Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Combinatorial Theory, Series B 96 (2006) 614–620
Journal of Combinatorial Theory
Series B
www.elsevier.com/locate/jctb

Note

Sparse halves in triangle-free graphs ☆

Peter Keevash [a], Benny Sudakov [b]

[a] *Department of Mathematics, Caltech, Pasadena, CA 91125, USA*
[b] *Department of Mathematics, Princeton University, Princeton, NJ 08544, USA*

Received 21 March 2005
Available online 5 January 2006

Abstract

One of Erdős' favourite conjectures was that any triangle-free graph G on n vertices should contain a set of $n/2$ vertices that spans at most $n^2/50$ edges. We prove this when the number of edges in G is either at most $n^2/12$ or at least $n^2/5$.

Keywords: Extremal graph theory; Triangle-free graphs; Local density

1. Introduction

A fundamental result in extremal graph theory is Turán's theorem from 1941, that the unique largest graph on n vertices not containing a copy of K_t (the complete graph on t vertices) is the Turán graph $T_{t-1}(n)$, which is the complete $(t-1)$-partite graph with part sizes as equal as possible. (The case $t = 3$ was proved by Mantel in 1907.) A generalisation that takes into account edge distribution, or local density, was introduced by Erdős [2] who asked the following question. Suppose $0 \leqslant \alpha, \beta \leqslant 1$ and that G is a K_t-free graph on n vertices in which every set of αn vertices span at least βn^2 edges. How large can β be as a function of α? Erdős, Faudree, Rousseau and Schelp [5] studied this problem and conjectured that there is a constant $c_t < 1$ so that if $c_t \leqslant \alpha \leqslant 1$ then the largest possible β is $\frac{t-2}{t-1}(\alpha - 1/2)$ (which is attained by the Turán graph $T_{t-1}(n)$). They proved this for triangle-free graphs ($t = 3$) and the general case was proved by the authors [7].

☆ Research supported in part by NSF grant DMS-0355497, USA–Israeli BSF grant, and by an Alfred P. Sloan fellowship.

E-mail addresses: keevash@caltech.edu (P. Keevash), bsudakov@math.princeton.edu (B. Sudakov).

0095-8956/$ – see front matter
doi:10.1016/j.jctb.2005.11.003

Moreover, for triangle-free graphs and general α it was conjectured in [5] that β is determined by a family of extremal triangle-free graphs. Besides the complete bipartite graph $T_2(n)$ already mentioned, another important graph is $C_5(n/5)$, which is obtained from a 5-cycle by replacing each vertex i by an independent set V_i of size $n/5$ (assuming for simplicity that n is divisible by 5), and each edge ij by a complete bipartite graph joining V_i and V_j (this operation is called a 'blow-up'). It is not difficult to see that a set S of αn edges spanning as few edges as possible will either contain V_i or be disjoint from V_i for all but at most one i; indeed, if S intersects V_i and V_j non-trivially it is possible to increase one intersection and decrease the other without increasing the number of edges spanned by S. For $2/5 \leqslant \alpha \leqslant 3/5$ it follows that every αn vertices in $C_5(n/5)$ span at least $\frac{5\alpha-2}{25}n^2$ edges. This is larger than $\frac{2\alpha-1}{4}n^2$ (the value for $T_2(n)$) when $\alpha < 17/30$. Erdős et al. conjectured that above this value the largest β is always $\frac{2\alpha-1}{4}$, i.e., the constant c_3 (defined above) can be taken equal to 17/30. The best-known bound for this problem is due to Krivelevich [8], who showed that one can take $c_3 < 3/5$. They also conjectured that β is $\frac{5\alpha-2}{25}$ for a certain range of α below 17/30, including $\alpha = 1/2$.

The case $\alpha = 1/2$ is an old question of Erdős that he returned to often in his problems papers, starting with [2], through to [3] where he offerred a \$ 250 prize for its solution. Here the conjecture is that any triangle-free graph on n vertices should contain a set of $n/2$ vertices that spans at most $n^2/50$ edges. Krivelevich [8] has shown that this holds when $n^2/50$ is replaced by $n^2/36$. In this paper we prove the following result, which establishes the conjecture under an additional assumption on the total number of edges in the graph, and shows that $C_5(n/5)$ is the unique extremal example in the range that we consider.

Theorem 1.1. *Let G be a triangle-free graph on n vertices with at least $n^2/5$ edges such that every set of $\lfloor n/2 \rfloor$ vertices of G spans at least $n^2/50$ edges. Then $n = 10m$ for some integer m and $G = C_5(2m)$.*

Also, it is not difficult to obtain an analogous result for graphs with few edges.

Proposition 1.2. *Let G be a triangle-free graph on n vertices with at most $n^2/12$ edges. Then some set of $\lfloor n/2 \rfloor$ vertices of G spans at most $n^2/50$ edges.*

Another problem, that is similar in spirit, is to determine how many edges one may need to delete from a triangle-free graph on n vertices in order to make it bipartite. A long-standing conjecture of Erdős [2] is that at most $n^2/25$ edges need to be deleted, and $C_5(n/5)$ shows that this would be best possible. A related conjecture, that any K_4-free graph on n vertices can be made bipartite with the omission of at most $n^2/9$ edges was proved recently by the second author [9]. For triangle-free graphs, the best-known bound is $(1/18 - \epsilon)n^2$ for some calculable constant $\epsilon > 0$, obtained by Erdős, Faudree, Pach and Spencer [4].

Krivelevich [8] noticed that for regular graphs a bound in the local density problem implies a bound for the problem of making the graph bipartite. Indeed, suppose n is even, G is a d-regular graph on n vertices and S is a set of $n/2$ vertices. Then $dn/2 = \sum_{s \in S} d(s) = 2e(S) + e(S, \bar{S})$ and $dn/2 = \sum_{s \notin S} d(s) = 2e(\bar{S}) + e(S, \bar{S})$, i.e., $e(S) = e(\bar{S})$. Deleting the $2e(S)$ edges within S or $\bar{S}$ makes the graph bipartite, so if we could find S spanning at most $n^2/50$ edges we would delete at most $n^2/25$ in making G bipartite. The converse reasoning does not work, as may be seen from considering the blow-up $P(n/10)$ (supposing n is divisible by 10), where P is the Petersen graph (see Fig. 1). In this graph every set of $n/2$ vertices spans at least $n^2/50$ edges (illustrated by the black circles), but it can be made bipartite by deleting only $3n^2/100$ edges (illustrated by

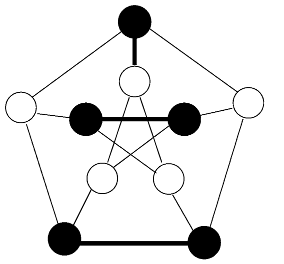

Fig. 1. The Petersen blow-up $P(n/10)$.

the bold lines). This seems to indicate that the local density problem may be harder, especially since it appears that there is not a unique extremal example. Noting that $P(n/10)$ has $3n^2/20$ edges, it seems interesting to extend our bound of $n^2/12$ in Proposition 1.2 to $3n^2/20$. We expect that graphs with between $3n^2/20$ and $n^2/5$ edges will be the most challenging to deal with.

It is not difficult to show (see [4]) that if G is a triangle-free graph on n vertices with $e \geqslant n^2/5$ edges then it can be made bipartite by deleting at most $n^2/25$ edges. Some extensions of this result can be found in [6]. With Theorem 1.1 we solve the local density problem in the same range.

Notation. Suppose G is a graph. For a vertex v we let $N(v)$ denote its neighbourhood and $d(v) = |N(v)|$ its degree. If X is a set of vertices then $G[X]$ is the restriction of G to X, i.e., the graph with vertex set X whose edges are edges of G with both endpoints in X. We write $e(X) = e(G[X])$ for the number of edges in X. If X and Y are sets of vertices then $e(X, Y)$ is the number of edges with one endpoint in X and the other in Y. For an integer t, the blow-up $G(t)$ is the graph obtained from G by replacing each vertex i by an independent set V_i of size t, and each edge ij by a complete bipartite graph joining V_i and V_j.

2. Proof of Proposition 1.2

Throughout this note we will repeatedly use the two following standard averaging arguments. Suppose we have a graph containing a set X of x vertices, such that there are e_1 edges with exactly one endpoint in X and e_2 edges with both endpoints in X. Then for any integer y with $0 \leqslant y \leqslant x$, by considering a random y-subset of X we see that:

(i) there is a subset of X of size y containing at most $(y/x)^2 e_2$ edges, and
(ii) there is a subset Y of X of size y so that the number of edges incident to Y, i.e., having at least one endpoint in Y, is at most $(y/x)e_1 + (y/x)^2 e_2 \leqslant (y/x)(e_1 + e_2)$.

Suppose G is a triangle-free graph on n vertices and θn^2 edges. As a warm-up, we note that the case $\theta \leqslant 2/25$ is easy to deal with: by averaging there is a set of $n/2$ vertices spanning at most $(1/2)^2 \theta n^2 \leqslant n^2/50$ edges. With a little more work we can deal with the range $\theta \leqslant 1/12$ as follows. Let I be an independent set of size $2\theta n$ (we can choose such a set inside the neighbourhood of a vertex of maximum degree). Let $I' = V(G) - I$ and suppose that there are ϕn^2 edges in $G[I']$. By averaging some set of $n/2$ vertices in I' spans at most $(\frac{1/2}{1-2\theta})^2 \phi n^2 \geqslant n^2/50$ edges, so we can assume $\phi \geqslant \frac{2}{25}(1 - 2\theta)^2$. Now by averaging there is a subset $J \subseteq I'$ of $(1/2 - 2\theta)n$ vertices so that

$$\begin{aligned} n^{-2}e(I\cup J) &\leqslant n^{-2}\left(\frac{1/2-2\theta}{1-2\theta}\big(e(G)-e(I')\big)+\left(\frac{1/2-2\theta}{1-2\theta}\right)^2 e(I')\right)\\ &= \frac{1/2-2\theta}{1-2\theta}(\theta-\phi)+\left(\frac{1/2-2\theta}{1-2\theta}\right)^2\phi\\ &= \frac{1/2-2\theta}{1-2\theta}\theta-\frac{1/2-2\theta}{2(1-2\theta)^2}\phi\\ &\leqslant \frac{1/2-2\theta}{1-2\theta}\theta-(1/2-2\theta)/25\leqslant 1/50. \end{aligned}$$

Here the last inequality follows from the fact that the function $f(t)=\frac{1/2-2t}{1-2t}t-(1/2-2t)/25$ is increasing when $t\leqslant 1/12$.

3. Proof of Theorem 1.1

We will need a result of Andrásfai, Erdős and Sós [1], which states that if G is a triangle-free graph on n vertices with minimum degree at least $2n/5$ then either G is bipartite or $G=C_5(n/5)$ is the blow-up of a 5-cycle.[1]

Now let G be a triangle-free graph on n vertices with at least $n^2/5$ edges such that every set of $\lfloor n/2\rfloor$ vertices of G spans at least $n^2/50$ edges. We will show that n is divisible by 10 and $G=C_5(n/5)$.

First suppose that we have proved the theorem under the assumption that n is divisible by 10, and consider the case when n is not divisible by 10. Then every set of $\lfloor n/2\rfloor$ vertices of G spans at least $\lceil n^2/50\rceil > n^2/50$ edges, since $n^2/50$ is not an integer. Consider the blow-up $H=G(10)$, obtained by replacing each vertex i of G by a set V_i of size 10. Then H has $h=10n$ vertices and at least $h^2/5$ edges. As noted in the introduction, it is not difficult to see that there is a set S of $h/2$ vertices spanning the minimal number of edges which either contains or is disjoint from the set V_i for all but at most one index i. Then S contains $\lfloor n/2\rfloor$ sets V_i, so it spans at least $100\lceil n^2/50\rceil > h^2/50$ edges. Applying the theorem to H we see that $H=C_5(2n)$. But then H has a set of $h/2$ vertices spanning $h^2/50$ edges, a contradiction. Therefore we can assume that n is divisible by 10.

Next we mention two simple consequences of the fact that G is triangle-free. One is that for any vertex v the neighbourhood $N(v)$ is an independent set. Another is that any set of vertices U spans at most $|U|^2/4$ edges, by Mantel's theorem.

We start the argument with the following useful observation. Suppose that G contains an independent set I of size $(2/5+t)n$ for some $t>0$, and that there is a vertex v with at least $(1/5-t)n$ neighbours in I. Then v has less than $(1/10-t)n$ neighbours outside I. For otherwise we can add a set of size $(1/10-t)n$ from $N(x)\backslash I$ to obtain a set of size $n/2$ in which the only edges are those going between $N(x)\backslash I$ and $I\backslash N(x)$.[2] The number of such edges is at most $(1/10-t)n\cdot(1/5+2t)n=(1/50-2t^2)n^2<n^2/50$, a contradiction.

Suppose that the maximum degree of G is $(2/5+t)n$, where $t\geqslant 0$. We divide the proof into two cases according to the value of t.

[1] Note that this theorem immediately implies our theorem when restricted to regular graphs, i.e., any regular triangle-free graph on n vertices with at least $n^2/5$ edges contains a set of $n/2$ vertices that spans at most $n^2/50$ edges.

[2] Since n is divisible by 10 and $|I|=(2/5+t)n$ is an integer we see that tn is an integer, and so $(1/10-t)n$ is an integer. Similar comments apply throughout the proof, but we will not labour the point.

Case 1. $t \geqslant 1/135$.

Let A be the neighbourhood of a vertex of maximum degree, so $|A| = (2/5 + t)n$. Since A is an independent set, we certainly have $t < 1/10$. Let B be the set of vertices with at least $(1/5 - t)n$ neighbours in A, and let C be the set of vertices not in A or B. By the previous observation every vertex in B has less than $(1/10 - t)n$ neighbours outside A.

By definition every vertex in C has less than $(1/5 - t)n$ neighbours in A. We claim that $|C| < (1/10 - t)n$. Otherwise let $X \subseteq C$ has size $(1/10 - t)n$ and considers $A \cup X$, which has size $n/2$. Now

$$\begin{aligned} n^2/50 \leqslant e(A \cup X) &< |X| \cdot (1/5 - t)n + |X|^2/4 \\ &= (1/10 - t)(1/5 - t)n^2 + (1/10 - t)^2 n^2/4 \\ &= n^2/50 + \big(5t^2/4 - 7t/20 + 1/400\big)n^2, \end{aligned}$$

which gives $5t^2/4 - 7t/20 + 1/400 > 0$. Solving the quadratic and recalling that $t < 1/10$ we obtain $t < (7 - 2\sqrt{11})/50 < 1/135$, a contradiction.

It follows that $|B| = n - |A| - |C| > n/2$. Consider $Y \subseteq B$ of size $n/2$. Recalling that every vertex in B has less than $(1/10 - t)n$ neighbours outside A we have $n^2/50 \leqslant e(Y) < \frac{1}{2} \cdot n/2 \cdot (1/10 - t)n$, so $t < 1/50$.

Let D be the set of vertices $v \in B$ with at most $(1/5 + 2t)n$ neighbours in A. Vertices in B have less than $(1/10 - t)n$ neighbours outside A, so vertices in D have degree less than $(3/10 + t)n$. Recall that the maximum degree of G is $(2/5 + t)n$ and let $|D| = pn$. Now we have

$$\begin{aligned} 2/5 \leqslant 2e(G)/n^2 = n^{-2} \sum_v d(v) &\leqslant n^{-2}\big(|D|(3/10 + t)n + \big(n - |D|\big)(2/5 + t)n\big) \\ &= p(3/10 + t) + (1 - p)(2/5 + t) = 2/5 + t - p/10, \end{aligned}$$

so $p \leqslant 10t$. Now we note that the only edges of G within B are those internal to D. Indeed, in any other pair of vertices, both have at least $(1/5 - t)n$ neighbours in A, and one vertex has more than $(1/5 + 2t)$ neighbours in A. Since $|A| = (2/5 + t)n$ they must have a common neighbour in A. Therefore these vertices are non-adjacent, as G is triangle-free. This implies that any set of $n/2$ vertices in B spans at most $e(D) \leqslant |D|^2/4 \leqslant 25t^2n^2$ edges. Since $t \leqslant 1/50$ this is less than $n^2/50$, so we have a contradiction. This completes the analysis of the first case.

Case 2. $t < 1/135$.

Let $e \geqslant n^2/5$ be the number of edges in G. By the Cauchy–Schwartz inequality, we have

$$\frac{1}{n} \sum_v \sum_{u \in N(v)} d(u) = \frac{1}{n} \sum_v d(v)^2 \geqslant \left(\frac{\sum_v d(v)}{n}\right)^2 = 4e^2/n^2.$$

Moreover equality holds only if all degrees in G are exactly $2e/n \geqslant 2n/5$. In this case we can apply the theorem of Andrásfai, Erdős and Sós mentioned earlier. G cannot be bipartite, as then one of its parts would have size at least $n/2$ and contain no edges at all! It follows that $G = C_5(n/5)$, as required. We can assume then that equality does not hold, so there is a vertex v for which $\sum_{u \in N(v)} d(u) > 4e^2/n^2$. Let $A = N(v)$ (an independent set) and suppose $|A| = (2/5 + s)n$, where $s \leqslant t < 1/135$.

There is some vertex $a \in A$ with $d(a) \geqslant (2/5 - s)n$. Otherwise we would have

$$\sum_{a \in A} d(a) < (2/5 + s)n \cdot (2/5 - s)n = \big(4/25 - s^2\big)n^2 \leqslant 4e^2\big/n^2,$$

a contradiction. Let B be a subset of $N(a)$ of size $(2/5 - s)n$ and let C be the set of vertices not in $A \cup B$. By definition, B is an independent set disjoint from A and $|C| = n - |A| - |B| = n/5$.

By our construction the number of edges between A and $B \cup C$ equals $\sum_{a \in A} d(a) > 4e^2/n^2$ and therefore the set $B \cup C$ spans less than $e - 4e^2/n^2$ edges. Since $e \geqslant n^2/5$ and $f(t) = t - 4t^2$ is a decreasing function for $t \geqslant 1/5$ we have that the number of edges of $G[B \cup C]$ is less than $n^2/25$. Then, by the second averaging argument mentioned in the previous section, there is some subset $X \subseteq C$ of size $n/2 - |B| = (1/10 + s)n$ incident to less than $(n/2 - |B|)/|C| \cdot n^2/25 = (1/2 + 5s)n^2/25$ edges of $G[B \cup C]$. Since B is an independent set, we have that $1/50 \leqslant n^{-2}e(B \cup X) < 1/50 + s/5$ and so $s > 0$.

Suppose that there are βn^2 edges with both endpoints in C. Then $e(C, B) \leqslant (1/25 - \beta)n^2$ and taking X to be a random subset of C of size $(1/10 + s)n$ we can improve the previous computation as follows:

$$\begin{aligned}1/50 &\leqslant n^{-2}e(B \cup X) \leqslant \big((1/2 + 5s)e(C, B) + (1/2 + 5s)^2 e(C)\big)/n^2 \\ &\leqslant (1/2 + 5s)(1/25 - \beta) + (1/2 + 5s)^2\beta \\ &= (1/2 + 5s)/25 - \big(1/4 - 25s^2\big)\beta.\end{aligned}$$

Since $s < 1/135$, this gives $\beta \leqslant (4s/5)/(1 - 100s^2) \leqslant 81s/100$.

Let K be the set of vertices in C with at least $(1/5 - s)n$ neighbours in A, and let $L = C \backslash K$. Recalling the observation made in the third paragraph of the proof, we see that every vertex in K has at most $(1/10 - s)n$ neighbours outside A.

Suppose that $|K| \geqslant 4sn$. Let $K' \subseteq K$ be a set of size $4sn$. Suppose there are αn^2 edges of $G[B \cup C]$ incident to K', where we have $\alpha \leqslant 4s(1/10 - s)$. There is some set $K'' \subseteq C \backslash K'$ of size $n/2 - |B \cup K'|$ incident to at most $(n/2 - |B \cup K'|)/|C \backslash K'| \cdot (1/25 - \alpha)n^2$ edges of $G[B \cup C]$. Then

$$\begin{aligned}1/50 &\leqslant n^{-2}e(B \cup K' \cup K'') \leqslant \alpha + \frac{1/10 - 3s}{1/5 - 4s}(1/25 - \alpha) \\ &\leqslant \alpha + (1/2 - 5s)(1/25 - \alpha) = 1/50 - s/5 + (1/2 + 5s)\alpha \\ &\leqslant 1/50 - s/5 + (1/2 + 5s) \cdot 4s(1/10 - s) \\ &= 1/50 - 20s^3.\end{aligned}$$

This contradiction shows that $|K| < 4sn$.

Recall that, by definition of K, each vertex in $L = C \backslash K$ has less than $(1/5 - s)n$ neighbours in A. Since $s < 1/135$, we also have that $|L| = |C| - |K| > (1/5 - 4s)n > (1/10 - s)n$. By averaging there is some $L' \subseteq L$ of size $(1/10 - s)n$ so that $e(A \cup L') \leqslant |L'|(1/5 - s)n + (|L'|/|L|)^2 e(C)$. Now $|L'|/|L| < (1/10 - s)/(1/5 - 4s) < 1/2 + 10s$ and $(1/2 + 10s)^2 < 1/4 + 11s$. Recalling that $n^{-2}e(C) = \beta < 81s/100$ we have

$$\begin{aligned}1/50 &\leqslant n^{-2}e(A \cup L') \leqslant (1/10 - s)(1/5 - s) + (1/4 + 11s) \cdot 81s/100 \\ &\leqslant 1/50 - 39s/400 + 10s^2.\end{aligned}$$

This gives $10s^2 > 39s/400 > s/11$, i.e., $s > 1/110$, a contradiction that completes the proof.

References

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

[2] P. Erdős, Problems and results in graph theory and combinatorial analysis, in: Proceedings of the Fifth British Combinatorial Conference, Univ. Aberdeen, Aberdeen, 1975, Congr. Numer. 15 (1976) 169–192.
[3] P. Erdős, Some old and new problems in various branches of combinatorics, in: Graphs and Combinatorics, Marseille, 1995, Discrete Math. 165/166 (1997) 227–231.
[4] P. Erdős, R. Faudree, J. Pach, J. Spencer, How to make a graph bipartite? J. Combin. Theory Ser. B 45 (1988) 86–98.
[5] P. Erdős, R. Faudree, C. Rousseau, R. Schelp, A local density condition for triangles, Discrete Math. 127 (1994) 153–161.
[6] P. Erdős, E. Győri, M. Simonovits, How many edges should be deleted to make a triangle-free graph bipartite? in: Sets, Graphs and Numbers, Budapest, 1991, in: Colloq. Math. Soc. János Bolyai, vol. 60, North-Holland, Amsterdam, 1992, pp. 239–263.
[7] P. Keevash, B. Sudakov, Local density in graphs with forbidden subgraphs, Combin. Probab. Comput. 12 (2003) 139–153.
[8] M. Krivelevich, On the edge distribution in triangle-free graphs, J. Combin. Theory Ser. B 63 (1995) 245–260.
[9] B. Sudakov, Making a K_4-free graph bipartite, submitted for publication.