Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 1 July 2000, pp. 30 – 43
Submitted: May 22, 1999. Published: July 15, 2000

Computational Complexity of Integrated Models of Network Design and Facility Location

J. Bhadury, R. Chandrasekharan, L. Gewali

Abstract

Transportation Network Design and Location Theory are two related areas that have been extensively studied in the past, primarily in isolation of each other. This is unfortunate because when the network is designed improperly, customers get extremely poor service even when service facilities are located optimally. Thus there is a need to investigate models where transportation networks are designed with future facility locations in consideration. In a partial attempt to address this gap, we develop a framework for an integrated model of transportation network design and facility location. In this model, transportation network is to be designed to allow the residents of the network to avail of the service provided by a single facility whose location is predetermined. Our principal contributions are to show that (i) our model reduces to a constrained minimal spanning tree problem and that (ii) it is NP Complete. We demonstrate that this complexity result holds even in the simple case where the customer sites are assumed to be given by the vertices of a convex polygon. This result thus shows that integrated models of facility location and network design can be computationally hard even in the simplest case, where only one facility needs to be located.

AMS subject classification: Subject Classification Codes 05, 11, 68
Keywords: Computational complexity; network design; facility location.

J. Bhadury, Department of Management and Finance, California State University-Hayward
E-mail: jbhadury@barney.sbe.csuhayward.edu
R. Chandrasekharan, School of Management, University of Texas at Dallas, Texas.
E-mail: chandra@utdallas.edu
L. Gewali, Department of Computer Science, University of Nevada, Las Vegas
E-mail: laxmi@cs.unlv.edu

1 Introduction

Transportation Network Design and *Facility Location Theory* are two (seemingly) different areas of study. The former refers to the body of methods that exist for designing a transportation network that is supposed to be used by a set of customers. When conceived in its broadest perspective, the study of network design captures virtually all important aspects of transportation planning: strategic, tactical and operational. As a result, it is an area that has been heavily studied by researchers in the past few decades - see Magnanti and Wong (1984)[9] for a comprehensive review. As mentioned in the above paper, most network design problems are computationally hard (NP Hard); however special instances, such as the Minimal Spanning Tree (MST) problem can be solved efficiently; see Prim (1957)[13], Kruskal (1956)[8]. On the other hand, *Facility Location Theory* seeks to develop methods that are designed to find optimal locations of facilities, both private and public, in order to serve a set of customers. These customers are assumed to avail of the facilities by traveling on a transportation network *whose design is already given*. Location theory is also another well established area with over 1000 published papers - see Mirchandani and Francis (1990)[12] and Drezner (1995)[3]. As mentioned in these texts, most Location Models are also NP Hard.

However, despite their popularity and umbilical connection, these two fields have been studied almost entirely in isolation. This is unfortunate because the very definition of optimal locations is constrained by the structure of the designed transportation network to be used by the customers. Furthermore, as shown in Bandelt and Labbe (1988)[1], when the network is designed improperly, customers get extremely poor service even when service facilities are located optimally. Thus there is a need to investigate models where transportation networks are designed with future facility locations in consideration.

In a partial attempt to address this gap, in this paper we begin by proposing a model that relates these two areas by incorporating facility location in the decision-making process involved in the design of a transportation network. Since this is essentially a first step in the confluence of these two areas, we begin by developing a simple model that seeks to design a transportation network that will be used by the customers in order to avail of the service provided by a *single* facility whose location has been pre-determined. Our principal contributions are to show that **(i)** our proposed network design problem reduces to a constrained MST problem and **(ii)** this problem is NP Complete. In fact, we demonstrate the computational complexity of this problem by showing that it remains NP Complete even in the simple case where all the users are assumed to be located on the vertices of a convex polygon. This result is surprising for two reasons: first and foremost, this result firmly establishes, for the first time, the computational complexity of integrated models of network design and facility location even in the simplest case where only one facility needs to be located. Second, in contrast to our problem, other NP Complete constrained MST problems are sometimes polynomially solvable in the Euclidean case (see for example, Ho et. al. (1989)[7]). By starting out with a different problem, we have also managed to demonstrate a Euclidean Constrained MST problem that remains NP Complete even in a simple instance where all the given points are vertices of a convex polygon.

The remaining paper is divided into three sections. The next section formally introduces our problem and discusses the preliminaries. Section three then demonstrates our primary result, namely, the NP Completeness of our model. Finally, the fourth section summarizes our findings.

2 Preliminaries

In this paper we study a problem of transportation network design around a given fixed location of a user-friendly facility. We begin by assuming that a Central Planning Agency (CPA), such a government, is responsible for constructing a transportation network to facilitate the usage of a given facility by a group of customers. The customers themselves are assumed to be dispersed on the N vertices of a given complete graph $G = (V, A)$ where V and A represent the set of vertices and arcs of G respectively. The arc connecting vertices v_i and v_j is denoted as (i, j). Associated with each vertex $v_i \in V$, is a non-negative weight w_i, which is referred to as the "weight" of the vertex v_i and represents the number of customers located at v_i. The total number of customers in G will be denoted by W. We assume that initially, there is no transportation network to allow travel by these customers. Thus, every arc $(i,j) \in A$, represents a *potential* link that can be constructed by the CPA to connect vertices v_i and v_j directly. The length of this link will be represented by l_{ij} and the cost of constructing it by c_{ij}.

The facility itself is assumed to be located at a vertex $v_m \in V$ and the CPA has to design the transportation network with this given location in mind - in practice such choices for facility location would probably have been made on the basis of factors such as location costs of the facility, available labor and material resources at this site or other externalities such as the case when governments target specific under-developed regions for development. This assumption however is not restrictive; it is possible to restate our model in the framework of an integrated model of facility location and network design where the facility location is also a decision variable. However, that problem can be tackled by repeatedly solving our model, once for each potential site of the facility. Hence, as it is a simple extension and does not alter the results of this paper, we do not discuss it here.

Suppose now that the CPA constructs a set of links $A' \subseteq A$; the resulting transportation network will be denoted by the subgraph $G' = (V, A')$ of G. We will say that an arc (i,j) is *present* in G' if and only if it represents a constructed link i.e., if and only if $(i,j) \in A'$ and the total cost of constructing G' will refer to the cost of constructing the links in A', and is denoted by $c(G')$. Thus,

$$c(G') = \sum_{(i,j) \in A'} c_{ij}$$

Given this, $d^{G'}(i,j)$ will be used to represent the length of any shortest path between vertices v_i and v_j in V, if travel is only allowed on the arcs that are present in G'. Thus, if G' is not a connected subgraph of G, then it is possible that such a path between v_i and v_j may not exist at all; when this happens, $d^{G'}(i,j)$ will be assumed to be infinity. Thus, with the given transportation network G' that has been constructed by the CPA, the total weighted distance of the facility at v_m to all the customers will be given by

$$F(m, G') = \sum_{i=1}^{N} w_i d^{G'}(i,m)$$

It is well known from location theory (Mirchandani and Francis (1990)[12]) that v_m is referred to as the *1-median* of G' if

$$F(m, G') \leq F(i, G') \ \forall \ v_i \in V$$

With these notations, we now turn to a description of two desirable properties of the transportation network that will be constructed by the CPA.

Connectedness Property: Any transportation network const-ructed by the CPA must be a connected subgraph of G. This property is obviously desirable to allow for communication with all the customer sites that this facility should service.

Median Constrained Property: Any transportation network constr-ucted by the CPA must have the property that the facility site v_m is the 1-median of this transportation network. In other words, the design of the transportation network is constrained by the location of this facility; hence the name of this problem. This property is desirable for two reasons: first, as evidenced by its definition, transportation networks with this property ensure that the average travel distance of all the customers to the service facility is minimized. The second reason is based on the voting paradigm ; see Hansen, Thisse and Wendell (1990)[6]. As we will shortly prove, the optimal transportation network is always given by a spanning tree of G'. Given that, the median constrained property ensures that the location of this facility at v_m is a "stable" location (see Hansen, Thisse and Wendell (1990)[6] for a proof). This means that it has the desirable property that once the transportation network has been constructed by the CPA, no alternate location of this facility will be strictly closer to more than half the total number of customers and thus be able to "defeat" the location at v_m in a pairwise election.

Given this, the objective of our model is to find the cheapest transportation network that has the connectedness and median constrained property. In other words, we assume that the CPA wishes to find the optimal set of arcs $A^* \subseteq A$ which, when constructed to yield the transportation network $G^* = (V, A^*)$ will satisfy

$$c(G^*) \leq c(G')$$

for all subgraphs G' of G that satisfy the connectedness and the median constrained property. We begin with the following interesting characterization of this optimal transportation network G^*.

Lemma 1 *It can be assumed without loss of generality that G^* is a spanning tree of G.*

Proof: It is readily seen that if only the connectedness property needs to be satisfied, then G^* can be assumed to be a spanning tree. However, what is non-trivial is to demonstrate that the assumption is still is true even if the median-constrained condition is imposed; that will be the focus of the proof.

Assume therefore, to the contrary, that G^* is not a tree. Clearly, v_m is a 1-median of G^*. Consider T', the shortest path tree from v_m to every other node in G^*. Since $d^{T'}(m,i) = d^{G^*}(m,i)$ for all vertices v_i, we have:

$F(m,T') = \sum_{i=1}^{n} w_i d^{T'}(m,j) = \sum_{i=1}^{n} w_i d^{G^*}(m,i) = F(m,G^*)$.

For any alternative vertex $v_p \in V$,

$F(p,T') = \sum_{i=1}^{n} w_i d^{T'}(p,i) \geq \sum_{i=1}^{n} w_i d^{G^*}(p,i) = F(p,G^*)$.

(since $d^{T'}(p,i) \geq d^{G^*}(p,i) \;\; \forall v_i \in V$)

Since $F(p,G^*) \geq F(m,G^*) = F(m,T')$ this implies that

$F(p,T') \geq F(m,G^*) = F(m,T') \;\; \forall p \in V$

which implies in turn that v_m is also a 1-median of T'. Thus T' satisfies both the connected and median-constrained property and further, $c(T') \leq c(G^*)$. ■

It is interesting to note that a similar result also exists for the UFLNDP - see for example, Property 2 in Melkote and Daskin (1997a)[10]. Given the above result however, henceforth we will only consider spanning trees of G. To that extent, let T be a given spanning tree of G. Then, the subtrees of T that are obtained by removing the vertex v_m will be denoted by $T_{m,1}, T_{m,2}, ..T_{m,f}$ where f is the degree of v_m in T. Without loss of generality, we will assume that the numbering of the subtrees obeys:

$$w(T_{m,1}) \geq w(T_{m,2}) \geq ...w(T_{m.f})$$

where $w(T_{m,k}), k = 1,..f$ represents the "weight" of the k^{th} subtree and is given by the sum of the weights of all vertices in $T_{m,k}$. With this definition, Goldman (1971)[5] has given an elegant property of the 1-median of a tree in term of the total weight W of all vertices of G.

Goldman's Property: The vertex v_m is a 1-median of T if and only if $w(T_{m,1}) \leq W/2$.

Since every transportation network in consideration must satisfy the median constrained property and hence, the Goldman's property, define $T(m)$ as the set of all spanning trees of G that have the vertex v_m as their 1-median and let $T^*(m)$ be the cheapest one among them, i.e.,

$$c(T^*(m)) \leq c(T)$$

for all spanning trees $T \in T(m)$. Then the optimization problem of the CPA can be stated as the following decision problem:

Median Constrained Minimum Spanning Tree Problem (MCMST)

Instance: A graph $G = (V, A)$ and a specified vertex $v_m \in V$.
Question: Find $T^*(m)$

When nodes are specified points in the plane, and the weight of an arc (i, j) is taken as the Euclidean distance between its nodes we get the Euclidean median constrained minimum spanning tree problem (**EMCMST**).

3 Complexity of MCMST

It is well known that while the MST problem is easily solvable, (Kruskal (1956)[8], Prim (1957)[13]), constrained MST problems are usually computationally hard; refer to section A2.1 of Garey and Johnson (1979)[4] for a detailed list of NP Complete constrained MST problems. To establish the complexity of MCMST we reduce two partitioning problem (2-P) to EMCMST.

2-Partitioning Problem (2-P)
Instance: Set D of n numbers $a_1, a_2, \ldots, a_n$

Question: Does there exist a subset B of D such that

$$\sum_{a_i \in B} a_i = \sum_{a_i \in D \setminus B} a_i$$

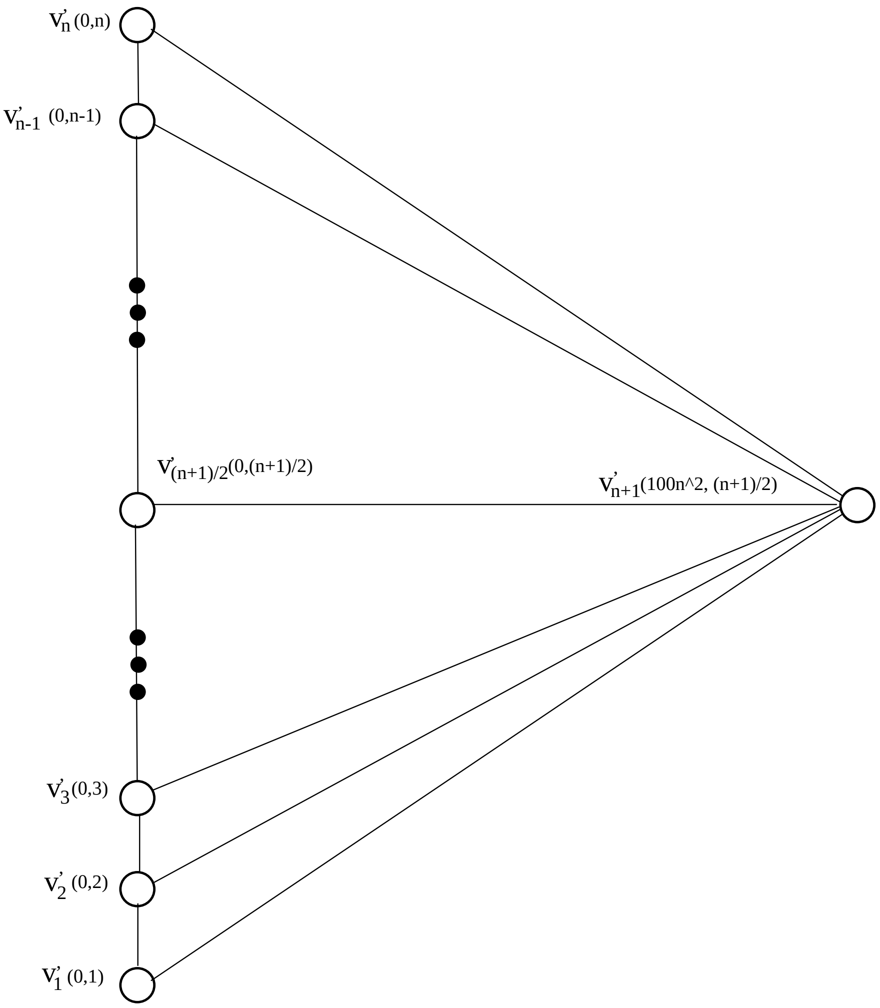

Figure 1: Illustrating the Construction of Graph G'

From an instance of 2-P we construct a complete Euclidean graph $G'(V', A')$ of $n+1$ vertices, where $V' = \{v_i' | a_i \in D\} \cup \{v_{n+1}'\}$. The first n nodes are placed equally spaced apart along a vertical line segment h with end points at coordinates $(0,1)$ and $(0,n)$. Specifically, the coordinates of $v_1', v_2', \ldots, v_n'$ are $(0,1), (0,2), \ldots, (0,n)$, respectively. The vertex v_{n+1}' is placed sufficiently far away to the the right of the vertical line segment. It is enough to place v_{n+1}' at $(100n^2, \frac{n+1}{2})$. This is illustrated in Figure 1 (the figure is not to the scale.) The weight of arc (i,j) is taken as $d(v_i', v_j')$ - the Euclidean distance between v_i' and v_j'. It can be easily verified that the graph $G'(V', A')$ as constructed above satisfies the following properties.

Property 1: $d(v_{n+1}', v_1') = d(v_{n+1}', v_n') = \sqrt{(\frac{n-1}{2})^2 + 10000n^4}$.

Property 2: $d(v_{n+1}', v_{\frac{n+2}{2}}') = 100n^2$

Property 3: $\max_{1 \le i \le n} d(v_{n+1}', v_i') = d(v_{n+1}', v_1') = d(v_{n+1}', v_n')$

Property 4: $\min_{1 \le i \le n} d(v_{n+1}', v_i') = d(v_{n+1}', v_{\frac{n+1}{2}}') = 100n^2$

Property 5: Let G^c denote the complete graph with nodes $v_1', v_2', \ldots, v_n'$. The maximum spanning tree of G^c is given by arcs $(1,n), (n,2)$, $(2, n-1), \ldots, (\frac{n+1}{2}, \frac{n+1}{2}+1)$. The weight of the maximum spanning tree is no more than $(n-1) + (n-2) + (n-3) + \ldots + 1 = \frac{n(n+1)}{2}$. This is shown in Figure 2.

Property 6: Any spanning tree with v_{n+1}' as median must have at least two arcs incident at v_{n+1}'.

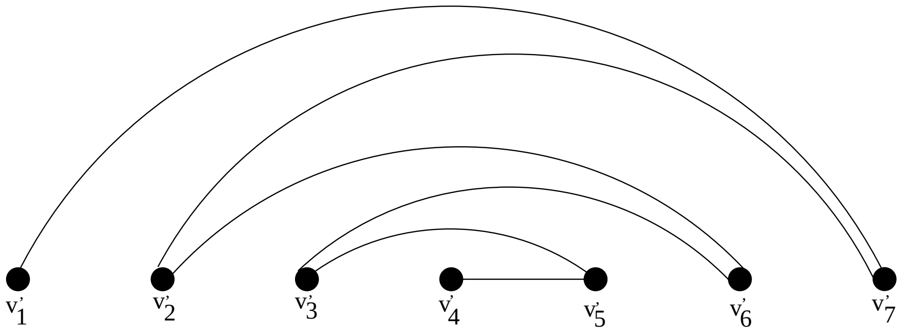

Figure 2: Illustrating the Maximum Spanning Tree

We use $T(m, v'_{n+1})$ to denote a spanning tree of $G'(V', A')$ with v'_{n+1} as median and with exactly m arcs incident at v'_{n+1}. Property 6 implies that $T(1, v'_{n+1})$ does not exist. The following lemma establishes that it is enough to consider only $T(2, v'_{n+1})$'s for finding a minimum weight spanning tree constrined to have median at v'_{n+1}.

Lemma 2 *Length of any $T(2, v'_{n+1})$ is less than the length of any $T(m, v'_{n+1})$, $m > 2$.*

Proof: Let U_2 be an upper bound of the weight of $T(2, v'_{n+1})$. Similarly, let L_3 be a lower bound of the weight of $T(m, v'_{n+1})$, $m > 2$. We observe that the weight of each arc of any $T(2, v'_{n+1})$ incident at v'_{n+1} must have length no more than the length of arcs $(n+1, 1)$ or $(n+1, n)$. Also the total length of the remaining $n-3$ arcs of any $T(2, v'_{n+1})$ is no more than the length of the maximum spanning tree of G^c (Property 5). Hence we can write:

$$U_2 \quad = \quad 2\sqrt{\frac{(n-1)^2}{4} + 10,000n^4} + \frac{n(n+1)}{2} \tag{1}$$

The total length of m arcs of $T(m, v'_{n+1})$, $m > 2$, incident at v'_{n+1} must be at least three times the length of the arc $(n+1, \frac{n+1}{2})$. Hence we can write:

$$L_3 \quad = \quad 300n^2 \tag{2}$$

Combining the above two relations we get:

$$\begin{aligned} L_3 - U_2 \quad &\geq \quad 300n^2 - 2\sqrt{\frac{(n-1)^2}{4} + 10,000n^4} \qquad (3) \\ &\geq \quad 98n^2 \text{ (a positive quantity)} \end{aligned}$$

∎

Let A be the set of numbers representing the capacities of n nodes $v'_1, v'_2, \ldots, v'_n$of $G'(V', E')$. The following lemma reduces the partitioning problem (2-P) to EMCMST.

Lemma 3 *There exist a 2-partition of D if and only if there exist a minimum spanning tree of $G'(V', A')$ with v'_{n+1} as median and total length $\leq U_2$.*

Proof: (2-P $\Rightarrow$ EMCMST): Suppose that we are given a 2-partitioning of D as $D_1 = \{ a'_1, a'_2, \ldots, a'_k \}$ and $D_2 = \{ a"_1, a"_2, \ldots, a"_r \}$, $k+r = n$. We refer to the nodes corresponding to D_1 as "first subset" and those corresponding to D_2 as "second subset". Given D_1 and D_2, we construct a spanning tree T' as follows. We construct two "special arcs" incident at v'_{n+1}. One of the special arcs is connected to a node in the first subset and the other to a node in the second subset. T' is obtained by adding special arcs to the spanning trees of the first subset and the second subset. Since the total weights of the first subset and the second subset are $\frac{W}{2}$ each, v'_{n+1} must be the median of T' (Goldman's condition); and the length of T' is no more than U_2.

(EMCMST $\Rightarrow$ 2-P): Suppose we are given a spanning tree of $G'(V', A')$ with v'_{n+1} as its median and total length $\leq U_2$. Since the total length is $\leq U_2$, Lemma 2 implies that there must be exactly two arcs incident at v'_{n+1}. Also by Goldman's condition each of the two subtrees of v'_{n+1} must have weight $\leq \frac{W}{2}$. Since the weight of both subtrees of v'_{n+1} can not be less than $\frac{W}{2}$, we must have that each subtree is of weight exactly $\frac{W}{2}$. Thus the capacities of the nodes of the subtrees of v'_{n+1} in T' give a two partitioning of D. ∎

All the construction described in Lemma 3 take polynomial time. Furthermore, since 2-P is known to be NP-complete (Garey and Johnson, 1979[4]), Lemma 3 implies that EMCMST is NP-hard. Since a proposed solution for EMCMST can be verified easily in polynomial time, EMCMST belongs to the class NP. Hence we have the following theorem.

Theorem 1 *EMCMST is NP-Complete.*

ECMST remains NP-complete even when we require all the nodes to lie on the vertices of a convex polygon. This can be verified by the following simple observation. The nodes in Figure 2 lie on the boundary of a convex polygon such that all nodes except one (v'_{n+1}) are degenerate i.e., $v'_1, v'_2, ..., v'_n$ are collinear. The collinear degeneracy is not necessary, and can be removed easily, without violating Lemma 2 and Lemma3. We can slightly perturb the collinear vertices appropriately so that all vertices are in a general position (no three vertices are collinear) and also on the boundary of a convex polygon. Equation 3 is still valid in the modified construction and Lemma 3 follows. Hence we have the following corollary.

Corollary 1: *EMCMST remains NP-complete even when all nodes are required to lie on the vertices of a convex polygon.*

Remark: Since EMCMST is a restricted version of MCMST, it follows that MCMST is also NP-complete.

4 Conclusion

In this paper we have studied the problem of designing a transportation network to facilitate the use of a given service facility by the residents of the network. Although we have assumed throughout that the location of this facility is fixed, our model can also be used to solve a problem where both the facility location and the network design are decision variables. Our first major result is to show that under certain assumptions, our model reduces to a new variant of the well known family of constrained MST problems. Then we demonstrate that this problem is computationally hard, i.e. NP Complete, and remains so even when we make many simplifying assumptions.

Future work on this topic could examine the development of exact algorithms and efficient heuristics for this problem. Another strand of research could be to look for polynomially solvable instances of this problem and develop efficient solution procedures for them. This is the focus of the ongoing work in Bhadury et. al. (1998)[2].

References

1. Bendelt, H.J. and Labbe, M. *How Bad Can A Voting Location Be ?*. Journal: Social Choice and Welfare, Vol. 3, (1986) 125-145.

2. Bhadury, J., Chandrasekaran, R., Gewali L.P., Badr, M.H. *An Integrated Model of Facility Location and Network Design* (work in progress). Also presented at the *CORS/INFORMS Conference*, Montreal, (Spring 1998).

3. Drezner, Z. (editor), *Facility Location: A Survey of Applications and Methods*, Springer-Verlag New York, Inc, (1995).

4. Garey M.R. and Johnson D.S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York (1979).

5. Goldman A.J. (1971). "Optimum Center Location In Simple Networks". *Transportation Science*, 5, (1979) 212-221.

6. Hansen, P., Thisse, J.-F. and Wendell, R.E., *Equilibrium Analysis for Voting and Competitive Location Problems*, in *Discrete Location Theory* (P.B. Mirchandani and R.L. Francis (eds.)), John Wiley and Sons, New York, (1990) 479-501.

7. Ho, J., Lee, D.T., Chang, C.H., Wong, C.K. *Bounded-Diameter Minimum Spanning Trees and Related Problems*, *Proceedings of ACM Symposium on Computational Geometry*, (1989) 276-282.

8. Kruskal, J.B., *On The Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem*, Proceedings of the American Mathematical Society, 7, (1956) 48-50.

9. Magnanti, T.L. and Wong, R.T., *Network Design and Transportation Planning: Models and Algorithms*, Transportation Science, Vol. 18, No. 1, (February 1984) 1-55.

10. Melkote, S., and Daskin M.S., An Integrated Model of Facility Location and Network Design, Working Paper, School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, (1997a).

11. Melkote, S., and Daskin, M.S., *Polynomially Solvable Cases of Combined Facility Location-Network Design Problems*, Working Paper, School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, (1997).

12. Mirchandani P.B. and Francis, R.L. (1990) eds. *Discrete Location Theory*, John Wiley and Sons, New York (1990).

13. Prim, R.C., *Shortest Connection Network and some generalizations*, Bell System Technical Journal, 36, (1957) 1389-1401.