Independence numbers of locally sparse graphs and a Ramsey type problem

Noga Alon *

Abstract

Let $G = (V, E)$ be a graph on n vertices with average degree $t \geq 1$ in which for every vertex $v \in V$ the induced subgraph on the set of all neighbors of v is r-colorable. We show that the independence number of G is at least $\frac{c}{\log(r+1)} \frac{n}{t} \log t$, for some absolute positive constant c. This strengthens a well known result of Ajtai, Komlós and Szemerédi. Combining their result with some probabilistic arguments, we prove the following Ramsey type theorem, conjectured by Erdös in 1979. There exists an absolute constant $c' > 0$ so that in every graph on n vertices in which any set of $\lfloor \sqrt{n} \rfloor$ vertices contains at least one edge, there is some set of $\lfloor \sqrt{n} \rfloor$ vertices that contains at least $c' \sqrt{n} \log n$ edges.

1 Introduction

All graphs considered here are finite and undirected, and contain no loops and no parallel edges. An easy, well known consequence of Turán's theorem asserts that any graph on n vertices with average degree t contains an independent set of size at least $n/(t+1)$. Ajtai, Komlós and Szemerédi [1] proved that this can be improved for triangle-free graphs; if G is triangle-free and has n vertices and average degree t (> 1) then its independence number is at least $c \frac{n}{t} \log t$ for some absolute positive constant c. Shearer [7] gave a simpler proof (and also improved the constant c). (All logarithms here and in the rest of the paper are in base 2, unless otherwise specified.) Ajtai, Erdös, Komlós and Szemerédi [2] showed that for any $r > 2$, the independence number of any K_{r+1}-free graph on n vertices and average degree t is at least $c(r) \frac{n}{t} \log \log(t+1)$ and conjectured that in fact it is at least

$$c(r) \frac{n}{t} \log t. \tag{1}$$

This conjecture is still open, despite some recent progress by Shearer [8], who improved the estimate in [2] to $c'(r) \frac{n}{t} \frac{\log t}{\log \log(t+1)}$. Note that a K_{r+1}-free graph is a graph in which the neighborhood of any

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: noga@math.tau.ac.il. Research supported in part by a USA Israeli BSF grant and by the Fund for Basic Research administered by the Israel Academy of Sciences.

vertex is K_r-free. A stronger assumption is that each such neighborhood is r-colorable. Our first result here is that under this stronger requirement, the independence number is indeed as large as in (1).

Theorem 1.1 *Let $G = (V, E)$ be a graph on n vertices with average degree $t \geq 1$ in which for every vertex $v \in V$ the induced subgraph on the set of all neighbors of v is r-colorable. Then, the independence number of G is at least $\frac{c}{\log(r+1)} \frac{n}{t} \log t$, for some absolute positive constant c.*

Although this is weaker than the conjecture of [2] mentioned above, it is clearly stronger than the main result of [1] and may indicate that this conjecture is likely to be true.

In the second part of the paper we combine the main result of [1] with some additional arguments and prove the following Ramsey type result.

Theorem 1.2 *Let G be a graph on n vertices whose independence number is smaller than $\lfloor \sqrt{n} \rfloor$, that is, any set of $\lfloor \sqrt{n} \rfloor$ vertices of G contains at least one edge. Then there is some subset of $\lfloor \sqrt{n} \rfloor$ vertices of G that contains at least $\Omega(\sqrt{n} \log n)$ edges.*

This is tight and settles a problem of Erdös [4].

The proofs rely heavily on probabilistic arguments. In Section 2 we prove Theorem 1.1 by combining the technique of [8] with an argument in [6]. In Section 3 we prove Theorem 1.2 and discuss the analogous question when we replace the parameter $\lfloor \sqrt{n} \rfloor$ by a general integer m.

2 Locally sparse graphs

Let $\alpha(G)$ denote the independence number of G, that is, the maximum number of vertices in an independent set in G. The main step in the proof of Theorem 1.1 is given below. We make no attempt to optimize the multiplicative constant here and in the rest of the paper.

Proposition 2.1 *Let $G = (V, E)$ be a graph on n vertices with maximum degree $d \geq 1$, in which the neighborhood of any vertex is r-colorable. Then*

$$\alpha(G) \geq \frac{n \log d}{160 d \log(r+1)}.$$

In order to prove the above proposition we need the following simple but somewhat technical fact. A similar estimate is used for a related purpose in [6].

Lemma 2.2 *Let $\mathcal{F}$ be a family of $2^{\epsilon x}$ distinct subsets of an x-element set X. Then the average size of a member of $\mathcal{F}$ is at least*

$$\frac{\epsilon x}{10 \log(1/\epsilon + 1)}.$$

Proof. Clearly we may assume that $\mathcal{F}$ consists of the $2^{\epsilon x}$ subsets of smallest cardinality of X. Since $\frac{6}{7} \cdot \frac{1}{8} > \frac{1}{10}$ it suffices to show that at least $6/7$ of the members of $\mathcal{F}$ are of size at least

$$\frac{\epsilon x}{8 \log(1/\epsilon + 1)}. \tag{2}$$

Since the last quantity is at most $x/8$ (as $\log(1/\epsilon+1) \geq 1$), and since for any $r \leq x/8$, $\binom{x}{r}$ is at least $7\binom{x}{r-1}$, and hence at least $\frac{6}{7}\sum_{i=0}^{r}\binom{x}{i}$, it suffices to show that

$$\sum_{i=0}^{\frac{\epsilon x}{8\log(1/\epsilon+1)}} \binom{x}{i} \leq 2^{\epsilon x},$$

since this would imply that at least $6/7$ of the sets in $\mathcal{F}$ must be of size at least (2). As $\sum_{i=0}^{r}\binom{x}{r} \leq 2^{H(r/x)x}$, where $H(y) = -y\log y - (1-y)\log(1-y)$ is the binary entropy function, it follows that it suffices to check that

$$H\left(\frac{\epsilon}{8\log(1/\epsilon+1)}\right) \leq \epsilon.$$

As for $y < 1/2$, $H(y) < 2y\log(1/y)$ it suffices to check that

$$\frac{2\epsilon}{8\log(1/\epsilon+1)} \log\left(\frac{8\log(1/\epsilon+1)}{\epsilon}\right) \leq \epsilon,$$

or, equivalently, that

$$\log(1/\epsilon) + 3 + \log(\log(1/\epsilon+1)) \leq 4\log(1/\epsilon+1).$$

This, however, is certainly true, since $\log(1/\epsilon) < \log(1/\epsilon+1)$ and since $s = \log(1/\epsilon+1) \geq 1$ and $3 + \log s \leq 3s$ for all real $s \geq 1$. □

Proof of Proposition 2.1. If, say, $d < 16$ the result follows easily from the general bound $\alpha(G) \geq n/(d+1)$ (with room to spare) and hence we may and will assume that $d \geq 16$. Let W be a random independent set of vertices in G, chosen uniformly among all (not necessarily maximal) independent sets in G. For each vertex $v \in V$ define a random variable $X_v = d|\{v\} \cap W| + |N(v) \cap W|$, where $N(v)$ denotes the set of all neighbors of v. Note that $X_v = d$ if $v \in W$ and $X_v = |N(v) \cap W|$ otherwise. We claim that the expectation of X_v satisfies

$$E(X_v) \geq \frac{\log d}{80\log(r+1)}.$$

To prove this claim, let H denote the induced subgraph of G on $V - (N(v) \cup \{v\})$, and fix an independent set S in H. It suffices to show that the conditional expectation

$$E(X_v |\ W \cap V(H) = S) \geq \frac{\log d}{80\log(r+1)} \tag{3}$$

for each possible S. Let X denote the set of all non-neighbors of S in the set $N(v)$, $|X| = x$. Let $2^{\epsilon x}$ denote the precise number of independent subsets in the induced subgraph $G_{|X}$ of G on X. Note that since this induced subgraph is r-colorable, it contains an independent set of size at least x/r and hence $\epsilon \geq 1/r$. Conditioning on the intersection $W \cap V(H) = S$ there are precisely $2^{\epsilon x} + 1$ possibilities for W: one in which $W = S \cup \{v\}$ and $2^{\epsilon x}$ in which $v \notin W$ and W is the union of S with an independent set in $G_{|X}$. It follows that the conditional expectation considered in (3) is precisely $\frac{d}{2^{\epsilon x}+1}$ plus the sum of the sizes of all independent sets in $G_{|X}$ divided by $2^{\epsilon x} + 1$. This last sum is, crucially, at least

$$\frac{\epsilon x 2^{\epsilon x}}{10 \log(1/\epsilon + 1)},$$

by Lemma 2.2. It follows that the conditional expectation (3) is at least

$$\frac{d}{2^{\epsilon x} + 1} + \frac{\epsilon x 2^{\epsilon x}}{10 \log(1/\epsilon + 1)(2^{\epsilon x} + 1)} \geq \frac{d}{2^{\epsilon x+1}} + \frac{\epsilon x}{20 \log(r + 1)}.$$

If $2^{\epsilon x+1} \leq \sqrt{d}$ then the first term in the right hand side of the last inequality is at least $\sqrt{d} > \log d/2$ where here we used the fact that $d \geq 16$. Otherwise, $\epsilon x \geq \frac{1}{2} \log d - 1 \geq \frac{1}{4} \log d$. Therefore, in any case the sum of the two terms is at least as large as their maximum and hence

$$E(X_v | \, W \cap V(H) = S) \geq \frac{\log d}{80 \log(r + 1)},$$

establishing the claim.

By linearity of expectation we conclude that the expected value of the sum $\sum_{v \in V} X_v$ is at least $\frac{n \log d}{80 \log(r+1)}$. On the other hand, this sum is clearly at most $2d|W|$, since each vertex $u \in W$ contributes d to the term X_u in this sum, and $deg_G(u) \leq d$ to the sum of all other terms X_v. It follows that the expected size of W is at least $\frac{n \log d}{160 d \log(r+1)}$, and hence there is an independent set of size at least this expectation, completing the proof. □

Proof of Theorem 1.1. Any graph on n vertices with average degree t has at least $n/2$ vertices of degree at most $2t$. Hence it contains an induced subgraph on $n/2$ vertices with maximum degree at most $2t$, and the result follows by applying the last proposition to this induced subgraph. □

Remark. A close look at the proof of Theorem 1.1 shows that its assertion is valid even if we replace the assumption that every neighborhood is r-colorable by the weaker assumption that for any x, any set of x vertices with a common neighbor contains an independent set of size at least x/r. In particular it suffices to assume that the chromatic number (or even the fractional chromatic number) of the induced subgraph of G on the neighborhood of any vertex is at most r. Moreover, it suffices, in fact, to assume that any set of $x \leq r \log^2 t$ vertices with a common neighbor contains an independent subset of size at least x/r. This is because in the proof of Proposition 2.1, if $x \geq r \log^2 d$, then there is an independent set of size at least $\log^2 d$ in X, and the number of independent subsets in $G_{|X}$ is at least $d^{\log d}$. The average size of a set in a collection of at least $d^{\log d}$ subsets of a set of size at most d is at least $\Omega(\log d)$ and hence in this case the contribution of the term $|N(v) \cap W|$ in the definition of X_v to the conditional expectation is at least $\Omega(\log d)$, as needed. If, on the other hand, $x \le r \log^2 d$, then the assumption, for this case, is identical to the one in the proposition and the proof can proceed as in its proof.

Remark. It is worth noting that unlike the proofs in [1] or [7] the above proof is non-constructive in the sense that it supplies no efficient algorithm that finds in any graph satisfying the assumptions of Theorem 1.1 an independent set of the size mentioned in the conclusion of the theorem.

3 A Ramsey type problem

Proof of Theorem 1.2. Let G be a graph on n vertices with independence number $\alpha(G) < \lfloor \sqrt{n} \rfloor$. We must show that there is a set of $m = \lfloor \sqrt{n} \rfloor$ vertices of G containing $\Omega(m \log n)$ edges. Throughout the proof here and in the rest of the section we assume, whenever this is needed, that n is sufficiently large. We also omit all floor and ceiling signs whenever these are not crucial.

If the average degree in G is at least $\sqrt{n} \log n$, then G has at least $\frac{1}{2} n^{3/2} \log n$ edges, and the expected number of edges in a random subset of m vertices of G is at least

$$\frac{\binom{m}{2}}{\binom{n}{2}} \frac{n^{3/2} \log n}{2} > \frac{m \log n}{4},$$

where in the last inequality we used the fact that n is large. Thus, in this case, there is an m-set containing sufficiently many edges, as needed. Hence we may and will assume that the average degree in G is at most $\sqrt{n} \log n$. Therefore, at least $n/2$ vertices have degrees that do not exceed $2\sqrt{n} \log n$. Let G' be the induced subgraph of G on some $n/2$ vertices whose degrees (in G) do not exceed $2\sqrt{n} \log n$.

If in G' there is a vertex v with at least $\sqrt{n} \log^3 n$ edges in its neighborhood $N(v)$, we argue as follows. If $|N(v)| \le m$ then any set of m vertices containing $N(v)$ has sufficiently many edges, as needed. Otherwise, take a random subset of m members of $N(v)$. Since $|N(v)| \le 2\sqrt{n} \log n$ the expected number of edges in such a random subset exceeds $\frac{1}{8} m \log n$, implying the desired result in this case too. Thus we may assume there is no vertex v as above, and hence the total number of triangles in G' is at most

$$\frac{1}{3}(n/2)\sqrt{n} \log^3 n < n^{3/2} \log^3 n.$$

Let U be a random subset of vertices of G' obtained by picking each vertex of G' to be a member of U, randomly and independently, with probability $p = n^{-0.4}$. Let $X = X(U)$ be the random variable whose value is the cardinality of U minus the number of triangles contained in the induced subgraph of G' on U. The expected value of X is at least

$$\frac{n}{2} p - (n^{3/2} \log^3 n) p^3 \ge \frac{n^{0.6}}{2} - \frac{n^{3/2} \log^3 n}{n^{1.2}} \ge \frac{n^{0.6}}{4}.$$

In particular, there is a choice of U for which the value of $X(U)$ is at least this quantity, and by throwing a vertex from each triangle in the induced subgraph $G'_{|U}$, (and by omitting some additional vertices, if needed) we obtain a triangle-free induced subgraph G_2 of G' (and hence of G) on a set U' of $\frac{n^{0.6}}{4}$ vertices. Let t denote the average degree in G_2. By the main result of [1] mentioned in the introduction and extended in Theorem 1.1 above, G_2 (and hence G) contains an independent set of size at least

$$c\frac{n^{0.6}}{4t}\log t,$$

where $c > 0$ is an absolute constant. This independent set must be of size smaller than $m \leq \sqrt{n}$, implying that $t \geq c' n^{0.1} \log n$ for some absolute positive c'. It follows that the induced subgraph of G on U' has $\Omega(n^{0.7} \log n)$ edges, and hence the expected number of edges in a random subset of $m = \lfloor \sqrt{n} \rfloor$ vertices in it is $\Omega(m \log n)$, completing the proof. □

Next, we prove the following simple result, which shows that the assertion of Theorem 1.2 is tight (in a more general setting).

Proposition 3.1 *There exists an absolute constant c so that for any $1 < m \leq n$ there exists a graph on n vertices whose independence number is smaller than m, in which every set of m vertices contains at most $cm \ln(en/m)$ edges.*

Proof. Without trying to optimize our estimates we prove the proposition with $c = 6$, say. If $m < 12\ln(en/m)$, even a complete graph on m vertices contains less than $6m\ln(en/m)$ edges, and hence we can simply define G to be a complete graph. Otherwise, let $G = (V, E)$ be the standard random graph on n labeled vertices obtained by picking each pair of vertices as an edge, randomly and independently, with probability $p = \frac{4}{m-1}\ln(en/m)$ (< 1). The expected number of independent sets of size m in G is

$$\binom{n}{m}(1-p)^{\binom{m}{2}} < (en/m)^m e^{-pm(m-1)/2} = (\frac{en}{m}e^{-2ln(en/m)})^m = (en/m)^{-m} < 1/2.$$

Let M be a fixed set of m vertices of G. By the standard estimates for Binomial distributions (see, e.g., Theorem A.12 in Appendix A of [3]), the probability that the value of a Binomial random variable with parameters N and p exceeds βpN for $\beta > 1$ is at most $(e^{\beta-1}\beta^{-\beta})^{pN}$. Substituting $N = \binom{m}{2}$, $p = \frac{4}{m-1}\ln(en/m)$, and $\beta = 3$ we conclude that the probability that the induced subgraph of G on M has more than $6m\ln(en/m)$ edges is at most

$$(\frac{e^2}{3^3})^{2m\ln(en/m)} < (\frac{en}{m})^{-2m}.$$

Therefore, the expected number of sets M on which G has more than $6m\ln(en/m)$ edges is at most

$$\binom{n}{m}(\frac{en}{m})^{-2m} < (en/m)^{-m} < 1/2.$$

It follows that the expected number of independent sets of size m plus the expected number of sets of size m containing more than $6m\ln(en/m)$ edges is smaller than 1. Thus, there is a graph containing no sets of these two types, completing the proof. □

By the known bounds for the usual Ramsey numbers (see., e.g., [5]), if a graph on n vertices contains no independent set of size $\log n$ then it contains a complete graph on $\Omega(\log n)$ vertices. This implies the following simple statement.

Proposition 3.2 *For $m \le \log n$, in any graph on n vertices in which every set of m vertices contains an edge there is some set of m vertices that contains $\Omega(m^2)$ edges.*

For each $1 < m \le n$, let $f(m,n)$ denote the largest integer f so that in any graph G on n vertices satisfying $\alpha(G) < m$ there is some set of m vertices containing at least $f(m,n)$ edges. By the above proposition $f(m,n) = \Theta(m^2)$ for $1 < m \le \log n$, and by Theorem 1.2 and Proposition 3.1, $f(m,n) = \Theta(n \log(en/m))$ ($= \Theta(m \log n)$) for $m = \lfloor \sqrt{n} \rfloor$. It is easy to see that $f(m,n) = \Theta(n-m)$ for $m \ge n/2$, and it is not too difficult to extend the proof of Theorem 1.2 and show that $f(m,n) = \Theta(m \log(em/n))$ for all $n^{\epsilon} \le m \le n/2$. P. Valtr [9] proved the following result that generalizes Theorem 1.2 and determines, together with Proposition 3.1, the asymptotic behaviour of $f(m,n)$ in all other cases.

Theorem 3.3 (Valtr) *For every m and n satisfying $\log n \le m \le n/2$,*

$$f(m,n) = \Omega(m \log(n/m)) \ (\ = \Omega(m \log(en/m)) \ ).$$

Therefore, $f(m,n) = \Theta(n-m)$ for $n/2 \le m \le n$, $f(m,n) = \Theta(m \log(en/m))$ for $\log n \le m \le n/2$ and $f(m,n) = \Theta(m^2)$ for $1 \le m \le \log n$. The precise determination of $f(m,n)$ for every m and n seems extremely difficult, as it requires, in particular, to determine precisely the diagonal Ramsey numbers $R(m,m)$, since, clearly, $f(m,n) = \binom{m}{2}$ iff $R(m,m) \le n$.

Acknowledgment Part of this work was done during a DIMANET workshop on some trends in Discrete Mathematics in Matrahaza, Hungary, organized by Ervin Györi and Vera T. Sós in October, 1995. I would like to thank the organizers and the participants of the workshop, including, in particular, P. Cameron, P. Erdös, D. Gunderson, M. Krivelevich and P. Valtr, for helpful discussions and comments.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, *A note on Ramsey numbers*, J. Combinatorial Theory, Ser. A 29 (1980), 354-360.

[2] M. Ajtai, P. Erdös, J. Komlós and E. Szemerédi, *On Turan's theorem for sparse graphs*, Combinatorica 1 (1981), 313-317.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons Inc., New York, 1992.

[4] P. Erdös, *Some old and new problems in various branches of Combinatorics,* in: Proc. 10^{th} Southeastern Conf. on Combinatorics, Graph Theory and Computing, Boca Raton 1979 Vol I, Congressus Numerantium 23 (1979), 19-37.

[5] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory* (Second Edition), Wiley, New York, 1990.

[6] A. Nilli, *The average size of an independent set in graphs with a given chromatic number*, J. Combinatorial Theory Ser. B 45 (1988), 112-114.

[7] J. B. Shearer, *A note on the independence number of triangle-free graphs*, Discrete Math. 46 (1983), 83-87.

[8] J. B. Shearer, *On the independence number of sparse graphs*, Random Structures and Algorithms 7 (1995), 269-271.

[9] P. Valtr, in preparation.