Journal of the Korean
OR/ MS Society
Vol. 20, No. 1, April 1995

Structural Study of the K-Median Problem*

Sang Hyung Ahn**

Abstract

The past three decades have witnessed a tremendous growth in the literature on location problem. A mathematical formulation of uncapacitated plant location problem and the k-median as an integer program has proven very fruitful in the derivation of solution methods. Most of the successful algorithms for the problem are based on so-called "strong" linear programming relaxation. This is due to the fact that the strong linear programming relaxation provides a tight lower bound. In this paper we investigate the phenomenon with a structural analysis of the problem.

1. Introduction

The past threee decades have witnessed a tremendous growth in the literature on location problems. However, among the myriads of formulations the uncapacitated plant location problem and the k-median problem have a wide range of real-world applications. A mathematical formulation of these problems as an integer program has proven very fruitful in the derivation of solution methods.

Consider an index set $I=\{1, 2, \cdots, n\}$ of n points, and integer $k \leq n$, and let c_{ij} be the shortest distance between two points $i, j \in I$.

We introduce integer variables. Let $y_j=1$ if a point j is selected as a median, otherwise 0 and $x_{ij}=1$ if a 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.

* 이 논문은 1992년도 교육부지원 한국학술 진흥재단의 자유공모과제 학술연구조성비에 의해 연구되었음
** 서울대학교 경영대학

Integer Program Formulation :

$$Z_{IP}=Min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \quad (1)$$

subject to

$$\sum_{j=1}^{n} x_{ij}=1 \quad i, j\in I \quad (2)$$

$$\sum_{j=1}^{n} y_j=k \quad (3)$$

$$x_{ij}\leq y_j \quad i, j\in I \quad (4)$$

$$y_j\leq 1 \quad j\in I \quad (5)$$

$$x_{ij}, y_j\geq 0 \quad i, j\in I \quad (6)$$

$$x_{ij}, y_j, \text{ integral } i, j\in I \quad (7)$$

When we drop integer constranint set (7), the integer program becomes a linear program.

Linerar Program Formulation

$$Z_{LP}=Min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

subject to

(2), (3), (4), (5), (6)

A vast number of algorithms were proposed for the k-median probelm. We refer readers to Ahn et al [1], Beasley [2], Boffey [3], Christofieds [4], Beasel and Christofides [5], Cornuejols [6] [7] [8], Fisher and Hochbaum [9], Francis and White [10], Handler and Mirchandani [11], Jacobsen and Pruzen [12], Kolen[13], Krarup and Pruzan [14], Papadimitriou [15], ReVelle [16], Rosing [17].

Most of the successful algorithms for the k-median problem are based on the strong linear programming relaxtion. In Ahn et al [16] we presented and explained why the strong linear programming relaxation provides a tight lower bound in the probabilistic sense. In this paper we investigate the phenomenon with a structural study of the problem.

2. Structural Analysis

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

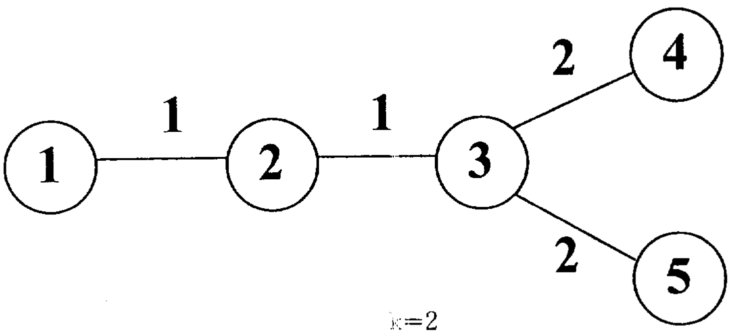
Kolen[13] proved that the linear programming relaxation of the uncapacitated plant location problem defined on a graph has an integer optimal solution when the underlining graph is a tree. However this does not hold for the k-median problem. We state this observation as a proposition below.

Proposition 1 :

Even 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.//



k=2

figures : lengths of edges

⟨Figure 1⟩ A Tree of Duality Gap

For te above thee $Z_{IP}=5$ with an optimal solution of $y_3=y_4=1$, $y_j=0$ for j=1, 2, 5. x_{ij} is defined to satisfy (2), (4), (6).

$Z_{LP}=4.5$ with an 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 $x_{ij}=0$.

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 soluion is always fractional. Let G(V,E) be graph such that length of every edge is unit and $|V|=n$. We shall assume thast $c_{ij}=1$ for all pairs of i, j $\in$ V which are joined by an edge hereafter.

We introduce a notion of a dominating set.

Definition 1 :

A subset D of V is a dominating set if for every node that does not belongs 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

$$Z_{IP} \geq Z_{LP} \geq n-k \tag{8}$$

In fact, when there exists a dominating set in a graph, Z_{IP} achieves its lower bound and equals Z_{LP}. We present this as a lemma below.

Lemma 2 :

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

Proof :

Suppose there exists a dominating set $D \subseteq V$ in the graph.

Let $y_j=1$, $x_{ij}=1$, for all $j \in D$ and $y_j=0$, $x_{ij}=0$ for all other j.

The value of this integer solution is $n-k$, which is its low bound. Hence Lemma 2 follows (8).//

We derive the dual of the linear programming relaxation of the k-median problem Let. V_i, U, W_{ij}, t_j be the dual variables associated with the constraints set (2), (3), (4), (5) respectively.

The dual formulation is :

$$Z_{LP}(D)=Max \sum_{i=1}^{n} V_i-(k)(U)-\sum_{j=1}^{n} t_j \tag{9}$$

subject to

$$V_i - W_{ij} \leq C_{ij} \qquad i,\ j \in I \tag{10}$$

$$\sum_{i=1}^{n} W_{ij} - U - t_{ij} \leq 0 \quad j \in I \tag{11}$$

$$W_{ij},\ t_j \geq 0 \quad i,\ j \in I \tag{12}$$

$$V_i,\ U\ ;\ unrestricted\ i \in I \tag{13}$$

For any given $V=(V_i : i=1,\cdots,n)$, define

$\rho_i(V)=\sum_{i=1}^{n} (V_i-c_{ij})^+$ for $j=1,\cdots,n$, where a^+ denotes $Max(0, a)$.

Lemma 3 :

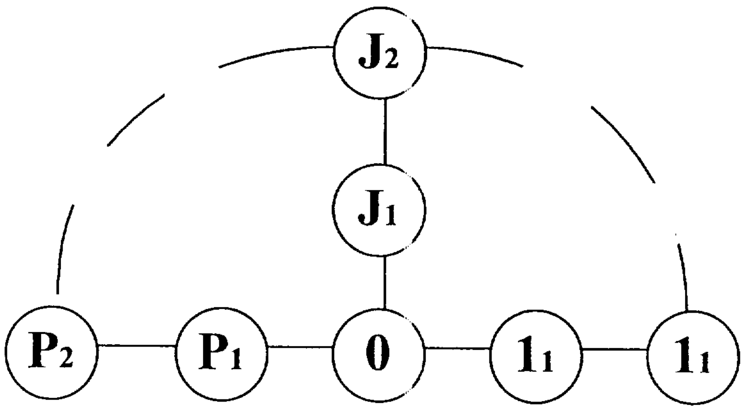
$$Z_{LP} \geq \sum_{i=1}^{n} V_i - k \times Max_{j=1,\cdots,n}\ \rho_i(V)$$

Proof :

If can be checked that, a feasible soluion of $Z_{LP}(D)$ is obtained by setting $\mathbf{W_{ij}=(V_i-c_{ij})^+}$, $\mathbf{t_j}=0$ and $U=Max_{j=1,\cdots,n}\rho_i(V)$. //

We let $Z_D(V)=\sum_{i=1}^{n} V_i - k \times Max_{j=1,\cdots,n}\ \rho_i(V)$ be this dual bound.

We present a tree where the linear progrmming relaxation always has a fractional optimal solution at ⟨Figure 2⟩.



p : # of spokes
each spoke consists of
1 non-leaf and 1 leaf node
besides center node

⟨Figure 2⟩ A Tree with a Fractional Optimal LP Solution : 1

Theorem 4 :

For $2 \leq k \leq p$ the optimal LP solution to the above three is : $y_0=(p-k)/(p-1)$, $y_1=(k-1)/(p-1)$, $y_{12}=0$ for j = 1,⋯,p.

$Z_{LP}(k)=(3p^2-2pk-p+k-1)/(p-1)$

Proof :

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

$V_0=1$, $V_{i_1}=1$, $V_{i_2}=2+1/(p-1)$, $t_{i_1}=t_{i_2}=0$, for $i=1,\cdots,p$

$W_{00}=1$, $W_{0i_1}=W_{0i_2}=0$, for $i=1,\cdots,p$

$W_{i_1 0}=1$, $W_{i_1 j_2}=0$, for $i=1,\cdots,p$

$W_{i_1 j_1}=$ 1 if $i=j$, 0 *otherwise* for $i, j=1,\cdots,p$

$W_{i_2 0}=1/(p-1)$ for $i=1,\cdots,p$

$W_{i_2 j_2}=$ $2+1/(p-1)$ if $i=j$, 0 *otherwise* for $i, j=1,\cdots,p$

$U=2+1/(p-1)$,

The value of the above solution, which is dual feasible, is ;

$Z_{LP}(D)=\sum_{i=1}^{n} V_i-kU=(3p^2-2pk-p+k-1)/(p-1)$, which is Z_{LP}.

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

Proposition 5 :

For $2\le k\le p$, an optimal integer solution for ⟨Figure 2⟩ is $y_0=1$, $y_{j1}=1$ for an $k-1$ spokes.

Proof :

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

The proposition 5 implies that even though a duality gap, $Z_{IP}-Z_{LP}$, always exists for the three 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 treee is that for $k=p$, there is no duality gap.

Proposition 6 :

For $k=p$, the duality gap vanishes for the tree of ⟨Figuere 2⟩. That is, $Z_{IP}=Z_{LP}$.

Proof :

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

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

Theorem 7 :

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

Proof :

Since the 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

$$\{3p^2 - 2p(k^*+a) - p + (k^*+a) - 1\}/(p-1) = Z_{LP}(k^*) - aU^*,$$

which is the optimal value according to the Theorem 4.//

We generalize the Theorem 4 by adding nodes to each spoke of tree of ⟨Figure 2⟩ in two different ways. First by adding one non-leaf node to each spoke, we have the following results.

Theorm 8 :

For $2 \leq k \leq p$, the optimal LP solution to the tree of ⟨Figure 3⟩ is :

(a) $y_0=(p-k)(p-1)$, $y_{j_1}=0$, $y_{j_2}=(k-1)/(p-1)$, $y_{j_3}=0$, for j=1,···,p

(b) $Z_{LP}(k)=\{2(3p^2-2pk-p+k-1)\}/(p-1)$

Proof :

The proof is same as the proof for the Theorem 4. Here we only give the dual variables V's and U

$V_0=2$, $U=4+2/(p-1)$

$V_{i_1}=1$, $V_{i_2}=2+1/(p-1)$, $V_{i_3}=3+1/(p-1)$, for i=1,···,p.//

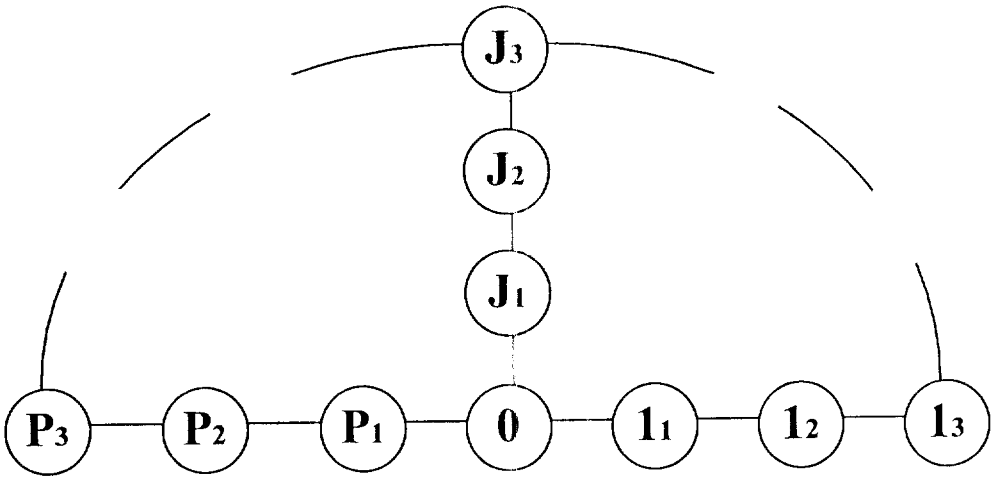
When we add m leaf-nodes to each spoke of the tree in ⟨Figure 2⟩ we have an example of infinitely large duality gap as n goes to infinity.

Theorem 9 :

For $2 \leq k \leq p$.

(a) optimal LP and IP solutions to ⟨Figure 4⟩ are sme as for the Theorem 3 and the Theorem 4 respectivley.

(b) $\lim_{m \to \infty} (Z_{IP}-Z_{LP})/Z_{IP}=(k-1)/\{k(k+3)\}$, and the maximum value is 1/9 attained when k=3



p : the number of spokes
each spoke consists of 3
nodes besides center node

〈Figure 3〉 A Tree with a Fracional Optimal LP Solution : 2

Proof :

For part (a), proof is same as for the Theorem 4. We just give the optimal dual solution V's and U;V's are same as those of the Theorem 4, U=(m+1)+m/p−1).

For part (b).

$Z_{LP}=\{pm(2p-k-1)+p^2-kp+k-1\}/(p-1)$,

$Z_{IP}=\{m(2p-k+1)+(p-k+1)$

Hence the relative gap is

$(Z_{IP}-Z_{LP})/Z_{IP}=\{m(k-1)\}+\{(p-1)(m(2p-k+1)+(p-k+1))\}$

$\rightarrow(k-1)/(p-1)(2p-k+1)$ as m $\rightarrow\infty$

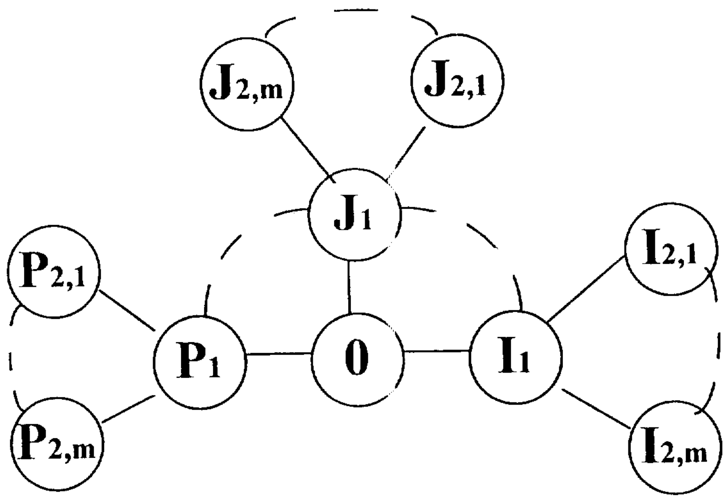
clearly the last fraction takes it maximum when p=k+1.//

Here we generalize the results of the Proposition 7 and the Theorem 9, and provide it in the following theorem.

Theorem 10 :

(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 $Z_{IP}\neq Z_{LP}$



p : # of spokes, m : # of leaves
each spoke consists of m+1
nodes besides center node

〈Figure 4〉 A tree with Infinitely Large Gap

(c) There is an infinite family of trees such that $(Z_{IP}-Z_{LP})/Z_{LP}\rightarrow r(k)$ where $r(2)=1/4$, $r(3)=1/3$ and $r(k)\rightarrow 1/2$

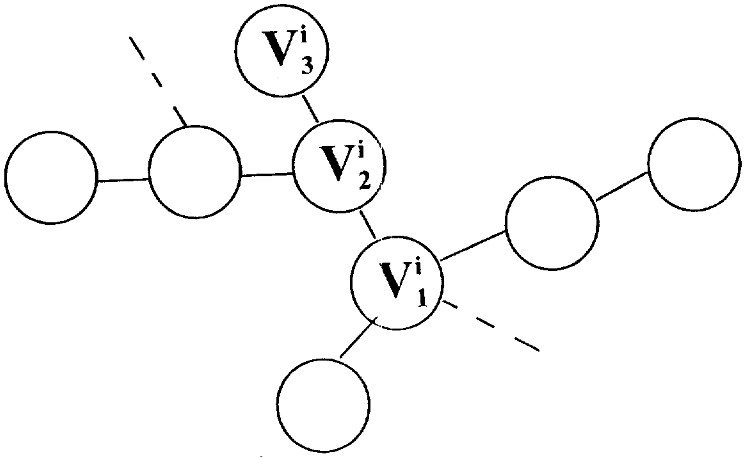
Proof :

For the 1-median problem, it is well-known that $Z_{IP}=Z_{LP}$ for every choice of c_{ij}, $1\leq i, j\leq n$.

When $k\geq[(n-1)/2]$ $Z_{IP}=Z_{LP}=n-k$ follows from the claim that every tree on n nodes has a dominating set of cardinality at most $[(n-1)/2]$. This claim is proved by inducion, It is true for $n=2$, or 3. Any tree on $n\geq 3$ nodes has at least one node which is not a leaf and is adjacent to at most one other non leaf node. Removing such a node and the adjacent leaves yields a tree with at most $n-2$ nodes. Putting V in the dominating set proves the claim.

To complete the proof of Theorem 10(a), it suffices to consider the case where n is even and $k=(n/2)-1$. A closer look at the proof of the above clami shows that the only trees which do not have a dominating set k are constructed inductively from a path with 4 nodes by adding paths $P_i=(V^i_1, V^i_2, V^i_3)$ where V^i_1 is one of the nonleaf nodes of the current tree and V^i_2, V^i_3) are two new nodes.(See 〈Figure 5〉). Form the construction $Z_{IP}=n-k+1=(n/2)+2$. Using the dual values $V_j=2$ if j is a leaf, 1 if not, Lemma 3 yields $Z_{LP}\geq(n/2)+2$. Therefore $Z_{IP}=Z_{LP}$.

To prove Theorem 10(b) when n is odd, consider the tree of 〈Figure 2〉. The number of nodes in the tree is 2p+1 and (n−1)/2=p. Hence Theorem 15(b) when n is odd follows the Theorem 4.



〈Figure 5〉 A tree constructed from a path with 4 nodes

To prove Theorem 10(b) when n is even, $n \neq 8$, we first consider the case $k \geq 3$. Add a node P_3 adjacent to P_2 to the tree of 〈Figure 2〉. Then it is optimum to choose P_2 in S and we can also choose $P_2=1$ in the LP solution. Removing P_1, P and P_3 we are back to the case where n is odd and $k \geq 2$. Now consider the case $n \geq 10$ and k=2. Add three nodes to the graph of 〈Figure 2〉, namely (i_{3+1}), for i=1, 2, 3.

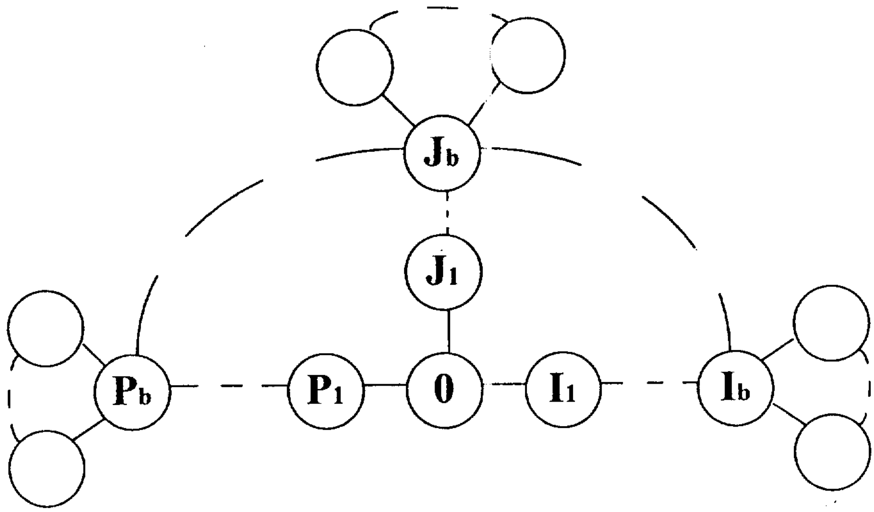
Then $Z_{IP}=3p+3$ but there is a better LP solution, namely, $y_0=1$ and $y_{1_1}=x_{2_1}=x_{3_1}=1/3$. This yields $Z_{LP}=3p+1$.

Finally, to prove Theorem 10(c). consider the tree of 〈Figure 6〉 which is modified of 〈Figure 4〉 and 〈Figure 5〉 in the following way. There are p spokes, and for each spoke, there are b number of non-leaf nodes besides the center node and m leaves.

The optimal LP solution is $y_0=(p-r)/(p-1)$, $y_{j_1}=(k-1)/(p-1)$, for j=1, 2,…,p. One obvious optimal IP solution is $y_0=1$, $y_{j_1}=1$ for any $k-1$ spokes. The values of the LP and IP solution are

$$Z_{LP}=b(k-1)/(p-1)+p[\{b(b+1)p-2bk-b(b-1)\}/2(p-1)+m\{(b+1)p-bk-1/(p-1)\}]$$

Z_{IP} is ;



〈Figure 6〉 A Tree where Relative Gap converges to 1/ 2

(i) when b=2a, $(a^2+m)(k-1)+(p-k+1)\{a(2a+1)+(2a+1)m\}$

(ii) when b=2a+1, $\{a(a+1)+m\}(k-1)+(p-k+1)\{(2a+1)(a+1)+2(a+1)m\}$

Hence the relative gap with p=k+1 goes to r(k) as m goes to ∞ (Note that we let m grow much faster than b, k and p)

$r(k) \to \{2a(k-1)\}/\{k(4a+k+1)\}$ when b=2a

$\to \{(2a+1)(k-1)\}/\{k(4a+k+3)\}$ when b=2a+1

In either case $r(k) \to 1/2$ as $k \to \infty$.

As stated previously the linear programming relaxation of the uncapacitated plant location problem on a graph has an integer optimal solution when the underlining graph is a tree. A similar but restictive result is known for k-median problem defined on a straight-line and on a cycle graph [8]. However, for more general graph, the appropriate conditions on distance matrix for attaining an integer optimal solution are not known. We conclude with the following proposition.

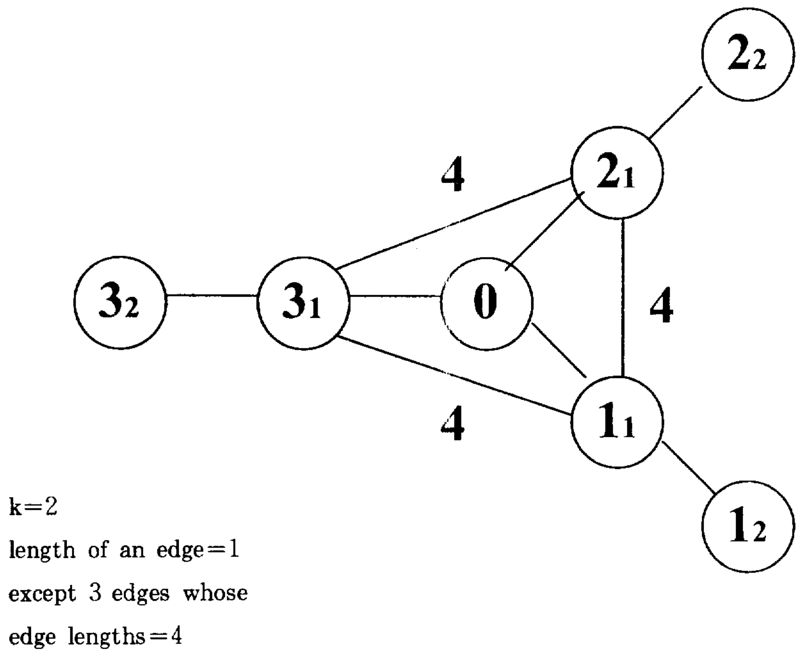
Proposition 11 :

Even when the underlying graph is a tree, a line graph, or a claw-free and triangulated graph, the linear programming relaxation of the k-median problem can have a fractional optimal solution.

Proof :

By a graph of ⟨Figure 7⟩.//

The unique optimal linear and integer solution is same as that of ⟨Figure 2⟩ with p=3.



⟨Figure 7⟩ A Graph of Duality Gap

References

[1] Ahn, S.H., C. Cooper, G, Cornuejols and A. Frieze, "Probabilistianalysis of a relaxtion for the k-median problem," *Mathematics of Operations Research* Vol. 13, No.1(Feb, 1988), pp. 1-31.

[2] Basley J.E., "A note on solving large p-median problems. technical report SW7 2BX," Imperial College, September, 1984.

[3] Boffey, T.B. and J. KarKazis, "P-median and multi-medians," *J. Oper, Res. Society* 35(1984), pp.57-64.

[4] Christofides, N., *Graph Theory: An Algorithmic Approach,* Academic Press, New York, 1975.

[5] Christofides, N. and J.E. Beasley, "A tree search algorithm for the p-mdeian problem," *European Journal of O.R., 10*(1982), pp.196-204.

[6] Cornuejols, G. and J.M. Thizy, "Some Facets of the Simple Plant Location Problem," *SIAM J. Algebraic and descrete Mathematics 3* (1982), pp.504-510.

[7] Cornuejols, G., G.L. Nemhauser and L.A. Wolsey, "Worst-case and probablilistic anaysisof algorithms for a location problem," *Operations Res., 28*(1980), pp.847-858.

[8] Cornuejols, G., M.L. Nemhauser, "Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms, Management Sci.," *Management Sci., 23*(1977), pp.789-810.

[9] Fisher M.L. and D.S. Hochbaum, "Probabilistic analysis of the planar k-median problem," *Mathematic of Operationd Research,* Vol.15, No.1(Feb,1980), pp.27-34.

[10] Francis, R.L. and J.A. White, *Facility Layout and Location:An Analytic Approach,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

[11] Handler, G.Y. and P.B. Mirchandani, *Location and Networks: Theory and Algorithms, MIT-press,* Cambridge, MA., 1979.

[12] Jacobsenm S.K. and P.K. Pruzan, Lokalisering-modeller & *Losningsmetoder, Studentinlitteratur,* Lund, Sweden, 1978.

[13] Kolen, A., Solving covering problems and the uncapacitated plant location problem on trees, *European Journal of O.R., 12*(1983), pp.266-278

[14] Krarup, J. and P.M. Pruzan, "The simple plant location problem: Survey and synthesis," Eurpoean Journal of O.R., 12(1983), pp.36-81

[15] Papadimitriou, C., "Worst-case and probabilistic analysis of a geometric location problem," *SIAM Jour. Computing,* Vol.10, No.3(1981), pp.542-557

[16] ReVelle, C.S. and R.S. Swain, "Central facilities location," *Geographical Anal,* Vol.2(1970), 30-42.

[17] Rosing, K.E. and C.S. ReVelle, Rosing-Vogelaar, "The p-median and its linear progrmming relaxion: an approach to large problems," *J. Operational Res. Soc. 30*(1979), pp.815-823.