}^\Pi$. Since we removed at most $\epsilon d n/2$ edges, this contradicts to the fact that G is ϵ -far from Π . $\square$

3.1.2 Final analysis of the tester. We are now ready to show the following main theorem.

THEOREM 3.1. *Let $\mathcal{F}$ be a hereditary and non-expanding family of graphs. Then every hereditary graph property Π is non-uniformly testable for $\mathcal{F}$ with one-sided error. Furthermore, Π is uniformly testable with one-sided error if $\psi_\Pi(\epsilon)$ and $n_0(\epsilon)$ are computable (or if their approximations are computable, where the quality of the approximation must be independent of the input graph size).*

Proof : Let us first fix $\mathcal{F}$, Π , and $\mathcal{H}_{\text{forb}}^\Pi$. Let G be any bounded degree graph from $\mathcal{F}$. We will show that ϵ -TESTER(G, n, Π) (see the beginning of Section 3) is an ϵ -tester for G for property Π and then we will discuss its complexity. (We will chose the value for s at the end of the proof.)

It is obvious that if G has property Π then ϵ -TESTER(G, n, Π) accepts G . Hence, we only have to show that if G is ϵ -far from Π then ϵ -TESTER(G, n, Π) rejects G with probability at least $\frac{2}{3}$.

For a graph $G \in \mathcal{F}$ that is ϵ -far from Π , let H be constructed as described at the beginning of Section 3.1 (see also Corollary 3.1 for the properties of H) and let $H_1, \dots, H_k$ be the non-trivial connected components of H . For each graph H_i , let us choose one vertex v_i from H_i to be its *representative*.

Consider the algorithm ϵ -TESTER(G, n, Π). Observe that if a representative v_i is in S , then H_i is an induced subgraph of $G|_U$, and (by Corollary 3.1) if certain graph $\mathfrak{G}_j$ is an induced subgraph of H_i , then $\mathfrak{G}_j$ is also an induced subgraph of $G|_U$. Therefore, our goal is to show that (with probability at least $\frac{2}{3}$) the sample set S will contain a set of representatives $v_{i_1}, v_{i_2}, \dots$ such that

the graph $H_{i_1} \cup H_{i_2} \cup \dots$ will contain as an induced subgraph a graph $\mathcal{G} \in \mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$; this would imply that ϵ -TESTER(G, n, Π) is an ϵ -tester.

Let $\mathcal{G}$ be a subgraph of $\mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ that satisfies Lemma 3.4. Let $f = \langle f_1, \dots, f_c \rangle$ be its characteristic vector and let I be the set of indices for which $f_i > 0$. It follows from Lemma 3.4 that for every $i \in I$ we have $g_i > \gamma n$, where $\gamma = \frac{\epsilon \cdot d}{2 \cdot c \cdot \zeta^2}$.

Let i be any index in I . Since H has more than γn induced copies of $\mathfrak{G}_i$, let us find one and remove from H the connected component containing that copy of $\mathfrak{G}_i$. Notice that by removing this connected component, we can remove at most 2^ζ different induced subgraphs from H . Therefore, even if we remove $\varrho = \frac{\gamma n}{2^\zeta c}$ disjoint induced copies of $\mathfrak{G}_i$ from H , still, for any index $j \in I$, the resulting graph will have more than $\gamma n - 2^\zeta \cdot \varrho = \gamma n - \frac{\gamma n}{c}$ induced copies of $\mathfrak{G}_j$. Therefore, we can apply this procedure to remove from H ϱ disjoint induced copies of all the graphs $\mathfrak{G}_i$ with $i \in I$, and still, for every $j \in I$, the obtained graph will contain a positive number of induced copies of $\mathfrak{G}_j$. Therefore, if we consider the connected components of H that we removed in this procedure, then our analysis implies that H contains $|I| \cdot \varrho$ disjoint connected components that we can enumerate as $H_{i,j}$ with $i \in I, 1 \leq j \leq \varrho$, such that $H_{i,j}$ contains $\mathfrak{G}_i$ as an induced subgraph.

For any $i \in I$, let X_i be the number of distinct representatives of the sets $H_{i,j}$ with $1 \leq j \leq \varrho$ that have been chosen to the sample set S . Since S is a randomly chosen subset of V of size s , it is easy to see that we have $\mathbf{E}[X_i] \geq s \varrho / (2n)$. Now, we can use Chernoff bound to prove that $\Pr[X_i \leq s \varrho / (4n)] \leq \exp(-s \varrho / (16n))$. Hence, for $s \geq 2^{\zeta+5} c^2 \zeta^2 \ln(3c) / (\epsilon d)$, we have

$$(3.1) \quad \Pr[X_i \leq s \varrho / (4n)] \leq \exp(-s \varrho / (16n)) \leq \frac{1}{3c}.$$

Next, we observe that if we choose $s \geq 2^{\zeta+3} c^2 \zeta^2 \psi_\Pi(\epsilon) / (\epsilon d)$, then $\psi_\Pi(\epsilon) \leq s \varrho / (4n)$. Therefore, in that case (3.1) yields $\Pr[\exists i \in I X_i \leq \psi_\Pi(\epsilon)] \leq |I| \cdot \frac{1}{3c} \leq \frac{1}{3}$, and therefore, $\Pr[\forall i \in I X_i > \psi_\Pi(\epsilon)] \geq \frac{2}{3}$. Hence, by the fact that $\psi_\Pi(\epsilon) \geq f_i$ for every i , and by our arguments above, we have proven that with probability at least $\frac{2}{3}$, in the algorithm ϵ -TESTER(G, n, Π) the graph $G|_U$ will contain $\mathcal{G} \in \mathcal{H}_{\text{forb}}^\Pi \cap \mathcal{J}$ as the induced subgraph. This implies that ϵ -TESTER(G, n, Π) is an ϵ -tester.

In our analysis, we can set $s = 2^{\zeta+5} c^2 \zeta^2 \ln(3c) / (\epsilon d) + 2^{\zeta+3} c^2 \zeta^2 \psi_\Pi(\epsilon) / (\epsilon d)$. Since all the parameters defining s are independent of n , the complexity of ϵ -TESTER(G, n, Π) is independent of n . Therefore, we have just proven that property Π is testable for the graph family $\mathcal{F}$.

Since we can give upper bounds on ζ and c in terms of $n_0(\epsilon/(dc))$, it follows that we have a uniform tester, if

the functions $\psi_{\Pi}(x)$ and $n_0(x)$ are computable. $\square$

4 Conclusions

In this paper we started a first attempt to characterize the graph properties that can be tested in time independent of the size of the graph in the bounded degree graph model for sparse graphs. We showed that all hereditary graph properties are (non-uniformly) testable, if the graph comes from a family of graphs that contains no weak-expander. Some interesting open questions include.

- 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 non-expanding 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)?

References

- [1] N. Alon, E. Fischer, M. Krivelevich, M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4): 451–476, 2000.
- [2] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 251–260, 2006.
- [3] N. Alon, T. Kaufman, M. Krivelevich, and D. Ron. Testing triangle-freeness in general graphs. *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 279–288, 2006.
- [4] N. Alon, P. Seymour, and R. Thomas. A separator theorem for graphs with an excluded minor and its applications. *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 293–299, 1990.
- [5] N. Alon and A. Shapira. Every monotone graph property is testable. *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 128–137, 2005.
- [6] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 429–438, 2005.
- [7] N. Alon and A. Shapira. Homomorphisms in graph property testing - A survey. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 85, 2005.
- [8] M. A. Bender and D. Ron. Testing properties of directed graphs: acyclicity and connectivity. *Random Structures and Algorithms*, 20(2): 184–205, 2002.
- [9] A. Bogdanov, K. Obata, and L. Trevisan. A lower bound for testing 3-colorability in bounded-degree graphs. *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 93–102, 2002.
- [10] C. Borgs, J. Chayes, L. Lovász, V. T. Sos, B. Szegedy, and K. Vesztegombi. Graph limits and parameter testing. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 261–270, 2006.
- [11] A. Czumaj and C. Sohler. Abstract combinatorial programs and efficient property testers. *SIAM Journal on Computing*, 34(3): 580–615, 2005.
- [12] A. Czumaj and C. Sohler. Sublinear-time algorithms *Bulletin of the EATCS*, 89: 23–47, June 2006.
- [13] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the EATCS*, 75: 97–126, October 2001.
- [14] J. R. Gilbert, J. P. Hutchinson, and R. E. Tarjan. A separator theorem for graphs of bounded genus. *Journal of Algorithms*, 5: 391–407, 1984.
- [15] O. Goldreich. Combinatorial property testing (a survey). In P. Pardalos, S. Rajasekaran, and J. Rolim, editors, *Proceedings of the DIMACS Workshop on Randomization Methods in Algorithm Design*, volume 43 of *DIMACS, Series in Discrete Mathematics and Theoretical Computer Science*, pp. 45–59, 1997. American Mathematical Society, Providence, RI, 1999.
- [16] O. Goldreich. Property testing in massive graphs. In J. Abello, P. M. Pardalos, and M. G. C. Resende, editors, *Handbook of massive data sets*, pp. 123–147. Kluwer Academic Publishers, 2002.
- [17] O. Goldreich. Contemplations on testing graph properties. August 2005.
- [18] O. Goldreich and D. Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.
- [19] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 7, 2000.
- [20] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2): 302–343, 2002.
- [21] M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, 1980.
- [22] R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34: 57–67, 2003.
- [23] T. Kaufman, M. Krivelevich, and D. Ron. Tight bounds for testing bipartiteness in general graphs. *SIAM Journal on Computing*, 33(6): 1441–1483, 2004.
- [24] R. J. Lipton and R. E. Tarjan. A separator theorem for planar graphs. *SIAM Journal of Applied Mathematics*, 36(2):177–189, 1979.
- [25] L. Lovász and B. Szegedy. Graph limits and testing hereditary graph properties. Technical Report, MSR-TR-2005-110, Microsoft Research, August 2005.
- [26] M. Parnas and D. Ron. Testing the diameter of graphs. *Random Structures and Algorithms*, 20(2): 165–183, 2002.
- [27] J. L. Ramirez-Alfonsin and B. A. Reed, editors. *Perfect*

- Graphs. Wiley and Sons, 2001.
- [28] D. Ron. Property testing. In P. M. Pardalos, S. Rajasekaran, J. Reif, and J. D. P. Rolim, editors, *Handbook of Randomized Algorithms*, volume II, pp. 597–649. Kluwer Academic Publishers, 2001.