A Ramsey-type problem and the Turán numbers

N. Alon[*] P. Erdős[†] D. S. Gunderson[‡] M. Molloy[§¶]

28 February 2000

Abstract

For each n and k, we examine bounds on the largest number m so that for any k-coloring of the edges of K_n there exists a copy of K_m whose edges receive at most $k-1$ colors. We show that for $k \geq \sqrt{n} + \Omega(n^{1/3})$, the largest value of m is asymptotically equal to the Turán number $t(n, \lfloor \binom{n}{2}/k \rfloor)$, while for any constant $\epsilon > 0$, if $k \leq (1-\epsilon)\sqrt{n}$ then the largest m is asymptotically larger than that Turán number.

1 Introduction

For any finite set S and positive integer k, we use the notation $[S]^k = \{T \subset S : |T| = k\}$. A graph G is an ordered pair $(V, E) = (V(G), E(G))$ where V is a finite set and $E \subset [V]^2$. Elements of V are called vertices and elements of E are called edges. If G is a graph with $|V(G)| = n$ and $E(G) = [V]^2$, then we say that G is a complete graph on n vertices, denoted by K_n. It will be convenient to use $V(K_n) = [n] = \{1, \ldots, n\}$ and $E(K_n) = [n]^2$. The complement of G will be denoted by $\overline{G} = (V(G), [V(G)]^2 \backslash E(G))$; so, for example, $\overline{K_n}$ is an independent set of n vertices.

[*]Dept. of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Research supported in part by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. Email: noga@math.tau.ac.il

[†]Hungarian Academy of Sciences, Budapest, Hungary.

[‡]Dept of Mathematics and Statistics, University of Calgary and Dept of Mathematics, Physics and Engineering, Mount Royal College, Calgary, Canada. Email: triangle@math.ucalgary.ca

[§]Dept. of Computer Science, University of Toronto, Toronto, Canada. Research supported in part by NSERC. Email: molloy@cs.toronto.edu

[¶]The remaining authors would like to dedicate this paper to the memory of Paul Erdős.

The standard Ramsey arrow notation $n \to (m)^2_k$ means that for every k-coloring of the edges of K_n, there exists a copy of K_m all of whose edges receive the same color. In two early papers on Ramsey theory ([3], [5]) it was shown that

$$n \to \left(\frac{\log n}{2\log 2}\right)^2_2 \quad \text{and} \quad n \not\to \left(\frac{2\log n}{\log 2}\right)^2_2.$$

When more than two colors are used, techniques from these two papers show that there are universal constants c_1 and c_2, so that for each $k \geq 2$,

$$n \to \left(\frac{c_1 \log n}{k\log k}\right)^2_k \quad \text{and} \quad n \not\to \left(\frac{c_2\log n}{\log k}\right)^2_k.$$

Erdős, Hajnal, and Rado [4] introduced another Ramsey arrow notation, $n \to [m]^2_k$, meaning that for any k-coloring of $E(K_n)$, there exists a copy of K_m whose edges receive at most $k-1$ colors. (When $k = 2$, this is the standard Ramsey arrow.) This concept was also examined in [8]. It was shown in [6] that there exists a universal constant c_1 so that for every n and $k \leq n$,

$$n \to \left[\frac{c_1 k}{\log k}\log n\right]^2_k, \tag{1}$$

and techniques from [3] give a universal constant c_2 so that

$$n \not\to [c_2 k \log n]^2_k. \tag{2}$$

Rather than being given an m and k and finding bounds on n, we shall be concerned with bounds on m given n and k.

Definition 1.1 *Let $f(n,k)$ be the largest integer so that $n \to [f(n,k)]^2_k$.*

Thus for a given n, under any k-coloring of $E(K_n)$, there is always a clique of size at least $f(n,k)$ which is not *completely multi-colored.* Restating equations (1) and (2), there are universal constants c_1 and c_2 so that for $k \leq n$:

$$\frac{c_1 k}{\log k}\log n \leq f(n,k) < c_2 k \log n. \tag{3}$$

In particular, for k constant, $f(n,k) = \Theta(\log n)$. In this paper, we are interested in the asymptotic order of $f(n,k)$ when k is a function of n.

2 Preliminaries

Lemma 2.1 *If* $2 \leq k < n$*, then* $f(n,k) \leq f(n+1,k)$ *and* $f(n,k) \leq f(n,k+1)$*.*

Proof: That $f(n,k) \leq f(n+1,k)$ follows immediately from the fact that K_{n+1} has K_n as a subgraph.

To prove that $f(n,k) \leq f(n,k+1)$, let $f(n,k) = m$ and let a $(k+1)$-coloring of $E(K_n)$ be given. Examine an auxiliary k-coloring defined by identifying two color classes. By the choice of m, there exist m vertices which induce at most $k-1$ colors from the auxiliary coloring, hence at most k colors of the original $(k+1)$-coloring. Hence, $f(n,k+1) \geq m$. □

It is not difficult to check that $f(6,3) < 4$ (by giving a 3-coloring of $E(K_6)$ under which every K_4 is 3-colored—one such coloring uses three paths of length 5). It easily follows that $f(6,3) = 3$. Since $f(6,2) = 3$ as well, the inequality $f(n,k) \leq f(n,k+1)$ is not always a strict inequality.

Let $t(n,m)$ denote the maximum t so that every graph with n vertices and m edges has an independent set of size at least t. This function is known precisely for all n and m, by Turán's Theorem. It is the number of cliques in the appropriate Turán graph, which we denote here by $T(n,m)$. This is the disjoint union of the minimum possible number of nearly equal cliques on n vertices whose total number of edges is at most m. We define

$$g(n,k) = t(n, \lfloor \binom{n}{2} / k \rfloor).$$

This yields an easy lower bound for $f(n,k)$, which was also observed in [8]:

Lemma 2.2 $f(n,k) \geq g(n,k)$*.*

Proof: Fix any k-coloring $E(K_n) = E_1 \cup \cdots \cup E_k$. Some color class, say E_1, has at most $\lfloor \binom{n}{2}/k \rfloor$ edges. By Turán's theorem the subgraph formed by E_1 contains an independent set of t vertices, and this independent set is a t-clique in G that is at most $(k-1)$-colored. □

The main focus of this paper will be to determine for which values of n, k, equality holds, or at least nearly holds, in Lemma 2.2. Note that if there is a coloring of $E(K_n)$ in which each color class is a copy of the appropriate Turán graph, then equality holds. Similarly, for equality to nearly hold, we must find a coloring in which each color class is very close to the Turán graph.

Here we show that $g(n,k)$ is the precise value of $f(n,k)$ for all $k \geq n$, (as well as in several other cases), that for every $k \geq (1+\epsilon)\sqrt{n}$

$$f(n,k) = (1+o(1))g(n,k),$$

where the $o(1)$ term tends to 0 as n (and hence k) tend to infinity, and that for every $\epsilon > 0$ there exist n_0 and $\delta > 0$ such that if $k < (1-\epsilon)\sqrt{n}, n > n_0$ then

$$f(n,k) > (1+\delta)g(n,k).$$

It will be helpful to observe that if both k and n/k tend to infinity then $g(n,k) = (1+o(1))k$.

3 Case 1: $k \geq n$

In Theorem 3.1 we shall employ a result about *balanced* edge-colorings. A *proper* edge-coloring of a graph is one where no two incident edges have the same color. An edge-coloring of a graph G using k colors is called *balanced* if each color class has either $\lfloor |E(G)|/k \rfloor$ or $\lceil |E(G)|/k \rceil$ edges. The following fact is well known:

Fact: If a graph G has a k-edge coloring, then it has a balanced proper k-edge coloring.

Theorem 3.1 *For each $k \geq n$, $f(n,k) = g(n,k)$.*

Proof: We use a balanced proper k-coloring of $E(K_n)$ and observe that each color class contains the appropriate Turán graph, which in this case is simply a matching of size $\lfloor \binom{n}{2}/k \rfloor$ or $\lceil \binom{n}{2}/k \rceil$. □

4 Case 2: $k = n/\alpha$ for a fixed real $\alpha \geq 1$

We need the following well known result of Wilson [12] on graph decomposition. We say that a graph G has an H-decomposition if the set of its edges can be colored such that each color class forms a copy of H. These colored copies are called the members of the decomposition.

Theorem 4.1 (Wilson) *Let $H = (V,E)$ be a graph with q edges and let g denote the greatest common divisor of the degrees of H. Then there is an $n_0 = n_0(H)$ such that for every $n > n_0$ for which $\binom{n}{2}$ is divisible by q and $n-1$ is divisible by g, K_n has an H-decomposition.*

We need to show that we can find complete graphs which have H-decompositions with additional regularity properties.

Lemma 4.2 *For every graph H with h vertices and q edges, there is a $2q$-regular graph H_1 on h^4 vertices which has an H-decomposition, such that each vertex of H_1 is incident with precisely h members of the decomposition.*

Proof: We begin with the following:
Claim There exist $g_1, ..., g_h \in \{0, ..., h^4 - 1\}$ such that the $h^2 - h$ differences $g_i - g_j \pmod{h^4}$ are pairwise distinct.

To see this, we take a uniform random choice of $g_1, ..., g_h$. The probability that $(g_{i_1}, g_{j_1}), (g_{i_2}, g_{j_2})$ with, say, $j_2 \neq i_1, j_1$, have the same difference is at most $\frac{1}{h^4-3}$ since after fixing $g_{i_1}, g_{j_1}, g_{i_2}$, there are still at least $h^4 - 3$ choices for g_{j_2}, at most one of which yields the same difference. The case $j_1 = i_2, j_2 = i_1$ (for even h) has an even lower probability. Therefore, the expected number of pairs which have the same difference is at most $\binom{h^2-h}{2}\frac{1}{h^4-3} < 1$ and so there is at least one choice for which this number is zero.

Now we simply take the copy of H formed by mapping its vertex number i to g_i, and let H_1 consist of this copy and all the $h^4 - 1$ cyclic shifts of it. $\square$

Corollary 4.3 *For every graph H there is a complete graph K_m which has an H-decomposition so that each vertex of K_m is incident with the same number of members of the decomposition.*

Proof: This follows from applying Wilson's Theorem to H_1 from Lemma 4.2. $\square$

Corollary 4.4 *For every graph H there are $c = c(H)$ and $n_0 = n_0(H)$ such that for every $n > n_0(H)$ there is some n' satisfying $n \leq n' \leq n + c$ for which $K_{n'}$ has an H-decomposition in which each vertex is incident with the same number of members of the decomposition, and any two vertices lie in at most c common members of the decomposition.*

Proof: First apply Wilson's Theorem to K_m from Corollary 4.3 to decompose $K_{n'}$ to copies of K_m, for an appropriate value of $n \leq n' \leq n+m(m-1)$. Then decompose each such K_m copy into copies of H. Note that each pair of vertices can lie only in copies of H that lie in the same K_m. $\square$

This yields the main theorem of this section:

Theorem 4.5 *If $k = n/\alpha$ and $\alpha \geq 1$ is a fixed real, then $f(n,k) = (1 + o(1))g(n,k)$.*

Proof: It is useful to note that for this case, $g(n,k) = \Theta(n)$.

Consider any $\epsilon > 0$. We will prove that $f(n', k') \leq (1+\epsilon)g(n,k)$ for a suitable $n' \geq n$ and $k' \geq k$. Our result then follows from Lemma 2.1. Our goal is to find a k'-colouring of $E(K_{n'})$ such that each colour class contains a Turán graph on n' vertices with independence number $t \leq (1+\frac{\epsilon}{2})g(n,k)$. Suppose that $a < \alpha+1 \leq a+1$ for an integer a. A straightforward calculation shows that we can choose $(1+\frac{\epsilon}{4})g(n,k) \leq t \leq (1+\frac{\epsilon}{2})g(n,k)$ so that the corresponding Turán graph will contain rn' disjoint a-cliques and sn' disjoint $(a+1)$-cliques, so long as rn', sn' are integers, where r, s are rationals with denominators bounded by some constant function of α, ϵ.

Now we define H to be a collection of rh disjoint a-cliques and sh disjoint $(a+1)$-cliques, where h is the LCM of the denominators of r, s. Take some n' as in Corollary 4.4 along with the corresponding decomposition of $K_{n'}$. Define $\mathcal{H}$ to be the $|H|$-uniform hypergraph whose vertices are the vertices of $K_{n'}$, and whose edges are the vertex sets of the copies of H. This hypergraph is κ-regular for some $\kappa \geq (1+\delta)k$ where $\delta = \delta(\epsilon)$, and any pair of vertices lies in at most $c = c(\alpha)$ edges. Therefore, by the main theorem of [9], it has a proper edge-colouring C using $c = \kappa(1+o(1))$ colours.

We use C to find our edge coloring of $K_{n'}$ as follows. Since $\mathcal{H}$ is $|H|$-uniform and κ-regular, $|E(\mathcal{H})| = \frac{\kappa n'}{|H|}$ and no colour class contains more than $\frac{n'}{|H|}$ hyperedges. We remove from C all color classes which contain fewer than $\frac{n'}{|H|}(1-\gamma)$ hyperedges (i.e., we uncolor all hyperedges belonging to one of those classes), where γ is a positive constant such that $\gamma n' < \frac{\epsilon}{4}g(n,k)$. A simple calculation yields that we remove $o(\kappa)$ colour classes, and so the number of remaining colour classes is $k' = \kappa - o(\kappa) > k$. Furthermore, the subgraph induced by any remaining colour class has independence number at most $t + \gamma\frac{n'}{|H|} \times |H| < t + \frac{\epsilon}{4}g(n,k) < (1+\epsilon)g(n,k)$. We complete our edge colouring of K_n by assigning to any uncoloured edge, an arbitrary colour from amongst the remaining colour classes of C. As this won't increase the independence number of a colour class, we obtain our desired colouring. $\square$

5 Case 3: $k = o(n), k^2 > (1+o(1))n$

For an integer $q \geq 2$, Lemma 2.2 yields $f(q^2, q+1) \geq q$. The next lemma shows that this bound is tight when q is a power of a prime.

Lemma 5.1 *If q is a power of a prime, then $f(q^2, q+1) = g(q^2, q+1) = q$.*

Proof: The coloring is given by the affine plane. For completeness we describe it in details. Let F_q denote the field of order q. The affine plane of order q is the geometry on the q^2 points $\{(x, y) : x, y \in F_q\}$, where lines are solutions to equations of the form $ax + by = c$, where a, b, and c are constants from the field with a and b both not zero (so there are $q^2 + q$ lines, each with q points). Let $n = q^2$ and define a $(q+1)$-coloring of the complete graph on the points of the affine plane of order q by assigning to each edge the slope of the line containing its endpoints. Each line of the plane induces a monochromatic clique of size q and every color class is determined by a parallel class of lines, so is formed by a union of q cliques each of size q. So each color class determines a subgraph of K_{q^2} with no independent set of size $q + 1$. Thus every set of $q + 1$ points induces every color, that is, $f(q^2, q + 1) < q + 1$, and so, by the preceding discussion, $f(q^2, q + 1) = q$. □

This, together with Lemma 2.1 and well known results about the distributions of primes implies the following

Corollary 5.2 *For every $k \geq \sqrt{n} + \Omega(n^{1/3})$, $f(n, k) \leq (1 + o(1))k$. Therefore, if, in addition, $k = o(n)$, then $f(n, k) = (1+o(1))g(n, k) = (1+o(1))k$.*

Remark: The assertion of Lemma 5.1 can be easily generalized. In fact, any resolvable balanced incomplete block design supplies an example in which $f(n, k) = g(n, k)$ precisely. In particular, one can use the lines in an affine geometry of dimension d to prove the following

Claim 1: For any prime power q and any integer $d > 1$

$$f(q^d, q^{d-1} + q^{d-2} + \ldots + 1) = g(q^d, q^{d-1} + q^{d-2} + \ldots + 1) = q^{d-1}.$$

Similarly, the existence of Kirkman's triple systems (c.f., e.g., [2]) gives the following

Claim 2: For every $n \equiv 3(mod\ 6)$,

$$f(n, (n-1)/2) = g(n, (n-1)/2) = n/3.$$

The existence of resolvable Steiner systems of the form $S(2, 4, n)$ for all $n \equiv 4(\ mod\ 12)$ (c.f., e.g., [2]) implies

Claim 3: For every $n \equiv 4(mod\ 12)$,

$$f(n, (n-1)/3) = g(n, (n-1)/3) = n/4.$$

6 Smaller values of k

We will need the following lemma:

Lemma 6.1 *Let $d_1, d_2, \ldots, d_n$ be n non-negative reals whose average value is d, and suppose that at least γn of them are at least $(1+\gamma)d$. Then*

$$\sum_{i=1}^{n} \frac{1}{d_i+1} \geq \frac{n}{(d+1)}(1+\Omega(\gamma^3)).$$

Proof: By the convexity of $f(z) = 1/z$, the minimum of the above sum, subject to $\sum d_i = dn$, is obtained when $d_i = d$ for all i. As long as there is some $d_i > (1+\gamma)d$ there is some other $d_j \leq d$, and replacing each of them by their average decreases the sum by $\frac{1}{d_i+1} + \frac{1}{d_j+1} - \frac{4}{d_i+d_j+2} = \Omega(\gamma^2/(d+1))$. Since we can perform at least γn steps of this form the desired estimate follows. □

Let $\alpha(G)$ denote the maximum size of an independent set in G.

Lemma 6.2 *For every $\epsilon > 0$ there is a $\delta > 0$ and d_0 such that for any graph $G = (V, E)$ on n vertices with average degree at most d, $d > d_0$, and maximum clique size at most $(1-\epsilon)d$,*

$$\alpha(G) \geq (1+\delta)\frac{n}{d+1}.$$

Proof: Let $d_1, \ldots, d_n$ be the degrees of the vertices of G. If there are at least $\frac{\epsilon}{4}n$ degrees which exceed $(1+\frac{\epsilon}{4})d$, then the result, with $\delta = \Omega(\epsilon^3)$, follows from Lemma 6.1 together with the well known fact that

$$\alpha(G) \geq \sum_{i=1}^{n} \frac{1}{d_i+1},$$

(see, e.g., [1], page 81, for a short proof.)

Otherwise, omit all vertices of degree that exceeds $(1+\frac{\epsilon}{4})d$. The remaining graph has at least $(1-\frac{\epsilon}{4})n$ vertices, has maximum degree $\Delta \leq (1+\frac{\epsilon}{4})d$ and has maximum clique size $\omega \leq (1-\epsilon)d$. Fajtlowicz[7] has shown that for every graph G,

$$\alpha(G) \geq \frac{2|V(G)|}{\omega(G)+\Delta(G)+1},$$

(see also [10]). The result now follows with $1+\delta = (1-\frac{\epsilon}{4})/(1-\frac{3\epsilon}{8})$. □

Theorem 6.3 *For every $\epsilon > 0$ there is a $\delta > 0$ and n_0 such that for every $n > n_0$ and $2 \leq k \leq (1-\epsilon)\sqrt{n}$,*

$$f(n,k) \geq (1+\delta)g(n,k).$$

Proof: In the above range $g(n,k) = \Theta(k)$. Since for every $k \geq 2$, $f(n,k) = \Omega(\log n)$ the assertion is trivial for $k = o(\log n)$, and we thus may and will assume that $k \geq \Omega(\log n)$. Thus $g(n,k) = (1+o(1))k$. Consider a coloring of $E(K_n)$, $n > n_0$, by k colors, and let G be the graph consisting of all edges of the least popular color. Then the average degree of G is at most $(n-1)/k < n/k$. Fix a $\gamma = \gamma(\epsilon) > 0$ such that $(1-\gamma)\frac{n}{k} > (1+\gamma)k$. If G contains a clique of size at least $(1-\gamma)n/k$, then the induced subgraph on the vertices of this clique misses a color (in fact, misses all colors but one) and is of size at least $(1+\gamma)k > (1+\gamma/2)g(n,k)$ for $n > n_0$. Otherwise, by Lemma 6.2, $\alpha(G) \geq (1+\delta)k$ for some $\delta = \delta(\gamma) > 0$, and so G has a large independent set corresponding to a complete subgraph of size $(1+\delta)k$ which does not contain the least popular color. □

7 Concluding remarks

Turán's theorem implies lower bounds on $f(n,k)$, and we have found that these lower bounds are tight (or nearly tight) whenever k is slightly bigger than $\sqrt{n}$, and are not nearly tight whenever k is slightly smaller than $\sqrt{n}$. The problem of finding an asymptotic formula for $f(n,k)$ for smaller values of k is more difficult (and in particular that of finding an asymptotic formula for $f(n,2)$ is a well known, difficult open problem in Ramsey theory). It would be interesting to find an estimate, up to a constant factor, for $f(n,k)$, for smaller values of k. Thus, for example, when $k \sim \sqrt{n}/\log n$, $f(n,k)$ is at least $\frac{\sqrt{n}}{\log n}$ and at most $O(\sqrt{n})$, and it would be interesting to find a more accurate estimate.

Acknowledgments: Thanks to A. Rosa for helpful discussions.

Remark: The research in this paper began with discussions between Erdős, Gunderson and Molloy in the summer of 1997 at a workshop in Annecy, France. It continued the following week with discussions between Alon and Erdős at a conference in Balatonlelle, Hungary. By the end of those collaborations, we had obtained some significant results and so planned to write a paper. A few months later, Erdős passed away, and due mostly to procrastination on the part of the remaining authors, the paper was not completed until March 2000. During the intervening years, some results in the paper were improved, but we strongly feel that the the crucial early collaborations fully justify the inclusion of Erdős as a coauthor.

References

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

[2] A. E. Brouwer, Block Designs, in: "Handbook of Combinatorics", R.L. Graham, M. Grötschel and L. Lovász, eds, North Holland (1995), Chapter 14, pp. 693-745.

[3] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.

[4] P. Erdős, A. Hajnal, and R. Radó, Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 93–196.

[5] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470.

[6] P. Erdős and A. Szemerédi, On a Ramsey type theorem, *Periodica Mathematica Hungarica* **2** (1972), 295–299.

[7] S. Fajtlowicz, On the size of independent sets in graphs, *Congressus Numerantium* **21** (1978), 296 - 274.

[8] H. Harborth and M. Möller, Weakened Ramsey numbers, Discrete Applied Math. 95 (1999), 279-284.

[9] N. Pippenger and J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs, *J. Combin. Th. Ser. A* **51** (1989), 24–42.

[10] B. Reed, ω, Δ , and χ, *J. Graph Th.* **27** (1998), 177 - 212.

[11] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz Lapook* **48** (1941), 436–452.

[12] R. M. Wilson, Decomposition of complete graphs into subgraphs isomorphic to a given graph, *Congressus Numerantium* **XV** (1975), 647-659.