Electronic Colloquium on Computational Complexity, Report No. 83 (2007)

Testing Hereditary Properties of Non-Expanding Bounded-Degree Graphs $^{*\,\dagger}$

Artur Czumaj[‡] Asaf Shapira[§] Christian Sohler[¶]

Abstract

We study property testing in the model of bounded degree graphs. It is well known that in this model many graph properties cannot be tested with a constant number of queries and it seems reasonable to conjecture that only few are testable with $o(\sqrt{n})$ queries. Therefore in this paper we focus our attention on testing graph properties for special classes of graphs, with the aim of proving the testability of general families of graph properties under the assumption that the input graph belongs to a (natural) family of graphs. We call a graph family *non-expanding* if every graph in this family has a weak expansion (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. 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.*

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, very few graph properties have been known to be testable for general graph classes in the bounded degree graph model.

---

[*] A preliminary version of this paper, entitled "On Testable Properties in Bounded Degree Graphs," authored by the first and third authors, appeared in the Proc. of 18^{th} Symposium on Discrete Algorithm (SODA), New Orleans, Louisiana, 2007, 494-501.

[†] Research 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.

[‡] Department of Computer Science, University of Warwick, Coventry, CV4 7AL, U.K. czumaj@dcs.warwick.ac.uk.

[§] Microsoft Research. asafico@tau.ac.il.

[¶] Department of Computer Science, University of Paderborn, 33102 Paderborn, Germany. csohler@upb.de. Work done while the author was visiting Rutgers University.

ISSN 1433-8092

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 adjacency matrix and the property is bipartiteness, then a graph is ϵ-far from bipartite if one has to delete more than $\epsilon\, n^2$ 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 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 considered for graphs and hypergraphs, functions, point sets, formal languages, and many other structures (for the references, see the excellent surveys [14, 16, 17, 24, 30]). 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, a model that was introduced by Goldreich, Goldwasser and Ron [19]. 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]). Furthermore, it has been shown in [6] 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 [11, 27] for alternative proofs). These results, imply that in the adjacency matrix model, essentially any "natural" graph property can be tested with a constant number of queries.

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 [22]. 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} 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 $\epsilon\, d\, n$ edges in G to obtain a graph having property Π.

Unlike the adjacency matrix model discussed above, in the bounded degree graph model only few graph properties are known to be testable in constant time, see [22] 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, $\mathcal{O}(\sqrt{n})$ tester for bipartiteness [20]). Even more, it has been demonstrated that unlike in the adjacency matrix model, in the bounded degree model many basic properties have a non-constant query complexity. For example, acyclicity in directed graphs has $\Omega(n^{1/3})$ query complexity [9], the property of being bipartite has query complexity $\Omega(\sqrt{n})$ [22], and the query complexity of testing 3-colorability is $\Omega(n)$ [10]. In fact, it seems reasonable to conjecture that very few properties can be tested in the bounded degree model with $o(\sqrt{n})$.

In this paper, we take a new approach and we 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 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 family of graphs is called *non-expanding* if it does not contain graphs with expansion larger than $1/\log^2 n$; (this is informally equivalent to the families of graphs with some good 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 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.

The reader is referred to Theorem 1 for the precise statement of our main result. 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 [6]). The class of hereditary graph properties contains also trivially all *monotone graphs properties* (properties closed under removal of edges and vertices). Many interesting graph properties are hereditary, for example, being acyclic, stable (independent set), planar, perfect, bipartite, k-colorable, chordal, perfect, interval, permutation, having no induced subgraph H, etc. (see also [16, 29]). 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 which is hereditary and non-expanding. Examples of natural hereditary non-expanding families 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 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 $f_G : [n] \times [d] \to [n] \cup \{+\}$, such that $f_G(v, i)$ denotes the i^{th} 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 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 with probability at least $\frac{2}{3}$ any graph $G \in \mathcal{F}$ that has property Π.

(b) Rejects with probability at least $\frac{2}{3}$ 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 $0 < \epsilon < 1$ 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 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 $|S| \le n/2$, we have $|\mathcal{N}(S)| \ge \lambda |S|$. With this, we can now define non-expanding graph families.

Definition 2.2 *A family of graphs $\mathcal{F}$ is called **non-expanding** if there exists a constant $n_{\mathcal{F}}$ such that all graphs in $\mathcal{F}$ of size at least $n_{\mathcal{F}}$ are not $(1/\log^2 n)$-expanders* [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 [2].

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 [26] 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, $\frac{1}{3}n \le |A| \le \frac{1}{2}n$, such that $|\mathcal{N}(A)| \le \mathcal{O}(\sqrt{n})$. Therefore, every planar graph with n vertices ($n \ge n_0$ 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. [15] 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 $\mathcal{O}(h^{3/2}n^{1/2})$.

[1] The choice of the factor $1/\log^2 n$ can be relaxed. In fact, using known bounds one can replace $1/\log^2 n$ with $1/(\log n \log^2 \log n)$.

[2] There is, of course, no difference between a graph property and a family of graphs. We use the different terms in order to distinguish between the property we want to test and the family of graphs to which the input is assumed to belong to.

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 hereditary non-expanding family of graphs $\mathcal{F}$.

```
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
```

Clearly, the number of queries to f_G is upper bounded by $2\, s_1\, d^{s_2}$, which for s_1 and s_2 being constants independent of n, gives the number of queries to be 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 and Π, at the end of our analysis, in the proof of Theorem 1.

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 $\mathcal{F}$ be any family of graphs that is both hereditary and non-expanding, and let $n_{\mathcal{F}}$ be the constant from Definition 2.2. Let δ be an arbitrary positive parameter. If $G = (V, E) \in \mathcal{F}$ satisfies $n = |V| \geq \max\{2n_{\mathcal{F}}, 2^{2/\delta^2}\}$ then one can partition V into two sets V_1 and V_2, such that $|V_1|, |V_2| \geq \frac{n}{4}$ and $e(V_1, V_2) \leq \delta\, d\, n/\log^{1.5} n$.*

Proof : Since $\mathcal{F}$ is non-expanding, every graph $G \in \mathcal{F}$ on $n \geq n_{\mathcal{F}}$ vertices is not a $1/\log^2 n$-expander. Therefore, there exists a set $S \subseteq V$ of cardinality at most $\frac{n}{2}$ such that $|\mathcal{N}(S)| \leq |S|/\log^2 n$. We first observe that if $|S| \geq \frac{n}{4}$, then we can take $V_1 = S$ and $V_2 = V \setminus S$. Indeed, since $|\mathcal{N}(S)| \leq |S|/\log^2 n$, there are at most $dn/\log^2 n$ edges between V_1 and V_2. Therefore, if in addition $n > 2^{2/\delta^2}$, we can infer that

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

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 $\mathcal{F}$ is hereditary, $G_{|V\setminus S} \in \mathcal{F}$, and $|V \setminus S| > n_{\mathcal{F}}$ (recall that $n > 2n_{\mathcal{F}}$), 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 $|\mathcal{N}(S')| \leq 2|S'|/\log^2 n$. 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 $|\mathcal{N}(S \cup S')| \leq 2\,|S \cup S'|/\log^2 n$. 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 *non-trivial* if it has more than a single vertex. The following is a corollary of Lemma 3.1.

Corollary 3.2 *For every hereditary and non-expanding family of graphs $\mathcal{F}$, there exists a positive constant $c = c_{\mathcal{F}}$, such that one can remove from any graphs $G \in \mathcal{F}$ a set of at most $\epsilon\, d\, n/2$ edges, such that*

(i) Their removal partitions G into connected components $C_1, C_2, \ldots$ of size at most $2^{c/\epsilon^2}$.

(ii) Each connected component C_i is an induced subgraph of G.

(iii) No edge connects in G two non-trivial connected components C_i and C_j.

Proof : Let $n_{\mathcal{F}}$ be the constant associated with $\mathcal{F}$ as in Definition 2.2, let G be any graph in $\mathcal{F}$, 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 $\max\{2n_{\mathcal{F}}, 2^{2/\delta^2}\}$. 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 non-trivial connected components in our new graph. By definition, every C_i has size $|C_i| \le \max\{2n_{\mathcal{F}}, 2^{2/\delta^2}\}$. 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:

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

Since $Q(n) = \Theta(\delta\, d^2\, n)$, 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'. This yields the proof by setting $\delta = \epsilon/(2dc')$. Finally, recall that all the connected components of H had size $|C_i| \le \max\{2n_{\mathcal{F}}, 2^{2/\delta^2}\} \le 2^{c/\epsilon^2}$ if we take $c = c_{\mathcal{F}} = 2\, d\, c'\, n_{\mathcal{F}}$. □

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 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 $Q_{i_1}, Q_{i_2}, \ldots, Q_{i_\ell}$ 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 $\widehat{Q}$ with ℓ connected components, where the j^{th} connected of $\widehat{Q}$ is isomorphic with Q_{i_j}, then $\widehat{Q}$ 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 $\mathcal{H}_{forb}^{\Pi}$ 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. Observe that, 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.

For simplicity of presentation (but without loss of generality) we will assume that the graphs in $\mathcal{H}^{\Pi}_{forb}$ 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. Therefore, in such cases, all large enough graphs will not satisfy $\mathcal{H}^{\Pi}_{forb}$, and thus testing $\mathcal{H}^{\Pi}_{forb}$ becomes trivial.

Next, let us consider an arbitrary graph $G \in \mathcal{F}$ that is ϵ-far from Π. By Corollary 3.2, 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 $r = 2^{c/\epsilon^2}$ 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 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 $G \in \mathcal{F}$ is ϵ-far from Π, then H (define 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 $\epsilon\, d\, n/2$ edges.* $\square$

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 most r vertices by $\mathfrak{G}_1, \dots, \mathfrak{G}_{\mathfrak{c}(r)}$. Then, we can define any graph $\mathcal{G}$ in $\mathcal{H}^{\Pi}_{forb} \cap \mathcal{J}_r$ as a $\mathfrak{c}(r)$-ary integer vector $f = \langle f_1, \dots, f_{\mathfrak{c}(r)} \rangle$, 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 $\mathcal{H}^{\Pi}_{forb}$ and $\mathcal{J}_r$).

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

Lemma 3.4 *Let $\mathcal{F}$ be a fixed hereditary non-expanding family of graphs and let Π be a fixed hereditary property. Suppose that $G \in \mathcal{F}$ is a 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 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 $f = \langle f_1, \dots, f_{\mathfrak{c}(r)} \rangle$ such that for all $1 \le i \le \mathfrak{c}(r)$ it holds that if $f_i > 0$ then $\mathfrak{g}_i^{\langle H \rangle} \ge \gamma\, n$, where $\gamma = \epsilon \cdot d/2^{r^2}$.*

Proof : Let $\mathfrak{G}_1, \dots, \mathfrak{G}_{\mathfrak{c}(r)}$ be all connected graphs of size at most r. We will first construct a graph H' by removing some edges from H so that for any graph $\mathfrak{G}_i$ either H' contains no copy of $\mathfrak{G}_i$ or it contains at least $\gamma\, n$ such copies. We proceed sequentially over the graphs $\mathfrak{G}_1, \dots, \mathfrak{G}_{\mathfrak{c}(r)}$. For each $\mathfrak{G}_i$ we do the following: if the number of induced copies in the current graph obtained from H is smaller than $\gamma\, n$, then we remove all the edges of any connected component that contains $\mathfrak{G}_i$ as an induced subgraph. 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 $\mathfrak{c}(r) \cdot \binom{r}{2} \cdot \gamma n < \epsilon\, d\, n/2$. 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 copy of $\mathfrak{G}_i$ or it contains at least $\gamma\, n$ such copies.

Since G was assumed to be ϵ-far from Π, and H was obtained from G by removing at most $\epsilon\, d\, n/2$ edges, we have that H is $\frac{\epsilon}{2}$-far from Π. Also, since H' is obtained from H by removing less than $\epsilon\, d\, n/2$ 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 for which $f_i > 0$ we must have that H' contains at least $\gamma\, n$ copies of $\mathfrak{G}_i$. Finally, observe that from the definition of the process of obtaining H' it follows that H must contain at least this many induced copies of $\mathfrak{G}_i$. Hence, for every i for which $f_i > 0$ we have $\mathfrak{g}_i^{\langle H \rangle} \geq \gamma\, n$. $\square$

3.2 Function Ψ_Π

We now introduce a key notion that we will use to test a hereditary property Π. Note, that the discussion below does not relate to the family of graphs $\mathcal{F}$ to which the input instance should belong. Given a family of pairwise non-isomorphic 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 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 non-isomorphic 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:*

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

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 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 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 : 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 graph $\mathcal{G}$ that does not satisfy Π with the property that all its connected components $\mathfrak{G}_i$ are of size at most $r = 2^{c/\epsilon^2}$ and each of these connected component appears as an induced subgraph of H at least γn times, where $\gamma = \epsilon d/2^{r^2}$. Observe that since each connected component of H is of size at most r, each of these connected components contains at most 2^r copies of each of the connected components $\mathfrak{G}_i$ of $\mathcal{G}$. Therefore, for each $\mathfrak{G}_i$ we have that at least $\gamma n/2^r$ of the connected components of H contains an induced copy of $\mathfrak{G}_i$.

Consider now the set of distinct connected components of $\mathcal{G}$, denoted $\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\}$. Since $\mathcal{G} \notin \Pi$ we have that $m(\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\}) < \infty$ (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 first paragraph of the proof, a randomly chosen vertex belongs to a connected component of H which contains a copy of $\mathfrak{G}_i$ with probability at least $\gamma/2^r$. Therefore, by Markov's inequality a randomly chosen sample of size $10 \cdot c(r) \cdot 2^r \cdot \Psi_\Pi(r)/\gamma$ will, with probability at least $2/3$, contain $k \cdot \Psi(r)$ vertices $\{v_{i,j}\}_{1\le i\le k}^{1\le j\le \Psi_\Pi(r)}$ that belong to distinct connected component of H, with the property that for every $1 \le j \le \Psi_\Pi(r)$, the connected component of H to which $v_{i,j}$ belongs, is an induced copy of $\mathfrak{G}_i$. In particular, the graph that is obtained by taking the disjoint union of the connected components to which the vertices $v_{i,j}$ belong does not satisfy Π.

Finally, since G does not contain edges connecting vertices from distinct non-trivial connected components of H, we get that any graph that is obtained by taking the union of non-trivial connected components of H is also an induced subgraph of G. Therefore, with probability at least $2/3$ the tester will reject G. Also, from the above analysis one can see that we can set $s_2 = r = 2^{c/\epsilon^2}$ and $s_1 = 10 \cdot c(r) \cdot 2^r \cdot \Psi_\Pi(r)/\gamma$. $\square$

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[3] 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 hereditary family of graphs $\mathcal{F}$).

When does Π have a uniform tester: The proof of Theorem 1 shows that when the function Ψ_Π is computable then one can design a one-sided error uniform tester for Π. Using arguments similar to those used in [8], 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 [8], 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 $q = Q(\epsilon)$, 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 1 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 [4]. Suppose then that Π can be tested with query complexity $Q(\epsilon)$. 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 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

[3] That is, if $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ satisfy the property, then so does $G_3 = (V_1 \cup V_2, E_1 \cup E_2)$.

[4] 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 non-zero probability of rejecting this graph even though it satisfies the property.

the graph and by m the number of vertices in each cluster C_i, and observe that $m \leq r2^{\binom{r}{2}}$. 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 $m(\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\}) < \infty$ 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 $1/(4mr) \leq 1/2^{r^2}$). However, since the algorithm must find a graph that does not satisfy Π, it must ask at least $m(\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\})$ queries in order to succeed on such graphs. Therefore, $m(\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\}) \leq Q(1/2^{r^2})$ for any set $\{\mathfrak{G}_1, \ldots, \mathfrak{G}_k\} \in \Pi_r$ and by the definition of Ψ_Π this means that $\Psi_\Pi \leq Q(1/2^{r^2})$ 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 (non-uniformly) testable, if the input graph comes from a family of graphs that is hereditary and non-expanding. 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. *Topics in Discrete Mathematics*, dedicated to Jarik Nesetril on the occasion of his 60th Birthday, edited by M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas and P. Valtr, pp. 281–313.

[8] N. Alon and A. Shapira. A separation theorem in property testing. *Combinatorica*, to appear.

[9] M. A. Bender and D. Ron. Testing properties of directed graphs: acyclicity and connectivity. *Random Structures and Algorithms*, 20(2): 184–205, 2002.

[10] 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.

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

[12] A. Czumaj and C. Sohler. Abstract combinatorial programs and efficient property testers. *SIAM Journal on Computing*, 34(3): 580–615, 2005.

[13] A. Czumaj and C. Sohler. Sublinear-time algorithms *Bulletin of the EATCS*, 89: 23–47, June 2006.

[14] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the EATCS*, 75: 97–126, October 2001.

[15] 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.

[16] 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 Mathetaics and Theoretical Computer Science*, pp. 45–59, 1997. American Mathematical Society, Providence, RI, 1999.

[17] 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.

[18] O. Goldreich. Contemplations on testing graph properties. Manuscript, August 2005.

[19] O. Goldreich, S. Goldwasser and D. Ron. Property testing and its connection to learning and approximation. JACM 45(4): 653-750 (1998).

[20] O. Goldreich and D. Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.

[21] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 7, 2000.

[22] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2): 302–343, 2002.

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

[24] R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34: 57–67, 2003.

[25] T. Kaufman, M. Krivelevich, and D. Ron. Tight bounds for testing bipartiteness in general graphs. *SIAM Journal on Computing*, 33(6): 1441–1483, 2004.

[26] R. J. Lipton and R. E. Tarjan A separator theorem for planar graphs. *SIAM Journal of Applied Mathematics*, 36(2):177–189, 1979.

[27] L. Lovász and B. Szegedy. Graph limits and testing hereditary graph properties. Technical Report, MSR-TR-2005-110, Microsoft Research, August 2005.

[28] M. Parnas and D. Ron. Testing the diameter of graphs. *Random Structures and Algorithms*, 20(2): 165–183, 2002.

[29] J. L. Ramirez-Alfonsin and B. A. Reed, editors. *Perfect Graphs*. Wiley and Sons, 2001.

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/