Dense H-free graphs are almost $(\chi(H)-1)$-partite

Peter Allen*

Submitted: Jul 24, 2009; Accepted: Jan 19, 2010; Published: Jan 29, 2010
Mathematics Subject Classification: 05C35

Abstract

By using the Szemerédi Regularity Lemma [10], Alon and Sudakov [1] recently extended the classical Andrásfai-Erdős-Sós theorem [2] to cover general graphs. We prove, without using the Regularity Lemma, that the following stronger statement is true.

Given any $(r+1)$-partite graph H whose smallest part has t vertices, there exists a constant C such that for any given $\varepsilon > 0$ and sufficiently large n the following is true. Whenever G is an n-vertex graph with minimum degree

$$\delta(G) \geqslant \left(1 - \frac{3}{3r-1} + \varepsilon\right) n \ ,$$

either G contains H, or we can delete $f(n,H) \leqslant Cn^{2-\frac{1}{t}}$ edges from G to obtain an r-partite graph. Further, we are able to determine the correct order of magnitude of $f(n,H)$ in terms of the Zarankiewicz extremal function.

1 Introduction

We define the graph $K_r(s)$ to be the complete r-partite graph whose parts each have s vertices. Given a graph H, whose chromatic number is $\chi(H)$, we examine all the proper $\chi(H)$-colourings of H. We choose one whose smallest colour class is of smallest possible size; then $\sigma(H)$ is the size of this smallest colour class. Otherwise, our notation is standard.

We recall the classical theorem of Zarankiewicz [12]:

Theorem 1. *If the n-vertex graph G has minimum degree exceeding $\left(1-\frac{1}{r}\right)n$ then G contains K_{r+1}.*

*DIMAP and Mathematics Institute, University of Warwick, Coventry, CV4 7AL, U.K. Email: p.d.allen@warwick.ac.uk . Research supported by the Centre for Discrete Mathematics and its Applications, EPSRC award EP/D063191/1.

This theorem is an immediate corollary of Turán's theorem [11]. As is well known, it is best possible, the extremal example being a complete balanced r-partite graph (sometimes called a Turán graph). An old result of Andrásfai, Erdős and Sós [2], which amounts to a (very strong) stability result for Zarankiewicz' theorem, is the following.

Theorem 2. *Suppose $r \geqslant 2$. If the n-vertex graph G has minimum degree exceeding $\left(1 - \frac{3}{3r-1}\right) n$ and G does not contain K_{r+1}, then G is r-partite.*

This theorem is best possible; however the extremal example is a little more complex than the Turán graph. We construct a graph $E_r(n)$ as follows: we partition n vertices into $r-2$ sets $X_1, \ldots, X_{r-2}$ each containing $\frac{3n}{3r-1}$ vertices and five sets $Y_1, \ldots, Y_5$ each containing $\frac{n}{3r-1}$ vertices. Each of these sets is independent; we set every vertex in each X_i adjacent to all vertices outside X_i, and we make $(Y_i, Y_{i+1 \mod 5})$ a complete bipartite graph for each i (so that the five sets form a blow-up of C_5). It is straightforward to check that each vertex has degree $\left(1 - \frac{3}{3r-1}\right) n$; since $\chi(C_5) = 3$ the chromatic number of $E_r(n)$ is $r+1$, but $E_r(n)$ does not contain K_{r+1}.

Erdős and Stone [6] extended Zarankiewicz' theorem, showing that for any fixed graph H, the chromatic number of H governs the minimum degree threshold at which H appears in a large graph G:

Theorem 3. *Let H be any fixed graph with chromatic number $r+1$. If the n-vertex graph G has minimum degree exceeding $\left(1 - \frac{1}{r} + o(1)\right) n$ then G contains H.*

Although the extremal graphs for this theorem are not necessarily r-partite, it is true that one may delete $o(n^2)$ edges from any extremal graph to obtain an r-partite graph. Indeed, it is not hard to show that there exists $\varrho = \varrho(H) > 0$ such that deletion of only $O(n^{2-\varrho})$ edges from an extremal graph yields an r-partite graph.

Quite recently, Alon and Sudakov [1] gave an extension of Andrásfai, Erdős and Sós' result to cover all fixed graphs H (Erdős and Simonovits [5] had previously considered the case when H is critical, i.e. when there is an edge of H whose removal decreases the chromatic number):

Theorem 4. *Let any fixed graph H with chromatic number $r+1$ and constant $\varepsilon > 0$ be given. Then there exist $\varrho = \varrho(H) > 0$ and $n_0 = n_0(H, \varepsilon)$ such that the following holds. If $n \geqslant n_0$ and G is an n-vertex graph with minimum degree exceeding $\left(1 - \frac{3}{3r-1} + \varepsilon\right) n$ which does not contain H, then one can delete at most $O(n^{2-\varrho})$ edges from G to yield an r-partite graph.*

Alon and Sudakov gave a value for the constant $\varrho(H)$. They showed that if we have $H \subseteq K_{r+1}(s)$ then we may take $\varrho(H) = 1/4r^{2/3}s$. The purpose of this paper is to give a simpler proof (avoiding the use of the Regularity Lemma) which gives the correct order of magnitude of the number of edges that must be deleted (albeit in terms of the Zarankiewicz problem).

Recall that given a family $\mathcal{H}$ of graphs, $\mathrm{ex}(n, \mathcal{H})$ is defined to be the maximum number of edges in an n-vertex graph which does not contain a copy of any graph $H \in \mathcal{H}$.

Given a graph H, we define a quantity $\mathrm{biex}(n, H)$ as follows. Let $c : V(H) \to [\chi(H)]$ be any proper $\chi(H)$-colouring of H. Let $S_c = c^{-1}(\{1, 2\})$ be the vertices receiving colours 1 and 2 in this colouring. Consider the family of graphs $\mathcal{F}$ containing all graphs of the form $H[S_c]$ for some proper $\chi(H)$-colouring c of H. Then we set $\mathrm{biex}(n, H) = \mathrm{ex}(n, \mathcal{F})$.

We note that if H is a complete r-partite graph, whose smallest part has t vertices and whose next smallest part has s vertices, then $\mathrm{biex}(n, H) = \mathrm{ex}(n, K_{t,s})$.

The problem of estimating $\mathrm{ex}(n, H)$ when H is bipartite (or, more generally, $\mathrm{ex}(n, \mathcal{H})$ for a family $\mathcal{H}$ of bipartite graphs) is the Zarankiewicz problem; for most H it is quite far from being solved. However an upper bound is provided by the following classical theorem of Kövári, Sós and Turán [8].

Theorem 5. *Let $1 \leqslant t \leqslant s$ be fixed integers. If G is any n-vertex graph with $\Omega(n^{2-\frac{1}{t}})$ edges, then G contains $K_{t,s}$.*

We note that for $t = 1, 2, 3$ and when $s \geqslant t! + 1$ there exist lower bound constructions matching the upper bound of Theorem 5 (see [9, 3, 7]); for $t \geqslant 4$ the best known lower bound is $\Omega(n^{2-\frac{2}{t+1}})$, but it is conjectured that the correct bound is $\Theta(n^{2-\frac{1}{t}})$.

We can now state our main theorem.

Theorem 6. *To any graph H with chromatic number $r + 1$ there is associated a constant $C = C(H)$ such that whenever $\varepsilon > 0$ is given, there is n_0 for which the following holds. Whenever $n \geqslant n_0$ and G is an n-vertex graph with minimum degree exceeding $\left(1 - \frac{3}{3r-1} + \varepsilon\right) n$ which does not contain H, then one can delete at most $C\,\mathrm{biex}(n, H)$ edges from G to obtain an r-partite graph.*

This theorem is best possible up to the value of C. For comparison with the result of Alon and Sudakov, suppose $H \subseteq K_{t,s,s,\ldots,s}$ has chromatic number $r + 1$, where $t \leqslant s$. Then, applying Theorem 5, we have

$$\mathrm{biex}(n, H) \leqslant \mathrm{ex}(n, K_{t,s}) = O(n^{2-\frac{1}{t}}) \ .$$

It follows that if G satisfies the conditions of Theorem 6, then Theorem 4 guarantees that G can be made r-partite by deleting $O(n^{2-\frac{1}{4r^{2/3}s}})$ edges; Theorem 6 strengthens this to $Cn^{2-\frac{1}{t}}$ edges. On the assumption that the conjectured bound in the Zarankiewicz problem is correct, this is best possible up to the value of the multiplicative constant. Furthermore, the constant hidden behind the $O(\cdot)$ notation in Theorem 4 depends upon ε; specifically, it grows as a polynomial function of $1/\varepsilon$, whereas the constant C in Theorem 6, while surely much larger than it 'should' be, does not depend on ε. Finally, owing to the use of the Regularity Lemma, the constant n_0 in Theorem 4 has an exceptionally unpleasant dependence on ε, r and s.

We give two constructions which demonstrate the tightness of our theorem.

Given H, let E be an n-vertex graph with $\mathrm{biex}(n, H)$ edges and not containing any of the forbidden bipartite subgraphs. Let E' be an n/r-vertex subgraph of E containing the maximum possible number of edges. Note that $e(E') > e(E)/2r^2 = \Omega(\mathrm{biex}(n, H))$.

Consider the graph G obtained from the complete balanced r-partite graph by replacing one part with E'. This graph has minimum degree $\frac{r-1}{r}n$, and does not contain a copy of H. However to make G r-partite we must delete $\Omega(\mathrm{biex}(n,H))$ edges.

Alon and Sudakov asked whether it is possible to replace the term εn in the minimum degree of their theorem with an $O(1)$ term. It is not possible; indeed, for any $\mu > 0$ there are graphs H such that the corresponding term must be larger than $n^{1-\mu}$.

Consider the following modification of $E_r(n)$. Let c be some sufficiently small positive quantity. We let each of the independent sets $Y_1, \ldots, Y_5$ have $\frac{n}{3r-1} + (r-2)cn^{1-2/t}$ vertices. We let each of the independent sets $X_1, \ldots, X_{r-2}$ have $\frac{3}{3r-1} - 5cn^{1-2/t}$ vertices. Finally, we take a $K_{t,t}$-free graph E' on $|Y_1|$ vertices with minimum degree $(3r-1)cn^{1-2/t}$: provided $c > 0$ is chosen sufficiently small, such a graph exists. We replace each of the independent sets $Y_1, \ldots, Y_5$ with E' to obtain $E'_{r,t}(n)$. Now observe that the minimum degree of $E'_{r,t}(n)$ is $\frac{3r-4}{3r-1}n + 5cn^{1-2/t}$. However it is not possible to find a copy of $K_{r+1}(2t)$ in $E'_{r,t}(n)$. The reason is that it would be necessary to find a copy of $K_3(2t)$ within the graph induced by $Y_1 \cup \ldots \cup Y_5$; this would require that one of the Y_i contained $K_{t,t}$, which by construction is false. Finally, it is clear that to make $E'_{r,t}(n)$ r-partite requires the removal of $\Omega(n^2)$ edges.

2 Constructing $(r+1)$-partite graphs

Given an $(r+1)$-partite graph H, a large graph G, and a family $\mathcal{F}$ consisting of the bipartite subgraphs of H whose removal decreases the chromatic number of H by two, we describe a construction of the graph H from a suitably well-structured set of copies of K_{r+1} in G. Alon and Sudakov made use of a related construction: the difference is that their construction as its first step finds (by use of the Kövári-Sós-Turán theorem) one specific bipartite subgraph of G and proceeds to build H using it. Our construction avoids this, relying instead on counting the number of suitable objects until the final step in the construction. This difference is primarily responsible for our improved bounds.

Given a graph G and a vertex $v \in G$, let G_v be the *neighbourhood graph* obtained by deleting from G every edge which is not contained in the neighbourhood of v.

We give first a counting variant of a lemma of Erdős [4]; this is essentially a statement about dense hypergraphs generalising the Kövári-Sós-Turán theorem.

Lemma 7. *For every r, s and $\varepsilon > 0$ there exists $\delta = \delta_{r,s}(\varepsilon) > 0$ such that the following holds for sufficiently large n. If the n-vertex graph G contains at least εn^r copies of K_r, then G contains $\delta_{r,s}(\varepsilon)n^{rs}$ copies of $K_r(s)$.*

Proof. For $r = 1$ the statement holds trivially. We complete the proof by induction.

Let G be an n-vertex graph containing εn^r copies of K_r: then there are some $\varepsilon n/2$ vertices D of G which are each contained in $\varepsilon n^{r-1}/2$ copies of K_r in G. By construction, for each $d \in D$, G_d contains $\varepsilon n^{r-1}/2$ copies of K_{r-1}; by induction it contains $\delta_{r-1,s}(\varepsilon/2)n^{(r-1)s}$ copies of $K_{r-1}(s)$.

For a given copy S of $K_{r-1}(s)$, let d_S be the number of vertices of D whose neighbourhoods contain S. Then we have (using the convention that $\binom{a}{b} = 0$ when $a < b$) at least $\frac{1}{r}\sum_S \binom{d_S}{s}$ copies of $K_r(s)$ contained in G. Since the mean value of d_S is at least $\delta_{r-1,s}(\varepsilon/2)|D|$, applying Jensen's inequality the number of copies of $K_r(s)$ in G is at least

$$\frac{1}{r}\sum_S \binom{d_S}{s} \geqslant \frac{1}{r}\delta_{r-1,s}(\varepsilon/2)n^{(r-1)s}\binom{\delta_{r-1,s}(\varepsilon/2)|D|}{s} = \delta_{r,s}(\varepsilon)n^{rs} ,$$

as required. □

Note that the value of $\delta_{r,s}(\varepsilon)$ obtained by the above method is polynomial in ε.

To complete our construction, we give the following corollary of Lemma 7.

Corollary 8. *Given $\varepsilon > 0$ and H there exists C such that for sufficiently large n the following is true. Every n-vertex graph G in which there are more than $C\mathrm{biex}(n, H)$ edges E of G, each contained in εn^{r-1} copies of K_{r+1}, contains H.*

Proof. Let G be a graph with a set E of edges each of whose common neighbourhoods contains εn^{r-1} copies of K_{r-1}. Suppose that n is large enough to permit us to conclude, by Lemma 7, that the common neighbourhood of each edge of E contains at least $\delta_{r-1,v(H)}(\varepsilon)n^{(r-1)v(H)}$ copies of $K_{r-1}(v(H))$. Let $C = 1/\delta_{r-1,v(H)}(\varepsilon)$. Suppose furthermore that $|E| > C\mathrm{biex}(n, H)$.

By averaging, there is one copy S of $K_{r-1}(v(H))$ in G which lies in the common neighbourhood of each of the edges $E' \subseteq E$, with $|E'| > \mathrm{biex}(n, H)$. By definition of $\mathrm{biex}(n, H)$, the edges E' must contain a copy of some bipartite subgraph of H in $\mathcal{F}$. Let this subgraph be B. Then $B \cup S$ contains H. □

Note that the value of $\delta_{r,s}(\varepsilon)$ given by Lemma 7 is clearly far smaller than the truth; but this affects only the constant C; furthermore, the dependence on ε is polynomial.

3 Proof of Theorem 6

We first prove a density version of Theorem 2. We note that Alon and Sudakov [1] proved a similar lemma; however their method (while in most ways similar to ours) obtained a first 'coarse' version by application of the Szemerédi Regularity Lemma. We avoid this by making use of an induction argument.

Lemma 9. *Given r and ε, let $\mu = \varepsilon^r/r!$ and $\eta = \varepsilon^{r+1}/(r+1)!$. Then whenever n is sufficiently large, the following is true. Any n-vertex graph G with $\delta(G) > \left(1 - \frac{3}{3r-1} + 4\varepsilon\right)n$ either contains more than ηn^{r+1} copies of K_{r+1}, or has a partition into $D \cup V_1 \cup \ldots \cup V_r$, with the properties that $\Delta(G[V_i]) \leqslant \varepsilon n$ for each i, each vertex of D is contained in more than μn^r copies of K_{r+1}, and $|D| \leqslant \varepsilon n$.*

Note that when $\varepsilon = 0$ we have $\mu = \eta = 0$, and we obtain the statement of Theorem 2. The intuition is that since we are looking at graphs which do not contain a high density of copies of K_{r+1}, rather than not containing any at all, we must expect that there may be some small set of vertices, and a few edges leaving every vertex, which 'misbehave'. These are, respectively, the set D and the replacement of the independent sets of Theorem 2 with sets which simply have restricted maximum degree.

Proof. We prove the lemma by induction. The $r = 1$ case is a triviality: either there are more than εn vertices of degree exceeding μn, in which case G certainly contains more than ηn^2 edges, or we can let D be the set of all vertices of degree exceeding μn, and together with $V_1 = V(G) \setminus D$ the partition conclusion is satisfied.

Suppose $r \geqslant 2$. We assume as our induction hypothesis that the lemma holds for $r-1$.

Let G be an n-vertex graph with minimum degree $\left(1 - \frac{3}{3r-1} + 4\varepsilon\right) n$. We presume G contains at most ηn^{r+1} copies of K_{r+1}.

Let $D \subseteq V(G)$ be the set of all vertices $d \in G$ such that there are more than μn^r copies of K_r in $\Gamma(d)$. Then $|D| \leqslant \varepsilon n$ since G contains at most ηn^{r+1} copies of K_{r+1}.

Let $G' = G[V(G) - D]$. This graph has minimum degree greater than $\left(\frac{3r-4}{3r-1} + 3\varepsilon\right) n$; none of its vertices are contained in more than μn^r copies of K_r.

Let X_1 be a maximum cardinality set in $V(G')$ with the property that $\Delta(G'[X_1]) \leqslant \varepsilon n$. Let $v \in X_1$.

Consider the graph $N = G'[\Gamma(v) \setminus X_1]$. Because $v \notin D$, the neighbourhood graph G_v contains at most μn^r copies of K_r, and so in particular N contains at most μn^r copies of K_r. Because $\Delta(G'[X_1]) \leqslant \varepsilon n$, $v(N) > \frac{3r-4}{3r-1} n + 2\varepsilon n$. Now consider $u \in N$. We have

$$d_N(u) > v(N) - \left(\frac{3}{3r-1} - 4\varepsilon\right) n$$

$$> v(N) - \left(\frac{3}{3r-1} - 4\varepsilon\right) \frac{3r-1}{3r-4} v(N) > \left(\frac{3r-7}{3r-4} + 4\varepsilon\right) v(N) \ .$$

By induction, we have that N has a partition $V(N) = B \cup X_2 \cup \ldots \cup X_r$, where $|B| \leqslant \varepsilon n$ and $\Delta(N[X_i]) \leqslant \varepsilon n$ for each of the $r-1$ sets $X_2, \ldots, X_r$.

Because X_1 has maximum cardinality subject to $\Delta(G'[X_1]) \leqslant \varepsilon n$, $|X_1| \geqslant |X_i|$ for each i. In particular, we have

$$|X_1| + \ldots + |X_r| \geqslant \left(\frac{3r-4}{3r-1} + \varepsilon\right) \frac{rn}{r-1} \geqslant \frac{(3r-4)rn}{(3r-1)(r-1)} + \varepsilon n \ .$$

Since every vertex in G has more than $\frac{3r-4}{3r-1} n + 4\varepsilon n$ neighbours in G, and since for each i we have $\Delta(G[X_i]) \leqslant \varepsilon n$, it follows that $|X_i| < \frac{3}{3r-1} n$ for each i.

Now suppose that for some i we have $|X_i| \leqslant \frac{2}{3r-1} n$. Because X_1 was chosen to be maximal, we may assume $2 \leqslant i \leqslant r$; without loss of generality let us suppose $i = r$. We have $|B| + |X_2| + \ldots + |X_r| = v(N) \geqslant \frac{3r-4}{3r-1} n + 2\varepsilon n$, and since also $|B| \leqslant \varepsilon n$, we have $|X_2| + \ldots + |X_{r-1}| \geqslant \frac{3r-6}{3r-1} n + \varepsilon n$. It follows that among the $r-2$ sets $X_2, \ldots, X_{r-1}$, there

must be one whose size exceeds $\frac{3r-6}{(3r-1)(r-2)}n = \frac{3}{3r-1}n$, which is a contradiction. Thus we have that for each i, $\frac{2}{3r-1}n < |X_i| < \frac{3}{3r-1}n$.

Now, if we have any two adjacent vertices u and v of G' whose codegree exceeds $\frac{3r-6}{3r-1}n + \varepsilon n$, then we may construct a clique K_{r+1} extending uv greedily by simply picking any common neighbour of the so far chosen vertices at each step. At the final step (and therefore at all steps) we have at least εn choices. It follows that any edge uv of G in which the common neighbourhood of u and v exceeds $\frac{3r-6}{3r-1}n + \varepsilon n$ lies in more than $\varepsilon^{r-1}n^{r-1}/(r-1)!$ cliques K_{r+1}.

Furthermore, if u has more than εn neighbours with each of which its codegree exceeds $\frac{3r-6}{3r-1}n+\varepsilon n$, then u lies in more than $\varepsilon^r n^r/r! = \mu n^r$ copies of K_{r+1}. This contradicts $u \notin D$.

Since $\Delta(G[X_i]) \leqslant \varepsilon n$, if a vertex u outside X_i has less than $|X_i| - \frac{n}{3r-1}$ neighbours in X_i, then the codegree of u and any neighbour $v \in X_i$ exceeds $\frac{3r-6}{3r-1}n + \varepsilon n$. It follows that any vertex of G' outside X_i has either fewer than εn neighbours in X_i or more than $|X_i| - \frac{n}{3r-1}$ neighbours in X_i.

Consider the set L_i of vertices of L which all have less than εn neighbours in X_i. Any one of these vertices has codegree exceeding $\frac{3r-6}{3r-1}n + \varepsilon n$ with any other, and with any vertex of X_i. It follows that $L_i \cup X_i$ has maximum degree εn. Let this set be V_i. Let the vertices of G' not in any X_i' be L'.

If $L' = \emptyset$ then we have $V(G) = D \cup V_1 \cup \ldots \cup V_r$ is the desired partition. So we may assume there is a vertex $l \in L'$. This vertex is non-adjacent to fewer than $\frac{n}{3r-1}$ vertices of each set V_i. It is convenient to assume that the sets $V_1, \ldots, V_r$ are in order of decreasing size.

Finally, consider the following greedy construction. We start with the vertex $l \in L'$. We now choose vertices $v_1, \ldots, v_r$ from the respective sets $V_1, \ldots, V_r$, such that after each choice the vertices chosen together with l form a clique.

At the first step we have more than $|V_1| - \frac{n}{3r-1}$ choices for v_1. At the second step we have more than

$$|V_2| - \frac{n}{3r-1} - \left(\frac{3}{3r-1} - 4\varepsilon\right)n + (|V_1| - \varepsilon n) = |V_1| + |V_2| - \frac{4}{3r-1}n + 3\varepsilon n$$

choices for v_2; there are less than $\frac{n}{3r-1}$ non-neighbours of l in V_2, and at most $\frac{3n}{3r-1} - 4\varepsilon n$ non-neighbours of v_1 in G, of which at least $|V_1| - \varepsilon n$ are in V_1. In general, for each $2 \leqslant i \leqslant r$, we have at the ith step more than

$$|V_1| + \ldots + |V_i| - \frac{3i-2}{3r-1}n + 3\varepsilon n$$

choices for v_i. Because the sets $V_1, \ldots, V_r$ are in order of decreasing size, the number of choices is least when choosing either v_1 or v_r. Since $|V_1| \geqslant |X_1| > \frac{3r-4}{(3r-1)(r-1)}n \geqslant \frac{2}{3r-1}n$, the number of choices for v_1 is greater than $\frac{n}{3r-1}$. Since

$$|V_1| + \ldots + |V_r| \geqslant |X_1| + \ldots + |X_r| \geqslant \frac{(3r-4)r}{(3r-1)(r-1)}n + \varepsilon n \ ,$$

the number of choices for v_r is at least $\frac{r-2}{(3r-1)(r-1)}n + 4\varepsilon n$. It follows that at each step there are more than εn choices; therefore l is contained in more than $\varepsilon^r n^r \geqslant \mu n^r$ copies of K_{r+1} in G, which contradicts $l \notin D$. □

At last, we can complete the proof of our main theorem. Again, our method is similar to that of Alon and Sudakov [1]; we take a little more care in order to ensure that the constant C in our theorem is independent of ε.

Proof of Theorem 6. Given $r \geqslant 2$ and $\varepsilon > 0$, let G be a sufficiently large n-vertex graph with $\delta(G) \geqslant \left(1 - \frac{3}{3r-1} + \varepsilon\right) n$ which does not contain the $(r+1)$-partite graph H.

By Lemma 9 there exist positive constants η, μ such that either G contains ηn^r copies of K_{r+1} or $V(G)$ may be partitioned as $V(G) = D \cup V_1 \cup \ldots \cup V_r$, where $\Delta(G[V_i]) \leqslant \varepsilon n/4$ for each i, each vertex of D is contained in at least μn^r copies of K_{r+1}, and $|D| \leqslant \varepsilon n/4$.

When n is sufficiently large, by Lemma 7 every graph G with ηn^{r+1} copies of K_{r+1} contains $K_{r+1}(v(H))$ and thus H. It follows that $V(G)$ possesses the given partition.

As in the proof of Lemma 9, for each i, since $\Delta(V_i) \leqslant \varepsilon n/4$ and $\delta(G) > \frac{3r-4}{3r-1}n + \varepsilon n$, we have $|V_i| < \frac{3}{3r-1}n - 3\varepsilon n/4$. Again, if for some i we have $|V_i| \leqslant \frac{2}{3r-1}n$ then among the $r-1$ sets $V_1, \ldots, V_r$ remaining there must be one whose size is at least

$$\left(n - \varepsilon n/4 - \frac{2}{3r-1}n\right)/(r-1) > \frac{3}{3r-1}n - \varepsilon n/2 \ ,$$

which again is a contradiction. Thus for each i we have $\frac{2}{3r-1}n < |V_i| < \frac{3}{3r-1}n$.

We alter slightly the partition given by Lemma 9 as follows. For each $1 \leqslant i \leqslant r$, let W_i be the set of vertices with at most $\frac{n}{4(3r-1)}$ neighbours in V_i. Let Y_i be the vertices of D with more than $\frac{n}{4(3r-1)}$ neighbours, but less than $|V_i| - \frac{3}{2(3r-1)}n$ neighbours in V_i. Let X be the vertices of D not contained in any set W_i or Y_i. By definition of V_i, we have $V_i \subseteq W_i$ for each i.

Consider the vertex $x \in X$. We make use of a greedy construction as in the proof of Lemma 9. We presume that the sets $V_1, \ldots, V_r$ are in order of decreasing size. We choose greedily vertices $v_1, \ldots, v_r$ in sets $V_1, \ldots, V_r$ (in that order), such that the set $\{x, v_1, \ldots, v_r\}$ are the vertices of an $(r+1)$-clique in G. As in the proof of Lemma 9, at the ith step we have at least

$$|V_1| + \ldots + |V_i| - \frac{3}{2(3r-1)}n - \frac{3i-3}{3r-1}n + 3\varepsilon n/4$$

choices for v_i. As before, since the sets V_i are in order of decreasing size the number of choices is fewest at either the first or the last step. The number of choices at the first step is at least $|V_1| - \frac{3}{2(3r-1)} > \frac{1}{2(3r-1)}n$; since the sets $V_1, \ldots, V_r$ together cover all of G except the at most $\varepsilon n/4$ vertices of D, the number of choices at the last step is at least

$$n - \varepsilon n/4 - \frac{3}{2(3r-1)}n - \frac{3r-3}{3r-1}n + 3\varepsilon n/4 > \frac{1}{2(3r-1)}n \ .$$

It follows that at every step there are at least $\frac{1}{2(3r-1)}n$ choices, and hence x is contained in at least

$$\left(\frac{n}{2(3r-1)}\right)^r$$

copies of K_{r+1} in G.

Consider the vertex $y \in Y_i$. Let u be any neighbour of y in V_i. The common neighbourhood of u and y contains at least

$$2\left(\frac{3r-4}{3r-1}+\varepsilon\right)n - \left(n - \frac{3}{2(3r-1)}n + \varepsilon n/4\right) > \frac{6r-11}{2(3r-1)}n$$

vertices. Now we construct an $(r+1)$-clique greedily starting from uy. At the final step, and thus at every step, we have at least $\frac{n}{2(3r-1)}$ choices. It follows that uy lies in at least $\left(\frac{n}{2(3r-1)}\right)^r/(r-1)!$ copies of K_{r+1} in G. Since y has at least $\frac{n}{4(3r-1)}$ neighbours in V_i, y lies in at least $\left(\frac{n}{4(3r-1)}\right)^r/r! = \gamma n^r$ copies of K_{r+1} in G.

Finally we have that every vertex of $Z = Y_1 \cup \ldots \cup Y_r \cup X$ lies in at least γn^r copies of K_{r+1} in G.

Now by Lemma 7 there exists $\delta > 0$ such that whenever n is sufficiently large, every graph G with γn^r copies of K_r contains $\delta n^{rv(H)}$ copies of $K_r(v(H))$. If $|Z| > (\sigma(H)-1)/\delta$, then there is one copy S of $K_r(v(H))$ in G which is in the neighbourhood of each of $\sigma(H)$ vertices B of G. But then $H \subseteq G[B \cup S]$, which is a contradiction. It follows that $|Z| \leqslant (\sigma(H)-1)/\delta$. It is important to note that γ, and hence δ, are independent of ε.

Finally, let E be the set of edges of G which are contained in any one of the sets W_i.

For any edge $uv \in E$, there is i such that $u, v \in W_i$. Then the common neighbourhood of u and v in $V(G)$ contains at least

$$2\left(\frac{3r-4}{3r-1}+\varepsilon\right)n - \left(n - |V_i| + \frac{n}{2(3r-1)}\right) \geqslant \frac{6r-11}{2(3r-1)}n + 2\varepsilon n$$

vertices, since both u and v are adjacent to at most $\frac{n}{4(3r-1)}$ vertices of V_i. As before, we can extend uv to a clique K_{r+1} by choosing vertices greedily; at each stage we have at least $\frac{n}{2(3r-1)}$ choices, and hence uv is contained in at least $\frac{n^{r-1}}{(6r-2)^{r-1}(r-1)!}$ copies of K_{r+1}. By Lemma 8, since G does not contain H, there exists C' such that $|E| \leqslant C'\mathrm{biex}(n,H)$. Observe that C' does not depend on ε.

If $\mathrm{biex}(n,H) < n-1$, then it must be the case that there is some bipartite subgraph F of H such that $F \subseteq K_{1,n-1}$ and the graph $H[V(H) \setminus V(F)]$ is $(r-1)$-colourable. But then there is a proper $(r+1)$-colouring of H in which one colour class has size one; so $\sigma(H) = 1$.

Upon deleting from G all edges incident to Z or contained in E, one obtains an r-partite graph. The total number of edges deleted is at most $n(\sigma(H)-1)/\delta + C'\mathrm{biex}(n,H)$. Since $n|Z| > 0$ only if $\sigma(H) > 1$, i.e. only if $\mathrm{biex}(n,H) \geqslant n-1$, we have $n|Z| + C'\mathrm{biex}(n,H) \leqslant C\mathrm{biex}(n,H)$, and C is as required independent of ε since C' and δ are. □

4 Concluding remarks

Perhaps the main conclusion of this paper is that (if such is necessary) there is a further motivation for solving the Zarankiewicz problem of determining $\mathrm{ex}(n, \mathcal{F})$ for all families $\mathcal{F}$ of bipartite graphs.

However there remain some open questions which are independent of the Zarankiewicz problem.

First, it would be interesting to know what the best possible value of $\mu(H)$ is such that the following statement is true.

Given H, with $\chi(H) = r + 1$, there exists C such that for all sufficiently large n, if G is an n-vertex H-free graph with minimum degree at least $\frac{3r-4}{3r-1}n + \Theta(n^{1-\mu})$, then G can be made r-partite by deleting at most $C\mathrm{biex}(n, H)$ edges.

It follows (by careful analysis of the proof given) that $\mu(H)$ must always be positive: but it seems likely that the value so obtained is much smaller than optimal.

Second, although we have shown that the correct number of edges which we should delete from a dense H-free graph G to obtain a $(\chi(H)-1)$-partite graph is $\Theta(\mathrm{biex}(n, H))$, it seems certain that the multiplicative constants proved for our upper and lower bounds are not best possible. We have made no particular effort to optimise our upper bound: but probably such effort using our techniques would produce only a somewhat less bad upper bound.

It would be interesting to know whether there exists a best possible value for the constant C, and if so, what it is. It seems likely that (despite the result of this paper) the best possible value will depend upon ε.

Acknowledgement

The author would like to thank Daniela Kühn and Deryk Osthus for suggesting this nice problem.

References

[1] N. Alon and B. Sudakov, *H-free graphs of large minimum degree*, Elec. J. Combin. **13** (2006), R19.

[2] B. Andrásfai, P. Erdős, and V. T. Sós, *On the connection between chromatic number, maximal clique and minimal degree of a graph*, Discrete Math. **8** (1974), 205–218.

[3] W. G. Brown, *On graphs that do not contain a Thomsen graph*, Canad. Math. Bull. **9** (1966), 281–285.

[4] P. Erdős, *On extremal problems of graphs and generalized graphs*, Israel J. Math. **2** (1964), 183–190.

[5] P. Erdős and M. Simonovits, *On a valence problem in extremal graph theory*, Discrete Math. **5** (1973), 323–334.

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

[7] J. Kollár, L. Rónyai, and T. Szabo, *Norm-graphs and bipartite Turán numbers*, Combinatorica **16** (1996), 399–406.

[8] T. Kövári, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloquium Math. **3** (1954), 50–57.

[9] I. Reiman, *Über ein problem von K. Zarankiewicz*, Acta. Math. Acad. Sci. Hungar. **9** (1958), 269–279.

[10] E. Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Orsay, 1976), Colloques Internationaux CNRS, vol. 260, CNRS, 1978, pp. 399–401.

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

[12] K. Zarankiewicz, *Sur les relations symétriques dans l'ensemble fini*, Colloq. Math. **1** (1947), 10–14.