EXTREMAL PROBLEMS IN GRAPH THEORY

P. ERDÖS

In the present paper $\mathfrak{G}(n; l)$ will denote a graph of n vertices and l edges. K_p will denote the complete graph of p vertices $\mathfrak{G}\left(p; \binom{p}{2}\right)$ and $K(p, p)$ will denote the complete bipartite graph, more generally $K(p_1, \ldots, p_r)$ will denote the complete r-chromatic graph with p_i vertices of the i-th colour, in which every two vertices of different colour are adjacent. C_n will denote a circuit having n edges.

In 1940 TURÁN [1] posed and solved the following question: Determine the smallest integer $m(n, p)$ so that every $\mathfrak{G}(n; m(n, p))$ contains a K_p. Turán in fact showed that the only $\mathfrak{G}(n; m(n, p) - 1)$ which contains no K_p is $K(m_0, \ldots m_{p-2})$ where the m_i are all as nearly equal as possible, i.e. for $0 \leqq i \leqq p - 2$ $m_i = [(n + i - 1)/(p - 1)]$. Thus a simple computation gives that if $n \equiv r \pmod{p - 1}$ then

$$m(n, p) = \frac{p - 2}{2(p - 1)} (n^2 - r^2) + \binom{r}{2}.$$

Turán further asked: How many edges must a graph contain that it should certainly have subgraphs of a prescribed structure? In particular he asked: Determine the smallest $h(k, n)$ so that every $\mathfrak{G}(n; h(k, n))$ should contain a path of length k. GALLAI and I [2] and ANDRÁSFAI [2] investigated these and related questions and solved them nearly completely. In the present paper we shall try to investigate as systematically as possible the following question: What is the smallest integer $f(n; k, l)$ for which every graph $\mathfrak{G}(n; f(n; k, l))$ contains a $\mathfrak{G}(k; l)$ as a subgraph? These problems become very much more difficult, but in my belief also more interesting, if we also consider the structure of the graphs $\mathfrak{G}(k; l)$. We now define three functions $f_i(n; k, l)$, $1 \leqq i \leqq 3$. $f_1(n; k, l)$ is the smallest integer for which every $\mathfrak{G}(n; f_1(n; k, l))$ contains at least one $\mathfrak{G}(k; l)$. $f_2(n; k, l)$ is the smallest integer for which there is a $\mathfrak{G}(k; l)$ of given structure so that every $\mathfrak{G}(n; f_2(n; k, l))$ contains this $\mathfrak{G}(k; l)$. $f_3(n; k, l)$ is the smallest integer so that even the $\mathfrak{G}(k; l)$ which requires most edges occurs in $\mathfrak{G}(n; f_3(n; k, l))$. Clearly $\mathfrak{G}(n; f_3(n; k, l))$ contains all the graphs of k vertices and l edges. Trivially

$$f_1(n; k, l) \leqq f_2(n; k, l) \leqq f_3(n; k, l).$$

It is easy to see that in general $f_1(n; k, l) < f_2(n; k, l)$, since it is not hard to see that for $n > 7$ $f_1(n; k, [k^2/4] + 2) = [n^2/4] + 2$ but $\mathfrak{G}(n; [n^2/4] + 2)$ does not have to contain any $\mathfrak{G}(k; [k^2/4] + 2)$ of any given configuration. Further it is easy to see that in general $f_2(n; k, l) < f_3(n; k, l)$. Further I recently proved that

$$(1) \qquad f_3(n; k, l) = f_1\left(n; u, \binom{u}{2}\right)$$

where $\binom{u}{2} \leqq l < \binom{u+1}{2}$.

Now we will try to determine as systematically as possible the values of $f_1(n; k, l)$ for fixed k as l increases from 1 to $\binom{k}{2}$, as far as possible we will investigate $f_i(n; k, l)$ too (for $2 \leqq i \leqq 3$), in other words we will investigate structural problems too.

We will give no proofs in this paper; if no reference is given to a result then it is not yet published.

Assume first $l < k$. If $l \leqq \frac{1}{2}k$ then trivially (where there is no danger of misunderstanding we write $f(n; k, l)$ for $f_1(n; k, l)$)

$$(2) \qquad f(n; k, l) = l\,.$$

If $\frac{1}{2}k < l < k$ then it is easy to see that

$$(3) \qquad f(n; k, l) = f(n;\ 2l + 2 - k,\ 2l + 1 - k)\,.$$

Finally

$$(4) \qquad f(n;\ k, k-1) = \left[\frac{(k-2)\,n}{k-1}\right] + 1\,.$$

The structural problems are very much more difficult: GALLAI and I proved [3] that every

$$(5) \qquad \mathfrak{G}(n; e(k, n)),\ e(k, n) = \max\left[\binom{2k-1}{2} + 1,\ (k-1)\,n - (k-1)^2 + \binom{k-1}{2} + 1\right]$$

contains k independent edges and that this result is the best possible. The proof is not easy. Trivially every $\mathfrak{G}(n; [\frac{1}{2}(k-1)\,n] + 1)$ contains a star of valency k. Further Gallai and I [2] proved that every $\mathfrak{G}(n; [\frac{1}{2}(k-1)\,n] + 1)$ contains a path of length k. V. T. Sós and I conjectured that every $\mathfrak{G}(n; [\frac{1}{2}(k-1)\,n] + 1)$ contains all trees having k edges and that every $\mathfrak{G}(n; e(k, n))$ contains all forests (i.e. graphs all whose components are trees) of k edges, but we did not succeed to prove any of these conjectures.

For $l = k$ there is a sharp jump in the behaviour of $f(n; k, l)$ since $f(n; k, k)/n \to \infty$ for every fixed k as $n \to \infty$. Before we continue our investigations for general k, we discuss as completely as possible the cases $k = 3$, 4 and 5. For $k = 3$ there is only one graph $\mathfrak{G}(3; 3)$, the triangle, and by Turán's theorem [1]

$$f(n; 3, 3) = \left[\frac{n^2}{4}\right] + 1 .$$

For $k = 4$ there are two graphs $\mathfrak{G}(4; 4)$, the square and the triangle, with an edge. A simple argument shows that for $n \geqq 4$

$$f_3(n; 4, 4) = \left[\frac{n^2}{4}\right] + 1 .$$

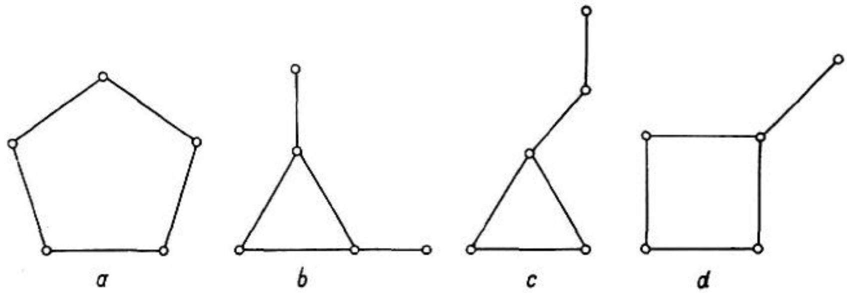

Fig. 1.

On the other hand the determination of $f_1(n; 4, 4) = f_2(n; 4, 4)$ seems to be a very difficult problem (i.e. how many edges does a graph of n vertices have to have in order to contain a square?).

E. KLEIN and I proved [4] that (the c's denote suitable absolute constants)

$$c_1 n^{3/2} < f_1(n; 4, 4) < c_2 n^{3/2} .$$

The sharpest estimates at present are due to REIMAN [5]; he proved

$$\frac{1}{2\sqrt{2}} \leqq \lim_{n=\infty} f(n; 4, 4)/n^{3/2} \leqq \tfrac{1}{2} . \tag{5}$$

It seems likely that the limit in (5) equals $\frac{1}{2}\sqrt{2}$ but it is not even known whether the limit in question exists.

It is easy to see that $f(n; 4, 5) = [n^2/4] + 1$ [6], there is only one graph $\mathfrak{G}(4; 5)$; K_4 minus an edge. More generally DIRAC and I [7] proved (independently) that every $\mathfrak{G}(n; m(n, k))$ already contains a K_{k+1} from which at most one edge is missing.

$f(n; 4, 6)$ is given by Turán's theorem.

It is not difficult to see that for $n > n_0, f_3(n; 5, 5) = [n^2/4] + 1$. The graphs $\mathfrak{G}(5; 5)$ are in Fig. 1 and **c** gives for $n > n_0, f_1(n; 5, 5) = f_2(n; 5, 5) = f_1(n; 4, 4)$.

CAVALLIUS [8] obtains an upper bound for $f_1(n; 5, 6) = f_2(n; 5, 6)$ (more generally he gives an upper bound for $u(n; 2, k)$ where $u(n; 2, k)$ is the smallest integer for which every $\mathfrak{G}(n; u(n; 2, k))$ contains a $K(2, k)$).

Besides $K(2, 3)$ the other graphs $\mathfrak{G}(5; 6)$ are given by Fig. 2 and it is easy to see that if $n > n_0$ all of them appear in an $\mathfrak{G}(n; [n^2/4] + 1)$, thus $f_3(n; 5, 6) = [n^2/4] + 1$.

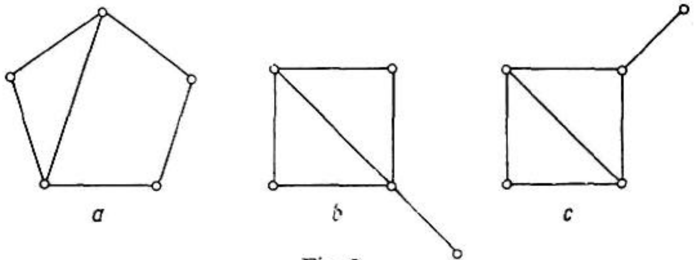

Fig. 2.

There are four types of graphs $\mathfrak{G}(5; 7)$, Fig. 3. Dirac and I showed (independently) that for $n > n_0$ every $\mathfrak{G}(n; [n^2/4] + 1)$ contains graphs of types ***a*** and ***b***, i.e. $f_1(n; 5, 7) = f_2(n; 5, 7) = [n^2/4] + 1$. I showed that every $\mathfrak{G}(n; l_n)$, $l_n = [n^2/4] + [n/4] + [(n+1)/4] + 1$ contains also subgraphs of the type ***c*** and it is easy to determine all the graphs $\mathfrak{G}(n; l_n - 1)$ which do not contain graphs of the above type.

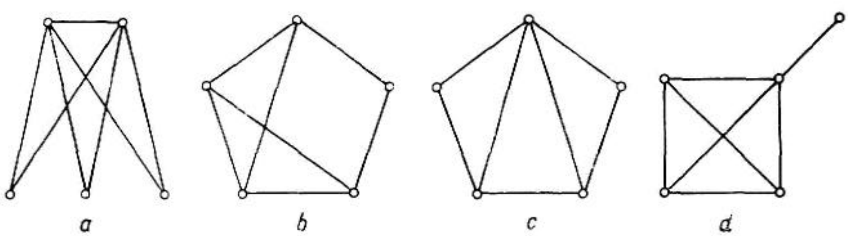

Fig. 3.

As already stated, Dirac and I showed that every $\mathfrak{G}(n; [n^2/4] + 1)$ contains ***d***, in fact it even contains a $\mathfrak{G}(5; 9)$.

There are two types of $\mathfrak{G}(5; 8)$, Fig. 4. ***a*** is settled by the sharpening of Turán's theorem due to Dirac and myself, Dirac and I further showed (independently) that every $\mathfrak{G}(n; l_n)$ contains ***b***.

$\mathfrak{G}(5; 9)$ and $\mathfrak{G}(5; 10)$ present no new difficulties.

I did not carry out a similar discussion for graphs having 6 vertices. I only state one result which seems interesting:

$$f_1(n; 6, 12) > \frac{n^2}{4} + c_3 n^{3/2}, \quad f_2(n; 6, 12) < \frac{n^2}{4} + c_4 n^{3/2}$$

and in fact every $\mathfrak{G}(n; [n^2/4 + c_4 n^{3/2}])$ contains as subgraph an octahedron. I cannot prove that $\lim_{n=\infty} (f_1(n; 6, 12) - n^2/4)/n^{3/2}$ exists.

Now we return to the discussion of $f_1(n; k, l)$ for general k. First of all I proved that for $k \geqq 3$

$$(6) \qquad f_1(n; k, k) < c'_k n^{1+1/[k/2]} .$$

It seems likely that

$$(7) \qquad f_1(n; k, k) > c''_k n^{1+1/[k/2]}$$

but I can prove (7) only for $3 \leqq k \leqq 5$. I can prove the weaker result

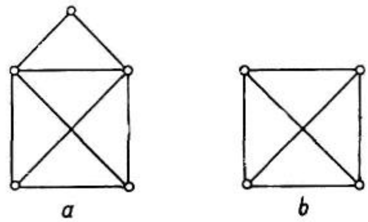

Fig. 4.

$$(8) \qquad f_1(n; k, k) > n^{1+\varepsilon_k}$$

for a certain $\varepsilon_k > 0$.

I can also prove that every $\mathfrak{G}(n; [c'''_k n^{1+1/k}])$ contains a C_{2k}; the proof is more difficult than the proof of (6).

For large values of k our three functions $f_i(n; k, l)$ do not suffice to describe completely the many problems, since there are very many graphs $\mathfrak{G}(k; l)$, but we have not succeeded in solving or even in classifying the many problems which can be posed here; I will try to state here all the results which are known.

I showed that for $n > n_0(k)$, every $\mathfrak{G}(n; [n^2/4] + 1)$ contains a C_{2k+1}; in fact there exists an absolute constant c so that every $\mathfrak{G}(n; [n^2/4] + 1)$ contains every C_m for $3 \leqq m \leqq cn$. The proof is not trivial. $K([n/2], [(n+1)/2])$ shows that $[n^2/4] + 1$ is best possible [13].

Now we investigate the range $k < l \leqq k^2/4$. Kövári, the Turáns [9] and (independently) I proved that every $\mathfrak{G}(n; [c_k n^{2-1/k}])$ contains a $K(k, k)$. It seems likely that this result is best possible and in fact we conjectured

$$f_1(n; 2k, k^2) > \alpha_k n^{2-1/k} ;$$

but this is proved only for $k = 2$, and we could not even prove that

$$\lim_{n=\infty} f(n; 6, 9)/n^{3/2} = \infty .$$

Further I proved that every $\mathfrak{G}(n; [\beta_k n^{2-1/k}])$ contains a $K(k+1, k+1)$ from which one edge is perhaps missing (the structure of this graph is uniquely determined).

In the range $k < l \leqq [k^2/4]$ I do not have good estimates for $f(n; k, l)$, I cannot even prove that for fixed k and sufficiently large n, $f(n; k, l)$ is a strictly monotone function of l.

I proved that for every $\varepsilon > 0$ there is an $\eta = \eta(\varepsilon) > 0$ so that for $k > k_0(\eta)$ and $n > n_0(k, \varepsilon, \eta)$,

$$f(n; k, [(1 + \eta) k] < n^{1+\varepsilon} \tag{9}$$

(the opposite inequality, with a different ε, follows from (8)).

Further for every $k \geqq 2c$ and $n > n_0(k)$, $l < ck$,

$$f(n; k, l) < n^{2-\varepsilon} \tag{10}$$

where ε depends only on c. In fact the following stronger result holds: Every $\mathfrak{G}(n; [n^{2-\varepsilon}])$ contains a subgraph of $K([k/2], [k/2])$ which has at least ck edges for every $2c < k \leqq n$ if $n > n_0(k)$ and $\varepsilon = \varepsilon(c)$.

On the other hand, to every $\varepsilon > 0$ there is a $C = C(\varepsilon)$ so that

$$f(n; k, Ck) > n^{2-\varepsilon}\,. \tag{11}$$

Instead of (11) the following sharper result can be proved: Let $\varepsilon' < \varepsilon$ and $C = C(\varepsilon)$ be sufficiently large. If $n > n_0(\varepsilon, \varepsilon', C)$ then there exists a $\mathfrak{G}(n; [n^{2-\varepsilon}])$ which does not contain a subgraph $\mathfrak{G}(k; Ck)$ for every $k < n^{\varepsilon'}$. This result is nearly the best possible, since it is not hard to show that every $\mathfrak{G}(n; [n^{2-\varepsilon}])$ contains a $\mathfrak{G}(k; Ck)$ for some $k < C_1 n^{\varepsilon}$ where $C_1 = C_1(\varepsilon)$ is sufficiently large.

I would like to state here one further result which can be proved by probabilistic methods [10]: Let $\varepsilon > 0$, $C > 1$ be arbitrary. There is a graph $\mathfrak{G}(n; Cn)$ so that every subgraph of it spanned by $m < \eta n$ vertices has fewer than $m(1 + \varepsilon)$ edges; $\eta = \eta(\varepsilon, C)$ could easily be estimated explicitly.

It is not hard to show that [11] for $\alpha > 1$ every $\mathfrak{G}(n; [\alpha n])$ contains a circuit of length $< \beta \log n$, where β depends only on α. Probably every $\mathfrak{G}(n; [\alpha n])$ $(\alpha > 1)$ contains a subgraph $\mathfrak{G}([\beta_t \log n]; [\beta_t \log n] + t)$ where β_t depends only on t.

Now we give a very short discussion of $l > [k^2/4]$. Dirac and I showed independently that every $\mathfrak{G}(n; [n^2/4] + 1)$ contains, for every $k \leqq n$, a $\mathfrak{G}(k; [k^2/4] + 1)$. In fact Dirac proved a more general theorem.

Considerably more difficult is the proof of the following result: To every k there is an $n_0(k)$ so that for every $n > n_0(k)$ every $\mathfrak{G}(n; [n^2/4] + 1)$ contain sa $K(k, k)$ with an extra edge (the structure of these graphs is uniquely determined) [13].

It is not hard to show by complete induction that for $[(k + 1)/4] \geqq u$,

$$f_1\left(n; k, \left[\frac{k^2}{4}\right] + u\right) = \left[\frac{n^2}{4}\right] + u\,. \tag{12}$$

It is easy to see that (12) no longer holds for $u > [(k + 1)/4]$, but the discontinuity is not very sharp since it is easy to see by induction that if $n \geqq k$ then

$$f\left(n; k, \left[\frac{k^2}{4}\right] + \left[\frac{k-1}{2}\right]\right) = \left[\frac{n^2}{4}\right] + \left[\frac{n-1}{2}\right] \tag{13}$$

and now there is a sharp discontinuity since it is not difficult to show that

$$(14)\qquad f\left(n; k, \left[\frac{k^2}{4}\right] + \left[\frac{k+1}{2}\right]\right) \geqq \left[\frac{n^2}{4}\right] + f\left(\left[\frac{n}{2}\right]; k, k\right) \geqq \frac{n^2}{4} + n^{1+\varepsilon_k}$$

by (8). We can determine the values of $f(n; k, [k^2/4] + l)$ for $[(k+1)/4] \leqq l < < [(k-1)/2]$, but the formulas are complicated and we omit them.

By very complicated arguments I can show that every $\mathfrak{G}(n; [n^2/4 + n^{1+1/[k/2]}])$ contains a $K(k, k)$ and a circuit whose vertices are all the vertices of one of the k-tuples.

In 1946 STONE and I proved [12] that for every $\varepsilon > 0$ and every k if $n > n_0(\varepsilon, k)$, $\mathfrak{G}(n; [(n^2/4)(1+\varepsilon)])$ contains a $K(k, k, k)$; by a refinement of our method I can in fact show that for sufficiently large C every $\mathfrak{G}(n; [n^2/4 + Cn^{2-1/k}])$ already contains a $K(k, k, k)$.

I do not pursue the investigations of $f(n; k, l)$ further since a complete analysis is hopeless at present and so far I have succeeded to find no new phenomena in the interval $k^2/3 \leqq l \leqq \binom{k}{2}$.

Before completing the discussion of $f_i(n; k, l)$ I would like to mention two further problems: It is not hard to show that

$$f(n; 7, 15) = \left[\frac{n^2}{4}\right] + \left[\frac{n+1}{3}\right] + 1\,.$$

But I cannot decide the question, how many edges must a graph of n vertices have in order that it contain a $K(1, 3, 3)$? Perhaps $[n^2/4] + n + 1$ edges suffice for this purpose. It is easy to see that $[n^2/4] + n$ edges are not sufficient. (Added in proof: I succeeded in proving this conjecture.) Very many other such problems could be stated.

Turán asked in a conversation to determine the smallest number of edges that a graph of n vertices must have in order that it contain the various regular bodies. For the tetrahedron the answer is $m(n, 3)$ by [1], the octahedron has already been discussed. The problem of the cube seems difficult. I can show that for sufficiently large c every $\mathfrak{G}(n, [cn^{3/2}])$ contains a hexagon and a vertex joined to three non adjacent vertices of the hexagon but I cannot decide whether it contains a cube. The icosahedron, dodecahedron and higher dimensional cubes have not been investigated so far.

Before completing the paper I would like to state a few related results which cannot be described in terms of our functions $f_i(n; k, l)$. PÓSA and I [11] proved that for $n > 24k$, every $\mathfrak{G}(n; (2k-1)\, n - 2k^2 + k + 1)$ contains k vertex independent circuits (i.e. k circuits which pairwise have no common vertex). But we have not succeeded in solving the extremal problem for $3k \leqq n \leqq 24k$, except for a few special values of k.

Pósa proved that every $\mathfrak{G}(n; 2n - 3)$ contains a circuit with at least one diagonal and that the result is false for $\mathfrak{G}(n; 2n - 4)$. CZIPSZER found a very simple and ingenious proof of this result; by his method one can easily show that for a certain c and $n > n_0(k)$ every $\mathfrak{G}(n; kn + c)$ contains a circuit with at least $k - 1$ diagonals emenating from a vertex. It is easy to see that $c \geqq 1 - k^2$. Perhaps $c = 1 - k^2$? For $k = 2$ this is Pósa's result, and I can prove it for $k = 3$ and $k = 4$ also.

Finally I proved that for every ε and r there is an $n_0 = n_0(\varepsilon, r)$ so that for every $n > n_0(\varepsilon, r)$ every $\mathfrak{G}(n; [n^2/4] + n(1 + \varepsilon))$ contains a circuit and r vertices not on this circuit each of which is adjacent to every vertex of the circuit. It is easy to see that this result does not hold for every $\mathfrak{G}(n; [n^2/4] + n)$.

References

[1] *P. Turán:* Eine Extremalaufgabe aus der Graphentheorie (written in Hungarian). Mat. és Fiz. Lapok *48* (1941) 436—452. See also *P. Turán*: On the theory of graphs. Coll. Math. *3* (1954) 19—30.

[2] *P. Erdös* and *T. Gallai:* On the maximal paths and circuits of graphs. Acta Math. Hung. Acad. Sci. *10* (1959) 337—356. *B. Andrásfai:* On the paths, circuits and loops of graphs (written in Hungarian). Matematikai Lapok *13* (1962) 65—107.

[3] See *P. Erdös* and *T. Gallai,* [2], p. 354—356.

[4] *P. Erdös:* On sequences of integers no one of which divides the product of two others and on some related problems. Mitt. Forschungsinst. Math. u. Mech. Tomsk *2* (1938) 74—82.

[5] *I. Reiman:* Über ein Problem von K. Zarankiewicz. Acta Math. Acad. Sci. Hungar. *9* (1958) 269—279.

[6] *P. Erdös:* Some theorems on graphs (in Hebrew). Riveon Lematematika *9* (1955) 13—17.

[7] *G. Dirac:* Extensions of Turán's theorem on graphs. Acta Hung. Acad. Sci. *14* (1963) 417—422.

[8] *H. Cavallius:* On a combinatorial problem. Coll. Math. *6* (1958) 59—65.

[9] *T. Kövári, V. T. Sós* and *P. Turán:* On a problem of K. Zarankiewicz. Coll. Math. *3* (1954) 50—57.

[10] *P. Erdös:* On circuits and subgraphs of chromatic graphs. Mathematika *9* (1962) 170—175. The result in question is not proved explicitly in this paper, but can easily be deduced by the method used in the proof of Lemma 2, p. 173.

[11] *P. Erdös* and *L. Pósa:* On the maximal number of disjoint circuits of a graph. Publ. Math. Debrecen *9* (1962) 3—12, see p. 11—12.

[12] *P. Erdös* and *A. H. Stone:* On the structure of linear graphs. Bull. Amer. Math. Soc. *52* (1946) 1087—1091.

[13] *P. Erdös:* On the structure of linear graphs. Israel Journal of Math. *1* (1963) 156—160.