ON THE STRUCTURE OF EDGE GRAPHS

BÉLA BOLLOBÁS AND PAUL ERDŐS

Every graph appearing in this note is a finite edge graph without loops and multiple edges. Denote by $G(n, m)$ a graph with n vertices and m edges. $K_r(t)$ denotes a graph with r groups of t vertices each, in which two vertices are connected if and only if they belong to different groups.

By dividing n vertices into $r-1$ almost equal groups and connecting the points in different groups one obtains a graph on n vertices with $\left((r-2)/2(r-1)+\sigma(1)\right)n^2$ edges which does not contain a $K_r(1)$. On the other hand, it was shown by Erdős and Stone [7] that $\left((r-2)/2(r-1)+\varepsilon\right)n^2 (\varepsilon > 0)$ edges assure already the existence of a $K_r(t)$, where $t \to \infty$ as $n \to \infty$. This result is the most essential part of the theorems on the structure of extremal graphs, see e.g. **[3]**, **[4]**, **[6]**, **[9]**.

Let us formulate the result of Erdős and Stone more precisely. Given n, r and ε, put $m = \left[\left((r-2)/2(r-1)+\varepsilon\right)n^2\right]$ ($[x]$ denotes the integer part of x) and define

$$g(n, r, \varepsilon) = \min\{t : \text{every } G(n, m) \text{ contains a } K_r(t)\}.$$

Erdős and Stone proved that if n is large enough then

$$\left(l_{r-1}(n)\right)^{\frac{1}{2}} \leqslant g(n, r, \varepsilon),$$

where l_s denotes the s times iterated logarithm. They also stated that for any fixed $\delta > 0$ and large enough n the same method gives

$$\left(l_{r-1}(n)\right)^{1-\delta} \leqslant g(n, r, \varepsilon).$$

In [7] Erdős and Stone also expected that $l_{r-1}(n)$ is, in fact, the proper order of $g(n, r, \varepsilon)$ if ε is small enough. For $r = 2$ this was stated in **[1]**. In **[2]** Erdős announced that given $\varepsilon > 0$ and $r \geqslant 2$ there exists a constant $c' > 0$ such that

$$c'(\log n)^{1/(r-1)} < g(n, r, \varepsilon),$$

and thought that $g(n, r, \varepsilon)$ will turn out to be of order $(\log n)^{1/(r-1)}$.

The aim of this note is to show that for $r > 2$ the situation is rather different from what seemed likely. The two theorems we prove (of which the second is an easy exercise in the vein of **[5]**) show that for any r and $0 < \varepsilon < 1/2(r-1)$ there are constants c_1 and $c_2 > 0$ such that

$$c_1 \log n \leqslant g(n, r, \varepsilon) \leqslant c_2 \log n$$

if n is sufficiently large and $c_2 \to 0$ as $\varepsilon \to 0$.

The following lemma is needed in the proof of Theorem 1.

Received 1 January, 1973; revised 1 April, 1973.

LEMMA 1. *Let G be a graph with n vertices. Suppose G does not contain a $K_r(t)$ but it contains a $K_{r-1}(q)$, say $\tilde{K}$. Put $N = n-(r-1)q$. Then G has at most*

$$((r-2)q+t)N+2qN^{1-(1/t)}$$

edges of the form (x, y), where x is a vertex of $\tilde{K}$ and y is a vertex of $G-\tilde{K}$.

For the sake of convenience we use the notation $C(a, b) = \binom{a}{b}$. Let $q \geqslant t$ and $r \geqslant 2$ be natural numbers. Suppose $S = \bigcup_{1}^{r-1} S_i$, $|S_i| = q$, $|S| = (r-1)q$. Call a set $K \subset S$ a K_t*-set* if $|K \cap S_i| = t$ for all i. We prove Lemma 1 in the following equivalent form.

LEMMA 1′. *Let $A_1, \ldots, A_N$ be (not necessarily distinct) subsets of S, such that every K_t-set is contained in at most $t-1$ sets A_i. Then*

$$\sum_{1}^{N}|A_i| \leqslant ((r-2)q+t)N+2qN^{1-(1/t)}.$$

Proof. Suppose $|A_i| \geqslant (r-2)q+t$ for $i \leqslant M$ and $|A_i| < (r-2)q+t$ for $i > M$. For $i \leqslant M$ and $j \leqslant r-1$ put $|A_i \cap S_j| = a_{ij}$. Then $a_{ij} \geqslant t$ and A_i contains $\prod_{j=1}^{r-1} C(a_{ij}, t)$ K_t-sets. Therefore, by the assumption,

$$\sum_{i=1}^{M}\prod_{j=1}^{r-1} C(a_{ij}, t) \leqslant (t-1)(C(q, t))^{r-1}.$$

Putting $a_i = \sum_{j=1}^{r-1} a_{ij}-(r-2)q$, one has

$$\prod_{j=1}^{r-1} C(a_{ij}, t) \geqslant (C(q, t))^{r-2}\, C(a_i, t),$$

so

$$\sum_{i=1}^{M} C(a_i, t) \leqslant (t-1)\, C(q, t).$$

As $C(x, t)$ is a convex function of x, on putting $a = \sum_{1}^{M} a_i/M$, this implies

$$MC(a, t) \leqslant (t-1)\, C(q, t).$$

Thus

$$a \leqslant (t-1)^{1/t}qM^{-1/t}+t < 2q\,M^{-1/t}+t,$$

$$\sum_{1}^{M}|A_i| < 2qM^{-1/t}+Mt+M(r-2)q.$$

Since

$$\sum_{M+1}^{N}|A_i| \leqslant (N-M)((r-2)q+t),$$

the lemma follows.

THEOREM 1. *Let $r \geqslant 2$ be an integer and let $\varepsilon > 0$. Then there exists $k = k(\varepsilon, r) > 0$ such that if n is sufficiently large and $m \geqslant \big((r-2)/2(r-1)+\varepsilon\big)n^2$ then every $G(n, m)$ contains a $K_r(t)$ with $t \geqslant [k \log n]$.*

Proof. Let $r = 2$. We show that any $k < -1/\log(2\varepsilon)$ will do for $k(\varepsilon, 2)$. For put $j = j(n) = [k \log n]$ and suppose that there are arbitrary large values of n for which there is a graph $G(n, m)$ without a $K_2(j)$, where $m \geqslant \varepsilon n^2$. Then by a result of Kővári, Sós and Turán [**8**] (improved by Znám [**10**]),

$$\varepsilon n^2 \leqslant \tfrac{1}{2}(j-1)^{1/j} n^{2-(1/j)} + \tfrac{1}{2} jn.$$

That is, after dividing by $\frac{1}{2}n^{2-(1/j)}$, we obtain that for any $\eta > 1$ there is an arbitrary large n, such that

$$2\,\varepsilon n^{1/j} \leqslant \eta,$$

which contradicts the choice of k. This proves the result for $r = 2$.

To prove the result for $r > 2$ we use induction on r. Suppose the theorem holds for $r = r'-1$ and take $r = r'$. For the sake of convenience we denote by $\tilde{G}_\varepsilon(n)$ a graph with n vertices and *at least* $n^2\big((r-2)/2(r-1)+\varepsilon\big)$ edges. If x is any positive real number, we also put $\tilde{K}_r(x) = K_r([x])$, where $[x]$ is the smallest integer not less than x. We suppose that $\varepsilon < 1/2(r-1)$, for otherwise there is nothing to prove.

By the induction hypothesis there exists a positive number e_r, depending on r, such that every $\tilde{G}_\varepsilon(n)$ contains a $\tilde{K}_{r-1}\big(q(n)\big)$, where $q(n) = e_r \log n$. We shall show that *every $\tilde{G}_\varepsilon(n)$ contains a $\tilde{K}_r\big(\varepsilon(r-1)/4\,q(n)\big) = \tilde{K}_r(d_r\,\varepsilon \log\, n) = \tilde{K}_r\big(t(n)\big)$ if n is large enough.*

Suppose that, contrary to this assertion, there are arbitrarily large values of n for which some $\tilde{G}_\varepsilon(n)$ does not contain a $K_r\big(t(n)\big)$. Given any N, let

$$G_0 = \tilde{G}_\varepsilon(n) \equiv \tilde{G}_\varepsilon(n_0),\, n_0 \geqslant N/\varepsilon,$$

be such a graph. Define a sequence of graphs $G_0 \supset G_1 \supset \ldots, G_k = \tilde{G}_\varepsilon(n_k)$, as follows. If G_k has a vertex x_k of degree less than $n_k(r-2/r-1+\varepsilon)$, put $G_{k+1} = G_k - x_k$. One can easily check that for $k \geqslant n - \varepsilon n$ the graph G_k would be the complete graph, so the sequence must stop with a graph $G_k = \tilde{G}_\varepsilon(n_k)$, where $n_k \geqslant \varepsilon n \geqslant N$. Furthermore, $\tilde{G}_\varepsilon(n_k)$ does not contain a $K_r\big(t(n)\big)$ so it does not contain a $K_r\big(2t(n_k)\big)$ either if N is large enough (e.g. if $t(N) \leqslant 2t(N/\varepsilon)$).

Consequently there are arbitrarily large values of n for which some graph $\tilde{G} = \tilde{G}_\varepsilon(n)$ does not contain a $\tilde{K}_r\big(2t(n)\big)$ and has only vertices of degree at least $n(r-2/r-1+\varepsilon)$. Put $q = q(n)$, $t = 2t(n)$. Then $\tilde{G}$ contains a $\tilde{K}_{r-1}(q)$, say $\tilde{K}_{r-1}$ but does not contain a $\tilde{K}_2(t)$ and so by the result of Kővári, Sós and Turán [**8**] there are at most

$$A = \binom{r-1}{2} q^2 + (r-1)\{\tfrac{1}{2}(t-1)^{1/t} q^{2-1/t} + \tfrac{1}{2}qt\}$$

edges in the subgraph spanned by $\tilde{K}_{r-1}$. Furthermore, by the lemma, at most $B = ((r-2)q+t)(n-(r-1)q)+2qn^{1-1/t}$ edges connect $\tilde{K}_{r-1}$ to $\tilde{G}-\tilde{K}_{r-1}$. Finally, as every vertex of $\tilde{G}$ has degree at least $n(r-2/r-1+\varepsilon)$, we must have

$$2A+B \geqslant (r-1)\,qn\,(r-2/r-1+\varepsilon),$$

and so

$$2qn^{1-(1/t)}+(r-1)\,q^2+tn \leqslant (r-1)\,\varepsilon qn.$$

As this inequality does not hold if n is sufficiently large, the theorem is proved.

THEOREM 2. *Let $0 < \varepsilon < \frac{1}{2}$ and $c > -2/\log(2\varepsilon)$. Then for every sufficiently large n there exists a graph $G(n, m)$ not containing a $K_2(t)$, where $m = [\varepsilon n^2]$ and $t = [c \log n]$.*

Proof. The number of $K_2(t)$ graphs on n distinguishable vertices is

$$C(n, 2t)\, C(2t, t)/2$$

and there are $C(n(n-1)/2-l,\, m-l)$ graphs with m edges containing a given set of l edges. Thus the result follows if we show that for large enough n one has

$$C(n(n-1)/2-t^2,\, m-t^2)\; C(n, 2t)\; C(2t, t)/2\; C(n(n-1)/2, \mathrm{m}) < 1.$$

As the left hand side is bounded by

$$(n(n-1)/2)^{-t^2}\, m^{t^2}\, n^{2t} \leqslant \left(\frac{2\,\varepsilon\, n}{n-1}\right) t^2\, n^{2t},$$

which tends to zero since $c > -2/\log(2\varepsilon)$, the proof is complete.

Remarks. 1. Denote by $c_r(\varepsilon)$ the *supremum* of the possible values for $k(\varepsilon, r)$

Then Theorem 2 and the first part of the proof of Theorem 1 show that

$$-1/\log(2\varepsilon) \leqslant c_2(\varepsilon) \leqslant -2/\log(2\varepsilon),$$

and

$$d_r\, \varepsilon \leqslant c_r(\varepsilon),$$

where $d_r > 0$ depends only on r.

Remarks. 2. If $0 < \varepsilon < \frac{1}{2}(r-1)^2$ then $c_r(\varepsilon) \leqslant c_2((r-1)^2\,\varepsilon)$

$$\leqslant -2/\log\left(2(r-1)^2\,\varepsilon\right), \tag{1}$$

so in particular $c_r(\varepsilon) \to 0$ as $\varepsilon \to 0$ for every $r \geqslant 0$. To prove (1) note that for every $\eta > 0$ we can construct the following graph if n is sufficiently large. Take an $(r-1)$-partite graph on n vertices with maximal number of edges (there are $[n+i-1/r-1]$ vertices in the ith class). Add $\varepsilon n^2 = (r-1)^2\,\varepsilon(n/r-1)^2$ edges to a class of it in such a way that the class contains no $K_2(t)$ if $t \geqslant \left(c_2((r-1)^2\,\varepsilon)+\eta\right) \log n/r-1$. Then the graph obtained in this way has no $K_r(t)$ if

$$t \geqslant \left(c_2\left((r-1)^2\,\varepsilon\right)+\eta\right) \log n.$$

Remarks. 3. It is very likely that inequality (1) gives, in fact, the right order of $c_r(\varepsilon)$, i.e. there exists a $c_r^* > 0$ such that

$$-c_r^*/\log \varepsilon \leqslant c_r(\varepsilon) \tag{2}$$

as $\varepsilon \to 0$. For $r = 2$ inequality (2) follows from Theorem 2, as we have already remarked.

References

1. P. Erdős, " On extremal problems of graphs and generalised graphs ", *Israel J. Math.*, 2 (1965) 183–190.

2. ——, " Some recent results on extremal problems in graph theory ", *Actes des journées d'études sur la théorie des graphes* (I.C.C. Dunod, 1967), 117–130.

3. ——, " Of some new inequalities concerning extremal properties of graphs ", *Theory of graph*, ed. by P. Erdős and G Katona (Acad. Press, N.Y., 1968), 83–98.

4. ——, " On some extremal problems of *r*-graphs ", *Discrete Math.*, 1 (1971), 7–18.

5. —— and A. Rényi, " On the evolution of random graphs ", *Publ. Math. Inst. Hung. Acad.*, 5 (1960), 17–61.

6. —— and M. Simonovits, "A limit theorem in graph theory ", *Studia Sci. Math. Hung. Acad.*, 1 (1966), 51–57.

7. —— and A. H. Stone, " On the structure of linear graphs ", *Bull. Amer. Math. Soc.*, Vol. LII (1946), 1087–1091.

8. T. Kővári, V. T. Sós and P. Turán, " On a problem of Zarankievicz ", *Coll. Mat.*, 3 (1954), 50–57.

9. M. Simonovits, "A method for solving extremal problems in graph theory, stability problems ", *Theory of Graphs*, ed. by P. Erdős and G. Katona (Acad. Press, N.Y., 1968), 279–319.

10. Š. Znám, " Two improvements of a result concerning a problem of K. Zarankiewicz ", *Colloq. Math.*, 13 (1965), 255–258.

University of Cambridge

and

Hungarian Academy of Sciences,
Budapest.