Hierarchy Theorems for Property Testing*

Oded Goldreich[†] Michael Krivelevich[‡] Ilan Newman[§] Eyal Rozenberg[¶]

January 20, 2010

Abstract

Referring to the query complexity of property testing, we prove the existence of a rich hierarchy of corresponding complexity classes. That is, for any relevant function q, we prove the existence of properties that have testing complexity $\Theta(q)$. Such results are proven in three standard domains often considered in property testing: generic functions, adjacency predicates describing (dense) graphs, and incidence functions describing bounded-degree graphs. While in two cases the proofs are quite straightforward, the techniques employed in the case of the dense graph model seem significantly more involved. Specifically, problems that arise and are treated in the latter case include (1) the preservation of distances between graph under a blow-up operation, and (2) the construction of monotone graph properties that have local structure.

1 Background

In the last decade, the area of property testing has attracted much attention (see, e.g., a couple of recent surveys [R1, R2]). Loosely speaking, property testing typically refers to sub-linear time probabilistic algorithms for deciding whether a given object has a predetermined property or is far from any object having this property. Such algorithms, called testers, obtain local views of the object by making adequate queries; that is, the object is seen as a function and the testers get oracle access to this function (and thus may be expected to work in time that is sub-linear in the length of the object).

Following most work in the area, we focus on the query complexity of property testing, measured as a function of the size of the object as well as the desired proximity (parameter). Interestingly, many natural properties can be tested in complexity that only depends on the proximity parameter; examples include linearity testing [BLR], and testing various graph properties in two natural models (e.g., [GGR, AFNS] and [GR1, BSS], respectively). On the other hand, properties for which testing requires essentially maximal query complexity were proved to exist too; see [GGR] for artificial examples in two models and [BHR, BOT] for natural examples in other models. In between these two extremes, there exist natural properties for which the query complexity of testing is logarithmic (e.g., monotonicity [EKK+, GGL+]), a square root (e.g., bipartiteness in the bounded-degree model [GR1, GR2]), and possibly other constant powers (see [FM, PRR]).

---

*A preliminary version has appeared as TR08-097 of ECCC, and an extended abstract has appeared in the proceedings of *RANDOM'09*.

[†]Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, ISRAEL. Email: oded.goldreich@weizmann.ac.il. Partially supported by the Israel Science Foundation (grant No. 1041/08).

[‡]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, ISRAEL. Email: krivelev@post.tau.ac.il. Partially supported by a USA-Israel BSF Grant, by a grant from the Israel Science Foundation, and by Pazy Memorial Award.

[§]Department of Computer Science, Haifa University, Haifa, ISRAEL. Email: ilan@cs.haifa.ac.il. Partially supported by an Israel Science Foundation (grant number 1011/06).

[¶]Department of Computer Science, Technion, Haifa, ISRAEL. Email: eyalroz@technion.ac.il

2 Our Main Results

One natural question that arises is whether there exist properties of arbitrary query complexity. We answer this question affirmatively, proving the existence of a rich hierarchy of query complexity classes. Such hierarchy theorems are easiest to state and prove in the generic case: Loosely speaking, for every sub-linear function q, *there exists a property of functions over* $[n]$ *that is testable using* $q(n)$ *queries but is not testable using* $o(q(n))$ *queries.*

Similar hierarchy theorems are proved also for two standard models of testing graph properties: the adjacency representation model (of [GGR]) and the incidence representation model (of [GR1]). For the incidence representation model (a.k.a the bounded-degree graph model), we show that, for every sub-linear function q, *there exists a property of bounded-degree* N*-vertex graphs that is testable using* $q(N)$ *queries but is not testable using* $o(q(N))$ *queries.* Furthermore, one such property corresponds to the set of N-vertex graphs that are 3-colorable and consist of connected components of size at most $q(N)$.

The bulk of this paper is devoted to hierarchy theorems for the adjacency representation model (a.k.a the dense graph model), where the complexity is stated as a function of the number of vertices (rather than as a function of the number of all vertex pairs, which is the representation size). Our main results for the adjacency matrix model are:

1. For every sub-quadratic function q, *there exists a graph property* Π *that is testable in* q *queries, but is not testable in* $o(q)$ *queries.* Furthermore, for "nice" functions q, it is the case that Π is in $\mathcal{P}$ and the tester can be implemented in poly(q)-time.

2. For every sub-quadratic function q, there exists a *monotone* graph property Π that is testable in $O(q)$ queries, but is not testable in $o(q)$ queries.

Additional results regarding the adjacency representation model are outlined in Section 4.

Conventions. For sake of simplicity, we state all results while referring to query complexity as a function of a size parameter that is polynomially related to the object's size (i.e., in the case of generic Boolean functions the size parameter is the size of the function's domain, but in the case of graphs the size parameter is the number of vertices). In other words, we consider a fixed (constant) value of the proximity parameter, denoted ϵ. In such cases, we sometimes use the term ϵ-testing, which refers to testing when the proximity parameter is fixed to ϵ. All our lower bounds hold for any sufficiently small value of the proximity parameter, whereas the upper bounds hide a (polynomial) dependence on (the reciprocal of) this parameter. In general, bounds that have no dependence on the proximity parameter refer to some (sufficiently small but) fixed value of this parameter.

A remotely related prior work. In contrast to the foregoing conventions, we mention here a result that refers to graph properties that are testable in (query) complexity that only depends on the proximity parameter. This result, due to [AS], establishes a (very sparse) hierarchy of such properties. Specifically, [AS, Thm. 4] asserts that for every function q there exists a function Q and a graph property that is ϵ-testable in $Q(\epsilon)$ queries but is *not* ϵ-testable in $q(\epsilon)$ queries. (We note that while Q depends only on q, the dependence proved in [AS, Thm. 4] is quite weak (i.e., Q is lower bounded by a non-constant number of compositions of q), and thus the hierarchy obtained by setting $q_i = Q_{i-1}$ for $i = 1, 2, ...$ is very sparse.)

3 Our Techniques

The proofs of the hierarchy theorems for the generic case and for the incidence representation graph model, are quite straightforward. In contrast, the treatment of the dense graph model is significantly more involved. We discuss the source of trouble next.

Given that properties of maximal query complexity are known in each of the testing models that we consider, a natural idea towards proving hierarchy theorems is to construct properties that correspond to repetitions of the original properties; that is, each object in the new property consists of an adequate number of objects, each belonging to the original property. Straightforward implementations of this idea work in the generic case and in the incidence representation graph model, but not in the dense graph model. The point is that a naive repetition of a graph, in this model, necessarily creates a graph that is not dense.

Nevertheless, the graph blow-up operation does seem to be the adequate construction that we seek. Loosely speaking, the graph blow-up operation replaces each vertex by an independent set (of a predetermined size), and replaces edges by corresponding complete bipartite graphs. One source of trouble is that the blow-up operation does not necessarily preserve distances; indeed the relative distance between the blow-up of G_1 and G_2 is at most the relative distance between the original graphs, but the naive assumption that it may not be smaller is false. We overcome this difficulty by showing that for certain graphs, which we call dispersed, the blow-up does preserve the original distances (up to a constant factor).[1] Thus, we first reduce the testing of the original property to testing a corresponding property that refers to dispersed graphs. (An n-vertex graph is called dispersed if the neighbor sets of any two vertices differ on at least $\Omega(n)$ elements.)

Using dispersed graphs also allows us to overcome another technical difficulty, which relates to the complexity of our tester. In particular, the use of dispersed graphs allows us to recover the canonical labeling of an unlabeled graph, which is helpful whenever a graph property (viewed as a set of labeled graphs) is obtained by a closure under isomorphism of some set of labeled graphs (cf. [GGR]).

When trying to obtain a result for monotone graph properties, we encounter another technical difficulty. The difficulty is that standard constructions of monotone graph properties (cf. [GT]) tend to lack any local structure, since the property should be preserved under arbitrary edge additions. We demonstrate that the latter conclusion is a bit hasty, by showing that a local structure can be essentially maintained as long as the edge density does not exceed some threshold, whereas we can include in the property all graphs that have edge density that exceeds this threshold.

A third type of difficulty arises when we try to obtain one-sided error testers (see Section 4). Towards this end, we use a different type of graph blow-up, which we call *generalized blow-up*. While under the aforementioned (balanced) blow-up operation each vertex is replaced by an independent set of the same size, in a generalized blow-up these independent sets may have different sizes.

4 Additional Results

The bulk of our work provides hierarchy theorems for graph properties in the adjacency matrix model. In particular, we have already mentioned the basic hierarchy theorem regarding this model and our related theorem for monotone graph properties. (These theorems are incomparable, see discussion below.)

We also address a refined issue that has been ignored above. Specifically, we note that all our lower bounds refer to two-sided error testers, whereas the upper bounds in the generic case and in the bounded-degree graph model are demonstrated using one-sided error testers (which only make these separations stronger). In contrast, the aforementioned upper bounds for the adjacency matrix model use two-sided error testers. Seeking a hierarchy of one-sided error testing also in this model, we modify the basic construction in order to obtain one-sided error testers (while the lower bounds still hold for two-sided error testers). However, the latter theorem loses some features of the former theorems; see discussion below.

[1] Our result was superseded by Oleg Pikhurko, who showed that *for any two graphs, the distance is actually preserved up to a constant factor* [P, Sec. 4].)

We mention that our results for graph properties in the adjacency matrix model use the existence of graph properties that are in $\mathcal{P}$ and have maximal query complexity. We prove the existence of such graph properties, by building on a prior construction of [GGR], which only asserted such properties in $\mathcal{NP}$.

Discussion: Three incomparable results regarding graph properties in the adjacency matrix model. As mentioned above, we proved three hierarchy theorems for testing graph properties in the adjacency matrix model.

1. The basic theorem is established by non-monotone properties (in $\mathcal{P}$), while the tester demonstrating the upper bound is relatively efficient in the sense that its running time is polynomial in its query complexity. Both the lower and upper bounds refer to two-sided testers.
2. The second theorem refers to monotone properties (in $\mathcal{NP}$). Again, both the lower and upper bounds refer to two-sided testers.
3. The third theorem refers to properties in $\mathcal{P}$, but the tester demonstrating the upper bound is not relatively efficient (i.e., its decision predicate is in $\mathcal{NP}$). However, in this case the tester has one-sided error (whereas the lower bound holds also for two-sided testers).

Obtaining a single theorem that combines all good features is left as an open problem.

Acknowledgments

We are grateful to Ronitt Rubinfeld for asking about the existence of hierarchy theorems for the adjacency matrix model. Ronitt raised this question during a discussion that took place at the Dagstuhl 2008 workshop on sub-linear algorithms. We are also grateful to Arie Matsliah, Dana Ron, and Yoav Tzur for helpful discussions. In particular, we thank Arie Matsliah for providing us with a proof that the blow-up operation does not preserve distances in a perfect manner.

References

[ABI] N. Alon, L. Babai and A. Itai. A fast and Simple Randomized Algorithm for the Maximal Independent Set Problem. *J. of Algorithms*, Vol. 7, pages 567–583, 1986.

[AFKS] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy. Efficient Testing of Large Graphs. *Combinatorica*, Vol. 20, pages 451–476, 2000.

[AFNS] N. Alon, E. Fischer, I. Newman, and A. Shapira. A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity. In *38th STOC*, pages 251–260, 2006.

[AGHP] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Journal of Random structures and Algorithms*, Vol. 3 (3), pages 289–304, 1992.

[AS] N. Alon and A. Shapira. Every Monotone Graph Property is Testable. *SIAM Journal on Computing*, Vol. 38, pages 505–522, 2008.

[BSS] I. Benjamini, O. Schramm, and A. Shapira. Every Minor-Closed Property of Sparse Graphs is Testable. In *40th STOC*, pages 393–402, 2008.

[BLR] M. Blum, M. Luby and R. Rubinfeld. Self-Testing/Correcting with Applications to Numerical Problems. *JCSS*, Vol. 47, No. 3, pages 549–595, 1993.

[BHR] E. Ben-Sasson, P. Harsha, and S. Raskhodnikova. 3CNF Properties Are Hard to Test. *SIAM Journal on Computing*, Vol. 35 (1), pages 1–21, 2005.

[BOT] A. Bogdanov, K. Obata, and L. Trevisan. A lower bound for testing 3-colorability in bounded-degree graphs. In *43rd FOCS*, pages 93–102, 2002.

[EKK+] F. Ergun, S. Kannan, S. R. Kumar, R. Rubinfeld, and M. Viswanathan. Spot-checkers. *JCSS*, Vol. 60 (3), pages 717–751, 2000.

[FM] E. Fischer and A. Matsliah. Testing Graph Isomorphism. In *17th SODA*, pages 299–308, 2006.

[GGL+] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samorodnitsky. Testing Monotonicity. *Combinatorica*, Vol. 20 (3), pages 301–337, 2000.

[GGR] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, pages 653–750, July 1998.

[GR1] O. Goldreich and D. Ron. Property Testing in Bounded Degree Graphs. *Algorithmica*, Vol. 32 (2), pages 302–343, 2002.

[GR2] O. Goldreich and D. Ron. A Sublinear Bipartitness Tester for Bounded Degree Graphs. *Combinatorica*, Vol. 19 (3), pages 335–373, 1999.

[GT] O. Goldreich and L. Trevisan. Three theorems regarding testing graph properties. *Random Structures and Algorithms*, Vol. 23 (1), pages 23–57, August 2003.

[LNS] O. Lachish, I. Newman, and A. Shapira. Space Complexity vs. Query Complexity. *Computational Complexity*, Vol. 17, pages 70–93, 2008.

[NN] J. Naor and M. Naor. Small-bias Probability Spaces: Efficient Constructions and Applications. *SIAM J. on Computing*, Vol 22, 1993, pages 838–856.

[PRR] M. Parnas, D. Ron and R. Rubinfeld. Testing Membership in Parenthesis Laguages. *Random Structures and Algorithms*, Vol. 22 (1), pages 98–138, 2003.

[P] O. Pikhurko. An Analytic Approach to Stability. Manuscript, 2009. Available from `http://arxiv.org/abs/0812.0214`

[R1] D. Ron. Property Testing: A Learning Theory Perspective. *Foundations and Trends in Machine Learning*, Vol. 1 (3), pages 307–402, 2008.

[R2] D. Ron. Algorithmic and Analysis Techniques in Property Testing. *Foundations and Trends in TCS*, to appear.

[RS] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2), pages 252–271, 1996.

[S] R. Shaltiel. Recent Developments in Explicit Constructions of Extractors. In *Current Trends in Theoretical Computer Science: The Challenge of the New Century, Vol 1: Algorithms and Complexity*, World scietific, 2004. Preliminary version in *Bulletin of the EATCS 77*, pages 67–95, 2002.