Studia Scientiarum Mathematicarum Hungarica 1 (1966) 215—235.

ON A PROBLEM OF GRAPH THEORY

by

P. ERDŐS[1], A. RÉNYI[1] and V. T. SÓS[2]

§ 0. Introduction

Let G_n be a non-directed graph having n vertices, without parallel edges and slings. Let the vertices of G_n be denoted by $P_1, \dots, P_n$. Let $v(P_i)$ denote the valency of the point P_i and put

$$V(G_n) = \max_{1 \leqq i \leqq n} v(P_i). \tag{0. 1}$$

Let $E(G_n)$ denote the number of edges of G_n. Let $\mathbf{H}_d(n, k)$ denote the set of all graphs G_n for which $V(G_n)=k$ and the diameter $D(G_n)$ of which is $\leqq d$, $(k=1, 2, \dots, n-1;\ d=2, 3, \dots, n-1)$.

In the present paper we shall investigate the quantity

$$F_d(n, k) = \min_{G_n \in \mathbf{H}_d(n, k)} E(G_n). \tag{0. 2}$$

Thus we want to determine the minimal number N such that there exists a graph having n vertices, N edges and diameter $\leqq d$ and the maximum of the valencies of the vertices of the graph is equal to k.

To help the understanding of the problem let us consider the following interpretation. Let be given in a country n airports; suppose we want to plan a network of direct flights between these airports so that the maximal number of airports to which a given airport can be connected by a direct flight should be equal to k (i.e. the maximum of the capacities of the airports is prescribed), further it should be possible to fly from every airport to any other by changing the plane at most $d-1$ times; what is the minimal number of flights by which such a plan can be realized? For instance, if $n=7$, $k=3$, $d=2$ we have $F_2(7, 3)=9$ and the extremal graph is shown by Fig. 1.

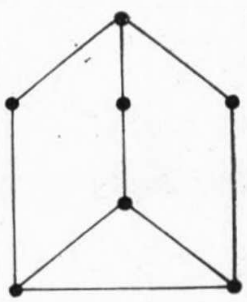


Fig. 1

The problem of determining $F_d(n, k)$ has been proposed and discussed recently by two of the authors (see [1]). In § 1 we give a short summary of the results of the paper [1], while in § 2 and 3 we give some new results which go beyond those of [1]. Incidentally we solve a long-standing problem about the maximal number of edges of a graph not containing a cycle of length 4.

In § 4 we mention some unsolved problems.

Let us mention that our problem can be formulated also in terms of $0-1$ matrices as follows: Let $M=(\varepsilon_{ij})$ be a symmetrical n by n zero-one matrix such

[1] Mathematical Institute of the Hungarian Academy of Sciences.
[2] Eötvös L. University, Budapest.

that $\varepsilon_{ii}=1$, $\max\limits_{1\leq i\leq n}\sum\limits_{j=1}^{n}\varepsilon_{ij}=k+1$ and all elements of M^d are $\geqq 1$. We want to determine

$$M_d(n, k) = \min \sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_{ij}.$$

Clearly

(0. 3) $$M_d(n, k)=2F_d(n, k)+n.$$

This formulation shows the connection of our problem with non-linear programming.

We give for the case $d=2$ a third formulation of our problem which displays its connection with the theory of block designs.

Let be given a sequence $A_1, A_2, \ldots, A_n$ of subsets of the elements $1, 2, \ldots, n$ such that if $j\in A_i$ then $i\in A_j$. Let us suppose that denoting by $|A|$ the cardinal number of the set A, we have $\max\limits_{1\leq j\leq n}|A_j|=k$. Let us suppose that for any i $(1\leqq i\leqq n)$ and any $j\neq i$ such that $j\notin A_i$ there is a set A_h which contains both i and j (this is equivalent by our supposition of symmetry to the statement that the sets A_i and A_j are not disjoint). The problem is to determine

(0. 4) $$\min \sum_{i=1}^{n}|A_i| = 2F_2(n, k).$$

§ 1. Some Basic Inequalities, and some Asymptotic Results

It is easy to see that if there exists a graph G_n with $V(G_n)=k$ and diameter $\leqq d$, then

(1. 1) $$n \leqq 1+k\frac{(k-1)^d-1}{k-2}.$$

(1. 1) can be proved as follows: if $V(G_n)=k$ the number of points which can be reached from a given point, say, P_1 by an edge is $\leqq k$; the number of points which can be reached from P_1 by a path of length 2 is $\leqq k(k-1)$ and finally the number of points which can be reached by a path of length d is $\leqq k(k-1)^{d-1}$. Thus if the graph has diameter $\leqq d$ we have

$$(n-1)\leqq k\bigl(1+(k-1)+(k-1)^2+\ldots+(k-1)^{d-1}\bigr).$$

This proves (1. 1). If both n and k are odd, then G_n must contain at least one point of valency $\leqq k-1$ (because the number of points of odd valency cannot be odd); thus in this case we get

(1. 2) $$n \leqq 1+(k-1)\frac{(k-1)^d-1}{k-2}.$$

Note that for the graph shown by Fig. 1, equality stands in (1. 2). For the graph shown on Fig. 2 (the so-called Petersen-graph) equality stands in (1. 1) with $n=10$, $k=3$, $d=2$.

As regards $F_d(n, k)$ we obtain easily the lower bound

$$F_d(n, k) \geqq \frac{n(n-1)(k-2)}{2((k-1)^d-1)}. \tag{1.3}$$

(1. 3) can be proved as follows: every edge is itself a path of length 1; it can be contained in at most $2(k-1)$ paths of length 2, but in this way each path of length 2 is counted twice, thus the number of paths of length 2 cannot exceed $E(G_n)(k-1)$. In general each edge can be contained in at most $3(k-1)^2$ paths of length 3, but in this way each path of length 3 is counted three times, thus the number of paths of length 3 cannot exceed $E(G_n)(k-1)^2$, etc. As in case G_n has diameter $\leqq d$ the number of paths of length $\leqq d$ has to be at least $\binom{n}{2}$, we obtain

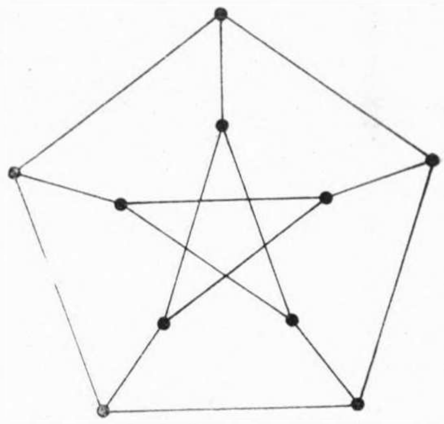


Fig. 2

$$E(G_n)\left(1+(k-1)+\ldots+(k-1)^{d-1}\right) \geqq \binom{n}{2}, \tag{1.4}$$

which implies (1. 3). Note that one has equality in (1. 4) for the Petersen graph shown on Fig. 2., further for $n=5$, $k=2$, $d=2$ because a cycle of length 5 has 5 vertices, each of which has valency 2, it has diameter 2 and the number of its edges is $5=\frac{5\cdot 4}{2\cdot 2}$.

It is clear from the above proof that one can have equality in (1. 4) only for a regular graph of order k, i. e. if $E(G_n)=\frac{nk}{2}$ and if any two points are joined by one and only one path of length $\leqq d$.

The first condition implies that if equality stands in (1. 4) then there is equality in (1. 1) too. For the case $d=2$ this means that a necessary condition of equality in (1. 4) is $n=k^2+1$. It has been shown by A. J. HOFFMAN and R. R. SINGLETON [4] that a regular graph of order k, having k^2+1 points and diameter 2 exists only for $k=2, 3, 7$ and perhaps for $k=57$. Thus for $d=2$ except for these values of k one has strict inequality in (1. 3). However it has been shown in [1] that there exists an infinite sequence of pairs (k_j, n_j) such that $k_j \to \infty$, $n_j \to \infty$ and

$$\lim_{j\to\infty} \frac{F_2(n_j, k_j)\, k_j}{n_j(n_j-1)} = \frac{1}{2}. \tag{1.5}$$

This is a consequence of the following

THEOREM 1. *If P is any prime power, there exists a graph G_n of order $n=P^2+P+1$ for which $V(G_n)=P+1$, which has diameter 2 and for which $E(G_n) \leqq \frac{1}{2}(n^{3/2}+n)$. The graph G_n has also the property that it does not contain any cycle of length 4.*

To make this paper self-contained we reproduce the proof of Theorem 1 given in [1].

PROOF OF THEOREM 1. Let $GF(P)$ be the Galois field with P elements. Let us represent the points of the finite plane geometry $PG(P, 2)$ by triples (a, b, c) where a, b, c are elements of $GF(P)$, not all three equal to 0, and $(\lambda a, \lambda b, \lambda c)$ with $\lambda \in GF(P)$, $\lambda \neq 0$ represents the same point as (a, b, c). The number of different points of $PG(P, 2)$ is P^2+P+1. A straight line in $PG(P, 2)$ is the set of all points (x, y, z) which satisfy the equation $ax+by+cz=0$; we denote this line by $[a, b, c]$. The point (a, b, c) and the line $[a, b, c]$ are clearly conjugate elements with respect to the conic $x^2+y^2+z^2=0$. As well known there are $P+1$ points on each line, any two different lines have exactly one point in common and through any two given points there is exactly one straight line. Now we define the mapping T which maps the point $A=(a, b, c)$ into the line $\alpha=[a, b, c]$ and conversely. We write $TA=\alpha$, $T\alpha=A$. This mapping has evidently the properties: if the point B lies on the line $\alpha=TA$ then the point A lies on the line $\beta=TB$; if C is the point of intersection of the lines TA and TB then TC is identical with the line passing through the points A and B; $A=(a, b, c)$ is on TA if and only if $a^2+b^2+c^2=0$, i.e. if A lies on the conic $x^2+y^2+z^2=0$. Now let us define a graph G_n $(n=P^2+P+1)$ as follows: the vertices of G_n are the points of $PG(P, 2)$; the vertices $A=(a, b, c)$ and $A'=(a', b', c')$ are joined in G_n by an edge if and only if A' is lying on TA (and thus A is lying on TA'). Clearly a vertex A in G_n has the valency P or $P+1$ according to whether A is on the conic $x^2+y^2+z^2=0$ or not.* Thus

$$\frac{1}{2}(n^{3/2}-n) \leqq \frac{1}{2}P(P^2+P+1) \leqq E(G_n) \tag{1.6}$$

and

$$E(G_n) \leqq \frac{1}{2}(P+1)(P^2+P+1) \leqq \frac{1}{2}(n^{3/2}+n).$$

Finally the diameter of G_n is equal to 2. As a matter of fact any two points A and B can be joined by the path ACB where C is the point of intersection of the lines TA and TB. Besides this A and B can be joined by a single edge if A lies on TB. But the point C such that the edges AC and BC both belong to G_n is in any case unique; thus G_n does not contain any cycle of length 4.

Thus our Theorem is proved.

We deduce from Theorem 1 the following corollaries.

COROLLARY 1. Put $n_k=k^2-k+1$; then

$$\liminf_{k\to\infty} \frac{F_2(n_k, k)k}{n_k(n_k-1)} = \frac{1}{2}. \tag{1.7}$$

* If P is prime, there are $P+1$ points on the conic and thus

$$E(G_n) = \frac{P(P+1)^2}{2} > \frac{1}{2}n^{3/2} \quad \text{if} \quad n \geqq n_0.$$

PROOF OF COROLLARY 1. By (1. 3)

$$\frac{F_2(n, k)k}{n(n-1)} \geqq \frac{1}{2} \tag{1.8}$$

further if $k=P+1$, $n_k=P^2+P+1$, by Theorem 1

$$F_2(P^2+P+1,\, P+1) \leqq \tfrac{1}{2}(P+1)(P^2+P+1); \tag{1.9}$$

thus in this case

$$\frac{F_2(n_k, k)k}{n_k(n_k-1)} \leqq \frac{1}{2}\left(1+\frac{1}{P}\right); \tag{1.10}$$

this proves our assertion.

Theorem 1 enables us also to solve — at least asymptotically — a problem which was raised by one of us 27 years ago (see [2]).*

Let $\mathbf{C}_n$ denote the class of graphs having n vertices and containing no cycle of order 4. Put

$$\mu(n) = \max_{G_n \in \mathbf{C}_n} E(G_n). \tag{1.11}$$

The problem is to determine the value of $\mu(n)$. From Theorem 1 we deduce the following

COROLLARY 2. We have

$$\lim_{n\to\infty} \frac{\mu(n)}{n^{3/2}} = \frac{1}{2}. \tag{1.12}$$

PROOF OF COROLLARY 2. It follows clearly from Theorem 1 that if P is a prime power, then putting $n=P^2+P+1$

$$\mu(n) \geqq \frac{1}{2}(n^{3/2}-n). \tag{1.13}$$

It is possible that for these n the graph of Theorem 1 is extremal but we cannot prove this. Clearly $\mu(n)$ is an increasing function of n, and thus it follows that for any n we have

$$\mu(n) \geqq \frac{1}{2}[P^2+P+1)^{3/2}-(P^2+P+1)] \tag{1.14}$$

where P is the largest prime power such that $P^2+P+1 \leqq n$. Now evidently for $n \geqq n_1$ one can choose a prime p so that

$$\sqrt{n} - \frac{\sqrt{n}}{\log n} \leqq p \leqq \sqrt{n} - 1 \tag{1.15}$$

* After having written this paper we have been informed by W. G. BROWN that independently of us he has proved (1.12), in the same way as we did. His paper will be published in the Bulletin of the Canadian Mathematical Society.

which implies for $n \geqq n_1$

$$n\left(1-\frac{2}{\log n}\right) \leqq p^2+p+1 \leqq n.$$

Thus we have for any $n \geqq n_1$

$$\mu(n) \geqq \frac{1}{2} n^{3/2}\left(1-\frac{3}{\log n}\right) \tag{1.16}$$

and thus

$$\liminf_{n\to\infty} \frac{\mu(n)}{n^{3/2}} \geqq \frac{1}{2}. \tag{1.17}$$

On the other hand it is easy to see (this follows also from the results of I. REIMAN in [3]) that

$$\limsup_{n\to\infty} \frac{\mu(n)}{n^{3/2}} \leqq \frac{1}{2}. \tag{1.18}$$

As a matter of fact, let G_n be a graph containing no cycle of order 4. Let $P_1, P_2, \ldots, P_n$ be the vertices of G_n and let us denote their valencies by $v_1, v_2, \ldots, v_n$. Now clearly one can select from the set E_i of vertices joined by an edge to P_i $\binom{v_i}{2}$ pairs, and no pair (P_j, P_h) can be contained in both E_i and E_l with $l \neq i$ because otherwise $P_i P_j P_l P_h$ would be a cycle contained in G_n. Thus we must have

$$\sum_{i=1}^{n} \binom{v_i}{2} \leqq \binom{n}{2}. \tag{1.19}$$

Now we have

$$\left(\sum_{i=1}^{n} v_i\right)^2 \leqq n \sum_{i=1}^{n} v_i^2 \tag{1.20}$$

and thus

$$\left(\sum_{i=1}^{n} v_i\right)^2 - n\left(\sum_{i=1}^{n} v_i\right) \leqq 2n \sum_{i=1}^{n} \binom{v_i}{2} \leqq 2n\binom{n}{2} \leqq n^3. \tag{1.21}$$

As clearly $\sum_{i=1}^{n} v_i = 2E(G_n)$, we have

$$4E^2(G_n) - 2nE(G_n) \leqq n^3 \tag{1.22}$$

which implies

$$E(G_n) \leqq \frac{n^{3/2}}{2}\sqrt{1+\frac{1}{4n}} + \frac{n}{4}. \tag{1.23}$$

Thus

$$\frac{\mu(n)}{n^{3/2}} \leqq \frac{1}{2}\sqrt{1+\frac{1}{4n}} + \frac{1}{4\sqrt{n}}. \tag{1.24}$$

which implies (1.18). Thus Corollary 2 is proved.

Let us note that weaker results have been obtained previously by E. KLEIN (see [2]) and I. REIMANN [3], who proved $\liminf_{n\to\infty} \frac{\mu(n)}{n^{3/2}} \geqq \frac{1}{2\sqrt{2}}$. REIMANN's extremal graph does not contain triangles either; it is possible that among such graphs it is optimal.

Note that for the pairs (n_j, k_j) for which according to Corollary 1 one has

$$\lim_{j\to\infty} \frac{F_2(n_j, k_j)k_j}{n_j(n_j-1)} = \frac{1}{2} \tag{1.25}$$

one has $k_j \sim \sqrt{n_j}$. It was shown in [1] that there exists another sequence of pairs (k_j, n_j) such that

$$\lim_{j\to\infty} \frac{F_2(n_j, k_j)k_j}{n_j(n_j-1)} = 1 \tag{1.26}$$

but for this sequence of pairs one has $\lim_{j\to\infty} \frac{k_j^2}{n_j} = +\infty$.

It remains an open question what is the value of the function $g(c)$ defined by

$$g(c) = \liminf_{\substack{k^2 > nc \\ n\to\infty}} \frac{F_2(n,k)k}{n(n-1)} \tag{1.27}$$

for $1 < c < +\infty$; we know only that $g(c)$ is nondecreasing, $\frac{1}{2} \leqq g(c)$ and $\lim_{c\to\infty} g(c) \geqq 1$.

§ 2. Some Exact Results for $d = 2$.

In this § we deal with the exact value of $F_2(n, k)$ for $\frac{n}{2} \leqq k \leqq n-1$. Evidently, $F_2(n, n-1) = n-1$, because the graph G_n in which one vertex is joined by an edge with all others, has diameter 2, further $V(G_n) = n-1$ and $E(G_n) = n-1$. It has been shown in [1] that $F_2(n, n-2) = 2n-4$ (a graph G_n with $V(G_n) = n-2$ and $E(G_n) = 2n-4$ and having the diameter 2 is shown by Fig. 3; another graph with the same properties is shown by Fig. 4), further that $F_2(n, n-3) = F_2(n, n-4) = 2n-5$. (The corresponding extremal graphs are shown by Figs. 5 and 6.)

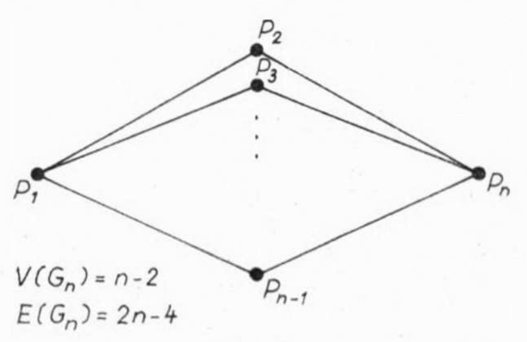


Fig. 3

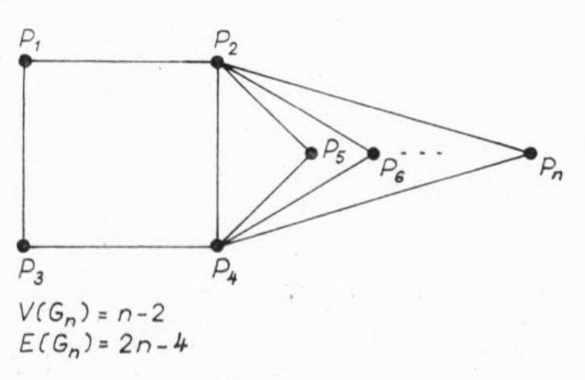


Fig. 4

We shall prove now

THEOREM 2. *We have for* $n \geqq 13$

$$F_2(n, k) = 2n-4 \quad \text{for} \quad \frac{2n-2}{3} \leqq k \leqq n-5. \tag{2.1}$$

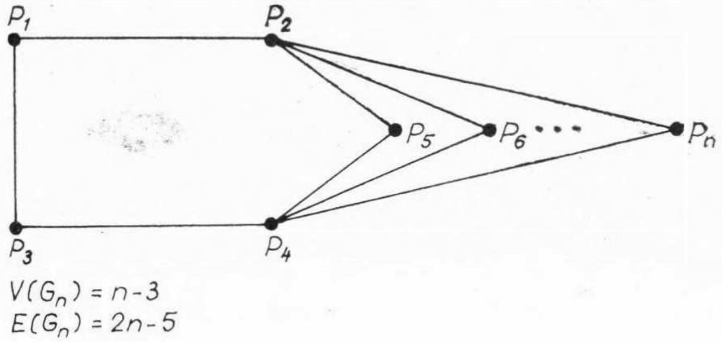


Fig. 5

PROOF OF THEOREM 2. The extremal graph G_n with

$$V(G_n) = k = n-l, \qquad \left(5 \leqq l \leqq \frac{n+2}{3}\right)$$

and $E(G_n) = 2n-4$ and having diameter 2 is exhibited by Fig. 7.

$$V(G_n) = n-l, \quad E(G_n) = 2n-4, \quad 5 \leqq l \leqq \frac{n+2}{3}; \qquad n \geqq 13.$$

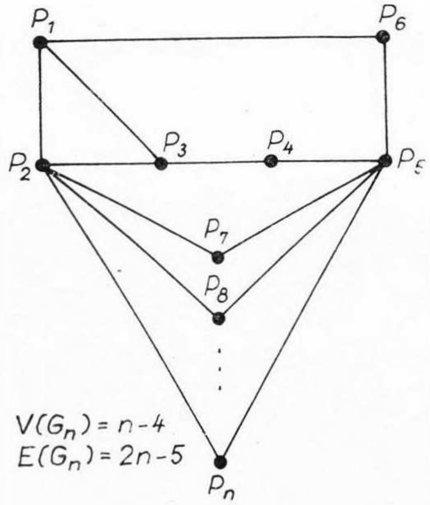


Fig. 6

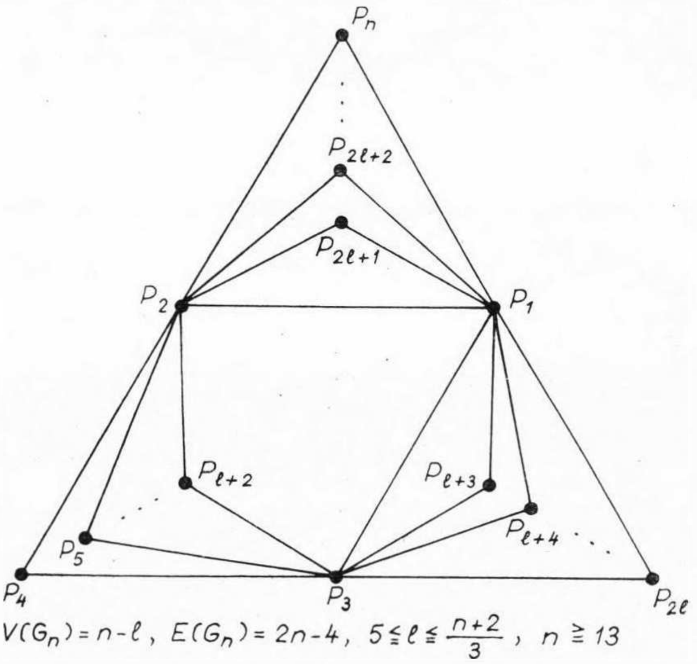


Fig. 7

Note that all vertices of G_n except P_1, P_2 and P_3 have the valency 2, further $v(P_1)=n-l$, $v(P_2)=n-l$, $v(P_3)=2l-2$ and by supposition $2l-2 \leqq n-l$. Thus $V(G_n)=n-l$. Clearly G_n has diameter 2 and the number of edges of G_n is

$$E(G_n) = \frac{2(n-l)+2l-2+2(n-3)}{2} = 2n-4.$$

We prove that for any G_n with $n \geqq 13$, $V(G_n)=n-l\left(5 \leqq l \leqq \frac{n+2}{3}\right)$ and diameter 2 one has $E(G_n) \geqq 2n-4$.

As $E(G_n)=\frac{1}{2}\sum_{i=1}^{n} v(P_i)$ we may suppose that G_n contains at least one point of degree $\leqq 3$. If G_n would contain no point of degree $\leqq 2$, then let us choose a point of degree 3; let this point be P_1. Let the points connected by an edge with P_1 be denoted by P_2, P_3 and P_4. As every point can be reached from P_1 by a path of length $\leqq 2$, we must have $v(P_2)+v(P_3)+v(P_4) \geqq n-1$.

Now if there would be a point among the points $P_5, \ldots, P_n$ which would be connected with more than one of the points P_2, P_3, P_4 we would have $v(P_2)+$ $+v(P_3)+v(P_4) \geqq n$; as all other points have degree $\geqq 3$ it would follow

$$\sum_{i=1}^{n} v(P_i) \geqq n+3(n-3) = 4n-9$$

and thus $E(G_n)>2n-5$ i.e. $E(G_n) \geqq 2n-4$, which was to be proved. Thus we may suppose that all points $P_i (5 \leqq i \leqq n)$ are connected with one and only one of P_2, P_3 and P_4; similarly we can suppose that P_2, P_3 and P_4 are not connected with each other because this would again imply $v(P_2)+v(P_3)+v(P_4) \geqq n$ and thus $E(G_n) \geqq$ $\geqq 2n-4$. If there is at least one among the P_i with $5 \leqq i \leqq n$ which has degree >3, it again follows that $E(G_n) \geqq 2n-4$. If however all have degree 3, let us suppose that $v(P_2)=\min\big(v(P_2), v(P_3), v(P_4)\big)$ which implies $v(P_2) \leqq \frac{n-1}{3}$. Let P_5 be connected with P_2. Then $v(P_5)=3$ and let the three points connected with P_5 be P_2, P_i and P_j; clearly $i>5$ and $j>i>5$. But then $v(P_2)+v(P_i)+v(P_j) \geqq n-1$ and thus

$$6 = v(P_i)+v(P_j) \geqq \frac{2(n-1)}{3}$$

that is $n \leqq 10$.

As we supposed $n \geqq 13$, this case is settled.

The case when there is a point P_i of valency 1 is easily settled, because if this point is P_1, and P_1 is connected with P_2 only, then P_2 has to be connected with the remaining $n-2$ points too, and thus would have valency $n-1$. Thus the only case which remains to be settled is when $\min_{1 \leqq i \leqq n} v(P_i)=2$. Suppose $v(P_1)=2$ and let P_1 be connected with P_2 and P_3. Then all remaining points have to be connected either with P_2 or with P_3 or with both.

Let $\mathbf{C}_1$ denote the class of points P_i with $i \geqq 4$ connected only with P_2, and c_1 the number of elements of $\mathbf{C}_1$; let $\mathbf{C}_2$ be the class of points P_i with $i \geqq 4$ connected only with P_3 and c_2 the number of elements of $\mathbf{C}_2$; finally let $\mathbf{C}_3$ be the class of points connected with both P_2 and P_3, and c_3 the number of elements of $\mathbf{C}_3$. Clearly

$c_1+c_2+c_3=n-3$. As the valency of P_3 cannot exceed $n-l$ and P_3 is connected with every point in G_n except itself and the points in $\mathbf{C}_1$, we have $c_1 \geqq l-2 \geqq 3$. Similarly $c_2 \geqq l-2 \geqq 3$.

The number of edges in G_n the existence of which is already established is clearly $c_1+c_2+2c_3+2=n+c_3-1$. Let us call these the edges of the first kind, and the remaining edges those of the second kind. As the graph has diameter 2, every point of $\mathbf{C}_1$ has to be connected by a path of length $\leqq 2$ with every point of $\mathbf{C}_2$. Such a path can not contain an edge of the first kind. Thus the graph G' consisting of the edges of the second kind has to be connected. Now three cases are possible. Either G' contains besides the points of $\mathbf{C}_1$ and $\mathbf{C}_2$ at least one further point from the class $\mathbf{C}_3$; in this case it contains at least c_1+c_2+1 points and thus there are at least c_1+c_2 edges of the second kind, and thus the total number of edges is $E(G_n) \geqq n+c_3-1+c_1+c_2=2n-4$. Or P_2 and P_3 are connected by an edge; in this case we get again $E(G_n) \geqq 2n-4$. Or P_2 and P_3 are not connected and G' consists only of the points of $\mathbf{C}_1$ and $\mathbf{C}_2$. In this case the connected graph G' is either a tree or not. If it is not a tree, it contains at least c_1+c_2 edges and thus we obtain again $E(G_n) \geqq 2n-4$. If G' is a tree, it must have at least two end-points. We may suppose that $\mathbf{C}_1$ contains an endpoint of G'. Let x be the total number of endpoints of G' in $\mathbf{C}_1$. Then the sum of valencies (in G') of the points of $\mathbf{C}_1$ is at least $x+2(c_1-x)$. As G_n has diameter 2 and P_2 and P_3 are not directly connected, any endpoint of G' in $\mathbf{C}_1$ has to be connected by a path of length 2 to P_3, it follows that for every endpoint P of G' in $\mathbf{C}_1$ the single edge starting from P ends in $\mathbf{C}_2$. Let y denote the number of points in $\mathbf{C}_2$ which are connected with an endpoint of G' in $\mathbf{C}_1$. If Q is such a point, clearly Q has to be connected with every other point of $\mathbf{C}_2$, because otherwise there would not exist a path of length 2 from P to these points. Now clearly no point of $\mathbf{C}_2$ can be an endpoint of G', because it must be connected to at least one point in $\mathbf{C}_1$ and also to Q. Thus the sum of valencies in G' of the points of $\mathbf{C}_2$ ist at least $2(c_2-y)+y(c_2-1)+x$. It follows that the number of edges of the second kind is at least

$$\frac{1}{2}\left(x+2(c_1-x)+2(c_2-y)+y(c_2-1)+x\right)=c_1+c_2+y\left(\frac{c_2-3}{2}\right) \geqq c_1+c_2,$$

because, as we have shown, $c_2 \geqq 3$.

Thus we have shown that $E(G_n) \geqq 2n-4$ and the proof of Theorem 2 is complete.

Note that the restriction $n \geqq 13$ in Theorem 2 is necessary, because for $n<13$ there is no value of k between $\frac{2n-2}{3}$ and $n-5$.

As regards the value of $F_2(n,k)$ for $k<\frac{2n-2}{3}$ we can show that for $n \geqq 15$

$$F_2(n,k)=\begin{cases} 3n-k-6 & \text{for } \frac{3n-3}{5} \leqq k < \frac{2n-2}{3} \\ 5n-4k-10 & \text{for } \frac{5n-3}{9} \leqq k < \frac{3n-3}{5} \\ 4n-2k-13 & \text{for } \frac{n+1}{2} \leqq k < \frac{5n-3}{9}. \end{cases} \tag{2.2}$$

We give in what follows the extremal graphs for these 3 cases. That these are really extremal can be proved in a way similar to the proof of Theorem 2, therefore we leave the details to the reader.

THE EXTREMAL GRAPH FOR $\frac{3n-3}{5} \leqq k < \frac{2n-2}{3}$.

The graph has four points of high degree; let us denote them by A, B, C, D and four groups of points.

There is a group denoted by AB, the points of which are joined to A and to B. The group contains $2k-n$ points. In the group BCD (connected with B, C and D) there are $n-k-1$ points. In the group AC (whose points are connected with A and C) there are $\left[\frac{n-k-3}{2}\right]$ points; finally in the group AD (the points of which are connected with A and D) there are $n-k-3-\left[\frac{n-k-3}{2}\right]$ points. Further the graph contains the edges AB, AC, AD. The points A and B have the degree k. The whole graph has $3n-k-6$ edges.

THE EXTREMAL GRAPH FOR $\frac{5n-3}{9} \leqq k < \frac{3n-3}{5}$.

There are 5 points of high order, A, B, C, D, E.

The group AB has $2k-n$ points,

The group BCD has $\left[\frac{n-k-1}{2}\right]$ points.

The group BCE has $n-k-1-\left[\frac{n-k-1}{2}\right]$ points.

The group AC has $2k-n$ points.

The group ADE has $2n-3k-4$ points.

Further the edges AB, AC, AD, AE, DE belong to the graph. The points A, B and C have the valency $n-k$; the total number of edges is $5n-4k-10$.

THE EXTREMAL GRAPH FOR $\frac{n+1}{2} \leqq k < \frac{5n-3}{9}$.

There are 6 points of high order, A, B, C, D, E, F.

The group AB contains $2k-n$ points.

The group BCE contains $\left[\frac{n-k-1}{2}\right]$ points.

The group BDF contains $n-k-1-\left[\frac{n-k-1}{2}\right]$ points.

The group ADC contains $\left[\frac{n-k-5}{2}\right]$ points.

The group AEF contain; $n-k-5-\left[\frac{n-k-5}{2}\right]$ points.

The graph contains further the edges AB, AC, AD, AE, AF. The graph has $4n-2k-13$ edges.

For $k<\frac{n+1}{2}$ we cannot determine $F_2(n, k)$ exactly. However, we can get a fairly good upper bound by constructing graphs of diameter 2 by the following principles. We divide all but $\binom{r}{2}$ of the points of a graph G_n into r groups of approximately the same size. We connect the points of each pair of groups with one of the remaining points, and connect as many of these points with each other

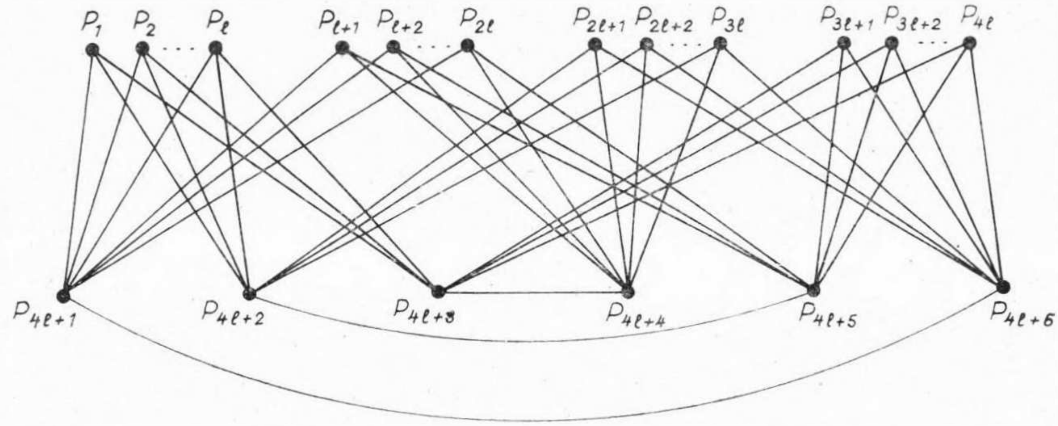


Fig. 8

as needed. For instance if $r=4$, $n=4l+6$, we put l points in each of 4 groups, connect each of the 6 pairs of groups with one of the remaining 6 points, and connect each of these points with that point which is connected with the other two groups. The graph obtained is shown by Fig. 8. It follows that

$$F_2(4l+6,\, 2l+1) \leqq 12l+3. \tag{2. 3}$$

§ 3. Some Results for $d \geqq 3$.

We prove first

THEOREM 3. *We have for every n, every $k \leqq n-1$ and $d \geqq 3$*

$$F_d(n, k) \geqq \frac{n^2}{k^{d-1}}\left(1-4\sqrt[3]{\frac{n}{k^d}}\right). \tag{3. 1}$$

PROOF OF THEOREM 3. Let us put

$$\delta = 4\sqrt[3]{\frac{n}{k^d}}. \tag{3. 2}$$

Clearly we may suppose $\delta<1$, because otherwise (3. 1) is trivially fulfilled. We have evidently

$$\delta > \frac{4}{k^{d/3}}. \tag{3. 3}$$

We may suppose $n > k^{d-1}$, because any graph G_n with diameter $\leqq d$ is connected and thus has at least $n-1$ edges; thus $F_3(n, k) \geqq n-1$ and if $n \leqq k^{d-1}$ the inequality (3. 1) is trivial. Thus we have to prove (3. 1) only for $k^{d-1} < n < \frac{k^d}{64}$ i.e. for $(64n)^{1/d} < < k < n^{1/d-1}$. *

Let G_n be a graph having n vertices, diameter d and such that $V(G_n) = k$. Let us denote by $X_1, \ldots, X_s$ those vertices of G_n the valency of which is $< \frac{4n}{k^{d-1}\delta}$; let $Y_1, \ldots, Y_{n-s}$ be the remaining vertices of G_n. We have clearly

$$E(G_n) = \frac{1}{2}\left(\sum_{i=1}^{s} v(X_i) + \sum_{j=1}^{n-s} v(Y_j)\right) > \frac{2(n-s)n}{k^{d-1}\delta}. \tag{3.4}$$

Thus if

$$s \leqq n\left(1 - \frac{\delta(1-\delta)}{2}\right)$$

we have

$$E(G_n) \geqq \frac{n^2}{k^{d-1}}(1-\delta). \tag{3.5}$$

Thus we have to consider only the case

$$s > n\left(1 - \frac{\delta(1-\delta)}{2}\right). \tag{3.6}$$

We distinguish two cases. Either every X_i $(1 \leqq i \leqq s)$ is connected with at least $\left(1-\frac{\delta}{2}\right)\frac{n}{k^{d-1}}$ of the vertices Y_j, or not. In the first case we have

$$E(G_n) \geqq s\left(1-\frac{\delta}{2}\right)\frac{n}{k^{d-1}} \geqq (1-\delta)\frac{n^2}{k^{d-1}}. \tag{3.7}$$

Thus we may suppose that there is an X_i — say X_1 — which is connected with less than $\left(1-\frac{\delta}{2}\right)\frac{n}{k^{d-1}}$ Y_j-s. We shall show that this case is impossible. By supposition we can reach, starting from X_1, every vertex of G by a path of length $\leqq d$. Let us consider first those paths starting from X_1, the next vertex of which is an Y_j. As Y_j can be chosen in $< \left(1-\frac{\delta}{2}\right)\frac{n}{k^{d-1}}$ ways, and all vertices of G_n have valency $\leqq k$, the number of such pathes is at most

$$\left(1-\frac{\delta}{2}\right)\frac{n}{k^{d-1}}\left(1+(k-1)+(k-1)^2+\ldots+(k-1)^{d-1}\right) \leqq \left(1-\frac{\delta}{2}\right)n. \tag{3.8}$$

* We may also suppose that $k > 64$.

Let us count now the pathes of length $\leqq d$ starting from X_1, on which the point next to X_1, is an X_i. The number of such pathes is clearly at most

$$\frac{4n}{k^{d-1}\delta}\left(1+\frac{4n}{k^{d-1}\delta}\left(1+(k-1)+\ldots+(k-1)^{d-2}\right)\right) < \frac{\delta n}{3}. \tag{3.9}$$

It follows from (3. 8) and (3. 9) that the total number of vertices which can be reached from X_1 by a path of length $\leqq d$, can not exceed $n\left(1-\frac{\delta}{6}\right)$ which is $\leqq n-2$ if $n\geqq\frac{12}{\delta}$, and this is true if $n-1\geqq k\geqq 64$; thus we arrived to a contradiction and this proves our theorem.

To show that the order of magnitude $\frac{n^2}{k^2}$ of the lower estimate of $F_3(n,k)$ is best possible, consider the following graph G_n: Take a complete graph G_r having r vertices, and connect each vertex of G_r with $r-1$ new points. Thus we obtain a graph G_n with $r(r-1)+r=r^2=n$ vertices. Clearly one has $k=V(G_n)=2r-2$, $D(G_n)=3$ and $E(G_n)=\frac{3}{2}r(r-1)$. Thus $E(G_n)\sim\frac{6n^2}{k^2}$.

In this example $k=2(\sqrt{n}-1)$; by slightly modifying this example we obtain that

$$F_3(n,k) < \frac{n^2}{k^2}(c+1)^2\left(\frac{1}{2}+\frac{1}{c}\right)$$

if $k\sim cn$ where $0<c<1$.

To show that $F_3(n,k)$ is of order of magnitude $\frac{n^2}{k^2}$ for $k\sim\lambda\sqrt{n}$ where $0<\lambda<1$ we have to apply a more involved construction. Let us consider a graph G_n which has the vertices P_{gij} where $1\leqq g\leqq l$, $1\leqq i\leqq s$, $1\leqq j\leqq s$ and the vertices Q_{ghi} where $1\leqq g<h\leqq l$ and $1\leqq i\leqq s$; thus $n=ls^2+\binom{l}{2}s$. Suppose that the edges of G_n are as follows:

a) P_{gij} and P_{hij} are both connected with Q_{ghi} for $1\leqq g<h\leqq l$, $i,j=1,2,\ldots,s$.

b) Q_{ghi_1} is connected with Q_{ghi_2} for $1\leqq i_1\leqq s$, $1\leqq i_2\leqq s$, $i_1\neq i_2$, $1\leqq g<h\leqq l$;

c) $Q_{g_1h_1i}$ and $Q_{g_2h_2i}$ are connected for $1\leqq g_1<h_1\leqq l$ and $1\leqq g_2<h_2\leqq l$, $i=1,2,\ldots,s$.

Clearly

$$E(G_n) = 2\binom{l}{2}s^2+\binom{l}{2}\binom{s}{2}+\binom{\binom{l}{2}}{2}s$$

further $v(P_{gij})=s-1$ and

$$v(Q_{ghi}) = s+l-1+\binom{\binom{l}{2}}{2}-1$$

and thus $V(G_n)=s+l+\binom{\binom{l}{2}}{2}-2$. Thus we obtain

$$F_3\left(ls^2+\binom{l}{2}s,\ s+l+\frac{(l+1)l(l-1)(l-2)}{8}-2\right)\leqq 2\binom{l}{2}s^2+\binom{l}{2}\binom{s}{2}+\binom{\binom{l}{2}}{2}s.$$

By other words by choosing for l an arbitrary fixed natural number and for s tending to $+\infty$, we obtain an infinite sequence of pairs n, k such that

$$k\sim\frac{\sqrt{n}}{\sqrt{l}}\quad\text{and}\quad F_3(n,k)\leqq\frac{5}{4}\frac{n^2}{k^2}l(l-1).$$

Thus for arbitrary small $\lambda>0$ there exists an infinity of pairs n, k such that $k\sim\lambda\sqrt{n}$ and

$$F_3(n,k)<\frac{5}{4\lambda^4}\cdot\frac{n^2}{k^2}.$$

Let us study now the behaviour of $F_3(n,k)$ for large values of k. Clearly $F_3(n,k)=n-1$ if $k\geqq\frac{n}{2}$ because the graph G_n shown on Fig. 9 has diameter 3 $V(G_n)=k$ and G_n is a tree, thus it has $n-1$ edges; this result is best possible because a connected graph G_n cannot have less than $n-1$ edges.

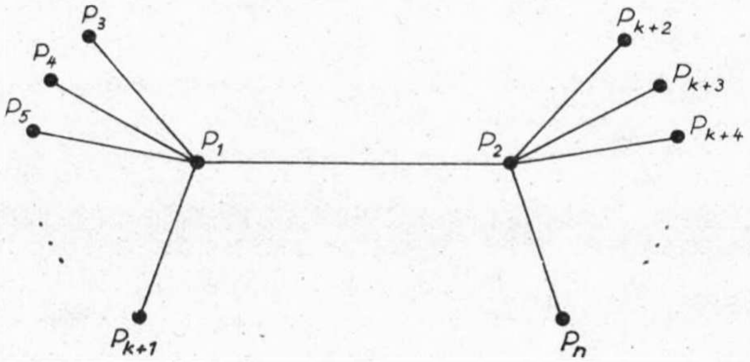


Fig. 9

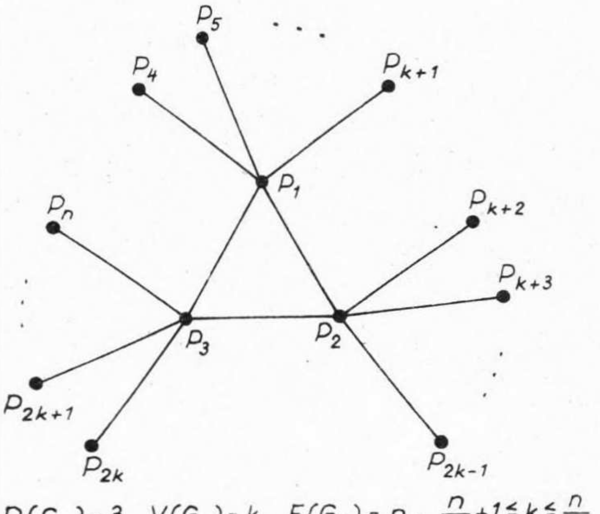
We prove now the following

THEOREM 4. *If* $\frac{n}{s+1}+s-1\leqq k\leqq\frac{n}{s}+s-2$ *where* $s=1,2,3,\dots,\left[\sqrt[3]{\frac{n}{2}}\right]$ *then* $F_3(n,k)=n+\binom{s}{2}-1$.

PROOF OF THEOREM 4. The case $s=1$ has been settled above. Let us consider first the case $s=2$. Suppose G_n would be a tree of diameter 3 and $V(G_n)=k\leqq\frac{n}{2}$, and let P_1 be an endpoint of G_n (such a point exists as every tree has at least two endpoints). Let P_2 denote the single point connected with P_1 by an edge, and let $P_3, \ldots, P_l$ be all the other points connected with P_2; as $V(G_n)=k$ we have $l \leqq k+1$. The remaining $n-k-1 \geqq k-1 \geqq l-2$ points have to be connected with one of the points $P_3, \ldots, P_l$ because otherwise it would be impossible to reach them from P_1 by a path of length $\leqq 3$. But they can not be all connected with the same point P_j $(3 \leqq j \leqq l)$ because this point would have valency $>k$. Let P_r and P_s be two points $(l<r<s \leqq n)$ such that P_r is connected with P_i and P_s with P_j $(3 \leqq i<j \leqq l)$. Then the (unique) path from P_r to P_s has length 4; this contradiction shows that $F_3(n, k) \geqq n$ for $k \leqq \frac{n}{2}$.

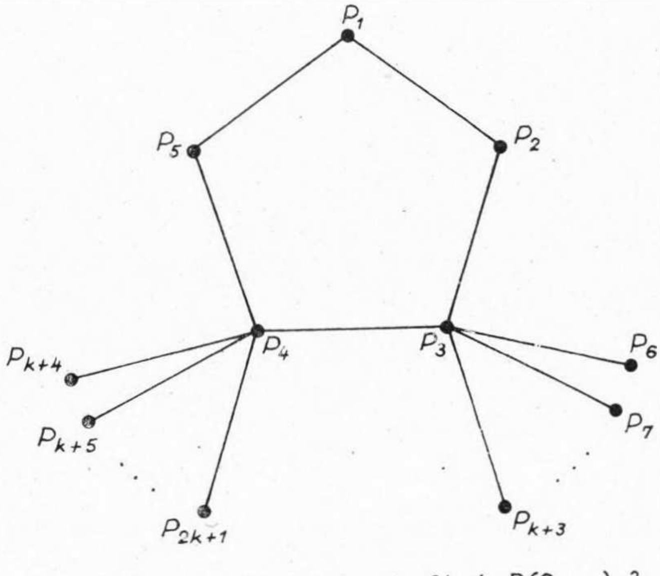
On the other hand Fig. 10 shows a graph G_n with $V(G_n)=k$ where $\frac{n}{3}+1 \leqq k \leqq \frac{n}{2}$ which has diameter 3 and contains exactly one cycle (a triangle) and thus $E(G_n)=n$. This completes the proof of the fact that $F_3(n, k)=n$ for $\frac{n}{3}+1 \leqq k \leqq \frac{n}{2}$.

Note that for $n=2k+1$ there is another extremal graph G_{2k+1} of diameter 3, for which $V(G_{2k+1})=k$ and $E(G_{2k+1})=2k+1$, shown by Fig. 11.



$D(G_n)=3,\ V(G_n)=k,\ E(G_n)=n,\ \frac{n}{3}+1 \leqq k \leqq \frac{n}{2}$

Fig. 10



$V(G_{2k+1})=k,\ E(G_{2k+1})=2k+1,\ D(G_{2k+1})=3$

Fig. 11

Now we pass to the case $s \geqq 3$.

Let G_n be a graph with $V(G_n)=k$ $\left(\frac{n}{s+1}+s-1 \leqq k \leqq \frac{n}{s}+s-2;\ s<\sqrt[3]{\frac{n}{2}}\right)$ and $D(G_n)=3$. Let $X_1, \ldots, X_l$ be the endpoints of G_n. As the remaining $n-l$ points all have valency $\geqq 2$, and at least one among them has valency k, we have

$$E(G_n) \geqq \frac{1}{2}\left(l+k+2(n-l-1)\right) = n-\frac{l}{2}+\frac{k}{2}-1.$$

Now if $E(G_n) \geqq n + \binom{s}{2} - 1$, we have nothing to prove; if however $E(G_n) < n + \binom{s}{2} - 1$ we get

$$l > k - s(s-1) \geqq s - 1$$

thus $l \geqq 2$. Let $Y_1, \dots, Y_v$ denote those vertices of G_n which are connected with at least one X_j $(1 \leqq j \leqq l)$.

Clearly Y_i and Y_j are connected by an edge $(1 \leqq i < j \leqq v)$ because otherwise there would not exist a path of length 3 connecting the X_h-s. Thus it is sufficient to consider the case $v \leqq s$, because every connected graph G_n containing a complete $s+1$-graph has at least $n - 1 + \binom{s}{2}$ edges. Let us suppose therefore that $v \leqq s$.

We prove first that $v \geqq s$. Let the endpoint X_1 be connected to Y_1. Let $Z_1, \dots Z_r$ denote all the points connected with Y_1 which are not endpoints of G_n. As every point of G_n can be reached from X_1 by a path of length $\leqq 3$, if Y_1 is connected with p endpoints then we have $\sum_{h=1}^{r} v(Z_h) \geqq n - p - 1$ thus

$$E(G_n) \geqq \frac{1}{2}(n - p - 1 + p + r + l + 2(n - l - r - 1)) = \frac{3n-3}{2} - \frac{r+l}{2}$$

thus in case $E(G_n) < n + \binom{s}{2} - 1$ we get

$$l \geqq n - r - s(s-1).$$

As however each Y_j has valency $\leqq k$, it can be connected to at most k of the X_i-s, and Y_1 only to $k - r$ X_i-s; thus

$$(v-1)k + k - r \geqq n - r - s(s-1)$$

and therefore, in view of $s \leqq \sqrt[3]{\frac{n}{2}}$, we obtain $v > s - 1$ i.e. $v \geqq s$. Thus we have only to consider the case $v = s$. Now if $v = s$ there exist in G_n at least s points which are not connected to any of the Y_j-s because these have valencies $\leqq k$ and thus the total number of points connected with them is $\leqq s(k - (s-1)) \leqq n - s$. Let W be such a point.

Now clearly W has to be connected with each X_h by a path of length 3 and therefore with each Y_j by a path of length 2. Let $U_1, \dots, U_t$ be the points connected with W, then each Y_j is connected with some U_z Thus it follows

$$E(G_n) \geqq \frac{1}{2}(2l + s(s-1) + 2s + 2t + 2(n - l - s - t - 1)) =$$

$$= l + \binom{s}{2} + s + t + n - l - s - t - 1 = n + \binom{s}{2} - 1.$$

Thus $F_3(n, k) \geqq n + \binom{s}{2} - 1$. On the other hand consider the graph G_n of the following structure: let us take a complete graph G_{s+1} having $s+1$ points, and connect each of these points except one with $k-s$ endpoints, and the last with $n-s(k-s)-$ $-(s+1)$ points. (Clearly $0 \leqq n-s(k-s)-(s+1) \leqq k-s$).

Thus we obtain a graph G_n with $V(G_n)=k$, $D(G_n)=3$ and $E(G_n)=n+\binom{s}{2}-1$. This completes the proof of Theorem 4.

Let us consider now $F_4(n, k)$. Clearly

$$F_4(n, k)=n-1 \quad \text{if} \quad k \geqq \sqrt{n-1}.$$

This can be seen as follows. Fig. 12 exhibits a tree of diameter 4 showing that $F_4(k^2+1, k)=k^2$

Clearly if $(k-1)^2+1<n<k^2+1$, we obtain a graph G_n exhibiting $F_4(n, k)=$ $=n-1$ by omitting from the graph on Fig. 12 k^2+1-n endpoints. We shall prove now

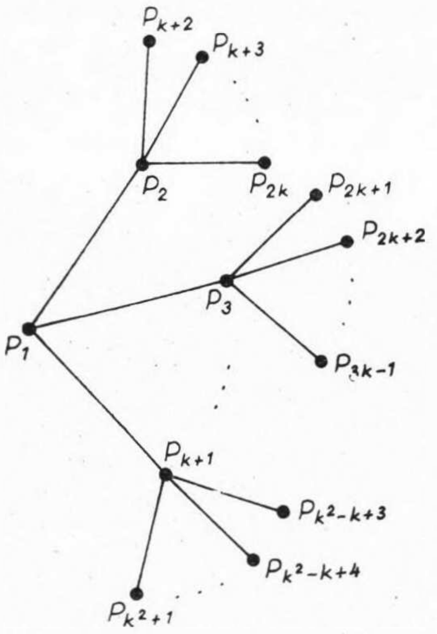


Fig. 12

THEOREM 5.

$$F_4(k^2+2, k) \geqq k^2+1+\frac{1}{2}\sqrt[4]{k} \qquad (k=2, 3, \ldots).$$

PROOF OF THEOREM 5. Let G_{k^2+2} be an extremal graph i.e. one which has k^2+2 points, diameter 4, satisfies the condition $V(G_{k^2+2})=k$ and has $F_4(k^2+2, k)$ edges.

Let $X_1, \ldots, X_m$ be the points of G_{k^2+2} having valency $\geqq 2$, and let G_m^* be the subgraph of G_{k^2+2} spanned by these points. We assert that each point X_i has the valency $\geqq 2$ in G_m^* too. Suppose that X_1 is an endpoint of G_m^*, and that X_2 is the only point of G_m^* to which X_1 is connected. Clearly X_1 is connected with at least one endpoint Y_1 of G_{k^2+2} because it has valency $\geqq 2$ in G_{k^2+2}, thus it is connected with some point of G_{k^2+2} different from X_2 and this point cannot be in G_m^* and thus is an endpoint of G_{k^2+2}. Every point of G_{k^2+2} can be reached by supposition from Y_1 by a path of length $\leqq 4$. However the number of points which can be reached from Y_1 by such a path is clearly

$$\leqq 2k-1+(k-1)^2=k^2$$

which is a contradiction. Thus in G_m^* each point has valency $\geqq 2$. As the diameter of G_m^* is $\leqq 4$, it follows from (1. 1) that G_m^* contains at least one point of valency $\sqrt[4]{m-1}$; thus the number of edges of G_m^* exceeds $(m-1)+\frac{1}{2}(\sqrt[4]{m-1})$. Each point in G_m^* can be connected with at most $k-2$ endpoints of G_{k^2+2} thus $k^2+2 \leqq m+$ $+m(k-2)=m(k-1)$ and therefore $m \geqq \dfrac{k^2+2}{k-1} \geqq k+1$; thus

$$E(G_{k^2+2}) \geqq k^2+1+\frac{1}{2}(\sqrt[4]{m-1}) \geqq k^2+1+\frac{1}{2}\sqrt[4]{k}.$$

Thus Theorem 5 is proved.

Note that the statement of Theorem 5 is trivial for $k \leqq 16$, because it states only what we know already that if $D(G_{k^2+2})=4$ then G_{k^2+2} can not be a tree.

To get an upper estimate for $F_4(k^2+2, k)-k^2$ consider the following graph. Take a graph G_{k+5} with $V(G_{k+5})=k$, $D(G_{k+5})=2$ and $E(G_{k+5})=2k+6$; such a graph exists according to Theorem 2 if $k \geqq 8$ (see Fig. 7 with $l=5$). This graph has $k+2$ points of valency 2. Connect k out of these points with $k-2$ new points each and one with $k-3$ new points. Thus we get a graph G_n with $n=k^2+2$ points, such that $V(G_n)=k$, $D(G_n)=4$ and $E(G_n)=k^2+k+3$. Thus

$$F_4(k^2+2, k) \leqq k^2+k+3.$$

§ 4. Some further Remarks and Unsolved Problems

First we formulate some general principles of construction which were implicitely used above.

If G_n is a graph of diameter d, and such that $V(G_n)=k$, then if G_n is not regular, we may construct from G_n a graph G_N of order $N=n+kn-E(G_n)$ with $V(G_N)=k$ and diameter $d+2$, by connecting each vertex P_i of G_n which has valency $v(P_i)<k$ with $k-v(P_i)$ new points. Thus

$$F_{d+2}(n+kn-2F_d(n, k), k) \leqq kn-F_d(n, k). \tag{4. 1}$$

For instance we have shown that $F_2(n, n-5)=2n-4$. It follows immediately from (4. 1) that

$$F_4(n^2-8n+8, n-5) \leqq n^2-7n+4.$$

Notice that for each value of d, the extremal graphs G_n with $V(G_n)=k$, $D(G_n)=d$ and having a minimal number of edges, are trees if k is sufficiently large, $k \geqq U_d(n)$ say.

We have implicitely shown that

$$U_2(n)=n-1 \tag{4. 2}$$

$$U_3(n)=\frac{n}{2} \tag{4. 3}$$

$$U_4(n)=\sqrt{n-1}. \tag{4. 4}$$

It can be shown that

$$U_5(n)=\frac{1+\sqrt{2n-3}}{2} \tag{4. 5}$$

further that for any fixed $s \geqq 3$ and $n \to \infty$

$$U_{2s}(n) \sim \sqrt[s]{n} \tag{4. 6}$$

and

$$U_{2s+1}(n) \sim \sqrt[s]{\frac{n}{2}}. \tag{4. 7}$$

The extremal tree of diameter $2s$ has a center, while the extremal tree of diameter $2s+1$ has a central edge.

Notice that if k decreases by one below the critical value $U_d(n)$, i.e. to $U_d(n)-1$, there is a considerable increase in the value of $F_d(n, k)$ if d is even, but not if d is odd. As a matter of fact

$$F_2(n, U_2(n)-1)-F_2(n, U_2(n)) = (2n-4)-(n-1) = n-3$$

$$F_3(2k+1, k)-F_3(2k+1, k+1) = (2k+1)-2k = 1$$

and

$$F_3(2k+2, k)-F_3(2k+2, k+1) = (2k+3)-(2k+1) = 2$$

further as proved by Theorem 5

$$F_4(k^2+2, k)-F_4(k^2+1, k) \geqq \frac{1}{2}\sqrt[4]{k}.$$

The situation is similar for $d>4$.

We call attention to the following problems, left open in this paper:

PROBLEM 1. Is the graph of Theorem 1 extremal in the sense that among all graphs with n vertices and not containing any cycle of length 4 does it have the maximal number of edges? (We have proved only that it is asymptotically extremal.)

We can prove the following result, which is connected with Problem 1.

THEOREM 6. If G_n is a graph in which any two points are connected by a path of length 2 and which does not contain any cycle of length 4, then $n=2k+1$ and G_n consists of k triangles which have one common vertex (see Fig. 13).

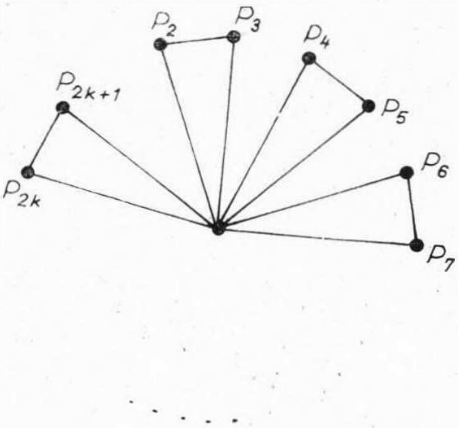


Fig. 13

PROOF OF THEOREM 6. Let G_n be a graph with the required properties. Let P_1 be a point of G_n having maximal valency. If P_1 is connected with all the remaining points of G_n then evidently these have to be connected by pairs, and G_n is of the type described in Theorem 6. Thus we may suppose that G_n contains at least one point P_2 which is not connected with P_1. It is easy to see that in this case $V(P_2)=V(P_1)$.

As a matter of fact there is a point P_3 in G_n which is connected with both P_1 and P_2. As there must be a path of length 2 between P_1 and P_3 there is a point P_4 which is connected with both P_1 and P_3. As there has to be a path of length 2 between P_2 and P_3, there is a point P_5 connected with both P_2 and P_3, which is clearly different from P_1, P_2, P_3 and P_4. Let $Q_1, Q_2, \ldots, Q_{k-2}$ be the remaining points (besides P_3 and P_4) which are connected with P_1. Clearly P_2 and P_5 are not among the Q_i; we have $k \geqq 4$ because $v(P_3) \geqq 4$ and by supposition P_1 has the maximal valency.

Now from each of the points Q_i there is a path of length 2 to P_2; thus for each Q_i $(i=1, 2, \ldots, k-2)$ there exists a point R_i which is connected with both Q_i and P_2. Clearly $R_i \neq R_j$ if $i \neq j$ because otherwise G_n would contain the cycle $P_1Q_iR_iQ_j$. Further R_i is different from P_3 because if R_i would be identical with P_3 G_n would contain the cycle $P_1Q_iP_3P_4$. Finally R_i is different from P_5 because otherwise G_n

would contain the cycle $P_1Q_iP_5P_3$. Thus $v(P_2) \geqq k$ and as $k = V(G_n)$ we obtain $v(P_2) = k = v(P_1)$. Thus any point of G_n which is not connected with P_1 has the valency $k = v(P_1)$. Repeating the same argument with P_2 instead of P_1 it follows that $v(Q_i) = k$ $(i = 1, 2, \ldots, k-2)$. As P_3 is not connected with Q_1 (because otherwise G_n would contain the cycle $P_1\, Q_1\, P_3\, P_4$) repeating the same argument for Q instead of P_1 it follows that $v(P_3) = k$. Thus the graph G_n is regular.

Now if $V(P_i) = k$ $(i = 1, 2, \ldots, n)$ and G_n does not contain a cycle of length 4 and between any two points there is a path of length 2, then clearly if S_i denotes the set of points connected with P_i then the sets S_i and S_j have exactly one point in common, and for any two points P_i and P_j $(j \neq i)$ there is exactly one point P_h such that S_h contains both P_i and P_j. Thus if we define the sets of points S_i as lines we obtain a finite plane geometry, with $k = P+1$ points on a line, and thus having $n = P^2 + P + 1$ points. But then in this geometry there would exist a one-to-one mapping beween points and lines such that no line contains the point corresponding to it, and such a mapping is known [5] to be impossible. This proves Theorem 6.

PROBLEM 2. To determine the exact value of $F_2(n, k)$ for $k < \frac{n}{2}$, or at least the asymptotic value of $F_2(n, [nc])$ with $0 < c < \frac{1}{2}$.

PROBLEM 3. Is the lower estimate in Theorem 3 asymptotically best possible, i.e. do there exist for each $d \geqq 3$ a sequence of graphs G_n $(n \to \infty)$ with $V(G_n) = k \sim cn^{\frac{1}{d-1}}$ where $c > 0$ is a constant, $D(G_n) = d$ and $E(G_n) \sim \frac{n^2}{k^{d-1}} \sim \frac{n}{c^{d-1}}$?

PROBLEM 4. Determine asymptotically $F_4(k^2+2, k) - k^2$.

Problems similar to those considered in this paper can be asked for directed graphs. We hope to return to these problems in an other paper.

(Received February 1, 1966.)

REFERENCES

[1] ERDŐS, P.—RÉNYI, A.: On a problem in the theory of graphs (in Hungarian, with English and Russian summaries), *Publ. Math. Inst. Hung. Acad. Sci.* **7/A** (1962) 623—641.

[2] ERDŐS, P.: On sequences of integers no one of which divides the product of two others and on some related problems, *Mitteilungen des Forschungsinstitutes für Math. und Mechanik, Tomsk,* **2** (1938) 74—82.

[3] REIMAN, I.: Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hung.* **9** (1958) 269—278.

[4] HOFFMAN, A. J.—SINGLETON, R. R.: On Moore graphs with diameter 2 and 3, *IBM Journal of Research and Development* **4** (1960) 497—504.

[5] BAER, R.: Polarities in finite projective planes, *Bulletin of the American Math. Soc.* **52** (1946) 77—93.