ON INDEPENDENT CIRCUITS CONTAINED IN A GRAPH

P. ERDÖS AND L. PÓSA

A family of circuits of a graph G is said to be *independent* if no two of the circuits have a common vertex; it is called *edge-independent* if no two of them have an edge in common. A set of vertices will be called a *representing set* for the circuits (for the sake of brevity we shall call it a representing set), if every circuit of G passes through at least one vertex of the representing set. Denote by $I(G) = k$ the maximum number of circuits in an independent family and by $R(G)$ the minimum number of vertices of a representing set. Dirac and Gallai asked whether there is any relation between $I(G)$ and $R(G)$ (trivially $R(G) \geqslant I(G)$). B. Bollobás (unpublished) proved that if $I(G) = 1$, then $R(G) \leqslant 3$ and the complete graph of five vertices shows that $R(G) \leqslant 3$ is best possible.

Consider now all graphs with $I(G) = k$. Denote by $r(k)$ the maximum value of $R(G)$ for all graphs with $I(G) = k$. It is not immediately obvious that $r(k)$ is finite and the theorem of Bollobás states that $r(1) = 3$. The value of $r(2)$ does not seem to be known. We are going to prove the following

THEOREM. *There are absolute constants c_1 and c_2 such that*

$$c_1\, k \log k < r(k) < c_2\, k \log k. \tag{1}$$

We cannot determine

$$\lim_{k\to\infty} r(k)/k \log k$$

and in fact cannot even prove that the limit exists.

First we prove the lower bound in (1). In fact we shall prove a somewhat stronger result. Denote by $E(G)$ the maximum number of edge-independent circuits of G. We shall show that for every k there is a graph G with $I(G) = k$ and

$$r(k) > c_3\, E(G) \log E(G). \tag{2}$$

(2) is stronger than the lower bound in (1) since clearly $E(G) \geqslant I(G) = k$.

We shall prove (2) by a probabilistic argument and cannot at present give an explicit example of a graph satisfying (2). Our proof will be very similar to the one used in (**1**, **2**, and **3**).

First we introduce a few notations. Vertices of G will be denoted by $x_1, \ldots, y_1, \ldots$; circuits will be denoted by C_i; the subgraph of G spanned by the vertices $x_1, \ldots, x_l$ will be denoted by $G(x_1, \ldots, x_l)$; $G(n; m)$ will denote

Received January 10, 1964.

a graph of n vertices and m edges; $\Pi(G)$ denotes the number of edges of G; the edges of G will be denoted by e_i, or by (x_i, x_j); and $G - e_1 - \ldots - e_m$ will denote the graph from which the edges $e_1, \ldots, e_m$ have been omitted. The length of a circuit C_i is the number of its edges.

Consider all graphs $G(n; 100n)$ with n labelled vertices $x_1, \ldots, x_n$. The number of these graphs is clearly

$$\binom{\binom{n}{2}}{100\,n} = A_n. \tag{3}$$

First we state two lemmas.

LEMMA 1. *All but $o(A_n)$ graphs $G(n; 100n)$ have the property that for every choice $x_{i_1} \ldots, x_{i_p}$, $p = [n/2]$, of p vertices,*

$$\Pi(G(x_{i_1}, \ldots, x_{i_p}) \geqslant 2n. \tag{4}$$

LEMMA 2. *Put $l = [(\log n)/100]$. All but $o(A_n)$ graphs $G(n; 100n)$ have fewer than n circuits of length $\leqslant l$.*

Assume that the lemmas have already been proved. Then we prove (2) as follows. By Lemmas 1 and 2 for $n > n_0$ there is a $G(n; 100n)$ which satisfies (4) and for which the number of circuits of length not exceeding l is less than n. Denote by C_i, $1 \leqslant i \leqslant m < n$, these circuits, and let e_i be an arbitrary edge of C_i. The e's are not necessarily different. Put

$$G' = G - e_1 - \ldots - e_m.$$

Clearly each circuit of G' has more than l edges and since G' has at most $100n$ edges, we evidently have

$$E(G') < 100n/l < 20{,}000\, n/(\log n). \tag{5}$$

On the other hand

$$R(G') \geqslant n/2. \tag{6}$$

To prove (6) observe that if $x_1, \ldots, x_k$, $k < n/2$, would represent all circuits of G', then $G'(x_{k+1}, \ldots, x_n)$ would not contain any circuits, hence would have fewer than $n - k$ edges, or

$$\Pi(G'(x_{k+1}, \ldots, x_n)) < n - k < n.$$

But we evidently have by (4) and $m < n$ $(n - k > n/2)$

$$\Pi(G'(x_{k+1}, \ldots, x_n)) \geqslant \Pi(G(x_{k+1}, \ldots, x_n)) - m > 2n - n = n,$$

an evident contradiction. Hence (6) is proved. (5) and (6) easily imply (2). To see this, let n be the largest integer with $20{,}000n/(\log n) \leqslant k$. For our graph G' we have by (5) and (6)

$$E(G') \leqslant k, \qquad R(G') > ck \log k$$

and by perhaps adding to G' some (at most k) independent circuits we clearly obtain a graph G_1' with

$$I(G_1') = k, \qquad E(G_1') \leqslant 2k, \qquad R(G_1') > c_3 E(G_1') \log E(G_1'),$$

which completes the proof of (2), if (5) and (6) are assumed.

Thus to complete the proof of (2) we only have to prove our lemmas. To prove Lemma 1, observe that the number of graphs $G(n; 100n)$ which have p vertices $x_{i_1}, \ldots, x_{i_p}$ with

$$\Pi(G(x_{i_1}, \ldots, x_{i_p})) < 2n$$

is at most (**1**, pp. 35–6)

$$(7) \qquad I_n = \binom{n}{p} \sum_{l<2n} \binom{\binom{p}{2}}{l} \binom{\binom{n}{2} - \binom{p}{2}}{100n - l} < 2n . 2^n \binom{\binom{p}{2}}{2n} \binom{\binom{n}{2} - \binom{p}{2}}{98n}$$

since

$$\binom{n}{p} < 2^n,$$

and a simple computation shows that the terms in the sum (7) are increasing for $l \leqslant 2n$. Now $e^{2n} > (2n)^{2n}/(2n!)$ and $p = [n/2]$ imply that

$$(8) \qquad \binom{\binom{p}{2}}{2n} < \left(\frac{p^2}{4n}\right)^{2n} . e^{2n} \leqslant \left(\frac{en}{16}\right)^{2n}$$

and for $n > n_0$ we obtain by a simple computation and (3)

$$(9) \qquad \binom{\binom{n}{2} - \binom{p}{2}}{98n} < \binom{\binom{n}{2}}{98n} \left(1 - \frac{(p-1)^2}{(n-1)^2}\right)^{98n}$$

$$< (1 + o(1))^n \left(\frac{3}{4}\right)^{98n} \binom{\binom{n}{2}}{100n} \left(\frac{100n}{n^2/2}\right)^{2n}$$

$$= (1 + o(1))^n A_n \left(\frac{3}{4}\right)^{98n} \left(\frac{200}{n}\right)^{2n}.$$

From (7), (8), and (9) we have

$$I_n < (1 + o(1))^n A_n (\tfrac{3}{4})^{98n} 200^{2n} = o(A_n).$$

which proves Lemma 1.

Now we prove Lemma 2 (**1**, p. 36). The number of graphs $G(n; 100n)$ which contain a given circuit $(x_1, x_2), (x_2, x_3), \ldots, (x_{r-1}, x_r), (x_r, x_1)$ clearly equals

$$\binom{\binom{n}{2} - r}{100n - r}.$$

A circuit is determined by its vertices and their order. Thus there are $n(n-1)\ldots(n-r+1) < n^r$ such circuits. Therefore the expected number of circuits of length $r \leqslant l = [(\log n)/100]$ is less than

$$\binom{\binom{n}{2}}{100n}^{-1} \sum_{3\leqslant r\leqslant l} n^r \binom{\binom{n}{2}-r}{100n-r} < (1+o(1)) \sum_{3\leqslant r\leqslant l} n^r \left(\frac{100n}{\binom{n}{2}}\right)^r = o(n).$$

Therefore by a simple and well-known argument the number of graphs $G(n; 100n)$ having n or more circuits of length not exceeding l is $o(A_n)$, which proves Lemma 2 and hence the proof of (2) is complete.

To complete the proof of our theorem we now have to prove that $r(k) < c_2 k \log k$. We are going to use two theorems, the first, due to ourselves **(1**, p. 9**)**, which states: *There exists an absolute constant* c_3 *so that every* $G(n, n+l)$ *contains at least* $c_3 l/\log l$ *edge-independent circuits.*

Assume now that every vertex of our graph has valency $\leqslant 3$. Then clearly it contains $c_3 l/\log l$ independent circuits; since if two circuits are edge-independent and not independent, then every common vertex of the two circuits must have valency 4.

The second theorem is due to T. Gallai **(4)**. Let G be a graph. Designate some of its vertices, say $x_1, \ldots, x_u$, as principal vertices; the other vertices, $y_1, \ldots, y_r$ of G, will be the subsidiary vertices. A path is called a principal path if its end points are principal vertices and it contains no other principal vertices. (A circuit having only one principal vertex is not allowed.) Denote by $V_{\max}$ the maximum number of independent principal paths (two principal paths are called independent if they have no vertex [principal or subsidiary] in common). $\Pi_{\min}$ denotes the smallest integer such that there are $\Pi_{\min}$ vertices representing all the principal paths—in other words there are $k = \Pi_{\min}$ vertices $x_{l_1}, \ldots, x_{l_k}$ (principal or subsidiary) so that every principal path contains one of the x_{l_i}'s and one cannot find fewer than k vertices with this property. Gallai's theorem asserts that

$$\text{(10)} \qquad \Pi_{\min} \leqslant 2V_{\max}.$$

Now we are ready to prove the right-side inequality of (1). Assume that in G the maximum number of independent circuits is k and let

$$\text{(7)} \qquad C_i, \qquad 1 \leqslant i \leqslant k,$$

be a maximal system of independent circuits of G. Omit all the edges of C_i, $1 \leqslant i \leqslant k$, but retain the vertices of C_i. Thus we obtain the graph G_1. Let the principal vertices of G_1 be the vertices of C_i, $1 \leqslant i \leqslant k$, all other vertices being subsidiary ones. Consider now a maximal system of independent principal paths of G_1. The circuits C_i and the maximal system of independent paths define a graph G^* every vertex of which has valency not exceeding three. (G^* is a subgraph of G but not of G_1.) Let m denote the number of vertices of G^*. Then clearly the number of edges of G^* is

$$m + V_{\max} \tag{11}$$

since each principal path gives an excess of 1 of the number of edges over the number of vertices. Thus by our theorem G^* (and therefore G) contains at least

$$c_3 \, V_{\max}/\log V_{\max}$$

independent circuits. Hence

$$\frac{c_3 \, V_{\max}}{\log V_{\max}} \leqslant k \quad \text{or} \quad V_{\max} \leqslant c_4 \, k \log k. \tag{12}$$

Now let $y_1, \ldots, y_t$ be a minimal system of vertices representing all the principal paths of G_1. By (12) and Gallai's theorem

$$t \leqslant 2c_4 \, k \log k. \tag{13}$$

For some i, $1 \leqslant i \leqslant k$, there may exist a circuit D_i which has one (and only one) common vertex x_i with C_i, which is independent of $C_j (1 \leqslant j \leqslant k, j \neq i)$ and does not pass through any of the y_j, $1 \leqslant j \leqslant t$. But for a given i there cannot be two such D_i's, say D_{i_1} and D_{i_2}, whose unique common vertex with C_i is x_{i_1} and x_{i_2}, where x_{i_1} and x_{i_2} are distinct. To see this, observe that if D_{i_1} and D_{i_2} are independent, then the $k + 1$ circuits

$$C_j (1 \leqslant j \leqslant k, j \neq i), \qquad D_{i_1}, \qquad D_{i_2}$$

would be independent, which contradicts the maximality property of k. If D_{i_1} and D_{i_2} are not independent, then their union contains a principal path connecting x_{i_1} and x_{i_2}; hence it contains one of the vertices $y_j (1 \leqslant j \leqslant t)$, which by assumption represent all principal paths; but this contradicts our assumption that D_{i_1} and D_{i_2} do not contain any of the $y_j (1 \leqslant j \leqslant t)$.

If C_i is such that there is a D_i corresponding to it, adjoin their common vertex x_i to the y's; otherwise choose any vertex of C_i, denote it by x_i, and adjoin it to the y's. Some of the x_i's might have already occurred amongst the y's; but in any case the system

$$y_j (1 \leqslant j \leqslant t), \qquad x_i \; (1 \leqslant i \leqslant k) \tag{14}$$

contains at most

$$2c_4 \, k \log k + k < c_2 \, k \log k$$

vertices. Our proof will be complete if we show that the system (14) represents every circuit of G. Let C be any circuit of G. We have to show that it contains at least one of the vertices (14). The circuits C_i are clearly represented by the vertices (14); thus we can assume that $C \neq C_i$, $1 \leqslant i \leqslant k$. If C contains at least two of the vertices of C_i, $1 \leqslant i \leqslant k$, then C contains a principal path of G_1 and hence one of the vertices y_j, $1 \leqslant j \leqslant t$. If C contains only one of the vertices of C_i and does not contain any of the y_j $(1 \leqslant j \leqslant t)$, then it contains

x_i, $1 \leqslant i \leqslant k$. Finally, C cannot be disjoint of all the C_i's because of the maximality property of the C_i, $1 \leqslant i \leqslant k$. This completes the proof of our theorem.

It would be easy to obtain explicit inequalities for c_1 and c_2 but they would be very far from being best possible.

References

1. P. Erdös and L. Pósa, *On the maximal number of disjoint circuits of a graph*, Publ. Math. Debrecen, *9* (1962), 3–12.

2. P. Erdös, *Graph theory and probability*, Can. J. Math., *11* (1959), 34–8.

3. ——— *On circuits and subgraphs of chromatic graphs*, Mathematika, *9* (1962), 170–5.

4. T. Gallai, *Maximum-minimum Sätze and verallgemeinerte Faktoren von Graphen*. Acta Math. Hung. Acad. Sci., *12* (1961), 131–73; cf. pp. 158 and 161.

University of British Columbia, Vancouver,
and Budapest, Hungary