Submitted to *Transportation Science*
manuscript (Please, provide the mansucript number!)

A New Heuristic Formulation for a Competitive Maximal Covering Location Problem

Tolga H. Seyhan
Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, tolgaseyhan@lehigh.edu,

Lawrence V. Snyder
Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, larry.snyder@lehigh.edu,

We consider a competitive facility location problem in which two firms engage in a leader–follower game. Both firms would like to maximize the customer demand that they capture. Given the other player's decision, each player's problem is the classical Maximal Covering Location Problem (MCLP). That is, the leader has to solve a bilevel problem in which the second-level problem is NP-hard. To overcome this, we use the greedy add algorithm as a proxy for the follower's response and formulate a mixed-integer programming (MIP) model that embeds the follower's heuristic response into the leader's constraints, and solve it as a single-level problem. The resulting formulation is tractable and provides near-optimal solutions for the leader's decision. We analyze the performance of the heuristic both theoretically and computationally. We also provide alternate formulations for the same problem and compare them.

Key words: Facility location; Stackelberg game; bilevel programming

History:

1. Introduction

Facility location problems are pivotal in strategic decision making. Most firms face some sort of facility location problem, and these problems are often difficult to solve. Classical facility location models tend to ignore the effect of competition on the location decision and act as if there were a single decision maker. This is reasonable for locating public facilities like airports or fire stations, which have virtually no competitors. However, in a market, taking competitors' responses into account in the location decisions can significantly alter the resulting solutions and improve the realized outcomes.

In this paper, we consider a competitive version of the Maximal Covering Location Problem (MCLP). There are two firms, who engage in a Stackelberg game, thus sequentially enter a new market. Each firm enters the market by locating multiple facilities, which are selected from a discrete set of potential facilities. Once the locations are decided, customers patronize their preferred facilities. The goal of each firm is to maximize the total customer demand that they serve at the end of the game. The second mover (i.e the follower) observes the first mover's (i.e. leader's) decision and responds optimally. We assume foresight on the part of the leader, and we solve her problem, taking the follower's optimal (i.e. best) response into account. The customers are represented as a discrete set of points and their demands are assumed to be deterministic.

The difficulty in solving this model comes from two sources. To begin with, MCLP is an NP-hard problem (Megiddo et al. 1983), therefore no polynomial time algorithm is known to solve it. However, this is really not the greater problem as large instances of these problems can be dealt with efficiently. The major difficulty in solving our problem is that it has a bilevel structure, with conflicting objectives, and each of the levels can be viewed as instances of MCLP given the other.

Our approach aims to overcome this difficulty by replacing the follower's problem with a reasonable heuristic, embedding this heuristic into the leader's problem as constraints, and thus solving the initially bilevel problem as a single-level problem.

We provide an introductory review of the literature on facility location problems that deal with competition in Section 2. In Section 3, we introduce our problem and our approach to modeling and solving it. We present our model in Section 4 and discuss two alternate formulations. Section 5 briefly analyzes the worst-case performance of our approach, while Section 6 provides computational results on the solution performance and the quality of the solutions generated by the proposed models. Section 7 concludes.

2. Literature Review

In this section, we provide a concise summary of significant works on competitive facility location and draw a historical timeline introducing the predecessors of our model. Facility location problems are extensively studied and the literature on them is vast, and a general review of the facility location literature is outside the scope of this paper. Interested readers may refer to the books by Daskin (1995), Drezner (1995) and Drezner and Hamacher (2002). Snyder (2010) and ReVelle and Eiselt (2005) provide broad overviews of facility location problems. ReVelle et al. (2008) provide a recent bibliography on the fundamental problem types in location science. Owen and Daskin (1998) discuss mainly dynamic and stochastic problems in facility location. Snyder (2006) discusses stochastic and robust facility location models. Klose and Drexl (2005) discuss mixed integer programming models in a wide range of facility location problems.

We can describe competitive facility location problems as any facility location problem that incorporates spatial competition among at least two firms. The decisions of these firms are interdependent and affect each other's market share. Competitive problems within the facility location field can be traced back to the seminal work by Hotelling (1929). Hotelling introduced a market that is represented by a line segment. Two firms, whose locations on this market are known, compete by setting their prices to maximize their profits. He also discussed a case, in which the first firm's location is fixed. The result asserts that the second firm locates itself arbitrarily close to the first one, which extends to the sameness of the product features and the positions taken by political parties. Considering location as an attribute of the product (e.g., store) gave rise to a plethora of studies. Hotelling's model has been revisited several times with different settings for strategic variables, moving sequences, customer preferences and number of firms; including a correction by (d'Aspremont et al. 1979). Reviews of these models can be found in Gabszewicz and Thisse (1992), Anderson et al. (1992), Eiselt (2011), and Younies and Eiselt (2011). These location models on a line are very effective in providing theoretical insights about spatial competition but their representative power fails to deal with the facility location problems that operations research mainly works on. Therefore, in the early 1980's several studies were introduced to incorporate competition among decision makers into the classical facility location problems.

One of the earliest attempts was by Hakimi (Hakimi 1983), who extended the linear decision space in the Hotelling model to a network. He studied locating a fixed number of new facilities on a network where there were some existing facilities. Customers are located on the vertices and they have associated weights. Each customer prefers the closest open facility. Finally, the decision maker's objective is to maximize the weighted sum of customers who prefer his facilities.

Eiselt et al. (1993) provide a five-criteria (space, number of players, pricing policy, number of players and behavior of customers) taxonomy of competitive facility location problems and classify more than one hundred articles accordingly. Several papers survey parts of this classification. Plastria (2001) presents a survey on static competition in facility location and a large section on classification. Eiselt and Laporte (1996) report on sequential models, mostly of the leader–follower type of facility location problems. Serra and ReVelle (1995) provide a discussion of competitive location in discrete space, which also brings together several of the models that will be explained below. Most recently, Kress and Pesch (2012) surveyed recent developments in sequential competitive location problems.

The problem we address originates from the maximal covering location problem (MCLP) introduced by Church and ReVelle (1974). In the classical MCLP, customers are represented as demand points and they can be served only by a facility within a pre-specified service distance. The decision maker locates a fixed number of facilities in order to maximize the total demand covered by the facilities. The MCLP is one of the pillars of facility location and numerous variants have been studied. Snyder (2011) reviews covering problems including a detailed discussion of the MCLP.

Note that this original problem does not consider competition. ReVelle (1986) extends the MCLP to the competitive domain and introduces the maximum capture (MAXCAP) problem. This is an example of static competition. The decision maker locates a fixed number of facilities in a market where competing facilities already exist, just as the follower in our problem does. Each customer prefers the closest open facility and the demand is equally shared if the closest open facilities of each firm are at equal distance from the customer. This problem is, in fact, equivalent to the p-median problem as covering, capturing and sharing are expressable as the weight of an edge between a facility and a customer. In a subsequent paper Serra et al. (1996) consider demand uncertainty for the MAXCAP problem.

Serra and ReVelle (1994) reformulate the MAXCAP as a Stackelberg-type problem (competition with foresight). The game described in the MAXCAP stays the same but now the leader's problem is solved. This model is named the preemptive capture problem. For reasons we explain in the coming sections, this problem is quite hard to formulate and solve as an IP. Serra and ReVelle describe a simple heuristic procedure to solve the problem. The leader locates her facilities first. The follower responds by solving the MAXCAP problem. Then the leader follows a one-opt procedure, exchanging a single open facility with a closed one at each iteration and re-solving for the follower's response. The new solution is accepted if the leader's demand increases and the procedure is continued until no more improvements are achieved. An interesting variant of the problem is described in (Serra and ReVelle 1995), in which the number of facilities to be opened by the opponent is defined probabilistically, and the probability monotonically decreases as the number of leader facilities increases.

Finally, Plastria and Vanhaverbeke (2008) revisit the problem and formulate three MIP models, robust models MaxMin, MinRegret, and Stackelberg, where they restrict the follower decision to a single facility. This assumption allows them to formulate the problem without needing a bilevel formulation. In the models we present below, we extend their Stackelberg approach to allow multiple follower facilities. However, we do not solve for an exact Stackelberg-Nash equilibrium; rather, we describe the follower response heuristically.

From a mathematical programming perspective, our problem is a 0-1 *integer bilevel linear programming problem* (IBLP), where the upper and lower levels correspond to the leader's and the follower's decision stages, respectively. (See (Colson et al. 2007) for a review of bilevel programming.) There are a few solution approaches, including classical branch-and-bound (Bard and Moore 1992, Moore and Bard 1990), cutting planes, and branch-and-cut variants (DeNegre and Ralphs 2009), that have been proposed to solve problems of this type. These algorithms typically struggle when applied to problems with more than a few variables and constraints. More efficient algorithms exist either for problems with continuous variables (which is still an NP-hard optimization problem (Jeroslow 1985)) and problems with special exploitable structures. In general, the existing practical needs for IBLP significantly exceed current solution capabilities (Bard 2009). We propose to reformulate the problem as a single-level 0-1 integer linear problem, and represent the follower's problem by a heuristic that is formulated using additional variables and constraints, whose cardinality is polynomial in the original number of variables and constraints. As the heuristic does not guarantee optimal solutions for the follower, the result is not guaranteed to be optimal for the leader. On the other hand, computational experiments show that the model performs quite well in terms of computational effort and solution quality.

3. The Problem

Before discussing the model, we introduce the problem and related terminology. Then, we describe the mechanics of the underlying Stackelberg game and how we embed the follower's problem into the leader's constraints.

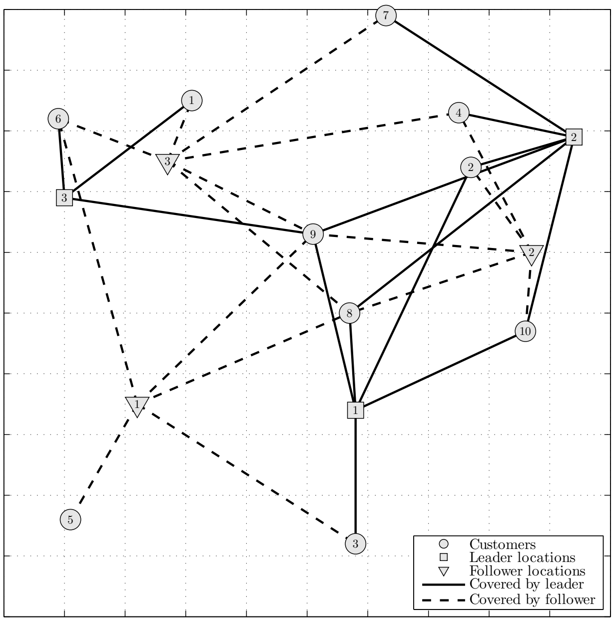


Figure 1 Customers, potential facilities and coverage relations.

In our model there are two competing firms, the leader (L) and the follower (F). Both firms have a number of potential locations where they can build and operate facilities to serve customers. We denote the leader's potential locations with $s \in S$ and the follower's potential locations with $t \in T$. We assume $S \cap T = \emptyset$ for the moment and in the numerical analysis in Section 6. However, this assumption can be altered without changing the model as we explain later in this section. The leader and the follower, respectively, open B and K facilities. Customer demand is assumed to accumulate at a number of discrete points and we denote these points with $i \in I$ and the associated demand with d_i. Both firms would like to maximize their profits, thus, the total demand they serve.

We define two relations among the components: *coverage* and *preference*. It is customary to use a distance analogy for these relations. An open facility *covers* all the customers that are at most a given service distance (i.e. radius) away from it. Each customer *prefers* the closest open facility that covers her. If there is no open facility that covers a particular customer then that customer is not served. Figure 1 illustrates a simple instance and the coverage relationship among problem entities on a grid assuming a radius of 5 units. We represent these relations as subsets. For a given customer $i \in I$ the leader [follower] locations that cover her are denoted as the subset $S_i \subseteq S$

$[T_i \subseteq T]$. We denote the set of leader facilities that customer i would prefer over follower location t by S_{it}.

The model assumes that the leader is the the first mover. She makes her decision and locates her facilities, then after some time, the follower enters the market and locates his facilities. The follower's problem is to maximize the demand he serves, given the leader's solution. When both parties locate their facilities, the winner for each customer demand is decided according to the outcome matrix presented in Table 1. The winner firm is said to *capture* the customer. Figure 2 illustrates the optimal solution to the specified problem for $B = K = 2$, and the customers captured by each player.

Table 1 Outcome matrix for customer i : Winner is Leader (L), Follower (F), none (X) or situation is impossible (I).

		Leader, s		
			$s \in S_i$	
		$s \notin S_i$	$s \in S_{it}$	$s \notin S_{it}$
Follower, t	$t \notin T_i$	1 X	2 L	3 I
	$t \in T_i$	4 F	5 L	6 F

Since the leader is not going to alter her solution after the follower acts, the follower does not need to take into account any strategic response and his problem is relatively easy. It reduces to the MCLP. On the other hand, the problem of the leader is a lot harder as she has to take into account her opponent's strategic response, too. This is difficult for two reasons. The follower, even though we restrict him to open exactly K facilities, needs to solve a combinatorial optimization problem which is known to be NP-hard. As almost all facility location problems (including the MCLP) are in this category, this is not surprising. The bad news is that the follower's objective conflicts with the leader's, yet has to be optimized in the leader's problem. Since we are not able to put the follower's objective in the objective function next to the leader's, we ensure that the

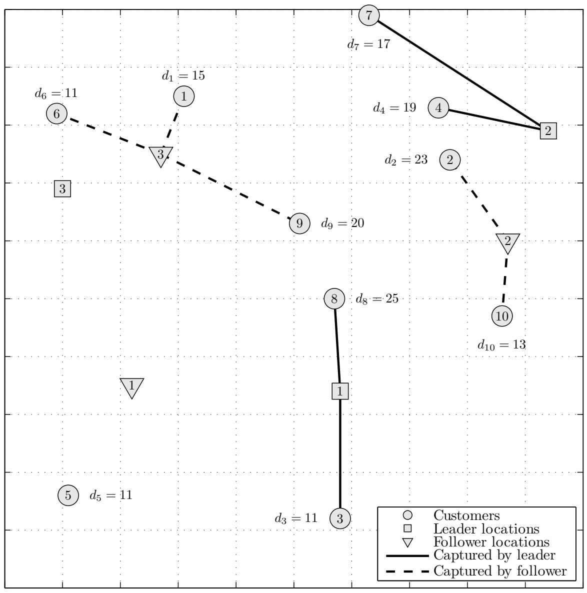


Figure 2 Solution: Opened facilities and captured customers.

follower problem is solved optimally through constraints, basically by making each suboptimal follower solution infeasible to the leader's problem, as follows:

$$\max \quad LeaderCapture(L, F) \tag{1}$$

$$\text{s.t.} \quad \ldots \tag{2}$$

$$FollowerCapture(L, F) \geq FollowerCapture(L, F') \qquad \forall F' \tag{3}$$

Here, L and F denote the leader's and follower's location vectors, respectively. The follower has $C(|T|, K)$ possible choices. This value is very large for large $|T|$ and gets larger as K approaches $|T|/2$. Therefore the number of required constraints (3) increases very fast, becoming hard to manage even for problems of moderate size. Plastria and Vanhaverbeke (2008) solve this problem by setting $K = 1$. They assume that the follower responds by opening a single facility, therefore keeping the number of constraints small.

Instead, we propose the following approach. We don't solve the follower's problem optimally, but replace it with a heuristic solution. We assume that the follower implements the *greedy add algorithm* given the locations of the leader facilities. The algorithm is described in Section 4.2. The leader views this heuristic as a proxy for her opponent's true optimal response, thus maximizes the total demand covered by her facilities assuming that the follower responds by applying his heuristic. We now introduce the rest of the notation and explain the proposed model and the accompanying heuristic.

4. The Model

4.1. Terminology and Notation

We revisit the sets and parameters, and introduce the variables of the model.

Sets

I: customers

S: potential leader facilities. Subset S_i denotes the set of facilities that can serve (*cover*) customer $i \in I$. Subset S_{it} denotes the set of leader facilities that customer i would prefer to the follower facility t.

T: potential follower facilities. Subset T_i denotes the set of follower facilities that cover customer $i \in I$.

Parameters

d_i: demand of customer $i \in I$

B: number of facilities that the leader will open

K: number of facilities that the follower will open, hence the number of greedy add algorithm iterations.

Decision Variables

L_s: binary variable that equals 1 if leader opens facility s.

F_{tk}: binary variable that equals 1 if follower opens facility t at the k^{th} iteration of the greedy add algorithm.

x_i^L: binary variable that equals 1 if customer i is captured by the leader.

x_{ik}^F: binary variable that equals 1 if customer i is captured by the follower facility opened at the k^{th} iteration of the greedy add algorithm.

y_{itk}: binary variable that equals 1 if customer i is still capturable by follower facility $t \in T_i$ in the k^{th} iteration of the greedy add algorithm (definition of "capturable" to be made more precise later).

Although we define the last three variables as binary variables, we will show that if they are defined as continuous variables, the problem will still have an all-integer solution.

4.2. Follower's Response

Let us now analyze our assumption on the follower's response. The greedy add procedure (Kuehn and Hamburger 1963) is as follows. At the beginning, all follower locations are closed. The follower receives the decision vector $L = [L_s]$ from the leader and updates the y variables. If a customer i can be covered by a particular follower facility t, then the customer is still *capturable* at iteration k ($y_{itk} = 1$) if there is no better open leader facility, otherwise it is not ($y_{itk} = 0$). Then for each facility that is still closed, the follower calculates the total demand that the facility could capture if it were open ($\sum_{i \in I} d_i y_{itk}$). The facility that can capture the largest demand (t') is opened. The customers who were capturable by this facility are captured ($x_{ik}^F \leftarrow 1$) and are removed from further consideration ($y_{itk'} \leftarrow 0$ for $k' > k$). This is repeated until the follower opens K facilities. This algorithm is included as a feasibility problem through constraints in the leader's model, as described in Section 4.3.

We selected this algorithm as the follower's response mainly for two reasons. First, the greedy add algorithm is a well known procedure for facility location problems. The quality of the solution it provides depends heavily on the characteristics of the underlying graph for the follower's problem. It does not guarantee optimality in the majority of the cases. The leader's solution is optimal if, after fixing the leader's solution, the follower's best response (solution of the resulting MCLP) and the result of the greedy add algorithm give the same result. The second reason is that it enables us to embed the follower's response into the leader's problem using a polynomial number of new constraints. (In contrast, recall from Section 3 that a straightforward embedding of the follower's problem into the leader's would require an exponential number of constraints.) In the computational study, we discuss this and the effectiveness of the approach. Next we introduce and explain our mathematical model.

4.3. Leader's Problem

The leader aims to maximize the demand she captures. The objective function is therefore the total demand captured by the leader.

$$\sum_{i \in I} d_i x_i^L \tag{4}$$

The leader is going to open at most B facilities.

$$\sum_{s \in S} L_s \le B \tag{5}$$

(B can also be interpreted as an investment budget if we introduce a fixed cost of investment for each facility.) The greedy add algorithm requires exactly one follower facility to be opened at each iteration $k = 1, \dots, K$.

$$\sum_{t \in T} F_{tk} = 1 \quad \forall k = 1, \dots, K \tag{6}$$

A particular follower facility $t \in T$ can be opened at most once throughout the algorithm.

$$\sum_{k=1}^{K} F_{tk} \le 1 \quad \forall t \in T \tag{7}$$

A particular customer $i \in I$ can be captured by at most one of the firms.

$$x_i^L + \sum_{k=1}^{K} x_{ik}^F \leq 1 \quad \forall i \in I \tag{8}$$

The leader has to have an open and covering facility to capture customer i. That is, $x_i^L = 1 \Rightarrow \exists s \in S_i$ such that $L_s = 1$, or equivalently $L_s = 0$ $\forall s \in S_i \Rightarrow x_i^L = 0$.

$$x_i^L \leq \sum_{s \in S_i} L_s \quad \forall i \in I \tag{9}$$

The above requirement is valid for the follower, too. $x_{ik}^F = 1 \Rightarrow \exists t \in T_i$ such that $F_{tk} = 1$ or equivalently $F_{tk} = 0$ $\forall t \in T_i \Rightarrow x_{ik}^F = 0$.

$$x_{ik}^F \leq \sum_{t \in T_i} F_{tk} \quad \forall i \in I, k = 1, \ldots, K \tag{10}$$

After the leader's decision is finalized we need to determine whether a customer i is still capturable by a covering facility $t \in T_i$. This is the initialization step of the greedy add algorithm. The following two constraints determine if a customer is still capturable by a facility. Given a customer i and a covering potential follower facility t, if the leader has no better facility than t, then i can be capturable by the follower by opening facility t. We denote this situation as $y_{it1} = 1$.

$$1 - y_{it1} \leq \sum_{s \in S_{it}} L_s \quad \forall i \in I, t \in T_i \tag{11}$$

On the other hand, if leader opens a better facility $s \in S_{it}$ then the follower cannot capture customer i through facility t. We denote this situation as $y_{it1} = 0$.

$$1 - y_{it1} \geq L_s \quad \forall i \in I, t \in T_i, s \in S_{it} \tag{12}$$

For the remaining iterations ($k \geq 2$) we update the value of y_{itk} as follows: if i is initially capturable ($y_{it1} = 1$) and it is not captured by the follower yet ($\sum_{k' < k} x_{ik'}^F = 0$) it is still capturable ($y_{itk} = 1$).

$$y_{itk} \geq y_{it1} - \sum_{k'=1}^{k-1} x_{ik'}^F \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \tag{13}$$

A customer becomes noncapturable once it is captured.

$$y_{itk} \leq 1 - \sum_{k'=1}^{k-1} x^F_{ik'} \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \tag{14}$$

Finally, a noncapturable customer stays noncapturable.

$$y_{itk} \leq y_{i,t,k-1} \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \tag{15}$$

The greedy add algorithm is implemented through the following constraints. At each iteration k, the follower captures a total demand that is at least equal to the largest capturable demand by a single follower facility.

$$\sum_{i \in I: t \in T_i} d_i y_{itk} \leq \sum_{i \in I} d_i x^F_{ik} \quad \forall t \in T, k = 1, \ldots, K \tag{16}$$

At iteration k, customer i is captured if the follower opens a covering facility t and the customer is capturable by t.

$$x^F_{ik} \geq F_{tk} + y_{itk} - 1 \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \tag{17}$$

On the other hand, customer i is not captured if the follower opens a covering facility t but the customer is not capturable by t.

$$x^F_{ik} \leq 1 - F_{tk} + y_{itk} \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \tag{18}$$

Taken together, (10), (17), and (18) ensure that $x^F_{ik} = 1$ if and only if i was captured by the follower facility opened at iteration k. Finally, the variables have the following restrictions.

$$L_s, F_{tk} \in \{0, 1\} \quad \forall s \in S; t \in T, k = 1, \ldots, K \tag{19}$$

$$x^L_i, x^F_{ik}, y_{itk} \geq 0 \quad \forall i \in I; t \in T_i, k = 1, \ldots, K \tag{20}$$

THEOREM 1. *The formulation above ensures the outcomes of the game depicted in Table 1.*

Proof. Each of the six cases (cells) in the outcome matrix are satisfied as follows.

1. If $L_s = 0$ for all $s \in S_i$ and $F_{tk} = 0$ for all $t \in T_i$, then $x_i^L = 0$ and $x_{ik}^F = 0$ by constraints (9) and (10), respectively.

2. If $F_{tk} = 0$ for all $t \in T_i$ and $k = 1, \ldots, K$ and there is at least one $s \in S_i$ such that $L_s = 1$, then $x_{ik}^F = 0$ for all k by constraint (10), and $x_i^L \leq 1$ by (9) and (8). Then, $x_i^L = 1$ because the leader maximizes total captured demand.

3. The case is impossible because if $s \in S_i$ but $t \notin T_i$, customer prefers s to t, thus $s \in S_{it}$.

4. If $L_s = 0$ for all $s \in S_i$ and there is at least one $t \in T_i$ such that $F_{tk} = 1$, then $x_{ik}^F = 1$ by constraints (17) and (8).

5. If there is an $s \in S_{it}$ such that $L_s = 1$, for each $t \in T_i$ with $F_{tk} = 1$, then $y_{itk} = 0$ for such t by constraints (12) and (15). Then $x_{ik}^F = 0$ by (18). $x_i^L = 1$ as in case 2.

6. If there is a $t \in T_i$ such that $F_{tk} = 1$ but $L_s = 0$ for all $s \in S_{it}$ then $y_{it1} = 1$, and $y_{itk} = 0$ if $x_{ik'} = 1$ for some $k' < k$ by (14), or $y_{itk} = 1$ and $x_{i,k} = 1$ by (17)

Theorem 2. *Variables x_i^L, x_{ik}^F and y_{itk} assume binary values in the optimal solution.*

Proof. Constraints (11) and (12), and nonnegativity constraint (20) imply that $y_{it1} \in [\max\{0, 1 - \sum_{s \in S_{it}} L_s\}, 1 - \max_{s \in S_{it}}\{L_s\}]$. Thus, given binary L_s, each y_{it1} is assigned a binary value.

At iteration k', assume we have binary L_s, $F_{tk'}$ and $y_{itk'}$. If there is no $t \in T_i$ such that $F_{tk'} = 1$, $x_{ik'}^F = 0$ by constraint 10 and $y_{itk'+1} = y_{itk'}$. If there is at least one such t, then given binary $y_{itk'}$, constraints (17) and (18) imply that $x_{ik'}^F = y_{itk'} = 1$, thus binary. Therefore $y_{itk'+1}$ is also binary. Then, by induction, all x_{ik}^F and y_{itk} are binary.

If there is a k such that $x_{ik}^F = 1$, by constraint (8) $x_i^L = 0$. If there is no $s \in S_i$ such that $L_s = 1$, by constraint (9) $x_i^L = 0$. If neither of these are true, $x_i^L = 1$ as in case 5 of Theorem 1.

Theorem 3. *Any follower response enforced by the above model corresponds to a greedy add algorithm result.*

Proof. At each iteration k, the follower facility which can capture the largest remaining demand is opened. That is, $F_{tk} = 1 \Rightarrow t = \arg\max_{t'}\{\sum_{i \in I} d_i y_{itk}\}$. The converse is also true if the maximizer is unique. Otherwise, there would be some $t' \neq t$ such that $\sum_{i \in I} d_i y_{it'k} > \sum_{i \in I} d_i y_{itk}$. Then

$\sum_{i\in I} d_i x^F_{ik} > \sum_{i\in I} d_i y_{itk}$ by (16). However, if $F_{tk}=1$, by (17) and (18), $x^F_{ik}=y_{itk}$. These constraints are only for (i,t) pairs such that $t\in T_i$, yet constraints (6) and (10) imply that for all i such that $t\notin T_i$, $x^F_{ik}=0$. Therefore, $\sum_{i\in I} d_i x^F_{ik} = \sum_{i\in I} d_i y_{itk}$, which is a contradiction.

Also it follows that all of the demand capturable by t is captured and no noncapturable demand is captured. A total of K facilities are opened in this manner and the result is a greedy add algorithm solution.

Note that the greedy add algorithm may return different solutions and total captured demands for the follower if one does not use a stable selection procedure between the multiple maximizers of an iteration. In the model stated above, if there are multiple maximizers at a step of the greedy add algorithm, then the optimal solution selects the one that would result in the greatest objective value for the leader. To avoid this optimistic outcome, we can introduce a stable selection procedure for the multiple maximizers of an iteration. An example procedure could be defining a precedence order for the follower (e.g., index of the facility) such that the ties are resolved according to that order (e.g., select the one with the least index). We can enforce such a selection rule by introducing the following constraint.

$$\sum_{i\in I} d_i y_{itk} - \sum_{i\in I} d_i x^F_{ik} + \alpha_t F_{tk} \le \sum_{t'\in T} \alpha_{t'} F_{t'k} \quad \forall t\in T, k=1,\dots,K \tag{21}$$

where $\alpha_t \in (0, \min_i\{d_i\})$ and $\alpha_t > \alpha_{t'}$ if and only if t precedes t' on the selection rule. This way the ties would be solved in favor of the first alternative according to the precedence order.

The leader solves Problem 1. We refer to the following formulation as CMCLP1.

Problem 1 (Competitive MCLP with Heuristic Follower Response).

$$\max \quad \sum_{i\in I} d_i x^L_i \tag{4}$$

$$\text{s.t.} \quad \sum_{s\in S} L_s \le B \tag{5}$$

$$\sum_{t\in T} F_{tk} = 1 \quad \forall k=1,\dots,K \tag{6}$$

$$\sum_{k=1}^{K} F_{tk} \le 1 \quad \forall t\in T \tag{7}$$

$$x_i^L + \sum_{k=1}^{K} x_{ik}^F \leq 1 \quad \forall i \in I \quad (8)$$

$$x_i^L \leq \sum_{s \in S_i} L_s \quad \forall i \in I \quad (9)$$

$$x_{ik}^F \leq \sum_{t \in T_i} F_{tk} \quad \forall i \in I, k = 1, \ldots, K \quad (10)$$

$$1 - y_{it1} \leq \sum_{s \in S_{it}} L_s \quad \forall i \in I, t \in T_i \quad (11)$$

$$1 - y_{it1} \geq L_s \quad \forall i \in I, t \in T_i, s \in S_{it} \quad (12)$$

$$y_{itk} \geq y_{itk-1} - x_{ik-1}^F \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \quad (13)$$

$$y_{itk} \leq 1 - \sum_{k'=1}^{k-1} x_{ik'}^F \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \quad (14)$$

$$y_{itk} \leq y_{itk-1} \quad \forall i \in I, t \in T_i, k = 2, \ldots, K \quad (15)$$

$$\sum_{i \in I} d_i y_{it} \leq \sum_{i \in I} \sum_{k} d_i x_{ik}^F \quad \forall t \in T, k = 1, \ldots, K \quad (16)$$

$$x_{ik}^F \geq F_{tk} + y_{itk} - 1 \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \quad (17)$$

$$x_{ik}^F \leq 1 - F_{tk} + y_{itk} \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \quad (18)$$

$$L, F \in \{0, 1\} \quad (19)$$

$$x^L, x^F, y \geq 0 \quad (20)$$

4.4. Alternative Formulations

CMCLP1 has the following two features that we might alter to obtain better formulations. First, it does not recognize which follower facility captures the customer at a given iteration. Second, it does not assume transitivity in customer preference. It has a follower point of view in defining the capturability, which we can define alternatively from the leader's point of view using transitivity.

4.4.1. CMCLP2: Identifying The Capturing Follower Facility In CMCLP1, there is no variable that identifies which follower facility captures a given customer at a given iteration. This information is retrievable using x_{ik}^F and F_{tk} once the problem is solved. We introduce the variable x_{itk}^F, which equals 1 if customer i is captured by facility t at iteration k. This alters the previous formulation slightly. For constraints, (8), (13), (14) and (16), we simply replace x^F_{ik} with $\sum_{t\in T_i} x^F_{itk}$ because only one facility can capture a given customer in a given iteration.

On the other hand, the new variable allows us to define stronger relationships between y, F and x^F. Facility t captures a customer only if it is open and the customer is capturable in the given iteration.

$$x^F_{itk} \geq F_{tk} + y_{itk} - 1 \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \tag{22}$$

If any of these conditions are not satisfied, the customer cannot be captured by facility t.

$$x^F_{itk} \leq F_{tk} \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \tag{23}$$

$$x^F_{itk} \leq y_{itk} \quad \forall i \in I, t \in T_i, k = 1, \ldots, K \tag{24}$$

Therefore, we can replace constraints (10), (17) and (18) with (23), (22) and (24), respectively. Finally, we can obtain an improvement in constraint (15), which can be altered as $y_{itk} \leq y_{itk-1} - x^F_{itk-1}$. Introducing the new index increases the number of variables and constraints, but the tighter constraints lead to improved performance; see Section 6 for more details.

4.4.2. CMCLP3: Capturability From Leader's Point Of View CMCLP1 takes the follower's point of view in formulating the capturability of a customer—variable y. It relies on the set S_{it} that is defined for each (i,t) pair. Note that this set definition does not assume transitive customer preference among members of $S \cup T$, even though consumer preferences are generally transitive. If we assume transitivity, we can use an alternative and tighter formulation. Therefore, let $\succeq_i$ be a strict total preference relation (complete, reflexive, transitive and antisymmetric) over the choice set $S \cup T$. Let ties be broken (e.g. in favor of the smaller-indexed facility) before we apply the preference relation. We introduce a new variable a_{isk} which equals 1 if customer i patronizes leader facility s before iteration k—in a sense the game is played in $1+K$ stages, which represent the leader's move and the follower's K consecutive moves, respectively. Then we make the following changes in CMCLP1.

At the beginning of each iteration k , customer i is either not covered by any of the opened facilities, or patronizes a covering leader facility or has been captured by the follower at some earlier iteration.

$$\sum_{s\in S_i} a_{isk} + \sum_{k'=1}^{k-1} x^F_{ik'} \leq 1 \quad \forall i \in I, k = 1, \ldots, K \tag{25}$$

At the leader's decision stage, a customer can only patronize an open leader facility.

$$a_{is1} \leq L_s \quad \forall i \in I, s \in S_i \tag{26}$$

When leader facility s is opened the following two conditions hold. A covered customer cannot be assigned to a less preferred facility (27), but it can be assigned to s or a more preferred facility (28).

$$L_s \leq 1 - \sum_{s'\in S_i : s \succ_i s'} a_{is'1} \quad \forall i \in I, s \in S_i \tag{27}$$

$$L_s \leq \sum_{s'\in S_i : s' \succeq_i s} a_{is'1} \quad \forall i \in I, s \in S_i \tag{28}$$

A customer is considered capturable by a given follower facility t if it is not currently held by a better leader facility and has not already been captured by the follower. At each step of the greedy add algorithm, the left-hand side of the inequality is the total demand capturable by each follower facility and the right-hand side equals the demand captured by the follower in that iteration. This guarantees that the demand captured at each iteration is the largest capturable demand by a single open facility.

$$\sum_{i\in I : t\in T_i} d_i \left(1 - \sum_{s\in S_i : s\succ_i t} a_{isk} - \sum_{k'=1}^{k-1} x^F_{ik'}\right) \leq \sum_{i\in I} d_i x^F_{ik} \quad \forall t \in T, k = 1, \ldots, K \tag{29}$$

Then, we have the following three counterparts of the previous model ensuring that (29) holds at equality for the opened facility. At a given iteration k, if customer i is captured, a covering follower facility t should be open.

$$x^F_{ik} \leq \sum_{t\in T_i} F_{tk} \quad \forall i \in I, k = 1, \ldots, K \tag{30}$$

Thus if customer i is capturable by t, and t or a better follower facility is open, i is captured (31). However, if it is not capturable by t, yet the opened follower facility is t or worse, then it cannot be captured (32). (This can be incorporated in the same way into CMCLP1 (but not to CMCLP2) by replacing F_{tk} in (17) and (18) with $\sum_{t'\in T_i:t'\succeq_i t} F_{t'k}$ and $\sum_{t'\in T_i:t\succeq_i t'} F_{t'k}$, respectively.)

$$x^F_{ik} \geq \sum_{t'\in T_i:t'\succeq_i t} F_{t'k} - \sum_{s\in S_i:s\succ_i t} a_{isk} - \sum_{k'=1}^{k-1} x^F_{ik'} \quad \forall i\in I, t\in T_i, k=1,\dots,K \tag{31}$$

$$x^F_{ik} \leq 2 - \sum_{t'\in T_i:t\succeq_i t'} F_{t'k} - \sum_{s\in S_i:s\succ_i t} a_{isk} - \sum_{k'=1}^{k-1} x^F_{ik'} \quad \forall i\in I, t\in T_i, k=1,\dots,K \tag{32}$$

The assignment status of customer i to a particular leader facility s, after the initial assignment in (27)–(28) is updated in one of the following three ways. If i is not assigned to (held by) s at a given iteration, it would not be in the subsequent iterations (33). If the leader was holding i and a better follower facility was not opened, the leader continues to hold it (34).

$$a_{isk} \leq a_{isk-1} \quad \forall i\in I, s\in S_i, k=2,\dots,K \tag{33}$$

$$a_{isk} \geq a_{isk-1} - \sum_{t\in T_i:t\succ s} F_{tk-1} \quad \forall i\in I, s\in S_i, k=2,\dots,K \tag{34}$$

The last assignment status is converted to the leader's updated covering relation as follows. If the leader does not hold the customer before the last iteration (35), or the follower has captured it (36), the leader cannot capture it. However, if both are false, the leader captures it (37).

$$x^L_i \leq \sum_{s\in S_i} a_{isK} \quad \forall i\in I \tag{35}$$

$$x^L_i \leq 1 - \sum_{k=1}^{K} x^F_{ik} \quad \forall i\in I \tag{36}$$

$$x^L_i \geq \sum_{s\in S_i} a_{isK} - x^F_{iK} \quad \forall i\in I \tag{37}$$

As a final note, $L, F\in\{0,1\}$ whereas $a, x\in[0,1]$.

5. Theoretical Performance

How accurate is the greedy add algorithm as an approximation of the follower's optimal response, and what impact does the inaccuracy have on the leader's decision? Let L and F be the leader's and follower's location vectors, respectively, and let $LC(L,F)$ and $FC(L,F)$ be the demand captured by the leader and the follower, respectively, for given values of L and F. Let $F^*(L)$ be the follower's optimal response to the leader locations L, and let $F^G(L)$ be the follower's response if he uses the greedy add algorithm. Finally, let L^* be the leader's optimal solution if she assumes that the follower acts optimally, and let L^G be the leader's optimal solution if she assumes that the follower uses the greedy add algorithm. We start with an ideal, and extreme, case, in which the model solves the bilevel program optimally.

PROPOSITION 1. *If the customers covered by the follower facilities do not overlap,* $F^*(L) = F^G(L)$, $L^* = L^G$.

Proof. In this case, the follower's coverage network (e.g., Figure 1) is a collection of stars which have a follower facility as the internal node and customers covered by that facility as the leaves. The leader's decision can remove edges from this graph, but cannot add edges to it. Opening a follower facility does not affect the stars associated with other facilities. Since there is no overlap, summing the demand captured by individual facilities gives the objective. The objective is maximized by selecting the best K facilities in order, which is equivalent to the greedy add algorithm. Since the greedy add algorithm identifies the follower's best response, the model solves the bilevel program optimally.

This result is tight in the sense that the algorithm can fail to find the optimal solution even when each facility shares at most one covered customer, as Figure 3 shows.

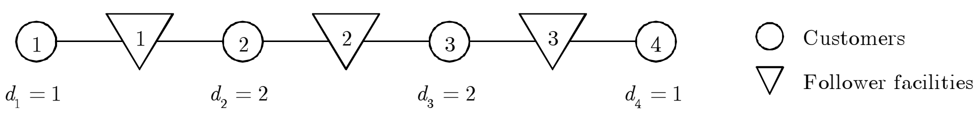


Figure 3 Greedy algorithm fails: $F^*(L) = [1,0,1], F^G(L) = [1,1,0]$.

The following theorem states an upper bound on the worst-case performance of the greedy add algorithm as a heuristic for solving the follower's problem. It is a straightforward application of a result whose proof can be found in (Hochbaum 1997).

Theorem 4. *If the follower opens K facilities, the following worst-case bound holds:*

$$\frac{FC(L,F^*(L)) - FC(L,F^G(L))}{FC(L,F^*(L))} \leq \left(1 - \frac{1}{K}\right)^K. \tag{38}$$

Moreover, $\lim_{K\to\infty}\left(1-\frac{1}{K}\right)^K = \frac{1}{e} \approx 0.37$, *and it approaches from below.*

Theorem 5. *The worst-case bound in Theorem 4 is tight.*

Proof. Worst-case examples can be generated by manipulating the the worst-case examples in (Cornuejols et al. 1977). One such rule is as follows. Given K, let $|I| = K^2$, $|S| = K$, $|T| = 2K-1$ and $B = K-1$. Let each leader facility s cover customer $i = K(K-1)+s$, respectively. Let each follower facility $t < K$ cover customers $i \in \{(t-1)K+1, ..., (t-1)K+K\}$, and each $t \geq K$ cover customers $i \in \{t+1-K, t+1, t+1+K, ..., t+1+(K-2)K\}$. Assume the follower's facilities are all better than the leader's. Finally, let the customers' demand be $d_i = K^{K-2}(\frac{K-1}{K})^{\lceil \frac{i}{K} \rceil - 1}$ for $i \leq K(K-1)$ and $d_i = (K-1)^{K-1} - \epsilon$ otherwise, where ϵ is an arbitrarily small positive number. An illustration is presented in Figure 4.

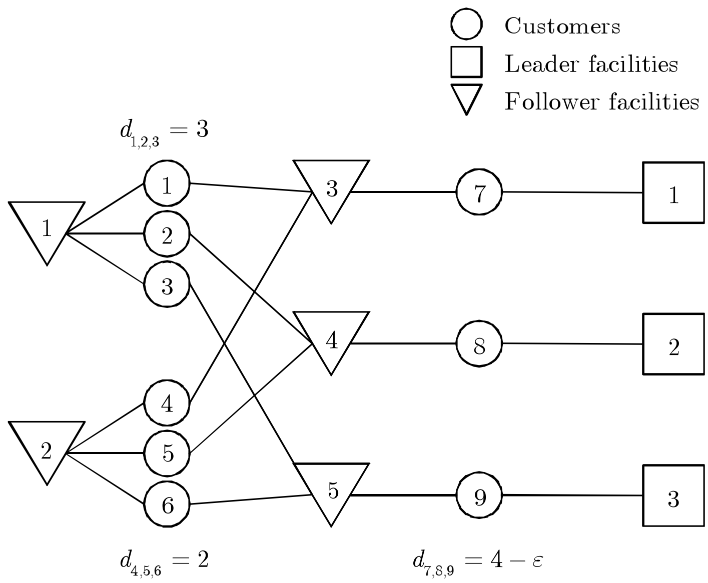


Figure 4 Worst case of greedy add algorithm, $K = 3$.

The greedy add algorithm would select locations $t \in \{1, \dots, K\}$ for each possible leader decision. However, the optimal follower decision is to open $t \in \{K, ..., 2K-1\}$ and capture all customer demand. This would give the following expressions for greedy and optimal objectives:

$$\begin{aligned} FC(L, F^G(L)) &= \sum_{i=1}^{K(K-1)} K^{K-2} (\frac{K-1}{K})^{\lceil \frac{i}{K} \rceil - 1} + (K-1)^{K-1} - \epsilon \\ &= \sum_{n=0}^{K-2} K^{K-1} (\frac{K-1}{K})^n + (K-1)^{K-1} - \epsilon \\ &= \sum_{n=0}^{K-1} K^{K-1-n} (K-1)^n - \epsilon = K^K - (K-1)^K - \epsilon \end{aligned}$$

$$\begin{aligned} FC(L, F^*(L)) &= FC(L, F^G(L)) + \sum_{i=K(K-1)+2}^{K^2} \left[(K-1)^{K-1} - \epsilon \right] \\ &= K^K - (K-1)^K + (K-1)^K - K\epsilon = K^K - K\epsilon \end{aligned}$$

Therefore, as $\epsilon \to 0$,

$$\frac{FC(L, F^*(L)) - FC(L, F^G(L))}{FC(L, F^*(L))} \to \frac{(K-1)^K}{K^K} = \left(1 - \frac{1}{K}\right)^K \tag{39}$$

We can use this result to assess the worst-case performance of the greedy assumption from the leader's perspective. Our first such result is the following bound on the leader's loss as a percentage of demand that the follower greedily captures.

THEOREM 6. *By assuming a greedy follower response and selecting facilities L, the leader loses at most $\frac{1}{1-\alpha} FC(L, F^G(L))$, where $\alpha = \left(1 - \frac{1}{K}\right)^K$.*

Proof. Let A be the subset of customers that the leader covers by selecting L, and B [C] be the subset of customers that the follower captures if he responds greedily [optimally]. See Figure 5 for an illustration. Therefore, the leader expects to capture $A \setminus B$, but ends up capturing $A \setminus C$. This corresponds to losing $(A \cap C) \setminus B$ and gaining $(A \cap B) \setminus C$. The net loss is maximized if all [none] of the customers that the optimizing [greedy] follower captures are stolen from the leader; that is, if $(A \cap C) \setminus B = C$ and $(A \cap B) \setminus C = \emptyset$. The total demand of set C is $FC(L, F^*(L))$. Then, from Theorem 4, we obtain $FC(L, F^*(L)) < \frac{1}{1-\alpha} FC(L, F^G(L))$.

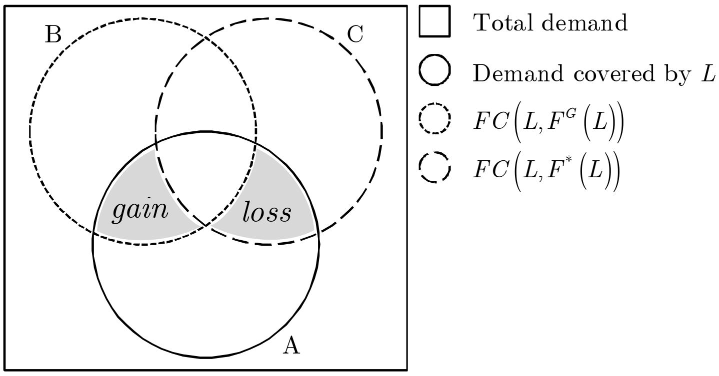


Figure 5 Set diagram for game outcome.

On the other hand, no fixed worst-case bound exists for the leader's percentage error when she assumes a greedy response from the follower but the follower responds optimally. Similarly, there is no such bound for the leader's percentage error between the heuristic model solution and the optimal solution to the problem.

Theorem 7. *No fixed worst-case bounds exist for:*

(a) $\frac{LC(L,F^G(L))-LC(L,F^*(L))}{LC(L,F^*(L))}$, *and*

(b) $\frac{LC(L^G,F^G(L^G))-LC(L^*,F^*(L^*))}{LC(L^*,F^*(L))}$

Proof. For case (a), consider the worst-case example in Theorem 4 with $K=3$, and let $L=[0,1,1]$. The greedy response of the follower is $F^G(L)=[1,1,1,0,0]$ and it yields $LC(L,F^G(L))=8$. If the follower responded optimally, he would pick $F^*(L)=[0,0,1,1,1]$ and it would yield $LC(L,F^*(L))=0$ since the follower facilities are better than the leader facilities for customers 7, 8 and 9. Thus, the leader expects to capture a positive demand whereas she cannot capture any demand in the end. This is illustrated in Figure 6.

$$\frac{LC(L,F^G(L))-LC(L,F^*(L))}{LC(L,F^*(L))}=\infty.$$

For case (b), consider the same example, but now introduce two new leader locations $s=4,5$, and two new customers $i=10,11$ as in Figure 6. Let $S_{10}=4, S_{11}=5, T_{10}=T_{11}=\emptyset$, and each customer have a demand of $\gamma<4-\epsilon$. The solution is $L^G=[0,1,1,0,0]$, $F^G(L^G)=[1,1,1,0,0]$, $L^*=$

$[0,0,0,1,1]$ and $F^*(L^*) = [0,0,1,1,1]$. These yield $LC(L^G, F^G(L^G)) = 8 - 2\epsilon$ and $LC(L^*, F^*(L^*)) = 2\gamma$. When γ approaches 0:

$$\frac{LC(L^G, F^G(L^G)) - LC(L^*, F^*(L^*))}{LC(L^*, F^*(L))} \to \infty.$$

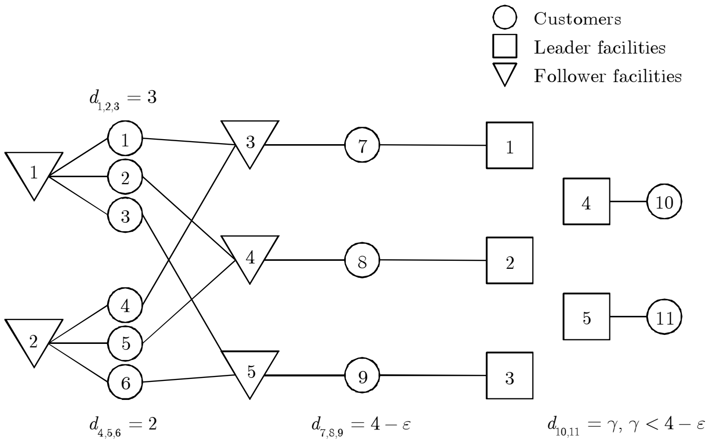


Figure 6 Worst case of greedy add algorithm, $K = 3$**.**

These results are for the limited-coverage case. However, they still hold if we assume an infinite coverage radius, too. The follower always solves a limited-coverage problem, because the follower's coverage is updated (and becomes limited) once the leader makes his decision. The same worst-case examples can be generated for the limited coverage case by carefully constructing the preference relations.

6. Computational Study

6.1. Experimental Design

We tested our deterministic model using two data sets from (Daskin 1995). The first set includes the geographical coordinates and populations of 88 cities. These are the 50 largest US cities according to the 1990 census and the 48 continental state capitals, less the double entries. The second set includes the geographical coordinates of the 150 largest US cities according to the 1990 census. The first set represent the customers, thus $|I| = 88$, and we used their populations as customer demands. We randomly selected the potential facilities from the second set. We selected 11 settings for the number of potential facilities, changing $|S|$ and $|T|$. We further decided on 20 problem types by changing the model size parameters (K, B) and radius (r). The radius (given in miles) is the key to generating the coverage and preference relationships. Each location is assumed to have a coverage radius given by r, and each customer prefers facilities in order of distance. For each problem type, we generated 50 instances by randomly picking potential leader and follower locations. We discarded duplicate locations, thus each instance consists of distinct locations for the leader and follower. In total, we solved 1000 instances. The problem types and their properties are summarized in Table 2.

Table 2 Instance settings.

Type	$\|I\|$	$\|S\|$	$\|T\|$	K	B	r
T1	88	10	10	4	4	300
T2	88	10	20	4	4	300
T3	88	20	10	4	4	200
T4	88	20	10	4	4	300
T5	88	20	10	4	4	400
T6	88	20	10	4	4	500
T7	88	20	10	1	5	300
T8	88	20	10	2	5	300
T9	88	20	10	3	5	300
T10	88	20	10	4	5	300
T11	88	20	10	5	5	300
T12	88	20	20	4	4	300
T13	88	20	20	4	4	400
T14	88	20	30	4	4	300
T15	88	30	10	4	4	300
T16	88	30	20	4	4	300
T17	88	30	30	4	4	300
T18	88	40	10	4	4	300
T19	88	40	20	4	4	300
T20	88	40	20	4	8	300

6.2. Solution Statistics

We coded the model using AMPL and solved the instances using CPLEX 12.2 on a Pentium Xeon 3.0 GHz (x2) 64 bit computer. We used a 1 hour (3600 second) time limit per problem. All of the instances were solved within the time limit. The statistics are tabulated in Tables 3 and 4. The leftmost column indicates the instance type. Under a given instance type, rows 1-3 correspond to the results for formulations CMCLP1-3, respectively. The values are the aggregated results from

Table 3 Solution Statistics: Comparison of Formulations.

		Solution Time			Simplex Iterations			Node Evaluations			%LPGap		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
T1	1	0.41	0.04	2.10	728	0	5817	15	0	175	16.4	0.0	42.7
	2	0.39	0.04	2.79	686	0	8723	10	0	145	15.5	0.0	40.2
	3	0.39	0.04	1.44	778	0	5038	13	0	128	14.8	0.0	40.2
T2	1	2.94	0.19	38.58	4245	53	54582	90	0	908	38.6	11.9	77.3
	2	1.82	0.15	11.98	2681	35	19697	28	0	158	37.2	11.7	75.1
	3	1.96	0.24	9.24	2234	79	11724	42	0	208	35.0	7.2	70.8
T3	1	0.21	0.02	0.98	335	0	2106	6	0	58	9.6	0.0	32.9
	2	0.22	0.02	0.98	316	0	1898	4	0	27	8.5	0.0	32.9
	3	0.27	0.04	0.94	571	0	3461	14	0	116	7.8	0.0	32.9
T4	1	2.11	0.13	11.83	4589	254	25430	98	0	411	14.9	1.2	39.8
	2	1.70	0.18	7.20	3757	120	25484	51	0	245	14.3	1.2	36.0
	3	2.06	0.17	8.84	5522	175	35199	88	0	395	14.0	1.2	35.7
T5	1	17.15	0.25	103.52	30708	365	164064	369	0	1672	18.9	7.4	43.8
	2	11.23	0.28	66.61	20083	357	109892	153	0	820	18.5	7.4	43.4
	3	16.81	1.34	116.48	33436	2076	197455	288	5	1249	18.3	7.4	41.3
T6	1	70.90	4.86	349.41	99338	12259	544882	750	139	2264	20.7	7.3	40.2
	2	45.61	6.72	268.24	62933	8514	373206	318	68	1351	20.4	7.3	40.1
	3	62.64	8.74	316.26	97645	10874	502173	541	82	1901	20.1	7.3	40.1
T7	1	0.11	0.01	0.51	151	16	593	2	0	32	3.3	0.0	24.6
	2	0.08	0.01	0.42	144	0	748	2	0	31	2.9	0.0	20.7
	3	0.11	0.02	0.61	325	30	2183	8	0	92	3.3	0.0	23.1
T8	1	0.28	0.02	0.92	818	36	4416	26	0	153	7.6	0.0	17.6
	2	0.27	0.03	1.20	618	0	3350	14	0	90	6.9	0.0	17.5
	3	0.44	0.03	1.93	1276	0	4672	37	0	167	6.7	0.0	16.4
T9	1	0.74	0.06	3.75	2211	109	15442	61	0	539	10.9	0.2	36.9
	2	0.67	0.06	2.57	1595	0	9383	27	0	146	10.3	0.2	35.9
	3	0.94	0.08	3.04	2849	111	11079	57	0	213	9.9	0.2	33.0
T10	1	2.62	0.16	27.45	6888	170	70071	148	0	1086	12.8	3.8	33.5
	2	1.76	0.14	9.81	4595	100	41966	70	0	627	12.3	3.8	33.2
	3	2.28	0.15	7.92	6507	238	29673	121	0	562	12.0	3.7	33.2

50 instances of each instance type. We report the CPU time used to solve the problem (Solution Time), number of simplex iterations (Simplex Iterations), and number of branch and bound nodes evaluated (Node Evaluations). We compare the LP relaxation at the root node and the optimal MIP solution and report the percent integrality gap (%LPGap) calculated as $\frac{(z_{LP}-z_{MIP})}{z_{MIP}}$. For each of these performance measures, we report the minimum, maximum and average of the results.

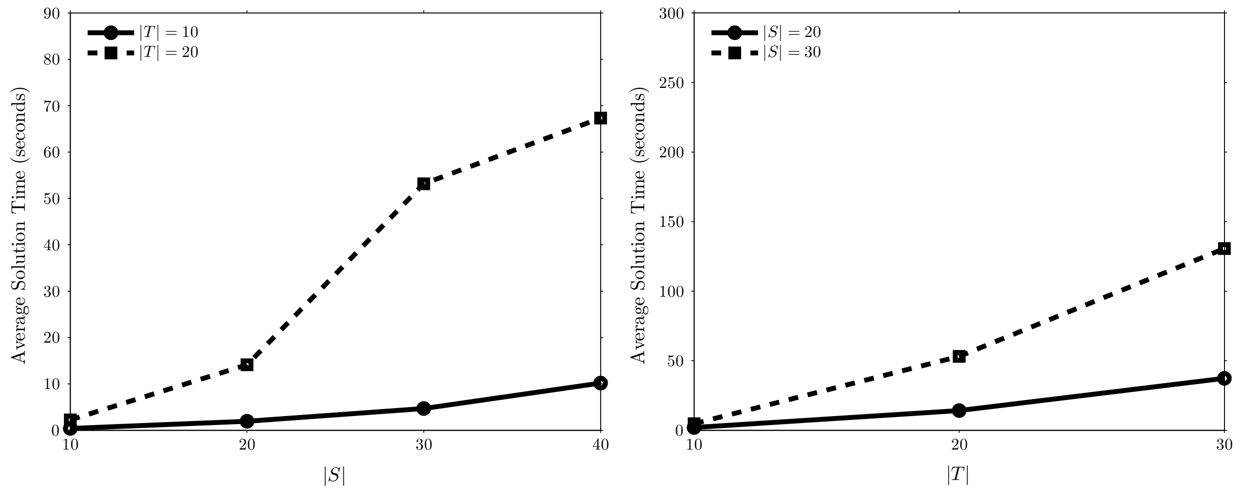
We isolate the effects of the problem parameters and illustrate them in Figures 7 and 8 below. Comparing T1, T4, T15, T18 with T2, T12, T16, T19, and T4, T12, T14 with T15, T16, T17; we see that an increase in the number of potential leader and follower facilities increases the average solution time. Their cross comparison shows that at higher levels of the other parameter, this increase is magnified. These comparisons are plotted in Figure 7. If we compare T7-T11 we notice

Table 4 Solution Statistics: Comparison of Formulations (continued).

		Solution Time			Simplex Iterations			Node Evaluations			%LPGap		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
T11	1	5.52	0.34	39.28	11986	473	71108	218	0	922	14.0	3.4	31.1
	2	4.03	0.41	18.12	9045	601	37972	116	0	474	13.6	3.4	30.5
	3	5.00	0.32	31.61	12758	666	78364	169	0	856	13.3	3.4	30.5
T12	1	18.90	0.73	82.16	31605	784	134944	539	3	2146	29.7	6.7	78.7
	2	11.95	0.57	55.30	21889	849	118361	217	0	1327	29.0	6.7	78.7
	3	11.52	1.15	61.31	24085	1654	131407	347	8	2591	28.5	6.7	78.6
T13	1	201.72	12.33	706.81	212994	11668	663059	1789	180	5911	32.6	12.5	68.2
	2	128.74	12.87	501.36	144339	10525	643126	682	57	3570	31.9	12.5	66.8
	3	117.99	9.78	431.41	175111	16026	648300	1255	156	3670	31.9	12.5	67.4
T14	1	50.32	1.42	284.75	63508	2475	386579	848	9	4019	39.6	10.4	74.0
	2	37.24	1.99	210.87	51574	3678	254825	368	27	1445	39.0	10.3	73.6
	3	24.54	1.46	119.57	41508	1880	150910	585	16	2168	38.0	10.3	68.3
T15	1	4.62	0.13	54.46	11152	211	129107	199	0	1489	13.3	3.2	27.8
	2	3.59	0.14	47.80	8963	221	108045	113	0	905	12.9	2.8	27.8
	3	5.88	0.40	45.21	15202	815	112504	205	0	1141	12.7	0.1	27.8
T16	1	69.10	0.94	640.31	100937	1408	596622	1354	31	8284	26.9	11.4	61.5
	2	50.90	1.01	557.96	81098	1540	655455	660	32	3361	26.3	11.4	59.9
	3	39.43	1.43	123.70	74529	2299	216763	909	60	3277	25.8	11.4	59.6
T17	1	160.04	10.32	725.09	180034	11939	655482	1913	150	6987	30.5	11.4	73.3
	2	139.64	8.83	522.29	150563	10405	528393	948	74	2979	30.1	11.4	72.0
	3	92.12	6.91	283.64	132119	8455	395635	1385	97	5053	29.6	11.3	72.7
T18	1	8.61	0.30	41.63	22097	475	85834	413	0	1432	13.6	3.2	40.0
	2	6.35	0.43	32.43	16636	600	91117	239	0	1494	13.3	3.2	38.2
	3	15.60	0.82	63.65	35319	982	140768	452	7	1729	13.2	3.2	35.7
T19	1	71.52	0.52	250.49	112918	754	420088	1449	0	6047	23.8	1.0	53.6
	2	61.64	0.53	217.65	101998	657	374626	842	0	3090	23.4	1.0	53.6
	3	68.78	0.61	194.02	112543	920	357041	1181	0	3689	23.1	1.0	53.6
T20	1	206.50	5.45	946.11	373380	13710	1694477	4717	265	31368	14.2	3.1	25.6
	2	168.39	4.46	1012.34	316145	10061	2032369	2868	129	17995	14.0	3.0	25.6
	3	250.67	15.20	1268.38	489369	21078	2691435	5578	339	35269	13.9	3.0	25.6

also that the number of follower facilities to open affects the solution effort. An increase in the number of facilities multiplies the size of the instance and the solution time. Another important factor is the coverage radius of the facilities. Although not a model parameter, it is the main factor in determining the coverage relationship. Increasing the radius increases the number of covering facilities, thus supply alternatives, for each customer. This generates a denser graph and the problem size increases, as comparing T3-T6 shows. The effects of the number of follower facilities and the coverage radius are plotted in Figure 8.

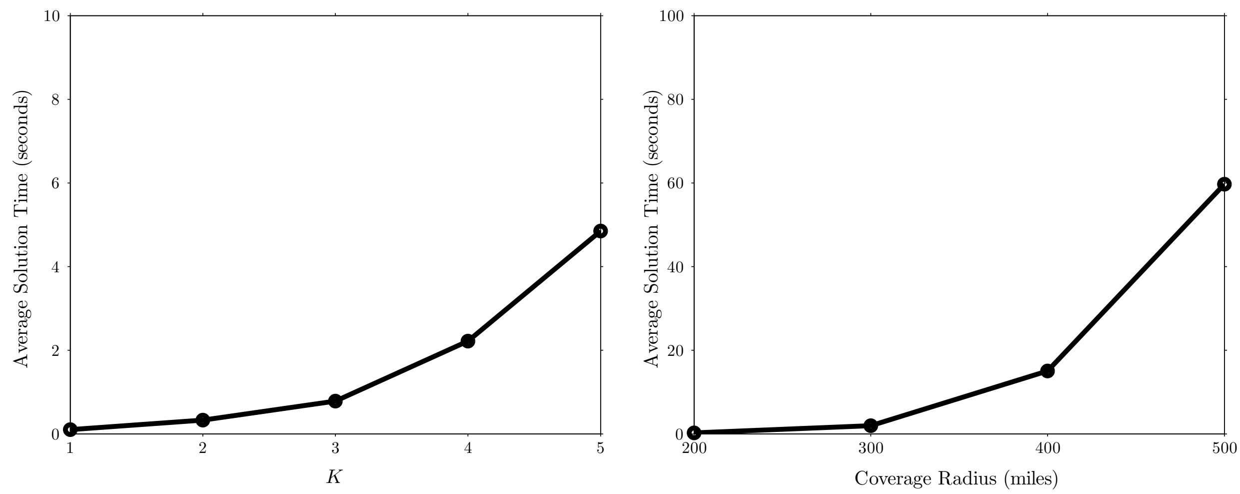
One final factor is the number of leader facilities to open (i.e. budget). An increase in the budget does not increase the size of the instance but increases the number of possible facility combinations for the leader, increasing the feasible solution space. A comparison between T19 and



(a) Effect of $|S|$ (b) Effect of $|T|$

Figure 7 Effect of number of potential locations on solution times.

T20 demonstrates this effect. When the budget is increased from 4 to 8, average solution time increases 3.1 times.

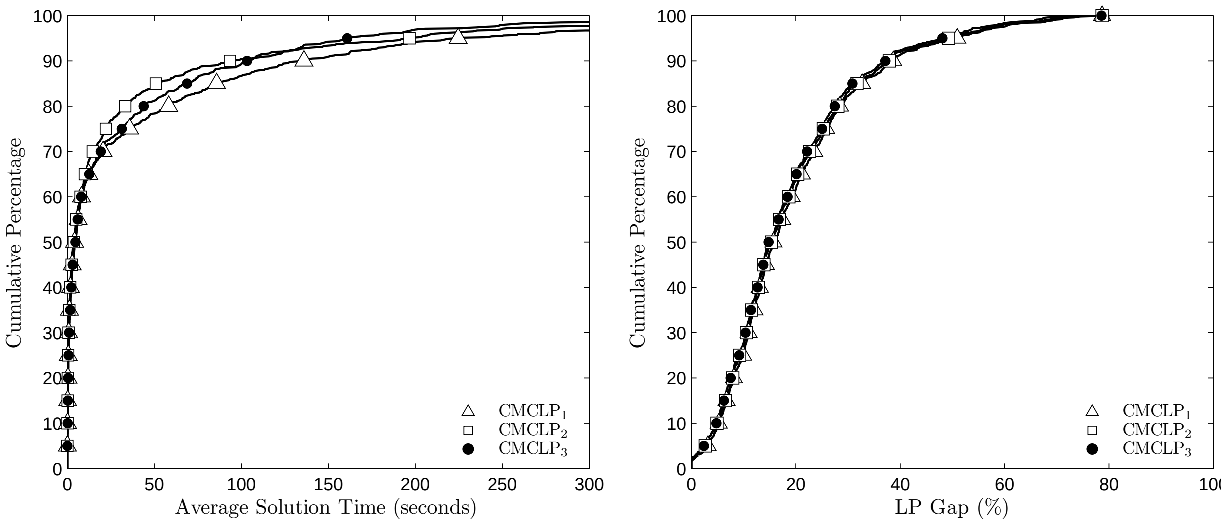


(a) Effect of K (b) Effect of coverage radius

Figure 8 Effects of the number of greedy iterations and coverage radius on solution times.

The same comparisons, when made for the %LPGap column, reveal different results. Increasing the number of follower facilities to open or the coverage radius increases the gap. However, when we increase the number of potential leader facilities and the investment budget, we see that the gap decreases. Comparing T1, T4, T15 and T18; and T2, T12, T16 and T19, we see that the LP gap decreases as $|S|$ increases. Similarly, comparing T4 and T10, and T19 and T20, we see that an increase in budget decreases the gap. These are also expected as the former would make it possible to have more good locations and the latter would make it possible to select more facilities from the same set, keeping the opponent's response capacity the same. The results for the number of follower facilities have a opposite effect, as gap increases with increasing number of facilities.

The tables and the accompanying figures above provide a comparison among instance types, which represent different difficulty levels. The cumulative histograms below illustrate the solution performance of each formulation over all instances. Figure 9a shows that around 75% of the instances were solved in under 30 seconds. The performance of the three formulations are close but begin to differ as more difficult instances appear. Then, CMCLP2 and 3 dominate CMCLP1. Interestingly, CMCLP2 performs the best for instances that were solved under 2 minutes, after that CMCLP3 takes over. Figure 9b indicates that about 65% of the instances had an LP gap of less than 20% and about 2% of the instances had no integrality gap. It also shows that the tightness of the formulations increase from CMCLP1 to 3, which is observable on Tables 3 and 4 as well.



(a) Solution times (b) LP Gap

Figure 9 Comparison of three formulations.

6.3. Solution Quality

Replacing the follower's optimization problem with the greedy add algorithm leaves us with likely suboptimal solutions for the leader's problem. In general, the model is expected to perform well if greedy add algorithm performs well for the restructured MCLP problem of the follower. By restructuring, we refer to the loss of coverage over previously covered customers as some customers would prefer open leader facilities. We would like to see how meaningful the solutions of the model are. To do so, upon solving Problem 1, we fix the leader's open facilities. Then we restructure the follower's coverage relations and solve optimally for the follower's problem alone. This yields the *best response* of the follower to the decision that the leader makes solving Problem 1. The best response is important because we expect that this would be the follower's true response, and the customers that the follower serves in the solution of this problem would be the ones that the follower actually would serve. Finally, we find the customers that are covered by the leader's open facilities but are not served by the follower in the subsequent solution and obtain the expected captured demand for the leader. We then compare this expected result with the suggested result of Problem 1. Note that if the suggested and expected capture for each player are the same, than the model has found the overall optimal solution, even to the problem with the exact optimal follower response.

In Table 5 we summarize the results for each of the 20 instance types. Table 5 is organized as follows. Avg |Diff| (%) shows the average difference (in absolute value) between the suggested and expected objectives and Max |Diff| (%) shows the maximum of these differences, both as a percentage of the suggested objective. %Opt shows in what percent of the results the model yielded the overall optimal solution. The experiments show that for these instances the model performed quite well, finding the overall optimal solution at least 74% of the time and 93.4% o the average. The maximum percent difference was 15.84% whereas on average the percent difference was 0.28%. In light of Table 2, we see that the performance deteriorates with increasing numbers of leader and follower facilities, coverage radius, and number of greedy iterations. Note that when $K = 1$, finding

Table 5 Solution quality statistics.

Type	Avg |Diff| (%)	Max |Diff| (%)	%Opt
T1	0.19	9.38	98
T2	0.55	14.0	92
T3	0.00	0.0	100
T4	0.01	0.30	98
T5	0.15	5.19	94
T6	0.44	6.53	88
T7	0.00	0.0	100
T8	0.00	0.0	100
T9	0.01	0.28	98
T10	0.13	4.19	94
T11	0.23	6.11	92
T12	0.10	5.24	98
T13	1.56	13.80	74
T14	0.76	15.84	92
T15	0.13	6.37	98
T16	0.22	7.43	94
T17	0.25	8.36	94
T18	0.21	3.83	88
T19	0.46	7.33	86
T20	0.28	5.88	90
Summary	0.28	15.84	93.4

the overall optimum is certain. The performance, on the other hand, is enhanced by an increasing investment budget of the leader. Figure 10 shows the histogram of the optimality errors arising from the heuristic models. Note that all give the same result since the generated instances have all transitive consumer choices (shortest distance). The error is the absolute value of the difference between the optimal objective value suggested by the model, which is the captured demand assuming greedy follower response, and the expected result after follower problem is optimally solved, given the same leader decision.

7. Conclusions

In this paper we introduced a mathematical model in order to devise a strategy for the leader firm in a leader–follower version of the maximal covering problem. Our primary contribution is to embed the follower's response into the constraints of the leader's optimization problem so that the leader's problem is reduced from a bilevel program to a single-level one. This keeps the leader's problem *tractable* and the problem size at a *manageable* level. Our model of the follower employs the greedy add algorithm rather than solving his problem exactly, but we demonstrate that this assumption comes at the expense of very little loss of accuracy.

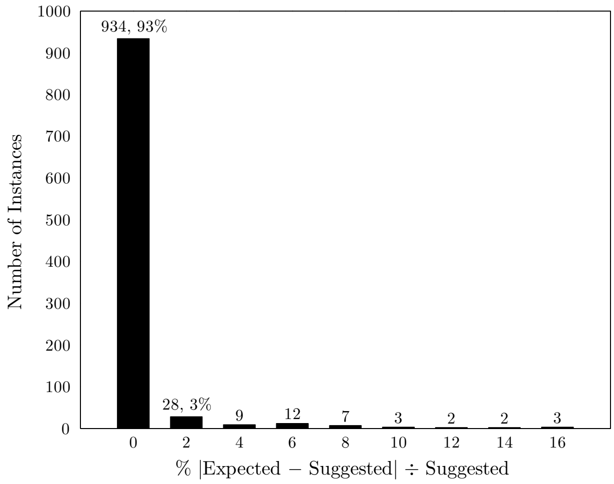


Figure 10 Histogram of absolute optimality errors.

Our computational study demonstrates that the model is able to:

- Generate near optimal (often optimal) solutions for the leader
- Provide a reasonable assessment of the follower's response that can be embedded into the leader's optimization model.

Our numerical studies also indicated that the model size and the solution effort increase quickly with increasing problem components—potential sites, customers and number facilities to open. One interesting avenue for future research would be the introduction of demand uncertainty, but the resulting scenarios would further increase the computational burden. Therefore investigation of solution methods that do not require MIP solvers is in our future research agenda.

Finally, the problem was stated with the most common features for competitive location problems. Variants of the problem with different patronizing rules and competition over potential locations (i.e. from a single set of potential locations) as well as customers would be important to study. It would also be worthwhile to study the applicability of other simple algorithms as proxies for the follower's response.

Acknowledgments

This research is supported in part by NSF grant CMMI-0726822.

References

Anderson, S. P., A. de Palma, J.-F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*, chap. 8. MIT Press, 273–342.

Bard, J. F., J. T. Moore. 1992. An algorithm for the discrete bilevel programming problem. *Naval Research Logistics* **39** 419–435.

Bard, Jonathan F. 2009. Bilevel programming in management. C. A. Floudas, P. M. Pardalos, eds., *Encyclopedia of Optimization*. Springer US, 269–274.

Church, Richard, Charles S. ReVelle. 1974. The maximal covering location problem. *Papers in Regional Science* **32**(1) 101–118.

Colson, Benoit, Patrice Marcotte, Gilles Savard. 2007. An overview of bilevel optimization. *Annals of Operations Research* **153** 235–256.

Cornuejols, G., M. L. Fisher, G. L. Nemhauser. 1977. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science* **23**(8) 789.

Daskin, Mark S. 1995. *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley and Sons, New York, NY.

d'Aspremont, C., J. J. Gabszewicz, J.-F. Thisse. 1979. On hotelling's "stability in competition". *Econometrica* **47**(5) 1145–1150.

DeNegre, S. T., T. K. Ralphs. 2009. A branch-and-cut algorithm for integer bilevel linear programs. J.W. Chinneck et al., ed., *Operations Research and Cyber-Infrastructure*. Springer, 65–78.

Drezner, Zvi, ed. 1995. *Facility Location: A Survey of Applications and Methods*. Springer-Verlag, New York, NY.

Drezner, Zvi, Horst W. Hamacher, eds. 2002. *Facility Location: Applications and Theory*. Springer-Verlag, New York, NY.

Eiselt, H. A. 2011. Equilibria in competitive location models. H. A. Eiselt, V. Marianov, eds., *Foundations of Location Analysis*, chap. 7. Springer US, 139–162.

Eiselt, Horst A., Gilbert Laporte. 1996. Sequential location problems. *European Journal of Operational Research* **96**(2) 217–231.

Eiselt, Horst A., Gilbert Laporte, Jacques-Franois Thisse. 1993. Competitive location models: A framework and bibliography. *Transportation Science* **27**(1) 44–54.

Gabszewicz, J. J., J.-F. Thisse. 1992. Location. R. Aumann, S. Hart, eds., *Handbook of Game Theory with Economic Applications*, vol. 1, chap. 9. Elsevier, 281–304.

Hakimi, S. Louis. 1983. On locating new facilities in a competitive environment. *European Journal of Operational Research* **12**(1) 29–35.

Hochbaum, D. S. 1997. Approximating covering and packing problems: Set cover, vertex cover, independent set, and related problems. D. S. Hochbaum, ed., *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company, 94–143.

Hotelling, Harold. 1929. Stability in competition. *The Economic Journal* **39**(153) 41–57.

Jeroslow, Robert G. 1985. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming* **32**(2) 146–164.

Klose, Andreas, Andreas Drexl. 2005. Facility location models for distribution system design. *European Journal of Operational Research* **162**(1) 4–29.

Kress, Dominik, Erwin Pesch. 2012. Sequential competitive location on networks. *European Journal of Operational Research* **217**(3) 483–499.

Kuehn, Alfred A., Michael J. Hamburger. 1963. A heuristic program for locating warehouses. *Management Science* **9**(4) 643–666.

Megiddo, Nimrod, Eitan Zemel, S. Louis Hakimi. 1983. The maximum coverage location problem. *SIAM Journal on Algebraic and Discrete Methods* **4**(2) 253–261.

Moore, J. T., J. F. Bard. 1990. The mixed integer linear bilevel programming problem. *Operations Research* **38**(5) 911–921.

Owen, Susan H., Mark S. Daskin. 1998. Strategic facility location: A review. *European Journal of Operational Research* **111**(3) 423–447.

Plastria, Frank. 2001. Static competitive facility location: An overview of optimisation approaches. *European Journal of Operational Research* **129**(3) 461–470.

Plastria, Frank, Lieselot Vanhaverbeke. 2008. Discrete models for competitive location with foresight. *Computers & Operations Research* **35**(3) 683–700.

ReVelle, Charles S. 1986. The maximum capture or "sphere of influence" location problem: Hotelling revisited on a network. *Journal of Regional Science* **26**(2) 343–358.

ReVelle, Charles S., Horst A. Eiselt. 2005. Location analysis: A synthesis and survey. *European Journal of Operational Research* **165**(1) 1–19.

ReVelle, Charles S., Horst A. Eiselt, Mark S. Daskin. 2008. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research* **184**(3) 817–848.

Serra, Daniel, Samuel Ratick, Charles S. ReVelle. 1996. The maximum capture problem with uncertainty. *Environment and Planning B: Planning and Design* **23**(1) 49–59.

Serra, Daniel, Charles S. ReVelle. 1994. Market capture by two competitors: The preemptive location problem. *Journal of Regional Science* **34**(4) 549–561.

Serra, Daniel, Charles S. ReVelle. 1995. Competitive location in discrete space. Zvi Drezner, ed., *Facility Location: A Survey of Applications and Methods*. Springer-Verlag, New York, NY, 367–386.

Snyder, L. V. 2011. Covering problems. H. A. Eiselt, V. Marianov, eds., *Foundations of Location Analysis*, chap. 6. Springer US, 109–135.

Snyder, Lawrence V. 2006. Facility location under uncertainty: a review. *IIE Transactions* **38**(7) 547–564.

Snyder, Lawrence V. 2010. Introduction to facility location. James J. Cochran, Louis A. Cox, Pinar Keskinocak, Jeffrey P. Kharoufeh, J. Cole Smith, eds., *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc.

Younies, H., H. A. Eiselt. 2011. Sequential location models. H. A. Eiselt, V. Marianov, eds., *Foundations of Location Analysis*, chap. 8. Springer US, 163–178.