Seoul Journal of Business
Volume 6, Number 1/2 (December 2000)

K-Median Problem on Graph

Sang-Hyung Ahn
Seoul National University

In past decades there has been a tremendous growth in the literature on location problems. However, among the myriad of formulations provided, the simple plant location problem and the k-median problem have played a central role. This phenomenon is due to the fact that both problems have a wide range of real-world applications, and a mathematical formulation of these problems as an integer program has proven very fruitful in the derivation of solution methods.

In this paper we investigate the k-median problem defined on a graph. That is, each point represents a vertex of a graph.

1. Introduction

In past few decades, there has been a tremendous growth in the literature on location problems. However, among the myriad of formulations provided, the simple plant location problem and the k-median problem have played a central role. This phenomenon is due to the fact that both problems have a wide range of real-world applications, and a mathematical formulation of these problems as an integer program has proven very fruitful in the derivation of solution methods.

Consider an index set $I = \{1, 2, \ldots, n\}$ of n points, and a positive integer $k \leq n$, and let C_{ij} be the shortest distance between two points $i, j \in I$. The k-median problem consists of identifying a subset $S \subseteq I$, $|S| = k$ so as to minimize $\Sigma_{i \in I} \text{Min}_{j \in S} C_{ij}$ (Here $|S|$ denotes the cardinality of the set S).

We introduce integer variables. Let $Y_j = 1$, if point j is selected as a median, otherwise 0 and X_{ij} = 1; if point j is the closest median to point i, otherwise 0. With X, Y variables, the k-median problem is formulated as an integer program as follows.

Integer Program Formulation:

$$Z_{IP} = Min\sum_{i\in I}\sum_{j\in I} C_{ij}X_{ij} \quad (1)$$

$$\sum_{j\in I} X_{ij} = 1 \qquad i \in I \quad (2)$$

$$\sum_{j\in I} Y_j = k \quad (3)$$

$$0 \le X_{ij}, Y_j \le 1 \qquad i, j \in I \quad (4)$$

$$X_{ij}, Y_j \quad \text{integral} \quad i, j \in I \quad (5)$$

A vast number of algorithms were proposed and probabilistic analyses were presented for the k-median problem. We refer readers to Ahn et al. [1], Beasley [2], Boffey [3], Christofides [5], Christofides and Beasley[4], Cornuejols [6] [7] [8], Even[9], Fisher and Hochbaum [10], Francis and White [11], Handler and Mirchandani [12], Jacobsen and Pruzan [13], Krarup and Pruzan [15], ReVelle [17], Rosing [18].

In this paper we investigate the k-median problem defined on a graph. That is, each point represents a vertex of a graph. Unless otherwise specified, it is assumed that C_{ii} = 0, C_{ij} = C_{ji} (symmetry of distance) and $C_{ij} \le C_{il} + C_{lj}$ (triangular inequality).

Kolen [14] proved that the linear programming relaxation of the simple plant location problem defined on graphs has an integer optimal solution when the underlying graph is a tree. However, this does not hold for the k-median problem. We state this observation as a proposition below.

Proposition 1: When the underlying graph is a tree, the linear programming relaxation of the k-median problem on a graph can have a fractional optimal solution.

Proof:

By an example in Figure 1.

Numbers on the edges in the following graph are the length of edges.

For the following tree with k = 2,

Z_{IP} = 5 for with an optimal solution of Y_3 = Y_4 = 1, Y_j = 0 for j = 1, 2, 5 and X_{ij} is defined to satisfy (2) - (4).

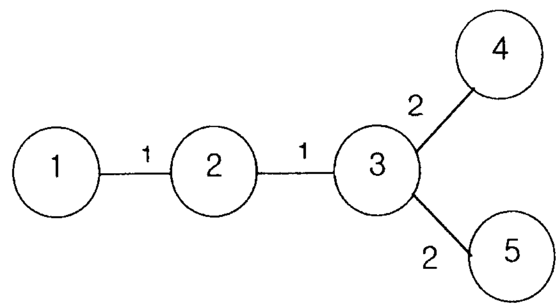

Figure 1. Tree of Duality Gap

Z_{LP} = 4.5 with a unique optimal solution of $Y_1 = 0$, $Y_j = 1/2$ for j = 2, 3, 4, 5 and $X_{12} = X_{13} = X_{22} = X_{23} = X_{32} = X_{33} = X_{43} = X_{44} = X_{53} = X_{55} = 1/2$, all other $X_{ij} = 0$.//

2. A tree model

Since the linear programming relaxation of the k-median problem on a tree can have a fractional optimal solution, here we further investigate a tree in which the optimal linear program solution is always fractional.

We introduce a notion of a dominating set.

Definition 1: A subset $D \subseteq I$, $|D| = k$ is a dominating set if for every node that does not belong to D, there exists at least one edge which connects it to any node in D. If the length of each edge, $C_{ij} \geq 1$ for all $i \neq j$, then we must have

$$Z_{IP} \geq Z_{LP} \geq n - k \quad (6)$$

Lemma 2:

If there exists a dominating set in a graph, then $Z_{IP} = Z_{LP} = n - k$

Proof:

If a dominating set exists in a graph, $Z_{IP} = |n| - k$. Hence Lemma 2 follows (6). //

We derive the dual of the linear programming relaxation of k-median problem. Let V_i, U, W_{ij}, t_j be the dual variables associated with the following LP relaxation constraints set (7)-(11) respectively.

$$\sum_{j \in I} X_{ij} = 1 \qquad i \in I \quad (7)$$

$$\sum_{j \in I} Y_j = k \tag{8}$$
$$X_{ij} \le Y_j \qquad i, j \in I \tag{9}$$
$$Y_j \le 1 \qquad j \in I \tag{10}$$
$$X_{ij},\ Y_j \ge 0 \tag{11}$$

The dual formulation is:

$$Z_{LP} = \text{Max} \sum_{i \in I} V_i - k * U - \sum_{j \in I} t_j \tag{12}$$
$$V_i - W_{ij} \le C_{ij} \qquad i, j \in I \tag{13}$$
$$\sum_{i \in I} W_{ij} - Ut_j \le 0 \qquad j \in I \tag{14}$$
$$W_{ij},\ t_j \ge 0 \qquad i, j \in I$$
$$V_i \text{ and } U: \text{ unrestricted}$$

We present a tree where linear programming relaxation always has fractional optimal solution. Consider following a graph where p is the number of spokes and each spoke consists of two nodes except node 0.

Theorem 3

For $2 \le k \le p$, the optimal solution to the above tree is,
$Y_0 = (p - k)/(p - 1)$, $Y_{j1} = (k - 1)/(p - 1)$, $Y_{j2} = 0$ for each spoke,
$Z_{LP}(k) = (3p^2 - 2pk - p + k - 1)/(p - 1)$.

Proof:

Let V_i, W_{ij}, U, t_j be dual variables and we construct a dual feasible solution as follows.

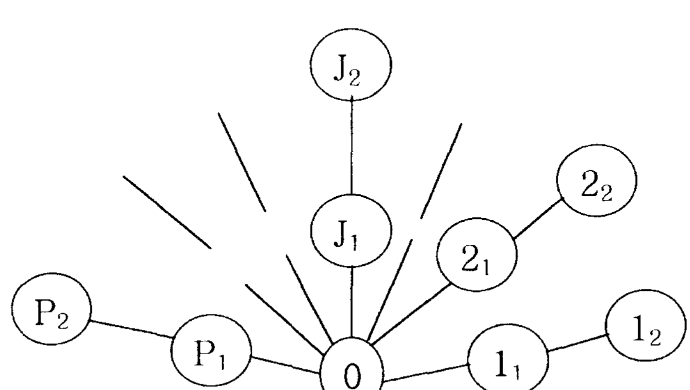

Figure 2. The Tree with unit edge cost

$V_0 = 1$, $V_{j1} = 1$, $V_{j2} = 2 + 1/(p - 1)$, $t_{i1} = t_{i2} = 0$ for each spoke, $U = 2+1/(p - 1)$.

$W_{00} = 1$, $W_{0j1} = W_{0j2} = 0$, $W_{j10} = 0$, $W_{j1j1} = 1$, $W_{j1j2} = 0$, and
$W_{j20} = 1/(p - 1)$, $W_{j2j1} = 1+1/(p - 1)$, $W_{j2j2} = 2+1/(p - 1)$

The value of the above solution, which is dual feasible, is:
$Z_{LP}(D) = \Sigma_{i\in I}V_i - kU = (3p^2 - 2pk - p+ k- 1)/(P - 1)$, which is Z_{LP}.

By strong duality theorem, both primal and dual solutions are optimal.//

Proposition 4

For $2 \leq k < p$, an optimal integer solution is $Y_0 = 1$, $Y_{j1} = 1$ for any k-1 spokes.

Proof:

The value of above solution $Z_{LP} = (k - 1)+3(p - k +1) = 3p - 2k + 2$, and
$Z_{IP} - Z_{LP} = (k - 1)/(p -1)<1$.//

Proposition 4 implies that even though a duality gap, $Z_{IP} - Z_{LP}$, always exists for the tree given in Figure 2, the duality gap is less than 1 and goes to 1 when p goes to infinity for $k = p - 1$. One interesting feature of the above tree is that for $k = p$, there is no duality gap.

Proposition 5

For $k = p$, duality vanishes for the above tree. That is, $Z_{IP} = Z_{LP}$

Proof:

Let J^* be a set of j_1 of each spoke. Then J^* is a dominating set, so $Z_{IP} = Z_{LP} = p + 1$ with $Y_{j1} - 1 = 1$ for each spoke. //

Since dual feasible region is independent of the value of k, we have the following results.

Theorem 6

Let $S^* = \{U^*, V^*, W^*\}$ be an optimal LP solution of $2 \leq k = k^* - p$. Then S^* is also an optimal LP solution of $2 - k = k^* + a - p$. and $Z_{LP}(k^* + a) = Z_{LP}(k^*) - aU^*$.

Proof:

Since dual feasible region does not depend on the value of k, S^* is a feasible LP solution to $k = k^* + a$. The value of this solution S^* to $k = k^* + a$ is $\{3p^2 - 2p(k^* + a) - p + (k^* + E) - 1\}/(p - 1) = Z_{LP}(k) - aU^*$, which is optimal value according to theorem 3. //

Consider a random tree T_n with node set $I = \{1, 2, ..., n\}$ where each of the n_{n-2} different trees is equally likely to occur. The distance d_{ij} is the number of edges in the unique path from i to j in T_n. Then we have random trees on n nodes, the number of values of k such that $z_{IP} \neq z_{LP}$ is almost surely at least cn, for some constant $c > 0$.

Theorem 7.

(a) For $k = 1$ or $k \geq [(n-1)/2]$, $Z_{IP} = Z_{LP}$ for every tree on n nodes.

(b) For $2 \leq k < [(n-1)/2]$, and $n \neq 8$, there is a tree on n nodes such that $Z_{IP} \neq Z_{LP}$.

Proof:

For the 1-median problem, it is well known that $Z_{IP} = Z_{LP}$ for every choice of d_{ij}, $1 \leq i, j \leq n$. For example, this result appears in Mukendi [16].

When $k \geq [n/2]$, $z_{IP} = Z_{LP} = n - k$ follows from the fact that every tree on n nodes has a dominating set of cardinality at most $[n/2]$.

To complete the proof of Theorem 7(a), it suffices to consider the case where n is even and $k = n/2 - 1$. By induction, one can show that the only trees which do not have a dominating set of size k are constructed inductively from a path with 4 nodes by adding paths $P_i = (v_1^i, v_2^i, v_3^i)$ where v_1^i is one of the non-leaf nodes of the current tree and v_2^i, v_3^i are two new nodes. (See Figure 3-a) From the construction $Z_{IP} = n - k + 1 = n/2 + 2$. Using the dual values $u_j = 2$ if X_j is a leaf, 1 if not, $Z_{LP} = n/2 + 2$. Therefore $Z_{IP} = Z_{LP}$.

To prove Theorem 7(b) when n is odd, consider the tree of Figure 3-b. Let $p = (n - 1)/2$. An optimal solution of the k-median problem is to take $S = \{1, 2, 4, 6, ..., 2(k - 1)\}$. Then $Z_{IP} = 3p - 2(k - 1)$. We get a feasible solution of the LP relaxation by setting $x_1 = (p - k)/(p - 1)$ and $x_{2i} = (k - 1)/(p - 1)$ for $i = 1, ..., p$. This yields

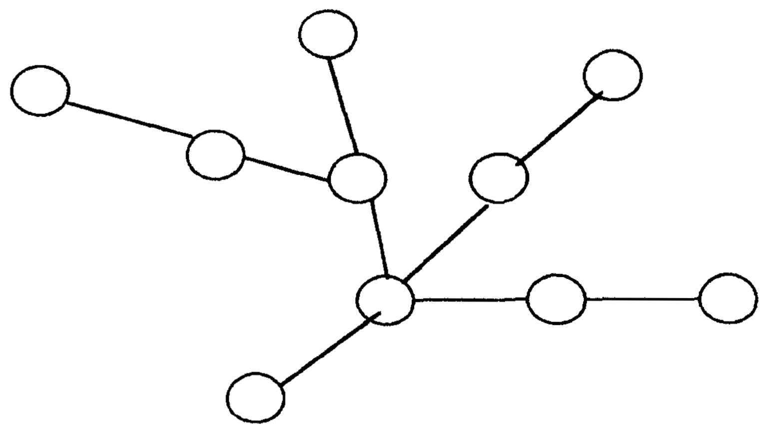

Figure 3-a.

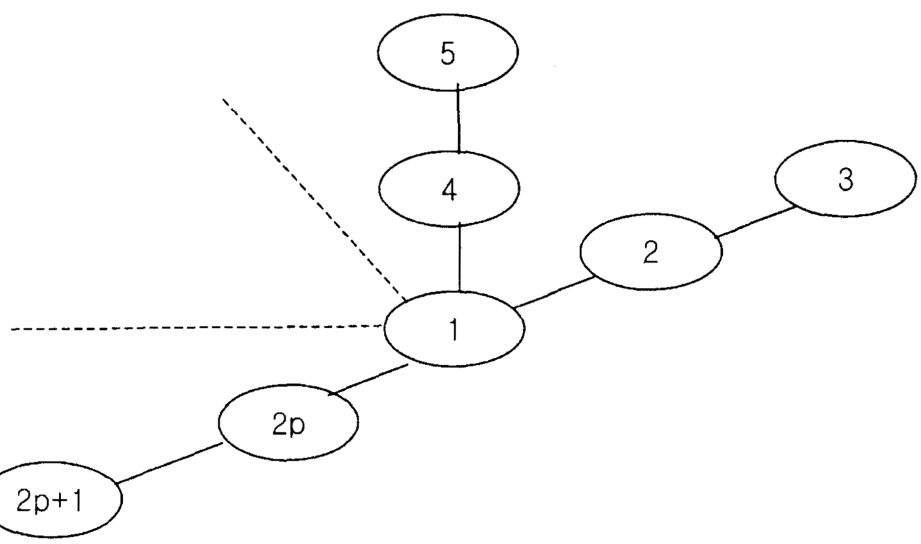

Figure 3-b.

$Z_{LP} \leq (3p^2 - 2pk - p + k - 1)/(p - 1)$. Therefore $Z_{IP} - Z_{LP} \geq (k - 1)/(p - 1) > 0$

To prove Theorem 7(b) when n is even, $n \neq 8$, we first consider the case $k \geq 3$. Add a node p_2+1 adjacent to p_2 to the tree of Figure 3-b. Then it is optimum to choose p_1 in S, and we can also choose $p_1 = 1$ in the *LP* solution. Removing p_1, p_2 and p_2+1, we are back to the case where n is odd and $k \geq 2$. Now consider the case $n \geq 10$ even and $k = 2$. Add three nodes to the graph of Figure 3-b, namely i_1+1 adjacent to i_1 for $i = 1, 2, 3$. Then $Z_{IP} = 3p + 3$, but there is a better *LP* solution, namely $y_1 = 1$ and $y_2 = y_4 = y_6 = 1/3$. This yields $Z_{LP} = 3p + 1$.//

3. Conclusion

In this paper we investigated the k-median problem defined on graphs whose linear programming relaxation can have a fractional optimal solution. We further presented the k-median problem on graphs whose linear programming relaxation always has fractional optimal solution even though the underlying graph is a tree.

We conclude with following observation. The linear programming relaxation of the k-median problem defined on graphs can have fractional optimal solution even when the underlying graph is a perfect graph.

Proposition 8:

When the underlying graph is a tree, line graphs, or claw-free and triangulated graphs (perfect graph), the linear programming relaxation of the k-median problem can have fractional optimal solution.

Proof:

Consider the following graph. The length of three edges connecting nodes 1_1, 2_1, 3_1 is 4, and the length of other edges is 1 where length of each edge is 1. The unique optimal linear and integer solution for k = 2 is the same as that of Figure 2 with p=2.//

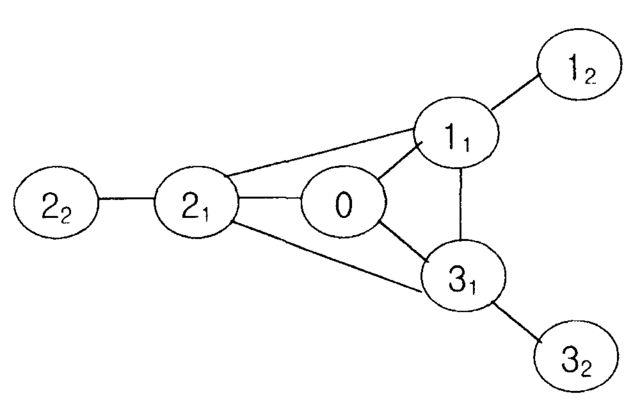

Figure 4. Graph of Duality Gap

References

1. Ahn, S. H. and Cornuejols, G. "Probabilistic Analysis of relaxation of p-median problem." Math. Of Operations Research No.1, 1988.
2. Beasley, J. E. "A note on solving large p-median problems." technical report SW7 2BX, Imperial College, September, 1984.
3. Boffey, T.B. and Karkazis, J. "p-median and multi-medians." J. Oper. Res. Society 35 (1984), 57-64.
4. Christofides, N. "Graph Theory: An Algorithmic Approach." Academic Press, New York, 1975.
5. Christof ides, N. and Beasley, J. E. "A tree search algorithm for the p-median problem," European J. of Oper. Res. 10 (1982), 196-204.
6. Cornuejols, O., Fisher, M.L. and Nemhauser, G. L. "Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms, Management Sci." Management Sci. 23 (1977), 789-810.
7. Cornuejols, G., Fisher. M. L. and Nemhauser, G. L. "On the uncapacitated location problem." Ann. of Discete Math. 1 (1977), 163-177.
8. Cornuejols, G., Nemhauser, G. L. and Wolsey, L. A. "Worst Case and probabilistic analysis of algorithms for a location problem." Operations Res. 28 (1980), 847-858.
9. Even, Shimon. Graph Algorithms. Computer Science Press, Rockville, Maryland, 1979.
10. Fisher, M.L. and Hochbaum, D. S. "Probabilistic analysis of the planar k-median problem." Math. of Operations Research 5, 1 (Feb. 1980), 27-34.
11. Francis, R. L. and White, J. A. "Facility Layout and Location: An Analytic Approach". Prentice-Hall, Englewood Cliffs, NJ, 1974.
12. Handler, G. Y. and Mirchandani, P. B. "Location on Networks: Theory and Algorithms". MIT-press, Cambridge, MA., 1979.
13. Jacobsen, S. K. and Pruzan, P. M. "Lokalisering-modeller & Losningsmetoder". Studentilitteratur, Lund, Sweden, 1978.
14. Kolen, A. "Location problems on trees and in the rectilinear plane." Mathematisch Centrum, Amsterdam, Netherland, 1982.
15. Krarup, J. and Pruzan, P. M. "The simple plant location problem: Survey and synthesis". European Jour. of O. R. 12 (1983), 36-81.
16. Mukendi, C. "Sur L'implantation d'equipement dans un reseau:le probleme de m-center", Thesis, University of Grenoble, France (1975)
17. ReVelle, C. S., Swain, R. S. Central facilities location._ Geographical

Anal. 2 (1970), 30-42.
18. Rosing, K. E., ReVelle, C. S. and Rosing-Vogelaar. The p-edian and its linear programming relaxation: an approach to large problems. J. Operational Res. Soc. 30(1979), 815-823.