ON THE STRUCTURE OF LINEAR GRAPHS

BY

P. ERDÖS

ABSTRACT

Denote by $G(n; m)$ a graph of n vertices and m edges. We prove that every $G(n; [n^2/4] + 1)$ contains a circuit of l edges for every $3 \leqq l < c_2 n$, also that every $G(n; [n^2/4] + 1)$ contains a $k_e(u_n, u_n)$ with $u_n = [c_1 \log n]$ (for the definition of $k_e(u_n, u_n)$ see the introduction). Finally for $t > t_0$ every $G(n; [tn^{3/2}])$ contains a circuit of $2l$ edges for $2 \leqq l < c_3 t^2$.

$G(n; m)$ will denote a graph of n vertices and m edges, $K(p)$ will denote the complete graph of p vertices, and $K(p, p)$ will denote the complete bipartite graph of $2p$ vertices. More generally $K(p_1, \cdots, p_r)$ denotes the r-chromatic graph where there are p_i vertices of the i-th color and any two vertices of different color are adjacent. $K_e(p_1, \cdots, p_r)$, $p_1 \leqq p_2 \leqq \cdots \leqq p_r$, will denote a $K(p_1, \cdots, p_r)$ where two vertices of the first color are adjacent, i.e. $K_e(p_1, \cdots, p_r)$ is a $K(p_1, \cdots, p_r)$ with an extra edge. The vertices of G will be denoted by $x, x_1, y, \cdots$; the edge connecting x and y will be denoted by (x, y). $(G - x_1 - \cdots - x_r)$ denotes the graph G from which the vertices $x_1, \cdots, x_r$ and all edges which are incident to them have been deleted. $v(x)$, the valency of x, is the number of edges adjacent to x. C_l will denote a circuit having l edges. $c_1, c_2, \cdots$ denote suitable positive absolute constants. $[t]$ is the greatest integer not exceeding t.

A special case of a well known theorem of Turán [1] states that every $G(n; [n^2/4] + 1)$ contains a $K(3)$ (i.e. a triangle). Dirac and I observed (independently) that every $G(n; [n^2/4] + 1)$ contains for every $4 \leqq k \leqq n$ a subgraph $G(k; [k^2/4] + 1)$ and in fact Dirac proved a more general theorem [2].

In the present paper we continue the investigation of the structure of the graphs $G(n; [n^2/4] + 1)$ and we are going to prove the following theorems:

THEOREM 1. *Put* $[c_1 \log n] = u_n$. *Every* $G(n; [n^2/4] + 1)$ contains a $K_e(u_n, u_n)$.

REMARK. The structure of $K_e(u_n, u_n)$ is clearly uniquely determined. It is the $G(2u_n; u_n^2 + 1)$ which contains a $K(u_n, u_n)$ as a subgraph.

THEOREM 2. *Every* $G(n; [n^2/4] + 1)$ *contains a* C_l *for every* $3 \leqq l \leqq c_2 n$.

THEOREM 3. *Let* $t > t_0$, *then every* $G(n; [tn^{3/2}])$ *contains a* C_{2l} *for every* $2 \leqq l < c_3 t^2$.

Received September 28, 1963.

This work was done while the author received support from the National Science Foundation, N.S.F. G.88.

Apart from the value of c_1 Theorem 1 is best possible. In fact we can show the following

THEOREM 4. *To every $\varepsilon > 0$ there is a $c(\varepsilon)$ so that for every n there is a $G(n; [\binom{n}{2}(1-\varepsilon)])$ which does not contain a $K([c(\varepsilon)\log n], [c(\varepsilon)\log n])$.*

We suppress the proof of Theorem 4 since it uses the methods used in [3]. A theorem of A. H. Stone and myself [4] implies that every $G(n; [\varepsilon n^2])$ contains a $K([c_1(\varepsilon)\log n], [c_1(\varepsilon)\log n])$. The exact determination of $c(\varepsilon)$ and $c_1(\varepsilon)$ seems difficult.

I would expect that the exact determination of c_2 in Theorem 2 will be difficult.

Theorem 3 is best possible in the sense that E. Klein [5] showed that there is a $G(n; [c_4 n^2])$ which contains no C_4. For $t > t_0$ perhaps every $G(n; [tn^{3/2}])$ contains a C_{2l} for every $2 \leqq l < c_5 t n^{1/2}$; if true, then apart from the value of c_5 this is easily seen to be best possible.

By the same method as used in the proof of Theorem 1 we can prove

THEOREM 5. *To every k there is an $n_0 = n_0(k)$ and a c_k so that, for $n > n_0$, $G(n; [n^2/4] + k)$ always contains a $K([c_k \log n], [c_k \log n])$ and k further edges.*

We suppress the proof of Theorem 5. Put $r_k = [c_k \log n]$. For $k > 1$ the structure of our $G(2r_k; r_k^2 + k)$ is of course not uniquely determined. Perhaps the following result holds: Let $n \geqq 8$. Then every $G(n; [n^2/4] + n - 1)$ contains a $K([c \log n]$, $[c \log n])$ and two edges which have no vertex in common and all four vertices of which have the same color. It is easy to see that a $G(n; [n^2/4] + n - 2)$ does not have to have this property. To see this consider a $K([n/2], [(n+1)/2])$ where further one vertex of each color is adjacent to all the vertices of our graph i.e., the vertices of our $G(n; [n^2/4] + n - 2)$ are $x_1, \cdots, x_k; y_1, \cdots, y_l$ $k = [n/2]$, $l = [(n+1)/2]$ and its edges are

$$(x_i, y_j);\ 1 \leqq i \leqq k,\ 1 \leqq j \leqq l \text{ and } (x_1, x_i), (y_1, y_j);\ 2 \leqq i \leqq k,\ 2 \leqq j \leqq l.$$

Put

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

Turán proved that every $G(n; m(n,p))$ contains a $K(p)$ and Dirac and I [2] observed (independently) that it contains a $K(p+1)$ from which one edge is missing . By very much more complicated methods I can prove that for $n > n_0(p,k)$ $G(n; m(n,p))$ contains a p chromatic subgraph $K(k, \cdots, k)$ and one further edge (i. e., a $K_e(k, \cdots, k)$); for $p = 2$ this is a weakened form of Theorem 1.

Now we prove Theorem 1. First we need two Lemmas.

LEMMA 1. Every $G(n; m)$ contains a subgraph $G(N, M)$ every vertex of which has valency greater than $[m/n]$. Further

$$M \geqq m - (n - N) \left[\frac{m}{n}\right] \tag{1}$$

(The Lemma of course means that every vertex of $G(N,M)$ has valency in $G(N,M)$ greater than $[m/n]$).

If every vertex of $G(n,m)$ has valency $> [n/m]$, there is nothing to prove. Hence we can assume that $G(n,m)$ has a vertex x_1 of valency $\leqq [m/n]$. If $G(n;m) - x_1$ has a vertex x_2 with $v(x_2) \leqq [m/n]$ we consider $G(n;m) - x_1 - x_2$. We repeat this process and obtain a sequence of vertices $x_1, \cdots, x_k$ so that the valency of x_i in $(G(n;m) - x_1 - \cdots - x_{i-1})$ is $\leqq [m/n]$ for every $1 \leqq i \leqq k-1$, but every vertex of

$$(G(n;m) - x_1 - \cdots - x_k) = G(N;M) \tag{2}$$

has valency $> [m/n]$.

Clearly $M > 0$ for otherwise, since $(G(n;m) - x_1 - \cdots - x_{n-1})$ has only one vertex and thus no edges, we can put in (2) $k \leqq n-1$ and by our construction we would have

$$m \leqq (n-1) \left[\frac{m}{n}\right] < m$$

an evident contradiction. Further by our construction $(k = n - N)$

$$M \geqq m - (n-N) \left[\frac{m}{n}\right]$$

which proves (1), and the proof of Lemma 1 is complete.

LEMMA 2. *Let $m > [n^2/4]$. Then every $G(n;m)$ contains a $K_e(2,k)$ where $k = [c_5 n]$.*

Lemma 2 is known [6].

Now we can prove Theorem 1. In fact we shall prove the stronger statement: To every $\varepsilon > 0$ there is a $c_1 = c_1(\varepsilon)$ so that every $G(n;[n^2/4]+1)$ contains a $K_e([c_1 \log n], [n^{1-\varepsilon}])$.

By Lemma 1 our $G(n;[n^2/4]+1)$ contains a subgraph $G(N,M)$ every vertex of which has valency $> [\frac{[n^2/4]+1}{n}] = [n/4]$. Further (1) implies by a simple computation

$$M \geqq \left[\frac{n^2}{4}\right] + 1 - (n-N)\left[\frac{n}{4}\right] > \left[\frac{N^2}{4}\right]. \tag{2}$$

Further since every vertex of $G(N,M)$ has valency $> [n/4]$ we have

$$N > \frac{n}{4}. \tag{3}$$

By (2) Lemma 2 can be applied to $G(N,M)$ and by Lemma 2 and (3) we obtain that $G(N,M)$ contains a $K_e(2,k)$ with $k = [c_5 n/4]$. Let the vertices of our $K_e(2,k)$ be (we choose $c_5 < 1/3$)

$$x_1, x_2; y_1, \cdots, y_k, \qquad k = \left[\frac{c_5 n}{4}\right] < \left[\frac{n}{8}\right] - 1 \tag{4}$$

Denote by $z_1, \cdots, z_r$ the other vertices of $G(N, M)$. Each y has by Lemma 1 valency $> [n/4]$ (in $G(N, M)$), hence each $y_i, 1 \leqq i \leqq k$ is connected with more than

$$\frac{n}{4} - 2 - k + 1 > \frac{n}{8} \tag{5}$$

z's. ((5) follows immediately from (4) since the number of x's and y's is $k + 2 < [n/8] + 1$ and in the worst case y_i is connected with all of them).

Let $z_j^{(i)}, 1 \leqq j \leqq t_i, t_i > n/8$, be the z's adjacent to y_i. Form all the $(u_n - 2)$-tuples ($u_n = [c_1 \log n]$ of Theorem 1) of these vertices for each $i, 1 \leqq i \leqq k = [c_5 n/4]$. By a simple computation we obtain (we use $\binom{a}{b} > (a/b)^b$)

$$\sum_{i=1}^{k} \binom{t_i}{u_n - 2} \geqq \frac{c_5 n}{4} \binom{[n/8] + 1}{u_n - 2} > \frac{c_5 n}{4} \left(\frac{n}{8(u_n - 2)}\right)^{u_n - 2} \tag{6}$$

Further trivially

$$\binom{n}{u_n - 2} < \frac{n^{u_n - 2}}{(u_n - 2)!} < \frac{n^{u_n - 2} e^{u_n - 2}}{(u_n - 2)^{u_n - 2}} < \left(\frac{3n}{u_n - 2}\right)^{u_n - 2} \tag{7}$$

Hence from (6) and (7)

$$\sum_{i=1}^{k} \binom{t_i}{u_n - 2} > \frac{c_5 n}{4} \binom{n}{u_n - 2} \frac{1}{24^{u_n - 2}} > n^{1 - \varepsilon} \binom{n}{u_n - 2} \tag{8}$$

for every $\varepsilon > 0$ if $c_1 = c_1(\varepsilon)$ is sufficiently small. The number of the z's is clearly less than n, hence the number of the $(u_n - 2)$-tuples formed from z's is less than $\binom{n}{u_n - 2}$. Thus from (8) there is a $(u_n - 2)$-tuple which occurs more than $n^{1-\varepsilon}$ times—in other words there is a set of $u_n - 2$ z's which are adjacent to the same $[n^{1-\varepsilon}]$ y's. If we adjoin to these z's x_1 and x_2 (which are adjacent and are adjacent to all y's) we obtain that $G(N; M)$ and hence our $G(n; [n^2/4] + 1)$ contains a $K_e(u_n, n^{1-\varepsilon})$ for every $\varepsilon > 0$ if $c_1 = c_1(\varepsilon)$ is sufficiently small. This completes the proof of our assertion and hence Theorem 1 is proved.

Proof of Theorem 2. As in the proof of Theorem 1 our $G(n; [n^2/4] + 1)$ contains a $K_e(2, [c_5 n/4])$, $c_5 < 1/3$, having the vertices $x_1, x_2, y_1, \cdots, y_k$, $k = [c_5 n/4]$. Each of the k vertices $y_1, \cdots, y_k$ are adjacent to more than $n/8$ z's (we use the notations of Theorem 1). Consider now the bipartite graph whose vertices are $y_1, \cdots, y_k$; $z_1, \cdots, z_r$ and whose edges are the edges (y_i, z_j) of $G(n; m)$. This bipartite graph has fewer than n vertices and more than

$$\frac{n}{8}\left[\frac{c_5 n}{4}\right] = c_6 n^2$$

edges. Hence by a theorem of Gallai and myself [7] it has a path of length $c_2 n$ (the length of a path is the number of its edges). Since our graph is bipartite every second of its vertices is a y. Now since x_1 and x_2 are adjacent and they are adjacent to each of the y's we immediately obtain that our $G(n; [n^2/4] + 1)$ contains a C_l for each $3 \leqq k \leqq [c_2 n]$, which proves Theorem 2.

Proof of Theorem 3. By Lemma 1 $G(n; [tn^{3/2}])$ contains a subgraph $G(N; M)$ every vertex of which has valency $\geqq [tn^{1/2}]$. Let x be one such vertex and let $y_1, \cdots, y_k$, $k = \frac{1}{2}[tn^{1/2}]$ be some of the vertices adjacent to x and denote by $z_1, \cdots$ the other vertices of $G(N, M)$. Every y has valency $\geqq [tn^{1/2}]$, thus since the number of y's is $\frac{1}{2}[tn^{1/2}]$ there are at least $\frac{1}{2}[tn^{1/2}]$ z's adjacent to each y. Hence the bipartite graph whose vertices are $y_1, \cdots, y_k$: $z_1, \cdots$ and whose edges are the edges (y_i, z_j) of $G(n, m)$ has at least

$$k\,\tfrac{1}{2}[tn^{1/2}] = \tfrac{1}{4}[tn^{1/2}]^2 > \frac{t^2}{8} n$$

edges. The number of its vertices is clearly $< n$. Thus by the theorem of Gallai and myself [7] it has a path of length $> 2c_3 t^2$ and as in the proof of Theorem 2 every second vertex of this graph is a y. Since x is adjacent to every y this path together with the vertex x gives the required circuits C_{2l}, $2 \leqq l \leqq c_3 t^2$, which proves Theorem 3.

REFERENCES

1. Turán, P., 1941, *Mat. Lapok*, **48**, 436–452 (Hungarian), see also Turan, P., 1955, On the theory of graphs, *Coll. Math.*, **3**, 19–30.

2. Dirac, G., will appear in *Acta Math. Acad. Sci. Hungar*

3. Erdös, P., 1947, Some remarks on the theory of graphs, *Bull. Amer, Math. Soc.*, **53**, 292–294; see also Erdös, P. and Renyi, A., 1960, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad*, **5**, 17–61.

4. Erdös, P. and Stone, A. H., 1946, On the structure of linear graphs, *Bull. Amer. Math. Soc.*, **52**, 1087–1091.

5. Erdös, P., 1938, On sequences of integers no one of which divides the product of two others and on some related problems, *Izv. Nauk. Inst. Mat. Mech. Tomsk.*, **2**, 74–82.

6. Erdös, P., 1962, On the theorem of Rademacher-Turán, *Illinois J. of Math.*, **6**, 122–127.

7. Erdös, P. and Gallai, T., 1959, On maximal paths and circuits of graphs, *Acta. Math. Acad. Sci. Hungar.*, **10**, 337–356.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICHIGAN, U.S.A.