A PROPERTY OF THE P-MEDIANS OF A GRAPH

by

Chuan Chen and C.K. Butt

Memorandum No. ERL-M355

1 September 1972

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A PROPERTY OF THE P-MEDIANS OF A GRAPH

Chuan Chen[†] and C. K. Butt[*]

ABSTRACT

Given an edge-weighted and node-weighted directed graph G with n nodes, we can define for every $1 \leq m \leq n$ a subset of m nodes called a p-median[1,2] of G. (If we interpret edge-weights as distances and node-weights as masses, the case m=1 corresponds roughly to the "center of mass" in G). We show that every p-median induces a partition of the nodes set into p subsets, exhibiting a certain optimality (theorem 1).

An algorithm for finding all 2-medians of a graph is also presented. As a side benefit, the solution method also yields all 1-medians.

[†]Department of Electrical Engineering and Computer Sciences, and the Electronics Research Laboratory, University of California, Berkeley, California 94720.

Research sponsored by the Joint Services Electronics Program, Contract F44620-71-C-0087 and the National Science Foundation, Grant GK-10656X2.

[*]Department of Business Administration, University of California, Berkeley, California, 94720.

Preliminary Definitions

$\mathbb{R}^+$ is the set of strictly positive real numbers.

$|S|$ is the cardinality of the set S.

An edge-weighted node-weighted directed graph is a 4-tuple $G = (N, \mathcal{A}, d, w)$ where

N is a finite set; $|N| = n$. Elements of N are called nodes.

$\mathcal{A} \subseteq N \times N$ Elements of $\mathcal{A}$ are called arcs.

$d: \mathcal{A} \to \mathbb{R}^+ \cup \{0\}$ If $(x,y) \in \mathcal{A}$, we call $d(x,y)$ the distance from x to y.

$w: N \to \mathbb{R}^+$ If $x \in N$, then $w(x)$ is the weight of node x.

If x and y are nodes, a path from x to y is a set of edges $\{(x,x_1), (x_1,x_2), (x_2,x_3),\ldots,(x_k,y)\}$. The length of a path P is the sum of all edge-weights in P. The shortest distance from x to y, denoted $\bar{d}(x,y)$, is the length of a shortest path from x to y.

Definition The shortest distance from a subset of nodes S to a node x, denoted $\bar{d}(S,x)$, is defined by

$$\bar{d}(S,x) \triangleq \min_{y \in S} \bar{d}(y,x).$$

In the following, p is an integer satisfying $1 \leq p \leq m$.

Definition[1,2] A subset of nodes S is a p-median for G if

(i) $|S| = p$

(ii) $$\sum_{x \in N} w(x)\bar{d}(S,x) \leq \sum_{x \in N} w(x)\bar{d}(T,x) \quad \text{for every } T \subseteq N$$

with $|T| = p$.

Partition induced by a p-median.

Let $S \subseteq N$ be a p-median of G.

Let $\{x_1,x_2,\ldots, x_p\}$ be an enumeration of S.

We can define subsets $X_1,X_2,\ldots, X_p$ of N in the following way.

(i) $x_k \varepsilon X_k$ and $x_k \notin X_j$ if $j \neq k$.

(ii) For each $x \varepsilon N-S$, compute the sequence

$d(x_1,x), d(x_2,x),\ldots., d(x_p,x)$.

The minimum element in this sequence may occur more than once. We let $x \varepsilon X_k$ if $d(x_k,x)$ is the first minimum in the sequence.

It is seen that each $x \varepsilon N$ belongs to precisely one X_k. Hence we have that

$$\bigcup_{k=1}^{p} X_k = N$$

and $$X_i \cap X_j = \phi \quad \text{if } i \neq j.$$

Definition The set $\{X_k\}_{k=1}^{p}$ constructed by (i) and (ii) above is called the partition of N induced by S (with respect to the given enumeration of S).

Definition A p-partition of N is a partition of N into p subsets.

Definition If $\{Y_i\}_{i=1}^{p}$ is a p-partition of N, we can find a 1-median for each of the p subgraphs created by the partition. Let y_i be a 1-median of the subgraph corresponding to Y_i. The value of the partition $\{Y_i\}_{i=1}^{p}$, denoted $V(\{Y_i\})$, is defined by

$$V(\{Y_i\}) \triangleq \sum_{i=1}^{p} \sum_{y \varepsilon Y_i} w(y)\bar{d}(y_i,y).$$

The following result shows why p-medians can be useful in various applications.

<u>Theorem 1</u> Let $\{X_i\}_{i=1}^{p}$ be a p-partition of N induced by a p-median. Let $\{Y_i\}_{i=1}^{p}$ be any other p-partition of N.

Then

$$V(\{X_i\}) \leq V(\{Y_i\}).$$

<u>Proof</u> Let $\{X_i\}$be induced by the p-median $S = \{x_1, x_2, \ldots, x_p\}$.

Let $T = \{y_1, y_2, \ldots, y_p\}$ be the set of 1-medians of the subgraphs corresponding to $Y_1, Y_2, \ldots, Y_p$ respectively.

We have

$$V(\{Y_i\}) \triangleq \sum_{i=1}^{p} \sum_{y \varepsilon Y_i} w(y)\bar{d}(y_i, y)$$

Now $\bar{d}(y_i, y) \geq \min_{y_j \varepsilon T} \bar{d}(y_j, y) \triangleq \bar{d}(T, y)$.

Hence
$$V(\{Y_i\}) \geq \sum_{y \varepsilon N} w(y)\bar{d}(T, y) \tag{1}$$

But
$$\sum_{y \varepsilon N} w(y)\bar{d}(T, y) \geq \sum_{x \varepsilon N} w(x)\bar{d}(S, x) \text{ since S is a p-median.} \tag{2}$$

Also
$$\sum_{x \varepsilon N} w(x)\bar{d}(S, x) \triangleq \sum_{x \varepsilon N} w(x) \min_{y \varepsilon S} \bar{d}(y, x)$$

$$= \sum_{i=1}^{p} \sum_{x \varepsilon X_i} w(x)\bar{d}(x_i, x) \tag{3}$$

since $\{X_i\}$ is induced by S.

Let x_i^* be the 1-median of X_i for each i.

By definition of 1-median we must have for each i that

$$\sum_{x\varepsilon X_i} w(x)\bar{d}(x_i,x) \geq \sum_{x\varepsilon X_i} w(x)\bar{d}(x_i^*,x) \tag{4}$$

Thus, summing (4) over all i, we have

$$\sum_{i=1}^{p} \sum_{x\varepsilon X_i} w(x)\bar{d}(x_i,x) \geq \sum_{i=1}^{p} \sum_{x\varepsilon X_i} w(x)d(x_i^*,x) \tag{5}$$

But by definition of $V(\{X_i\})$ we have

$$\sum_{i=1}^{p} \sum_{x\varepsilon X_i} w(x)\bar{d}(x_i^*,x) \triangleq V(\{X_i\})$$

Putting (1),(2),(3),(5),(6) together, we get the desired inequality. ⌑

An algorithm to find all the 2-medians of G.

Notice that if {u,v} is a 2-median for G, then condition (ii) in the definition of p-median above reduces to

$$\sum_{x\varepsilon N}^{p} w(x) \min \{\bar{d}(u,x),\bar{d}(v,x)\} \leq \sum_{x\varepsilon N} w(x) \min \{\bar{d}(u',x),\bar{d}(v',x)\} \tag{*}$$

for every $u',v'\varepsilon N$ with $u'\neq v'$

Thus, if we compute the right-hand side of (*) for every pair (u',v'), the pair that yields a smallest value will be a 2-median. The method we present below is such an exhaustive scheme, but is computationally efficient and conceptually simple.

Given two nxn matrices A,B with real entries, let us define a binary matrix operation * by

$$C=A*B \Longleftrightarrow C \text{ is } n x n \text{ and } c_{ij} = \sum_k \min\{a_{ik}, b_{kj}\}$$

Given the graph G, order the node set in some fashion, say

$$N = \{x_1, x_2, \ldots\ldots, x_n\}.$$

A very efficient algorithm[3,4] can be used to compute the matrix of shortest distances $\bar{D}$ whose entries are

$$\bar{d}_{ij} = \bar{d}(x_i, x_j)$$

Defining $w_i \triangleq w(x_i)$, we can form the matrix

$$W \triangleq \begin{bmatrix} w_1 & & & & \\ & w_2 & & \bigcirc & \\ & & w_3 & & \\ & \bigcirc & & \ddots & \\ & & & & w_n \end{bmatrix}$$

The computation of the right-hand side of (*) for all pairs (u',v') then reduces to the computation of the matrix

$$(\bar{D}W)*(\bar{D}W)^T \qquad (\dagger)$$

Since (†) is symmetric, every minimum element a_{ij} above the diagonal in (†) corresponds to a 2-median $\{x_i, x_j\}$. Further, all 2-medians can be found this way.

Remark 1 Since every undirected graph G can be considered as a directed graph with a symmetric distance matrix, the above method works for undirected graphs as well.

Remark 2 Putting u=v and u'=v' in (*), we have a necessary and sufficient condition for {u} to be a 1-median. Thus each 1-median corresponds to a diagonal in (†) which is minium among all diagonal elements.

Example Let G be the undirected graph below (figure 1). The number inside each node is its weight; the number beside each arc is its length.

$$\bar{D} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 3 & 4 & 8 \\ 1 & 0 & 2 & 4 & 3 & 7 \\ 1 & 2 & 0 & 3 & 4 & 8 \\ 3 & 4 & 3 & 0 & 4 & 6 \\ 4 & 3 & 4 & 4 & 0 & 4 \\ 8 & 7 & 8 & 6 & 4 & 0 \end{bmatrix} \end{array}$$

$$W = \begin{bmatrix} 1 & & & & & \\ & 2 & & & \bigcirc & \\ & & 3 & & & \\ & \bigcirc & & 4 & & \\ & & & & 5 & \\ & & & & & 6 \end{bmatrix}$$

$$\bar{D}W = \begin{bmatrix} 0 & 2 & 3 & 12 & 20 & 48 \\ 1 & 0 & 6 & 16 & 15 & 42 \\ 1 & 4 & 0 & 12 & 20 & 48 \\ 3 & 8 & 9 & 0 & 20 & 36 \\ 4 & 6 & 12 & 16 & 0 & 23 \\ 8 & 14 & 24 & 24 & 20 & 0 \end{bmatrix}$$

$$(\bar{D}W)*(\bar{D}W)^T = \begin{bmatrix} 85 & 72 & 82 & 61 & 41 & \textcircled{37} \\ & 80 & 70 & 58 & 47 & 38 \\ & & 85 & 61 & 41 & \textcircled{37} \\ & & & 76 & 42 & 40 \\ & \text{symmetric} & & & \textcircled{62} & 38 \\ & & & & & 90 \end{bmatrix}$$

The minimum above-diagonal element is 37, corresponding to the two 2-medians

$$\{x_1,x_6\} \text{ and } \{x_3,x_6\}$$

The minimum entry on the diagonal is 62, corresponding to the single 1-median $\{x_5\}$.

References

1. S. L. Hakimi, "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," Operations Research, 12, 450, (1964).
2. S. L. Hakimi, "Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph-Theoretic Problems," Operations Research, 13, 462, (1965).
3. R. W. Floyd, "Algorithm 97: Shortest Path," Comm. ACM, 5, 345, (1962).
4. S. Warshall, "A Theorem on Boolean Matrices," JACM, 9, 11-12, (1962).
5. P. Järvinen, J. Rajala, and H. Sinervo, "A Branch-and-Bound Algorithm for Seeking the p-median," Operations Research, (Jan. 1972).

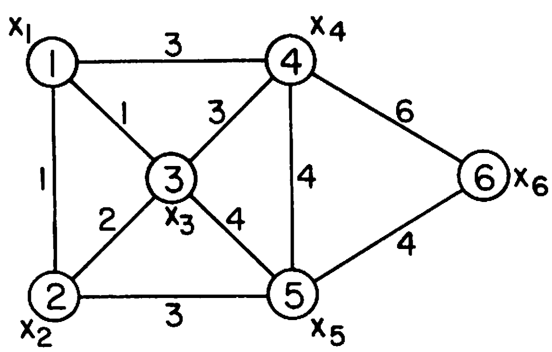

Figure 1: A node-weighted edge-weighted undirected graph.