Operations Research Letters 14 (1993) 79–84
North-Holland

September 1993

Complexity results for the p-median problem with mutual communication

Arie Tamir

Department of Statistics and Operations Research, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel

Received January 1993
Revised May 1993

The p-median problem with mutual communication is defined as follows: Let $G = (V, E)$ be an undirected graph with positive edge lengths. Suppose that each node represents an existing facility. The objective is to locate p new facilities (medians) while minimizing the sum of weighted distances between all pairs of new facilities and pairs of new and existing facilities. We present new complexity results for this model and for some of its variants.

facility location; p-median; tree networks

1. Introduction

The p-median problem with mutual communication is defined as follows:

Let $G = (V, E)$ be an undirected graph with a node set $V = \{v_1, \dots, v_n\}$ and an edge set E. Suppose that each edge in E has a positive weight (length). For each pair of nodes v_i and v_j let $d(v_i, v_j)$ denote the length of a shortest path connecting v_i and v_j. Given nonnegative reals a_{ij}, $i = 1, \dots, n$; $j = 1, \dots, p$, and b_{jk}, $j = 1, \dots, p$; $k = 1, \dots, p$, locate p points, *medians*, $x_1, \dots, x_p$, at the nodes of V to minimize the expression

$$\sum_{\substack{i=1,\dots,n\\ j=1,\dots,p}} a_{ij} d(v_i, x_j) + \tfrac{1}{2} \sum_{\substack{j=1,\dots,p\\ k=1,\dots,p}} b_{jk} d(x_j, x_k). \tag{1}$$

(We assume that $b_{jk} = b_{kj}$, $j = 1, \dots, p$; $k = 1, \dots, p$.) To justify the term 'median' consider the case when G is a simple path and there is no mutual communication, i.e., $b_{jk} = 0$, for all pairs j, k. In this case the nodes are viewed as points on the real line, and (1) decomposes into p 1-median problems. The optimal solution is to let x_j, $j = 1, \dots, p$, be the weighted median of the points $v_1, \dots, v_n$, using the weights $a_{1j}, \dots, a_{nj}$, respectively.

Problem (1) has been discussed in several papers dealing with location models [1–4,6,8,11,13,14]. A generalization of it, known as the *Module Allocation* problem, arises in the allocation of tasks to processors in a distributed system to minimize total communication and execution costs [7].

When the graph G is a single edge, i.e., $n = 2$, (1) reduces to the classical minimum cut problem on a complete graph with $p + 2$ nodes. The two nodes of G, v_1 and v_2 correspond to the source and the sink respectively and the remaining p nodes represent the points $x_1, \dots, x_p$. The real numbers a_{ij}, $i = 1, 2$;

Correspondence to: A. Tamir, Department of Statistics and Operations Research, Raymond and Beverly Sackler, Faculty of Exact Sciences, Tel Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel.

$j = 1, \dots, p$, and b_{jk}, $j = 1, \dots, p$; $k = 1, \dots, p$, are the respective edge capacities. Thus, when $n = 2$ and p is variable (1) is solvable in $O(p^3)$ time [12]. It is shown in [11] that (1) is strongly NP-hard when both n and p are variable. (It is certainly solvable in polynomial time when p is fixed.) We will use the results in [5] to prove that (1) is strongly NP-hard even when $n = 3$ and p is variable. [2–4,7] have discussed some polynomially solvable cases of (1), based on the graph induced by the support of the $\{b_{jk}\}$ numbers.

Extending the case $n = 2$, [1,13,14] have shown that if G is a simple path, (1) decomposes into $n - 1$ minimum cut problems, one for each edge of the path. This observation has led to an $O(np^3)$ time algorithm to solve (1). Exploiting the relationships between these $n - 1$ cut problems, [8] have applied a parametric approach to the minimum cut problems and solved (1) in $O(p^3 + p^2 \log n + pn)$ time. The same complexity bound is also reported in [9]. [11,13] have shown that the decomposition into $n - 1$ minimum cut problems applies also to the case where G is a tree, hence yielding an $O(np^3)$ bound for (1) on tree graphs. However, the parametric approach of [8,9] does not seem to be extendible to the tree case.

In this paper we apply a centroid decomposition to the given tree, and improve the $O(np^3)$ complexity bound to $O(p^3 \log n + pn + n \log n)$. We will then focus on the case when the points $x_1, \dots, x_p$, are restricted to a prespecified proper subset of nodes S. We will refer to this model as the *restricted* case. In the unrestricted case $S = V$. Unlike the polynomial time results stated above we demonstrate that the restricted variant of (1) becomes strongly NP-hard even when the tree has 4 nodes. However, when G is a path we will show that even the restricted model can be solved in polynomial time. We note that this is one of the few examples of location problems that we know of where the problem on a path is solvable in polynomial time while the tree case is strongly NP-hard.

2. NP-completeness results

It is shown in [5] that the following 3 Multiway Cut problem is NP-hard: Given a graph $G = (V, E)$, a set of 3 specified nodes x, y and z in V find a minimum cardinality subset of edges E', such that the removal of E' from E disconnects each of the above three nodes from the other two. (See Theorem 3 in [5].)

We show that the p-median problem with mutual communication, defined on a triangle (a complete graph with three nodes) is a generalization of the 3 Multiway Cut problem.

Consider the 3 Multiway Cut problem. Suppose that G has $p + 3$ nodes and $V = \{v_1, v_2, v_3, x_1, \dots, x_p\}$. The 3 specified nodes are v_1, v_2 and v_3. Every 3 multiway cut corresponds to a solution to the p-median problem with mutual communication on a triangle defined by the following sets of reals $a_{ij} = 1$ if and only if nodes v_i and x_j are connected by an edge in G, and $a_{ij} = 0$ otherwise. Similarly, $b_{jk} = 1$ if and only if nodes x_j and x_k are connected by an edge in G, and $b_{jk} = 0$ otherwise.

The 3 nodes of the triangle, v_1', v_2' and v_3' correspond respectively to the specified nodes of G, v_1, v_2 and v_3. The length of each edge of the triangle is taken to be 1. It is now a simple matter to verify that each solution to the p-median problem with mutual communication on the triangle corresponds to a 3 multiway cut in the above $p + 3$ node graph G. An optimal solution to the p-median problem with mutual communication is an optimal solution to the 3 Multiway Cut problem.

The above reduction of the Multiway Cut problem can also be used to show that the p-median model, which is polynomially solvable on trees, is NP-hard even on a tree with 4 nodes if we restrict the sites of $x_1, \dots, x_p$ to a proper subset of nodes. For the reduction use a star tree with 3 leaves (tips) and restrict $x_1, \dots, x_p$ to these leaves. Assuming that the 3 edges of this star are of equal length, we note that the star is equivalent to the above triangle. Nevertheless, in Section 4 we show that unlike the tree case the restricted model on a path is polynomially solvable. We note in passing that using the above reduction it follows from [5] that the restricted model on a k leaf star tree with equal edge lengths has a polynomial time $2 - 2/k$ approximation algorithm. For example, there is a polynomial time algorithm which yields a $\frac{4}{3}$ approximation to the star tree with 3 leaves.

3. Solving the unrestricted tree case

In this section we discuss the case where G is a tree and the points $x_1, \ldots, x_p$ can be located anywhere in V, and improve the $O(np^3)$ bound in [11,13]. To facilitate the discussion we introduce the following notation. Let (v_i, v_j) be an edge. By removing this edge the node set V is partitioned into two subsets, say, $V(i, j)$ and $V(j, i)$, where v_i is in $V(i, j)$ and v_j is in $V(j, i)$. Following [13] define the $[i, j]$-local problem by

$$\text{Minimize} \sum_{\substack{v_t \in V(i,j) \\ m=1,\ldots,p}} a_{tm} d(v_i, x_m) + \sum_{\substack{v_t \in V(j,i) \\ m=1,\ldots,p}} a_{tm} d(v_j, x_m) + \tfrac{1}{2} \sum_{k,m=1,\ldots,p} b_{km} d(x_k, x_m), \tag{2}$$

where $x_1, \ldots, x_p$ are in $\{v_i, v_j\}$. Problem (2) is the reduction of the original problem obtained by identifying all the nodes in $V(i, j)$ with the single node v_i, and the nodes in $V(j, i)$ with the single node v_j.

As noted above the $[i, j]$-local problem is equivalent to the classical minimum cut problem. The following result is proven in [13].

Theorem 3.1. *Let G be a tree. Suppose that $x_1, \ldots, x_p$ is an optimal solution to the $[i, j]$-local problem* (2). *Then there exists an optimal solution to problem* (1) *$x_1^*, \ldots, x_p^*$, such that for $k = 1, \ldots, p$, x_k^* is in $V(i, j)$ if $x_k = v_i$, and x_k^* is in $V(j, i)$ if $x_k = v_j$.*

Since the solution to the $[i, j]$-local problem is independent of the length of the edge (v_i, v_j) it follows, as shown in [13], that the solution to problem (1) on a tree is independent of the edge lengths. It depends only on the topology of the tree. Furthermore, since the objective value varies continuously with the edge lengths, it follows that the optimality of a solution is not affected when we contract edges. We use this latter property to justify our centroid decomposition approach.

In the preprocessing we compute for each $k = 1, \ldots, p$, and edge (v_i, v_j) in E the quantities

$$A[i, j\colon k] = \sum_{v_t \in V(i,j)} a_{tk} \quad \text{and} \quad A[j, i\colon k] = \sum_{v_t \in V(j,i)} a_{tk}.$$

To compute the above we root the tree at some node, say v_1, and then use a standard 'bottom-up' approach. With this approach we first compute $A[i, j\colon k]$ for the edges (v_i, v_j) where v_i is a leaf of the rooted tree, and then recursively proceed towards the root of the tree. The total preprocessing time is easily observed to be $O(np)$.

We start by finding a centroid of the given tree, i.e., a node whose removal splits the tree into connected components none of which contains more than $n/2$ nodes. A centroid can be found in $O(n)$ time [10]. If v_i is a centroid then V can be split into two subsets U and W such that their intersection is v_i and $|U|, |W| \le \frac{2}{3}n + \frac{1}{3}$. Consider the tree obtained by connecting the subtree induced by U with the subtree induced by W by a new edge. This step is called splitting the node v_i according to U and W. Arbitrarily label one copy of v_i as v_i and the second v_i^*. To define problem (1) for the new tree we assign zero a coefficients to the node v_i^*.

From the above it follows that solving problem (1) on the new tree and contracting the edge connecting the subtrees induced by U and W, will yield the optimal solution to the original tree. We first solve the local problem corresponding to the augmented edge in $O(p^3)$ time using any of the classical algorithms for the minimum cut problem [12]. Using Theorem 3.1 we are then left with two problems defined on the subtrees induced by U and W. Having computed the terms $A[i, j\colon k]$ for $i, j = 1, \ldots, n$, and $k = 1, \ldots, p$, it is easy to see that the marginal effort needed to define these two subproblems is only $O(p^2)$.

Thus the recursive equation for the effort $t(n, p)$ to solve problem (1) is

$$t(n, p) \le c(p^3 + n) + t(n_1, p_1) + t(n_2, p_2)$$

where $n_1 + n_2 = n + 1$, $n_i \le \frac{2}{3}n + \frac{1}{3}$, $i = 1, 2$, and $p_1 + p_2 = p$ (c is some constant independent of p and n.)

The solution to the above is $t(n, p) = O((p^3 + n) \log n)$. Adding the preprocessing time we get $O((p^3 + n) \log n + np)$ as the complexity bound for solving the (unrestricted) p-median problem with mutual communication on a tree graph.

4. Locating in a proper subset of nodes

From the results in Section 2 it follows that if we allow the new points to be established only at a proper subset of the nodes of the tree, then the problem becomes strongly NP-hard. However, the case of a path is still solvable in polynomial time. We will show that in this case the solution depends on the edge distances, but still constitutes a parametric minimum cut problem satisfying the conditions in [8]. We assume that the nodes in V are points on the real line with $v_1 < \cdots < v_n$. Let S denote the subset of V where the new medians may be set up. For each k, $k = 1, \dots, p$, let

$$F_k(x_k) = \sum_{i=1,\dots,n} a_{ik} |v_i - x_k|.$$

Since x_k is restricted to the nodes in S we define and consider the function $f_k(x_k)$, the restriction of $F_k(x_k)$ to S. For each x_k in S let $f_k(x_k) = F_k(x_k)$. If x_k is not in S, $f_k(x_k)$ is defined as the respective linear interpolation of the values at the two adjacent points in S which bound x_k. The function $f_k(x_k)$ is convex, piecewise linear, coincides with $F_k(x_k)$ on S, and has its breakpoints at S only. Thus the objective function of the restricted model is to minimize

$$\sum_{k=1,\dots,p} f_k(x_k) + \tfrac{1}{2} \sum_{j,k=1,\dots,p} b_{jk} |x_j - x_k|, \tag{3}$$

where x_k is in S for $k = 1, \dots, p$.

The above is a convex minimization problem in the variables x_k. If we remove the requirement that x_k must be in S, and view $x_1, \dots, x_p$ as real variables, then there is still an optimal solution such that $x_1, \dots, x_p$ are all in S. In particular, Problem (3) can be solved in polynomial time. Furthermore, we show that (3) can be solved by applying the parametric approach, which is used in [8] to solve the unrestricted case.

Without loss of generality suppose that we relabel the nodes in S and assume that $S = \{v_1, v_2, \dots, v_m\}$ with $v_i < v_{i+1}$. S is also viewed as the node set of a path. Consider the edge (v_i, v_{i+1}) connecting v_i and v_{i+1}. Define the $[i, i+1]$ local problem corresponding to (3) by

$$\begin{aligned} &\text{Minimize} && \sum_{k=1,\dots,p} f_k(x_k) + \tfrac{1}{2} \sum_{j,k=1,\dots,p} b_{jk} |x_j - x_k| \\ &\text{Subject to} && x_1, \dots, x_p \in \{v_i, v_{i+1}\}. \end{aligned} \tag{4}$$

The next theorem follows from the same arguments used in [13] to prove Theorem 3.1 above. Therefore we skip its proof.

Theorem 4.1. *Suppose that $x_1, \dots, x_p$ is an optimal solution to the $[i, i+1]$ local problem (4). Then there exists an optimal solution to (3), $x_1^*, \dots, x_p^*$, such that for $k = 1, \dots, p$, x_k^* is in $V(i, i+1)$ if $x_k = v_i$ and x_k^* is in $V(i+1, i)$ if $x_k = v_{i+1}$.*

We note that the $[i, i+1]$ local problem can be formulated as a minimum cut problem on the following auxiliary graph G' with $p+2$ nodes. The source node is v_i and the sink is v_{i+1}. Then, for $k = 1, \dots, p$, there is a node corresponding to the point x_k. In this auxiliary graph connect the pair of nodes corresponding to x_j and x_k with an edge of capacity $b_{jk} |v_i - v_{i+1}|$. Connect a node corresponding to x_j by an edge to the source node, v_i, with capacity $f_j(v_{i+1})$, and by an edge to the sink node, v_{i+1}, with capacity $f_j(v_i)$.

We can further subtract the capacity $f_j(v_i)$ from both edges to obtain zero capacity for the edge connecting the node corresponding to x_k with the sink. Next divide all edge capacities by $|v_i - v_{i+1}|$.

Thus, the only edges in G' whose capacities depend on the distance $|v_i - v_{i+1}|$ are the edges that are incident to the source. Using the convexity of the f_j functions we now observe that the $[i, i+1]$ local problem and the $[i+1, i+2]$ local problem differ only in the capacities of the edges connecting the source to the nodes corresponding to the x_k variables. The capacity at the $[i, i+1]$ local problem is not larger than the respective capacity at the $[i+1, i+2]$ local problem. In particular, the monotonicity property in [8] is indeed satisfied, and we can apply the procedure in [8] to solve parametrically the sequence of the $m-1$ $[i, i+1]$ local problems (4), $i = 1,\ldots,m-1$. The total complexity will be $O(p^3 + p^2 \log m + np)$.

It is easily seen from the above that when S is a proper subset of V, the solution does depend on the distances $\{|v_i - v_{i+1}|\}$. Using the above framework for the unrestricted case $S = V$, we compute and note that

$$[f_j(v_{i+1}) - f_j(v_i)]/|v_{i+1} - v_i| = \sum_{t=1,\ldots,i} a_{tj} - \sum_{t=i+1,\ldots,m} a_{tj}.$$

In particular, the edge capacities of the $[i, i+1]$ local problem are independent of the distance between v_i and v_{i+1}. This is already shown in [13].

In our p-median model the functions f_j are explicitly defined above. However, we note that the minimum cut approach applies to any set of functions that are piecewise linear and convex with breakpoints at the nodes in S.

5. Extensions and concluding remarks

The results in [13] for the unrestricted case on a tree can be further extended.

Let $T = (V, E)$ be a given tree. Consider the following extension of the unrestricted p-median problem with mutual communication. Locate p points, $x_1,\ldots,x_p$, at V to minimize the expression

$$\sum_{j=1,\ldots,p} G_j(x_j) + \tfrac{1}{2} \sum_{j,k=1,\ldots,p} b_{jk} d(x_j, x_k), \tag{5}$$

where G_j, $j = 1,\ldots,p$, is a piecewise linear and convex function on T, with breakpoints at V. (Convexity on T is defined as convexity on each path of T [6].)

The results in [13] hold for this extension but the solution depends on the edge lengths. In view of the NP-completeness result of Section 2 it is essential that medians can be established anywhere in V. Note that if a certain node is not a potential site the value of the G functions will be infinite, hence the convexity is destroyed. In the case of a path this problem is resolved by eliminating that node. This elimination cannot be done on a node with degree greater than 2.

A natural extension of the location model on the line discussed in Section 4 is to allow setting up a median at each node of the path, but associate a setup cost which might depend on both the node and the median. (See [3,7] for a definition of a general model with setup costs.) The restricted model of Section 4 is a special case where the setup costs are 0 for each node in S and infinity otherwise. The complexity of the extended model is an interesting and important open problem.

References

[1] T.Y. Cheung, "Multifacility location problem with rectilinear distance by the minimum-cut approach", *ACM Trans. Math. Software* **6**, 549–561 (1980).

[2] D. Chhajed and T.J. Lowe, "Location facilities which interact: some solvable cases", *Ann. Oper. Res.* **40**, 101–124 (1992).

[3] D. Chhajed and T.J. Lowe, "m-Median and m-center problems with mutual communication: solvable cases", *Oper. Res.* **40** S56–S66 (1992).

[4] D. Chhajed and T.J. Lowe, "Solving structured multifacility location problems efficiently", forthcoming in *Transportation Science*.

[5] E. Dalhaus, D.S. Johnson, C.H. Papadimitriou, P.D. Seymour and M. Yanakakis, "The complexity of multiway cuts", *Proc. of the 24th annual ACM Symp. on Theory of Computing*, 1992, 241–251.

[6] P.M. Dearing, R.L. Francis and T.J. Lowe, "Convex location problems on tree networks", *Oper. Res.* **24**, 628–642 (1976).

[7] D. Fernandez-Baca, "Allocating modules to processors in a distributed system", *IEEE Trans. Software Engnr.* **15**, 1427–1436 (1989).

[8] D. Gusfield and C. Martel, "A fast algorithm for the generalized parametric minimum cut problem and applications", *Algorithmica* **7**, 499–519 (1992).

[9] D. Gusfield and E. Tardos, "A faster parametric minimum cut algorithm", Tech. Report #926, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, 1990.

[10] R. Hassin and A. Tamir, "Efficient algorithms for optimization and selection on series-parallel graphs", *SIAM J. Algebraic and Discrete Methods* **7**, 379–389 (1986).

[11] A.J.W. Kolen, "Location problems on trees and in the rectilinear plane", Stichting Mathematics Centrum, Amsterdam, The Netherlands, 1982.

[12] E.L. Lawler, *Combinatorial Optimization; Networks and Matroids*, Holt, Rinehart and Winston, New York, 1976.

[13] J.C. Picard and H.D. Ratliff, "A cut approach to the rectilinear distance facility location problem", *Oper. Res.* **28**, 422–433 (1978).

[14] V.A. Trubin, "Effective algorithm for the Weber problem with a rectangular metric", *Cybernetics* **14**, 874–878 (1978).