ON THE MEAN DISTANCE BETWEEN POINTS OF A GRAPH

Paul Erdős, János Pach, Joel Spencer

Given any graph G with vertex set V(G), let
$r(G) = \sum_{x \in V(G)} \frac{1}{\deg(x)}$, where deg(x) is the degree of x in G.
Further, let $\mu(G)$ denote the <u>mean distance</u> between vertices of G, i.e.,

$$\mu(G) = \frac{\sum d(x,y)}{\binom{|V(G)|}{2}},$$

where d(x,y) is the length of the shortest path connecting x and y in G, and the sum is taken over all pairs {x,y} of distinct vertices.

The following attractive conjecture was made by a computer instructed by Siemion Fajtlowicz (University of Houston-University Park) to search for hidden relations between various parameters of graphs.

<u>Conjecture</u>. $\mu(G) \leq r(G)$ for every connected graph G.

The authors of the present note are greatly impressed by the increasing number of contributions to mathematics made possible by clever computer investigations. Nevertheless, those who have already experienced the miserable feeling of being beaten by a chess program, will perhaps share the satisfaction we felt at the disproof of the above conjecture.

For any natural number n and r>0, set

$$\mu(n,r) = \max\{\mu(G) : G \text{ is connected}, |V(G)| = n, r(G) \leq r\},$$
$$\mathrm{diam}(n,r) = \max\{\mathrm{diam}(G) : G \text{ is connected}, |V(G)| = n, r(G) \leq r\},$$

where $\mathrm{diam}(G) = \max\{d(x,y) : x,y \in V(G)\}$ <u>diameter</u> of G.

We will show the following.

CONGRESSUS NUMERANTIUM **64**(1988), pp.121-124

Theorem. For any fixed $r \geq 3$,

$$\left(\frac{2}{3}[r3]+o(1)\right)\frac{\log n}{\log\log n} \leq \mu(n,r) \leq \mathrm{diam}(n,r) \leq (6r+o(1))\frac{\log n}{\log\log n},$$

where the o(1) term goes to 0 as n tends to infinity.

Proof. Let U_i, $0 \leq i \leq 2[r/3]k$, be disjoint sets with

$$|U_i| = k^{k-|j|},$$

where $j \in (-k,+k]$ is the unique integer satisfying $i \equiv j \pmod{2k}$. Define a graph G_k by

$$V(G_k) = \bigcup_{0 \leq i \leq 2[r/3]k} U_i,$$

$$E(G_k) = \{xy : x,y \in U_i \cup U_{i+1} \text{ for some } 0 \leq i < 2[r/3]k\}.$$

By simple calculations,

$$|V(G_k)| = n < rk^k,$$

$$r(G_{k)} < r,$$

$$\mu(G_k \geq \frac{2[r/3]k}{3} > \frac{2}{3}[/3]\frac{\log n}{\log\log n},$$

provided that k is large enough. This established the lower bound.

To prove the upper bound, fix a connected graph G with n vertices and $r(G) \leq r$, and fix a point $x \in V(G)$. Let

$$V_i = \{y \in V(G) : d(x,y)=i\}$$

and $|V_i| = n_i$ for any $0 \leq i \leq m = \max\{d(x,y) : y \in V(G)\}$. Then, setting $n_{-1}=n_{m+1}=0$,

$$r(G) = \sum_{x} \frac{1}{\deg(x)} \geq \sum_{0\leq i\leq m} \frac{n_i}{n_{i-1}+n_i+n_{i+1}} .$$

Thus, for all but at most 3r values of i,

$$\frac{n_i}{n_{i-1}+n_i+n_{i+1}} \leq \frac{1}{3} , \ n_i \leq \frac{1}{2}(n_{i-1}+n_{i+1}).$$

This implies that there exist $0\leq i_0<i_1\leq m$ such that

$$i_1 - i_0 \geq \frac{1}{2} \frac{m + 1 - 3r}{3r + 1}$$

and n_i is monotonic (say, monotone increasing) on the interval $i \in [i_0,i_1]$. In view of the fact that

$$\sum_{i_0<i<i_1} \frac{n_i}{n_{i+1}} \leq 3 \sum_{i_0<i<i_1} \frac{n_i}{n_{i-1}+n_i+n_{i+1}} < 3r ,$$

we obtain

$$n_{i_0+1} = \left(\prod_{i_0<i<i_1} \frac{n_i}{n_{i+1}}\right) n_{i_1} < \left(\frac{3r}{i_1-i_0-1}\right)^{i_1-i_0-1} n_{i_1} ,$$

whence

$$1 \leq n_{i_0+1} < \frac{3r(6r+2)}{m-9r-2}^{\frac{m-9r-1}{6r+2}} n.$$

This yields

$$m = \max\{d(x,y) : y \in V(G)\} \leq (6r+o(1))\frac{\log n}{\log\log n} ,$$

as desired. □

Note that the above estinmates remain valid (apart from the values of the constants) for all $r \le (\log n)^{1-\varepsilon}$, $\varepsilon>0$. On the other hand, if $r>\log n$, then our argument yields $m \ge rO(\log n)$.

For some other problems and results on mean distance see [1]-[7].

REFERENCES

[1] F. Buckley and L. Superville, Distance distributions and mean distance problems, *Proc. 3rd Caribbean Conf. on Combinatorics and Comp.*, Bridgetown, 1981, 67-76.

[2] F. Buckley and L. Superville, Mean distance in line graphs, *Congressus Numerantium* 32 (1981), 153-162.

[3] K. K. Doyle and J. E. Graver, Mean distance in a graph, *Discrete Math.* 17 (1977), 147-154.

[4] R. E. Jamison, On the average number of nodes in a subtree of a tree, *J. Combin. Th.* 35 (1983), 207-223.

[5] M. Paoli, Comparison of mean distance in superposed networks, *Discrete Appl. Math.* 8 (1984), 279-287.

[6] J. plesnik, On the sum of all distances in a graph or digraph, *J. Graph Th.* 8 (1984) 1-21.

[7] P. Winkler, Mean distance and the "four-thirds conjecture," to appear.