ELSEVIER

Information Processing Letters 58 (1996) 39–42

Information Processing Letters

On proving that a graph has no large clique: A connection with Ramsey theory

Richard J. Lipton [1]

Department of Computer Science, Princeton University, Princeton, NJ 08544, USA

Received 4 February 1995; revised 14 November 1995
Communicated by L.A. Hemaspaandra

Keywords: Computational complexity; Clique approximation; Ramsey theory

1. Introduction

Recently, there have been a number of results that show that certain NP-hard problems cannot be approximated in polynomial time [2,3,8,11,12,16]. For example, it is now known that, if there is a polynomial-time function that approximates the size of the largest clique of G, $\omega(G)$, within a factor of n^{ε_0}, then P = NP. There are upper bounds on ε_0; but the best value is unknown.

Our main result is the observation that there is a fundamental connection between these results and certain Ramsey-like results [7,10,13]. This seems to be surprising. Roughly the connection shows that if certain approximation problems are NP-hard, then either co-NP has "feasible" proofs or certain Ramsey results are true. It is, of course, believed that co-NP does not have polynomial size proofs. However, by "feasible" proofs we mean proofs that are at most $n^{O(\log(n))}$ long. This is still a reasonable assumption.

These Ramsey results are interesting in their own right. Moreover, they *appear* to be difficult to resolve independently [15]. Of course, if they are eventually shown to be false, then we would have either proved that co-NP has feasible proofs or that certain approximation problems are not NP-hard. Either case would be interesting.

In order to state our results, it is natural to introduce a new concept: the notion of an NP-hard *pair* of sets. This new definition allows us to simply state the known results about hard approximation problems. Further, it makes the connection with Ramsey theory easy to establish. The notion of an NP-hard pair is related to an old concept from recursion theory of effectively *inseparable* sets [14].

2. Definitions

We need a "Ramsey"-like function [13]. It is essentially the functional inverse of the usual one. For a graph G, define $R(G)$ to be the size of the largest "monochromatic" clique that must appear in any 2-coloring of the edges of G. Recall that a monochromatic clique is one that has all its edges colored in the same way. Clearly, $R(G)$ is at most $\omega(G)$ since the maximum is over the cliques of G. The usual Ramsey theorem is just the statement that $R(K_m)$ is unbounded as a function of m. K_m denotes

[1] Email: rjl@cs.princeton.edu.

0020-0190/96/$12.00
SSDI 0020-0190(95)00209-X

the complete graph on m vertices. It is known that $R(K_m)$ is $\Theta(\log(m))$ [10].

We can as usual extend this Ramsey function to l-colorings of t cliques instead of 2-colorings of 2 cliques, i.e. edges. A monochromatic clique is now required to have all its t subsets colored in the same manner. Then, define $R_l^{(t)}(G)$ to be the size of the largest monochromatic clique in any l-coloring of its t cliques. Clearly, $R(G) = R_2^{(2)}(G)$. Again $R_l^{(t)}(K_m)$ is known to be unbounded as a function of m. It grows at most $O(\log(m))$. Even stronger, for t large it behaves like an iterated logarithm [10].

As usual NP denotes the class of nondeterministic polynomial-time sets and co-NP denotes those sets whose complements are in NP. SAT is the NP-complete set of satisfiable formulas in disjunctive normal form. Also $NTIME(t(n))$ is the sets accepted in nondeterministic time $t(n)$ [4,5]. Say that a set is $f(n)$-*sparse* if for each n, there are at most $f(n)$ elements in the set of size n. Let $NTIME(t(n))/f(n)$ denote the sets accepted by nondeterministic machines that run in time $t(n)$ and have access to an oracle that is $f(n)$ sparse. Let $\mathcal{N}$ denote the sets in

$$NTIME(n^{O(\log(n))})/(\log\log(n)).$$

Our assumption is that co-NP does *not* lie in $\mathcal{N}$. Given our current understanding of complexity theory it would certainly be surprising if this assumption was false.

Finally, say that a pair of disjoint sets (A, B) is an NP-*hard pair* provided for any L in NP there is a polynomial-time computable function g so that the following are true:

(1) for each $x \in L$, $g(x) \in A$;

(2) for each $x \notin L$, $g(x) \in B$.

Note, if L is NP-complete in the usual sense of many-one polynomial-time reduction, then $(L, \bar{L})$ is an NP-hard pair. The new approximation results are statements that certain pairs are NP-hard. Let $s(n)$ and $b(n)$ be functions. Define the following sets:

$$S_{\leqslant s(n)} = \{G \mid \omega(G) \leqslant s(n),\ n = |G|\}$$

and

$$S_{\geqslant b(n)} = \{G \mid \omega(G) \geqslant b(n),\ n = |G|\}.$$

Then, it is of course classic that $(S_{\geqslant n/3}, S_{\leqslant n/2-1})$ is an NP-hard pair. On the other hand the new results [2,3,8,11,12,16] show that $(S_{\geqslant b(n)}, S_{\leqslant s(n)})$ is an NP-hard pair for $b(n) = n^{\delta_1}$ and $s(n) = n^{\delta_2}$ where $\delta_1 > \delta_2$ are known constants.

3. The main result

Our main result assumes, as stated earlier, that co-NP does not lie in the class $\mathcal{N}$.

Main Theorem. *Suppose that $(S_{\geqslant b(n)}, S_{\leqslant s(n)})$ is an* NP-*hard pair for some functions $b(n)$ and $s(n)$. Also suppose that $l > 1$ and t are fixed. Then, for arbitrarily large n, there are graphs G_n on n vertices so that $\omega(G_n) \leqslant s(n)$ and $R_l^{(t)}(G_n) \geqslant R_l^{(t)}(K_{b(n)})$.*

Thus, the graphs G_n have only small cliques yet from the point of view of Ramsey theory they "look" like they have large cliques.

Before we proceed with the proof of this theorem we will first prove a number of simple facts about NP-hard pairs.

Lemma 1. *The pair of disjoint sets (A, B) is* NP-*hard if and only if there is a polynomial-time computable function g such that the following are true:*

(1) *for each $x \in$ SAT iff $g(x) \in A$;*

(2) *for each $x \notin$ SAT iff $g(x) \in B$.*

Proof. Routine from the definition. □

Define $A \sqsubseteq B$ provided $A - B$ is finite, i.e., provided all but a finite number of elements from A are in B.

Lemma 2. *Suppose that (A, B) is an* NP-*hard pair. Also suppose that L is a non-empty set such that $A \cap L = \emptyset$ and $B \sqsubseteq L$. Then, (A, L) is an* NP-*hard pair.*

Proof. Just modify the reduction to avoid the finite part of B that is not contained in L. □

Lemma 3. *Suppose that (A, B) is* NP-*hard and that B is in $\mathcal{N}$. Then,* co-NP *is contained in $\mathcal{N}$.*

Proof. Follows directly from the definition of $\mathcal{N}$ and from Lemma 1. □

Proof of the theorem. The key is the set L defined by

$$L = \{G \mid R_l^{(t)}(G) < R_l^{(t)}(K_{b(n)}),\ n = |G|\}.$$

We first claim that L is in $\mathscr{N}$. Suppose that G is an n-vertex graph. Guess an l-coloring of the t cliques of G. Since t is fixed this can be done in polynomial time. Now assume that $r = R_l^{(t)}(K_{b(n)})$ is known. Since it is at most $O(\log(n))$ this can be encoded in an oracle that is $\log\log(n) + O(1)$ sparse. Find the largest monochromatic clique of size $\leqslant r$. Let its size be q. If $q < r$, then conclude that G is in L. Clearly, this last step can be done in time at most $n^{O(r)}$ which is at most $n^{O(\log(n))}$. Thus, L is in $\mathscr{N}$.

We claim next that $(S_{\geqslant b(n)}, L)$ is not an NP-hard pair. If it was, then by Lemma 3, it follows that co-NP is in $\mathscr{N}$. Since we have assumed that this is false, it follows that it is not an NP-hard pair.

Clearly, $S_{\geqslant b(n)}$ and L must be disjoint: if G has a clique of size at least $b(n)$, then $R_l^{(t)}(G) \geqslant R_l^{(t)}(K_{b(n)})$. Thus, $(S_{\geqslant b(n)}, L)$ satisfies the first part of the definition of an NP-hard pair. Thus, by Lemma 2, $S_{\leqslant s(n)} \sqsubseteq L$ must be false. Otherwise, $(S_{\geqslant b(n)}, L)$ would be an NP-hard pair. Thus, there must exist an infinite sequence of graphs $G_{n_1}, G_{n_2}, \ldots$ so that $\omega(G_{n_i}) \leqslant s(n_i)$ and $R_l^{(t)}(G_{n_i}) \geqslant R_l^{(t)}(K_{b(n_i)})$. This proves the theorem. □

4. Extensions

It is possible to extend the theorem so that the graphs G "work" for more than one $l > 1$ and t. It is easy to see that the following is true:

Theorem. *Suppose that $(S_{\geqslant b(n)}, S_{\leqslant s(n)})$ is an NP-hard pair for some functions $b(n)$ and $s(n)$. Then, for arbitrarily large n, there are graphs G_n on n vertices so that for all $1 < l < O(\log^*(n))$, and $t < O(\log^*(n))$, $\omega(G_n) \leqslant s(n)$ and $R_l^{(t)}(G_n) \geqslant R_l^{(t)}(K_{b(n)})$.*

The proof is the same except that now we must assume that co-NP does not lie in

$$NTIME(n^{O(\log(n))})/O(\log^*(n)\log\log(n)).$$

This theorem is clearly much stronger: a graph G must work for all nontrivial colorings of edges or triangles and so on up to cliques of size $O(\log^*(n))$.

One weakness of these results is that we have only proved that there is a graph G_n for an infinite number of n's. It is possible to improve this by making a stronger assumption about co-NP. The assumption we now need is that if TAUT (the set of unsatisfiable formulas) is accepted by a Nondeterministic Turing Machine, then it must take time at least $n^{\Omega(\log(n))}$ for an infinite number of n. These machines have access to an $O(\log^*(n)\log\log(n))$ sparse set. Under this assumption it follows that the following is true:

Theorem. *Suppose that $(S_{\geqslant b(n)}, S_{\leqslant s(n)})$ is an NP-hard pair for some functions $b(n)$ and $s(n)$. Then, for all n large enough, there are graphs G_n on n vertices so that for all $1 < l < O(\log^*(n))$, and $t < O(\log^*(n))$, $\omega(G_n) \leqslant s(n)$ and $R_l^{(t)}(G_n) \geqslant R_l^{(t)}(K_{b(n)})$.*

5. Discussion

The central question is: What do these results "mean"? What do they say about the relationship between Ramsey Theory and Complexity Theory?

It seems interesting that the Ramsey results are "pure" questions of combinatorics. They do not encode computation in any obvious way. Indeed, they are quite natural questions. For example, it is *known* that for any k there exists a graph G on n vertices so that $\omega(G) = k$ and $R(G) = k$. Thus, there is a graph with only small cliques of size k, yet in any 2-coloring of the edges of G, there is a monochromatic clique of size k. The problem with the known constructions of such graphs is that they appear to require very large n as a function of k [10]. Thus, they are not counterexamples to our Ramsey results.

Thus, the key question is whether or not the Ramsey results are true without any complexity hypothesis. If they are shown to be false, then our results would be very important. Either, they would show that our complexity assumptions about co-NP are false or that certain approximation problems are not intractable. Either would be quite interesting.

Thus, the critical question that must be resolved is whether or not the Ramsey questions have a simple direct proof, i.e. whether or not there are graphs

graphs G_n on n vertices so that $\omega(G_n) \leqslant s(n)$ and $R_l^{(t)}(G_n) \geqslant R_l^{(t)}(K_{b(n)})$. Szemeredi [15] has remarked that the general case of this question where $t > 1$ appears to be quite difficult. The reason is that our understanding of such Ramsey functions is fairly weak. The current situation is that for such Ramsey functions there is an exponential gap between the upper and the lower bounds [10]. This gap may make proving that these graphs do not exist difficult or even impossible. Recently, Alon [1] has also stated that such results may be quite impossible to prove or disprove.

The situation is somewhat different for $t = 2$. Here the upper and lower bound on the Ramsey function is tighter: it is known that $R(K_m)$ grows at least as fast as $c_1 \log(m)$ and at most as fast as $c_2 \log(m)$ for constants $c_1 < c_2$. However, even this smaller gap should make resolving the Ramsey questions for $b(n)$ much larger than $s(n)$ difficult.

References

[1] N. Alon, Private communication, November 1993.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy, Proof verification and intractability of approximation problems, in: *Proc. 33rd IEEE Symp. on Foundations of Computer Science* (1992).

[3] S. Arora and S. Safra, Approximating clique is NP-complete, in: *Proc. 33rd IEEE Symp. on Foundations of Computer Science* (1992).

[4] L. Babai, L. Fortnow and C. Lund, Non-deterministic exponential time has two-prover interactive protocols, in: *Proc. FOCS 90*, pp. 16–25.

[5] L. Babai, L. Fortnow, L. Levin and M. Szegedy, Checking computations in polylogarithmic time, in: *Proc. STOC 91*, pp. 24–31.

[6] M. Bellare, S. Goldwasser, C. Lund and A. Russell, Efficient probabilistically checkable proofs and applications to approximation; in: *Proc. 25th ACM Symp. on The Theory of Computing* (1993).

[7] P. Erdös, *The Art of Counting* (edited by J. Spencer) (MIT Press, Cambridge, MA, 1973).

[8] U. Frige, S. Goldwasser, L. Lovasz, S. Safra and M. Szegedy, Approximating clique is almost NP-complete, in: *Proc. FOCS 91*.

[9] M.R. Garey and D.S. Johnson, The complexity of near-optimal graph coloring; *J. ACM* **23** (1976) 43–49.

[10] R.L. Graham and B.L. Rothschild, Some recent developments in Ramsey theory, in: M. Hall, Jr. and J.H. van Lint, eds., *Combinatorics* (Reidel, Boston, MA, 1975) 261–276.

[11] C. Lund and M. Yannakakis, On the hardness of approximating minimization problems, in: *Proc. STOC 93*.

[12] C. Lund and M. Yannakakis, The approximation of maximum subgraph problems, in: *Proc. ICALP 93*.

[13] F.P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.* **30** (193) 264–286.

[14] H. Rogers, *Theory of Recursive Functions and Effective Computation* (McGraw-Hill, New York, 1967).

[15] E. Szemeredi, Private communication, September 1993.

[16] D. Zuckerman, NP-complete problems have a version that is hard to approximate, *Structure in Complexity Theory* (1993).