ON THE CAPACITY OF GRAPHS

by

P. ERDŐS and J. KOMLÓS (Budapest)[1]

To the memory of A. RÉNYI

The capacity of a graph $\mathcal{G}$ is the number of non-isomorphic (non-empty) subgraphs of $\mathcal{G}$, and is denoted by $\nu(n, \mathcal{G})$; here n indicates the number of vertices of $\mathcal{G}$ (or the order of $\mathcal{G}$, in notation $n = |\mathcal{G}|$). (Throughout this paper, subgraph will always mean induced subgraph.)

Put

$$\nu(n) = \max_{|\mathcal{G}|=n} \nu(n, \mathcal{G}) .$$

Obviously

$$n \leqq \nu(n) \leqq 2^n - 1 .$$

Goldberg conjectured that

$$\lim_{n\to\infty} \frac{\nu(n)}{2^n} = 0 .$$

In this paper we are going to disprove this conjecture, moreover, we establish the following

THEOREM 1.

$$\lim_{n\to\infty} \frac{\nu(n)}{2^n} = 1 .$$

Theorem 1 states that for large n there is a graph of whose subgraphs almost all are different (non-isomorphic).

Actually, we will show that almost all graphs have this property.

We say that almost all graphs have a given property, if the ratio of the number of graphs with n vertices not having this property to the total number of graphs with n vertices tends to 0 as $n \to \infty$. If this property does not depend on the labeling of the vertices of the graph (if there is any), then the above statement does not depend on the fact whether we consider the graphs labelled (and divide by $2^{\binom{n}{2}}$) or not-labelled (and divide by the number A_n of non

[1] This work was partially carried out while the second author spent an academic year at McGill University, Montreal.

isomorphic graphs with n vertices). This can be seen easily using the well-known asymptotic relation

$$A_n \sim \frac{2^{\binom{n}{2}}}{n!}.$$

We are going to prove the following sharpening of Theorem 1.

THEOREM 2. *For all graphs* $\mathcal{G}$

$$\nu(n, \mathcal{G}) \leq 2^n - 2^{\left[\frac{n}{2}\right]-1}, \tag{1}$$

and given $\varepsilon > 0$ *we have for almost all graphs*

$$\nu(n, \mathcal{G}) > 2^n - 2^{\left(\frac{1}{2}+\varepsilon\right)n}. \tag{2}$$

Hence

$$\nu(n) = 2^n - 2^{\left(\frac{1}{2}+o(1)\right)n}.$$

Actually, we will prove a bit sharper estimation:

$$\nu(n) = 2^n - 2^{\frac{1}{2}n + O(\sqrt{n \log n})}.$$

However, $\nu(n, \mathcal{G})$ can be very small, since for the complete graph $\mathcal{G}_n$ $\nu(n, \mathcal{G}_n) = n$. Theorem 2 is in accordance with the fact that almost all graphs are not symmetric.

DEFINITION. The consistency set of a subset U of the vertices of a graph is the set V of vertices $v \notin U$, for which v is connected by either all vertices in U or none of them.

LEMMA. *In a graph of order n there always exist two vertices such that the consistency set of this pair of vertices contains at least* $\left[\frac{n}{2}\right] - 1$ *vertices.*

Lemma 1 obviously implies part (1) of Theorem 2, since putting one of these vertices and an arbitrary subset of their consistency set, and the other vertex and the same subset, the two subgraphs defined this way are isomorphic. Theorem 4 will state that in almost all graphs the above pairs of subgraphs are practically all pairs of isomorphic subgraphs.

PROOF of the lemma. Denote the valencies of the vertices by $v_1, v_2, \ldots, v_n$. Thus the sum of the numbers of vertices of consistency sets of all the $\binom{n}{2}$ pairs of vertices is obviously equal to

$$\sum_{i=1}^{n}\left[\binom{v_i}{2} + \binom{n-1-v_i}{2}\right] \geq n\,\frac{(n-1)(n-3)}{4},$$

therefore at least one of these numbers is greater than or equal to $\frac{n-3}{2}$ proving the lemma.

The statement that in a (labelled) random graph of order n some event (property) has probability p will mean that the ratio of the number of graphs of order n having this property to the number $2^{\binom{n}{2}}$ is p. This can also be expressed (using the language of probability theory) by saying that this event has probability p if the labelled graphs of order n are obtained by drawing the edges at random, independently of each other, with probability $\frac{1}{2}$.

Now we formulate an auxiliary theorem which gives an interesting explicit formula for the probability that two subgraphs of a random graph are isomorphic under a given mapping.

Let us have two (labelled) — not necessarily different — subgraphs $\mathcal{G}_1, \mathcal{G}_2$ of a (labelled) random graphs, and a one-to-one mapping φ from the vertices $u_1, u_2, \ldots, u_m$ of $\mathcal{G}_1$ to those of $\mathcal{G}_2$, say $v_1, v_2, \ldots, v_m$ (i.e. $v_i = \varphi(u_i)$).

Put

$$|\mathcal{G}_1| = |\mathcal{G}_2| = m\,,$$
$$U = \{u_1, u_2, \ldots, u_m\},$$
$$V = \{v_1, v_2, \ldots, v_m\},$$
$$A = U - V, \quad M = U \cap V, \quad B = V - U$$

(A, B or M can be empty). Let us call a sequence of different vertices $u_{i_1}, u_{i_2}, \ldots, u_{i_d}, v_j$ a path of length d if

$$u_{i_2} = \varphi(u_{i_1}), \ldots, u_{i_d} = \varphi(u_{i_{d-1}}), v_j = \varphi(u_{i_d}), \quad u_{i_1} \in A, v_j \in B$$

(obviously $u_{i_2}, \ldots, u_{i_d} \in M$).

A sequence of different vertices $u_{i_1}, u_{i_2}, \ldots, u_{i_c}$ is called a cycle of length c if

$$u_{i_2} = \varphi(u_{i_1}), \ldots, u_{i_c} = \varphi(u_{i_{c-1}}), u_{i_1} = \varphi(u_{i_c})$$

(obviously $u_{i_1}, \ldots, u_{i_c} \in M$).

Thus the mapping φ can be "split" into disjoint paths and cycles of lengths $d_1, d_2, \ldots, d_s$ and $c_1, c_2, \ldots, c_r$, respectively.

Put $\sum_{i=1}^{s} d_i = d$, $\sum_{j=1}^{r} c_j = c$. The number $p = m - r$ will be called the parameter of the mapping. E.g. $p = 0$ means the identity map, in the case $p = 1$ at most two elements are moving (we have a path of length one, or a cycle of length 2), in the case $p = 2$ we either have a cycle of length 3, or a cycle of length 2 and a path of length 1, or a path of length 2, or two paths of length 1, and any other vertex is mapped into itself. The number p describes the number of moving vertices and the way they move, that is why we call it the parameter.

THEOREM 3. *The probability that the subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in a random graph are isomorphic under the map φ is equal to $2^{-\alpha}$, where*

$$\alpha = \frac{m(m-1)}{2} + \sum_{i=1}^{r} \left\{\frac{c_i}{2}\right\} - \frac{1}{2} \sum_{i,j=1}^{r} (c_i, c_j). \tag{3}$$

Here (a, b) stands for the greatest common divisor of a and b, and $\{x\} = x - [x]$ for the fraction part of x.

Since $(a, b) \leqq a$, Theorem 3 leads to the following estimation.

COROLLARY. The above probability is at most

$$2^{-\frac{m(m-r-1)}{2}} = 2^{-\frac{m(p-1)}{2}}. \tag{4}$$

The proofs of Theorems 2 and 4 will be based on this inequality.

If two subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are isomorphic, there might exist many isometries from $\mathcal{G}_1$ to $\mathcal{G}_2$, the parameter of the one having the greatest parameter is called the rank of isometry of the pair $\mathcal{G}_1, \mathcal{G}_2$.

Let us denote the number of pairs of isometric subgraphs of a graph $\mathcal{G}$ with n vertices by $N(n, \mathcal{G})$ and that of isometric subgraphs with rank of isometry p by $N_p(n, \mathcal{G})$.

Thus

$$N(n, \mathcal{G}) = \sum_{p=1} N_p(n, \mathcal{G}),$$

and obviously

$$\nu(n, \mathcal{G}) \geqq 2^n - 1 - N(n, \mathcal{G}). \tag{5}$$

We will see that for almost all graphs $\mathcal{G}$

$$2^{\frac{n}{2} - 4\sqrt{n \log n}} < N_1(n, \mathcal{G}) < 2^{\frac{n}{2} + 4\sqrt{n \log n}}, \tag{6}$$

or more generally for fix p

$$2^{\frac{n}{2^p} - 4\sqrt{n \log n}} < N_p(n, \mathcal{G}) < 2^{\frac{n}{2^p} + 4\sqrt{n \log n}}. \tag{7}$$

REMARK. The upper part of (7) says that we have for almost all graphs $\mathcal{G}$: If there is a pair of subsets and a map between them with parameter at least p, then the number of vertices not moving under the mapping is at most $\frac{n}{2^p} + 4\sqrt{n \log n}$.

(7) will lead to the following theorem:

Theorem 4. *For fixed P we have the expansion for almost all graphs*

$$N(n, \mathcal{G}) = \sum_{p=1}^{P} N_p(n, \mathcal{G}) + o\big(N_P(n, \mathcal{G})\big).$$

In particular, $N(n, \mathcal{G}) = (1 + o(1))N_1(n, \mathcal{G})$.

(5) and Theorem 4 imply Theorem 2.

For given two vertices the number C of vertices in the consistency set of this pair has a binomial distribution of parameters $n - 2, \frac{1}{2}$:

$$\mathsf{P}(C = i) = \binom{n-2}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{n-2-i}, \quad i = 0, 1, \ldots, n-2$$

proving (6). More generally, that of a set of k vertices follows a binomial distribution of parameters $n - k, \frac{1}{2^{k-1}}$ and applying it to fixed cycles of length at least 2, and paths, allowing to change only the not moving vertices, they can be chosen only as the subsets of the intersection of the consistency sets of these cycles and paths, whose number of elements has binomial distribution of parameters

$$n - \sum_{j=1}^{r} C_j = \sum_{i=1}^{s} (d_i + 1), \frac{1}{2^{\sum_{i=1}^{r}(C_j-1) + \sum_{r=1}^{s} d_i}},$$

or $n - m + s, \frac{1}{2^p}$. These cycles and paths can be chosen in at most $(4n)^{2p}$ ways.

The probability that at least one of these $(4n)^{2p}$ binomial distributed numbers differs from $\frac{n}{2^p}$ by more than $4\sqrt{\frac{p}{2^p}}\sqrt{n \log n}$ is less than $(4n)^{2p} \cdot e^{-8p \log n} = o(1)$.

Thus for given p in almost all graphs

$$2^{\frac{n}{2^{2p}} - 4\sqrt{n \log n}} < N_p(n, \mathcal{G}) < (4n)^{2p} \cdot 2^{\frac{n}{2^{2p}} + 4\sqrt{\frac{p}{2^p}}\sqrt{n \log n}}$$

proving (7).

For proving Theorem 4 we have to show that for fixed P in almost all graphs

$$\sum_{p=P+1} N_p(n, \mathcal{G}) = o\big(N_P(n, \mathcal{G})\big).$$

For this we use (4).

Choose $P_0 > P$ in such a way that

$$\binom{n}{\left(\frac{1}{2^{P_0}} + \varepsilon\right) n}^2 < 2^{\left(\frac{1}{2^P} - \varepsilon\right) n}$$

for some $\varepsilon > 0$. Thus, the number of pairs of subgraphs with $m \leq \left(\frac{1}{2^{P_0}} + \varepsilon\right) n$ vertices is less than

$$n \cdot 2^{\left(\frac{1}{2^P} - \varepsilon\right) n} = o(N_P(n, \mathcal{G})) .$$

For a given pair of subgraphs with

$$m > \left(\frac{1}{2^{P_0}} + \varepsilon\right) n$$

vertices and a given map φ between them we have either

a) $$r - r' < \frac{1}{2^{P_0}} n$$

or

b) $$r - r' \geq \frac{1}{2^{P_0}} n ,$$

where $r' = \sum_{c_i \geq 2} 1$, i.e. $r - r'$ is the number of non-moving vertices. Clearly $r + r' \leq m$.

In case a) we have

$$r' > r - \frac{n}{2^{P_0}},$$

thus

$$p = m - r \geq r' > r - \frac{n}{2^{P_0}},$$

whence

$$p > \frac{m - \frac{n}{2^{P_0}}}{2} .$$

Therefore

$$\frac{m(p-1)}{2} \geq \frac{\varepsilon}{2 \cdot 2^{P_0}} n^2.$$

The number of maps φ is $m! \leq n!$, thus — by (4) — the probability of having such a pair of subgraphs with such a map, is at most

$$2^{2n} n! \, 2^{-\frac{\varepsilon}{2 \cdot 2^{P_0}} n^2} = o(1) .$$

In case b) we have

$$r - r' \geq \frac{n}{2^{P_0}},$$

i.e. we have at least $\frac{n}{2^{P_0}}$ vertices which are not moving under the mapping. According to the Remark, this implies in almost all graphs that the parameter is at most P_0, thus, in almost all graphs the number of such pairs is at most

$$\sum_{p=P+1}^{P_0} N_p(n, \mathcal{G}) = o\big(N_P(n, \mathcal{G})\big).$$

Finally, we prove Theorem 3.

We have $|U \cup V| = m + s$ (s was the number of paths), therefore there are $\binom{m+s}{2}$ edges between the points of $U \cup V$. Let us call this graph $\mathcal{H}$. (Edges different from these ones are of no influence to the fact whether $\mathcal{G}_1$ and $\mathcal{G}_2$ are isomorphic or not). An edge (a, b) is said to be equivalent to an edge (c, d) if for some integer (possibly negative or zero)

$$a = \varphi^k(c), \quad b = \varphi^k(d)$$

or

$$a = \varphi^k(d), \quad b = \varphi^k(c),$$

i.e. if the repeated mapping φ^k takes the edge (a, b) to (c, d). This is obviously an equivalence relation among the edges of $\mathcal{H}$. Denote the number of equivalence classes by β and put

$$\alpha = \binom{m+s}{2} - \beta .$$

Since we can decide about one edge in each class arbitrarily (whether this edge should be drawn or not), and this determines all the edges of $\mathcal{H}$ (since we want $\mathcal{G}_1$ and $\mathcal{G}_2$ to be isomorphic), the probability that $\mathcal{G}_1$ and $\mathcal{G}_2$ are isomorphic equals $2^{-\alpha}$. Now we are going to determine the value of α and verify that it is the value given by (3).

This will be done by constructing a system S of edges of $\mathcal{H}$, which contains exactly one edge from each class.

We divide the edges of $\mathcal{H}$ into five groups: the edges within a path, between paths, between a path and a cycle, within a cycle and between cycles. We will refer to them as edges of category I, II, . . . , V, respectively. In a path of length d_i we choose the edges starting from the first vertex of the path, i.e. d_i edges. Thus we put

$$\text{(i)} \qquad \sum_{i=1}^{s} d_i = d$$

edges of category I to system S.

Between two paths of lengths d_i and d_j we choose the edges starting from the first vertices of the two paths (and going to the vertices of the other path), thus putting

$$\text{(ii)} \qquad \sum_{1 \leq i < j \leq s} (d_i + d_j + 1) = d(s-1) + \binom{s}{2}$$

edges of category II to system S.

Between a path of length d_i and a cycle of length c_j we choose the edges starting from the first vertex of the path (and going to the vertices of the cycle), thus we put

$$\text{(iii)} \qquad \sum_{i=1}^{s} \sum_{j=1}^{r} c_j = s \cdot c$$

edges of category III to system S.

Within a cycle of length c_i we choose one edge connecting two vertices of distance 1, one edge connecting two vertices of distance 2, etc. up to distance $\left[\frac{c_i}{2}\right]$.

(The distance of the vertices a and b is the smallest non-negative k for which $a = \varphi^k(b)$ or $b = \varphi^k(a)$.) Thus

$$\text{(iv)} \qquad \sum_{i=1}^{r} \left[\frac{c_i}{2}\right] = \frac{1}{2} c - \sum_{i=1}^{r} \left\{\frac{c_i}{2}\right\}$$

vertices are chosen to S.

Finally, between two cycles of lengths c_i and c_j we have $c_i \cdot c_j$ edges, and it is easy to see that these edges are split into equivalence classes of size $[c_i, c_j]$ (the least common multiple of c_i and c_j), thus we have

$$\frac{c_i \cdot c_j}{[c_i, c_j]} = (c_i, c_j)$$

equivalence classes, and choosing one edge from each class, we put

$$\text{(v)} \qquad \sum_{1 \leq i < j \leq r} (c_i, c_j) = \frac{1}{2} \sum_{i,j=1} (c_i, c_j) - \frac{1}{2} c$$

edges of category v to system S.

Summing the right-hand sides of (i)—(v) and using the relation $c + d = m$ we see that S contains

$$\beta = s \cdot m + \binom{s}{2} + \frac{1}{2} \sum_{i,j=1}^{r} (c_i, c_j) - \sum_{i=1}^{r} \left\{ \frac{c_i}{2} \right\}$$

edges and it is easy to see that S contains exactly one edge from each equivalence class. Thus

$$\alpha = \binom{m+s}{2} - \beta = \binom{m}{2} + \sum_{i=1}^{r} \left\{ \frac{c_i}{2} \right\} - \frac{1}{2} \sum_{i,j=1}^{r} (c_i, c_j),$$

proving Theorem 3.

(Received May 26, 1971)

MTA MATEMATIKAI KUTATÓ INTÉZETE,
BUDAPEST, V., REÁLTANODA U. 13–15.
HUNGARY