© 2008 Society for Industrial and Applied Mathematics

LARGE NEARLY REGULAR INDUCED SUBGRAPHS*

NOGA ALON†, MICHAEL KRIVELEVICH‡, AND BENNY SUDAKOV§

Abstract. For a real $c \geq 1$ and an integer n, let $f(n, c)$ denote the maximum integer f such that every graph on n vertices contains an induced subgraph on at least f vertices in which the maximum degree is at most c times the minimum degree. Thus, in particular, every graph on n vertices contains a regular induced subgraph on at least $f(n, 1)$ vertices. The problem of estimating $f(n, 1)$ was posed long ago by Erdős, Fajtlowicz, and Staton. In this paper we obtain the following upper and lower bounds for the asymptotic behavior of $f(n, c)$: (i) For fixed $c > 2.1$, $n^{1-O(1/c)} \leq f(n, c) \leq O(cn/\log n)$. (ii) For fixed $c = 1 + \varepsilon$ with $\varepsilon > 0$ sufficiently small, $f(n, c) \geq n^{\Omega(\varepsilon^2/\ln(1/\varepsilon))}$. (iii) $\Omega(\ln n) \leq f(n, 1) \leq O(n^{1/2} \ln^{3/4} n)$. An analogous problem for not necessarily induced subgraphs is briefly considered as well.

Key words. induced subgraphs, regular subgraphs, Ramsey-type problem

AMS subject classification. 05C35

DOI. 10.1137/070704927

1. Introduction. All graphs considered here are finite and simple. For a graph $G = (V, E)$, let $\Delta(G), \delta(G)$, and $d(G) = \frac{2|E|}{|V|}$ denote its maximum degree, minimum degree, and average degree, respectively. The *density* of G is $p = |E|/\binom{|V|}{2}$; clearly, this is a number between 0 and 1. For $U \subseteq V$, let $G[U]$ denote the subgraph of G induced on U.

DEFINITION 1. *A graph G is c-nearly regular if $\Delta(G) \leq c \cdot \delta(G)$.*

For a graph $G = (V, E)$ and a constant $c \geq 1$, let

$$f(G, c) = \max\{|U| : G[U] \text{ is a } c\text{-nearly regular graph}\}.$$

Define

$$f(n, c) = \min\{f(G, c) : |V(G)| = n\}.$$

Thus, every graph G on n vertices contains a c-nearly regular induced subgraph on at least $f(n, c)$ vertices. In particular, for $c = 1$ every such G contains a strictly regular induced subgraph on at least $f(n, 1)$ vertices.

The problem of estimating $f(n, 1)$ was posed by Erdős, Fajtlowicz, and Staton (cf. [3] or [2, page 85]). By the known estimates for graph Ramsey numbers (cf.,

---

e.g., [6]), every graph on n vertices contains either a clique or an independent set of order $\Omega(\ln n)$. This implies that $f(n, 1) \geq \Omega(\ln n)$. Erdős, Fajtlowicz, and Staton conjectured that the ratio $f(n, 1)/\ln n$ tends to infinity as n tends to infinity. We are unable to prove or disprove this conjecture and can only obtain several bounds, which are listed in the following results. The first deals with the case of large c.

PROPOSITION 1.1. *There exists an absolute constant b so that for all $K \geq 2.1$, $f(n, K) \geq n^{1-b/K}$.*

The problem of obtaining a nontrivial lower bound for values of c close to 1 is more interesting. Here we first deal with the case of graphs with positive density and show that any such graph must contain a nearly regular subgraph on a linear number of vertices.

THEOREM 1.2. *Let $\varepsilon > 0$ be a small real, and let p satisfy $0 < p < 1$. Then, for every sufficiently large n, any graph $G = (V, E)$ on n vertices with density at least p contains an induced $(1 + \varepsilon)$-nearly regular subgraph on at least*

$$0.5 \left(\frac{\varepsilon}{6}\right)^{\frac{144}{\varepsilon^2} \ln(\frac{1}{p})} \cdot n$$

vertices.

For general (possibly sparse) graphs we have the following.

THEOREM 1.3. *Let $\varepsilon > 0$ be a sufficiently small constant. Then*

$$f(n, 1 + \varepsilon) \geq n^{\frac{\varepsilon^2}{250 \ln(1/\varepsilon)}}$$

for all sufficiently large n.

Our upper bounds for $f(n, c)$ are rather far from the lower bounds. For the strictly regular case we prove the following.

THEOREM 1.4. $f(n, 1) \leq O(n^{1/2} \log^{3/4} n)$.

This is a slight improvement of an earlier estimate of Bollobás (cf. [2]), who showed that for every $\epsilon > 0$, $f(n, 1) \leq c(\epsilon)n^{1/2+\epsilon}$. For the nearly regular case we have the following.

PROPOSITION 1.5. *For every constant $K \geq 2$, $f(n, K) \leq 7K\frac{n}{\log n}$.*

The lower bounds are proved in the next section, and the upper bounds are presented in section 3. We conclude in section 4 with a few open problems and a brief discussion of an analogous problem for not necessarily induced subgraphs. Throughout this paper we assume, whenever needed, that the number of vertices n of the graphs discussed is sufficiently large. To simplify the presentation, we make no attempt to optimize the absolute constants, and we omit all floor and ceiling signs whenever these are not crucial. We also use the following standard asymptotic notation: For two functions $f(n)$, $g(n)$ of a natural valued parameter n, we write $f(n) = o(g(n))$ whenever $\lim_{n\to\infty} f(n)/g(n) = 0$; $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $f(n) \leq Cg(n)$ for all n; and $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$.

2. Lower bounds. For a graph $G = (V, E)$ and a subset $U \subseteq V$, the number of edges of G spanned by U in G is denoted by $e_G(U)$; the number of edges between disjoint subsets U, W of vertices of G is denoted by $e_G(U, W)$.

2.1. Large c. In this subsection we prove Proposition 1.1, which provides a lower bound for $f(n, c)$ when c is a relatively large constant.

We need the following rather standard argument, allowing one to pass from a graph with a large average degree to one with a large minimum degree.

PROPOSITION 2.1. *Let $K > 1, \alpha < 1/2$ be constants. Then every graph $G = (V, E)$ on $|V| = n$ vertices with $\Delta(G) \leq Kd(G)$ contains an induced (K/α)-nearly regular subgraph G^* with at least $\frac{1-2\alpha}{K-2\alpha}n$ vertices and at least $\frac{K-2K\alpha}{2K-4\alpha}nd$ edges.*

Proof. Denote the average degree $d(G)$ of G by d. We can obviously assume that $d > 0$. Start with G and delete repeatedly vertices of degree less than αd until there are none left. Denote the resulting graph by G^*. Then G^* is an induced subgraph of G, satisfying $\Delta(G^*) \leq \Delta(G)$, $\delta(G^*) \geq \alpha d$, implying

$$\Delta(G^*) \leq \Delta(G) \leq Kd \leq \frac{K}{\alpha}\delta(G^*),$$

and thus G^* is a (K/α)-nearly regular graph. We now estimate the number of vertices of G^*. Denote the latter by t. While creating G^* from G, we deleted less than $(n-t)\alpha d$ edges, and thus

$$\Delta(G^*) \geq d(G^*) > \frac{2(|E(G)| - (n-t)\alpha d)}{t} = \frac{nd - 2(n-t)\alpha d}{t}.$$

But $\Delta(G^*) \leq \Delta(G) \leq Kd$, implying

$$\frac{nd - 2(n-t)\alpha d}{t} \leq Kd.$$

Solving the above inequality for t, we get $t \geq \frac{1-2\alpha}{K-2\alpha}n$, supplying the required lower bound for the number of vertices of G^*. To bound the number of its edges, note that the number of vertices deleted is $n - t$, and hence the number of edges deleted is at most $(n-t)\alpha d$, leaving at least

$$\frac{1}{2}nd - (n-t)\alpha d = \left(\frac{1}{2} - \alpha\right)nd + t\alpha d \geq \left(\frac{1}{2} - \alpha\right)nd + \frac{1-2\alpha}{K-2\alpha}n\alpha d = \frac{K-2K\alpha}{2K-4\alpha}nd,$$

as needed. □

Remark 1. It is instructive to observe that the above argument breaks down completely for $\alpha \geq 1/2$. Therefore, when estimating $f(n,c)$ from below for small c, in particular for $c < 2$, we will adapt a different strategy.

We proceed with the following result, whose proof resembles that of one of the results in [4].

PROPOSITION 2.2. *Let $K > 1$ be a constant. Every graph $G = (V, E)$ on $|V| = |V(G)| = n$ vertices contains an induced subgraph G^* on at least $n^{1+\log_2\left(1-\frac{1}{K}\right)}$ vertices, for which $\Delta(G^*) \leq Kd(G^*)$.*

For large K the above estimate behaves like $n^{1-\Theta\left(\frac{1}{K}\right)}$. Therefore, the assertions of Propositions 2.1 and 2.2 imply that of Proposition 1.1.

Proof of Proposition 2.2. Set $G_0 = G$, $k^* = \log_2 n$. For $i = 0, \ldots, k^*$ repeat the following loop. Set

$$n_i = |V(G_i)|, \quad \Delta_i = \Delta(G_i), \quad d_i = d(G_i).$$

If $\Delta_i \leq Kd_i$, abort the loop. Otherwise, delete repeatedly vertices of degree at least $Kd_i/2$ from G_i until there are none left. Let G_{i+1} be the resulting graph and increment i.

Denote by G^* the resulting graph of the above described process. Observe that at iteration i we delete at most $|E(G_i)|/(Kd_i/2) = (n_id_i/2)(Kd_i/2) = n_i/K$ vertices,

and thus $n_{i+1} \geq (1 - 1/K)n_i$. It follows that

$$|V(G^*)| \geq \left(1 - \frac{1}{K}\right)^{k^*} n = n^{1 + \log_2\left(1 - \frac{1}{K}\right)} .$$

If G^* was created when the above loop was aborted due to $\Delta(G_i) \leq Kd(G_i)$, then obviously the obtained graph satisfies the assertion of the theorem. Otherwise, G^* was obtained after k^* iterations. At each such iteration i, we have $\Delta_{i+1} \leq Kd_i/2$ and $d_i \leq \Delta_i/K$, implying $\Delta_{i+1} \leq \Delta_i/2$. Therefore, in this case

$$\Delta(G^*) \leq \Delta_0 \cdot \left(\frac{1}{2}\right)^{k^*} < n \cdot \left(\frac{1}{2}\right)^{\log_2 n} = 1,$$

implying that G^* has no edges, and thus $\Delta(G^*) = d(G^*) = 0$, and G^* can again serve as the required graph.   □

2.2. Small c. Next we treat the more challenging case where the constant c in $f(n, c)$ is very close to 1. Throughout this subsection ε denotes a small positive real. We start with several lemmas.

LEMMA 2.3. *Let $G = (V, E)$ be a graph on n vertices with density p, and let $\varepsilon > 0$. Then G contains an induced subgraph G' of density $p' \geq p$ on a set of*

$$n' \geq \varepsilon^{\frac{2}{\varepsilon} \ln(\frac{1}{p})} n$$

vertices so that every set of $t \geq \varepsilon n'$ vertices of G' spans at most $\binom{t}{2} p'(1 + \varepsilon)$ edges.

Proof. Set $G_0 = G$. For $i = 0, 1, \ldots$ repeat the following loop. Set

$$n_i = |V(G_i)|, \quad m_i = |e(G_i)|, \quad p_i = \frac{m_i}{\binom{n_i}{2}} .$$

If G_i contains a subset $U \subseteq V(G_i)$ of at least εn_i vertices such that $e_{G_i}(U) \geq \binom{|U|}{2}p_i(1 + \varepsilon)$, then set $G_{i+1} := G_i[U]$, $i := i + 1$.

Observe that after k iterations of the above loop, the density p_k of the current graph G_k satisfies $p_k \geq (1+\varepsilon)^k p_0$. Thus, if the loop is repeated at least $\frac{2}{\varepsilon} \ln(\frac{1}{p})$ times, we have

$$p_k \geq (1 + \varepsilon)^{\frac{2}{\varepsilon} \ln(\frac{1}{p})} \cdot p > 1,$$

which is a contradiction. It follows that the above process concludes after less than $\frac{2}{\varepsilon} \ln(\frac{1}{p})$ iterations. The resulting graph G_k has n_k vertices and m_k edges. Observe that at each iteration the number of vertices of the new graph is at least an ε-proportion of the number of vertices of the previous graph. Therefore,

$$n_k \geq \varepsilon^k |V(G)| \geq \varepsilon^{\frac{2}{\varepsilon} \ln(\frac{1}{p})} n.$$

We can thus take $G' = G_k$, $n' = n_k$ to complete the proof.   □

LEMMA 2.4. *Let $G = (V, E)$ be a graph on $|V| = n$ vertices with $m = |E|$ edges and density $p = m/\binom{n}{2} \geq n^{-a}$ for some constant $0 < a < 1$. Suppose that*

(1)   *every $t \geq \varepsilon n$ vertice in G spans at most $\binom{t}{2}p(1 + \varepsilon)$ edges.*

Then for every subset $U \subseteq V$ of cardinality $|U| = \varepsilon n$ in G, there are at most

$$\varepsilon n^2 p(1 + 2\sqrt{\varepsilon})$$

edges between U and its complement in G.

Proof. Assume that $U \subset V$ contradicts the above statement. Denote

$$e_1 = e_G(U), \quad e_2 = e_G(U, V - U), \quad e_3 = e_G(V - U) .$$

Then $e_1 + e_2 + e_3 = m = \binom{n}{2}p$.

Choose uniformly at random a subset $X \subset V - U$ of cardinality $|X| = x = \sqrt{\varepsilon}|V - U| = \sqrt{\varepsilon}(1-\varepsilon)n$. Then the expected number of edges of G spanned by $U \cup X$ is

$$E[e_G(U \cup X)] = e_1 + \frac{x}{|V - U|}e_2 + \frac{x(x - 1)}{|V - U|(|V - U| - 1)}e_3$$

$$> e_1 + \frac{x}{n - \varepsilon n}e_2 + \frac{x^2}{(n - \varepsilon n)^2}\left(1 - \frac{1}{x}\right)e_3$$

$$= e_1 + \sqrt{\varepsilon}e_2 + \varepsilon e_3 - O(n) = e_1 + \sqrt{\varepsilon}e_2 + \varepsilon(m - e_1 - e_2) - O(n)$$

$$\geq (\sqrt{\varepsilon} - \varepsilon)e_2 + \varepsilon m - O(n) \geq (\sqrt{\varepsilon} - \varepsilon)\varepsilon n^2 p(1 + 2\sqrt{\varepsilon}) + \varepsilon m - O(n)$$

$$= (\sqrt{\varepsilon} - \varepsilon)\varepsilon(1 + 2\sqrt{\varepsilon})n^2 p + \frac{\varepsilon n^2 p}{2} - O(n) =: A .$$

On the other hand, by the assumption on G, every such set $U \cup X$ satisfies

$$e_G(U \cup X) \leq \binom{\varepsilon n + \sqrt{\varepsilon}(1 - \varepsilon)n}{2}p(1 + \varepsilon)$$

$$\leq \frac{\varepsilon n^2 p}{2}(1 + \sqrt{\varepsilon} - \varepsilon)^2(1 + \varepsilon) = \frac{\varepsilon n^2 p}{2}(1 + 2\sqrt{\varepsilon} - \varepsilon - 2\varepsilon^{3/2} + \varepsilon^2)(1 + \varepsilon)$$

$$= \frac{\varepsilon n^2 p}{2}(1 + 2\sqrt{\varepsilon} + O(\varepsilon^{5/2})) =: B .$$

Let us compare the asymptotic (in small ε) behavior of the two quantities A and B defined above. We have

$$A = \frac{\varepsilon n^2 p}{2}(1 + (2\sqrt{\varepsilon} - 2\varepsilon)(1 + 2\sqrt{\varepsilon})) - O(n)$$

$$= \frac{\varepsilon n^2 p}{2}(1 + 2\sqrt{\varepsilon} + 2\varepsilon - 4\varepsilon^{3/2}) - O(n) .$$

Since $n = o(n^2 p)$, we have that $A > B$ for ε small enough, which is a contradiction. $\square$

LEMMA 2.5. *Let $G = (V, E)$ be a graph on $|V| = n$ vertices with $m = |E|$ edges and density $p = m/\binom{n}{2} \geq n^{-a}$ for some constant $0 < a < 1$. Suppose that (1) holds. Then, for all sufficiently large n, G contains an induced subgraph G^* on at least $(1 - \varepsilon - 2\sqrt{\varepsilon})n > n/2$ vertices with maximum degree $\Delta(G^*) \leq (1 + 3\sqrt{\varepsilon})pn$ and minimum degree $\delta(G^*) \geq (1 - 2\sqrt{\varepsilon})np$. In particular, G^* is c-nearly regular for $c = (1 + 6\sqrt{\varepsilon})$.*

Proof. Let U be a set of εn vertices of highest degrees in G (ties are broken arbitrarily). Set $H = G[V - U]$. We claim that all vertex degrees in H are at most

$np(1 + 3\sqrt{\varepsilon})$. If this is not so, then the degrees of all vertices of U in G are at least $np(1 + 3\sqrt{\varepsilon})$, implying (through condition (1))

$$e_G(U, V - U) \geq \varepsilon n^2 p(1 + 3\sqrt{\varepsilon}) - 2e_G(U) \geq \varepsilon n^2 p(1 + 3\sqrt{\varepsilon}) - \varepsilon^2 n^2 p(1 + \varepsilon)$$

$$> \varepsilon n^2 p(1 + 2\sqrt{\varepsilon})$$

for small enough ε, thus contradicting Lemma 2.4. Therefore, H is an induced subgraph of G on $|V(H)| = (1 - \varepsilon)n$ vertices, of maximum degree $\Delta(H) \leq np(1 + 3\sqrt{\varepsilon})$, still satisfying condition (1), and having

$$|E(H)| \geq \binom{n}{2}p - \frac{\varepsilon^2 n^2}{2}p(1 + \varepsilon) - \varepsilon n^2 p(1 + 2\sqrt{\varepsilon})$$

$$(2) \qquad\qquad\qquad \geq \binom{n}{2}p - 2\varepsilon n^2 p$$

edges.

We now repeatedly delete from H vertices of degree less than $np(1 - 2\sqrt{\varepsilon})$ until there are no such vertices, or until we have deleted $2\sqrt{\varepsilon}n$ of them. Assume the latter case and denote the set of $2\sqrt{\varepsilon}n$ deleted vertices by W. Then the set $V(H) - W$ has $|V(H) - W| = (1 - \varepsilon - 2\sqrt{\varepsilon})n$ vertices and by (2) spans at least

$$e_H(V(H) - W) \geq |E(H)| - 2\sqrt{\varepsilon}n \cdot np(1 - 2\sqrt{\varepsilon})$$

$$\geq \binom{n}{2}p - 2\varepsilon n^2 p - 2\sqrt{\varepsilon}n^2 p(1 - 2\sqrt{\varepsilon})$$

$$= \binom{n}{2}p - 2\sqrt{\varepsilon}n^2 p + 2\varepsilon n^2 p = \frac{n^2 p}{2}\left(1 - 4\sqrt{\varepsilon} + 4\varepsilon\right) - O(np)$$

edges. On the other hand, by condition (1), the set $V(H) - W$ satisfies

$$e_H(V(H) - W) \leq \binom{(1 - 2\sqrt{\varepsilon} - \varepsilon)n}{2}p(1 + \varepsilon) \leq (1 - 2\sqrt{\varepsilon} - \varepsilon)^2(1 + \varepsilon)\frac{n^2 p}{2}$$

$$= \frac{n^2 p}{2}\left(1 - 4\sqrt{\varepsilon} + 3\varepsilon + O(\varepsilon^{3/2})\right).$$

Comparing the above two estimates for $e_H(V(H) - W)$, we get a contradiction for small enough ε.

It follows that the above deletion process stops before $2\sqrt{\varepsilon}n$ vertices have been deleted. Denote the resulting graph by G^*. Then G^* has $|V(G^*)| \geq n - \varepsilon n - 2\sqrt{\varepsilon}n > \frac{n}{2}$ vertices and has maximum degree $\Delta(G^*) \leq np(1 + 3\sqrt{\varepsilon})$ and minimum degree $\delta(G^*) \geq np(1 - 2\sqrt{\varepsilon})$. Hence

$$\Delta(G^*) \leq \frac{1 + 3\sqrt{\varepsilon}}{1 - 2\sqrt{\varepsilon}}\delta(G^*) < (1 + 6\sqrt{\varepsilon})\delta(G^*),$$

completing the proof of the lemma.   ☐

We are now ready to prove Theorems 1.2 and 1.3.

Proof of Theorem 1.2. By Lemma 2.3 (with $\frac{\varepsilon^2}{36}$ playing the role of ε), G contains an induced subgraph G' of density $p' \geq p$ on

$$n' \geq \left(\frac{\varepsilon^2}{36}\right)^{\frac{72}{\varepsilon^2} \ln\left(\frac{1}{p}\right)} \cdot n = \left(\frac{\varepsilon}{6}\right)^{\frac{144}{\varepsilon^2} \ln\left(\frac{1}{p}\right)} \cdot n$$

vertices, such that every set of $t \geq \frac{\varepsilon^2}{36}n'$ vertices of G' spans at most $\binom{t}{2}p'(1 + \frac{\varepsilon^2}{36})$ edges. Since $p' \geq p > n^{-1/2}$ (as $p > 0$ is a constant and n is large), Lemma 2.5 implies that G' contains an induced subgraph G^* on at least $n'/2$ vertices which is c-nearly-regular for $c = 1 + 6\sqrt{\frac{\varepsilon^2}{36}} = 1 + \varepsilon$, as needed. □

Proof of Theorem 1.3. Set

$$\varepsilon_0 = \frac{\varepsilon^2}{36}, \quad a = \frac{\varepsilon_0}{3 \ln(1/\varepsilon_0)} .$$

Let $G = (V, E)$ be a graph on n vertices. Denote by $p = |E|/\binom{n}{2}$ the density of G. The average degree of G is at most np and by Turán's theorem G contains an independent set U of size $n/(np + 1)$. Therefore we can assume that $p \geq n^{-a}$, as otherwise,

$$\frac{n}{np + 1} \geq \frac{n}{n^{1-a} + 1} \geq (1 - o(1))n^a = (1 - o(1))n^{\frac{\varepsilon_0}{3 \ln(1/\varepsilon_0)}}$$

$$= (1 - o(1))n^{\frac{\varepsilon^2}{108 \ln(36/\varepsilon^2)}} > n^{\frac{\varepsilon^2}{250 \ln(1/\varepsilon)}} ,$$

where here we used the assumptions that ε is sufficiently small and n is sufficiently large. This gives an induced 0-regular subgraph of G, and we can thus indeed assume that $p \geq n^{-a}$.

By Lemma 2.3, G contains an induced subgraph G' of density $p' \geq p$ on

$$n' \geq \varepsilon_0^{\frac{2}{\varepsilon_0} \ln(n^a)} \cdot n = \varepsilon_0^{\frac{2}{3 \ln(1/\varepsilon_0)} \ln n} \cdot n = n^{\frac{1}{3}}$$

vertices, in which the density of the induced subgraph on any set of at least $\varepsilon_0 n'$ vertices does not exceed $p'(1 + \varepsilon_0)$. By Lemma 2.5, G' (and hence G) contains an induced subgraph on at least $0.5n' \geq 0.5n^{1/3}$ vertices, which is $(1 + 6\sqrt{\varepsilon_0}) = (1 + \varepsilon)$-nearly regular, completing the proof. □

Remark 2. Note that we have actually proved the following result, which is stronger than the assertion of Theorem 1.3: Every graph on n vertices contains either an independent set of size at least $n^{\frac{\varepsilon^2}{250 \ln(1/\varepsilon)}}$, or a $(1 + \varepsilon)$-nearly regular induced subgraph on at least $0.5n^{1/3}$ vertices.

3. Upper bounds.

3.1. The strictly regular case.

Proof of Theorem 1.4. Fix an integer k satisfying

$$k \geq Cn^{1/2} \ln^{3/4} n,$$

where $C > 0$ is a sufficiently large constant to be set later.

We will work with the following model of random graphs on n vertices which we denote by $G(n, \bar{p})$. Let $\bar{p} = (p_1, \ldots, p_n)$, where

$$p_i = \frac{1}{4} + \frac{i}{2n}, \quad i = 1, \ldots, n .$$

Then $G(n, \bar{p})$ is the probability space of graphs with vertex set $[n] = \{1, \ldots, n\}$, where for each pair $1 \leq i \neq j \leq n$, (i, j) is an edge of $G(n, \bar{p})$ with probability $p_i p_j$, independently of all other pairs. Notice that the probability of each individual pair (i, j) to be an edge of $G(n, \bar{p})$ is strictly between $1/16$ and $9/16$.

PROPOSITION 3.1. *Let $X_1, \ldots, X_t$ be independent Bernoulli random variables, where $Pr[X_i = 1] = \rho_i$, $i = 1, \ldots, t$. Let $X = X_1 + \cdots + X_t$. Assume that $1/16 \leq \rho_i \leq 9/16$ for $i = 1, \ldots, t$. Then for every integer $0 \leq s \leq t$, $\Pr[X = s] \leq c_0/\sqrt{t}$ for some absolute constant $c_0 > 0$.*

Proof. For every $1 \leq i \leq t$, we represent X_i as a product $X_i = Y_i \cdot Z_i$, where $\{Y_i, Z_i\}$ is a collection of mutually independent Bernoulli random variables defined by $\Pr[Y_i = 1] = \frac{9}{16}$, $\Pr[Z_i = 1] = \frac{16}{9}\rho_i$, $i = 1, \ldots, t$.

Set $I_0 = \{1 \leq i \leq t : Z_i = 1\}$. Since $\rho_i \geq 1/16$ for all $1 \leq i \leq t$, we have that $\Pr[Z_i = 1] \geq \frac{1}{9}$ and $E[|I_0|] \geq \frac{t}{9}$. By standard large deviation arguments, $|I_0| \geq t/10$ with probability $1 - o(\frac{1}{\sqrt{t}})$. Thus

$$\Pr[X = s] = \sum_{I \subseteq [t]} \Pr[I_0 = I] \cdot \Pr\left[\sum_{i \in I} Y_i = s\right] = \sum_{I \subseteq [t]} \Pr[I_0 = I]\Pr[B(|I|, 9/16) = s]$$

$$\leq \Pr[|I_0| < t/10] + \sum_{|I| \geq t/10} \Pr[I_0 = I]Pr[B(|I|, 9/16) = s] .$$

Here $B(n, p)$ denotes the binomial random variable with parameters n and p. From known estimates on binomial random variables, we obtain that $\Pr[B(r, \frac{9}{16}) = s] \leq \frac{c}{\sqrt{r}} \leq \frac{4c}{\sqrt{t}}$ for every $r \geq \frac{t}{10}$, where $c > 0$ is an absolute constant. Plugging this estimate into the inequality above, we get the claimed result. $\square$

LEMMA 3.2. *Let $U \subseteq [n]$ be a fixed set of $|U| = k$ vertices. Then the probability that in $G(n, \bar{p})$, with $\bar{p}$ as defined above, the graph $G[U]$ is a regular graph is at most $n\left(\frac{c_1}{k}\right)^{k/2}$ for some absolute constant $c_1 > 0$.*

Proof. Fix the degree of regularity d of the regular subgraph $G[U]$ (this can be done in n ways). Let $U = \{u_1, u_2, \ldots, u_k\}$. We bound the probability that the induced subgraph $G[U]$ is d-regular as follows.

Expose the edges of $G[U]$ by first exposing the edges from u_1 to $U - \{u_1\}$, then from u_2 to $U - \{u_1, u_2\}$, etc. If vertex u_i gets t_i neighbors in $\{u_1, \ldots, u_{i-1}\}$, then u_i should have exactly $d - t_i$ neighbors in $\{u_{i+1}, \ldots, u_{k_0}\}$. Recall that all edge probabilities in $G(n, \bar{p})$ are between $1/16$ and $9/16$. Thus Proposition 3.1 applies, and the probability of the latter event is at most $c_0/\sqrt{k - i}$. Multiplying these probabilities for $i = 1, \ldots, k$, we derive that the probability that U is a d-regular graph is at most

$$\prod_{i=1}^{k-1} \frac{c_0}{\sqrt{k - i}} = \frac{(c_0)^{k-1}}{\sqrt{(k-1)!}} \leq \left(\frac{c_1}{k}\right)^{k/2} ,$$

where $c_1 > 0$ is an absolute constant and the last inequality follows by applying the Stirling formula. $\square$

We can now complete the proof of Theorem 1.4. We bound the probability that in $G(n, \bar{p})$ there exists a set U of size $|U| = k$ such that $U = (u_1 < \cdots < u_{k/\ln k} = a < \cdots < u_{k-k/\ln k} = b < \cdots < u_k)$ and $G[U]$ is a regular graph by considering two possible cases depending on the difference $t = b - a$ between a and b.

Case 1. $t \leq c_2 k^{3/2}$, where $c_2 > 0$ is a small positive constant to be determined later.

The probability that there exists such a U is at most

$$(3) \qquad \binom{n}{2k/\ln k} n^2 \binom{t}{k - 2k/\ln k} \cdot n \left(\frac{c_1}{k}\right)^{k/2}.$$

Indeed, there are less than $\binom{n}{2k/\ln k}$ ways to choose the vertices

$$u_1, \ldots, u_{k/\ln k - 1}, u_{k - k/\ln k + 1}, \ldots, u_k,$$

less than n^2 ways to choose the vertices a and b so that the difference between them, t, is at most $c_2 k^{3/2}$, and less than $\binom{t}{k - 2k/\ln k}$ ways to choose the vertices $u_{k/\ln k + 1}, \ldots, u_{k - k/\ln k - 1}$. For each such choice, the probability that the induced subgraph on $\{u_1, u_2, \ldots, u_k\}$ is regular is at most $n\left(\frac{c_1}{k}\right)^{k/2}$, by Lemma 3.2.

A simple computation shows that the expression in (3) is (much) smaller than, say, $1/n^2$ for an appropriate choice of c_2. Indeed, since $n \le k^2$ and $\binom{t}{k - 2k/\ln k} \le \binom{2t}{k} \le (2et/k)^k$, this expression is at most

$$n^3 k^{4k/\ln k} \left(\frac{2et}{k}\right)^k \left(\frac{c_1}{k}\right)^{k/2} \le n^3 2^k e^{5k} c_2^k c_1^{k/2} = n^3 [4e^{10} c_2^2 c_1]^{k/2},$$

implying the required estimate by choosing, for example, $c_2 = \frac{1}{3e^5 \sqrt{c_1}}$.

Case 2. $t \ge c_2 k^{3/2}$.

Let $U = (u_1 < \cdots < u_{k/\ln k} = a < \cdots < u_{k - k/\ln k} = b < \cdots < u_k)$. Denote the first (smallest) $k/2$ vertices of U by U_1, and the last $k/2$ vertices by U_2. Observe that if $G[U]$ is a regular graph, then the two induced subgraphs $G[U_1]$ and $G[U_2]$ have the same average degree. This is highly improbable. Indeed, by the definition of $G(n, \bar{p})$ the probability of each pair inside U_1 to become an edge is strictly less than the probability of each pair inside U_2 to become an edge. In addition, since $b - a = t$, the probability that each pair $i, j \in U_1$ with $i < j, i \le a$ is an edge is less than the probability that each pair $i', j' \in U_2$ with $i' < j', j' \ge b$ is an edge by $\Omega(t/n)$—this is because in this case,

$$\left(\frac{1}{4} + \frac{i'}{2n}\right) \left(\frac{1}{4} + \frac{j'}{2n}\right) - \left(\frac{1}{4} + \frac{i}{2n}\right) \left(\frac{1}{4} + \frac{j}{2n}\right) \ge \Omega \left(\frac{i' + j' - i - j}{2n}\right)$$

$$\ge \Omega \left(\frac{j' - i}{2n}\right) \ge \Omega \left(\frac{t}{n}\right).$$

It follows that the expected number of edges inside U_2 exceeds that inside U_1 by $\Omega(\frac{k^2 t}{n \ln k})$. By Chernoff's inequality (cf., e.g., [1, Appendix A]), we obtain that the probability that $G[U_1]$ and $G[U_2]$ have the same average degree in $G(n, \bar{p})$ is at most $\exp\{-c_3 k^2 t^2 / (n^2 \ln^2 k)\}$ for some absolute constant $c_3 > 0$.

Thus, recalling our assumption on t, we derive that the probability that there is U as above for which $G[U]$ is a regular graph is at most

$$\binom{n}{k} e^{-\frac{c_3 k^2 t^2}{n^2 \ln^2 k}} \le \left(\frac{en}{k}\right)^k e^{-\frac{c_3 c_2^2 k^5}{n^2 \ln^2 k}} = \left(\frac{en}{k} e^{-\frac{c_3 c_2^2 k^4}{n^2 \ln^2 k}}\right)^k.$$

Finally, as $k \ge C n^{1/2} \ln^{3/4} n$ and $\ln k = \Theta(\ln n)$, we can choose $C > 0$ to be large enough so that the above expression is (much) smaller than $1/n^2$.

Combining the two cases, we conclude that the probability that for any fixed k which is at least $Cn^{1/2} \ln^{3/4} n$ our graph contains an induced regular subgraph on k vertices is smaller than $2/n^2 < 1/n$, and as there are less than n choices for k, this shows that with positive probability the graph contains no such subgraph. This completes the proof of Theorem 1.4. □

3.2. The nearly regular case.

Proof of Proposition 1.5. We will prove that for every (large) n there exists a graph G on n vertices in which every K-nearly regular induced subgraph has at most $7Kn/\log n$ vertices.

Assume first that n is of the form $n = (s + 1)2^s$ for a positive integer s. Notice that $s = (1 - o(1)) \log_2 n$. Take a set V of n vertices and partition it into $s + 1$ disjoint equally sized subsets $V_0, \ldots, V_s$, $|V_i| = 2^s$. Now we define G as follows. For $i = 0, \ldots, s$ the set V_i spans 2^{s-i} disjoint cliques of size 2^i each. There are no edges between the cliques inside V_i and no edges between distinct subsets $V_i \neq V_j$ in G.

Assume now that a subset $U \subseteq V$ spans a graph $G[U]$ satisfying $\delta(G[U]) = d$, $\Delta(G[U]) \leq Kd$. Observe that the degrees of all vertices from V_i in G are $2^i - 1$, and thus if $2^i - 1 < d$, then $U \cap V_i = \emptyset$. Let now $2^i \geq d + 1$. Since $\Delta(G[U]) \leq Kd$, U has at most $Kd + 1$ vertices in each clique spanned by V_i, implying $|U \cap V_i| \leq 2^{s-i}(Kd + 1)$. Therefore,

$$|U| = \sum_{i=0}^{s} |U \cap V_i| = \sum_{i:2^i \geq d+1} |U \cap V_i|$$

$$\leq \sum_{i=\lceil \log_2(d+1) \rceil}^{s} 2^{s-i}(Kd + 1) < (Kd + 1)2^{s - \lceil \log_2(d+1) \rceil + 1}$$

$$\leq \frac{2(Kd + 1)}{d + 1} 2^s = \frac{2(Kd + 1)}{d + 1} \frac{n}{s + 1} \leq (2 + o(1))K \frac{n}{\log n},$$

implying the desired result.

For n not of the form $n = (s + 1)2^s$, choose a minimal s satisfying $(s + 1)2^s \geq n$. Let $n' = (s + 1)2^s$. It is easy to verify that $n' \leq 3n$. Now we can apply the above construction to create a graph G' on n' vertices in which every K-nearly induced subgraph has at most $(2 + o(1))Kn'/\log n' \leq 7Kn/\log n$ vertices, and then take G to be an arbitrary induced subgraph of G' on exactly n vertices. □

4. Open problems. The most intriguing open problem is that of obtaining a better estimate for $f(n, 1)$. In particular, the conjecture of Erdős, Fajtlowicz, and Staton that $f(n, 1)/\ln n$ tends to infinity as n tends to infinity remains open. The values of $f(n, 1)$ for $n \leq 17$ have been determined by the authors of [5] and by McKay, and these are indeed larger than the bounds that follow from the corresponding Ramsey numbers.

Our upper and lower bounds for $f(n, c)$ for $c > 1$ are also rather far from each other, and it will be nice to better understand the behavior of this function.

One can also study a variant of the problems considered here that deals with not necessarily induced subgraphs. Of course, every graph contains a regular subgraph on all vertices (the subgraph with no edges), and hence in this case it is natural to look for regular or nearly regular subgraphs with a large number of *edges*. For every two positive integers n, m with $m \leq \binom{n}{2}$ and a real $c \geq 1$, let $g(n, m, c)$ denote the largest g so that every graph with n vertices and m edges contains a (not necessarily

induced) c-nearly regular subgraph with at least g *edges*. The problems of determining or estimating the behavior of this function seems interesting. Here we can establish tighter estimates than the ones obtained for the induced case.

Consider first the case $c = 1$. Since the complete graph on n vertices can be covered by n matchings (and by $(n-1)$ for even n), it follows that every graph with m edges contains a matching of size at least m/n. This implies that $g(n, m, 1) \geq m/n$. The star $K_{1,n-1}$ shows that for some values of m and n this is essentially tight, and that $g(n, m, 1) = 1$ for all $1 \leq m < n$. By a simple application of Szemerédi's regularity lemma it can be shown (see [8]) that for every fixed $p > 0$ there is a $\delta = \delta(p) > 0$ so that $g(n, pn^2, 1) \geq \delta n^2$. This bound was significantly improved by Rödl and Wysocka [9], who proved that every graph with n vertices and pn^2 edges contains an r-regular subgraph with $r \geq \alpha p^3 n$ for some positive constant α.

For a larger constant c observe, first, that complete bipartite graphs show that for $m \geq n$, $g(n, m, c) \leq O(c(m/n)^2) = O(cd^2)$, where $d = 2m/n$ is the average degree of a graph with n vertices and m edges. Indeed, a complete bipartite graph $K_{k,n-k}$ with $k \leq n/2$ has average degree $d = \Theta(k)$. Every c-nearly regular subgraph in it has minimum degree at most k, and hence maximum degree at most ck. Thus it cannot have more than $k \cdot ck = ck^2$ edges. Therefore, for every fixed $c > 1$ there exists some $C = C(c)$ so that $g(m, n, c) \leq C(m/n)^2$ for all $m > n$. We can show that for $c > 2$ this is tight, up to a constant factor; namely, for any $c > 2$ there is a $b = b(c) > 0$ so that $g(n, m, c) \geq b(m/n)^2$ for all $m > n$. For simplicity we present the proof only for $c = 5$; the proof for any other $c > 2$ is similar.

THEOREM 4.1. *Let $G = (V, E)$ be a graph with $|V| = n$ vertices, $|E| = m > n$ edges, and average degree $d = d(G) = 2m/n$. Then G contains a 5-nearly regular subgraph with at least $\frac{d^2}{2^{12}}$ edges.*

Proof. We apply the method of Pyber [7] together with a few extra twists. Clearly we may assume that $d \geq 2^6$. First, omit from G repeatedly vertices of degree smaller than $d/2$, as long as there are such vertices. As this process can only increase the average degree, it ends with a nonempty graph G' with minimum degree at least $d/2$. Now take a spanning bipartite subgraph of G' with the maximum number of edges. It is easy to see and well known that the degree of every vertex in this bipartite subgraph is at least half its degree in G', giving a bipartite graph H with minimum degree at least $d/4$. Put $H_1 = H$. Let A and B denote the two vertex classes of H, where $|A| \geq |B|$. Let $A_1 \subseteq A$ be a nonempty subset of A which satisfies $|N_{H_1}(A_1)| \leq |A_1|$ and A_1 is minimal with respect to containment (subject to the condition above and to being nonempty). Clearly there is such an A_1, as $|N_{H_1}(A)| \leq |A|$ and $|N_{H_1}(v)| \geq d/4 > 1$ for all $v \in A$. Put $N_{H_1}(A_1) = B_1$ and note that by the minimality of A_1, $|A_1| = |B_1|$. By the minimality, again, and by Hall's theorem, there is a matching M_1 saturating A_1 and B_1. Let H_2 be the graph obtained from H_1 by deleting all edges of M_1. Now let $A_2 \subseteq A_1$ be a nonempty, minimal subset of A_1 satisfying $|N_{H_2}(A_2)| \leq |A_2|$. As before, there is such a set, as $|N_{H_2}(A_1)| \leq |N_{H_1}(A_1)| = |A_1|$. The minimality shows, again, that in fact $|N_{H_2}(A_2)| = |A_2|$, and that there is a matching M_2 saturating A_2 and $N_{H_2}(A_2) = B_2$. Proceeding in this manner we define a sequence of sets

$$\emptyset \neq A_{d/4} \subseteq A_{d/4-1} \subseteq \cdots \subseteq A_2 \subseteq A_1 \subseteq A$$

and

$$\emptyset \neq B_{d/4} \subseteq B_{d/4-1} \subseteq \cdots \subseteq B_2 \subseteq B_1 \subseteq B$$

where $|A_i| = |B_i|$ for all i, and a sequence of pairwise edge-disjoint matchings $M_{d/4}$, $\ldots, M_2, M_1$, where M_i is a perfect matching between A_i and B_i. Note that, indeed,

this process does not terminate before these $d/4$ phases, as initially all degrees in H are at least $d/4$, and with the omission of each matching the degrees drop by 1.

For convenience we assume, from now on, that d is a power of 2 (otherwise, simply consider only the first $d'/4$ sets A_i, B_i and matchings M_i, where $d' > d/2$ is the largest power of 2 that does not exceed d). Note that $|A_{d/8}| > d/8$, since every vertex of $B_{d/4}$ is incident with an edge of each of the matchings M_i for $d/8 \le i \le d/4$, and all these edges are incident with vertices of $A_{d/8}$. We consider two possible cases.

Case 1. For every i, $0 \le i \le \log_2 d - 4$,

$$|A_{d/2^{i+4}}| > 2|A_{d/2^{i+3}}|.$$

In this case,

$$|A_1| > 2^{\log_2 d - 3}|A_{d/8}| \ge \frac{d^2}{64},$$

and the matching M_1 is a regular subgraph with more than $d^2/64$ edges, supplying the desired result (with room to spare).

Case 2. There is an i, $0 \le i \le \log_2 d - 4$, such that

$$|A_{d/2^{i+4}}| \le 2|A_{d/2^{i+3}}|.$$

In this case, take the minimum i for which this holds. Then

$$|A_{d/2^{i+3}}| > 2^i |A_{d/8}| \ge 2^{i-3}d.$$

Let H' be the graph consisting of the $\frac{d}{2^{i+4}}$ matchings M_j for $\frac{d}{2^{i+4}} \le j < \frac{d}{2^{i+3}}$. The vertices of H' are all those saturated by the largest matching among those, namely, $M_{\frac{d}{2^{i+4}}}$. Then the maximum degree in H' is exactly $\frac{d}{2^{i+4}}$ (as every vertex of $A_{d/2^{i+3}}$ has that degree), and the average degree is at least half of that, since each of the $\frac{d}{2^{i+4}}$ matchings M_j above is of size at least half that of the largest one, which is spanning. As in H' the degree of every vertex of $A_{d/2^{i+3}}$ is exactly $\frac{d}{2^{i+4}}$, and the total number of edges of H' is at least

$$|A_{d/2^{i+3}}| \cdot \frac{d}{2^{i+4}} \ge 2^{i-3}d \cdot \frac{d}{2^{i+4}} = 2^{-7}d^2.$$

Thus, H' is a graph with maximum degree that exceeds the average degree by a factor of at most $K = 2$. We can now apply Proposition 2.1 with $K = 2$ and $\alpha = 0.4$ to conclude that H' contains a $K/\alpha = 5$-nearly regular subgraph with at least

$$\frac{K - 2K\alpha}{2K - 4\alpha}|E(H')| \ge \frac{1}{6}2^{-7}d^2$$

edges, completing the proof.   □

Acknowledgments. We thank Domingos Dellamonica and Vojta Rödl for pointing an error in an early version of the paper.

REFERENCES

[1] N. ALON AND J. H. SPENCER, *The Probabilistic Method*, 2nd ed., Wiley, New York, 2000.
[2] F. R. K. CHUNG AND R. L. GRAHAM, *Erdős on Graphs: His Legacy of Unsolved Problems*, A. K. Peters, Ltd., Wellesley, MA, 1998.

[3] P. ERDŐS, *On some of my favourite problems in various branches of combinatorics*, in Proceedings of the Fourth Czechoslovakian Symposium on Combinatorics, Graphs and Complexity (Prachatice, 1990), Ann. Discrete Math. 51, North–Holland, Amsterdam, 1992, pp. 69–79.

[4] P. ERDŐS AND M. SIMONOVITS, *Some extremal problems in graph theory*, in Combinatorial Theory and Its Applications I (Proc. Colloq., Balatonfüred, 1969), North–Holland, Amsterdam, 1970, pp. 377–390.

[5] S. FAJTLOWICZ, T. MCCOLGAN, T. REID, AND W. STATON, *Ramsey numbers for induced regular subgraphs*, Ars Combin., 39 (1995), pp. 149–154.

[6] R. L. GRAHAM, B. L. ROTHSCHILD, AND J. H. SPENCER, *Ramsey Theory*, 2nd ed., Wiley, New York, 1990.

[7] L. PYBER, *Regular subgraphs of dense graphs*, Combinatorica, 5 (1985), pp. 347–349.

[8] L. PYBER, V. RÖDL, AND E. SZEMERÉDI, *Dense graphs without 3-regular subgraphs*, J. Combin. Theory Ser. B, 63 (1995), pp. 41–54.

[9] V. RÖDL AND B. WYSOCKA, *Note on regular subgraphs*, J. Graph Theory, 24 (1997), pp. 139–154.