Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series B
All Rights Reserved by Academic Press, New York and London

Vol. 45, No. 1, August 1988
Printed in Belgium

How to Make a Graph Bipartite

PAUL ERDÖS,* RALPH FAUDREE,† JÁNOS PACH,*,‡ AND JOEL SPENCER‡

*Mathematical Institute of the Hungarian Academy, Budapest, Hungary,
†Department of Mathematics, Memphis State University, Memphis, Tennessee 38152, and
‡Department of Mathematics, State University of New York, Stony Brook, New York 11790*

Communicated by the Managing Editors

Received June 5, 1986

DEDICATED TO THE MEMORY OF PAUL TURÁN ON THE OCCASION OF THE 10TH ANNIVERSARY OF HIS DEATH

1. INTRODUCTION

Let G be a graph with vertex set $V(G)$ and edge set $E(G)$, respectively. The set of vertices adjacent to an $x \in V(G)$ is denoted by $\Gamma(x)$, and the degree of x is $d(x) = |\Gamma(x)|$. For any subset $V' \subseteq V(G)$, let $G[V']$ denote the subgraph of G *induced* by the vertices of V'. Further, let K_n stand for the complete graph on n vertices.

It is easily seen (e.g., Erdös [7]) that every graph G with n vertices and m edges contains a bipartite subgraph H such that $|E(H)| \geqslant |E(G)|/2 = m/2$, i.e., every graph can be made bipartite by the omission of at most half of its edges. Erdös and Lovász proved that if G has no triangle, then it can be made bipartite by the omission of $m/2 - m^{2/3} (\log m)^c$ edges. On the other hand, Erdös [9] showed by the probability method that for every r, there is a graph G with no cycle of length less than r which cannot be made bipartite by the omission of fewer than $m/2 - m^{1-\varepsilon_r}$ edges. The best exponent in $m^{1-\varepsilon_r}$ is not known even for $r = 3$, but ε_r approaches 0 as r becomes large.

However, the graphs constructed in [9] are "sparse" (i.e., $m = O(n^2)$), and the aim of this paper is to show that much stronger results can be obtained if we assume that our graph G is *not* sparse.

We will restrict our attention to families of graphs not containing some so-called *forbidden* subgraph F. (Such graphs are also said to be F-free.) In particular, for triangle-free graphs, i.e., when $F = K_3$, we will prove the following.

0095-8956/88 $3.00

THEOREM 1. *Every triangle-free graph G with n vertices and m edges can be made bipartite by the omission of at most*

$$\min\left\{\frac{m}{2}-\frac{2m(2m^2-n^3)}{n^2(n^2-2m)},\, m-\frac{4m^2}{n^2}\right\}$$

edges.

THEOREM 2. *There is a (calculatable) constant $\varepsilon>0$ such that every triangle-free graph G with n vertices can be made bipartite by the omission of at most $(1/18-\varepsilon+o(1))\,n^2$ edges.*

According to a long-standing conjecture of Erdös (see [3, 4, 8]), in the last assertion, $(1/18-\varepsilon)\,n^2$ can be replaced by $n^2/25$. This bound, if valid, would be best possible. (It is also conjectured that a K_4-free graph with n vertices can be made bipartite by the omission of $(\frac{1}{9}+o(1))\cdot n^2$ edges. The complete tripartite graph with $n/3$ vertices in each class shows that this conjecture, if true, is also the best possible.)

In the general case, when F can be an arbitrary graph, we have the following result.

THEOREM 3. *For every forbidden graph F and for every $c>0$ there is a constant $\varepsilon(F, c)>0$ such that any F-free graph G with n vertices and $m\geqslant cn^2$ edges can be made bipartite by the omission of at most $(m/2)-\varepsilon(F, c)\,n^2$ edges.*

The proof of the above results is largely based on the fact that triangle-free graphs contain relatively large *induced* bipartite subgraphs. More specifically, we will establish the following.

THEOREM 4. *Let $f=f(n, m)$ denote the maximum integer satisfying the condition that every triangle-free graph with n vertices and at least m edges contains an induced bipartite subgraph with at least f edges. Then*

(i) $\quad \frac{1}{2}m^{1/3}-1\leqslant f(n, m)\leqslant cm^{1/3}\log^2 m \quad$ *if* $\; m<n^{3/2}$,

(ii) $$\frac{4m^2}{n^4}\leqslant f(n, m)\leqslant c\,\frac{m^3}{n^4}\log^2\frac{n^2}{m} \quad \text{if } \; m\geqslant n^{3/2}.$$

In the next section we prove Theorems 1 and 3 and some basic properties of triangle-free graphs. Sections 3 and 4 contain the proofs of Theorems 4 and 2, respectively. In the last section we consider some related questions, generalizations, and unsolved problems.

2. Some Basic Properties of Triangle-Free Graphs

For any $x \in V(G)$ let $\bar{\Gamma}(x) = V(G) - (\{x\} \cup \Gamma(x))$, i.e., the set of those vertices distinct from x which are not connected to x by an edge of G.

Lemma 2.1. *Every triangle-free graph G has a vertex x such that*

$$|E(G[\bar{\Gamma}(x)])| \leqslant |E(G)| - \frac{4\,|E(G)|^2}{|V(G)|^2}.$$

Proof. By a simple averaging argument we obtain

$$\sum_{x \in V(G)} |E(G[\bar{\Gamma}(x)])| = \sum_{ab \in E(G)} (n - d(a) - d(b))$$
$$= n\,|E(G)| - \sum_{a \in V(G)} d^2(a) \leqslant n\,|E(G)| - \frac{4\,|E(G)|^2}{n}. \quad \blacksquare$$

A triangle-free graph is called *saturated* if the addition of any edge results in a graph with a K_3. That is, a triangle-free graph is saturated if and only if its diameter is 2.

Corollary 2.2. *Every triangle-free graph G with n vertices has a vertex x such that $|E(G[\bar{\Gamma}(x)])| \leqslant n^2/16$. Furthermore, for every large n one can find saturated triangle-free graphs G_n with n vertices such that* $\min_{x \in V(G_n)} |E(G_n[\bar{\Gamma}(x)])| = n^2/16 + O(n)$.

Proof. The first assertion follows immediately from Lemma 2.1.

To prove the second one, assume that $n = 4k$ and let H be a $k/4$ regular graph on the vertex set $\{1, 2, ..., k\}$. Define now a graph G_n as follows. Let

$$V(G_n) = \{x_i, y_i, u_i, v_i \colon 1 \leqslant i \leqslant k\}$$
$$E(G_n) = \{x_i y_i, u_i v_i \colon 1 \leqslant i \leqslant k\} \cup \{x_i u_j, y_i v_j \colon ij \in E(H)\}$$
$$\cup \{x_i v_j, y_i u_j \colon ij \notin E(H)\}.$$

It can readily be checked that G_n will meet both requirements. $\blacksquare$

Lemma 2.3. *Let G be a triangle-free graph with m edges and with chromatic number $\chi(G)$. Then $\chi(G) \leqslant 2m^{1/3} + 1$.*

Proof. By double induction on n (the number of vertices of G) and m. If $m = 0$ then the assertion is trivial.

If G has a vertex of degree at most $2m^{1/3}$, then, applying the induction hypothesis to $G - x$, we obtain that $G - x$ can be coloured by at most

$2m^{1/3}+1$ colours, and this colouration can be extended to x without using any new colour.

Assume next that $d(y)>2m^{1/3}$ for every $y\in V(G)$. Then $m=\Sigma\, d(y)/2>nm^{1/3}$. By Lemma 2.1, there is a vertex x in G such that

$$|E(G[\bar{\Gamma}(x)])|<m-\frac{4m^2}{n^2}<m-4m^{2/3}.$$

By the induction hypthesis, $G[\bar{\Gamma}(x)]$ can be coloured with at most $2(m-4m^{2/3})^{1/3}+1$ colours, and using two further colours (one for x and one for $\Gamma(x)$), we get a proper colouration of G. This completes the proof, since

$$2(m-4m^{2/3})^{1/3}+3\leqslant 2m^{1/3}+1. \quad \blacksquare$$

Remark. Although an improvement of Lemma 2.3 is not needed now, it is an interesting problem to try to estimate as exactly as possible the maximum of the chromatic number $f_3(m)$ of a triangle-free graph with m edges. The results in [1] give that for some $1\leqslant\alpha_2\leqslant\alpha_1\leqslant 2$,

$$c_1 m^{1/3}/(\log m)^{\alpha_2}<f_3(m)<c_2 m^{1/3}/(\log m)^{\alpha_1}.$$

The exact determination of $f_3(m)$ is probably hopeless, and even an asymptotic formula for $f_3(m)$ seems out of reach.

As usual, a cycle of length 4 is denoted by C_4 and is called a *quadrilateral.*

LEMMA 2.4. *Let G be a graph with n vertices and m edges. Then*

(i) *G has an edge which is contained in at least $8m^3/n^4-6m/n$ quadrilatrals.*

(ii) *If, in addition, G is triangle-free then it has an edge contained in at least $4m(2m^2-n^3)/n^2(n^2-2m)$ quadrilaterals.*

Proof. For any unordered pair $\{x,y\}$ of distinct vertices, let $t(\{x,y\})$ denote the number of vertices in G joined to both x and y. Then

$$\sum_{\{x,y\}} t(\{x,y\})=\sum_{x\in V(G)}\binom{d(x)}{2}\geqslant n\binom{2m/n}{2}.$$

This is a consequence of the well-known inequality (see, e.g., [3]) that if $d_1+d_2+\cdots+d_2=2m$ and $2m>n$, then

$$\binom{d_1}{2}+\binom{d_2}{2}+\cdots+\binom{d_n}{2}\geqslant n\binom{2m/n}{2}.$$

On the other hand, the total number of quadrilterals in G is

$$\frac{1}{2}\sum_{\{x,y\}}\binom{t(\{x,y\})}{2} \geqslant \frac{1}{2}\binom{n}{2}\binom{\Sigma t(\{x,y\})\Big/\binom{n}{2}}{2}$$

$$\geqslant \frac{1}{2}\binom{n}{2}\binom{n\binom{2m/n}{2}\Big/\binom{n}{2}}{2} \geqslant \frac{2m^4}{n^4}-\frac{3m^2}{2n}.$$

The last inequality follows from direct calculations and the fact that $m \leqslant n^2/2$. Since there exists an edge contained in at least $4/m$ times this many quadrilaterals, we obtain (i).

The proof of part (ii) is entirely similar. The only difference is that in this case the second sum should be taken over all *non-adjacent* pairs $\{x, y\}$. Thus, $\binom{n}{2}$ is to be replaced by $\binom{n}{2}$-m in the previous inequalities. The details are left to the reader. ∎

We shall need the following easy observation.

PROPOSITION 2.5. *Let G be a graph and $W \subseteq V(G)$ be a set of vertices such that the subgraph $G[W]$ induced by them can be made bipartite by the omission of δ edges. Then G can be made bipartite by the omission of at most* $|E(G)|/2 - |E(G[W])|/2 + \delta$ edges.

Proof. For any $X \subseteq V(G)$, put $e(X) = |E(G[X])|$. Let $W = W_1 \cup W_2$ be a partition of W satisfying $e(W_1) + e(W_2) \leqslant \delta$, and let $U = V(G) - W$. Since $G[U]$, just like any other graph, can be made bipartite by the deletion of at most half of its edges, there exists a partition $U = U_1 \cup U_2$ such that $e(U_1) + e(U_2) \leqslant e(U)/2$. Taking into account that

$$\begin{aligned}\sum_{1 \leqslant i,j \leqslant 2} e(U_i \cup W_j) &= e(V(G)) + 2e(U_1) + 2e(U_2) - e(U)\\ &\quad + 2e(W_1) + 2e(W_2) - e(W)\\ &\leqslant e(V(G)) + 2\delta - e(W),\end{aligned}$$

we obtain that either $e(U_1 \cup W_1) + e(U_2 \cup W_2)$ or $e(U_1 \cup W_2) + e(U_2 \cup W_1)$ is at most $e(V(G))/2 + \delta - e(W)/2$. ∎

Now we are in a position to prove Theorems 1 and 3.

Proof of Theorem 1. Let G be a triangle-free graph with n vertices and m edges. By Lemma 2.4, there exists an $xy \in E(G)$ which is contained in at least $(8m^3-4mn^3)/(n^4-2mn^2)$ quadrilaterals. That is, the set $W=\Gamma(x)\cup\Gamma(y)$ induces a bipartite subgraph of G such that $|E(G[W])|\geqslant(8m^3-4mn^3)/(n^4-2mn^2)$. Applying Proposition 2.5 with $\delta=0$, we get the first inequality of the theorem.

The second inequality follows directly from Lemma 2.1. We have to note only that the omission of all edges in $G[\bar{\Gamma}(x)]$ leaves G bipartite. ∎

The following statement, slightly weaker than our Theorem 2, follows immediately from Theorem 1 by considering two cases: $m\geqslant n^2/6$ and $m<n^2/6$, where m is the number of edges of G.

COROLLARY 2.6. *Every triangle-free graph G with n vertices can be made bipartite by the omission of at most $n^2/18+n/2$ edges.*

Proof of Theorem 3. It is obviously enough to prove the theorem in the case when $F=K_r$ $(r>3)$. We are going to show by induction on r that the assertion is true for $F=K_r$ and $\varepsilon(K_r,c)=c^{4^r}$. If $r=3$ then the result follows by Theorem 1.

Assume now that $r>3$, and let G be a K_r-free graph with n vertices and $m\geqslant cn^2$ edges. If n is sufficiently large then, by Lemma 2.4, we can find an edge $x_1x_2\in E(G)$ such that there are at least $8c^3n^2-6cn\geqslant 4c^3n^2$ edges running between $\Gamma(x_1)$ and $\Gamma(x_2)$. Put $e_i=|E(G[\Gamma(x_i)])|$, $i=1,2$.

If $e_1+e_2<2c^3n^2$ then, by Proposition 2.5, G can be made bipartite by the omission of at most

$$\frac{m}{2}-\frac{|E(G[\Gamma(x_1)\cup\Gamma(x_2)])|}{2}+e_1+e_2<\frac{m}{2}-c^3n^2$$

edges, and the result follows.

Suppose now that, say, $e_1\geqslant c^3n^2$. In view of the fact that $G[\Gamma(x_1)]$ does not contain a K_{r-1}, we can apply the induction hypothesis to obtain that $G[\Gamma(x_1)]$ can be made bipartite by the omission of at most

$$\delta=\frac{e_1}{2}-(c^3)^{4^{r-1}}\cdot|\Gamma(x_1)|^2<\frac{e_1}{2}-c^{4^r}\cdot n^2$$

edges. Thus, using Proposition 2.5 with $W=\Gamma(x_1)$, we conclude that G can be made bipartite by the deletion of at most

$$\frac{m}{2}-\frac{e_1}{2}+\delta\leqslant\frac{m}{2}-c^{4^r}n^2$$

edges, as desired. ∎

3. The Size of the Largest Induced Bipartite Subgraph

The aim of this section is to prove Theorem 4. We need some preparation.

Let $G_{n,p}$ denote a random graph of n vertices in which the edges are chosen independently and with probability p. A triangle-free subgraph H of a graph G is called *maximal* if the addition of any edge in $G-H$ results in a graph with a triangle. A quarter of a century ago Erdös [5, 6] found the following result, which provides fairly good lower bounds for some Ramsey numbers.

Theorem 3.1. *If $p=\frac{1}{2}n^{-1/2}$ then, with probability tending to* 1, *no maximal triangle-free subgraph of $G_{n,p}$ contains an independent set of more than $3n^{1/2}\log n$ vertices.*

We shall make use of the following

Lemma 3.2. *If $p=\frac{1}{2}n^{-1/2}$ then, with probability tending to* 1, *every maximal triangle-free subgraph $H\subseteq G_{n,p}$ has the following two properties.*

(i) $\frac{1}{5}n^{3/2}<|E(H)|<\frac{2}{5}n^{3/2}$;

(ii) *H does not contain an induced bipartite subgraph with more than $30n^{1/2}\log^2 n$ edges.*

Proof. The expected number of edges and triangles in $G_{n,p}$ is equal to $p\binom{n}{2}\cong\frac{1}{4}n^{3/2}$ and $p^3\binom{n}{3}\cong\frac{1}{48}n^{3/2}$, respectively. Observe that if H is a maximal triangle-free subgraph of $G_{n,p}$ then

$$|E(G_{n,p})|-\#(\text{triangles in } G_{n,p})\leqslant|E(H)|\leqslant|E(G_{n,p})|,$$

whence (i) follows by a routine application of the Chernoff Inequality for the tail of the binomial distribution (cf. [2, 12]).

In view of Theorem 3.1, to prove (ii) it is sufficient to show that the probability that there are two disjoint subsets A, $B\subseteq V(G_{n,p})$ such that $|A|=|B|=3n^{1/2}\log n=\lambda$ and there are at least $30n^{1/2}\log^2 n$ edges running between them tends to 0. But this probability is clearly at most

$$\binom{n}{\lambda}^2 \operatorname{Prob}\{S_{\lambda^2,p}>30n^{1/2}\log^2 n\}$$
$$<\exp(6n^{1/2}\log^2 n)\exp(-7n^{1/2}\log^2 n)\to 0,$$

where $S_{\lambda^2,p}$ denotes the number of edges connecting two fixed disjoint subsets A and B of size λ, which is a random variable of a binomial distribution with parameters λ^2 and p. ∎

The following assertion is trivial.

PROPOSITION 3.3. *Given a triangle-free graph H which does not contain an induced bipartite subgraph with more than t edges, let $H(k)$ denote the graph obtained from H replacing each vertex $x \in V(H)$ with an independent set V_x of size k and joining two vertices $x' \in V_x$ and $y' \in V_y$ by an edge if and only if $xy \in E(H)$. Then*

(i) *$H(k)$ is triangle-free;*

(ii) *$H(k)$ does not contain an induced bipartite subgraph of more than tk^2 edges.*

Proof of Theorem 4. First we establish the upper bounds.

For any natural number $\bar{n}$, set $\bar{p} = \frac{1}{2}\bar{n}^{(-1/2)}$, and let $H_{\bar{n}}$ denote a maximal triangle-free subgraph of $G_{\bar{n},\bar{p}}$ having the two properties in Lemma 3.2.

Assume first that $m < \frac{1}{5}n^{3/2}$, and let r denote the smallest integer such that $|E(H_r)| \geqslant m$. Let G be a graph of n vertices obtained from H_r by the addition of $n-r$ isolated vertices. Then $\frac{1}{5}r^{3/2} \leqslant m \leqslant |E(H_r)| < \frac{2}{5}r^{3/2}$, and G does not contain an induced bipartite subgraph with more than

$$30r^{1/2}\log^2 r < c'm^{1/3}\log^2 m$$

edges.

If $m \geqslant \frac{1}{5}n^{3/2}$ then let $k \geqslant 1$ be the smallest integer for which $H_{n/k}(k)$, (i.e., the graph obtained from $H_{n/k}$ by replacing each vertex with an independent set of size k) has at least m edges. Obviously,

$$\tfrac{1}{5}n^{3/2}(k-1)^{1/2} < m < \tfrac{2}{5}n^{3/2}k^{1/2}$$

and, by Lemma 3.2, $H_{n/k}(k)$ does not contain an induced bipartite subgraph with more than

$$k^2 30\left(\frac{n}{k}\right)^{1/2}\log^2\left(\frac{n}{k}\right) < 2\cdot 10^4\cdot\frac{m^3}{n^4}\log^2\left(\frac{n^2}{m}\right)$$

edges. This completes the proof of the upper bounds.

Every graph G with m edges and chromatic number $\chi(G)$ splits up into $\binom{\chi(G)}{2}$ induced bipartite subgraphs. Therefore, if G is triangle-free, then by Lemma 2.3 it contains an induced bipartite subgraph of at least $m/\binom{2m^{1/3}+1}{2} > m^{1/3}/2-1$ edges, which proves the lower bound in (i).

The lower bound in (ii) follows immediately from Lemma 2.4(ii). ∎

4. PROOF OF THEOREM 2

Assume, in order to obtain a contradiction, that there is a triangle-free graph G which requires the removal of at least $n^2/18 + o(n^2)$ edges to be made bipartite. By Theorem 1, G must have $m=n^2/6+o(n^2)$ edges. From the proof of Lemma 2.4, $\Sigma\binom{d(x)}{2}=n\binom{n/3}{2}+o(n^3)$ so that $d(x)=n/3+o(n)$ for all but $o(n)$ vertices, which we shall ignore. Fix a vertex x, and let $S=\Gamma(x)$, $T=\overline{\Gamma(x)\cup\{x\}}$ so that $|S|=n/3+o(n)$, $|T|=2n/3+o(n)$. As S is independent, $S\times T$, the bipartite subgraph of G induced by the sets S and T, has $n^2/9+O(n^2)$ edges. Suppose $d(y, T)>n/6+\varepsilon_1 n$ for $\varepsilon_2 n$ vertices $y\in T$. (Here $d(y, T)$ is the number of edges between y and vertices of T.) Move these vertices from T to S, forming S^*, T^*. Each vertex moving to S gives at least $2\varepsilon_1 n$ extra crossing edges, minus the at most $\varepsilon_2^2 n^2$ edges $\{y, y'\}$, where y, y' are both moved. Then $S^*\times T^*$ would have at least $n^2/q+\varepsilon_1\varepsilon_2 n^2-\varepsilon_2^2 n^2+o(n^2)$ edges. Replacing ε_2 by $\min[\varepsilon_2, \varepsilon_1/2]$, $S^*\times T^*$ would have $n^2/9+cn^2$ edges and G could be made bipartite by the deletion of only $n^2/18-cn^2$ edges, a contradiction. Hence $d(y, T)<n/6+o(n)$ for all but $o(n)$ vertices $y\in T$. As $S\times T$ has $n^2/9+o(n^2)$ edges, $d(y, T)=n/6+o(n)$ for all but $o(n)$ vertices $y\in T$. Again we ignore these $o(n)$ vertices.

Fix an edge $\{y, z\}\in E(G)$ with $y, z\in T$. Set $S_1=\Gamma(y)\cap S$, $S_2=\Gamma(z)\cap S$. Then $S_1\cap S_2=\varnothing$ as G is triangle-free and $|S_1|=n/6+o(n)=|S_2|$. Let $Y=\Gamma(z)\cap T$, $Z=\Gamma(y)\cap T$ so that $|Y|=n/6+o(n)=|Z|$ and $Y\cap Z=\varnothing$. For each $y'\in Y$, $(\Gamma(y')\cap S)\cap S_2=\varnothing$ and $|\Gamma(y')\cap S|=n/6+o(n)$ so $|(\Gamma(y')\cap S)\,\Delta S_1|=o(n)$. Similarly, $|(\Gamma(z')\cap S)\,\Delta S_2|=o(n)$ for each $z'\in Z$. Hence $S_1\cup Z$ has $o(n^2)$ edges, as does $S_2\cup Y$.

Suppose $Y\times Z$ had εn^2 edges. Then $S_1\cup S_2\cup Y\cup Z$ would have $n^2[1/18+\varepsilon+o(1)]$ edges, all but $o(n^2)$ of which were in $(S_1\cup Z)\times (S_2\cup Y)$. We extend to a partition of $V(G)$ so that at most half of the remaining edges are not crossing; i.e., at most $n^2[(\frac{1}{9}-\varepsilon)/2+o(1)]$ edges, a contradiction if ε is bounded from below. Hence $Y\times Z$ has $o(n^2)$ edges.

Pick $y'\in Y$ with $d(y', Z)=o(n)$ and set $Z'=\Gamma(y')-(S_1\cup S_2\cup Y\cup Z)$ so that $|Z'|=n/6+o(n)$. Then, as before, $|(\Gamma(z')\cap S)\,\Delta S_2|=o(n)$ for each $z'\in Z'$. Then $S_2\times Z'$ has $n^2/36+o(n^2)$ edges. Let Y' be the remaining points of T. Then $|Y'|=n/6+o(n)$ and so $Y'\times S$ has $n^2/36+o(n^2)$ edges. But $S_2\times Y'$ has only $o(n^2)$ edges so $S_1\times Y'$ has $n^2/36+o(n^2)$ edges.

Now G is nearly bipartite. All but $o(n)$ vertices may be partitioned into $S_1\cup Z\cup Z'$ and $S_2\cup Y\cup Y'$, both of which have $o(n^2)$ edges. This contradiction implies the claim.

5. Generalizations and Open Problems

Let $p\geqslant 2$ be a natural number. Then every graph G may be made p-partite with the removal of at most $|E(G)|/p$ edges. We also have the following straightforward generalization of Proposition 2.5.

Proposition 5.1. *Let G be a graph and $W\subseteq V(G)$ be a set of vertices such that the subgraph $G[W]$ induced by them may be made p-partite with* *the omission of δ edges. Then G can be made p-partite with the omission of at most $|E(G)/p-|E(G[W])|/p+\delta$ edges.*

Theorem 3 can now be generalized in the following way.

THEOREM 3'. *Let $p\geqslant 2$, $r\geqslant 3$ be natural numbers, $0<c<\frac{1}{2}$. Then there exists a constant $\varepsilon(p,r,c)>0$ such that any K_r-free graph with n vertices and $m\geqslant cn^2$ edges may be made p-partite with the omission of at most $m/p-\varepsilon(p,r,c)\,n^2$ edges.*

Proof. We only outline the proof, which is entirely similar to that of Theorem 3. Let $p>2$ be fixed. We will show by induction on r that the assertion is true with $\varepsilon(p,r,c)=c^{4^r}/p$. If $r=3$ then the result follows from Lemma 2.4 and Proposition 5.1. Assume now that $r>3$ and let G, x_i, e_i denote the same as in the proof of Theorem 3. $G[\Gamma(x_i)]$ can be made $\lfloor p/2\rfloor$-partite with the removal of $e_i/\lfloor p/2\rfloor$ edges, hence, by Proposition 5.1, G may be made p-partite with the omission of at most

$$\frac{m}{p}-\frac{4c^3n^2+e_1+e_2}{p}+\frac{e_1+e_2}{\lfloor p/2\rfloor}<\frac{m}{p}-\frac{c^3n^2}{p}$$

edges, provided that $e_1+e_2<2c^3n^2$. If, say, $e_1\geqslant c^3n^2$ then using the induction hypothesis we obtain that $G[\Gamma(x_1)]$ may be made p-partite with the removal of $\delta=e_1/p-\varepsilon(p,r-1,c^3)|\Gamma(x_1)|^2$ edges, and we are done by Proposition 5.1. ∎

In Section 3 we have proved that triangle-free graphs contain relatively large induced bipartite subgraphs. Similarly, one can ask the following question. Given a natural number $r>3$ and a real c $(0<c<\frac{1}{4})$, what is the maximal integer $f_{r,c}(n)=f$ such that every K_r-free graph with n vertices and at least cn^2 edges contains an induced bipartite subgraph with at least f edges. We are unable to prove asymptotically tight bounds for $f_{r,c}(n)$ even if $r=4$. Our only results in this direction can be summarized, as follows.

PROPOSITION 5.2. *Let $f_{r,c}(n)$ denote the same as above. Then there exist two constants $\lambda_1,\lambda_2>0$ depending only on r and c such that*

(i) $\lambda_1 n\log n<f_{4,c}(n)<\lambda_2 n^{3/2}\log n$ *if* $r=4$;

(ii) $\lambda_1 n^{2/(r-2)}(\log n)^{2-2/(r-2)}<f_{r,c}(n)<\lambda_2 n^{8(r+3)/(r+2)(r+4)}\log^2 n$ *if* $r>4$ *is even*;

(iii) $\lambda_1 n^{2/(r-2)}(\log n)^{2-2/(r-2)}<f_{r,c}(n)<\lambda_2 n^{8/(r+3)}\log^2 n$ *if* $r>4$ *is odd.*

Proof. Let G be a K_r-free graph with n vertices and at least cn^2 edges. By Lemma 2.4, we can choose an edge $x_1x_2\in E(G)$ such that there are at least $4c^3n^2$ edges between $\Gamma(x_1)$ and $\Gamma(x_2)$. But $G[\Gamma(x_i)]$ is K_{r-1}-free, so by an easy corollary to a well-known theorem of Ajtai, Komlós, and Szemerédi [1], $\chi(G[\Gamma(x_i)]) < \mu(n/\log n)^{1-1/(r-2)}$, i = 1, 2. Hence $G[\Gamma(x_1) \cup \Gamma(x_2)]$ splits up into $\chi(G[\Gamma(x_1)])\,\chi(G[\Gamma(x_2)])$ induced bipartite subgraphs, and at least one of them must have at least

$$\frac{4c^3n^2}{\chi(G[\Gamma(x_1)])\,\chi(G[\Gamma(x_2)])} \geqslant \lambda_1 n^{2/(r-2)}(\log n)^{2-2/(r-2)}$$

edges.

The upper bound can be established by the following construction. Let $r > 4$, and let $V(G)$ be divided into two equal classes V_1 and V_2. Let any pair of points in different classes be joined by an edge, and let V_i induce a K_{r_i}-free subgraph in G containing no independent set of size $\lambda' n^{2/(r_i+1)} \log n$, where $r_1 = \lfloor (r+1)/2 \rfloor$ and $r_2 = \lceil (r+1)/2 \rceil$. The existence of such graphs was proved by Spencer [14]. Obviously, every induced bipartite subgraph of G has at most

$$\lambda'^2 n^{2/(r_1+1)+2/(r_2+1)} \log^2 n$$

edges, which gives the upper bound if $r > 4$. The case $r = 4$ can be treated similarly. ∎

We end this paper by answering the following question of Füredi. Characterize the class of those graphs F which have the property that any F-free graph with n vertices and cn^2 edges has an induced bipartite subgraph with at least $\varepsilon_c n^2$ edges.

THEOREM 5.3. *Let F be a graph whose vertex set can be split into two disjoint parts A and B such that $F[A]$ is empty and $F[B]$ is a forest. Then, any F-free graph G with n vertices and cn^2 edges has an induced bipartite subgraph with at least εn^2 edges ($\varepsilon = \varepsilon(F, c) > 0$). Moreover, no other graphs have this property.*

Proof. Assume that G has cn^2 edges but no induced bipartite subgraph with εn^2 edges. With no loss of generality we can assume that each vertex of G has degree at least $cn/2$, because the deletion of vertices with smaller degree leaves a non-trivial graph of minimal degree at least $cn/2$.

Partition the vertices of G into sets R and S such that the number of edges between R and S is a maximum. Then, each vertex of R (or S) has at least $cn/4$ neighborhoods in S (or R). If not, the maximality of edges between R and S would be contradicted by moving a vertex to the other side.

For an appropriate $\delta = \delta(c)$ select a maximum number of vertex disjoint independent subsets $R_1, R_2, ..., R_p$ such that $|R_i| \geqslant \delta n$. Let R' be the remaining vertices of R. In the same manner select vertices $S_1, S_2, ..., S_q$ of S with corresponding set S'. By assumption there are less than εpqn^2 edges between the R_i's and the S_j's. Thus, with no loss of generality, we can assume that there are at least $c'n^2$ edges between R and S' (or equivalently between R' and S) for some $c' = c'(c) > 0$. In addition, S' contains no independent set of order δn.

If $k = |V(F)|$, then there are k vertices in R, say $x_1, x_2, ..., x_k$, such that if $S'' = \Gamma(x_1) \cap \cdots \cap \Gamma(x_n) \cap S'$, then $|S''| > c''n$ for some $c'' = c''(c) > 0$ (see [7]). If $|E(S'')| \geqslant 4k\,|S''|$, then $G[S'']$ contains all trees on k vertices and G contains F. On the other hand, if $|E(S'')| < 4k\,|S''|$, then S'' contains an independent set of order at least $|S''|/8k > \delta n$. This contradiction completes the proof that G contains a bipartite subgraph with n^2 edges.

To verify the last statement of the theorem consider the following graph G on n vertices. The vertices are partitioned into two equal parts R and S with all edges between R and S in G. The vertices of R are independent, and the graph $G[S]$ has no cycles of length as small as $|V(F)|$ and no independent set with more than $O(n^{1/2}\log(n))$ vertices (see [6]). If F is not of the required type, then clearly F is not in G. Also, G contains no induced bipartite graph with εn^2 edges. ∎

A. Hajnal pointed out that Theorem 5.3 can also be deduced by using arguments in [11].

Some other related problems can be found in Hedetniemi, Laskar, and Peters [13] and Erdös and T. Sós [10].

Acknowledgments

We have learned from Adrian Bondy that he also proved Corollary 2.6 in the special case when the graph G is regular. Our thanks are due to Fan Chung, A. Hajnal, Renu Laskar, Vojtech Rödl, Cecil Rousseau, Dick Schelp, and Miklós Simonovits for their valuable remarks.

References

1. M. Ajtai, J. Komlós, and E. Szemerédi, A dense infinite Sidon sequence, *Eur. J. Combin.* **2** (1981), 1–11.
2. B. Bollobás, "Random Graphs," Academic Press, London/New York, 1985.
3. B. Bollobás, "Extremal Graph Theory," p. 363, Academic Press, London/New York, 1978.
4. J. A. Bondy and U. S. R. Murty, "Graph Theory with Applications," Amer. Elsevier, New York, 1976.
5. P. Erdös, Graph theory and probability, II, *Canad. J. Math.* **13** (1961), 346–352.
6. P. Erdös, On circuits and subgraphs of chromatic graphs, *Mathematika* **9** (1962), 170–175.
7. P. Erdös, On some extremal problems in graph theory, *Israel J. Math.* **3** (1965), 113–116.

8. P. ERDÖS, Some new applications of the probability methods in combinatorial analysis and graph theory, *in* "Proceedings, Fifth S.-E. Conference on Combinatorics, Graph Theory and Computing," pp. 39–51, Utilitas Mathematica, Winnipeg, 1974.
9. P. ERDÖS, On bipartite subgraphs of graphs, *Math. Lapok.* **18** (1967), 283–288.
10. P. ERDÖS AND V. T. SÓS, Some remarks on Ramsey's and Turán's theorem, *in* "Combinatorial Theory and Its Applications" (P. Erdös *et al.*, Eds.), pp. 395–404, Colloq. Math. Soc. János Bolyai, Vol. 4, North-Holland, Amsterdam, 1970.
11. P. ERDÖS, A. HAJNAL, V. T. SOS, AND E. SZEMERÉDI, More results on Ramsey–Turan type problems, *Combinatorica* **3** (1983), 69–82.
12. P. ERDÖS AND J. SPENCER, "Probabilistic Methods in Combinatorics," Academic Press, New York, and Akad. Kiadó, Budapest, 1974.
13. S. T. HEDETNIEMI, R. LASKAR, AND K. PETERS, "On the Strength of a Graph," Technical Report No. 498, Department of Mathematics, Clemson University, Clemson, S. C, 1985.
14. J. SPENCER, Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20** (1977), 69–76.