On the density of subgraphs in a graph with bounded independence number*

Pavel Valtr

Department of Applied Mathematics, Charles University
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
e-mail: `valtr@kam.mff.cuni.cz`

DIMACS Center, Rutgers University
P.O.Box 1179, Piscataway, NJ 08855, U.S.A.
e-mail: `valtr@dimacs.rutgers.edu`

Abstract

Let $\sigma(n,m,k)$ be the largest number $\sigma \in [0,1]$ such that any graph on n vertices with independence number at most m has a subgraph on k vertices with at least $\sigma \cdot \binom{k}{2}$ edges. Up to a constant multiplicative factor, we determine $\sigma(n,m,k)$ for all n,m,k. For $\log n \leq m = k \leq n$, our result gives $\sigma(n,m,m) = \Theta(\frac{\log(n/m)}{m})$, which was conjectured by Alon.

1 Introduction

All logarithms in this paper are assumed to be in base e. Let $\alpha(G)$ denote the independence number of a graph G. Erdős [Er79] conjectured that every graph G on n vertices with $\alpha(G) < \lfloor\sqrt{n}\rfloor$ contains a subgraph on $\lfloor\sqrt{n}\rfloor$ vertices with at least $c \cdot \sqrt{n}\log n$ edges, where $c > 0$ is a constant independent of n. This conjecture was proved by Alon [Al96] who further conjectured [GS95] that, for $\log n \leq m \leq n$, every graph G on n vertices with $\alpha(G) < m$ contains a subgraph on

*Research was supported by Czech Republic Grant GAČR 0194/1996 and by Charles University grant No. 194/1996.

m vertices with at least $c \cdot m \log(n/m)$ edges. Alon [Al96] proved his conjecture for $\log n \le m \le C \log n$ and for $n^{\varepsilon} \le m \le n$, where $C, \varepsilon > 0$ are any two positive constants. Alon [Al96] also proved that if his conjecture is true, then the bound $c \cdot m \log(n/m)$ in it is optimal up to a constant factor. In this paper, we prove Alon's conjecture by showing the following more general results.

Theorem 1 *Let $2 \le k \le n$ and $2 \le m \le n/2$. Let $\sigma(n,m,k)$ be the largest number $\sigma \in [0,1]$ such that any graph on n vertices with $\alpha(G) \le m$ has a subgraph on k vertices with at least $\sigma \cdot \binom{k}{2}$ edges. Then*

$$\sigma(n,m,k) = \begin{cases} \Theta\left(\min\left\{\frac{\log(en/k)}{m}, 1\right\}\right), & \text{for } k \ge m, \\ \Theta\left(\min\left\{\frac{\log(en/k)}{k\log(em/k)}, 1\right\}\right), & \text{for } m \ge k. \end{cases}$$

($\sigma(n,m,k) = \Theta(f(n,m,k))$ means that there are two positive constants $c_1, c_2 > 0$ such that $c_1 f(n,m,k) \le \sigma(n,m,k) \le c_2 f(n,m,k)$ for all feasible n, m, k.)

Observation 2 *Let $2 \le k \le n$ and $n/2 \le m \le n$. Let $\sigma(n,m,k)$ be defined as in Theorem 1. Then*

$$\sigma(n,m,k) = \frac{\min\{n-m, \lfloor k/2 \rfloor\}}{\binom{k}{2}}.$$

Theorem 1 and Observation 2 quantitatively express the fact that any graph with no dense subgraphs has a large independence number. Alon's conjecture indeed follows from them, since for $m = k$ we get $\sigma(n,m,m) = \Theta\left(\min\{\frac{\log(en/m)}{m}, 1\}\right) = \Theta\left(\frac{\log(n/m)}{m}\right)$ (if $\log n \le m \le n/2$) or $\sigma(n,m,m) = \frac{\min\{n-m, \lfloor m/2 \rfloor\}}{\binom{m}{2}} = \Theta\left(\frac{n-m}{m^2}\right) = \Theta\left(\frac{\log(n/m)}{m}\right)$ (if $n/2 \le m \le n$).

Both the lower and the upper bounds on $\sigma(n,m,k)$ are proved differently for $m \le k$ than for $k \le m$. In both cases ($m \le k$ and $k \le m$), the lower bound is proved by "algorithmic" methods and the upper bound by probabilistic methods (using Alon's proof in the special case $m = k$).

and the lower bound in Theorem 1 follows from Observation 3.

Acknowledgments. I would like to thank Noga Alon for interesting discussions and for helpful comments on the draft of this paper. I also thank the referee for helpful comments and suggestions.

References

[Al96] N. Alon, Independence numbers of locally sparse graphs and a Ramsey type problem, Random Structures and Algorithms 9 (1996), 271–278.

[Bo85] B. Bollobás, Random graphs, Academic Press, London 1985.

[Er79] P. Erdős, 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.

[GS95] E. Győri and V.T. Sós (eds.), Problem collection of the DIMANET Mátraháza Workshop, Mátraháza, October 22–28, 1995.

[Ne95] J. Nešetřil, Ramsey theory, in: R.L. Graham, M. Grötschel, and L. Lovász (eds.), Handbook of combinatorics, Vol. II, Elsevier, New York 1995, pp. 1331–1403.

[Rö86] V. Rödl, Upper bounds on Ramsey numbers $R(k,l)$, unpublished manuscript (1986).

[Th87] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs, in: C. Whitehead (ed.), Survey in combinatorics, Cambridge Univesity Press, Cambridge 1987, 173-196.

[Th88] A. Thomason, An upper bound for some Ramsey numbers, J. Graph Theory 12 (1988), 509–517.

[Tu54] P. Turán, On the theory of graphs, Colloquium Math. 3 (1954), 19–30.

Our result is closely related to Ramsey theory. Recall that the classical (off-diagonal) Ramsey number $r(m,k)$ is defined as the smallest number r such that, for any 2-coloring of edges of K_r by red and blue, there is a subgraph (isomorphic to) K_m all of whose edges are red or a subgraph K_k all of whose edges are blue. One can ask how the number $r(m,k)$ changes if in the latter case one requires only some fraction of edges in K_k to be blue. More precisely, let $n_\sigma(m,k)$ be the smallest number n such that, for any 2-coloring of edges of K_n by red and blue, there is a subgraph K_m all of whose edges are red or a subgraph K_k with at least $\sigma\binom{k}{2}$ blue edges. For $\sigma \in [0,1]$, Theorem 1 and Observation 2 show that

$$n_\sigma(m,k) \approx \begin{cases} k \cdot 2^{c\cdot\sigma m}, & \text{for } k \geq m, \\ k \cdot 2^{c\cdot\sigma k \log(em/k)}, & \text{for } m \geq k. \end{cases}$$

We plan to study this question in a more general setting in a forthcoming paper.

Our proofs give relatively good constants of proportionality in Theorem 1. However, for the sake of simplicity, we do not try to optimize them.

Theorem 1 and Observation 2 might be also stated as results about densities of graphs. The *density* of a graph $G = (V,E)$ is the number $\frac{|E|}{\binom{|V|}{2}}$, i.e. the number of edges in G divided by the number of edges in the complete graph on the same vertex set. The graph density satisfies the following useful observation.

Observation 3 *Let G be a graph on m vertices of density σ. Then, for any $2 \leq k \leq m$, there is an (induced) subgraph of G on k vertices of density at least σ.*

Proof. Standard averaging argument. □

Theorem 1 is proved in Sections 2 (the upper bound) and 3 (the lower bound). Observation 2 has a simple proof:

Proof of Observation 2. Let $m \geq n/2$. The upper bound

$$\sigma(n,m,k) \leq \frac{\min\{n-m, \lfloor k/2 \rfloor\}}{\binom{k}{2}}$$

can be obtained from a matching of size $n-m$ on n vertices. To see the lower bound

$$\sigma(n,m,k) \geq \frac{\min\{n-m, \lfloor k/2 \rfloor\}}{\binom{k}{2}},$$

observe that any graph G on n vertices with $\alpha(G) \leq m$ has at least $n-m$ edges. Any $\min\{n-m, \lfloor k/2 \rfloor\}$ of them have at most k vertices altogether. The lower bound follows. □

2 The upper bound

2.1 The case $m \geq k$

The random graph $G_{n,p}$ is defined as a random graph on n labeled vertices obtained by picking each pair of vertices as an edge, randomly and independently, with probability p. Set

$$p = \frac{4}{m-1}\log(en/m).$$

As shown in [Al96, Prop. 3.1] by a "brute-force" argument, if $p \leq 1$ then the random graph $G_{n,p}$ satisfies $\alpha(G_{n,p}) < m$ with probability bigger than $\frac{1}{2}$. Consequently, the following lemma is the core of the proof of the upper bound in Theorem 1 for $m \geq k$:

Lemma 4 *For any $n \geq m \geq k \geq 2$ such that $p = \frac{4}{m-1}\log(en/m)$ is not bigger than 1, the random graph $G_{n,p}$ has a subgraph on k vertices of density at least $8e^2\frac{\log(en/k)}{k\log(em/k)}$ with probability smaller than $\frac{1}{2}$.*

Lemma 4 gives the upper bound in Theorem 1 for $m \geq k$ and $p = \frac{4}{m-1}\log(en/m) \leq 1$ (indeed, in this case Lemma 4 and Alon's result mentioned above it show that the graph $G_{n,p}$ demonstrates the bound with positive probability). If $m \geq k$ and $p > 1$, then $m < 4\log(en/m) + 1 \leq 5\log(en/k)$, and consequently $\frac{\log(en/k)}{k\log(em/k)} \geq \frac{\log(en/k)}{m} > \frac{1}{5}$. Thus, for $m \geq k$ and $p > 1$ Theorem 1 asserts that $\sigma(n,m,k) = \Theta(1)$, and the upper bound in it is trivial in this case. It remains to prove Lemma 4.

the set of its neighbors in G_{i-1}. Certainly, the vertices $v_1,\ldots,v_{m+1}$ are independent. To prove the correctness of the argument, it suffices to show that $|V_i| > 0$ for $i = 1,\ldots,m$.

Let $i \in \{1,\ldots,m\}$. If $|V_{i-1}| \geq k$, then the density of G_{i-1} is smaller than σ, and therefore the degree of v_i in G_{i-1} is smaller than $\sigma(|V_{i-1}|-1)$. From this we get

$$|V_i| > |V_{i-1}| - 1 - \sigma(|V_{i-1}|-1) > (1-2\sigma)|V_{i-1}|,$$

provided $|V_{i-1}| \geq k$ (since $\sigma|V_{i-1}| \geq 1$ follows from $|V_{i-1}| \geq k$, $\sigma = \min\{\frac{\log(en/k)}{4m}, \frac{1}{4}\}$, $k \geq 4$, and from our supposition $k \leq \frac{n}{e^3}$). Further, we get

$$|V_i| \quad > \quad (1-2\sigma)^i|V_0|,$$

provided $|V_{i-1}| \geq k$. Since the function $f(t) = 1-2t-e^{-3t}$ is concave and its values in $t=0$ and $t=\frac{1}{4}$ are non-negative, we have $f(\sigma) \geq 0$, thus

$$1-2\sigma \geq e^{-3\sigma}.$$

Thus,

$$\begin{aligned} |V_i| &> (1-2\sigma)^i|V_0| \\ &\geq (1-2\sigma)^m|V_0| \\ &\geq (e^{-3\sigma})^m n \\ &\geq e^{-\frac{3\log(en/k)}{4}} n^{\frac{3}{4}} n^{\frac{1}{4}} \\ &\geq \left(\frac{en}{k}\right)^{-\frac{3}{4}} n^{\frac{3}{4}} (e^3k)^{\frac{1}{4}} \\ &= k, \end{aligned}$$

provided $|V_{i-1}| \geq k$. From this and from $|V_0| = n \geq k$, we get by induction on i that

$$|V_1| \geq |V_2| \geq \cdots \geq |V_m| \geq k > 0,$$

which completes the argument for $k \leq \frac{n}{e^3}$.

If $k > \frac{n}{e^3}$ (and $m \leq \frac{n}{2}$) then, by Turán's theorem [Tu54], any graph G on n vertices with $\alpha(G) \leq m$ has density at least

$$\frac{\frac{n\left(\frac{n}{m}-1\right)}{2}}{\binom{n}{2}} > \frac{\frac{n\frac{n}{2m}}{2}}{\frac{n^2}{2}} = \frac{1}{2m} > \frac{\log(en/k)}{8m},$$

Case 3: $k < \frac{1}{2\sigma_0}$**.** Any graph on n vertices with $\alpha(G) \leq m$ has at least $n - m \geq m \geq k$ edges (we have $m \leq \frac{n}{2}$), and thus any subgraph on k vertices with $\lfloor k/2 \rfloor$ edges has density at least

$$\frac{\lfloor \frac{k}{2} \rfloor}{\frac{k(k-1)}{2}} > \frac{\frac{k}{4}}{\frac{k^2}{2}} = \frac{1}{2k} > \sigma_0.$$

Thus,

$$\boxed{\sigma(n,m,k) > \sigma_0 \quad \text{in Case 3.}}$$

We conclude that

$$\boxed{\sigma(n,m,k) \geq \min\left\{\sigma_0, \tfrac{1}{2}\right\} \quad \text{in each case.}}$$

Thus,

$$\sigma(n,m,k) \geq \min\left\{\sigma_0, \frac{1}{2}\right\} \geq \frac{1}{32}\min\left\{\frac{\log(en/k)}{k\log(em/k)}, 1\right\},$$

which is a desired lower bound in Theorem 1 for $m \geq k$.

3.2 The case $k \geq m$

Let $k \geq m$. We first suppose that $k \leq \frac{n}{e^3}$. We may assume that $k \geq 4$ (the case $k \leq 3$ is trivial). For a contrary, suppose that G is a graph on n vertices with $\alpha(G) \leq m$ such that the density of any subgraph on k vertices is smaller than $\sigma = \frac{1}{4}\min\{\frac{\log(en/k)}{m}, 1\} = \min\{\frac{\log(en/k)}{4m}, \frac{1}{4}\}$. By Observation 3, every subgraph of G on at least k vertices has density smaller than σ.

We now find $m+1$ independent vertices $v_1, \ldots, v_{m+1}$, which gives a contradiction to $\alpha(G) \leq m$. We start with the graph G, and proceed inductively in $m+1$ steps $S_1, \ldots, S_{m+1}$ so that in step S_i we choose a vertex v_i of minimum degree in the current graph and remove it and all its neighbors from the graph. Let V_i be the set of vertices which remain after step S_i. Thus, if V_0 denotes the vertex set of G, then $V_i = V_{i-1} \setminus (\{v_i\} \cup N_{i-1}(v_i))$, where v_i is a vertex of minimum degree in the subgraph G_{i-1} of G induced by V_{i-1} and $N_{i-1}(v_i)$ is

In the proof of Lemma 4, we use the following estimate on binomial coefficients.

Lemma 5 *If $n \geq k \geq 1$, then*

$$\binom{n}{k} < \left(\frac{en}{k}\right)^k.$$

Lemma 5 appears relatively often in the literature. We include its short proof for the sake of completeness.

Proof of Lemma 5. We fix n and proceed by induction on k. For $k = 1$ the lemma trivially holds. Suppose now that $k > 1$ and that $\binom{n}{k-1} < \left(\frac{en}{k-1}\right)^{k-1}$. Then

$$\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1} < \frac{n}{k}\left(\frac{en}{k-1}\right)^{k-1} =$$

$$= \frac{1}{e}\left(\frac{k}{k-1}\right)^{k-1}\left(\frac{en}{k}\right)^k < \left(\frac{en}{k}\right)^k.$$

□

Proof of Lemma 4. Set $z = \left\lceil \frac{8e^2 \cdot \log(en/k)}{k\log(em/k)}\binom{k}{2} \right\rceil$. We will show (also by a "brute-force" argument) that the graph $G = G_{n,p}$ has a subgraph on k vertices with at least z edges with probability smaller than $\frac{1}{2}$.

There are $\binom{n}{k}$ sets of k vertices of G. For any set V_0 of k vertices of G, there are $\binom{\binom{k}{2}}{z}$ collections (sets) of z pairs of vertices of V_0. Thus, there are at most $\binom{n}{k}\binom{\binom{k}{2}}{z}$ collections of z pairs of vertices chosen from among at most k distinct vertices in G. Denote the set of these collections by $\mathcal{C}$. G has a subgraph on k vertices with at least z edges if and only if $\mathcal{C}$ contains a subset of $E(G)$. The lemma holds if this happens with probability smaller than $\frac{1}{2}$. Since each member of $\mathcal{C}$ is a subset of $E(G)$ with probability p^z, the lemma follows from the inequality $|\mathcal{C}|p^z < \frac{1}{2}$,which can be shown from Lemma 5 as follows:

$$\begin{aligned}
|\mathcal{C}|p^z &\le \binom{n}{k}\binom{\binom{k}{2}}{z}p^z \\
&< \left(\frac{en}{k}\right)^k \left(\frac{e\binom{k}{2}}{z}\right)^z p^z \\
&\le \left(\frac{en}{k}\right)^k \left(\frac{e\binom{k}{2}}{\frac{8e^2\cdot\log(en/k)}{k\log(em/k)}\binom{k}{2}} \cdot \frac{4}{m-1}\log(en/m)\right)^z \\
&= \left(\frac{en}{k}\right)^k \left(\frac{1}{2e}\cdot\frac{k}{m-1}\cdot\frac{\log(en/m)}{\log(en/k)}\cdot\log(em/k)\right)^z \\
&< \left(\frac{en}{k}\right)^k \left(\frac{1}{2e}\cdot\frac{k}{m/2}\cdot 1\cdot\sqrt{em/k}\right)^z \\
&= \left(\frac{en}{k}\right)^k \left(\frac{k}{em}\right)^{\frac{z}{2}} \\
&= e^{k\log(en/k)-\frac{z}{2}\log(em/k)} \\
&< \frac{1}{2}.
\end{aligned}$$

□

2.2 The case $k \ge m$

It follows from the special case $m = k$ in Section 2.1 (and was also shown in [Al96]) that for any $s \ge m$ there is a graph $G(s,m)$ on s vertices with $\alpha(G(s,m)) \le m$ having no subgraph on m vertices of density at least $8e^2\frac{\log(es/m)}{m}$. Let $n \ge k \ge m \ge 2$. Put $s = \frac{nm}{k}$ and $t = \frac{k}{m}$. For simplicity, assume that s and t are integers (using roundings one can get the proof in a similar way also in other cases). Set $G_0 = G(s,m)$. Replace each vertex in G_0 by K_t, and denote the obtained graph by G, i.e., $G = (V,E)$, where

$$\begin{aligned}
V &= V(G_0)\times\{1,2,\ldots,t\}, \\
E &= \left\{\{(v_1,i_1),(v_2,i_2)\}\in\binom{V}{2} : v_1 = v_2 \text{ or } (v_1,v_2)\in E(G_0)\right\}.
\end{aligned}$$

$$\begin{aligned}
&\le 2k\left(\frac{2em}{k}\right)^{\frac{\log(en/k)}{4\log(em/k)}} \\
&= 2k\left(\frac{en}{k}\right)^{\frac{\log(2em/k)}{4\log(em/k)}} \\
&< 2k\left(\frac{en}{k}\right)^{\frac{1}{2}} \\
&= (4ekn)^{\frac{1}{2}} \\
&< n
\end{aligned}$$

(the last inequality follows from $4ekn < \frac{kn}{e}\cdot e^{16} \le \frac{kn}{e}\cdot e^{32k\sigma_0} = \frac{kn}{e}\cdot e^{\frac{\log(en/k)}{\log(em/k)}} \le \frac{kn}{e}\cdot\frac{en}{k} = n^2$). From Observation 6 we now get

$$\boxed{\sigma(n,m,k) \ge \sigma_0 \quad \text{in Case 1.}}$$

Case 2: $\sigma_0 > \frac{1}{2}$**.** Consequence 8 and Lemma 5 give

$$\begin{aligned}
n_{\frac{1}{2}}(m,k) &< 2\binom{m+k}{k} - 1 \\
&< 2\left(\frac{e(m+k)}{k}\right)^k - 1 \\
&< 2\left(\frac{2em}{k}\right)^{2\sigma_0\cdot k} - 1 \\
&= 2\left(\frac{2em}{k}\right)^{\frac{\log(en/k)}{16\log(em/k)}} - 1 \\
&= 2\left(\frac{en}{k}\right)^{\frac{\log(2em/k)}{16\log(em/k)}} - 1 \\
&< 2\left(\frac{en}{k}\right)^{\frac{1}{8}} - 1 \\
&< n.
\end{aligned}$$

Observation 6 now gives

$$\boxed{\sigma(n,m,k) \ge \tfrac{1}{2} \quad \text{in Case 2.}}$$

$$
\begin{aligned}
&< \frac{1}{\sigma}\binom{2\sigma(m-\frac{1}{2\sigma}+k)}{2\sigma k} - \frac{1}{2\sigma} \\
&\quad + \frac{1}{\sigma}\binom{2\sigma(m+k-\frac{1}{2\sigma})}{2\sigma(k-\frac{1}{2\sigma})} - \frac{1}{2\sigma} + \frac{1}{2\sigma} \\
&= \frac{1}{\sigma}\left(\binom{2\sigma(m+k)-1}{2\sigma k} + \binom{2\sigma(m+k)-1}{2\sigma k-1}\right) - \frac{1}{2\sigma} \\
&= \frac{1}{\sigma}\binom{2\sigma(m+k)}{2\sigma k} - \frac{1}{2\sigma}.
\end{aligned}
$$

□

We can now prove the lower bound in Theorem 1 for $m \geq k$. We assume $m \geq k$ and set

$$\sigma_0 = \frac{\log(en/k)}{32k\log(em/k)}.$$

We distinguish three cases such that at least one of them must occur.

Case 1: $k \geq \frac{1}{2\sigma_0}$, $\sigma_0 \leq \frac{1}{2}$. This is the basic and most difficult case. Set

$$\sigma_1 = \frac{1}{2\left\lfloor \frac{1}{2\sigma_0} \right\rfloor}, \quad m_1 = \frac{\lceil 2\sigma_1 m \rceil}{2\sigma_1}, \quad k_1 = \frac{\lceil 2\sigma_1 k \rceil}{2\sigma_1}.$$

Obviously,

$$0 < \sigma_1 \leq \frac{1}{2}, \quad \sigma_0 \leq \sigma_1 < 2\sigma_0, \quad m \leq m_1 < 2m, \quad k \leq k_1 < 2k.$$

From Observation 3, Consequence 8, Lemma 5, and from $m \geq k \geq \frac{1}{2\sigma_0}$, we get

$$
\begin{aligned}
n_{\sigma_0}(m,k) &\leq n_{\sigma_1}(m_1,k_1) \\
&< \frac{1}{\sigma_1}\binom{2\sigma_1(m_1+k_1)}{2\sigma_1 k_1} \\
&< \frac{1}{\sigma_0}\binom{8\sigma_0(m+k)}{8\sigma_0 k} \\
&< \frac{1}{\sigma_0}\left(\frac{e(m+k)}{k}\right)^{8\sigma_0 k}
\end{aligned}
$$

We now prove that G gives the upper bound in Theorem 1 in case $k \geq m$. Certainly, $|V| = n$ and $\alpha(G) = \alpha(G_0) \leq m$. It remains to show that G has no subgraph on k vertices of density at least $c \cdot \frac{\log(en/k)}{m}$, which we now show with $c = 8e^2 + 1$.

We say that a subgraph $H \subseteq G$ *splits* a vertex $v \in V(G_0)$ if there are two indices $i, j \in \{1, 2, \ldots, t\}$ with $(v,i) \in V(H), (v,j) \notin V(H)$. Let H be a subgraph of G on k vertices with the maximum density (i.e., with the maximum number of edges). If H splits two vertices $v_1, v_2 \in V(G_0)$, then we may find another (induced) subgraph $H' \subseteq G$ on k vertices which splits fewer vertices of G_0 than H and has density at least as big as H (H' can be obtained from H by replacing either some vertices (v_1, i) by some vertices (v_2, j) or some vertices (v_2, i) by some vertices (v_1, j) in H). Repeating this procedure, we get a subgraph of G on k vertices which splits at most one vertex of G_0 and has density at least as big as H. Therefore, we may assume that H itself splits at most one vertex of G_0. H cannot split exactly one vertex of G_0, since the number of vertices in H is divisible by t. Thus, we may assume that H splits no vertex of G_0. By the construction of G, the density of H is then smaller than

$$8e^2\frac{\log(es/m)}{m} + \frac{m\binom{t}{2}}{\binom{mt}{2}} < 8e^2\frac{\log(en/k)}{m} + \frac{1}{m} \leq (8e^2+1)\frac{\log(en/k)}{m}.$$

3 The lower bound

3.1 The case $m \geq k$

We now settle the most complicated case in the paper. The main idea (contained in Lemma 7(a) below) has some similarities with the ideas of Rödl [Rö86] and Thomason [Th87, Th88] leading to the currently asymptotically best upper bounds on the Ramsey off-diagonal numbers $r(k,m)$. These ideas are described in a simple way in [Ne95, pp. 1348-1349].

For $\sigma \in [0,1]$, $m \geq 1$, $k \geq 1$, we define a Ramsey-type number $n_\sigma(m,k)$ as the minimum number n such that any graph on n vertices with independence number at most m has a subgraph on k vertices of density at least σ. (The density of a graph with one vertex is assumed to be 1.)

Observation 6 *If $n_\sigma(m,k) \le n$, then $\sigma(n,m,k) \ge \sigma$.*

Hence, an upper bound on the function n_σ gives a lower bound on the function σ. Here is the key lemma:

Lemma 7 *Let $\sigma \in [0, \frac{1}{2}]$ and let $\frac{1}{2\sigma}$ be an integer. Then*

(a) $n_\sigma(m,k) < n_\sigma(m - \frac{1}{2\sigma}, k) + n_\sigma(m, k - \frac{1}{2\sigma}) + \frac{1}{2\sigma}$, for $m \ge \frac{1}{\sigma}$ and $k \ge \frac{3}{4\sigma}$,

(b) $n_\sigma(\frac{1}{2\sigma}, k) = k$, for $k \ge \frac{1}{\sigma}$,

(c) $n_\sigma(m, \frac{1}{2\sigma}) \le m + \frac{1}{8\sigma}$, for $m \ge \frac{1}{2\sigma}$.

Proof. **(a)** Suppose for a contrary that G is a graph on $n_\sigma(m - \frac{1}{2\sigma}, k) + n_\sigma(m, k - \frac{1}{2\sigma}) + \frac{1}{2\sigma} - 1$ vertices with $\alpha(G) \le m$ containing no subgraph on k vertices of density at least σ. Since G has more than $n_\sigma(m - \frac{1}{2\sigma}, k)$ vertices, it has $m - \frac{1}{2\sigma} \ge \frac{1}{2\sigma}$ independent vertices. Fix any set I of $\frac{1}{2\sigma}$ independent vertices in G. Partition the set $V(G) \setminus I$ into two subsets I_0 and I_1 containing the vertices of $V(G) \setminus I$ adjacent to no vertex in I and to at least one vertex in I, respectively. By the pigeonhole principle, $|I_0| \ge n_\sigma(m - \frac{1}{2\sigma}, k)$ or $|I_1| \ge n_\sigma(m, k - \frac{1}{2\sigma})$. In the first case I_0 contains $m - \frac{1}{2\sigma}$ independent vertices (otherwise G would have a subgraph on k vertices of density at least σ), which form, together with the vertices of I, m independent vertices in G – a contradiction. If $|I_1| \ge n_\sigma(m, k - \frac{1}{2\sigma})$, then there is a vertex set $I_2 \subseteq I_1, |I_2| = k - \frac{1}{2\sigma}$, such that the density of the subgraph of G induced by I_2 is at least σ. Since

$$\left|\binom{I}{2}\right| = \frac{1}{2\sigma}(\frac{1}{4\sigma} - \frac{1}{2}) < \frac{1}{2\sigma}(k - \frac{1}{2\sigma}) = \frac{1}{2\sigma}|I_2|,$$

the density of the subgraph induced by the k vertices of $I \cup I_2$ is then

$$\frac{\left|\binom{I_2}{2} \cap E(G)\right| + |(I \times I_2) \cap E(G)|}{\left|\binom{I}{2}\right| + \left|\binom{I_2}{2}\right| + |I \times I_2|} > \frac{\sigma\left|\binom{I_2}{2}\right| + |I_2|}{\frac{1}{2\sigma}|I_2| + \left|\binom{I_2}{2}\right| + \frac{1}{2\sigma}|I_2|} = \sigma,$$

which is again a contradiction. This completes the proof of (a).

(b) Let G be a graph on $k \ge \frac{1}{\sigma}$ vertices with independence number at most $\frac{1}{2\sigma}$. By Turán's theorem [Tu54], the density of G is at least

$$\frac{\frac{1}{2\sigma}\binom{2\sigma k}{2}}{\binom{k}{2}} = \frac{2\sigma k - 1}{k-1} = \frac{\sigma k}{k-1} + \frac{\sigma k - 1}{k-1} > \sigma + 0 = \sigma.$$

By Observation 3, G contains a subgraph on k vertices of density at least σ.

(c) Let G be a graph on $m + \frac{1}{8\sigma}$ vertices with $\alpha(G) \le m$. Certainly, G contains at least $\frac{1}{8\sigma}$ edges. The density of any subgraph of G on $\frac{1}{2\sigma}$ vertices with at least $\frac{1}{8\sigma}$ edges is at least

$$\frac{\frac{1}{8\sigma}}{\frac{\frac{1}{2\sigma}(\frac{1}{2\sigma}-1)}{2}} > \sigma.$$

□

Consequence 8 *If $\frac{1}{2\sigma}$ is an integer and m, k are two positive integers divisible by $\frac{1}{2\sigma}$, then*

$$n_\sigma(m,k) \quad < \quad \frac{1}{\sigma}\binom{2\sigma(m+k)}{2\sigma k} - \frac{1}{2\sigma}.$$

Proof. We fix σ such that $\frac{1}{2\sigma}$ is an integer, and proceed by induction on $m + k$.

For $k = \frac{1}{2\sigma}$, Consequence 8 follows from Lemma 7(c):

$$n_\sigma(m, \frac{1}{2\sigma}) \le m + \frac{1}{8\sigma} < \frac{1}{\sigma}(2\sigma m + 1) - \frac{1}{2\sigma} = \frac{1}{\sigma}\binom{2\sigma(m+k)}{2\sigma k} - \frac{1}{2\sigma}.$$

For $m = \frac{1}{2\sigma}$ and $k > \frac{1}{2\sigma}$, Consequence 8 follows from Lemma 7(b):

$$n_\sigma(\frac{1}{2\sigma}, k) = k < \frac{1}{\sigma}(2\sigma k + 1) - \frac{1}{2\sigma} = \frac{1}{\sigma}\binom{2\sigma(m+k)}{2\sigma k} - \frac{1}{2\sigma}.$$

To complete the proof, it suffices to exhibit the second inductive step for $m > \frac{1}{2\sigma}, k > \frac{1}{2\sigma}$:

$$n_\sigma(m,k) \quad < \quad n_\sigma(m - \frac{1}{2\sigma}, k) + n_\sigma(m, k - \frac{1}{2\sigma}) + \frac{1}{2\sigma}$$