On the maximal number of disjoint circuits of a graph

By P. ERDŐS and L. PÓSA (Budapest)

Throughout this paper $G_k^{(n)}$ will denote a graph with n vertices and k edges where circuits consisting of two edges and loops (i. e. circuits of one edge) are not permitted and $\bar{G}_k^{(n)}$ will denote a graph of n vertices and k edges where loops and circuits with two edges are permitted. $\nu(G)$ (respectively $\nu(\bar{G})$) will denote the number of edges of G (respectively $\bar{G}$).

If $x_1, x_2, \dots, x_k$ are some of the vertices of G, then $(G - x_1 - \dots - x_k)$ will denote the graph which we obtain from G by omitting the vertices $x_1, \dots, x_k$ and all the edges incident to them. By $G(x_1, \dots, x_k)$ we denote the subgraph of G spanned by the vertices $x_1, \dots, x_k$. The valency of a vertex $x - v(x) -$ will denote the number of edges incident to it. (A loop is counted doubly.) The edge connecting x_1 and x_2 will be denoted by $[x_1, x_2]$, edges will sometimes be denoted by $e_1, e_2, \dots$. $(x_1, x_2, \dots x_k)$ will denote the circuit having the edges $[x_1, x_2], \dots, [x_{k-1}, x_k], [x_k\, x_1]$.

A set of edges is called independent if no two of them have a common vertex. A set of circuits will be called independent if no two of them have a common vertex. They will be called weakly independent if no two of them have a common edge.

In a previous paper ERDŐS and GALLAI [1] proved that every

$$(1)\qquad G_{l+1}^{(n)} \text{ where } l = \max\left[\binom{2k-1}{2},\ (k-1)n - (k-1)^2 + \binom{k-1}{2}\right]$$

contains k independent edges.

In the present paper we shall investigate the following question of Turanian type (see 1): how many edges are needed that a graph should have to contain k independent or weakly independent circuits? Put

$$f(n,k) = (2k-1)n - 2k^2 + k.$$

Our principal result will be that for $n \geqq n_0(k)$, $k > 1$ every $G_{f(n,k)}^{(n)}$ contains k independent circuits except if it contains $2k-1$ vertices of valency $n-1$ (its structure is then uniquely determined). If $k=1$ trivially every $G_n^{(n)}$ contains a circuit, but there are of course graphs $G_{n-1}^{(n)}$ where no vertex has valency $n-1$ and the graph nevertheless contains no circuit. Thus the restriction $k > 1$ is necessary.

Clearly $n_0(k) \geqq 3k$ (since a circuit contains at least three vertices). For $k=2$ and $k=3$ $n_0(k) = 3k$, but in general $n_0(k) > 3k$, but we will prove $n_0(k) \leqq 24k$.

Perhaps the following result analogous to (1) holds: Every

$$(2) \qquad G^{(n)}_{l+1} \text{ with } l = \max\left[\binom{3k-1}{2} + n - 3k + 2,\ (2k-1)n - 2k^2 + k\right]$$

contains k independent circuits.

Denote by $g(k)$ the smallest integer so that every $\overline{G}^{(n)}_{n+g(k)}$ contains k weakly independent circuits. We will show that $g(2)=4$ and that for every k

$$c_1 k \log k < g(k) < c_2 k \log k$$

where c_1 and c_2 are suitable absolute constants (the c's throughout this paper will denote suitable absolute constants). The exact determination of $g(k)$ seems to be a very difficult problem and we cannot even show that $g(k)/\log k$ tends to a constant. Further we do not know the value of $g(3)$.

It is easy to see that $g(2) \geqq 4$ i. e. we will show that for every $n \geqq 6$ there exists a $G^{(n)}_{n+3}$ which does not contain two weakly independent circuits. To see this let the vertices of our graph be $x_1, \ldots, x_n$ and its edges

$$[x_i, x_j],\ 1 \leqq i \leqq 3 < j \leqq 6 \quad \text{and} \quad [x_k,\ x_{k+1}],\ 6 \leqq k \leqq n-1.$$

A simple argument shows that this $G^{(n)}_{n+3}$ does not contain two weakly independent circuits.

After completing our paper we found out that some of our results were known to G. Dirac but he published nothing on this subject. In particular he proved that for $n \geqq 6$ every $G^{(n)}_{3n-5}$ contains two independent circuits and that every $G^{(n)}_{n+4}$ contains two weakly independent circuits. He also proved that for $n \geqq 6$ every $G^{(n)}$ where the valency of every vertex is $\geqq 3$ and the valency of every vertex with at most one exception is $\geqq 4$ contains two independent circuits and conjectured that for $n \geqq 3k$ every $G^{(n)}$ which is $2k$-fold connected (i. e. which remains connected after the omission of any $2k-1$ of its vertices) contains k independent circuits.

Theorem 1. *Let $k > 1$, $n \geqq 24k$ then every $G^{(n)}_{f(n,k)}$ either contains k independent circuits or $2k-1$ vertices of valency $n-1$.*

Our Theorem clearly implies that for $n \geqq 24k$ every $G^{(n)}_{f(n,k)+1}$ contains k independent circuits (since a simple computation shows that a $G^{(n)}_{f(n,k)}$ which has $2k-1$ vertices of valency $n-1$ has all its other vertices of valency $2k-1$ and its structure is thus uniquely determined). $n \geqq 24k$ could easily be improved a great deal, but our method does not give any hope of best possible estimates.

Theorem 1. will be proved by a fairly complicated induction process and to make this as painless as possible we will restate Theorem 1. in a very much more complicated form but which will be more suitable for our induction process.

Theorem 1′. *Every $G^{(n)}_n$ contains a circuit ($k=1$). For $k > 1$ put*

$$g(n,k) = \begin{cases} f(n,k) + (24k-n)(k-1) & \text{for} \quad n \leqq 24k \\ f(n,k) & \text{for} \quad n \geqq 24k. \end{cases}$$

Then if $3k \leqq n \leqq 24k-1$ *every* $G^{(n)}_{g(n,k)}$ *contains k independent circuits, and if* $n \geqq 24k$ *and* $l_0 \geqq g(n,k)$ *then every* $G^{(n)}_{l_0}$[1]) *contains k independent circuits except only if* $l_0 = g(n,k)$ *and* $G^{(n)}_{l_0}$ *contains* $2k-1$ *vertices of valency* $n-1$ *and* $n-(2k-1)$ *vertices of valency* $2k-1$.

Since $g(n,k) = f(n,k)$ for $n \geqq 24k$ Theorem 1′ implies Theorem 1.

Theorem 1′ is trivial for $k=1$. It is also trivial for $k>1$ if $3k \leqq n \leqq 6k$ since by a simple calculation

$$g(n,k) \geqq \binom{n}{2}$$

and for $n \geqq 3k$ the complete graph contains k independent circuits.

First we prove two Lemmas.

Lemma 1. *Let* $n \geqq 6k$ *and assume that* $G^{(n)}$ *contains* $2k$ *vertices* $x_1, x_2, \ldots, x_{2k}$ *of valency* $v(x_i) \geqq n-k$ $(1 \leqq i \leqq 2k)$. *Then* $G^{(n)}$ *contains k independent quadrilaterals.*

Denote by $y_1, \ldots, y_{n-2k}$ the other $n-2k$ vertices of $G^{(n)}$. Consider a maxim a system of independent quadrilaterals of the form

$$(x_{2i-1}, y_{2i-1}, x_{2i}, y_{2i}), \; 1 \leqq i \leqq l, \; l \leqq k.$$

We shall show $l=k$. Assume $l<k$. Each of the vertices x_{2l+1} and x_{2l+2} are connected with at least $n-k$ vertices. Thus every vertex except possibly $2k$ vertices are connected with both x_{2l+1} and x_{2l+2}, i. e. these are at least $n-2k \geqq 4k$ of them which are connected with both x_{2l+1} and x_{2l+2}. Since $2k+2l \leqq 4k-2$ there are two further vertices y_{2l+1} and y_{2l+2} which are connected with both x_{2l+1} and x_{2l+2}. Thus the quadrilateral $(x_{2l+1}, y_{2l+1}, x_{2l+2}, y_{2l+2})$ is independent of the others, which contradicts our maximality assumption, which proves Lemma 1.

Lemma 2. *Let* $n \geqq 2k$ *and assume that every vertex of* $G^{(n)}$ *has valency* $\geqq 2k$, *then* $G^{(n)}$ *contains k independent edges.*

Lemma 2 can be proved from first principles in a few lines as follows:[2]) Let $e_i = [x_{2i-1}, x_{2i}]$, $1 \leqq i \leqq t$ be a maximal set of independent edges. Assume $t<k$ (otherwise there is nothing to prove). But then since $n \geqq 2k$ there are two vertices of $G^{(n)}$ y_1 and y_2 distinct from the x_i, $1 \leqq i \leqq 2t$. y_1 and y_2 can be joined only to the x_i $1 \leqq i \leqq 2t$ (by our maximality assumption), and by the same assumption if y_1 is connected to an endpoint of e_i $1 \leqq i \leqq t$ then y_2 can not be connected to the other endpoint. Thus $v(y_1)+v(y_2) \leqq 2t < 2k$ which contradicts $v(y_i) \geqq k$. This contradiction proved the Lemma.

Now we prove Theorem 1′ by induction. Let $k>1$ and assume that Theorem 1′ holds for $k-1$ and assume that it holds for every $6k<m<n$. (We already remarked that it trivially holds for $3k \leqq m \leqq 6k$). Then we shall prove it for n, and if we have succeeded in this the proof of Theorem 1′ and therefore Theorem 1 will be complete.

---

1) If $l_0 > \binom{n}{2}$ $G^{(n)}_{l_0}$ will denote the complete graph of n vertices.

2) This proof is due to G. Dirac (written communication).

Assume first that our $G^{(n)}_{g(n,k)}$ contains $2k$ vertices of valency $\geqq n-k$. Then by Lemma 1. our graph contains k independent quadrilaterals and thus Theorem 1′ is proved in this case.

Henceforth we can assume that G contains at most $2k-1$ vertices of valency $\geqq n-k$. If all the other vertices have valency $<2k$ then the number of edges of G is at most

$$\frac{(2k-1)(n-1)+(2k-1)(n-2k+1)}{2}=f(n,k)$$

and equality can occur only if G contains $2k-1$ vertices of valency $n-1$ (i. e. these vertices are connected with all the vertices of the graph), and G contains no other edge, otherwise it would contain another vertex of valency $\geqq 2k$). Thus the structure of our G is uniquely determined (G can have this structure only if $n \geqq 24k$) and Theorem 1′ is proved in this case too.

Therefore we can now assume that G has a vertex — say x_0 — of valency l' satisfying

$$2k \leqq l' < n-k.$$

Let $x_1, \ldots, x_{l'}$, $2k \leqq l' < n-k$ be the vertices of G connected with x_0 by an edge. Assume first that in the graph $G(x_1, \ldots, x_{l'})$ there is a vertex — say x_1 — of valency $<k$. It may be assumed that x_1 is not connected by an edge to any of the vertices $x_{r+1}, \ldots, x_{l'}$, where $r \leqq k$. Define the graph G_1 with $n-1$ vertices as follows: Omit the vertex x_0 and all the l' edges incident to it, and add the edges $[x_1, x_{r+i}]$ $1 \leqq i \leqq l'-r$ (i. e. x_1 is connected in G_1 to all the vertices to which x_0 is connected in G [except of course x_1]). Clearly

$$v(G_1) \geqq v(G)-k$$

or

$$v(G_1) \geqq g(n,k)-k \geqq g(n-1,k). \tag{3}$$

Thus by our induction hypothesis G_1 contains k independent circuits (from the first inequality of (3) it follows that $v(G_1) > f(n-1,k)$ thus G_1 cannot have $2k-1$ vertices of valency $n-1$ and $n-2k$ vertices of valency $2k-1$, i. e. the second alternative of Theorem 1′. is excluded). But then G must also contain k independent circuits. To see this let $C_1, \ldots, C_k$ be the k independent circuits of G_1, at most one of these circuits — say C_1 — contains one or two of the new edges $[x_1, x_{r+i}]$, $1 \leqq i \leqq l'-r$, (if none of these circuits contains any of these edges, then $C_1, \ldots, C_k$ are k independent circuits of G). If C_1 contains only one of the new edges — say $[x_1, x_{r+1}]$ — then we obtain C_1^* by omitting $[x_1; x_{r+1}]$ from C_1 and replacing it by $[x_0, x_1]$ and $[x_0\, x_{r+1}]$. If C_1 contains $[x_1, x_r]$ and $[x_1, x_{r+1}]$ then in C_1^* these are replaced by $[x_0, x_r]$ and $[x_0, x_{r+1}]$. In any case $C_1^*, \ldots, C_k$ are k independent circuits of G. Thus Theorem 1′ is proved in this case too.

Assume next that all the vertices of $G(x_1, \ldots, x_{l'})$ have valency $\geqq k$. Then (since $l' \geqq 2k$) by Lemma 2. $G(x_1, \ldots, x_{l'})$ contains k independent edges $e_i = [x_{2i-1}, x_{2i}]$, $1 \leqq i \leqq k$. Assume first that each of the e_i are contained in at least $k-1$ triangles $(x_{2i-1}, x_{2i}, y_t^{(i)})$, $1 \leqq t \leqq k-1$, $1 \leqq i \leqq k$ where the $y_t^{(i)}$ are all different from $x_0, x_1, \ldots, x_{2k}$.

In this case Theorem 1′ easily follows since G contains k independent triangles. To see this observe that it immediately follows from our assumptions that there are $k-1$ independent triangles (x_{2i-1}, x_{2i}, y_i), $1 \leqq i \leqq k-1$. (x_0, x_{2k-1}, x_{2k}) is the k-th independent triangle.

Henceforth we can thus assume that — say $e_1=[x_1, x_2]$ — is contained in at most $k-2$ triangles in the graph $(G-x_0-x_3-\ldots-x_{2k})$. Put $G_2=(G-x_0-x_1-x_2)$. Now we estimate $v(G_2)$ from below, by estimating from above the number of edges of G incident to x_0, x_1 and x_2. $v(x_0)<n-k$ by our assumption. In $G(x_1, \ldots, x_{2k})$ the vertices x_1 and x_2 are incident to at most $4k-3$ edges. Finally every vertex of $(G-x_0-x_1-\ldots-x_{2k})$ is connected with at most one of the vertices x_1 and x_2, except possibly $k-2$ vertices which might be connected with both. Thus we obtain at most $n-k-3$ further edges. Thus the total number of edges incident to x_0, x_1 and x_2 is at most

$$n-k-1+4k-3+n-k-3=2n+2k-7.$$

Hence for $k>2$

$$v(G_2) \geqq v(G)-2n-2k+7=(n-3)(2k-3)-2(k-1)^2+k+ \tag{4}$$
$$+g(n, k)-f(n, k)>g(n-3, k-1)$$

since clearly

$$g(n, k)-f(n, k) \geqq g(n-3, k-1)-f(n-3, k-1).$$

For $k=2$, we obtain

$$v(G_2) \geqq v(G)-2n+3 \geqq n-3. \tag{4′}$$

Thus (4) and (4′) imply that by our induction hypothesis G_2 contains $k-1$ independent circuits. These and (x_0, x_1, x_2) are together k independent circuits contained in G. Theorem 1′ is now proved.

If $k=2$ then the assertion of Theorem 1 holds for all $n \geqq 6$. The reader can verify it for $n=6$ and then adapt our induction process to prove it for $n>6$.

Perhaps the following result is of some interest.

Theorem 2. *Let $n \geqq 4k$, then every $G^{(n)}_{(2k-1)n-(2k-1)^2+1}$ which contains no triangle contains k independent circuits.*

For $k=1$ the Theorem is trivial. We use induction and assume that it holds for $k-1$. Let $(x_1, \ldots, x_{l_1})=C_1$ be the shortest circuit of G and denote by $x_{l_1+1}, \ldots, x_n$ the other vertices of G. No two non-neighbouring vertices of C_1 can be connected by an edge (for otherwise C_1 would not be the shortest circuit of G). Assume first that $l_1>4$. Then every x_r, $l_1<r \leqq n$ can be connected to at most one vertex of C_1 (for otherwise C_1 would not be the shortest circuit). Thus the vertices $x_1, \ldots, x_{l_1}$ are incident to at most n edges. Let $C_2=(x_{l_1+1}, \ldots, x_{l_1+l_2})$ be the shortest circuit of $(G-x_1-\ldots-x_{l_1})$ and $C_3=(x_{l_1+l_2+1}, \ldots, x_{l_1+l_2+l_3})$ the shortest circuit of $(G-x_1-\ldots-x_{l_1+l_2})$ etc. Thus we obtain the circuits $C_1, \ldots, C_r$ of length $4<l_1 \leqq \ldots \leqq l_r$ and we assume that the graph $(G-x_1-\ldots-x_{l_1+\cdots+l_r})$ contains no circuit. If $r \geqq k$ our Theorem is proved. Assume $r<k$. By our previous argument we obtain that in $(G-x_1-\ldots-x_{l_1+\ldots+l_i})$ the vertices of C_{i+1} are

incident to at most $n-\sum_{t=1}^{i} l_t$ edges. Since $(G-x_1-\ldots-x_{l_1+\ldots+l_r})$ has no circuit it has fewer than $n-\sum_{t=1}^{r} l_t$ edges. Thus the total number of edges of G is less than

$$n+\sum_{t=1}^{r}(n-l_1-\ldots-l_t)\leqq(r+1)n-\sum_{t=1}^{r} l_t\leqq kn-5<(2k-1)n-(2k-1)^2$$

for $n\geqq 4k$, an evident contradiction.

Assume next that $l_1=4$. Then every vertex x_r, $4<r\leqq n$ is connected with at most two of the vertices x_1, x_2, x_3, x_4. Thus the number of edges incident to x_1, x_2, x_3, x_4 is at most

$$4+2(n-4)=2n-4.$$

Hence

$$v(G-x_1-x_2-x_3-x_4)\geqq v(G)-2n+4=(2k-1)n-(2k-1)^2-2n+5=$$
$$=(2k-3)(n-4)-(2k-3)^2+1.$$

By our induction hypothesis $(G-x_1-x_2-x_3-x_4)$ contains $k-1$ independent circuits which together with (x_1, x_2, x_3, x_4) gives k independent circuits of G. Thus the proof of Theorem 2. is complete.

Now we show that Theorem 2. is best possible. Let G be a graph whose vertices are $x_1, \ldots, x_n$, and whose edges are $[x_i, x_j]$ where $1\leqq i\leqq 2k-1<j\leqq n$. Clearly G has $(2k-1)n-(2k-1)^2$ edges and does not contain k independent circuits. If $k>1$ then this graph is the only $G^{(n)}_{(2k-1)n-(2k-1)^2}$ which contains no triangles and does not contain k independent circuits, we leave the proof to the reader.

G. Dirac [2] proved that for $n\geqq 4$ every $G^{(n)}_{2n-2}$ contains a topological complete quadrilateral (i. e. it contains four vertices x_1, x_2, x_3, x_4 any two of which are connected by pairwise disjoint paths). We shall give a simple proof of this theorem by our method. For $n=4$ the theorem clearly holds. We will assume that it holds for $n-1$ and prove it for n. Our $G^{(n)}_{2n-2}$ clearly contains a vertex x_0 of valency not exceeding 3. If $v(x_0)<3$ then $v(G-x_0)\geqq 2n-4$ and thus by our induction hypothesis $(G-x_0)$ and therefore G contains a topological complete quadrilateral. Assume that $v(x_0)=3$ and let x_1, x_2, x_3 be the vertices connected with x_0 by an edge. If $[x_1, x_2]$, $[x_1, x_3]$, $[x_2, x_3]$ are all edges of G then G contains the complete quadrilateral $\{x_1, x_2, x_3, x_4\}$. Thus we can assume that one of these edges — say $[x_1, x_2]$ — does not occur in G. Add the edge $[x_1, x_2]$ to $(G-x_0)$, thus we obtain a graph G' having $n-1$ vertices and $2n-4$ edges. By our induction hypothesis G_1 contains a topological complete quadrilateral $\{y_1, y_2, y_3, y_4\}$. But it is immediate that $\{y_1, y_2, y_3, y_4\}$ is a topological complete quadrilateral of G. To see this observe that the new edge $[x_1, x_2]$ can occur in at most one of the connecting paths and there it can be replaced by $[x_1, x_0]$ and $[x_0, x_2]$. G. Dirac showed by simple examples that not every $G^{(n)}_{2n-3}$ contains a complete topological quadrilateral, e. g. the vertices are $x_1, \ldots, x_n$ the edges $[x_1, x_j]$, $2\leqq j\leqq n$, $[x_2, x_j]$, $3\leqq j\leqq n$.

One could perhaps conjecture that for $n\geqq 5$ every $G^{(n)}_{3n-5}$ contains a complete topological pentagon, but the above proof breaks down and we can not even show that there exists an absolute constant C so that for $n\geqq 5$ every $G^{(n)}_{Cn}$ contains complete topological pentagon.

Theorem 3. *Every $\overline{G}^{(n)}_{n+4}$ contains two weakly independent circuits.*

In other words $g(2)=4$ (see the introduction). We use induction on n. Our Theorem clearly holds for $n=1$. We will assume that it holds for $n-1$ and prove it for n. If our graph contains a circuit of four or fewer edges, then our Theorem is immediate, since by omitting the edges of this circuit a $G^{(n)}_{n+4-i}$ remains with $n+4-i \geqq n$, thus it contains another circuit thus giving our two weakly independent circuits. Thus we can suppose that our graph contains no circuit with fewer than five edges. If our graph contains a vertex of valency one we omit this vertex and obtain a $G^{(n-1)}_{n+3}$ which by our induction hypothesis contains two weakly independent circuits. If x_0 is a vertex of valency two and x_1, x_2 are the vertices connected to x_0 by an edge then we define G_1 as the graph which we obtain from $(G-x_0)$ by adding the edge $[x_1, x_2]$. Clearly G_1 has $n-1$ vertices and $n+3$ edges and thus our Theorem again follows. If all vertices of G have valency $\geqq 3$ then it has at least $\frac{3}{2}n$ edges, or $\frac{3}{2}n \leqq n+4$, which implies $n \leqq 8$. But it is well known and easy to show that every graph with fewer than 10 vertices every vertex of which has valency $\geqq 3$ contains a circuit of at most 4 edges (for 10 vertices this is false as is shown by the well known Petersen graph). This completes the proof of Theorem 3.

Theorem 4. *For every $k>1$*

$$c_1 k \log k < g(k) < c_2 k \log k \tag{5}$$

where c_1 and c_2 are suitable absolute constants.

First we prove the upper bound in (5), (no attempt will be made to get a good estimation for c_2). We shall use induction with respect to k. For $k=2$ the inequality follows from Theorem 3. Assume that it holds for $k-1$, we shall prove it for k. As in the proof of Theorem 3. we can assume that every vertex of our graph $G^{(n)}_{n+[c_2 k \log k]}$ has valency $\geqq 3$. But then

$$v(G) \geqq \tfrac{3}{2} n$$

or

$$n \leqq 2c_6 k \log k. \tag{6}$$

First we prove

Lemma 3. *Let $n \geqq 2$. Every graph $\overline{G}^{(n)}$ every vertex of which has valency $\geqq 3$ contains a circuit of at most $2\left[\dfrac{\log n}{\log 2}\right]$ edges.*

If our graph contains a loop or a circuit of two edges our Lemma is trivial. Thus assume that such circuits do not occur in our graph. Let x_1 be any vertex of $G^{(n)}$. If $G^{(n)}$ contains no circuit of $\leqq 2t$ edges, then all the vertices which can be reached from x_1 in t or fewer edges are all distinct. Since every vertex of $G^{(n)}$ has valency $\geqq 3$ a simple argument shows that in t steps we can reach at least

$$1+3+\ldots+3 2^{t-1} \geqq 2^{t+1} > n$$

vertices if $t=\left[\dfrac{\log n}{\log 2}\right]$. Thus $G^{(n)}$ contains a circuit of length not exceeding $2\left[\dfrac{\log n}{\log 2}\right]$ as stated.

From (6) and Lemma 3. our graph contains a circuit of length not greater than

$$2\left[\frac{\log n}{\log 2}\right] \leq 2\left[\frac{\log (2c_2 k \log k)}{\log 2}\right] < \tfrac{1}{2}c_2 \log k$$

for sufficiently large c_2. If we omit the edges of this circuit we obtain a graph of at most n vertices and more than

$$n + [c_2 k \log k] - \frac{c_2}{2}\log k > n + c_2(k-1)\log k$$

edges. By our induction hypothesis our new graph contains $k-1$ weakly independent circuits, thus together with our first circuit we have our required k weakly independent circuits, which completes the proof of the right side of (5).

To prove the lower bound in (5) we need

Lemma 4. *There exists a constant $c_3 > 0$ so that for every m there exists a $G_{2m}^{(m)}$ which contains no circuit of length less than $c_3 \log m$.*

The proof of the Lemma is implicitely contained in a paper by Erdős [3], but for the sake of completeness we give it here in full detail.

Consider all graphs of m labelled vertices having $2m$ edges. The number l of these graphs clearly equals $\binom{\binom{m}{2}}{3m}$. Denote these graphs by $G_1, \ldots, G_l$ and denote by $f(G_i)$ the number of distinct circuits of length not exceeding $[c_3 \log m]$ contained in G_i. We are going to estimate

$$M = \frac{1}{l}\sum_{i=1}^{l} f(G_i)$$

from above. A simple combinatorial argument shows that the number of graphs G_i which contain a given circuit of k edges equals

$$\binom{\binom{m}{2}-k}{3m-k}. \tag{7}$$

The number of circuits of length k is clearly less than

$$k!\binom{m}{k} < m^k. \tag{8}$$

Thus from (7) and (8) we obtain by a simple argument

$$M < \frac{1}{l}\sum_{k=3}^{[c_3 \log m]} m^k \binom{\binom{m}{2}-k}{3m-k} = \sum_{k=3}^{[c_3 \log m]} m^k \frac{3m(3m-1)\ldots(3m-k+1)}{\binom{m}{2}\left[\binom{m}{2}-1\right]\ldots\left[\binom{m}{2}-k+1\right]} < \tag{9}$$

$$< \sum_{k=3}^{[c_3 \log m]} 10^k < m^{1/2}$$

if c_3 is a sufficiently small absolute constant (again no attempt is made to get a good estimate for c_3 since as in the previous cases there seems no hope at present to obtain the best possible value for c_3). From (9) we obtain that at least one of our graphs — say G_1 — contains fewer than $m^{1/2}$ circuits of length $<[c_3 \log m]$. Omitting one edge from each of these circuits we obtain a graph of m vertices and more than $3m - m^{1/2} \geqq 2m$ edges which contains no circuit of length less than $c_3 \log m$, which proves Lemma 4.

Assume first $k > k_0$ and let $c_1 > 0$ be a sufficiently small absolute constant and put $m = [c_1 k \log k]$. Then by Lemma 4. there is a $G_{2m}^{(m)}$ which contains no circuit of length $< c_3 \log m$. Therefore our $G_{2m}^{(m)}$ contains at most

$$\frac{2m}{c_3 \log m} < k$$

weakly independent circuits, if c_1 is sufficiently small. On the other hand our graph has $2m = m + [c_1 k \log k]$ edges, which completes the proof of the left side of (5) for $k > k_0$. But clearly for $2 \leqq k \leqq k_0$ $g(k) \geqq g(2) = 4$, thus if c_1 is sufficiently small (5) holds for all $k \geqq 2$ and thus Theorem 4. is proved.

If we have already constructed a $G_{m+[c_1 k \log k]}^{(m)}$ which does not contain k weakly independent circuits, we can construct such a $G_{n+[c_1 k \log k]}^{(n)}$ for every $n > m$ by adding a path of $n - m$ new vertices and edges to our $G_{m+[c_1 k \log k]}^{(m)}$.

Finally we consider the following question: Let $m \geqq n$ and consider a graph $G_m^{(n)}$. Define $h(G_m^{(n)})$ as the length of the shortest circuit of our $G_m^{(n)}$. Put

$$f(n, m) = \max h(G_m^{(n)})$$

where the maximum is taken over all graphs $G_m^{(n)}$.

Trivially $f(n, n) = n$ and it is not difficult to show that

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

The determination, or even the estimation, of $f(n, m)$ for general n and m seems a difficult problem.

Theorem 5. *Put $m = n + d$, $d > 1$. Then we have*

$$f(n, n+d) < c_4 \frac{(n+d) \log d}{d} \tag{10}$$

and to every constant $C > 0$ there exists an $A(C)$ depending only on C so that

$$f(n, n+d) > A(C) \frac{(n+d) \log d}{d}. \tag{11}$$

(11) shows that for $d < Cn$ (10) gives the correct order of magnitude for $f(n, m)$.

From Theorem 4. every $G_{n+d}^{(n)}$ contains $c_5\, d/\log d$ weakly independent circuits, thus at least one of them has length not exceeding

$$\frac{(n+d)\log d}{c_5\, d} < c_4 \frac{(n+d)\log d}{d}$$

for sufficiently large c_4, which proves (10).

We shall only outline the proof of (11). Assume first $\frac{n}{4} \leqq d < Cn$. By the same method as used in Lemma 4. we can construct a $G_{n+d}^{(n)}$ the smallest circuit of which has more than $c_6 \log n$ edges (c_6 depends on C and tends to 0 as C tends to infinity), which implies (11) by a simple calculation.

Assume next $d < \frac{n}{4}$. By Lemma 4. there exists a $G_{2d}^{(d)}$ all circuits of which have length $\geqq c_3 \log d$. Put on each edge of this graph $\left[\frac{n}{2d}\right] - 1$ vertices of valency 2. Thus we obtain a graph of $m \leqq n$ vertices and $m+d$ edges the smallest circuit of which has length not less than

$$\left(\left[\frac{n}{2d}\right] - 1\right) c_3 \log d > c_7 \frac{(n+d)\log d}{d}. \tag{12}$$

By adding a path of $n-d$ new edges and vertices to this graph we obtain a $G_{n+d}^{(n)}$ the shortest circuit of which satisfies the inequality (12), thus (11) and therefore Theorem 5. is proved.

It would be easy to strengthen Lemma 3. as follows: Let $C \to \infty$ then there exists an ε_C which tends to 0 as C tends to infinity so that every $G_{[Cn]}^{(n)}$ contains a circuit of length less than $\varepsilon_C \log n$, but we are far from being able to determine the exact dependence of ε_C from C.

Bibliography

[1] P. Erdős and T. Gallai, On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hungar.* **10** (1959), 337–356.
[2] G. Dirac, In abstrakten Graphen vorhandene vollständige 4-Graphen und ihre Unterteilungen, *Math. Nachr.* **22** (1960), 61–85.
[3] P. Erdős, Graph theory and probability, *Canad. J. Math.* **11** (1959), 34–38.

(Received September 30, 1961.)