SEPARATUM

Studia Scientiarum Mathematicarum Hungarica

Tomus II Fasciculi 1—2

G. KATONA and E. SZEMERÉDI

ON A PROBLEM OF GRAPH THEORY

1967

Studia Scientiarum Mathematicarum Hungarica 2 (1967) 23—28.

ON A PROBLEM OF GRAPH THEORY

by

G. KATONA and E. SZEMERÉDI

1. Introduction

We say that a directed graph has the diameter 2 if any two vertices are connected by a directed path of length at most 2. Let $V(G)$ denote the number of vertices of a graph G. Let $E(G)$ denote the number of edges of G and $D(G)$ the diameter of G.

Put

$$F(n) = \min_{\substack{V(G)=n \\ D(G)=2}} E(G).$$

P. ERDŐS, A. RÉNYI and V. T. SÓS [1] proposed the question of determining the value of $F(n)$. The problem has the following interpretation. There are n airports. Any (ordered) pair A, B of these airports is connected by at most one (directed) flight from A to B. How many directed connections have to be established to assure the possibility to fly from every airport to any other by changing the plane at most once?

It was noticed also by P. ERDŐS, A. RÉNYI and V. T. SÓS that we can reduce this problem to the following one. A (non-directed) graph is called to be a complete even graph if we can split its vertices into two disjoint subsets, so that two vertices are connected if and only if they are in different subsets. At least how large is the sum of the numbers of the vertices of even complete graphs covering every edge of a complete graph having n vertices? We are going to prove in 2 of this paper that this number is at least $n \log n$. (Now and in what follows $\log n$ denotes $\overset{2}{\log}\, n$). In 3 we deduce from this result of 2 estimates for $F(n)$. In 4 we consider the sum of the numbers of the vertices of complete even graphs covering an arbitrary given graph.

2. On Covering of a Complete Graph by Complete Even Graphs

Let A and B be two finite disjoint sets. Let (A, B) denote the complete even graph in which x and y are connected if and only if $x \in A$ and $y \in B$, or if $x \in B$ and $y \in A$. We denote the number of elements of a set A by $|A|$. We say that the G graph is covered by the family of complete even graphs $(A_i\, B_i)$ $1 \leq i \leq m$ if any edge of G is the edge of some of the complete even graphs $(A_i B_i)$. The complete graph having n vertices is denoted by $\langle n \rangle$.

THEOREM 1. *If* $\langle n\rangle$ *is covered by the family of the complete even graphs* (A_iB_i) $1\leqq i\leqq m$ *then*

$$\sum_{i=1}^{m}|A_i|+|B_i|\geqq n\log n. \tag{1}$$

PROOF OF THEOREM 1. Let us denote the vertices of $\langle n\rangle$ by $x_1, x_2, \ldots, x_n$. We construct a matrix M of m rows and n columns each element of which is equal to one of the numbers 0, 1, 2.

We put $M=(a_{ij})$ where

$$a_{ij}=\begin{cases}0 & \text{if } x_j\in A_i\\ 1 & \text{if } x_j\in B_i\\ 2 & \text{if } x_j\notin A_i \quad x_j\notin B_i.\end{cases}$$

We denote the number of zeros and ones being in the j-th column by h_j. In case the family of complete even graphs (A_iB_i) $1\leqq i\leqq m$ cover $\langle n\rangle$ then there is to any of the pairs (x_k, x_l) an (A_iB_i), so that one of them is an element of A_i, and the other is an element of B_i. Concerning M this means that to any two different columns there is a row, so that in this row in one of the two columns there stands 0 and in the other 1. By other words to each j and k $(j\neq k)$ there is at least one i such that either $a_{ij}=0$ and $a_{ik}=1$ or $a_{ij}=1$ and $a_{ik}=0$. Thus the k-th column and l-th column are different if $k\neq l$, and they remain different even if we replace any 2 by either 0 or 1. There are in the k-th column $m-h_k$ elements equal to 2, so we obtain from the k-th column 2^{m-h_k} different columns, if we replace 2 whereever it occurs either by 0 or by 1. As, however, the number of columns having length m consisting of either 0 or 1 is 2^m, we get the inequality

$$\sum_{k=1}^{n}2^{m-h_k}\leqq 2^m \tag{2}$$

that is

$$\sum_{k=1}^{n}\frac{1}{2^{h_k}}\leqq 1. \tag{3}$$

The inequality between the geometrical and the arithmetical means and (3) imply

$$\sqrt[n]{\frac{1}{2^{\sum_{k=1}^{m}h_k}}}\leqq\sum_{k=1}^{n}\frac{1}{n}\frac{1}{2^{h_k}}\leqq\frac{1}{n}$$

wherefrom

$$\sum_{k=1}^{n}h_k\geqq n\log n. \tag{4}$$

However the total number of zeros and ones in the matrix M is equal to $\sum_{k=1}^{n}h_k$, and it is equal to $\sum_{i=1}^{m}|A_i|+|B_i|$ as well.

Otherwise it is obvious that we can always find such a family (A_iB_i) $1 \leqq i \leqq m$ which coverns $\langle n \rangle$ and satisfies

$$\sum_{i=1}^{m} |A_i|+|B_i| = n\{\log n\}$$

where $\{x\}$ denotes the smallest integer which is greater than x or equal to x. Let us construct the family in the form of a matrix. Let the j-th column consist of the sequence of digits of the number $j-1$ in the binary system. Since the number of digits of any $k \leqq n-1$ in the binary system is $\leqq [\log (n-1)]+1=\{\log n\}$ the number of the rows of the matrix will be $\{\log n\}$. M will entirely consist of 0-s and 1-s. Obviously the columns of M are all different and M has $n\{\log n\}$ elements. In this example was no 2 in the matrix. This corresponds to the case when the set is divided into two disjoint subsets by (A_iB_i). To cover $\langle n \rangle$ $\log n$ such pairs (A_iB_i) are necessary, as it was proved in [2]. As then $|A_i|+|B_i|=n$ the total number of vertices of the covering graphs is in this case $n\{\log n\}$. This fact led to conjecturing the result of Theorem 1.

3. Lower and Upper Bounds for $F(n)$

Now let us consider a directed graph of diameter 2, having n vertices $x_1x_2\ldots x_n$ and between any two different vertices there is at most one directed edge. Let us consider the vertex x_j. Let the set A_j consist of those vertices to which a directed edge leads from x_j, and let the set B_j consist of x_j and of those vertices wherefrom a directed edge leads into x_j. Let x_k and x_l be any two different verices then, according to our assumption, either a directed edge or a directed path of length two leads from x_k into x_l. In the first case $x_k \in B_k$ and $x_l \in A_k$ while in the second case, if the directed path of length 2 from x_k to x_l goes through x_j, then $x_k \in B_j$ and $x_l \in A_j$. That means that $\langle n \rangle$ is covered by the family of (A_i, B_i) $1 \leqq j \leqq n$ complete even graphs. Theorem 1 implies

$$\sum_{i=1}^{n} |A_i|+|B_i| \geqq n \log n.$$

Since the number of the edges starting from x_j is $|A_j|$ and the number of edges ending in x_j is $|B_j|-1$

$$E(G) = \sum_{j=1}^{n} \frac{|A_j|+|B_j|-1}{2} \geqq \frac{n\cdot \log n}{2} - \frac{n}{2}.$$

Thus we obtained the following theorem:

THEOREM 2. *Let G be a directed graph of diameter two between any two vertices of which there is at most one directed edge and the number of its edges is $E(G)$, then*

$$E(G) \geqq \frac{n}{2} \log \frac{n}{2}. \tag{5}$$

Thus for the function $F(n)$ defined in § 1 we obtain the inequality

$$F(n) \geqq \frac{n}{2} \log \frac{n}{2}. \tag{6}$$

On the other hand, it follows from the remark at the end of § 1 that $F(n) \leqq n\{\log n\}$.

4. The Covering of Arbitrary Graphs by Complete Even Graphs

THEOREM 3. *Let G be a nondirected graph having n vertices $x_1, x_2 \dots x_n$ and let the degree of x_j be f_j and let G be covered by the family of complete even graphs, $(A_i B_i)$ $1 \leqq i \leqq m$ then*

$$\sum_{i=1}^{m} |A_i| + |B_i| \geqq \sum_{i=1}^{n} \log \frac{n}{n-f_i}. \tag{7}$$

PROOF OF THEOREM 3. The following lemma is necessary for the proof.

LEMMA: Let H be a set of k elements, let $H_1 \dots H_n$ be subsets of H and suppose that from any fixed i the number of j-s for which $H_i \cap H_j \neq 0$ is at least g_i. Then

$$\sum_{i=1}^{n} \frac{|H_i|}{g_i} \leqq k. \tag{8}$$

PROOF OF THE LEMMA. It is easy to see that

$$\sum_{i=1}^{n} \frac{|H_i|}{g_i} = \sum_{i=1}^{n} \sum_{x \in H_i} \frac{1}{g_i} = \sum_{x \in H} \sum_{\{i:\, x \in H_i\}} \frac{1}{g_i}. \tag{9}$$

To any fixed x, however, if $x \in H_j$ we have $|\{i: x \in H_i\}| \leqq g_j$, because H_j has common elements at least with $|\{i: x \in H_i\}|$ sets H_i. Therefore $\sum_{\{i: x \in H_i\}} \frac{1}{g_i} \leqq 1$ and thus from (9) $\sum_{i=1}^{n} \frac{|H_i|}{g_i} \leqq k$.

Let us return to the proof of the theorem. Similarly as in the proof of Theorem 1 we construct a matrix M of m rows and n columns, whose elements a_{ij} are defined as follows:

$$a_{ij} = \begin{cases} 0 & \text{if} \quad x_j \in A_i \\ 1 & \text{if} \quad x_j \in B_i \\ 2 & \text{if} \quad x_j \notin A_i \quad x_j \notin B_i \end{cases}$$

where $x_1, x_2 \dots x_n$ are the vertices of the graph G. Let H_j be the set of those columns which we can get from the j-th column of M, so that we write instead of any 2 occurring there independently either 0 or 1. Further, let us examine at most with how many H_i has a fixed H_j a common element. If in a graph G between x_j and x_i there is an edge, then there is an (A_l, B_l) such that $x_i \in A_l$ $x_j \in B_l$ or $x_i \in B_l$ $x_j \in A_l$. That is there may be found such l-th row to the j-th and i-th column in which the j-th element is 0 and the i-th element is 1 or vice versa. However, if we write into the j-th and i-th columns in the place of 2 either 0 or 1, the columns remain different, that is $H_i \cap H_j = 0$. Since the degree of x_j is f_j, at most $n - f_j$ is the number of those H_i for which $H_i \cap H_j \neq 0$. We can get from the lemma

$$\sum_{i=1}^{n} \frac{|H_i|}{n-f_i} \leqq 2^m$$

If h_i denotes the sum of the number of 0-s and 1-s in the i-th column we get

$$\sum_{i=1}^{n} \frac{2^{m-h_i}}{n-f_i} \leqq 2^m$$

that is

$$\sum_{i=1}^{n} \frac{1}{2^{h_i}(n-f_i)} \leqq 1. \tag{10}$$

The inequality between the geometrical and arithmetical means and (10) imply

$$\sqrt[n]{\frac{1}{2^{\sum_{i=1}^{n} h_i} \prod_{i=1}^{n} n-f_i}} \leqq \sum_{i=1}^{n} \frac{1}{n} \frac{1}{2^{h_i}(n-f_i)} \leqq \frac{1}{n}$$

wherefrom, after some calculations using the equality

$$\sum_{i=1}^{m} |A_i|+|B_i| = \sum_{i=1}^{n} h_i$$

we get

$$\sum_{i=1}^{m} |A_i|+|B_i| \geqq \sum_{i=1}^{n} \log \frac{n}{n-f_i} \tag{11}$$

what was to be proved.

Remarks

It is easy to see that Theorem 1 and Theorem 3 can be generalized in such a way that we use as covering graphs (instead of even graphs) graphs of the form $(A_1, A_2, \dots A_s)$ which are defined as follows: The sets of vertices $A_1, A_2, \dots A_s$ are disjoint and two vertices are connected by an edge if and only if they do not belong to the same set A_i. In this case the result, according to Theorem 3 is:

$$\sum_{i=1}^{m} |A_1^i|+|A_2^i|\dots|A_s^i| \geqq \sum_{i=1}^{n} \overset{s}{\breve{\log}} \frac{n}{n-f_i}.$$

A good estimate can be obtained by Theorem 3 if f_i is large. The case is of special interest when $f_i=f\ (1 \leqq i \leqq n)$. In this case the right-hand side of (11) takes the form of $n \log \frac{n}{n-f}$. If, for instance, $f=n-c$ then $\sum_{i=1}^{m} |A_i|+|B_i| \geqq n \log \frac{n}{c}$. That is it does not differ essentially from the case of a complete graph. Taking the other interesting case when $f=c \cdot n$ then the right-hand side of (11) is $n \log \frac{1}{1-c}$.

REFERENCES

[1] ERDŐS, P., RÉNYI, A. and SÓS, V. T.: On a problem of graph theory, *Studia Sci. Math. Hung.* **1** (1966) 215—235.

[2] RÉNYI, A.: On random generating elements of a finite Boolean algebra, *Acta Sci. Math.* (Szeged) **22** (1961) 75—81.

MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST

(Received March 20, 1966.)