Heuristic Methods for Estimating the Generalized Vertex Median of a Weighted Graph
Author(s): Michael B. Teitz and Polly Bart
Reviewed work(s):
Source: *Operations Research*, Vol. 16, No. 5 (Sep. - Oct., 1968), pp. 955-961
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/168488

HEURISTIC METHODS FOR ESTIMATING THE GENERALIZED VERTEX MEDIAN OF A WEIGHTED GRAPH†

Michael B. Teitz and Polly Bart

University of California, Berkeley, California

(Received July 31, 1967)

The generalized vertex median of a weighted graph may be found by complete enumeration or by some heuristic method. This paper investigates alternatives and proposes a method that seems to perform well in comparison with others found in the literature.

THE PROBLEM of supplying some number of destinations, n, from a number of sources, p, has been attacked with a variety of assumptions and methods. If both sources and destinations are at fixed locations with given quantities available and required, the standard transportation problem of linear programming appears. With given locations for destinations, the determination of locations of sources in a euclidean plane may be called the generalized Weber problem, after the nineteenth century student of industrial location who examined this problem for the special case of the single location. Operational generalizations of this form of the problem have been investigated by COOPER[1, 2, 3] using iterative approximation methods that are appropriate where the locations for sources may be treated as continuous in the plane. If destinations consist of fixed nodes on a network, but sources may lie anywhere on network links and destination demands are fixed while source capacities are unconstrained, it has been shown by HAKIMI[4, 5] that the problem resolves itself into finding the generalized absolute median of the weighted shortest path graph corresponding to the network. Hakimi also demonstrated that there will exist such a generalized p-median that includes only vertices of the graph, that is, nodes on the network. Thus a solution to this problem will correspond to the case where both destinations and sources lie on nodes of a network, a situation like that investigated for fixed destination demands and unconstrained source capacity by MARANZANA.[6]

In this paper we deal with the problem of choice of location of p sources of unconstrained capacity from among n destinations having fixed demands and located at nodes of a network. The problem is essentially the same as

† The support of The California State Department of Public Health is gratefully acknowledged. We are indebted to Lee DeCola for aid in calculations.

that investigated by Hakimi and Maranzana. However, our concern is with alternative methods of solution.

THE DISCRETE ALLOCATION PROBLEM AS THE GENERALIZED MEDIAN OF A GRAPH

CONSIDER A network of n nodes, for each of which a demand h_i is given. For each pair of nodes i, j on the network, the minimum path distance d_{ij} is found. The network may now be conveniently represented by an edge and vertex weighted graph G. Each vertex v_i in G is weighted by the corresponding h_i, and each i-j edge by the shortest path d_{ij}.

The vertex median of the graph G is defined in the following way. Let the distance matrix D of G be the $n\times n$ symmetric matrix of d_{ij} between all pairs of vertices v_i, v_j. For a graph with equal vertex weights, the vertex median v_k will be that vertex for which the sum of the elements in the corresponding column of D is minimized. That is, let

$$d_j=\sum_{i=1}^{i=n} d_{ij}, \qquad (j=1, 2, \cdots, n) \quad (1)$$

then v_k is the vertex median if and only if

$$d_k=\min(d_1, d_2, \cdots, d_n). \quad (2)$$

If the graph G has unequal vertex weights, then it is necessary to redefine the distance matrix. Let H be the nth order diagonal matrix with vertex weights on the diagonal. Now the weighted distance matrix R of G is defined by

$$R=HD=[r_{ij}]=[h_i d_{ij}]. \quad (3)$$

We have now multiplied D by the corresponding row vertex weights. R is no longer generally symmetric. Each r_{ij} represents the weighted distance quantity associated with vertex v_i if v_j were the unique source. The vertex median is defined as above in (1) and (2) but over the matrix R.

A generalization of the vertex median follows logically from (1) and (2). Let V_p be some subset containing exactly p vertices of G. For an n vertex graph there will be $\binom{n}{p}$ possible subsets of this size; arbitrarily index them by $V_p^m\left[m=1, 2, \cdots, \binom{n}{p}\right]$. For each such subset we may construct a submatrix R_p^m of R by adjoining all columns of R for which the corresponding column vertices are contained in V_p^m. R_p^m is of order $n\times p$ and describes succinctly the sources and associated weighted distances for every destination if the set of sources is limited to vertices in V_p^m. If we assume that sources have no capacity constraints, each destination v_i will be served by that source v_k in V_p^m for which r_{ik} is a minimum. That is,

$$r_{ik} \leqq r_{ij} \qquad \text{for all } v_k,\ v_j \epsilon V_p{}^m. \tag{4}$$

The total movement (weighted distance) r_m for the $V_p{}^m$ set of sources will be the sum of the row minima of $R_p{}^m$:

$$r_m = \sum_{i=1}^{i=n} r_{ik}, \tag{5}$$

where k refers to the source for which r_{ij} is minimized. A generalized vertex p-median of G is now defined as some V_p^{m*} such that

$$r_{m*} = \min\left(r_1, r_2, \cdots, r_{\binom{n}{p}}\right), \tag{6}$$

so that

$$r_{m*} \leqq r_m. \qquad \left[m = 1, 2, \cdots, \binom{n}{p}\right]$$

The p-median is not necessarily unique. Clearly, it defines a set of sources that is in some sense 'closest' to all vertex destinations in the discrete locational system. The next task is to find it.

FINDING THE GENERALIZED MEDIAN

So FAR no completely satisfactory method for finding the generalized median of a graph has been put forward. In the following sections we describe two approaches that have been suggested and advance a third that seems to be a significant improvement although still subject to question.

THE DIRECT ENUMERATION METHOD

FOR SMALL systems the p-median may be found by direct enumeration and evaluation of all possibilities. For every possible subset $V_p{}^m$ the value of r_m is calculated and the minimum r_{m*} found by inspection. This procedure was used by Hakimi to find the 3-median of a graph with ten vertices. Direct calculation seems to be reasonable so long as n and either p or $n-p$ are not large, since the number of possible subsets $V_p{}^m$ equals $\binom{n}{p}$. For systems with many vertices the computational time demands grow very rapidly and shortcut methods become necessary.

THE PARTITION METHOD

A METHOD put forward by MARANZANA[6] parallels in several respects the iterative procedure used for the continuous location case by COOPER.[3] At first sight it appears attractive, but its performance may be quite erratic as we show below. In essence the method approaches the p-median by finding successive single vertex medians of p subsets of destination vertices each associated with one source, and then adjusting the subsets before repeating the process.

Select some initial subset V_1 of p vertices from the vertex set V of the graph G. Consider these as initial source locations. For each vertex v_j in V_1 define an associated set of destination vertices P_{1j} such that

$$P_{1j}=\{v_i|r_{ij}<r_{ik}\forall v_k\epsilon V_1\}. \tag{7}$$

The collection

$$P_1=\{P_{1j}|v_j\epsilon V_1\}$$

of vertex subsets P_{1j} constitutes a partition of V. Compute the total weighted distance for the system on the assumption that V_1 is the source set. Call this first weighted distance total

$$r_1=\sum\nolimits_{v_j\epsilon V_1}\sum\nolimits_{v_i\epsilon P_{1j}} r_{ij}. \tag{8}$$

Now for each P_{1j} find the single vertex median of its associated subgraph of G. In general this vertex median may be any vertex in P_{1j} including v_j. Let the set of vertices so found constitute a new source subset V_2 of V. Again define for each vertex v_j in V_2 its associated P_{2j} destination set and compute r_2. It is easily shown that $r_2\leqq r_1$. Labelling the vertex median of each destination set P_{1j} as its source must by definition either reduce or leave unchanged the weighted distance sum for that subset. Shifting any vertex into a new subset during the definition of the P_{2j} subsets must by (7) reduce the contribution to the total weighted distance made by that destination vertex. Thus neither step can increase the total weighted distance value.

This process continues until at some iteration t, every v_j is also the vertex median of its associated P_{tj}, and for all v_j the corresponding P_{tj} equals $P_{t-1,j}$. No further reduction in r_t is now possible. The system has achieved a form of equilibrium and the set V_t is an estimate of the generalized p-median of G. There is no assurance that the true p-median and global minimum r have been found. Maranzana claims that with several initial choices for V_1 a solution close to the optimum is likely. For the special example that he used this seems possible. For other cases it may not be so.

THE VERTEX SUBSTITUTION METHOD

THE PARTITION method lends itself to visualization as a series of geometric operations. A set of points is chosen as the initial source configuration. Destinations are assigned to 'market sets' similar to conventional market areas. The optimum source location within each market set is found. The market sets are redefined with reference to the new source locations, and the process iterated until no improvement can be achieved. An alternative approach concentrates upon the formal definition of the generalized vertex median and its associated weighted distance matrix. In some respects it is a discrete version of one of the methods in Cooper.[2]

Consider the definition of the generalized vertex median given by (5) and (6). For each possible subset of source vertices V_p^m we may construct a submatrix R_p^m of R by adjoining the p columns corresponding to the vertices v_j in V_p^m. The source from which any destination vertex v_i is served is defined as that v_k such that

$$r_{ik} \leqq r_{ij} \qquad \text{for all } v_k,\, v_j \epsilon V_p^m.$$

This is the ith row minimum in R_p^m. The total weighted distance r_m for the mth source subset will be the sum of these row minima.

Suppose that we decide to replace one vertex, v_j, in the source subset by another, v_b. Several kinds of resultant effect on total weighted distance may occur. If r_{ij} were not the ith row minimum of R_p^m, then no change in the ith row contribution to r would result. If r_{ij} were the ith row minimum, then its replacement by r_{ib} might have several outcomes depending upon whether:

$$r_{ib} \leqq r_{ij}, \tag{9a}$$

or

$$r_{ij} \leqq r_{ib} \leqq r_{is}, \tag{9b}$$

or

$$r_{ij} \leqq r_{is} \leqq r_{ib}, \tag{9c}$$

where r_{is} is the weight-distance quantity from v_i to that vertex v_s for which

$$r_{ij} \leqq r_{is} \leqq r_{is^*} \qquad \text{for } v_{s^*} \epsilon V_p^m, \quad \text{and } v_{s^*} \neq v_j,\, v_s. \tag{10}$$

In other words, r_{is} is the second smallest ith row element in R_p^m. In case (9a), the ith row contribution to r from the substitution of v_b for v_j is now incremented by

$$_i\Delta_{bj} = r_{ij} - r_{ib}, \quad \text{and} \quad _i\Delta_{bj} \leqq 0. \tag{11}$$

In case (9b), the ith row contribution to r is incremented by

$$_i\Delta_{bj} = r_{ij} - r_{ib}, \quad \text{and} \quad _i\Delta_{bj} \geqq 0. \tag{12}$$

In case (9c), the comparable increment is

$$_i\Delta_{bj} = r_{ij} - r_{is}, \quad \text{and} \quad _i\Delta_{bj} \geqq 0. \tag{13}$$

Whether it is worth substituting vertex v_b for v_j depends upon the net effect of the increments summed over all rows:

$$\Delta_{bj} = \sum_{i=1}^{i=n} {}_i\Delta_{bj}. \tag{14}$$

Substituting v_b for v_j reduces total weighted distance only if

$$\triangle_{bj} < 0. \tag{15}$$

These observations suggest an approach to finding the generalized median that involves an iterative process of single vertex substitution in the following steps.

(i) Select some initial source vertex subset V_1; for convenience in exposition let it contain vertices $v_1, v_2, \cdots, v_j, \cdots, v_p$.

(ii) For each vertex v_j find its associated P_{1j} destination subset as in (7), and compute the total weighted distance for the system, r_1, as in (8).

(iii) Select some vertex, v_b, not in the source subset.

(iv) For each vertex v_j in V_1 substitute v_b and compute $\triangle_{bj}$.

(v) Find that vertex v_k in V_1 such that

$$\begin{aligned} &\triangle_{bk} < 0, \\ \text{and} \qquad &\triangle_{bk} = \min \triangle_{bj}, \qquad (j = 1, 2, \cdots, p). \end{aligned} \tag{16}$$

(vi) If a vertex v_k meeting the conditions (16) can be found, substitute v_b for v_k in the source subset, label the new subset so formed V_2, and compute r_2. If no vertex meets the conditions (16), retain the source subset V_1 and continue to step (vii).

(vii) Select another vertex not contained in V_1 or V_2 and not previously tried, and repeat steps (iv) through (vi).

(viii) When all vertices in the complement of V_1 have been tried, define the resulting source subset V_t as a new V_1 and repeat steps (ii) through (vii). Call each such complete repetition a cycle.

(ix) When one complete cycle of steps (ii) through (viii) results in no reduction in r, terminate the procedure. The final V_t is an estimate of the vertex p-median of G.

At first sight this procedure appears laborious in comparison with the partition method. However, the component steps are simple and the computational time reasonable. In no case during our trials have more than four cycles been required to reach a stable estimate of the p-median.

As with the partition method, no assurance can be given that the final V_t and r_t will be global optimum values. It is possible to construct cases in which single vertex substitution yields no improvement in r while pairwise or higher substitutions will still reduce it. At present we are working on tests for the existence of such cases. The problems involved are difficult owing to the arbitrary nature of the R matrix.

TESTS OF THE METHODS

In the absence of a rigorous test for the optimum, we are forced back on experimental trials. Testing alternative methods presents some difficulty.

Statistically, we prefer a large number of tests on different matrices. At the same time, we would like to know whether either method is more sensitive to the choice of initial source set, and also whether variation for problems of a given size measured in terms of total weighted distance is significant. This would involve many trials on a given matrix or on matrices having identical weighted distance sums for the median set. Taken altogether, these involve extensive computations, and given our inability to construct examples with known weighted distance sums, they cannot be realized.

In this paper, we present results of tests of a simpler kind. Matrices of random numbers representing systems of twenty-five destination vertices were generated. For each matrix, the first ten vertices were taken as an initial source set. Estimates of the generalized ten-median were then made by both the partition and vertex substitution methods. Since the optimum in each case was unknown, the lower of the two associated weighted distance sums was chosen as the 'true' estimate and the percentage 'error' of the other method calculated from it.

Two runs of fifty random matrices each were made. In no case did the weighted distance sum of the vertex substitution method exceed that of the partition method. The mean per cent errors of the partition method were 103 and 105 per cent. The associated standard deviations of the per cent errors were 63 and 82. Thus, the partition method exhibited considerable variation in performance.

There is a good deal to question in this method of evaluation, but given the problem alternatives are hard to find. COOPER[3] and others have been forced to use similar approaches for the continuous case. Nevertheless, at this sample size the results seem to be unambiguous enough to suggest that vertex substitution may be a preferable heuristic. If the partition method is used, the high variance of its error suggests that great caution in selection of initial locations is necessary.

REFERENCES

1. L. COOPER, "Location-Allocation Problems," *Opns. Res.* **11,** 331–343 (1963).
2. ———, "Heuristic Methods for Location-Allocation Problems," *SIAM Review* **6,** 37–53 (1964).
3. ———, "Solutions of Generalized Locational Equilibrium Models," *J. Regional Sci.* **7,** 1–18 (1967).
4. S. L. HAKIMI, "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," *Opns. Res.* **12,** 450–459 (1964).
5. ———, "Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems," *Opns. Res.* **13,** 426–475 (1965).
6. F. E. MARANZANA, "On the Location of Supply Points to Minimize Transport Costs," *Opnal. Res. Quart.* **15,** 261–270 (1964)