Combinatorics, Probability and Computing (2003) 12, 139–153.
DOI: 10.1017/S0963548302005539 Printed in the United Kingdom

Local Density in Graphs with Forbidden Subgraphs

PETER KEEVASH[1] and BENNY SUDAKOV[2]†

[1] Department of Mathematics, Cambridge University, Cambridge, England
and
Department of Mathematics, Princeton University, Princeton, NJ 08540, USA
(e-mail: keevash@math.princeton.edu)

[2] Department of Mathematics, Princeton University, Princeton, NJ 08540, USA
and
Institute for Advanced Study, Princeton, NJ 08540, USA
(e-mail: bsudakov@math.princeton.edu)

Received 1 November 2001; revised 28 September 2002

A celebrated theorem of Turán asserts that every graph on n vertices with more than $\frac{r-1}{2r}n^2$ edges contains a copy of a complete graph K_{r+1}. In this paper we consider the following more general question. Let G be a K_{r+1}-free graph of order n and let α be a constant, $0 < \alpha \leqslant 1$. How dense can every induced subgraph of G on αn vertices be? We prove the following local density extension of Turán's theorem.

For every integer $r \geqslant 2$ there exists a constant $c_r < 1$ such that, if $c_r \leqslant \alpha \leqslant 1$ and every αn vertices of G span more than

$$\frac{r-1}{2r}(2\alpha - 1)n^2$$

edges, then G contains a copy of K_{r+1}. This result is clearly best possible and answers a question of Erdős, Faudree, Rousseau and Schelp [5].

In addition, we prove that the only K_{r+1}-free graph of order n, in which every αn vertices span at least $\frac{r-1}{2r}(2\alpha - 1)n^2$ edges, is a Turán graph. We also obtain the local density version of the Erdős–Stone theorem.

1. Introduction

Extremal problems are at the heart of modern graph theory. These problems have attracted a lot of attention during the last half century (*e.g.*, see [1] for a survey.) One of the central problems in extremal graph theory can be described as follows. Given a *forbidden graph* H, determine $ex(n, H)$, the maximal number of edges in a graph on n vertices that does

† Supported in part by NSF grants DMS-0106589, CCR-9987845 and by the State of New Jersey.

not contain H as a subgraph. In the case when H is K_{r+1}, a complete graph of order $r+1$, the value $ex(n, K_{r+1})$ was determined in 1941 by Turán. His celebrated theorem asserts that every graph on n vertices with more than $\frac{r-1}{2r}n^2$ edges contains a copy of a K_{r+1}. Let $T_r(n)$ denote a complete r-partite graph on n vertices with class sizes as equal as possible (usually called a *Turán graph*). More precisely, Turán proved that $T_r(n)$ is the only *extremal* K_{r+1}-free graph of order n, *i.e.*, it is the only graph of order n and of size $ex(n, K_{r+1})$ that contains no copy of K_{r+1}.

In this paper we consider the following more general question. Let G be a K_{r+1}-free graph of order n and let α be a constant, $0 < \alpha \leqslant 1$. Suppose that every αn vertices of G span at least βn^2 edges. How large can the function $\beta(\alpha)$ be? This problem was raised by Erdős, Faudree, Rousseau and Schelp in [5]. They conjectured that, in the case when $r = 2$, β is determined by a family of extremal triangle-free graphs. In particular, when $\alpha \geqslant 17/30$ they suggested that the complete bipartite graph with equal sides has the greatest *local density*, which is $\beta = \frac{2\alpha-1}{4}$. They proved that indeed this value of β is best possible for a certain range of α. Later, their result was extended by Krivelevich [7], who proved that this conjecture holds for $\alpha \geqslant \frac{3}{5}$. For the values of $\alpha \leqslant 17/30$ it appears that there are graphs with higher local density than the Turán graph. In particular, Erdős and his co-workers observed that the blow-up of a 5-cycle has this property. This led them to a more general conjecture about the dependence of β on α (see, *e.g.*, [5], [7], [2] for more details and discussions).

Also in [5] Erdős, Faudree, Rousseau and Schelp posed the problem of determining the best local density in K_{r+1}-free graphs for $r > 2$. They conjectured that for α sufficiently close to 1 the r-partite Turán graph has the highest local density. It is easy to check that for $\alpha \geqslant \frac{r-1}{r}$ every subset of $T_r(n)$ of size αn contains at least $\frac{r-1}{2r}(2\alpha-1)n^2$ edges. In this paper we prove the conjecture of [5]. We also obtain a slightly stronger statement, which extends the original result in [5] even in the triangle-free case, namely that the Turán graph is the only extremal graph for this problem. Our main result is as follows.

Theorem 1.1. *Let $r \geqslant 1$ and let G be a K_{r+1}-free graph on n vertices. If $1 - \frac{1}{2r^2} \leqslant \alpha \leqslant 1$, then G contains a set of αn vertices spanning at most $\frac{r-1}{2r}(2\alpha-1)n^2$ edges. Moreover, if G is a K_{r+1}-free graph of order n in which every αn vertices span at least $\frac{r-1}{2r}(2\alpha-1)n^2$ edges, then G is the Turán graph.*

This result can be also viewed as a modest first step towards a solution of a problem of Chung and Graham [3]. They conjectured that for $r \geqslant 3$ the Turán graph has the best local density even for α as low as $1/2$.

Another fundamental result in extremal graph theory is the Erdős–Stone theorem [4]. They proved that, if the density of a graph G is slightly larger than that of the r-partite Turán graph, then G not only contains a clique of size $r+1$ but also any fixed graph of chromatic number $r+1$. More precisely, let $K_{r+1}(t)$ denote a complete $(r+1)$-partite graph in which each class has size t. Then Erdős and Stone proved the following result.

Theorem 1.2. *For all fixed positive integers r and t, we have*

$$ex(n, K_{r+1}(t)) = \left(\frac{r-1}{2r} + o(1)\right)n^2.$$

The following corollary follows immediately from this theorem.

Corollary 1.3. *Let H be a fixed graph with chromatic number $\chi(H) = p \geqslant 2$. Then*

$$ex(n, H) = \left(\frac{p-2}{2(p-1)} + o(1)\right)n^2.$$

Using our main theorem together with Szemerédi's Regularity Lemma, we can deduce the following result about edge distribution in graphs with any fixed forbidden subgraph. This theorem is the natural local density generalization of the Erdős–Stone theorem.

Theorem 1.4. *Let $r \geqslant 2$, $t \geqslant 1$ and let G be a graph on n vertices not containing $K_{r+1}(t)$ as a subgraph. If $1 - \frac{1}{2r^2} \leqslant \alpha \leqslant 1$, then G has a subset of αn vertices containing at most $(\frac{r-1}{2r}(2\alpha - 1) + o(1))n^2$ edges.*

Corollary 1.5. *Let H be a fixed graph with chromatic number $\chi(H) = p \geqslant 2$ and let $1 - \frac{1}{2(p-1)^2} \leqslant \alpha \leqslant 1$. If G is a graph on n vertices not containing H as a subgraph, then G has a subset of αn vertices containing at most $(\frac{p-2}{2(p-1)}(2\alpha - 1) + o(1))n^2$ edges.*

The rest of this paper is organized as follows. In the next section we prove Theorem 1.1 in the simplest case $r = 3$, so as to illustrate the main idea of the proof, and also to get a better lower bound on α than the one guaranteed by this theorem. In Section 3 we prove our main result and also show that the Turán graph is the only extremal graph for the local density problem. Next, in Section 4 we use Szemerédi's Regularity Lemma to prove the local density version of the Erdős–Stone theorem. The final section contains some concluding remarks and open problems.

Throughout the paper we omit all floor and ceiling signs whenever these are not crucial, to simplify the presentation. We also want to present the following observation that justifies this, even without assuming that n is sufficiently large. Let H be a K_{r+1}-free graph of order k such that every $\lfloor \alpha k \rfloor$ vertices of H span at least $\lfloor \beta k^2 \rfloor + 1$ edges. Let G be a graph obtained from H by substituting for every vertex i an independent set V_i of size n/k, and for every edge $(i, j) \in E(H)$ connecting the sets V_i and V_j by a complete bipartite graph. By definition G is also a K_{r+1}-free graph and has order n. Consider a set S of $\lfloor \alpha n \rfloor$ vertices of G which spans the minimal number of edges. It is easy to see that S either contains or is disjoint from sets V_i for all but at most one index i. Otherwise there are two classes V_i and V_j such that $0 < |S \cap V_i| \leqslant |S \cap V_j| < n/k$. Then, by deleting any vertex of S from V_i and adding a new vertex to S from V_j, we clearly decrease the number of edges spanned by S. Therefore S contains $\lfloor \alpha k \rfloor$ sets V_i, so it spans at least $(\lfloor \beta k^2 \rfloor + 1)(n/k)^2 = \beta_1 n^2$ edges, for some $\beta_1 > \beta$. This shows that, in order to prove that every K_{r+1}-free graph of order n contains $\lfloor \alpha n \rfloor$ vertices which span at most $\lfloor \beta n^2 \rfloor$ edges, it is sufficient to show that it contains such a set spanning at most $\lfloor (\beta + o(1))n^2 \rfloor$ edges.

Since all rounding errors can change the number of edges only by at most $o(n^2)$, we can indeed ignore them. Also, by the above argument, we may and will assume that whenever we write $e(X) > \beta n^2$ we actually have $e(X) > \beta_1 n^2$ for some $\beta_1 > \beta$.

We close this section with some conventions and notation. Let $G = (V, E)$ be a graph and let X and Y be two disjoint subsets of G. Then we let $e(X, Y)$ denote the number of edges of G adjacent to exactly one vertex from X and one from Y. Similarly, $E(X)$ denotes the set of edges spanned by a subset X of G and $e(X)$ stands for $|E(X)|$. Also $e(G) = |E(G)|$. The neighbourhood $N(v)$ of a vertex v is the set of vertices of G adjacent to it. For a subset of vertices Y and a vertex v, we let $d_Y(v)$ denote the number of neighbours of v in the set Y. Finally, a function which tends to zero arbitrarily slowly with n is denoted by $o(1)$.

2. Main idea

In this section we illustrate the main idea that we are going to use for the general case by presenting the proof for $r = 3$. This proof also gives a better lower bound on α than the one guaranteed by Theorem 1.1. We obtain the following result.

Theorem 2.1. *Let G be a K_4-free graph on n vertices and let $0.8661 \leqslant \alpha \leqslant 1$. Then G contains a subset of αn vertices spanning at most $\frac{2\alpha-1}{3}n^2$ edges.*

To prove this theorem we need a few technical lemmas. First we obtain an upper bound on the number of edges spanned by a subset of G of size αn which has the following special structure.

Lemma 2.2. *Let H be a K_4-free graph of order αn. Suppose also that the vertex set of H is a union of three disjoint sets X, Y and Z such that X is an independent set and Y can be covered by a set of disjoint triangles. Denote $|X| = xn, |Y| = yn$ and $|Z| = zn$. Then*

$$\frac{e(H)}{n^2} \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}(2x - z)^2 + \frac{e(Z)}{n^2} - \frac{1}{4}z^2.$$

Proof. Let v be a vertex which does not belong to Y. Since H is K_4-free, then it is easy to see that v is adjacent to at most two vertices of any triangle in Y. Therefore $d_Y(v) \leqslant \frac{2}{3}|Y|$. In addition we can apply Turán's theorem to the K_4-free subgraph induced by Y and obtain $e(Y) \leqslant \frac{1}{3}|Y|^2$. Since, by definition, $x + y + z = \alpha$, we finally conclude that

$$\begin{aligned}
\frac{e(H)}{n^2} &= \frac{1}{n^2}(e(X) + e(Y) + e(Z) + e(X, Y) + e(Y, Z) + e(X, Z)) \\
&\leqslant \frac{1}{3}y^2 + \frac{e(Z)}{n^2} + \frac{2}{3}xy + \frac{2}{3}zy + xz \\
&= \frac{1}{3}(x + y + z)^2 - \frac{1}{3}x^2 - \frac{1}{3}z^2 + \frac{e(Z)}{n^2} + \frac{1}{3}xz \\
&= \frac{1}{3}\alpha^2 - \frac{1}{12}(2x - z)^2 + \frac{e(Z)}{n^2} - \frac{1}{4}z^2. \qquad \square
\end{aligned}$$

Next we need the following easy lower bound on the size of a maximum independent set in a K_4-free graph.

Lemma 2.3. *Let G be a K_4-free graph with n vertices and m edges. Then it contains an independent set of size at least $\frac{4m}{n} - n$.*

Proof. Consider the sum

$$\sum_{(x,y)\in E(G)} d(x) + d(y) = \sum_{x\in V(G)} d(x)^2 \geqslant n\left(\frac{\sum_{x\in V(G)} d(x)}{n}\right)^2 = \frac{4m^2}{n}.$$

Hence, there exists an edge (x, y) with $d(x) + d(y) \geqslant \frac{4m}{n}$. Since G is K_4-free, then clearly $N(x) \cap N(y)$ is an independent set of size at least $\frac{4m}{n} - n$. □

Corollary 2.4. *Let G be a K_4-free graph of order n such that every αn vertices of G span at least $\frac{2\alpha-1}{3}n^2$ edges, and let $0.8661 < \alpha \leqslant 1$. Then G contains an independent set of size at least $\frac{9}{4}(1-\alpha)n$.*

Proof. Let m denote the number of edges of G. Let W be a random subset of vertices of G of size αn. Then, for every edge $e \in E(G)$, the probability that $e \in E(W)$ is at most

$$\Pr(e \in E(W)) = \frac{\binom{n-2}{\alpha n-2}}{\binom{n}{\alpha n}} = \frac{\alpha n(\alpha n - 1)}{n(n-1)} \leqslant \alpha^2.$$

Therefore the expected value of $e(W)$ is at most $\alpha^2 m$. Hence we conclude that there exists a subset W of size αn which spans at most $\alpha^2 m$ edges. This implies that

$$m \geqslant \frac{e(W)}{\alpha^2} \geqslant \frac{2\alpha - 1}{3\alpha^2}n^2.$$

Next, by Lemma 2.3 we have that G contains an independent set of size at least

$$\frac{4m}{n} - n \geqslant \left(4\frac{2\alpha-1}{3\alpha^2} - 1\right)n.$$

Now some simple but tedious computations, which we omit here, show that $4(\frac{2\alpha-1}{3\alpha^2}) - 1 \geqslant \frac{9}{4}(1-\alpha)$ for all $\alpha > 0.8661$. This completes the proof. □

Finally we need the following result proved by Krivelevich [7].

Proposition 2.5. *Let G be a triangle-free graph on n vertices and let $3/5 \leqslant \alpha \leqslant 1$. Then G contains a subset of αn vertices spanning at most $\frac{2\alpha-1}{4}n^2$ edges.*

Having finished all the necessary preparations, we are now ready to complete the proof of our first result.

Proof of Theorem 2.1. We assume that there exists a K_4-free graph G on n vertices such that every αn vertices of G span strictly more than $\frac{2\alpha-1}{3}n^2$ edges, and obtain a contradiction.

By Corollary 2.4, G contains an independent set U of size $\frac{9}{4}(1-\alpha)$. Let T be the largest subset of $G - U$ which can be covered by vertex-disjoint triangles. Denote $|T|/n$ by t. Let

S be the complement of $U \cup T$ and let $s = |S|/n$. Note that by definition the subgraph induced by S is triangle-free, and that $\frac{9}{4}(1-\alpha) + t + s = 1$.

Let X_1 be any subset of U of size $\frac{5}{4}(1-\alpha)n$, and let $Y_1 = T$, $Z_1 = S$. Consider the subgraph H_1 induced by the disjoint union of X_1, Y_1 and Z_1. Then, by applying Lemma 2.2 to H_1, we obtain

$$\frac{1}{n^2}e(H_1) \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{5}{2}(1-\alpha) - s\right)^2 + \frac{e(S)}{n^2} - \frac{1}{4}s^2 \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{5}{2}(1-\alpha) - s\right)^2,$$

where in the last inequality we applied Turán's theorem to the triangle-free graph $G[S]$ to deduce $e(S) \leqslant \frac{1}{4}s^2$. On the other hand, since the vertex set of H_1 has size αn, we know that $\frac{1}{n^2}e(H_1) > \frac{2\alpha-1}{3} = \frac{1}{3}(\alpha^2 - (1-\alpha)^2)$ so we obtain $\frac{1}{12}(\frac{5}{2}(1-\alpha) - s)^2 < \frac{1}{3}(1-\alpha)^2$, that is,

$$\left|\frac{5}{2}(1-\alpha) - s\right| < 2(1-\alpha). \tag{2.1}$$

In particular, this implies $s < \frac{9}{2}(1-\alpha)$.

Define the value of q to be $q = \frac{(1-\alpha)-t}{3}$ if $t < (1-\alpha)$ and zero otherwise. Certainly $s \geqslant 3q$, since otherwise U would be independent of size at least αn, which contradicts our assumption about G. Let X_2 be a subset of U of size $\left(\frac{9}{4}(1-\alpha) - q\right)n$, let Z_2 be a subset of S of size $(s-2q)n$ and let Y_2 be a subset obtained by deleting $(\frac{1-\alpha}{3} - q)n$ vertex-disjoint triangles from T. Note that by definition $|X_2| + |Y_2| + |Z_2| = \alpha n$. As before, by applying Lemma 2.2 to the graph H_2 with the vertex set $X_2 \cup Y_2 \cup Z_2$, we conclude that

$$\frac{1}{n^2}e(H_2) \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{9}{2}(1-\alpha) - 2q - s + 2q\right)^2$$
$$+ \frac{e(Z_2)}{n^2} - \frac{1}{4}|Z_2|^2 \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{9}{2}(1-\alpha) - s\right)^2.$$

Again as before, we deduce $\frac{1}{12}(\frac{9}{2}(1-\alpha) - s)^2 < \frac{1}{3}(1-\alpha)^2$. This implies $|\frac{9}{2}(1-\alpha) - s| < 2(1-\alpha)$, so that $s > \frac{5}{2}(1-\alpha)$.

Next consider the triangle-free subgraph induced by the set S. Since

$$\alpha_1 = \frac{s-(1-\alpha)}{s} = 1 - \frac{(1-\alpha)}{s} > 1 - \frac{2}{5} = \frac{3}{5},$$

by Proposition 2.5 $G[S]$ contains a subset Z_3 of size $(s-(1-\alpha))n$ such that

$$e(Z_3) \leqslant \frac{2\alpha_1 - 1}{4}s^2 = \frac{s(s-2(1-\alpha))}{4}n^2 = \frac{1}{4}((s-(1-\alpha))^2 - (1-\alpha)^2)n^2.$$

Finally, let $X_3 = U$ and $Y_3 = T$. Then, by applying Lemma 2.2 to the graph H_3 with the vertex set $X_3 \cup Y_3 \cup Z_3$, we deduce

$$\frac{1}{n^2}e(H_3) \leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{11}{2}(1-\alpha) - s\right)^2 + \frac{e(Z_3)}{n^2} - \frac{1}{4}(s-(1-\alpha))^2$$
$$\leqslant \frac{1}{3}\alpha^2 - \frac{1}{12}\left(\frac{11}{2}(1-\alpha) - s\right)^2 - \frac{1}{4}(1-\alpha)^2.$$

Again, since H_3 has αn vertices it satisfies $\frac{1}{n^2}e(H_3) > \frac{1}{3}(\alpha^2 - (1-\alpha)^2)$. Hence

$$\frac{1}{12}\left(\frac{11}{2}(1-\alpha) - s\right)^2 < \frac{1}{12}(1-\alpha)^2.$$

Therefore

$$\left|\frac{11}{2}(1-\alpha) - s\right| < (1-\alpha),$$

which implies that $s > \frac{9}{2}(1-\alpha)$. This contradicts inequality (2.1) and completes the proof of the theorem. □

3. Edge distribution in K_{r+1}-free graphs

In this section we use the ideas of the proof for the case $r = 3$ to obtain our main result. For the sake of clarity of presentation we will make no attempt to optimize our estimates. At the end of this section we describe a recurrence relation that computes the best possible bounds on α which can be obtained using our proof. We also compute these bounds for a few small values of r.

To prove Theorem 1.1 we need a few technical lemmas. First we obtain an upper bound on the number of edges spanned by a subset of size αn of a K_{r+1}-free graph, which has the following special structure.

Lemma 3.1. *Let $r \geqslant 2$ and let H be a K_{r+1}-free graph with the vertex set of H a union of three disjoint sets X, Y and Z such that X is an independent set and Y can be covered by a set of disjoint copies of K_r. Denote $|X| = xn, |Y| = yn$ and $|Z| = zn$. Then*

$$\frac{1}{n^2}e(H) \leqslant \frac{r-1}{2r}(x+y+z)^2 - \frac{1}{2r(r-1)}((r-1)x - z)^2 + \frac{1}{n^2}e(Z) - \frac{r-2}{2(r-1)}z^2.$$

Proof. Let v be a vertex which does not belong to Y. Since H is K_{r+1}-free, then it is easy to see that v is adjacent to at most $r-1$ vertices of any copy of K_r in Y. Therefore $d_Y(v) \leqslant \frac{r-1}{r}|Y|$. In addition we can apply Turán's theorem to the K_{r+1}-free subgraph induced by Y and obtain $e(Y) \leqslant \frac{r-1}{2r}|Y|^2$. So we conclude that

$$\begin{aligned}
\frac{1}{n^2}e(H) &= \frac{1}{n^2}(e(X) + e(Y) + e(Z) + e(X,Y) + e(Y,Z) + e(X,Z)) \\
&\leqslant \frac{r-1}{2r}y^2 + \frac{1}{n^2}e(Z) + \frac{r-1}{r}xy + \frac{r-1}{r}zy + xz \\
&= \frac{r-1}{2r}(x+y+z)^2 - \frac{r-1}{2r}x^2 - \frac{r-1}{2r}z^2 + \frac{1}{n^2}e(Z) + \frac{1}{r}xz \\
&= \frac{r-1}{2r}(x+y+z)^2 - \frac{1}{2r(r-1)}((r-1)x - z)^2 + \frac{1}{n^2}e(Z) - \frac{r-2}{2(r-1)}z^2. \quad \square
\end{aligned}$$

For a graph G and integer t, let N_t denote the number of copies of a complete graph K_t contained in G. We need the following useful result, which was proved by Moon and Moser [8].

Proposition 3.2. *For any graph G of order n and integer $t \geqslant 2$, the numbers N_t satisfy*

$$\frac{N_{t+1}}{N_t} \geqslant \frac{1}{t^2-1}\left(t^2\frac{N_t}{N_{t-1}} - n\right).$$

Using this recursion we obtain a lower bound on the size of a maximum independent set in a K_{r+1}-free graph.

Lemma 3.3. *Let $r \geqslant 2$ be an integer and let G be a K_{r+1}-free graph with n vertices and m edges. Then it contains an independent set of size at least $2(r-1)\frac{m}{n} - (r-2)n$.*

Proof. Let $q \leqslant r$ be minimal such that G is K_{q+1}-free. Let N_t be the the number of copies of the complete graph K_t contained in G. This is nonzero for $t \leqslant q$, in which range we prove the following inequality by induction on t:

$$(t+1)\frac{N_{t+1}}{N_t} \geqslant 2t\frac{m}{n} - (t-1)n. \tag{3.1}$$

For $t = 1$ this inequality follows immediately from the definition, since $N_1 = n$ and $N_2 = m$. Now suppose that

$$t\frac{N_t}{N_{t-1}} \geqslant 2(t-1)\frac{m}{n} - (t-2)n.$$

Then by Proposition 3.2 we obtain

$$\begin{aligned}(t+1)\frac{N_{t+1}}{N_t} \geqslant \frac{t+1}{t^2-1}\left(t^2\frac{N_t}{N_{t-1}} - n\right) &\geqslant \frac{1}{t-1}\left(t\left(2(t-1)\frac{m}{n} - (t-2)n\right) - n\right) \\ &= 2t\frac{m}{n} - (t-1)n.\end{aligned}$$

Note that every complete subgraph of G of order q contains exactly q distinct copies of K_{q-1}. Therefore, by (3.1) there exists a particular clique in G of order $q-1$ which is contained in at least

$$q\frac{N_q}{N_{q-1}} \geqslant 2(q-1)\frac{m}{n} - (q-2)n$$

distinct copies of K_q. Since G is a K_{q+1}-free graph, then the set of common neighbours of the vertices in this clique forms an independent set of size at least $2(q-1)\frac{m}{n} - (q-2)n$. Of course $m < n^2/2$, so $2(q-1)\frac{m}{n} - (q-2)n \geqslant 2(r-1)\frac{m}{n} - (r-2)n$. This completes the proof of the lemma. □

Corollary 3.4. *Let $r \geqslant 2$ be an integer and let G be a K_{r+1}-free graph of order n such that every αn vertices of G span at least $\frac{r-1}{2r}(2\alpha - 1)n^2$ edges, and let $1 - \frac{1}{2r^2} \leqslant \alpha \leqslant 1$. Then G contains an independent set of size at least $(2r-1)(1-\alpha)n$.*

Proof. Let m denote the number of edges of G. Arguing as in Corollary 2.4 we can show that m is at least $\frac{r-1}{2r}\left(\frac{2\alpha-1}{\alpha^2}\right)n^2$. Therefore, by Lemma 3.3, G contains an independent set of size

$$2(r-1)\frac{m}{n}-(r-2)n \geqslant \left(\frac{(r-1)^2}{r}\left(\frac{2\alpha-1}{\alpha^2}\right)-(r-2)\right)n.$$

So our conclusion is valid whenever

$$\frac{(r-1)^2}{r}\left(\frac{2\alpha-1}{\alpha^2}\right)-(r-2) \geqslant (2r-1)(1-\alpha).$$

Setting $\theta = \frac{1}{\alpha}-1$, we can rewrite the latter inequality as

$$1-(r-1)^2\theta^2 \geqslant \frac{r(2r-1)\theta}{1+\theta} = r(2r-1)\left(1-\frac{1}{1+\theta}\right). \tag{3.2}$$

By our assumption on α we have that $\theta \leqslant \frac{1}{2r^2-1}$. Hence

$$1-(r-1)^2\theta^2 \geqslant 1-\left(\frac{r-1}{2r^2-1}\right)^2 \quad \text{and} \quad \frac{2r-1}{2r} \geqslant r(2r-1)\left(1-\frac{1}{1+\theta}\right).$$

So it suffices to show that

$$1-\left(\frac{r-1}{2r^2-1}\right)^2 \geqslant \frac{2r-1}{2r} = 1-\frac{1}{2r},$$

which can be easily verified. □

Having finished all the necessary preparations, we are now ready to complete the proof of our main result.

Proof of Theorem 1.1. To prove the first part of the theorem we use induction on r. For $r = 1$ the statement of the theorem is trivially true. We assume that there exists a K_{r+1}-free graph G on n vertices such that every αn vertices of G span strictly more than $\frac{r-1}{2r}(2\alpha-1)n^2$ edges, and obtain a contradiction.

By Corollary 3.4, G contains an independent set U of size $(2r-1)(1-\alpha)n$. Let T be the largest subset of $G-U$ which can be covered by vertex-disjoint complete graphs of size r. Denote $|T|/n$ by t. Let S be the complement of $U \cup T$ and let $s = |S|/n$. Note that by definition the subgraph induced by S is K_r-free, and that $(2r-1)(1-\alpha)+t+s = 1$.

Let X_1 be any subset of U of size $(2r-2)(1-\alpha)n$, and let $Y_1 = T, Z_1 = S$. Consider the subgraph H_1 induced by $X_1 \cup Y_1 \cup Z_1$. Then, by applying Lemma 3.1 to H_1, we obtain

$$\begin{aligned}\frac{1}{n^2}e(H_1) &\leqslant \frac{r-1}{2r}\alpha^2-\frac{1}{2r(r-1)}((r-1)(2r-2)(1-\alpha)-s)^2+\frac{1}{n^2}e(S)-\frac{r-2}{2(r-1)}s^2\\ &\leqslant \frac{r-1}{2r}\alpha^2-\frac{1}{2r(r-1)}((r-1)(2r-2)(1-\alpha)-s)^2, \end{aligned} \tag{3.3}$$

where in the last inequality we applied Turán's theorem to the K_r-free graph $G[S]$ to deduce that $e(S) \leqslant \frac{r-2}{2(r-1)}s^2$. On the other hand, since the vertex set of H_1 has size αn, we know that

$\frac{1}{n^2}e(H_1) > \frac{r-1}{2r}(2\alpha - 1) = \frac{r-1}{2r}(\alpha^2 - (1-\alpha)^2)$ so we obtain $\frac{1}{2r(r-1)}(2(r-1)^2(1-\alpha) - s)^2 < \frac{r-1}{2r}(1-\alpha)^2$, that is,

$$|2(r-1)^2(1-\alpha) - s| < (r-1)(1-\alpha). \tag{3.4}$$

In particular, this implies $s < (r-1)(2r-1)(1-\alpha)$.

Define the value of q to be $q = \frac{(1-\alpha)-t}{r}$ if $t < (1-\alpha)$ and zero otherwise. Certainly $s \geqslant rq$, since otherwise U would be independent of size at least αn, which contradicts our assumption about G. Let X_2 be a subset of U of size $((2r-1)(1-\alpha) - q)n$, let Z_2 be a subset of S of size $(s - (r-1)q)n$ and let Y_2 be a subset obtained by deleting $(\frac{1-\alpha}{r} - q)n$ vertex-disjoint copies of K_r from T. Note that by definition $|X_2| + |Y_2| + |Z_2| = \alpha n$. As before, by applying Lemma 3.1 to the graph H_2 with the vertex set $X_2 \cup Y_2 \cup Z_2$, we conclude that

$$\begin{aligned}\frac{1}{n^2}e(H_1) &\leqslant \frac{r-1}{2r}\alpha^2 - \frac{((r-1)((2r-1)-q)(1-\alpha) - s + (r-1)q)^2}{2r(r-1)} \\ &\quad + \frac{e(Z_2)}{n^2} - \frac{r-2}{2(r-1)}|Z_2|^2 \\ &\leqslant \frac{r-1}{2r}\alpha^2 - \frac{1}{2r(r-1)}((r-1)(2r-1)(1-\alpha) - s)^2.\end{aligned}$$

Again as before, we deduce that $\frac{1}{2r(r-1)}((r-1)(2r-1)(1-\alpha) - s)^2 < \frac{r-1}{2r}(1-\alpha)^2$. This implies $|(r-1)(2r-1)t - s| < (r-1)(1-\alpha)$, so that $s > 2(r-1)^2(1-\alpha)$.

Next consider the K_r-free subgraph induced by the set S. Since

$$\alpha_1 = \frac{s - (1-\alpha)}{s} = 1 - \frac{(1-\alpha)}{s} > 1 - \frac{1}{2(r-1)^2}, \tag{3.5}$$

by the induction hypothesis $G[S]$ contains a subset Z_3 of size $(s - (1-\alpha))n$ such that

$$e(Z_3) \leqslant \frac{r-2}{2(r-1)}(2\alpha_1 - 1)s^2 = \frac{r-2}{2(r-1)}((s - (1-\alpha))^2 - (1-\alpha)^2).$$

Finally, let $X_3 = U$ and $Y_3 = T$. Then, by applying Lemma 3.1 to the graph H_3 induced by $X_3 \cup Y_3 \cup Z_3$, we deduce

$$\begin{aligned}\frac{1}{n^2}e(H_3) &\leqslant \frac{r-1}{2r}\alpha^2 - \frac{(((r-1)(2r-1)+1)(1-\alpha) - s)^2}{2r(r-1)} \\ &\quad + \frac{e(Z_3)}{n^2} - \frac{r-2}{2(r-1)}(s - (1-\alpha))^2 \\ &\leqslant \frac{r-1}{2r}\alpha^2 - \frac{(((r-1)(2r-1)+1)(1-\alpha) - s)^2}{2r(r-1)} - \frac{r-2}{2(r-1)}(1-\alpha)^2.\end{aligned}$$

Again, since H_3 has αn vertices it satisfies $\frac{1}{n^2}e(H_3) > \frac{r-1}{2r}(\alpha^2 - (1-\alpha)^2)$. Hence

$$\frac{1}{2r(r-1)}(((r-1)(2r-1)+1)(1-\alpha) - s)^2 < \frac{1}{2r(r-1)}(1-\alpha)^2.$$

Therefore

$$|((r-1)(2r-1)+1)(1-\alpha) - s| < (1-\alpha), \tag{3.6}$$

which implies that $s > (r-1)(2r-1)(1-\alpha)$. This contradicts inequality (3.4) and completes the proof of the first part of the theorem.

Next we show that the Turán graph is the unique extremal graph for the local density problem. First of all we consider the conditions under which equality holds in Lemma 3.1. From the equality case of Turán's theorem we know that Y must be a Turán graph and contain $\frac{r-1}{2r}y^2n^2$ edges. We must also have as many cross-edges as possible, that is,

$$e(X,Y) = \frac{r-1}{r}xyn^2, \quad e(Y,Z) = \frac{r-1}{r}yzn^2, \quad e(X,Z) = xzn^2.$$

Now suppose that G is a K_{r+1}-free graph on n vertices such that every αn set spans at least $\frac{r-1}{2r}(2\alpha-1)n^2$ edges. Partition the vertices of G into a disjoint union of sets U, T and S as in the proof of the first part of the theorem. Then it is easy to see that we can avoid a contradiction only if the inequalities (3.4) and (3.6) hold as equalities. This implies that $|S| = (r-1)|U|$.

Note that H_1 was formed by deleting $(1-\alpha)n$ vertices from U to form X_1, and taking $Y_1 = T$, $Z_1 = S$. By the above discussion, we obtain a contradicting inequality in (3.3) unless $G[T]$ is a Turán graph with $\frac{r-1}{2r}t^2n^2$ edges and $G[S]$ is a Turán graph with $\frac{r-2}{2(r-1)}s^2n^2$ edges. We must also have

$$e(X_1,S) = |X_1|\,|S|, \quad e(X_1,T) = \frac{r-1}{r}|X_1|\,|T|, \quad e(S,T) = \frac{r-1}{r}|S|\,|T|.$$

All this should be valid for every subset X_1 of U of size $|U|-(1-\alpha)n$. Therefore, by choosing X_1 uniformly at random we obtain that the expected value

$$\mathbf{E}(e(X_1,S)+e(X_1,T)) = \frac{|X_1|}{|U|}(e(U,S)+e(U,T)) = |X_1|\,|S| + \frac{r-1}{r}|X_1|\,|T|.$$

This implies that $e(U,S)+e(U,T) = |U|\,|S| + \frac{r-1}{r}|U|\,|T|$. Now, similar computations to those in the proof of Lemma 3.1 show that

$$\begin{aligned} e(G) &= \frac{r-1}{2r}(|U|+|S|+|T|)^2 - \frac{1}{2r(r-1)}((r-1)|U|-|S|)^2 \\ &\quad + e(S) - \frac{r-2}{2(r-1)}|S|^2 = \frac{r-1}{2r}n^2. \end{aligned}$$

This completes the proof of the theorem, since the only K_{r+1}-free graph of order n with $\frac{r-1}{2r}n^2$ edges is a Turán graph. □

Let us now examine more carefully the first part of the proof to work out the best range of α given by our method. Suppose we want to prove that, if $c_r \leqslant \alpha \leqslant 1$, then any K_{r+1}-free graph of order n contains a set of αn vertices spanning at most $\frac{r-1}{2r}(2\alpha-1)n^2$ edges. We choose U to be an independent set of size $k_r(1-\alpha)n$, for some k_r. From (3.2) we see that we need to choose c_r such that, if $\alpha \geqslant c_r$, then $\theta = \frac{1}{\alpha} - 1$ satisfies

$$1-(r-1)^2\theta^2 \geqslant \frac{rk_r\theta}{1+\theta}.$$

From this choice of U we follow the proof to deduce that $s > (r-1)(k_r-1)(1-\alpha)$. Then (3.5) becomes

$$\alpha_1 = 1 - \frac{1-\alpha}{s} > 1 - \frac{1}{(r-1)(k_r-1)}.$$

Therefore, for the application of the inductive hypothesis to be valid we need

$$c_{r-1} \geqslant 1 - \frac{1}{(r-1)(k_r-1)}.$$

This implies that our proof works for c_r, satisfying the following recursion.

- $c_2 = \frac{3}{5}$.
- For $r \geqslant 3$, let $k_r = \frac{1}{(r-1)(1-c_{r-1})} + 1$ and let $\theta_r > 0$ be maximal such that

$$1 - (r-1)^2\theta_r^2 \geqslant rk_r \frac{\theta_r}{1+\theta_r}.$$

Then $c_r = \frac{1}{1+\theta_r}$.

We conclude this section by presenting the first few values of c_r, which can be computed from the above recursion: $c_3 = 0.8661$, $c_4 = 0.9318$, $c_5 = 0.9585$, $c_6 = 0.972$.

4. Local density version of the Erdős–Stone theorem

In this short section we show how to use Theorem 1.1 to deduce a similar statement about edge distribution in graphs with any fixed forbidden subgraph. This implies the local density version of the Erdős–Stone theorem. Our proof uses Szemerédi's Regularity Lemma and we refer the interested reader to the excellent survey by Komlós and Simonovits [6], which discusses various applications of this powerful tool.

We start with a few definitions, most of which follow [6]. Let $G = (V, E)$ be a graph, and let A and B be two disjoint subsets of $V(G)$. If A and B are non-empty, define the *density of edges* between A and B by

$$d(A, B) = \frac{e(A, B)}{|A|\,|B|}.$$

For $\epsilon > 0$ the pair (A, B) is called *ϵ-regular* if, for every $X \subset A$ and $Y \subset B$ satisfying $|X| > \epsilon|A|$ and $|Y| > \epsilon|B|$, we have

$$|d(X, Y) - d(A, B)| < \epsilon.$$

An *equitable partition* of a set V is a partition of V into pairwise disjoint classes $V_1, \dots, V_k$ of almost equal size, that is, $||V_i| - |V_j|| \leqslant 1$ for all i, j. An equitable partition of the set of vertices V of G into the classes $V_1, \dots, V_k$ is called *ϵ-regular* if $|V_i| \leqslant \epsilon|V|$ for every i and all but at most ϵk^2 of the pairs (V_i, V_j) are ϵ-regular. The above partition is called *totally ϵ-regular* if all the pairs (V_i, V_j) are ϵ-regular. The following celebrated lemma was proved by Szemerédi in [9].

Lemma 4.1. *For every $\epsilon > 0$ there is an integer $M(\epsilon)$, such that every graph of order $n > M(\epsilon)$ has an ϵ-regular partition into k classes, where $k \leqslant M(\epsilon)$.*

In order to apply the Regularity Lemma we need to show the existence of a complete multipartite subgraph in graphs with a totally ϵ-regular partition. This is established in the following well-known lemma: see, *e.g.*, [6].

Lemma 4.2. *For every $\delta > 0$ and integers m and t there exist an $0 < \epsilon = \epsilon(\delta, t, m)$ and $n_0 = n_0(\delta, t, m)$ with the following property. If G is a graph of order $n > n_0$ and $(V_1, \ldots, V_m)$ is a totally ϵ-regular partition of vertices of G such that $d(V_i, V_j) \geqslant \delta$ for all $i < j$, then G contains a complete m-partite subgraph $K_m(t)$.*

Proof of Theorem 1.4. Let r and t be fixed positive integers and let G be a graph of order n not containing a copy of $K_{r+1}(t)$. Suppose $\delta > 0$ and let $\epsilon = \min(\delta, \epsilon(\delta, t, r+1))$, where $\epsilon(\delta, t, r+1)$ is defined in the previous statement. Then, by Lemma 4.1, for sufficiently large n there exists an ϵ-regular partition $(V_1, \ldots, V_k)$ of vertices of G.

Consider a new graph G' on the vertices $\{1, \ldots, k\}$ in which (i, j) is an edge if and only if (V_i, V_j) is an ϵ-regular pair with density at least δ. We claim that G' contains no clique of size $r+1$. Indeed, any such clique in G' corresponds to $r+1$ parts in the partition of G such that any pair of them is ϵ-regular and has density at least δ. This contradicts our assumption on G, since by Lemma 4.2 the union of these parts will contain a copy of the complete $(r+1)$-partite graph $K_{r+1}(t)$.

Next, by applying Theorem 1.1 to graph G', we deduce that it contains a subset W' of size $\lfloor \alpha k \rfloor$ that spans at most $\frac{r-1}{2r}(2\alpha - 1)k^2$ edges of G'. Then $W = \bigcup_{i \in W'} V_i$ is a subset of G of size at least $(\frac{n}{k} - 1)\lfloor \alpha k \rfloor > \alpha n - \frac{2n}{k} > \alpha n - 2\delta n$ which contains at most $\frac{r-1}{2r}(2\alpha - 1)k^2$ ϵ-regular pairs with density at least δ. These pairs span at most $\frac{r-1}{2r}(2\alpha - 1)n^2$ edges. Note also that the total number of edges in G that lie within classes of the partition, or that belong to a non-regular pair, or that join a pair of classes of density less than δ is at most $n^2/k + \epsilon n^2 + \delta n^2 \leqslant 3\delta n^2$. Therefore W spans at most $\big(\frac{r-1}{2r}(2\alpha - 1) + 3\delta\big)n^2$ edges. Complete it arbitrarily to a subset of G size αn. Clearly, the new set spans at most

$$\left(\frac{r-1}{2r}(2\alpha - 1) + 3\delta\right)n^2 + 2\delta n^2 = \left(\frac{r-1}{2r}(2\alpha - 1) + 5\delta\right)n^2$$

edges. This completes the proof of the theorem. □

5. Concluding remarks

In this paper we have proved that, if G is a K_{r+1}-free graph on n vertices and $1 - \frac{1}{2r^2} \leqslant \alpha \leqslant 1$, then G contains a set of αn vertices spanning at most $\frac{r-1}{2r}(2\alpha - 1)n^2$ edges. The value $\frac{r-1}{2r}(2\alpha - 1)$ given for the local density is best possible, as it is attained by the corresponding Turán graph $T_r(n)$. On the other hand, the range of α in which we make our statement is relatively small. It would be very interesting to extend this range, even if one could only prove the first part of our theorem, namely, without characterizing the extremal graphs. As already mentioned in the introduction, the given formula describes the local density of $T_r(n)$ provided $\alpha \geqslant \frac{r-1}{r}$, so it is natural to believe that the following might be true.

Conjecture 5.1. *Let $r \geqslant 3$ and let G be a K_{r+1}-free graph on n vertices. If $\frac{r-1}{r} \leqslant \alpha \leqslant 1$, then G contains a set of αn vertices spanning at most $\frac{r-1}{2r}(2\alpha - 1)n^2$ edges.*

Note that the above statement is not true if $r = 2$, by the example given earlier of the blow-up of a 5-cycle. On the other hand for $r \geqslant 3$, Chung and Graham [3] made the even more ambitious conjecture that the corresponding Turán graph should be extremal for the problem of maximizing local density in K_{r+1}-free graphs for any $\alpha \geqslant 1/2$. In particular, for $\alpha = 1/2$ they conjectured the following.

Conjecture 5.2. *Let $r \geqslant 3$ and let G be a K_{r+1}-free graph of order n. Then G contains $n/2$ vertices which span at most $\frac{r-2}{8r}n^2$ edges if r is even and at most $\frac{(r-1)^2}{8r^2}n^2$ edges if r is odd.*

By analogy with the triangle-free case, it is natural to try to disprove this by considering the blow-up of an appropriate Ramsey graph. It is worth mentioning here that we have checked the blow-ups of the graphs that give the values of two small Ramsey numbers, $R(4,3)$ and $R(4,4)$, but that this construction fails to produce a counterexample to the Chung–Graham conjecture. This may be taken as weak supporting evidence. Finally we want to present an argument that gives a more compelling reason for there to be some room for improvement in our result.

In the proof of Theorem 1.1, the first crucial step is to find a sufficiently large independent set, which we deduce from a lower bound on the number of edges in the graph G given by a simple averaging argument. But, as observed in [7], given this independent set, we can iteratively improve our bound on the number of edges in G and so find a bigger independent set as follows.

Suppose that we have a graph G on n vertices such that every αn set spans more than βn^2 edges, and we also have some independent set A of size γn, for some $1 - \alpha \leqslant \gamma \leqslant \alpha$. Let B denote the complement of A in G. Define d to be the density of edges between A and B and let $\theta = e(B)/|B|^2$. Then it is easy to check

$$\frac{1}{n^2}e(G) = d\gamma(1-\gamma) + \theta(1-\gamma)^2.$$

Consider the two subsets obtained by taking all vertices of A and a random part of B, or all vertices of B and random part of A so that the total number of vertices is αn. Using a simple averaging argument we obtain the following inequalities:

$$d\gamma(\alpha-\gamma) + \theta(\alpha-\gamma)^2 > \beta,$$
$$d(1-\gamma)(\alpha+\gamma-1) + \theta(1-\gamma)^2 > \beta.$$

By taking the sum of the first inequality multiplied by $\frac{(1-\gamma)^2}{\alpha(\alpha-\gamma)}$ and the second inequality multiplied by $\frac{\gamma}{\alpha}$, after some algebraic manipulation, we get

$$e(G) = (d\gamma(1-\gamma) + \theta(1-\gamma)^2)n^2 > \left(\frac{1-2\gamma+\gamma\alpha}{\alpha(\alpha-\gamma)}\right)\beta n^2.$$

Now suppose, for instance, we are trying to prove the result in the K_4-free case, with α slightly below 0.8661 (the threshold for finding an independent set of size $\frac{9}{4}(1-\alpha)n$).

First we use the simple averaging argument to get a lower bound on the number of edges. Then we apply Lemma 3.3 to find an independent set, and denote its size by γn. Next we apply the above argument (with $\beta = \frac{2\alpha-1}{3}$) to get a better bound on the number of edges. Then we again use Lemma 3.3 and repeat the procedure. A simple but tedious computation with this iteration, which we omit, shows that in Theorem 2.1 it suffices to take $\alpha \geqslant 0.861$.

This gives a very minor improvement to the range, but it may also indicate that there is considerable room for further improvement.

References

[1] Bollobás, B. (1995) Extremal graph theory. In *Handbook of Combinatorics*, Vol. 2, Elsevier, Amsterdam, pp. 1231–1292.

[2] Brandt, S. (1998) The local density of triangle-free graphs. *Discrete Math.* **183** 17–25.

[3] Chung, F. and Graham, R. (1990) On graphs not containing prescribed induced subgraphs. In *A Tribute to Paul Erdős*, Cambridge University Press, Cambridge, pp. 111–120.

[4] Erdős, P. and Stone, A. H. (1946) On the structure of linear graphs. *Bull. Amer. Math. Soc.* **52** 1087–1091.

[5] Erdős, P., Faudree, R., Rousseau, C. and Schelp, R. (1994) A local density condition for triangles. *Discrete Math.* **127** 153–161.

[6] Komlós, J. and Simonovits, M. (1996) Szemerédi's Regularity Lemma and its applications in graph theory. In *Combinatorics: Paul Erdős is Eighty*, Vol. 2, János Bolyai Math. Soc., Budapest, pp. 295–352.

[7] Krivelevich, M. (1995) On the edge distribution in triangle-free graphs. *J. Combin. Theory Ser. B* **63** 245–260.

[8] Moon, J. W. and Moser, L. (1962) On a problem of Turán. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** 283–286.

[9] Szemerédi, E. (1978) Regular partitions of graphs. In Vol. 260 of *Proc. Colloque Inter. CNRS*, CNRS, Paris, pp. 399–401.