A Finite-Compactness Notion, and Property Testing (extended abstract)

Richard Lipton
Georgia Institute of Technology

Kenneth W. Regan
University at Buffalo

June 1, 2011

Abstract

We investigate properties Π of graphs G of size N such that possession of Π by all $f(N)$-node subgraphs implies possession of Π by some $g(N)$-node subgraph. Bounds are given on $f(N)$ and $g(N)$ for bipartiteness. This is compared to a situation in property testing, where edge-induced subgraphs are involved and different, intuitively "better," bounds are obtained. Bounds are discussed for other properties, and possible links to logical formulas describing the properties are raised.

1 Introduction

The classic notion of compactness in logic states roughly that if some property Π holds for all finite sub-structures of an infinite structure Ω, then it holds for all of Ω. This is a powerful notion that is true for first order logic—it is a corollary of Gödel's Completeness Theorem. Let us talk about this for structures that are undirected graphs.

Our rough idea is that even for finite objects there can be a kind of compactness. The idea is that if all *small* subgraphs have property Π, then there is a *large* subgraph that has Π. Here a size N intuitively thought of as being exponential in a complexity parameter n (e.g., $N = 2^n$) plays the role of "infinite," while "small" can mean polynomial or linear in n, or even finite independent of N. "Large" can mean $\Theta(N)$ as N grows, or $\geq (1-\epsilon)N$ where ϵ is an adjustable parameter, or even $N - o(N)$. The subgraphs are vertex-induced, unless otherwise indicated.

Here is a general definition, for arbitrary functions $f(N)$ standing for "small" and $g(N) \gg f(N)$ standing for "large":

Definition 1.1. *A graph property* Π *is* $(f(N), g(N))$-compact *if for all* N *and all graphs* G *of size* N*, if all* $f(N)$*-node subgraphs of* G *have* Π*, then there is a* $g(N)$*-node subgraph that has* Π.

To avoid caring how f and g are defined for small N, we can relax this to apply for all sufficiently large N.

Which interesting properties are $(f(N), g(N))$-compact, for natural functions f and g? This falls under the general heading of *local-global* (graph) properties, some results from which are itemized below. Several of those, however, allow the properties of small and large subgraphs to differ.

The notion can be extended to any reasonable family of mathematical objects defined as formal structures on a ground set Ω. Our work was initially motivated by the general topic of *property testing*, but for graphs the results yield large edge-induced not vertex-induced subgraphs. Our emphasis on the latter case appears to be fairly novel, and the examples that follow argue that it is non-trivial.

2 Some Background

Our idea falls under the more-general heading of "local-global graph properties." Here the properties need not be the same. A recent forum discussion [Zai09] enumerated the following examples and results:

(a) If $(\forall u \neq v \in V)(\exists! w)[E(u, w) \wedge E(v, w)]$, then $(\exists u)(\forall v \neq u)E(u, v)$.

(b) If every k-node subgraph is bipartite, then G can be colored with $N^{O(1/k)}$ colors (see [Lin93]).

(c) If every k-node subgraph is 3-colorable, then G can be colored with $n^{1/2+r(k)}$ colors, where $r(k) \longrightarrow 0$ as k grows.

(d) If $h : V \longrightarrow \mathbb{R}^+$ has average value at least μ on $N_t(v)$ for all $v \in V$ and $t \leq r$, $t \geq 1$, then its average on V is at least $\mu / n^{O(1/\log r)}$ (see [Lin93]).

(e) If G is s-connected and has no independent sets of size $s + 1$, then G has a Hamiltonian circuit [CE72].

Further examples are itemized there and in the blog post [Bro10]. Besides the local and global properties being different in these examples, the emphasis is on a property of all of G rather than a large subgraph of G. Thus our notion is really one of a local/*almost*-global property.

Our prime example takes Π to be the property of 2-colorability, i.e., bipartiteness.

3 Example: Bipartiteness

If all k-node subgraphs are bipartite, does it follow that some $(1 - \epsilon)N$-node subgraph is bipartite? Here k may depend on N and ϵ, but ϵ is fixed for all N. The following positive and negative results, communicated to us by Noga Alon [Alo11] and Luca Trevisan [Tre11], act as bookends on some possible values, asymptotically.

Theorem 3.1 ([Alo11]). *For every $\epsilon > 0$ there exists $C > 0$ depending only on ϵ such that bipartiteness is $(C \log N, (1 - \epsilon)N)$-compact. Moreover, we can construct the subgraph of size $(1 - \epsilon)N)$ in $N^{O(1)}$ time.*

Proof. Fix $\epsilon > 0$; fix $C > 0$ later, and let $k = \frac{1}{2}C \log_2 N$. A graph is non-bipartite if and only if it has an odd-length cycle. Hence if all k-node subgraphs of G are bipartite, then G has no odd cycles of length k or less.

To construct a bipartite subgraph H, start with any vertex v. For each $i \geq 1$ define $N_i(v)$ to be the neighborhood of vertices within i steps of v, and $S_i = N_i \setminus N_{i-1}$ to be the "shell" of those at exactly distance i. Now consider the least j such that

$$|S_j| < \epsilon |N_j|.$$

If $j > k$, then we have for all $i \leq k$, $|S_i| \geq \epsilon |N_i|$, so $|S_i| \geq \epsilon(|S_i| + |N_{i-1}|)$, so $|S_i| \geq (1/(1-\epsilon))|N_{i-1}|$, which in turn trivially implies $|S_i| \geq (1/(1-\epsilon))|N_{i-1}|$. This implies

$$|S_i| \geq (\frac{1}{1-\epsilon})^k,$$

which is $> N$ when $\frac{C}{2} \log_2 N \log_2(\frac{1}{1-\epsilon}) > \log_2 N$, so when

$$C > \frac{2}{\log_2(\frac{1}{1-\epsilon})}.$$

Fixing such C thus makes $j \leq k$. Now we observe that every S_i for $i < j$ is an independent set. If it has an edge (s_i, t_i), then the paths of length i from s_i and t_i back to v come together at v or some earlier node in a way that forms an odd cycle of length at most $2i+1 \leq C \log_2 n$, contradicting the assumption. It follows that the subgraph H_j induced by $N_j \setminus S_j$ is bipartite, since the S_i give the 2-coloring. Putting $n = |N_j|$, note that $|S_j| \leq \epsilon n$. By induction we have that the leftover graph induced by $V(G) \setminus N_j$ has a bipartite subgraph H' of size at least $(1-\epsilon)(N-n)$. Since $H_j \cup H'$ is separated, it is bipartite as well, and has size at least $(1-\epsilon)N$. Clearly this induction yields a polynomial-time algorithm. □

Theorem 3.2 ([Tre11]). *If $f(N)$ is such that bipartiteness is $(f(N), (1-\epsilon)N)$-compact, then $f(N) = \Omega(\log N)$.*

Proof. There exist size-N expanders G of girth $k = \Omega(\log N)$, so certainly any odd cycle has that size, thus all subgraphs of size $k-1$ are bipartite. By the expander mixing lemma (see [Tre08]), G cannot have even a bipartite subgraph of size $> N/2$, for large enough N. □

If we allow edge-induced subgraphs, then the resulting relaxed idea is related to *property testing*.

4 Relation to Property Testing

Property testing started explicitly with Blum, Luby, and Rubinfeld [BLR93], and since then there has been a vast literature of what properties can be tested and which cannot. A *tester* is a randomized algorithm A that probes edges of the graph, such that:

1. If G has the property Π, then after the probing the algorithm A says 'yes' with probability at least 2/3. In the *one-sided* model, the algorithm always says 'yes' in this case.

2. If the graph does not have Π, and is not "near" a graph in Π, then A returns a 'no' with probability at least 2/3.

Here "near" is defined via a parameter $\epsilon > 0$ that is also given to the algorithm. Two graphs G, G' are *ϵ-close* if they have the same size N and for all but $\epsilon\binom{N}{2}$ pairs $1 \leq i < j \leq N$, $E(i,j) \leftrightarrow E'(i,j)$. The *query complexity* q of the tester is the maximum number of edge probes, and is always a function of ϵ. Sometimes it is a function of N, or in case of an exponential-sized graph with $N = 2^n$, a function of n usually bounded by $n^{O(1)}$.

Our intent is that the hypothesis of (k, ϵ)-compactness leads the tester A to accept G, which implies that G is ϵ-close to a graph G' with property Π, and then use the existence of G' to produce H. There are two obstacles:

- The tester A need not work by probing small subgraphs for the property Π itself.
- The closeness condition counts edges, and might not carry over to *vertex*-induced subgraphs.

We address the former issue by strengthening a notion called "canonical" in [Gol10], with reference to [GT03].

Definition 4.1. *A property testing algorithm A for Π is* **cognizant** *if it generates one or more vertex-induced subgraphs H of G, probing only the edges in H, and accepts if and only if the majority of the probed graphs have property Π.*

A property Π is *closed under induced subgraphs* if whenever G has Π and H is a vertex-induced subgraph of G, then H has Π. Bipartiteness is such a property; indeed it is closed under edge-induced subgraphs as well. So is k-colorability for any k. Ignoring some polynomial slippage in bounds, the following is credited to Noga Alon in Appendix D of [GT03].

Theorem 4.2 (Alon cited in [GT03])**.** *Every testable property that is closed under induced subgraphs has a cognizant tester.*

This yields a weak "edge-induced" notion of finite compactness. An N-vertex graph is *dense* if it has δN^2 edges, where we intend $\delta > \epsilon$ and fixed. For the property of bipartiteness the weaker version of compactness is achieved this way:

Theorem 4.3. *There is an absolute constant $c > 0$ such that for all dense graphs G and sufficiently large k, if all k-node subgraphs of G are bipartite, then by deleting at most $(1/k^c)N^2$ edges we can obtain a bipartite graph H.*

Proof. Given $\epsilon > 0$, the cognizant tester in Oded's survey (Algorithm 2.3) selects $k = O(\log(1/\epsilon)/\epsilon^2)$ vertices uniformly at random, and accepts iff the subgraph R they induce is bipartite. This makes $1/k^3 < \epsilon$. Suppose G is a graph for which all k-node subgraphs are bipartite. Then the tester accepts with certainty. By definition of being a tester, G is near a graph G' in Π, which here entails that ϵN^2 edges can be deleted from G to yield H. (In fact, as observed in a comment after the proof, a suitable H can be described succinctly in terms of choices that accompany R.) □

The bounds are stronger than those in Theorem 3.1, insofar as k is constant here. The issue, however, is that H need not be a vertex-induced subgraph. Indeed, ϵN^2 edges can easily be incident on all N vertices, in such a way that all $(1 - \epsilon)N$-vertex induced subgraphs H have many of the bad edges from G. Theorem 3.2 above shows ithat Theorem 4.3 cannot carry over for vertex-induced subgraphs.

5 Some Other Properties

After bipartiteness, a natural next property Π to consider is 3-colorability. The known testers have polynomial-in-$(1/\epsilon)$ query complexity but exponential running time, and improvements would have unlikely consequences for NP (see [Gol10]). Per item (c) in Section 2 above, the known local-global behavior is not great, and the sources cited there show that basic questions in that line are wide open. Here we note that a result analogous to Theorem 3.1 for 3-colorability would have a consequence deemable unlikely.

Theorem 5.1. *If 3-colorability is $(f(N), \Theta(N))$-compact, any f, with a (random) polynomial-time algorithm for finding a $\Theta(N)$-sized subgraph and 3-coloring it, then 3-colorable graphs can be colored with $O(\log N)$ colors in (random) polynomial time.*

Proof. Removing the 3-colored subgraph always shrinks the graph by a constant factor, and since we can use fresh colors for the rest, the iteration uses $O(\log N)$ colors overall. □

To see why unlikely, Chlamtac [Chl07] showed how to color any given 3-colorable graph G in $O(N^{0.2072})$ colors, while Guruswami and Khanna [GK04] showed that it is NP-hard to find a 4-coloring. To our knowledge these are still the best upper and lower bounds of their kind. Zuckerman [Zuc07] showed that for all $\epsilon > 0$, approximating the chromatic number of a graph to wihin a factor of $N^{1-\epsilon}$ is NP-hard, and this seems to be reason to suspect that $O(N^c)$ for some fixed $c < 0.2072$ should be a lower bound, but the consequence doesn't immediately apply to 3-colorable graphs.

Similar unlikelihood considerations apply to k-coloring for $k > 3$ and other properties for which inapproximability results are known. We speculate that our finite-compactness notion may provide a way to sharpen the considerations about exact bounds here.

Finally we ask about the relation to logical formulas defining the properties. If Π is defined by a first-order sentence $\phi = (\forall x_1, \ldots, x_k)(\exists \ldots) \ldots$ (without constants), and ϕ holds for all k-node subgraphs, then it is true for the whole graph. However, it seems hard to say more than this in short order, or to make direct use of the weaker goal of needing ϕ to be true only of a large subgraph, in relation to the formula's structure.

6 Conclusions and Prospects

We have introduced a finite-compactness notion that appears to be new, at least in terms of attracting emphasis in its own right. It was motivated by property testing, but appears to involve different technical considerations—mainly in vertex-induced versus edge-induced subgraphs. We have given examples such as k-colorability for $k = 2$ and $k > 2$.

The main open question is to classify further properties that are finitely compact, with a high spread between $f(N)$ and $g(N)$. Can we show further dependence on the structure of logical formulas that define the property? Are there further relations to property testing?

Acknowledgments We thank Luca Trevisan and Gilbert Bernstein for helpful comments in the May 25, 2011 *Gödel's Lost Letter and P=NP* blog item, from which this extended abstract is adapted. We also thank Noga Alon, Terence Tao, Peter Bro Miltersen, and Peter Cameron for private communications.

References

[Alo11] N. Alon. Personal communication, 2011.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993.

[Bro10] H. Brown. What's a "locally determined graph property?". http://harrisonbrown.wordpress.com/2010/01/01/whats-a-locally-determined-graph-property/, 2010.

[CE72] V. Chvátal and P. Erdős. A note on Hamiltonian paths. *Discrete Mathematics*, 2:111–113, 1972.

[Chl07] E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 691–701, 2007.

[GK04] V. Guruswami and S. Khanna. On the hardness of 4-coloring a 3-colorable graph. *SIAM J. Disc. Math*, 18:30–40, 2004.

[Gol10] O. Goldreich. Introduction to testing graph properties. Technical Report TR10-082, Electronic Colloquium on Computational Complexity (ECCC), May 2010.

[GT03] O. Goldreich and L. Trevisan. Three theorems regarding testing graph properties. *Random Structures and Algorithms*, 23:23–57, 2003.

[Lin93] N. Linial. Local-global phenomena in graphs. In *Combinatorics, Paul Erdős is Eighty*, Bolyai Society Mathematical Studies, pages 491–503. North-Holland, Amsterdam, 1993.

[Tre08] L. Trevisan. CS 278 – Computational Complexity – Spring 2008, Lecture 11: Eigenvalues and random walks; the expander mixing lemma. http://www.cs.berkeley.edu/ luca/cs278-08/lecture11.pdf, 2008.

[Tre11] L. Trevisan. Comment in Gödel's Lost Letter item. http://rjlipton.wordpress.com/2011/05/25/finite-compactness/#comment-11977, 2011.

[Zai09] G. Zaimi. Local-global approach to graph theory. http://mathoverflow.net/questions/10241/local-global-approach-to-graph-theory, 2009. MathOverflow item.

[Zuc07] D. Zuckerman. Linear degree extractors and inapproximability of Max Clique and Chromatic Number. *Theory of Computing*, 3:103–128, 2007.