Roger C. Entringer; Douglas E. Jackson; D. A. Snyder
Distance in graphs

*Czechoslovak Mathematical Journal*, Vol. 26 (1976), No. 2, 283–296

Persistent URL: http://dml.cz/dmlcz/101401

Czechoslovak Mathematical Journal, 26 (101) 1976, Praha

# DISTANCE IN GRAPHS

R. C. ENTRINGER, Albuquerque, D. E. JACKSON, Los Alamos, D. A. SNYDER*), Albuquerque

(Received November 28, 1974)

## 1. INTRODUCTION

In this article we are dealing only with finite, undirected graphs without loops or multiple edges. We define the *distance of a vertex* $p$, $d(p)$, of such a connected graph $G$ by $d(p) = \sum d(p, q)$ where $d(p, q)$ is the distance between $p$ and $q$ and the summation extends over all vertices $q$ of $G$. (We will consistently use $\sum$ to represent summation over the vertex set of the graph under consideration.) If $G$ is not connected we set $d(p) = \infty$. We further define the *distance of a graph* $G$, $d(G)$, to be $\frac{1}{2} \sum d(p)$ if $G$ is connected and $\infty$ otherwise.

The digraph analogs of these concepts were investigated by HARARY [1] is a sociometric framework: that of precisely defining and measuring the "status" of an individual within an organization. See also [2].

ORE [3], p. 30, suggested the ratio $d(p)/n$, where $n$ is the number of vertices of $G$, as a measure of the "centrality" of a vertex $p$ in a graph. He suggested this concept be studied in the case of trees [3], p. 66, and this was consequently done by ZELINKA in [5] and [6]. We will quote several of Zelinka's results in a latter section.

The notion of "point-centrality" (= distance) of a vertex of a graph, having appeared in several psychometric studies, was studied in some detail by SABIDUSSI [4]. His principal purpose was that of refuting several intuitive conclusions drawn in these studies.

Our purpose in this paper is to collect and extend the results referred to above. We are interested only in the graph-theoretic properties of distance and we will not enter into any discussions concerning the appropriateness of the use of distance in measuring "centrality" in organizational structures.

Our first results, in the next section, deal generally with the size of $d(p)$ and $d(G)$. There, and throughout the paper, if $S$ is a set, $|S|$ will denote its cardinality.

All paths and cycles will be understood to be simple, i.e., no vertex will be repeated.

The cycle on $n$ vertices will be denoted by $C_n$ with $C_3$ sometimes being called

*) Research supported by the United States Atomic Energy Commission.

a triangle. $K_n$ will denote the complete graph on $n$ vertices and $K_{n,m}$ the complete bipartite graph; $K_1$ is the trivial graph. The complement of a graph $G$ is denoted by $\bar{G}$.

The degree of a vertex is the number of edges incident with it; a graph is regular if all vertices have the same degree.

The girth of a graph $G$ is the number of vertices in a smallest cycle of $G$ while the diameter of $G$ is max $d(p, q)$ where the maximum is taken over all vertices $p$ and $q$ of $G$; the diameter is said to be infinite if $G$ is not connected. We note that if $G$ has cycles then its girth is at most 1 plus twice its diameter.

$G$ is totally disconnected if it has no edges and connected if any two distinct vertices lie in a path. A component of $G$ is a maximal connected subgraph. A (An) vertex (edge) is a cut vertex (bridge) of $G$ if its deletion increases the number of components.

Finally, a connected graph $G$ is called $k$-connected if the deletion of any $k - 1$ vertices yields a connected nontrivial graph.

With these definitions in mind we proceed to our first results.

## 2. BOUNDS ON DISTANCE

We first note two useful relations concerning distance. In the first, $G - p$ is the graph obtained from $G$ by deleting vertex $p$ and all edges incident to $p$.

**Property 2.1.** *$d(G) \leqq d(p) + d(G - p)$ for each point $p$ of a graph $G$.*

Proof. We may write $d(G) = \sum' d(q, r)$ where $\sum'$ indicates summation over the set of unordered pairs of distinct vertices of $G$. Now $\sum' d(q, r) = d(p) + \sum'' d(q, r)$ where $\sum'' d(q, r)$ indicates summation over the set of unordered pairs of vertices of $G - p$. Then, since $d(q, r)$ is at most the distance between $q$ and $r$ in $G - p$ we have the desired inequality.

**Property 2.2.** *Suppose $a$ and $b$ are adjacent vertices of a connected graph $G$. Let $A$ be the set of vertices closer to $a$ than $b$, and $B$ the set of vertices closer to $b$ than $a$. Then $d(a) - d(b) = |B| - |A|$.*

Proof. $d(a) = \sum_{p \in A} d(a, p) + \sum_{p \in B} d(a, p) + \sum_{p \in C} d(a, p)$ where $C$ is the set of vertices equidistant from $a$ and $b$. Combining this with a similar equality for $b$ we have $d(a) - d(b) = \sum_{p \in A} [d(a, p) - d(b, p)] + \sum_{p \in B} [d(a, p) - d(b, p)]$ which in turn is equal to $-|A| + |B|$ since the adjacency of $a$ and $b$ implies $d(a, p) - d(b, p)$ is $-1$ if $p \in A$ and 1 if $p \in B$.

The distances of many well known graphs are easy to determine because of their symmetry. We have, for example, $d(P_n) = \frac{1}{6}(n^3 - n)$ and

$$d(C_n) = \begin{cases} \frac{1}{8}n(n^2 - 1) & n \text{ odd} \\ \frac{1}{8}n^3 & n \text{ even} \end{cases}$$

where $P_n$ and $C_n$ are the path and cycle on $n$ vertices respectively. We also have $d(K_{m,n}) = mn + m(m - 1) + n(n - 1)$ and $d(K_n) = \frac{1}{2}n(n - 1)$.

Our first theorem shows that paths and complete graphs achieve the extreme values of distance.

**Theorem 2.3.** *If $G$ is a connected graph with $n$ vertices and $k$ edges then* $n(n - 1) \leqq d(G) + k \leqq \frac{1}{6}(n^3 + 5n - 6)$.

Proof. To obtain the lower bound we write $d(G) = k + \frac{1}{2}\sum' d(p, q)$ where the summation is over the set $S$ of all ordered pairs of distinct non-adjacent vertices $p$ and $q$ of $G$. Since $|S| = n(n - 1) - 2k$ we have $d(G) \geqq k + \frac{1}{2}\{2[n(n - 1) - 2k]\} = n(n - 1) - k$.

The upper bound is obtained by induction on $k$ after showing by induction on $n$ that the bound holds for trees. To see the latter, observe that it holds for the tree $K_1$ and assume that for all trees $T$ on $n$ vertices $d(T) + n - 1 \leqq \frac{1}{6}(n^3 + 5n - 6)$. If $T'$ is a tree on $n + 1$ vertices of which $p$ is an end-vertex then $d(p) \leqq \sum_{i=1}^{n} i$ so that by Property 2.1 we have $d(T') \leqq \frac{1}{2}n(n + 1) + \frac{1}{6}(n^3 + 5n - 6) - n + 1 = \frac{1}{6}[(n + 1)^3 + 5(n + 1) - 6] - n$. We now have the upper bound of the theorem holding for all connected graphs with $n$ vertices and $n - 1$ edges and assume it holds for all connected graphs with $n$ vertices and $k \geqq n - 1$ edges. Let $G$ be any connected graph on $n$ vertices and $k + 1$ edges. Then, since $G$ is not a tree it contains an edge $e$ so that $G - e$ is a connected graph on $n$ vertices and $k$ edges and consequently satisfies $d(G - e) + k \leqq \frac{1}{6}(n^3 + 5n - 6)$. But obviously $d(G - e) \geqq 1 + d(G)$ so that $d(G) + k + 1 \leqq \frac{1}{6}(n^3 + 5n - 6)$.

A similar result holds for the distance of a vertex.

**Theorem 2.4.** *If $p$ is any vertex of a connected graph $G$ with $n$ vertices and $k$ edges then $n - 1 \leqq d(p) \leqq \frac{1}{2}(n - 1)(n + 2) - k$ and these bounds can be achieved for each $k$, $n - 1 \leqq k \leqq \binom{n}{2}$.*

Proof. The lower bound is obvious; that this bound is best is easily seen from consideration of any graph on $n$ vertices and $k$ edges one vertex of which has degree $n - 1$.

We prove $d(p) \leqq \frac{1}{2}(n - 1)(n + 2) - k$ by induction on $k$. Since $G$ is connected we first consider $k = n - 1$ i.e., $G$ is a tree. If $p$ is any vertex of $G$ and $d_i$ is the number of vertices whose distance from $p$ is $i$, then $d(p) \leqq \sum_{i=1}^{n-1} id_i$. But, since $\sum_{i=1}^{n-1} d_i = n - 1$ and $d_i = 0$ implies $d_{i+1} = 0$, $\sum_{i=1}^{n-1} id_i$ maximizes at $\sum_{i=1}^{n-1} i$ so that $d(p) \leqq \frac{1}{2}n(n - 1) = \frac{1}{2}(n - 1)(n + 2) - (n - 1)$.

Now assume that the upper bound holds for all connected graphs on $n$ vertices and $k$ edges and let $p$ be any vertex of a connected graph $G$ with $n$ vertices and $k + 1$ edges. Since $G$ is not a tree, it contains a vertex $q$ lying in a cycle $C$ and for which $d(p, q)$ is a minimum over all $q$ lying in cycles ($q$ may be $p$). Let $r$ be a vertex of $C$ adjacent to $q$ and consider the graph $G - qr$.

This graph is connected, has $n$ vertices and $k$ edges and the distance between $p$ and $r$ in it is greater than the distance between $p$ and $r$ in $G$. This last assertion may be seen by consideration of the shortest paths between $p$ and $r$ in $G - qr$. If every such path contains $q$ then the distance between $p$ and $r$ in $G - qr$ is greater than this distance in $G$. If there is a shortest path $A$ between $p$ and $r$ in $G - qr$ that does not contain $q$ we choose a shortest path in $G$, $B = (p = b_0, b_1, \ldots, b_m = q)$ between $p$ and $q$. Now choose the largest index $i$ for which $b_i$ is a vertex of $A$ also. This vertex is closer to $p$ than is $q$ and lies in a cycle consisting of $b_i, b_{i+1}, \ldots, b_m = q$, $r$ and the vertices of $A$ between $b_i$ and $r$. Since this cannot be we conclude, in view of the previous discussion, that the distance between $p$ and $r$ is greater in $G - qr$ than in $G$.

We now apply the induction hypothesis to obtain

$$d(p) \leqq -1 + [\tfrac{1}{2}(n - 1)(n + 2) - k] = \tfrac{1}{2}(n - 1)(n + 2) - (k + 1)$$

and the proof that the upper bound holds is complete.

To show this bound is actually achieved for each $k$, $n - 1 \leqq k \leqq \binom{n}{2}$, consider the graph of Fig. 2.1.

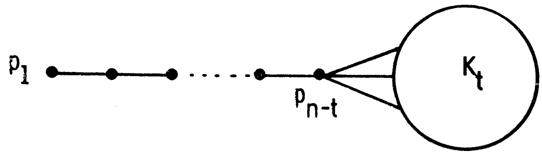


Fig. 2.1.

In this graph $t$ is chosen as the largest integer for which $s(t) = k - n + 1 - \tfrac{1}{2}t(t - 3)$ is non-negative. Then, since $t - 1 = s(t) - s(t + 1) \geqq s(t) - 1$ implies $s(t) \leqq t$, we make vertex $p_{n-t}$ adjacent to exactly $s(t)$ vertices of $K_t$. This graph then has $n$ vertices, $n - t - 1 + s(t) + \binom{t}{2} = k$ edges and vertex $p_1$ has distance $d(p_1) = \sum_{i=1}^{n-t-1} i + (n - t)\, s(t) + (n - t + 1)\,[t - s(t)] = \tfrac{1}{2}n(n - 1) - s(t) - \tfrac{1}{2}t(t - 3)$ so that $d(p_1) + k = \tfrac{1}{2}(n - 1)(n + 2)$ which is the upper bound on $d(p) + k$.

A natural refinement of this last theorem would involve determination of a function $f(n, k, d)$ for which every vertex of degree $d$ in any connected graph on $n$ vertices and $k$ edges has distance at most $f(n, k, d)$. However, a difficulty arises if one attempts to parallel the proof of Theorem 2.4. That proof utilized the existence of an edge $qr$ such that the vertex $p$ was closer to $r$ in $G$ than in $G - qr$ and this edge, $qr$, had to lie in a cycle so that $G - qr$ was connected. We showed that such an edge always existed; however, it was possible that $p$ and $q$ were identical. In this case the degree of $p$ in $G - qr$ was less than its degree in $G$ and this seems to be the essential difficulty in proving this extension of Theorem 2.4 by induction on the number of edges. Although we do not determine $f(n, k, d)$ we do obtain some allied results.

We first obtain a result analogous to that of Theorem 2.4; however, we are not able to claim our upper bound is achieved for all permissible values of $k$.

**Theorem 2.5.** *If $p$ is a vertex of degree $d$ in a connected graph with $n$ vertices and $k$ edges then $2n - d - 2 \leqq d(p) \leqq \frac{1}{2}(n-1)(n+2) - k$ and these bounds can be achieved for all $d$, $1 \leqq d \leqq n - 1$.*

Proof. The lower bound is obvious; that the upper bound holds is immediate from Theorem 2.4. They are simultaneously achieved in the graph $G$ consisting of $K_{n-1}$ together with an additional vertex $p$ adjacent to exactly $d$ vertices of $K_{n-1}$.

Our next result concerns graphs in which $d(p)$ is fixed. In the statement and proof of this we will consider a distinguished vertex $p$ of a graph and denote the set of vertices distance $i$ from $p$ by $D_i$. As usual we take $d_i = |D_i|$.

**Theorem 2.6.** *If over all graphs with $n$ vertices, one of which, $p$, has degree $d$ and given distance, $G$ is one with a maximum number of edges, then $d_{r+1} = 0$, $d_r > 0$, $d_i = 1$ implies $d_j = 1$ for $i \leqq j \leqq r - 2$.*

Proof. Since $G$ has a maximum number of edges every vertex of $D_s$, $s \geqq 1$ is adjacent to every other vertex of $D_s$ and to all vertices of $D_{s+1}$. We consider three cases. In each case we may take $t$ to be the largest index for which $d_i = d_{i+1} = \ldots$ $\ldots = d_t = 1$. If $t = r - 2$ we are done. Otherwise we have each of the following to consider:

i) Suppose $t = r - 3$. If a vertex is moved from $D_r$ to $D_{r-1}$ and made adjacent to all vertices in $D_r$, $D_{r-1}$ and $D_{r-2}$ there is a net gain of $d_{r-2}$ edges. If, now, a second vertex is moved from $D_{r-2}$ to $D_{r-1}$ and made adjacent to all the vertices now in $D_r$, $D_{r-1}$ and $D_{r-2}$ there is a net gain of $d_r - 2$ edges so that the total net gain is $d_r + d_{r-2} - 2$ which is positive since $d_r > 0$ and, by definition of $t$, $d_{r-2} > 1$. But this is impossible since $d(p)$ is not changed.

ii) Suppose $t = r - 4$. If vertices are moved, one from $D_r$ to $D_{r-1}$ and one from $D_{r-3}$ to $D_{r-2}$, the net gain in edges is $(d_{r-2} + d_{r-1} + 1) - 1 > 0$ which is impossible since $d(p)$ is not changed.

iii) Suppose $t \leqq r - 5$. If vertices are moved, one from $D_r$ to $D_{r-1}$ and one from $D_{t+1}$ to $D_{t+2}$, the net gain in edges now is $(d_{r-2} + d_{t+3}) - 1 > 0$ which also is impossible since $d(p)$ is not changed.

This completes the proof.

We next look at the relation between the distance of a graph $G$ and the distance of its complement $\bar{G}$. For this purpose we define $\bar{d}(p)$ to be the distance of $p$ in $\bar{G}$ and $\bar{d}(p, q)$ to be the distance between $p$ and $q$ in $\bar{G}$.

**Theorem 2.7.** *If $G$ is any graph on $n$ vertices then $d(G) + d(\bar{G}) \geqq \frac{3}{2}n(n-1)$ and this bound is best for $n \geqq 5$.*

Proof. Since two distinct vertices $p$ and $q$ cannot be adjacent in both $G$ and $\bar{G}$ we have $d(p, q) + \bar{d}(p, q) \geqq 3$ so that $d(G) + d(\bar{G}) = \frac{1}{2}\sum_{p\in G}\sum_{q\in G}[d(p, q) + \bar{d}(p, q)] \geqq$ $\geqq \frac{3}{2}n(n-1)$. To show the bound can be achieved we form a graph on $n \geqq 5$ vertices as follows. Partition the vertices in any manner into five sets $S_1, \ldots, S_5$. Now join every vertex in $S_i$ to every vertex in $S_{i+1}$, $i = 1, \ldots, 5$ ($S_6 = S_1$) as shown in Fig. 2.2.

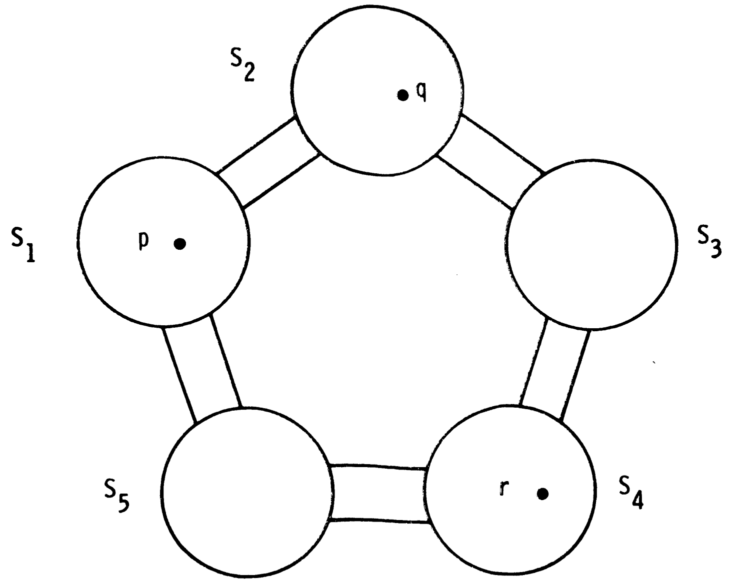


Fig. 2.2.

Suppose that $p$ and $q$ are distinct vertices of $G$. If they are adjacent in $G$ then they lie in $S_i$ and $S_{i+1}$ for some $i$. Now any vertex $r$ in $S_{i+3}$ is not adjacent to either $p$ or $q$ so that $d(p, q) + \bar{d}(p, q) = 3$. If $p$ is not adjacent to $q$ then either $p$ and $q$ both lie in $S_i$ for some $i$ or (relabelling $p$ and $q$ if necessary) $p$ is in $S_i$ and $q$ is in $S_{i+2}$. In each case any vertex $r$ in $S_{i+1}$ is adjacent to both $p$ and $q$ and again $d(p, q) +$ $+ \bar{d}(p, q) = 3$ so that $d(G) + d(\bar{G}) = \frac{3}{2}n(n-1)$.

The following is an immediate consequence of the theorem.

**Corollary 2.8.** *At least half of the graphs on $n$ vertices have distance $\frac{3}{4}n(n-1)$ or larger.*

We conclude this section with some miscellaneous samples of graphs illustrating certain properties (or lack thereof) with regard to distance.

I. A vertex of maximum distance in a connected graph need not have minimum degree. In the graph of Fig. 2.3 vertices $p, q, r$ and $s$ have distances $a + 3b + 1$, $a + 2b$, $2a + b - 1$ and $3a + 2b - 2$ respectively so that, for proper choice of $a$ and $b$, vertex $p$ has maximum distance yet relatively large degree.

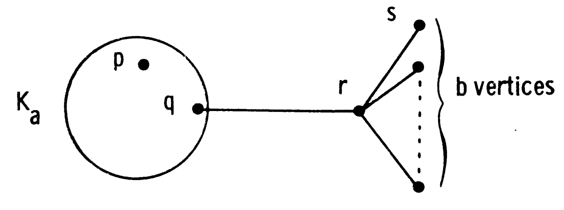


Fig. 2.3.

II. A vertex of maximum distance need not lie in a path of maximum length. In the graph of Fig. 2.4 vertices $p, q, r$ and $s$ have distances $5m + 3$, $4m + 1$, $4m - 1$ and $6m - 1$ respectively.

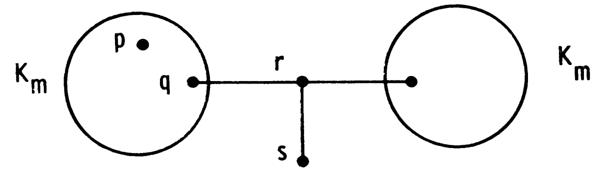


Fig. 2.4.

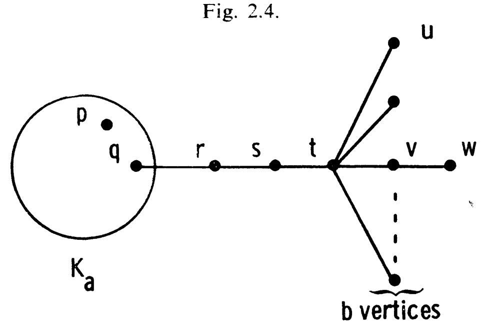


Fig. 2.5.

III. A central vertex $p$ of a connected graph $G$ is one for which $\max_{q \in G} (d(p, q))$ is a minimum. Our example here, Fig. 2.5, shows that a central vertex of a graph $G$ need not lie on any shortest path joining a vertex of maximum distance to a vertex of minimum distance. In this example, vertices $p, q, r, s, t, u, v$ and $w$ have distances

$a + 5b + 14$, $a + 4b + 10$, $2a + 3b + 6$, $3a + 2b + 4$, $4a + b + 4$, $5a + 2b + 6$, $5a + 2b + 4$ and $6a + 3b + 6$ so that if $a$ and $b$ are chosen to satisfy $a < b < \frac{5}{2}a - 4$ then $t$ has minimum distance, $w$ has maximum distance while $s$ is the central vertex.

IV. Our final examples shows that the set of distances of the vertices of a graph does not determine the graph uniquely. The two graphs of Fig. 2.6 each have two

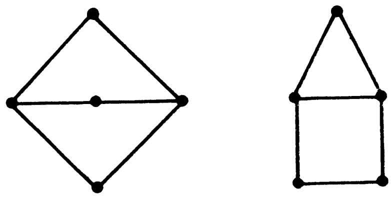

Fig. 2.6.

vertices with distance 5 and three vertices with distance 6 and the graphs are obviously not isomorphic. To obtain an infinite class of such pairs of graphs we simply take each of those of Fig. 2.6 and add in vertices, these added vertices being adjacent to all other added vertices and to each vertex of the original graph.

## 3. DISTANCE IN TREES

If the distance of a vertex is taken as a measure of its "centrality" in a graph it would seem appropriate to measure the disparity between "central" vertices and vertices at the other extreme. To this end we introduce the notion of the *variation* of a connected graph $G$ as $\max(d(p)) - \min(d(p))$ where both extrema are taken over all vertices $p$ of $G$. In our discussion of variation we will have use for the following result.

**Theorem 3.1.** (Oystein Ore [3], p. 103.) *If $p$ is any vertex of a connected graph $G$ then $G$ contains a spanning tree $T$ such that the distance of $p$ in $G$ equals the distance of $p$ in $T$.*

**Theorem 3.2.** *Of all connected graphs on $n$ vertices, trees have the maximum variation.*

Proof. Let $G$ be a connected graph on $n$ vertices with maximum variation. Let $p$ and $q$ be vertices of $G$ with minimum and maximum distance respectively. By the previous theorem of Ore, we can choose a spanning tree $T$ of $G$ in which the distance of $p$ is the same as it was in $G$. Since the distance of $q$ in $T$ cannot be smaller than its distance in $G$, $T$ must have variation as least as large as that of $G$ and the proof is complete.

In our study of the variation of trees we will occasionally digress to generalize various results of Sabidussi [4] and Zelinka [5] since we require certain corollaries of these results. The first instance of this is the following theorem.

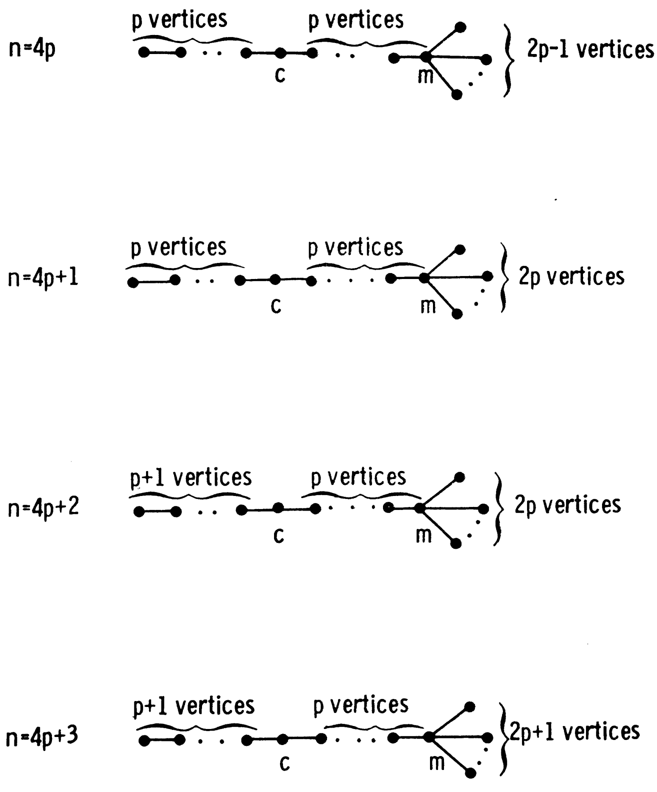


Fig. 3.1.

**Theorem 3.3.** *If $p_0, p_1, \ldots, p_k$ is a path in a tree $T$ and $p_0$ has minimum distance (of all vertices in $T$) then $d(p_1) < d(p_2) < \ldots < d(p_k)$.*

Proof. Suppose the assertion is false for some tree on $n$ vertices and let $i$ be the least positive index for which $d(p_i) \geqq d(p_{i+1})$. Then $d(p_{i-1}) \leqq d(p_i)$ so that, by Property 2.2 there are at least $[(n+1)/2]$ vertices closer to $p_{i-1}$ than to $p_i$ (since no vertex can be equidistant from adjacent vertices of a tree). Call this set of vertices $S$. Then the path joining a member of $S$ to $p_i$ contains $p_{i-1}$. Consequently, since $T$ is a tree, the path joining a vertex in $S$ to $p_{i+1}$ must contain $p_i$ so that there are at least $1 + |S| > \frac{1}{2}n$ vertices closer to $p_i$ than to $p_{i+1}$. By Property 2.2 we then have $d(p_i) < d(p_{i+1})$.

**Corollary 3.4.** (Zelinka [5], p. 90). *A tree has exactly one vertex with minimum distance or exactly two vertices with minimum distance and they are adjacent.*

Proof. If $p$ and $q$ are vertices with minimum distance then, by the theorem, they must be adjacent.

**Corollary 3.5.** (Zelinka [5], p. 94.) *A vertex of maximum distance in a tree $T$ is an end vertex of $T$.*

Proof. An immediate consequence of the theorem.

We note that it is easy to construct counterexamples to the converse of the above corollary.

Another immediate consequence of the theorem is the following result.

**Corollary 3.6.** (Zelinka [5], p. 94.) *An end vertex of a tree $T$ is not a vertex of minimum distance unless $T = K_1$ or $K_2$.*

Sabidussi [4], p. 586, has shown that if $a$ and $b$ are vertices of a connected graph with $n$ vertices then $|d(a) - d(b)| \leqq (n - 2)\, d(a, b)$. Our next theorem strengthens this result considerably.

**Theorem 3.7.** *If $a$ and $b$ are vertices of a connected graph $G$ on $n$ vertices then $|d(a) - d(b)| \leqq [n - 1 - d(a, b)]\, d(a, b)$ and this bound is achieved if and only if all vertices, other than $a$ and $b$, of a shortest path joining $a$ and $b$ have degree 2 and either $a$ or $b$ has degree 1.*

Proof. We first show that the bound cannot be exceeded.

Let $k$ be any fixed integer between 1 and $n$ and from all connected graphs on $n$ vertices having at least one pair of vertices distance $k$ apart, let $G$ be one, containing a pair of vertices $a$ and $b$ distance $k$ apart, for which $d(a) - d(b)$ is a maximum. Let $P = (a = p_0, p_1, \ldots, p_k = b)$ be a shortest path in $G$ between $a$ and $b$.

We wish to show $\deg(p_0) = 1$ and $\deg(p_i) = 2$ for $1 \leqq i \leqq k - 1$. Assume the contrary and let $r$ be the least index so that either $r = 0$ and $\deg(p_r) > 1$ or $1 \leqq r \leqq$ $\leqq k - 1$ and $\deg(p_r) > 2$. Let $c$ be a vertex adjacent to $p_r$ and different from $p_{r+1}$ (and different from $p_{r-1}$ if $r > 0$). If the edge $cp_r$ is a bridge we detach the branch containing $c$ from $G$ at $p_r$ and attach it to $p_{r+1}$ and so create a connected graph on $n$ vertices with $a$ and $b$ still distance $k$ apart but with a larger difference of distances. Consequently $cp_r$ is not a bridge and so $G - cp_r$ is a connected graph on $n$ vertices in which $a$ and $b$ are distance $k$ apart.

So also is the graph $G - cp_r + cp_{r+1}$ (which may be the same as $G - cp_r$). It is not difficult to see that vertex $a$ has strictly greater distance in $G - cp_r + cp_{r+1}$ that it had in $G$ and that vertex $b$ has the same or smaller distance in $G - cp_r + cp_{r+1}$ than it had in $G$. By our choice of $G$ this is impossible so that $\deg(p_0) = 1$ and

$\deg(p_i) = 2$ for $1 \leqq i \leqq k - 1$ and consequently no point of $G$ is equidistant from a pair of adjacent points of $P$.

Now $d(a) - d(b) = \sum_{i=0}^{k-1} [d(p_i) - d(p_{i+1})]$. By Property 2.2, since $p_i$ and $p_{i+1}$ are adjacent, $d(p_i) - d(p_{i+1}) = |P_{i+1}| - |P_i|$ where $P_i(P_{i+1})$ is the number of vertices closer to $p_i(p_{i+1})$ than $p_{i+1}(p_i)$. But $|P_i| = i + 1$ and $|P_{i+1}| = n - i - 1$ so that

$$d(a) - d(b) = \sum_{i=0}^{k-1} (n - 2i - 2) = k(n - k - 1) = [n - 1 - d(a, b)]\, d(a, b)\,.$$

We have shown, incidentally, in our proof that this bound cannot be exceeded, that if the bound is to be achieved then $a$ must have degree 1 and each vertex, other than $a$ and $b$, on a path joining $a$ and $b$ must have degree 2. It remains only to show that this bound can be achieved but this is easily verified for the tree consisting of a path of length $d(a, b)$ together with $n - d(a, b) - 1$ vertices adjacent to $b$.

**Corollary 3.8.** *Of all trees on n vertices, a tree T has maximum variation if and only if it contains a path of length* $[(n - 1)/2]$, *one end vertex of which is an end vertex of T, and all vertices of the path other than its end vertices have degree* 2 *in T.*

Proof. It is a simple matter to verify that $[n - 1 - d(a, b)]\, d(a, b)$ maximizes at $d(a, b) = [(n - 1)/2]$.

**Corollary 3.9.** *The variation of a connected graph on n vertices is at most* $[((n - 1)/2)^2]$ *and this bound is achieved.*

Proof. This follows immediately from Theorem 3.2 and the preceding corollary.

With Corollary 3.4 we have shown that a tree has exactly one vertex of minimum distance or exactly two vertices of minimum distance and they are adjacent.

Now it is well known [3] that a tree $T$ has exactly one center or exactly two centers and they are adjacent, and $T$ has exactly one mass center or exactly two mass centers and they are adjacent. Zelinka has shown [5], p. 91 that the vertices with minimum distance in a tree $T$ are precisely the mass centers of $T$. He later remarked, p. 95, and gave an example showing, that the vertices of minimum distance in a tree need not be its centers. Our final result shows just how far apart these vertices may be.

**Theorem 3.10.** *If T is a tree with* $n + 2$ *vertices. c is a center of T and m is a vertex of T with minimum distance (i.e. a mass center), then* $d(c, m) \leqq [\frac{1}{4}n]$ *and this bound is best.*

Proof. We may assume $T$ has at least three vertices since the assertion is trivial for $n = 1$. By Corollary 3.6 $m$ has degree at least 2 and so is adjacent to some vertex $q$ satisfying $d(c, q) = 1 + d(c, m)$. Since $c$ is a center there is some vertex $b$ in $T$ so that $d(b, c) \geqq d(c, q) - 1$ and the path joining $b$ and $c$ is disjoint from the path joining $c$ and $m$. Consequently $T$ contains a path $P$ joining $b$ and $m$ and passing through $c$ so that $P$ has length $d(b, c) + d(c, m) \geqq 2\, d(c, m)$. But, since $m$ is a mass center $P$ can have at most $\frac{1}{2}n$ edges ([3], p. 66). Consequently $d(c, m) \leqq \frac{1}{4}n$. This upper bound $[\frac{1}{4}n]$ is obviously achieved for $k_1$, and the path on three vertices. Taking $p$ to be any positive integer we see, in the following figure, that the upper bound is achieved for all $n \geqq 4$.

## 4. DISTANCE AND CONNECTIVITY

In this section we obtain a few miscellaneous results relating connectivity to distance. Our first theorem requires the following lemma.

**Lemma 4.1.** *Let $C$ be a smallest cycle containing a given edge $e$ of a graph $G$. If $u$ and $v$ are distinct vertices of $C$, there is a shortest path joining $u$ and $v$ that is a subgraph of $C$.*

Proof. Of all shortest paths joining $u$ and $v$, let $P = (u = p_0, p_1, \ldots, p_n = v)$ be one that uses the maximum number of vertices of $C$.

If $P$ contains only vertices of $C$, then $P$ is a subgraph of $C$ since each edge of $P$ must be an edge of $C$ due to the minimality of $C$.

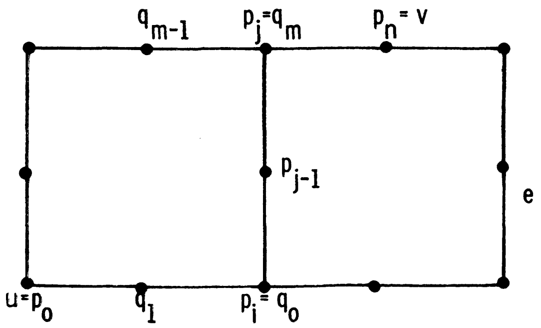


Fig. 4.1.

We may assume, then, that $P$ contains a vertex $p_{i+1}$ not in $C$ and that $p_1, p_2, \ldots, p_i$ are in $C$ and hence are consecutive vertices of $C$. Let $j$ be the least index larger than $i$ for which $p_j$ is also in $C$. Let $Q = (p_i = q_0, q_1, \ldots, q_m = p_j)$ be the path in $C$ between $p_i$ and $p_j$ and not containing the edge $e$ (see Fig. 4.1). Because of the minimality of $C$, $j - i \geqq m$.

Suppose, now, that $Q$ contains no vertex $p_k$ with $k > j$. Then, if $u$ does not lie in $Q$, the path $p_0, p_1, \ldots, p_i, q_1, \ldots, q_m, p_{j+1}, \ldots, p_n$ is a path joining $u$ and $v$, as short as $P$

and using more vertices of $C$ than used by $P$. This contradicts the definition; hence it must be so that $u$ does lie on $Q$, i.e. $u = q_s$ for some $s$, $0 \leqq s \leqq m$. But then $p_0 = q_s, q_{s+1}, \ldots, q_m, p_{j+1}, \ldots, p_n$ is again a path joining $u$ and $v$, as short as $P$, and using more vertices of $C$ than used by $P$. Since this cannot be we may conclude that $Q$ contains some vertex $p_k$ with $k > j$.

Let $r$ be the least index such that $q_r = p_k$ for some $k > j$. If $u$ does not lie on $Q$ then the path $p_0, p_1, \ldots, p_i, q_1, \ldots, q_r, p_{k+1}, \ldots, p_n$ is a path, shorter than $P$, joining $u$ and $v$; if $u$ does lie on $Q$ then $u = q_t$ for some $t$ and the path $p_0 = q_t, q_{t+1}, \ldots, q_r, p_{k+1}, \ldots, p_n$ is such a path (it should be noted here that $r > t$ for otherwise $q_r = p_s$ for some $s$ between 0 and $i$ but this cannot be since a shortest path cannot contain a vertex twice). Since such paths cannot exist we see that $P$ contains only vertices of $C$ and the proof is complete.

We are now able to characterize bridges of a connected graph in terms of distance.

**Theorem 4.2.** *An edge $e$ of a connected graph $G$ is a bridge if and only if each vertex of $G$ has smaller distance in $G$ than in $G - e$.*

Proof. Certainly if $e$ is a bridge of $G$ then each vertex of $G$ has larger (infinite) distance in $G - e$ than in $G$.

To prove the converse we let $e$ be any edge of $G$ that is not a bridge and proceed to show the existence of a vertex $u$ whose distance is the same in $G - e$ as in $G$. To this end choose a smallest cycle $C = (p_0, \ldots, p_n, p_0)$ containing $e$ and label $C$ so that $e = p_0p_n$. Let $u = p_{[n/2]}$ and let $w$ be any vertex of $G$ other than $u$. We will show the distance between $u$ and $w$ is the same in $G - e$ as in $G$. We may assume $w$ is not a vertex of $C$ for this is certainly true otherwise.

Let $R = (u = r_0, r_1, \ldots, r_m = w)$ be a shortest path between $u$ and $w$. We may assume $e = r_ir_{i+1}$ for some $i$ for otherwise we are done. Because of our choice of $u$ and $R$ none of the vertices $r_{i+2}, \ldots, r_m$ can lie in $C$. Since $r_0, r_1, \ldots, r_i$ is a shortest path joining $r_0$ and $r_i$ we may assume, by the previous lemma, that it is a subgraph of $C$. Consequently the path $r_0, \ldots, r_{i+1}$ is identical to either the path $p_{[n/2]}, p_{[n/2]-1}, \ldots, p_0, p_n$ or the path $p_{[n/2]}, p_{[n/2]+1}, \ldots, p_n, p_0$. In the first case, the path $u = p_{[n/2]}, p_{[n/2]+1}, \ldots, p_n, r_{i+2}, r_{i+3}, \ldots, r_m$ is as short as $R$ and does not contain $e$; and in the second, $p_{[n/2]}, p_{[n/2]-1}, \ldots, p_0, r_{i+2}, r_{i+3}, \ldots, r_m$ is as short as $R$ and does not contain $e$. Consequently the distance between $u$ and $w$ is the same in $G - e$ as in $G$ and the proof is complete.

We remark that no analogous result holds for cut vertices. Indeed, it is easy to construct connected graphs having a vertex $v$ that is not a cut vertex yet every vertex of $G$, other than $v$, has greater distance in $G - v$ than in $G$. Examples of such graphs are paths $P_n$ on $n$ vertices, $n \geqq 6$ together with a vertex $p$ adjacent to each vertex of $P_n$. Even though $p$ is not a cut vertex of the graph, its deletion increases the distance of each remaining vertex.

We conclude this section with a sufficient condition, in terms of distance, for a graph to be hamiltonian.

**Theorem 4.3.** *A graph G on n vertices is hamiltonian if* $d(p) \leqq [\frac{1}{2}(3n - 4)]$ *for each vertex p of G.*

Proof. In view of a well result of DIRAC (Ore [3], p. 56) it suffices to show that each vertex of $G$ has degree at least $\frac{1}{2}n$ but this follows immediately for if some vertex $p$ had degree $d_1 < \frac{1}{2}n$ this would imply $d(p) \geqq d_1 + 2(n - 1 - d_1) > > \frac{3}{2}n - 2$.

To see that the bound given cannot be increased consider the graph consisting of $K_{[n/2]}$ and $K_{[(n-1)/2]}$ together with one additional vertex adjacent to all the vertices of both $K_{[n/2]}$ and $K_{[(n-1)/2]}$. It is easy to see that a vertex of $K_{[(n-1)/2]}$ has distance

$$\left[\frac{n-1}{2}\right] + 2\left[\frac{n}{2}\right] \leqq \left[\frac{3n-4}{2}\right] + 1$$

and no vertex has greater distance. This graph is obviously not hamiltonian.

*References*


[1] *Harary, Frank:* Status and contrastatus, Sociometry, *22* (1959) 23—43.

[2] *Harary, F., Norman, R. Z.,* and *Cartwright, D.:* Structural Models: An introduction to the theory of directed graphs. Wiley, New York, 1965.

[3] *Ore, Oystein:* Theory of Graphs, American Mathematical Society. Providence, Rhode Island, 1962.

[4] *Sabidussi, Gert:* The centrality index of a graph, Psychometrika, *31* (1966) 581—603.

[5] *Zelinka, Bohdan:* Medians and peripherians of trees, Arch. Math. (Brno) *4* (1968) 87—95.

[6] *Zelinka, Bohdan:* Generalized centers of gravity of trees, Mathematics (Geometry and Graph Theory) (Czech) 127—136. Univ. Karlova, Prague, 1970.



*Authors' addresses:* R. C. Entringer, University of New Mexico, Albuquerque, N.M. 87131, U.S.A., D. E. Jackson, Los Alamos Scientific Laboratories, Los Alamos, N.M. 87544, U.S.A., D. A. Snyder, Sandia Laboratories, Albuquerque, N.M. 87115, U.S.A.