COMBINATORICA 1(4) (1981) 313—317

ON TURÁN'S THEOREM FOR SPARSE GRAPHS

M. AJTAI, P. ERDŐS, J. KOMLÓS

and

E. SZEMERÉDI

Mathematical Institute of the
Hungarian Academy of Sciences
Budapest, Hungary H-1053

Received 24 April 1981

For a graph G with n vertices and average valency t, Turán's theorem yields the inequality $\alpha \geqq n/(t+1)$ where α denotes the maximum size of an independent set in G. We improve this bound for graphs containing no large cliques.

0. Notation

$n=n(G)$=number of vertices of the graph G
$e=e(G)$=number of edges of G
$h=h(G)$=number of triangles in G
$\deg(P)$=valency (degree) of the vertex P
$\deg_3(P)$=triangle-valency of P=number of triangles in G adjacent to P
$t=t(G)=\frac{1}{n}\sum_P \deg(P)=2e/n$=average valency in G (we will tacitly assume $t\geqq 1$)
$T=T(G)$=maximum valency in G
$\alpha=\alpha(G)$=maximum size of independent set of vertices
(independence or stability number)
K_p=shorthand for p-clique
$\log x=\max\{1, \ln x\}$
$t_0, c_1, c_2, \ldots$ are absolute constants
when speaking of union, difference or partition of graphs, we work with the vertex-sets

1. Introduction

Let G be a graph of n vertices and e edges with average valency $t=2e/n$. It is an easy consequence of the celebrated Turán's theorem [6] (and can easily be proved directly) that G contains an independent set of size $n/(t+1)$, i.e.

$$\alpha \geqq n/(t+1). \tag{1}$$

AMS subject classification (1980): 05 C 35; 05 C 55

This estimation is best possible, as shown by the Turán graph: $n/(t+1)$ cliques of size $t+1$. This extreme graph is very stable, graphs which are not that crowded locally have a much higher independence number. This idea of Szemerédi has been formulated by Ajtai, Komlós and Szemerédi in [2] and [3] as follows:

Theorem 1. *If G is trianglefree then* (1) *can be improved to*

$$\alpha > 0.01\,(n/t)\log t \tag{2}$$

(2) *is best possible up to constant multiple.*

Denote by $f(n, t, p)$ the largest integer such that every graph of n vertices and average valency t that contains no K_p satisfies

$$\alpha \geqq f(n, t, p).$$

Theorem 1 states that

$$f(n, t, 3) > c\,(n/t)\log t. \tag{2'}$$

It is possible that for every fixed p we have

$$f(n, t, p) > c_p(n/t)\log t. \tag{3}$$

Perhaps (3) is too optimistic, but we feel that it is an interesting and challenging question. Here we make a modest but perhaps not quite insignificant contribution by proving that for any fixed p, $f(n, t, p)$ tends to infinity with n and t faster than n/t, i.e. the exclusion of K_p improves Turán's bound (1) significantly.

More precisely, we prove the following estimation.

Theorem 2. *There is an absolute constant c_1 such that*

$$f(n, t, p) > c_1(n/t)\log A, \tag{4}$$

where $A = (\log t)/p$.

Thus the exclusion of K_p improves on Turán's bound as long as $p = o(\log t)$. Theorem 2 gives no new information for $p > \log t$. There are two obvious gaps here in our knowledge. The first one is that $p = o(\log t)$ can perhaps be replaced by $p = o(t^\varepsilon)$. The second gap is that we cannot decide whether (3) is true or not even in the case $p = 4$.

The same questions can be asked for hypergraphs. Consider an r-graph with n vertices and e edges. Set $t = t_r$ to be the $(r-1)$-st of the average valency, i.e. $re = nt^{r-1}$. The probabilistic method shows (Spencer [5]) that $\alpha > cn/t$, i.e. G contains cn/t independent vertices. Ajtai, Komlós, Pintz, Spencer and Szemerédi [1, 4] improved this "Turán bound" by a factor $(\log t)^{1/(r-1)}$ by forbidding certain small subgraphs (the assumption is that the hypergraph G contains no cycles of length $\leqq 4$). Both this latter result and Theorem 1 proved to be essential tools in several applications.

Let us now assume that our r-graph G contains no $K^{(r)}(p)$ for some $p > r$. Does that improve the bound $\alpha > cn/t$? In particular, is it true that there is a function $g(t) \to \infty$ such that if G contains no $K^{(3)}(4)$ then $\alpha(G) > c(n/t)g(t)$? This is not even known if we exclude $K^{(3)}(4; 3)$.

This is perhaps the third big gap in this fascinating subject.

2. A sharper version of Theorem 1

A crucial point in the proof of Theorem 2 will be the application of the following sharper form of Theorem 1:

Theorem 1'. *If the number h of triangles in G is less than $\varepsilon n t^2$, where $\varepsilon > 1/(\log t)$, then*

$$\alpha > c_2(n/t)\log 1/\varepsilon. \tag{5}$$

In other words, for any graph G

$$\alpha > c_2(n/t)\min\{\log(nt^2/h);\ \log t\}.$$

Joel Spencer remarked that Theorem 1' is best possible up to constant factor. His example starts from a trianglefree graph G' on n' points with average valency t', $10 < t' < (n')^{1/3}$, and independence number

$$\alpha' < c(n'/t')\log t'.$$

(That such a graph exists is mentioned in [3] — take a random graph and delete the few vertices in triangles.) Now fix a number $s > \exp t'$ and blow up each point to an s-clique. Connecting the vertices of two s-cliques if and only if the original two points were connected in G', we get a graph G with $n = sn'$, $t = st'$. The number of triangles in G is at most

$$s^3n' + s^3n't' < 2snt = (2/t')nt^2 \stackrel{\text{def}}{=} \varepsilon nt^2, \quad \varepsilon = 2/t'.$$

On the other hand,

$$\alpha = \alpha' < c(n'/t')\log t' < 2c(n/t)\log 1/\varepsilon$$

and $\varepsilon > 1/\log t'$.

3. Sparse Subgraph Lemma

Lemma. *Let $p \geqq 2$, $0 < \delta < 1/2$ be arbitrary. If a graph H contains no K_p then it contains a (spanned) subgraph H' with*

$$n(H') \geqq (2\delta)^{p-2}n(H), \quad e(H') < \delta\big(n^2(H')\big)^2.$$

Indeed, for $p=2$ the lemma is trivial. Apply induction on p: If $e(H) < \delta\big(n^2(H)\big)^2$, choose $H' = H$. If $e(H) \geqq \delta n^2(H)$ then there is a point P with $\deg(P) > 2\delta n(H)$; let H' be the neighbourhood of P. It contains no K_{p-1} and $n(H') > 2\delta n(H)$, thus the induction applies.

The above lemma implies the following

Lemma*. *If H contains no K_p then it can be partitioned to $H = H_0 \cup H_1 \cup H_2 \dots$ in such a way that*

$$n(H_i) = \delta^{p-1}n(H), \quad e(H_i) < \delta n^2(H_i), \quad i = 1, 2, \dots .$$

and for the leftover H_0

$$n(H_0) < \delta n(H).$$

Indeed, apply the lemma with $\delta/2$ to get H' with

$$n(H') \geqq \delta^{p-2} n(H), \quad e(H') < (\delta/2) n^2(H').$$

Take a subgraph H_1 of H' with

$$n(H_1) = \delta^{p-1} n(H), \quad e(H_1) < \delta n^2(H_1)$$

(there is such an H_1 since, for any k, the average of $e(H'')\Big/\binom{n(H'')}{2}$ over all subgraphs H'' of H' with $n(H'')=k$, is equal to $e(H')\Big/\binom{n(H')}{2} < 2\delta)$. Then repeat this for $H-H_1$, etc., until we get H_0 with

$$n(H_0) < \delta n(H).$$

4. Proof of Theorem 2

The proof will use induction on n. We consider two cases according to the maximal valency.

If $T > t + 10t/(\log t)$, we pull out a vertex P with valency T and apply induction on on $G - \{P\}$. Since

$$t' = t(G - \{P\}) < (nt - 2t - 20t/(\log t))/(n-1)$$

we have with $A' = (\log t')/p$

$$\alpha(G) \geqq \alpha(G - \{P\}) > c_1((n-1)/t') \log A' > c_1(n/t) \log A.$$

Thus we can assume

(6) $$T \leqq t + 10t/(\log t).$$

We will partition the vertices of G to subsets $V_1, V_2, \ldots$ of size T. Select the point P with the largest triangle-valency $\deg_3(P)$. V_1 will consist of this point, its neighbourhood, and arbitrarily chosen other vertices so that V_1 will have exactly T points. Now in the remaining graph select the vertex with the largest triangle-degree (within this remainin graph), and let V_2 consist of this vertex, its neighbourhood, and some other vertices so that $|V_2| = T$, etc. We get a partition $V_1, V_2, \ldots, V_m$, $m \sim n/T$.

Let us have a closer look to what happens after $V_{m/2}$, when half the vertices have already been partitioned. At the next step we select from the other half of vertices the one with the largest triangle-degree H (within this half). Set $\varepsilon = A^{-3c_1/c_2}$. There are two possibilities:

Case I. $H < \varepsilon T^2$

Case II. $H \geqq \varepsilon T^2$

In Case I the number of triangles within the second half of vertices is less than $(n/2)\varepsilon T^2$, thus by Theorem 1′

$$\alpha > c_2(n/2T) \log(1/\varepsilon)$$

and we get (3) directly.

So we only have to consider Case II. Then at every step up to the $m/2$-th we pulled out a vertex with at least εT^2 triangles, i.e. each V_i, $1 \leqq i \leqq m/2$, contained at least εT^2 edges. Thus there are at most $(1-\varepsilon)nT/2$ edges between the classes $V_1, \ldots, V_m$.

Set $\delta = \varepsilon/10$ and subdivide each class V_i to $V_{i0}, V_{i1}, V_{i2}, \ldots$ according to Lemma* and delete all vertices of V_{i0}, $i=1, 2, \ldots$.

Now $|V_{ij}| = \delta^{p-1}|V_i|$, thus by taking average, we see that there is a choice function j_i such that the number of edges between the subclasses V_{ij_i}, $i=1, 2, \ldots$, is at most $\delta^{2p-2}(1-\varepsilon)nT/2$. The number of edges within a class V_{ij_i} is at most $\delta \cdot \delta^{2p-2}|V_i|^2 = \frac{\varepsilon}{10}\delta^{2p-2}T^2$, thus the number of edges in the graph G' whose vertex set is $\bigcup_i V_{ij_i}$, is less than $(1-0.8\varepsilon)\delta^{2p-2}nT/2$.

Since

$$n' = n(G') > (1-\delta)\delta^{p-1}n(G).$$

we have

$$t' = t(G') < (1-0.8\varepsilon)\delta^{2p-2}n(G)T/n(G')$$

$$< (1-0.7\varepsilon)\delta^{p-1}t(G)\big(1+10/\log t(G)\big) < (1-0.6\varepsilon)\delta^{p-1}t.$$

Applying induction, we have $(A' = (\log t')/p)$.

$$\alpha(G) \geqq \alpha(G') > c_1(n'/t')\log A' > c_1(1+\varepsilon/2)(n/t)\log\left[\frac{\log t-(p-1)\log(1/\delta)-0.6\varepsilon}{p}\right] >$$

$$> c_1(n/t)\log A \quad \text{for} \quad c_1 < c_2/10.$$

References

[1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer and E. Szemerédi, Extremal uncrowded hypergraphs, in manuscript.

[2] M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, *European Journal of Combinatorics*, **2** (1981) 1—11.

[3] M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers, *Journal of Combinatorial Theory A* **29** (1980), 354—360.

[4] J. Komlós, J. Pintz and E. Szemerédi, A lower bound for Heilbronn's problem, *Journal of the London Math. Soc.*, to appear.

[5] J. Spencer, Turán's theorem for k-graphs, *Discrete Math.* **2** (1972), 183—186.

[6] P. Turán, Egy gráfelméleti szélsőértékfeladatról, *Mat. Fiz. Lapok* **48** (1941), 436—452; see also: *On the Theory of Graphs, Colloq. Math.* **3** (1954), 19—30.