MEDIANS AND DISTANCE SEQUENCES IN GRAPHS

Zevi Miller

Abstract

The distance of a point v in a graph G is the sum of all distances from v to other points in the graph. The median, M(G), of G is the subgraph of G induced by points having minimum distance. A recent result of Slater shows that for any graph H there is a graph G such that $M(G) \cong H$ and the graph G constructed has $O(|H|^3)$ points. We simplify his construction and in the process show that if H has no isolated points, then G need have at most $2|H|$ points. Let d(G) be the maximum distance over all points in graph G. We consider the extremal problem corresponding to the one considered by Erdős and Renyi for the diameter, of determining the minimum number of edges in a graph G having a given number of points and a given value of d(G).

1. *Introduction.*

The notions of centrality, distance, and eccentricity in graphs have been the subject of several recent investigations. Apart from the intrinsic interest of these notions from a graph theoretic point of view, they are useful in the analysis of problems of optimum facility location in a generalized communication network. Among the results obtained so far are ones which concern both the center of a graph and its eccentricity sequence. It is the object of this paper to obtain corresponding results for the median of a graph and its distance sequence.

We begin with some notation.

Let $G = (V(G),E(G))$ be a connected graph with no loops or multiple edges, where V(G) is its set of points and E(G) is its set of edges. For each $v \in V(G)$, let $N_G(v)$ (or just N(v) when G is understood) be the set of points in G which are adjacent to v.

ARS COMBINATORIA, Vol. 15 (1983), pp. 169-177

Given $v,w \in V(G)$, define $d_G(v,w)$ (the *distance between* v *and* w) to be the least number of edges in a path joining v and w in G. For $v \in V(G)$, define the *eccentricity* of $v, e_G(v)$, by $e_G(v) = \max\{d_G(v,w)\colon w \in V(G)\}$. We also define the *distance* of v to be $d_G(v) = \sum_{w \in G} d_G(v,w)$. Now if $S \subseteq V(G)$, we let $\langle S \rangle$ (the *graph induced by* S) be the graph defined by

$$V(\langle S \rangle) = S, \quad \text{and}$$
$$E(\langle S \rangle) = \{xy\colon \; xy \in E(G),\ x \in S,\ y \in S\}.$$

We may then define the *center,* $C(G)$, of G and the *median,* $M(G)$, of G by

$$C(G) = \langle v \in V(G)\colon \; e_G(v) \le e_G(w),\ w \in V(G) \rangle, \quad \text{and}$$
$$M(G) = \langle v \in V(G)\colon \; d_G(v) \le d_G(w),\ w \in V(G) \rangle.$$

Denote by C_n the n-cycle graph

$$V(C_n) = \mathbb{Z}_n,\ E(C_n) = \{i(i+1)\colon \; i \in \mathbb{Z}_n\}.$$

Letting $[x]$ (resp. $\{x\}$) denote the greatest integer smaller than (resp. least integer larger than) x, define the graph C_n^j, $1 \le j \le [\frac{n}{2}]$, by $V(C_n^j) = V(C_n)$ and $E(C_n^j) = \{st\colon \; |s-t| \le j \text{ modulo } n\}$. Given graph G, let $\bar{G}$ be graph defined by $V(G) = V(\bar{G})$ and $E(\bar{G}) = \{xy\colon \; xy \notin E(G)\}$, let $q(G) = |E(G)|$, and let $|G| = |V(G)|$. Any terminology we use that is not mentioned above may be found in [1] or [3].

2. *Centers and Medians.*

The first known result concerning centers or medians of graphs, due to Jordan [4], states that for any tree T we have $C(T) = K_1$ or K_2 and $M(T) = K_1$ or K_2. A recent result of Kopylov and Timofeev [5], given without proof, states that given any H there is a graph G such that $C(G) \cong H$. A simple construction shows that G need have only four more points than H.

In analogy with this result on central embeddings, we consider median embeddings.

The fact that for any graph H there is a graph G such that $C(G) \cong H$ suggests the problem of determining whether there exists a graph G such that $M(G) \cong H$. This problem was solved by Slater [7] who showed that the construction of such a G is possible for any H. It is then natural to define for any graph H the parameter $\beta(H) = \min\{|G|: M(G) \cong H\}$. The construction of Slater provided an upper bound for $\beta(H)$ which is of order $O(|H|^3)$. We simplify his construction and give an improvement on this bound for β when H has no isolates (i.e. isolated points) in the following theorem.

THEOREM 1. *For any graph H with no isolated points we have $\beta(H) \leq 2|H|$.*

Proof. Let H be a graph with no isolates. We construct a graph G satisfying $|G| = 2|H|$ and $M(G) \cong H$.

Let $V(H) = \{x_1, x_2, \ldots, x_n\}$. Define G as follows:

$$V(G) = V(H) \cup \{x_1', x_2', \ldots, x_n'\},$$

$$E(G) = E(H) \cup \{x_i x_i': 1 \leq i \leq n\} \cup \left(\bigcup_{i=1}^{n} \{x_i' t: \ t \in V(H) \setminus N_H(x_i),\ t \neq x_i\}\right).$$

Thus G is formed by joining each x_i' to x_i and to precisely those points of H not adjacent to x_i. The portion of G "near" a pair of points x_i, x_i' is illustrated in Figure 1.

We now show that $M(G) \cong H$. For any $v \in V(G)$ let $S_j(v)$ be the set of points in G at distance j from v. Then we have

$$S_1(x_i) = N_H(x_i) \cup \{x_i'\} \cup \{x_j': \ x_j \notin N_H(x_i)\} \quad \text{and}$$

$$S_2(x_i) = ((V(H) \setminus \{x_i\}) \setminus N_H(x_i)) \cup \{x_j': \ x_j \in N_H(x_i)\}, \quad \text{so that}$$

$$S_j(x_i) = \emptyset \quad \text{for } j \geq 3. \quad \text{Hence}$$

$$\begin{aligned} d_G(x_i) &= |S_1(x_i)| + 2|S_2(x_i)| \\ &= |N_H(x_i)| + n - |N_H(x_i)| + 2(n - |N_H(x_i)| - 1 + |N_H(x_i)|) \\ &= n + 2n - 2 = 3n - 2. \end{aligned}$$

We also have

$S_1(x_i') = \{x_i\} \cup \{x_j\colon\ x_j \notin N_H(x_i)\}$. Hence we get

$$\begin{aligned} d_G(x_i') &\geq |S_1(x_i')| + 2(|G| - 1 - |S_1(x_i')|) \\ &= n - |N_H(x_i)| + 2(|N_H(x_i)| + n - 1) \\ &= 3n - 2 + |N_H(x_i)|. \end{aligned}$$

Since H has no isolates we have $d_G(x_i') > 3n - 2$ for all i, and hence $M(G) \cong H$. □

3. *Distance Sequences.*

Consider the set of integers $\{e_G(v)\colon\ v \in V(G)\}$ for a connected graph G. Arranging this set in ascending order we obtain the so called *eccentricity sequence* of G. Several results in the literature deal, at least implicitly, with this sequence. In [6] Lesniak-Foster gives a characterization of the integer sequences which are realizable as eccentricity sequences. The many extremal results concerning the diameter of a graph may be viewed as dealing with the maximum term in the eccentricity sequence. A typical example of such a result is the analysis of the parameter $e_d(n,k) = \min\{q(G)\colon\ |G| = n,\ \mathrm{diam}(G) \leq d,\ \Delta(G) = k\}$ (where Δ is maximum degree) carried out by Erdős and Renji in [2].

The set of distances $\{d_G(v)\colon\ v \in V(G)\}$ of a connected graph G on n points may be arranged to form the nondecreasing sequence $d_1(G) \leq d_2(G) \leq \ldots \leq d_n(G) = d(G)$, which we shall call the *distance sequence* of G. Notice that $n - 1 \leq d(G) \leq \frac{n(n-1)}{2}$, where the lower and upper extremes are realized by $G = K_n$ and $G = P_n$ respectively. We now consider the analogue of the Erdős-Renji problem for distance without the restriction on maximum degree. That is, we let $d(G)$, the maximum term in the distance sequence, play the role of $\mathrm{diam}(G)$, the maximum term in the eccentricity sequence. Letting $n > 1$ and k be positive integers with $n-1 \leq k \leq \frac{n(n-1)}{2}$, the problem is then to determine $f(n,k) = \min\{q(G)\colon\ |G| = n,\ G \text{ is connected},\ d(G) = k\}$. We will determine $f(n,k)$ exactly for those k lying in certain initial and terminal subintervals of the possible range $[n-1, \frac{n(n-1)}{2}]$. For the remaining k upper bounds for $f(n,k)$ are obtained.

We begin with a series of lemmas.

LEMMA 2.1. *$f(n,k) \geq n(n-1) - \{\frac{kn-1}{2}\}$.*

Proof. Observe that if G is a connected graph on n points with minimum degree $\delta(G)$, then $d(G) \geq 2(n-1) - \delta(G)$. It follows that if $d(G) = k$ then $\delta(G) \geq 2(n-1) - k$. Hence we get

$$q(G) \geq [\frac{\delta(G)n+1}{2}] \geq [(n-1)n - (\frac{kn-1}{2})]$$

$$= n(n-1) - \{\frac{kn-1}{2}\}. \quad \square$$

The following two lemmas are used in the extremal constructions of Theorem 2.

LEMMA 2.2: *$\overline{C_n^j}$ has diameter 2 if and only if $1 \leq j \leq [\frac{n-2}{3}]$.*

Proof. For convenience let the points of C_n^k be $\{1,2,\ldots,n\}$ so that $st \in E(C_n^k)$ if and only if $|s-t| \leq k$ modulo n. For $i \in V(\overline{C_n^k})$, let $\overline{N(i)}$ denote the neighborhood of i in $\overline{C_n^k}$.

Now the condition $\mathrm{diam}(\overline{C_n^k}) = 2$ is equivalent to $\overline{N(1)} \cap \overline{N(k+1)} \neq \emptyset$. Since $\overline{N(k+1)} = \{2k+2, 2k+3, \ldots, n\}$ and $\overline{N(1)} = \{k+2, k+3, \ldots, n-k\}$, the stated intersection holds if and only if $2k+2 \leq n-k$ or equivalently $k \leq [\frac{n-2}{3}]$. The lemma is thereby proved. $\square$

LEMMA 2.3. *For any k satisfying $\{\frac{5}{2}n\} - 4 \leq k \leq \frac{n(n-1)}{2}$ there exists a tree T on n points such that $d(T) = k$.*

Proof. Define a tree $S_\ell(r,t)$, where $\ell + r + t = n$, $\ell \geq 2$, as follows. Begin with the path P_ℓ and let z be an endpoint of P_ℓ and y the point of P_ℓ adjacent to z. Then $S_\ell(r,t)$ is obtained from P_ℓ by joining t endpoints to z and r endpoints to y. We note that if a tree T is of the form $S_\ell(r,t)$ then ℓ, r and t are uniquely determined by G with the exception of the case $\ell = 2$ where we have $S_2(r,t) = S_3(r-1, t)$. For what follows below, we $\cong S_2(r,t)$ for some $r \geq 1$, then we denote G by $S_2(r,t)$ if $r > 1$ and by $S_3(0,t)$ if $r = 1$.

We now construct the required trees inductively. Clearly $d(S_2([\frac{n-2}{2}], \{\frac{n-2}{2}\})) = \{\frac{5}{2}n\} - 4$, so the lemma is proved for the smallest value of k. Let k be in the given range, and suppose inductively that we have constructed a tree T of the form $S_\ell(r,t)$ satisfying $d(S_\ell(r,t)) = k - 1$. Then define a tree $\tilde{T}$ by

$$\tilde{T} = \begin{cases} S_\ell(r-1,\ t+1) & \text{if } r \geq 1 \\ S_{\ell+1}(t-2,\ 1) & \text{if } r = 0. \end{cases}$$

It is easy to check that $d(\tilde{T}) = d(T) + 1 = k$, and hence the lemma is proved. □

We may now proceed to the determination of $f(n,k)$.

THEOREM 2. *Let* $n \geq 6$ *be an integer, and let* $A = n - 1 + 2[\frac{n-2}{3}]$, $B = \{\frac{5}{2}n\} - 4$, $C = 2n - 3$. *Then*

$$f(n,k) = \begin{cases} n(n-1) - \{\frac{kn-1}{2}\} & \text{if } n - 1 \leq k \leq A \\ n - 1 & \text{if } B \leq k \leq \frac{n(n-1)}{2} \quad \text{or} \quad k = C \end{cases}$$

Furthermore we have the bounds

$$f(n,k) \leq \begin{cases} n - 1 + (C-k)(\frac{n-2}{2}) + [\frac{C-k}{2} + \frac{1}{2}(\frac{1+(-1)^n}{2})] & \text{if } A + 1 \leq k \leq C - 1 \\ n - 1 + 2(B-k) - (\frac{1+(-1)^{n+1}}{2}) & \text{if } C + 1 \leq k \leq B - 1. \end{cases}$$

Proof. Suppose first $n - 1 \leq k \leq A$. For the special case $k = n - 1$, define the graph G_k by $G_k = K_n$. Then clearly $d(G_k) = n - 1$ so that $f(n,n-1) \leq q(K_n) = \binom{n}{2}$. But by Lemma 2.1 we have $f(n,n-1) \geq n(n-1) - \{\frac{(n-1)n-1}{2}\} = \binom{n}{2}$, and hence the theorem follows for $k = n - 1$. For $k > n - 1$, define $E_k \subset E(C_n^{\{\frac{k-(n-1)}{2}\}})$ as follows. For n even let E_k be a perfect matching. For n odd, let E_k be the disjoint union $P \cup M$, where M is a matching having $\frac{n-3}{2}$ edges and P is a pair of edges forming a path of length 2 on the three points nonincident with M. Now define the graph G_k by

$$V(G_k) = V(C_n)$$

$$E(G_k) = \begin{cases} E(\overline{C_n^{\{\frac{k-(n-1)}{2}\}}}) \cup E_k & \text{if } k - (n-1) \text{ is odd} \\ E(\overline{C_n^{\{\frac{k-(n-1)}{2}\}}}) & \text{if } k - (n-1) \text{ is even.} \end{cases}$$

Since k lies in the range $n - 1 \le k \le A$ the graphs G_k so defined have diameter 2 by Lemma 2.2. It follows that $d(G_k) = n - 1 + n - 1 - \delta(G_k)$. Since $\delta(G_k) = 2(n - 1) - k$, we get $d(G_k) = k$. Finally since G_k is $2(n - 1) - k$ regular if n is even, while it has $n - 1$ points of degree $2(n - 1) - k$ and one point of degree $2(n - 1) - k + 1$ if n is odd, we have $f(n,k) \le q(G_k) = [\frac{(2(n - 1) - k)n + 1}{2}] = n(n - 1) - \{\frac{kn - 1}{2}\}$. Combining this with Lemma 2.1 we get $f(n,k) = n(n - 1) - \{\frac{kn - 1}{2}\}$ as desired.

Now suppose $B \le k \le \frac{n(n - 1)}{2}$. We have $f(n,k) \ge n - 1$ for all k since any connected graph has at least $n - 1$ edges. On the other hand, Lemma 2.3 implies that $f(n,k) \le n - 1$ for k in the given range, and hence the result is proved.

When $k = C$, let $G = K_{1,n-1}$. Then clearly $d(G) = 2n - 3 = k$ and $q(G) = n - 1$ so it follows as above that $f(n,C) = n - 1$.

The upper bounds follow from certain constructions. These are omitted for brevity. □

4. *Open Problems.*

It seems that the bound $\beta(H) \le 2|H|$ should be amenable to improvement if consideration is restricted to special classes of graphs H. Slater has shown (personal communication) that $\beta(P_n) \le n + 3$. It would be interesting to see if $\beta(H) \le 2|H|$ is best possible in the sense that there is an infinite class of graphs for which $\beta(H) = 2|H|$.

The determination of $f(n,k)$ is not quite complete. Although $f(n,k)$ has been "mostly" determined (in the sense that of the possible interval $[n - 1, \frac{(n-1)n}{2}]$ of length $0(n^2)$ for k, we have determined $f(n,k)$ for $k = 2n - 3$ and all $k \neq 2n - 3$ lying outside the open interval (A,B) of length approximately $\frac{5}{6}n = 0(n)$), it would still be nice to find $f(n,k)$ for all k. I would conjecture that the upper bounds given here for $f(n,k)$, $k \varepsilon (A,B)$, are in fact the exact values of $f(n,k)$.

Finally, there is the problem of characterizing the set of integer sequences which are distance sequences for graphs. This problem is of course the analogue of the one solved by Lesniak-Foster in [6], and it seems at present to be difficult. Even the more restricted problem of characterizing distance sequences for trees has so far eluded the author's many attempts at solution.

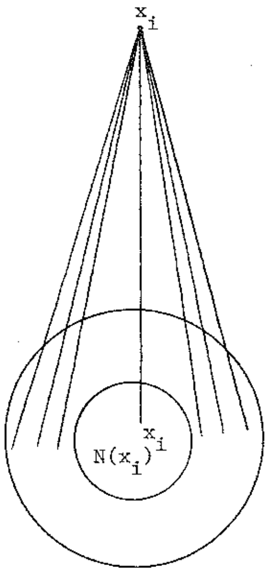

Fig. 1. The graph G at x_i and x_i'

References

[1] M. Behzad, G. Chartrand, and L. Lesniak-Foster, *Introduction to the Theory of Graphs*, Prindle, Weber, and Schmidt (1979).

[2] P. Erdős, A. Renyi, *On a problem in the theory of graphs* (in Hungarian), Publ. Math. Inst. Hungar. Acad. Sci. 7 (1962).

[3] F. Harary, *Graph Theory*, Addison Wesley (1969).

[4] C. Jordan, *Sur les assemblages des lignes*, J. Reine Angew. Math. 70 (1869) 185-190.

[5] G.N. Kopylov, E. Timofeev, *Centers and radii of graphs* (Russian), Uspekhi Mat. Nauk 32, No. 6, (1977) 226.

[6] L. Lesniak, *Eccentric sequences in graphs*, Periodica Mathematica Hungarica Vol. 6, 4 (1975) 287-293.

[7] P. Slater, *Medians of arbitrary graphs*, Journal of Graph Theory, Vol. 4, No. 4 (1980) 389-392.

Department of Mathematics and Statistics
Miami University
Oxford, Ohio 45056
U.S.A.