Decision Support

Hub location under competition

Ali Irfan Mahmutogullari, Bahar Y. Kara*

Department of Industrial Engineering, Bilkent University, 06800, Ankara, Turkey

ABSTRACT

Hubs are consolidation and dissemination points in many-to-many flow networks. Hub location problem is to locate hubs among available nodes and allocate non-hub nodes to these hubs. The mainstream hub location studies focus on optimal decisions of one decision-maker with respect to some objective(s) even though the markets that benefit hubbing are oligopolies. Therefore, in this paper, we propose a competitive hub location problem where the market is assumed to be a duopoly. Two decision-makers (or firms) sequentially decide locations of their hubs and then customers choose one firm with respect to provided service levels. Each decision-maker aims to maximize his/her own market share. We propose two problems for the leader (former decision-maker) and follower (latter decision-maker): $(r|X_p)hub - medianoid$ and $(r|p)hub - centroid$ problems, respectively. Both problems are proven to be NP-complete. Linear programming models are presented for these problems as well as exact solution algorithms for the $(r|p)hub - centroid$ problem. The performance of models and algorithms are tested by computational analysis conducted on CAB and TR data sets.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

1. Introduction

Hubs are consolidation and dissemination points in many-to-many flow networks. Consolidation generates economies of scale and thus, unit transportation cost is discounted between hubs. Hubbing also reduces number of required links to ensure that each flow is routed to its desired destination. Hub networks are used in many applications in airline, cargo and telecom industries.

The hub location problem is to determine location of hubs and allocation of non-hub nodes to these hubs with respect to a given allocation structure and objective(s). A single decision-maker can determine locations of hubs depending on problem parameters such as amount of flow and unit transportation cost between each pair of nodes, interhub transportation discount factor, allocation strategy (single- or multiple-allocation), and structure of hub network (incomplete, star network etc.). In single-allocation case, whole flow originating from and destined to a node is routed via a unique hub. On the other hand, in multiple-allocation case, different hubs can be used to route different flows with same origin. However, in a competitive environment, a decision-maker should also consider decisions of his/her rivals and customer preferences. In this study, we consider a duopolistic market -a special case of oligopoly- where there are two operating firms. The decision-maker who makes the initial location decision is called *leader* and the other one is called as *follower*. We assume a multiple-allocation structure.

Then by combining retail location from marketing, spatial competition in economics and location theory in operations research, in this paper, we propose a discrete Stackelberg hub location problem where firms make sequential decisions. Each decision-maker (or firm) decides location of hubs and allocation of non-hub nodes to maximize their market share.

2. Literature review

2.1. Competitive location literature

The pioneering study of competition in economics is due to Cournot (1838). He studied a market operated by two competing firms where each firm decides amount of production of a single product. Later, Bertrand (1883) considered a duopoly model where the competitors decide price of a single product. Hotelling (1929) presented the first competitive model that includes location decisions. He investigated location and price decisions of two ice cream vendors operating on a beach in which each customer prefers the vendor that offers lower price.

In duopoly models presented by Cournot (1838), Bertrand (1883) and Hotelling (1929), decisions of two firms are made simultaneously. Another streamline of research in competitive models deals with sequential decision making. The preliminary work of sequential decision making of location was first proposed in a book by Von Stackelberg (1951). Since sequential decisions result in an asymmetry

* Corresponding author. Tel.: +905336164923.
E-mail addresses: a.mahmutogullari@bilkent.edu.tr (A.I. Mahmutogullari), bkara@bilkent.edu.tr (B.Y. Kara).

between decision makers, we need to differentiate identities of the decision makers. Von Stackelberg (1951) studied a duopoly where the firm that makes the initial decision is called as *leader* and other one is called as *follower*. He suggested three major assumptions:

- Decisions are permanent.
- Decisions are made sequentially.
- The leader and the follower have full and complete knowledge about the system.

If leader's decisions are given, follower makes decisions with respect to his/her own objective. These decisions are called *reaction function* of the follower. Since both parties have complete information of the system, the leader observes reaction function of the follower and hence the leader makes his/her decisions based on this reaction function. These leader–follower situations can be modeled as bilevel optimization problems. Bilevel optimization models consider the follower's reaction function an input to the leader's decisions. Bard (1999) and Dempe (2002) gave detailed discussion on bilevel programming models and solution techniques.

Teitz and Bart (1968) studied sequential location problem on a line segment. Moreover, Teitz and Bart (1968) considered an extension of Hotelling's model by allowing each decision maker to locate more than one facility.

Drezner (1982) and Hakimi (1983) independently proposed sequential location problems with an OR point of view and attracted the community's attention. They both studied same competitive model but with different spaces. While Drezner (1982) considered locations on a plane, Hakimi (1983) dealt with network models. Their problem includes a number of customers with *inelastic demand*, that is, amount of demand of each customer is known *a priori* and is not affected by the decisions of leader and follower. A customer prefers the closest facility to buy a homogenous product. The decision-makers act sequentially, that is, the leader locates p facilities and then the follower locates r facilities.

In order to describe contributions of Drezner (1982) and Hakimi (1983), the following conventions are necessary. Assume that n customers (or demand points) are located on points $V = \{v_1, v_2, \ldots, v_n\}$ with the demand of customer i being $w(v_i)$. For any subset of points $Z \subseteq V$, let $D(v, Z) = min\{d(v, z) : z \in Z\}$ where $d(v, z)$ is the distance between v and z. The distance between two points is the Euclidean distance in a two-dimensional plane and the shortest path on a network. Assume that the leader's and follower's facilities are located on the set of points $X_p = \{x_1, x_2, \ldots, x_p\}$ and $Y_r = \{y_1, y_2, \ldots, y_r\}$, respectively. A customer v_i prefers the follower if and only if $D(v_i, Y_r) < D(v_i, X_p)$. Then, total demand captured by the follower can be expressed as $W(Y_r | X_p) = \sum_{i:D(v_i, Y_r) < D(v_i, X_p)} w(v_i)$.

Assume that the leader has already been operating with facilities located on X_p. Then, $(r|X_p)$ *medianoid* is the set Y_r^* such that $W(Y_r^* | X_p) \geq W(Y_r | X_p)$ for all sets of follower's possible facility locations Y_r. $(r|X_p)$ *medianoid* is the optimal set of facility locations for the follower to capture the highest market share given the set X_p.

Similarly, $(r|p)$ *centroid* is the set X_p^* such that $W(Y_r^*(X_p^*)|X_p^*) \leq W(Y_r^*(X_p)|X_p)$ for all sets of the leader's possible set of facility locations X_p where $Y_r^*(X_p)$ is the $(r|X_p)$ *medianoid* given X_p. $(r|p)$ *centroid* is the optimal set of facility locations for the leader to capture the highest market share under realistic assumption that the follower will respond by $(r|X_p)$ *medianoid*. Hakimi (1983) proved that both centroid and medianoid problems are NP-hard.

An interested reader may refer to surveys by Eiselt and Laporte (1997) and Dasci (2011) for a detailed discussion on competitive location problems.

2.2. Hub location literature

O'Kelly (1986a, 1986b) presented the hub location problem as system-wide transportation cost of a network is minimized by locating p hubs in a single-allocation structure (This problem is later referred to as the single-allocation p-hub median problem). Later, O'Kelly (1987) also proposed the first mathematical formulation of single-allocation p-hub median problem.

Later, Skorin-Kapov, Skorin-Kapov, and O'Kelly (1996) provided a new linear model for the problem and Ernst and Krishnamoorthy (1996) modeled the single allocation p-hub median problem as a multi-commodity flow problem. The single allocation p-hub median problem was proven to be NP-hard by Kara (1999).

Multiple-allocation p-hub median problem has also attracted attention. Campbell (1992) presented first multiple-allocation hub model. Skorin-Kapov et al. (1996) developed a linear model for the problem. Ernst and Krishnamoorthy (1998) modeled the multiple-allocation p-hub median problem based on the idea that they use for the single-allocation version.

Although hub location problem with median objective constitutes the main streamline of the literature, other objectives were also investigated by researchers. O'Kelly (1992), Campbell (1992) and Aykin (1994) proposed mathematical models for the hub location problem with fixed costs.

In some applications of hub networks, such as the cargo applications, service levels are considered as well as cost. The p-hub center problem is to locate p hub on a network where the distance or travelling cost between the most disadvantageous pair of nodes is minimized. Campbell (1994) proposed linear models for the hub location problems with center-type objectives. Kara and Tansel (2000) proved that p-hub center problem is NP-hard. They also proposed different mathematical models for the problem. Later, Ernst, Hamacher, Jiang, Krishnamoorthy, and Woeginger (2009) provided a new formulation for the p-hub center problem based on the value of maximum collection/distribution distance between a hub and a non-hub node.

Hub covering is another version of the hub location problem. Campbell (1994) presented mathematical models for different types of hub covering problem. After his contribution, Kara and Tansel (2003) studied single allocation hub set covering problem and proposed three different linearizations of the problem. Later, Ernst, Jiang, and Krishnamoorthy (2005) provided new formulations of the problem based on the idea that they use for the p-hub center problem.

Various extensions of hub location problems were also considered such as latest arrival problems (Kara & Tansel, 2001; Yaman, Kara, & Tansel, 2007), hub location with stopovers (Kuby & Gray, 1993; Yaman et al., 2007), hierarchical hub network models (Yaman, 2009) and hub location problems with ordered averaging objective functions (Puerto, Ramos, & Rodríguez-Chía, 2011; 2013).

An interested reader may refer to surveys by Campbell, Ernst, and Krishnamoorthy (2002), Alumur and Kara (2008), Kara and Taner (2011) and Campbell and O'Kelly (2012) for detailed discussion of hub location problems.

2.3. Hub location with competition

Although competition in location decisions has been studied in detail, competitive hub location studies in the literature are rare. Marianov, Serra, and ReVelle (1999) proposed first hub location problem with competition. They proposed mathematical models for follower's problem where the leader had already been operating the market with his/her existing hubs.

They also considered proportional capture levels in addition to all-or-nothing type capture. They assumed that the follower would capture half of the flow between nodes i and j if his/her service level is between $0.9C_{ij}$ and $1.1C_{ij}$, three–fourth of the flow if his/her service level is between $0.7C_{ij}$ and $0.9C_{ij}$ and the whole flow if his/her service level is less than $0.7C_{ij}$ where C_{ij} is the service level of the leader. A mathematical model was provided for proportional capture case by triplicating the capture variables and constraints. Later

Table 1
Competitive hub location literature

Paper	Decision space	Market type	Decisions	Capture type	Computational study	Solution techniques	Contributions
Marianov et al. (1999)	Network	Duopoly	Follower's hubs	Partial Discrete	Random AP(20,25,40,45)	Heuristic	Competitive hub location model
Wagner (2008)	Network	Duopoly	Follower's hubs	Partial Discrete	AP(50)	MIP	First exact solution for moderate size instances
Sasaki and Fukushima (2001)	Plane	Oligopoly	One hub for leader and each follower	Partial Continuous	CAB(25)	SQP	First Stackelberg competition in hub literature
Sasaki (2005)	Network	Duopoly	Leader's and follower's hubs	Partial Continuous	CAB(25)	Enumeration Heuristic	Application of Sasaki and Fukushima (2001) to a network
Eiselt and Marianov (2009)	Network	Oligopoly	Follower's hubs	Partial Continuous	AP(25)	Heuristic	Various customer preferences considered
Sasaki et al. (2014)	Network	Duopoly	Leader and follower's hubs	Partial Continuous	CAB(25)	Enumeration	Bilevel model, exact solutions for small size instances
Lüer-Villagra and Marianov (2013)	Network	Duopoly	Follower's hubs and prices	Partial Continuous	CAB(25)	Heuristic	Pricing decisions, Different network topologies
This study	Network	Duopoly	Leader and follower's hubs	Binary Discrete	CAB(25) TR(81)	MIP Enumeration	Formal definition, complexity results, exact solutions for large size instances

Wagner (2008) proposed a new capture set where the follower gets nothing in case of equal service levels for the same problem.

Sasaki and Fukushima (2001) proposed a new kind of competitive hub location model where the decision space is a plane. Route between any O–D pair on the plane visits only one hub. First, a big firm locates one hub, and then several medium size firms locate their hubs. There is no competition between medium size firms. They used logit functions for customer preferences to express proportional capture. They initially modeled the problem as a bilevel program and then use a sequential quadratic programming approach.

Sasaki (2005) applied the same idea in the study by Sasaki and Fukushima (2001) to a discrete environment with some modifications. Her model includes two decision-makers: one leader and one follower. The leader and follower locate p and q hubs on the network, respectively. Capture rule is as in Sasaki and Fukushima (2001) and each route contains one only hub.

Eiselt and Marianov (2009) proposed another hub location model with competition where an airline transportation company enters a market. It is assumed that some other companies has already been operating in the market. The entrant firm aims to capture as many customers as possible. Customer preferences are based on basic attractiveness of the firms (such as safety record, personal space, quality of the foods etc.), number of stopover on the trip, cost of the route and time required by the flight.

Another hub location problem with Stackelberg competition was studied by Sasaki, Campbell, Ernst, and Krishnamoorthy (2014). In their problem environment, the decision-makers do not locate hubs but hub arcs. One leader and one competitor airline companies locate q^a and q^b hub arcs on the network to maximize the total revenue. The leader can capture 0, 25, 50, 75 or 100% of flow between any O–D pair based on cost and travel time of the trip and the remaining customers prefer the follower. They proposed a bilevel program for the problem.

A study by Lüer-Villagra and Marianov (2013) considered both hub location and pricing decisions of an entrant firm where an other firm has already been operating on the market. They propose a nonlinear model where the objective is to maximize the entrant's profit. The customer preferences are represented as a logit function.

Although existing studies contribute to hub location and competition literature, both theoretical aspect of the problem and applications in industry require more effort. In order to motivate the studies in this area, in this paper, we formally define *hub-medianoid* and *hub-centroid* problems by following the terminology used by Hakimi (1983) for analogous competitive location problems. Moreover, we prove that both problems are NP-complete. Table 1 summarizes studies in the competitive hub location literature where the last row corresponds to this paper.

3. Problem definition

Given a network $G = (N, E)$ where N is set of nodes and E is set of edges, let w_{ij} be total flow and c_{ij} be transportation cost of a unit flow from node i to node j for all $i, j \in N$. Interhub transportation cost is discounted by the factor α, $0 \le \alpha \le 1$. (Later we use $< G = (N, E), w_{ij}, c_{ij}, \alpha >$ to refer this network.) The leader and follower would like to enter a market with a prespecified number of hubs. Let p and r be number of hubs to be opened by the leader and follower, respectively. We assume that both p and r are greater than or equal to 2 since otherwise economies of scale is not generated. Let $H \subseteq N$ be the subset of nodes that are available to locate hubs. Customers prefer the leader or follower with respect to provided service levels. Service level is defined as the cost of routing the flow from a node to its destination via hubs. A customer prefers the follower if the service level provided by the follower is strictly better than the one provided by the leader, otherwise the demand is captured by the leader. Ties are broken in the advantage of the leader in case of equal service levels since the customer was already operating with the leader when the follower entered the market and the customer has no incentive to deviate from current position.

First, assume that the leader has already been operating the market with hubs located at a subset of nodes $X_p = \{x_1, x_2, \ldots, x_p\}$, $X_p \subseteq H$. The flow from node $i \in N$ to node $j \in N$ visits one or two hubs. Therefore, we can easily compute service level, say β_{ij}, provided by the leader for the flow from node $i \in N$ to node $j \in N$.

$$\beta_{ij} = \min_{k, m \in X_p} \{c_{ik} + \alpha c_{km} + c_{mj}\} \quad \forall i, \quad j \in N. \tag{1}$$

Now, consider the follower enters the market by opening hubs on subset of nodes $Y_r = \{y_1, y_2, \ldots, y_r\}$, $Y_r \subseteq H$. Similarly, followers service levels, say γ_{ij}, for all node pairs i and j can be calculated as:

$$\gamma_{ij} = \min_{k, m \in Y_r} \{c_{ik} + \alpha c_{km} + c_{mj}\} \quad \forall i, j \in N. \tag{2}$$

Flow w_{ij} is captured by the follower if $\gamma_{ij} < \beta_{ij}$ for all $i, j \in N$. Given that the leader's and follower's hubs are located on the subset of nodes X_p and Y_r, respectively, total flow captured by the follower can be expressed by a mapping $f : \mathcal{P}_p(H) \times \mathcal{P}_r(H) \to [0, W]$ such that

$$f(X_p, Y_r) = \sum_{i, j \in N : \gamma_{ij} < \beta_{ij}} w_{ij} \tag{3}$$

where $\mathcal{P}_p(H)$ is the collection of subsets of H with cardinality p and W is the total flow over the network, that is, $W = \sum_{i, j \in N} w_{ij}$.

Given X_p, the follower aims to maximize $f(X_p, Y_r)$ over all $Y_r \in \mathcal{P}_p(H)$. We define set Y_r^* as $(r|X_p) hub - medianoid$ if $f(X_p, Y_r^*) \ge$

$f(X_p, Y_r)$, $\forall Y_r \in \mathcal{P}_r(H)$. Therefore, $(r|X_p)hub - medianoid$ is the set of follower's hubs with cardinality r that maximizes captured demand given X_p.

Now, we look at the problem from the leader's perspective. The leader wants to minimize the demand captured by the follower (or equivalently maximize demand captured by himself/herself) while deciding his/her hub set. The leader also has the information that the follower will respond rationally.

We define set X_p^* as $(r|p)hub - centroid$ if $f(X_p^*, Y_r^*(X_p^*)) \leq f(X_p, Y_r^*(X_p))$, $\forall X_p \in \mathcal{P}_p(H)$ where $Y_r^*(X_p)$ is the $(r|X_p)hub - medianoid$ given X_p. $(r|p)hub - centroid$ is the set of leader's hubs with cardinality p so that in the remaining scenario the follower can capture the least possible flow.

4. $(r|X_p)$ hub-medianoid

Let $< G = (N, E), w_{ij}, c_{ij}, \alpha >$ be a many-to-many flow network. At the time the follower makes decision, the leader has already located his/her hubs on the set $X_p \subseteq H$ and locations of these hubs are observed by the follower. Then, the follower has the information of the service levels provided by the leader for each pair of nodes $i, j \in N$. These service levels can be calculated as in Eq. (1).

4.1. Complexity of $(r|X_p)$ hub-medianoid

We prove that the problem of finding a $(r|X_p)hub - medianoid$ is NP-complete by using polynomial time reduction from *clique* problem, a known NP-complete problem due to Karp (1972).

Decision version of clique problem: Given an undirected graph $G = (N, E)$ and an integer r, determine if G has a r-clique, that is, there is a set of vertices K with $|K| \geq r$ such that for each pair of vertices in K there is an edge in E between them.

Theorem 1. $(r|X_p)hub - medianoid$ is NP-complete even if $\alpha = 0$.

Proof. Given an instance of *clique* problem, we construct a network $G' = (N', E')$ where $N' = N \cup X_p$, where $X_p = \{x_1, x_2, \ldots, x_p\}$ and $E' = E \cup \{(i, j) : i \in N \text{ and } j \in X_p\}$ where X_p is assumed to be the hub set of the leader. Let $c_{ij} = 1$ if $(i, j) \in E$ and $c_{ij} = 0.5$ if $i \in N$ and $j \in X_p$ and let $\alpha = 0$. The flow values for all pairs $i, j \in N$ are set to 1. Clearly, $\beta = 1$ for all $i, j \in N$.

We prove the theorem by showing that there exists a set of r points $Y_r(X_p)$ on G' such that $f(X_p, Y_r(X_p)) \geq \binom{r}{2} = (r^2 - r)/2$ if and only if there exists an $r - clique$ on G.

Assume that *clique* problem has solution $K \subseteq N$ and $|K| \geq r$. By letting $Y_r \supseteq K$, we can observe that $\gamma_{ij} = 0$ for all $i, j \in K$ since all flows on the clique benefit discounting where $\alpha = 0$ and the total flow among the clique is captured by the follower, that is, $f(X_p, Y_r(X_p)) \geq (r^2 - r)/2$.

On the other hand, suppose Y_r in G' is such that $f(X_p, Y_r(X_p)) \geq (r^2 - r)/2$. If for all $i, j \in Y_r$ there exists an edge $(i, j) \in E$, then Y_r itself form an $r - clique$ on G. Then set $K = Y_r$. Otherwise, assume that Y_r does not form an $r - clique$, then there must be $(r^2 - r)/2$ units of flow captured by the follower and at least one unit of flow should be routed via a spoke link. Equivalently, we can say that for $(r^2 - r)/2$ pairs of node $\gamma_{ij} < 1$. Then, none of the captured flow is routed via the spoke link of the follower which contradicts the assumption.

Hence, we conclude that $(r|X_p)hub - medianoid$ is reducible from *clique* problem in polynomial time. So, it is NP-complete. □

4.2. Mathematical model for $(r|X_p)$ hub-medianoid problem

To provide a mathematical model for the $(r|X_p)hub - medianoid$ problem, we define following decision variables:

$$h_k = \begin{cases} 1, & \text{if the follower locates a hub on node } k \in H \\ 0, & \text{otherwise} \end{cases}$$

$$u_{ijk} = \begin{cases} 1, & \text{if flow from node } i \in N \text{ to node } j \in N \text{ visits hub} \\ & k \in H \text{ as the first hub} \\ 0, & \text{otherwise} \end{cases}$$

$$v_{ijm} = \begin{cases} 1, & \text{if flow from node } i \in N \text{ to node } j \in N \\ & \text{visits hub } m \in H \\ & \text{as the second hub and this flow is captured by} \\ & \text{the follower} \\ 0, & \text{otherwise} \end{cases}$$

The following mixed integer problem, namely **H-MED**, solves the $(r|X_p)hub - medianoid$ problem:

$$\text{maximize} \sum_{i \in N} \sum_{j \in N} \sum_{m \in H} w_{ij} v_{ijm} \tag{4}$$

$$\text{subject to} \sum_{k \in H} h_k = r, \tag{5}$$

$$\sum_{k \in H} u_{ijk} = 1 \ \forall i, j \in N, \tag{6}$$

$$\sum_{m \in H} v_{ijm} \leq 1 \ \forall i, j \in N, \tag{7}$$

$$u_{ijk} \leq h_k \ \forall i, j \in N \ \forall k \in H, \tag{8}$$

$$v_{ijm} \leq h_m \ \forall i, j \in N \ \forall m \in H, \tag{9}$$

$$\sum_{k \in H} u_{ijk}(c_{ik} + \alpha c_{km}) + c_{mj} - \beta_{ij}$$
$$+ \epsilon \leq (1 - v_{ijm})M \ \forall i, j \in N \ \forall m \in H, \tag{10}$$

$$h_k, u_{ijk}, v_{ijk} \in \{0, 1\} \ \forall i, j \in N \ \forall k, m \in H \tag{11}$$

The objective (4) maximizes amount of flow captured by the follower. Constraint (5) ensures that follower locates r hubs on the set of available nodes. Constraints (6) guarantee that each flow is allocated to a first hub. Constraints (7) state flow from node $i \in N$ to node $j \in N$ can be captured by the follower using a hub located at node $m \in H$. Constraints (8) and (9) prevent allocation of flows to non-hub nodes. Constraints (10) determine captured flows in the following manner: LHS of the constraint is the difference of service level provided by the follower and service level provided by the leader plus ϵ. Let $\epsilon = 10^{-6}$ be a very small positive number used to break ties in favor of the leader. If this value is non-negative, the corresponding variable v_{ijm} is forced to be 0 that is the follower cannot provide a better service level for flow from $i \in N$ to node $j \in N$; otherwise there is no restriction on v_{ijm}. M is a large positive value but $M = (2 + \alpha) \max_{i,j \in N} c_{ij}$ value is large enough since the LHS can be at most $(2 + \alpha) max_{i,j \in N} c_{ij}$. Since the objective is to maximize the captured flow, corresponding v_{ijm} value is assigned to 1 when there is no restriction on v_{ijm}. Constraints (11) are domain Constraints.

If flow from $i \in N$ to node $j \in N$ visits only one hub $k \in H$, then $u_{ijk} = 1$. Additionally, if this flow is captured by the follower, corresponding variable v_{ijk} is set to 1 and 0 otherwise.

4.3. Computational analysis of $(r|X_p)$ hub-medianoid problem

Performance of **H-MED** is investigated by computational experiments conducted on two different data sets: CAB and TR. α values are chosen as either 0.6 or 0.8. Also, for TR data α is set to 0.9 due to Tan and Kara (2007). Nodes in the CAB data set are numbered based on the alphabetical order of the city names whereas nodes in the TR data sets are plate codes of cities in Turkey which ranges from 1 to 81. All instances are solved with CPLEX 12.4 (ILOG, 2012) and a 4 x AMD

Table 2
Summary of experiment instances $(r|X_p)hub - medianoid$ problem

Data set	CAB	TR
Hub set of the leader	**UMApHM** and **UMApHC**	**UMApHM**
p	2,3,4 and 5	6,8,10,12 and 14
r	2,3,4 and 5	6,8,10,12 and 14
α	0.6 and 0.8	0.6,0.8 and 0.9

Opteron Interlagos 16C 6282SE 2.6G 16M 6400MT computer running under a Linux operating system. Table 2 summarizes all 139 instances used in the computational study of $(r|X_p)hub - medianoid$ problem.

Since we need to take β_{ij} values as parameters of $(r|X_p)hub - medianoid$ problem, we have to make some assumptions for the leaders hub set in advance. Therefore, we assume that the leader locates his/her hubs on a set of nodes according to his/her optimal choices of well-studied multiple-allocation hub location problems: uncapacitated multiple-allocation p-hub median (**UMApHM**) and p-hub center (**UMApHC**). However, current models in the literature are not able to solve the **UMApHC** for the size of TR data set, so only **UMApHM** solutions are used as leaders hub set for this data set.

The distance matrices of both data sets are symmetric. Therefore, if from node $i \in N$ to node $j \in N$ is routed via the leader's (follower's) hubs then flow from node j to node i is also routed via the leader's (follower's) hubs. By using this fact, the constraints (6)–(11) of **H-MED** are imposed for only $i < j$ and the objective (4) is replaced with $\sum_{i<j:i,j\in N} \sum_{m\in H} (w_{ij} + w_{ji})v_{ijm}$ for computational studies.

Tables 3 and 4 summarize CPU time and market share of the follower in an optimal solution of $(r|X_p)hub - medianoid$ problem

where the leader has already located his/her hubs on the optimal solution of **UMApHM** and **UMApHC** on CAB and TR data sets, respectively.

Since the leader chooses his/her hub locations without being aware of competition, the follower can capture high amounts of flow even $p = r$. For example, if $p = r = 2$ the follower can capture more than 65% of total demand.

The proposed mathematical model **H-MED** can be regarded as the formulation of maximal hub cover problem so that covering radius for each pair of nodes $i, j \in N$ are defined as $\beta_{ij} - \epsilon$ where ϵ is a small positive real number. Having this property, CPLEX efficiently solves **H-MED** within reasonable times. All instances of CAB data set could be optimally solvable within 25 s.

For TR data set in case of equal number of hubs, that is $p = r$, the leader captures more than half of the market. The follower should open at least two more hubs to defeat the leader. Moreover, since the same discount factor applies for both firms, there is no important correlation between market shares and α value.

As p value gets closer to $|H|$, the leader can capture at least one half of the market even for the case $r > p$. As seen in TR instances, the leader locates his/her hubs on strategic locations and prevents good choices for the follower when $p \geq 10$. Then, for these instances, the follower is not able to capture one half of the market. Hence, if p is not a small value compared to $|H|$, the leader uses the advantageous of being the first mover. Therefore, the firms have incentive of competing to be the leader. For example, even if the leader chooses his/her hubs according to an optimal solution of **UMApHM** for $\alpha = 0.6$, $p = 10$ and $r = 14$, he/she can capture more flow than follower even without considering competition. However, in CAB instances, p is relatively small compared to $|H|$, so after the leader makes his/her decision, the follower still has big action space and being the latter decision-maker

Table 3
$(r|X_p)hub - medianoid$ experiment results for CAB data set given the leader's hubs are **UMApHM** or **UMApHC**

α	Leader's hubs	p	r	Follower's capture	CPU sec	α	Leader's hubs	p	r	Follower's capture	CPU sec
0.6	UMApHM	2	2	65.62%	6.15	0.6	UMApHC	2	2	75.86%	2.86
			3	78.25%	5.59				3	85.10%	4.32
			4	87.08%	7.95				4	90.98%	4.74
			5	92.26%	12.49				5	94.74%	3.69
		3	2	30.49%	11.16			3	2	51.81%	6.45
			3	45.13%	9.05				3	70.25%	4.63
			4	53.69%	14.15				4	79.08%	15.68
			5	62.02%	10.97				5	85.23%	18.30
		4	2	17.91%	23.44			4	2	36.56%	21.72
			3	28.39%	17.93				3	47.39%	22.67
			4	37.73%	20.79				4	57.38%	20.94
			5	46.18%	24.61				5	66.93%	19.91
		5	2	18.64%	11.28			5	2	45.62%	6.75
			3	28.14%	9.32				3	57.27%	13.39
			4	35.04%	12.63				4	69.34%	16.08
			5	42.32%	10.55				5	76.75%	10.96
0.8	UMApHM	2	2	65.84%	2.84	0.8	UMApHC	2	2	73.04%	2.75
			3	74.19%	11.89				3	82.43%	4.65
			4	80.69%	15.82				4	89.68%	9.47
			5	87.14%	13.63				5	92.32%	11.31
		3	2	29.04%	11.5			3	2	42.37%	11.6
			3	42.92%	6.7				3	55.89%	12.3
			4	52.83%	10.3				4	65.90%	13.01
			5	60.14%	10.39				5	75.00%	7.32
		4	2	21.06%	9.82			4	2	44.19%	4.17
			3	32.69%	10.47				3	58.80%	3.76
			4	42.10%	8.91				4	65.64%	10.52
			5	48.60%	19.02				5	73.18%	12.55
		5	2	18.19%	12.19			5	2	42.19%	9.29
			3	29.12%	7.5				3	52.65%	10
			4	36.93%	9.31				4	62.66%	7.92
			5	44.24%	8.78				5	71.62%	9.67

Table 4
$(r|X_p)hub - medianoid$ experiment results for TR data set given the leader's hubs are **UMApHM**

α	p	r	Follower's capture	CPU sec	α	p	r	Follower's capture	CPU sec	α	p	r	Follower's capture	CPU sec
0.6	6	6	39.31%	467.27	0.8	6	6	37.97%	449.76	0.9	6	6	40.86%	354.78
		8	49.19%	358.34			8	48.24%	330.15			8	49.44%	339.55
		10	56.94%	266.14			10	55.70%	158.63			10	56.06%	213.72
		12	64.02%	80.93			12	61.84%	85.34			12	61.54%	177.01
		14	68.91%	20.77			14	66.97%	35.69			14	66.45%	74.73
	8	6	28.58%	326.31		8	6	29.37%	393.24		8	6	31.11%	182.35
		8	37.09%	286.47			8	37.08%	263.35			8	38.69%	197.03
		10	44.37%	213.4			10	44.35%	174.8			10	44.83%	158.88
		12	51.77%	76.28			12	50.71%	84.56			12	50.49%	129.43
		14	57.97%	21.11			14	56.33%	39.49			14	55.77%	76.34
	10	6	19.91%	302		10	6	20.12%	385.03		10	6	20.74%	287.35
		8	27.13%	190.99			8	27.03%	222.86			8	27.77%	121.56
		10	34.10%	144.75			10	33.84%	200.72			10	33.86%	128.64
		12	40.48%	68.05			12	40.74%	72.85			12	39.89%	90.58
		14	45.73%	18.38			14	46.84%	26.06			14	44.90%	32.29
	12	6	15.83%	168.81		12	6	16.93%	232.28		12	6	18.45%	127.1
		8	21.79%	125.01			8	23.41%	104.74			8	24.59%	42.89
		10	27.06%	61.06			10	28.62%	107.7			10	29.08%	36.44
		12	31.37%	28.81			12	32.81%	13.03			12	32.98%	42.05
		14	35.48%	13.45			14	35.85%	15.83			14	36.18%	31.13
	14	6	13.04%	141.97		14	6	13.02%	109.78		14	6	13.66%	66.42
		8	17.87%	108.2			8	18.57%	13.44			8	18.81%	29.76
		10	22.25%	22.23			10	22.52%	13.66			10	22.50%	30.4
		12	26.00%	10.43			12	25.20%	34.18			12	25.60%	31.8
		14	28.42%	9.76			14	27.40%	25.14			14	28.18%	30.64

is more advantageous if the former one does not have information about the competition.

5. $(r|p)$ hub-centroid

Let $< G = (N, E), w_{ij}, c_{ij}, \alpha >$ be a many-to-many flow network. At the time the leader makes his/her decision, he/she knows that the follower is going to respond rationally, that is, the follower is going to choose the optimal solution of $(r|X_p)hub - medianoid$ problem after observing X_p. Therefore, $(r|X_p)hub - medianoid$ problem is embedded in $(r|p)hub - centroid$ problem. Due to this relation, $(r|p)hub - centroid$ problem has a bilevel structure.

5.1. Complexity of (r|p) hub-centroid

We prove that the problem of finding a $(r|p)hub - centroid$ is NP-complete by using polynomial time reduction from *vertex cover* problem, a known NP-complete problem due to Karp (1972).

Decision version of vertex cover problem: Given an undirected graph $G = (N, E)$ and an integer p, determine if G has a *vertex cover*, that is, if there is a set of vertices C with $|C| \leq p$ such that for each edge $(i, j) \in E$, either i or j is in C.

Theorem 2. *The problem of finding* $(r|p)hub - centroid$ *is NP-complete even if* $\alpha = 1$.

Proof. Given an instance of *vertex cover* problem, we construct a network $G' = (N', E')$ where $G' = G$. Let $c_{ij} = 1$ if $(i, j) \in E$. The flow values for all pairs $i, j \in N$ is set to 1 if $(i, j) \in E$ and 0 otherwise. Also, assume that $\alpha = 1$.

We prove the theorem by showing that there exists a set of p points X_p on G' such that $f(X_p, Y_r(X_p)) = 0$ if and only if there exists a *vertex cover* C with $|C| \leq p$.

Assume that *vertex cover* problem has solution $C \subseteq N$ and $|C| \leq p$. By letting $X_p \supseteq C$, we can observe that unit flow w_{ij} either i or j is in X_p. Therefore, for each flow w_{ij}, the service level provided by the leader $\beta_{ij} = 1$ noting that each flow is routed via only a single link. Since the follower cannot provide a strictly better service level for

any of the node pairs i and j, no flow is captured by the follower. Then, $f(X_p, Y_r^*(X_p)) = 0$.

On the other hand, suppose X_p in G' is such that $f(X_p, Y_r^*(X_p)) = 0$. Also, assume that X_p does not contain a subset which is a vertex cover C of G. So, there exists an edge $(i, j) \in E'$ where neither i nor j is in X_p. Then, the follower can capture the flow w_{ij} by location his/her hubs on both i and j which yields $\gamma_{ij} = 1$. On the other hand, the follower can provide a service level $\beta_{ij} \geq 2$ since the flow should visit a hub that is different from both i and j. Then, $f(X_p, Y_r^*(X_p)) \geq 1$ which contradicts with the assumption.

Hence, we conclude that $(r|p)hub - centroid$ is reducible from *vertex cover* problem in polynomial time. So, it is NP-complete. □

5.2. Mathematical model for (r|p) hub-centroid problem

To provide a bilevel mathematical model for the $(r|p)hub - centroid$ problem, we define following decision variables:

$$H_k = \begin{cases} 1, & \text{if the leader locates a hub on node } k \in H \\ 0, & \text{otherwise} \end{cases}$$

$$U_{ijk} = \begin{cases} 1, & \text{if flow from node } i \in N \text{ to node } j \in N \text{ visits hub} \\ & k \in H \text{ as the first hub} \\ 0, & \text{otherwise} \end{cases}$$

$$V_{ijm} = \begin{cases} 1, & \text{if flow from node} i \in N \text{ to node } j \in N \text{ visits hub} \\ & m \in H \text{ as the second hub} \\ 0, & \text{otherwise} \end{cases}$$

$$a_{ij} = \begin{cases} 1, & \text{if flow from node } i \in N \text{ to node } j \in N \text{ is captured} \\ & \text{by the follower} \\ 0, & \text{otherwise} \end{cases}$$

$\beta_{ij} = $ the service level for the node pair $i, j \in N$
provided by the leader.

Note that capital letter decision variables H_k, U_{ijk}, V_{ijm} of the leader are analogous to their lowercase versions defined for the follower and location variable H_k should not be confused set of possible hub locations H.

The following bilevel mixed integer problem **H-CEN-B** solves the $(r|p)hub - centroid$ problem:

$$\text{minimize} \sum_{i \in N} \sum_{j \in N} w_{ij} a_{ij} \tag{12}$$

$$\text{subject to} \sum_{k \in H} H_k = p, \tag{13}$$

$$\sum_{k \in H} U_{ijk} = 1 \; \forall i, j \in N, \tag{14}$$

$$\sum_{m \in H} V_{ijm} = 1 \; \forall i, j \in N, \tag{15}$$

$$U_{ijk} \leq H_k \; \forall i, j \in N \; \forall k \in H \tag{16}$$

$$V_{ijm} \leq H_m \; \forall i, j \in N \; \forall m \in H \tag{17}$$

$$\beta_{ij} \geq \sum_{k \in H} U_{ijk}(c_{ik} + \alpha c_{km}) + V_{ijm} c_{mj} \; \forall i, j \in N \; \forall m \in H \tag{18}$$

$$\beta_{ij} - \gamma_{ij} \leq a_{ij} M \; \forall i, j \in N, \tag{19}$$

where γ_{ij} is induced from optimal solution of

H − MED for $H_k, k \in H$ (20)

$$H_k, U_{ijk}, V_{ijk}, a_{ij} \in \{0, 1\} \text{ and } \beta_{ij} \geq 0 \; \forall i, j \in N \; \forall k, m \in H \tag{21}$$

The objective (12) minimizes amount of flow captured by the follower which is equivalent to maximizing amount of flow captured by the leader. Constraint (13) ensures the leader locates p hubs on the set of available nodes. Constraints (14), (15), (16) and (17) guarantee that flow from node $i \in N$ to $j \in N$ visits two (not necessarily different) hub nodes $k \in H$ and $m \in H$. Constraints (18) correctly calculate the service levels of the follower in the following manner: if $V_{ijm} = 0$, the constraint becomes redundant. However, if $V_{ijm} = 1$ the RHS of the constraint becomes the service level provided by the leader for flow from node $i \in N$ to $j \in N$. Constraints (19) correctly calculate whether a flow is captured by the follower or not in the following manner: If the LHS of the constraint is positive, that is the follower provides a service level for the flow from node $i \in N$ to $j \in N$ which is better than the service level provided by the leader, the RHS of the constraint must be positive and $a_{ij} = 1$. Otherwise, the constraint becomes redundant. Constraint (20) states that service levels of the follower are induced from another optimization problem, **H-MED** according to decisions of the leader and hence **H-CEN-B** is a bilevel problem due to Constraint (20). Constraints (21) are the domain constraints.

As stated by Bard (1999) and Dempe (2002), bilevel models are hard to solve even for a small number of decision variables. We use a mini–max approach to linearize **H-CEN-B** where the leaders choose a hub set so as to minimize the total captured flow by the follower in the remaining scenario. Let us define a new parameter:

γ_{ij}^S = service level for pair $i, j \in N$ provided by the follower

if he/she chooses $S \subseteq H$ as hub set

that is, $\gamma_{ij}^S = min_{k,m \in S}\{c_{ik} + \alpha c_{km} + c_{mj}\}$. Also define a new decision variable:

$$a_{ij}^S = \begin{cases} 1, & \text{if flow from node } i \in N \text{ to node } j \in N \text{ is captured by} \\ & \text{the follower when he/she chooses} \\ & S \subseteq H \text{ as hub set} \\ 0, & \text{otherwise} \end{cases}$$

Table 5
Summary of the experiments conducted for **H-CEN.**

n	Follower's capture (%)	Solution time (s)	Gap %	n	Follower's capture (%)	Solution Time (s)	Gap %
5	41.39	1.04	–	16	43.15	–	76.74
6	40.16	3.83	–	17	58.49	–	83.53
7	40.59	13.3	–	18	61.16	–	86.10
8	36.36	18.34	–	19	100	–	91.41
9	34.31	109.31	–	20	100	–	92.33
10	39.72	475.02	–	21	58.18	–	88.02
11	41.03	325.55	–	22	98.36	–	92.19
12	40.55	–	4.5	23	57.65	–	87.82
13	39.55	–	20.62	24	100	–	93.02
14	46.18	–	17.16	25	100	–	93.33
15	43.75	–	55.92				

Then, the following mixed integer problem **H-CEN** solves the $(r|p)hub - centroid$ problem with exponential number of decision variables and constraints:

$$\text{minimize} \; Z \tag{22}$$

subject to (13) − (18)

$$Z \geq \sum_{i \in N} \sum_{j \in N} a_{ij}^S w_{ij} \; \forall i, j \in N \; \forall S \subseteq H \text{ with } |S| = r, \tag{23}$$

$$\beta_{ij} - \gamma_{ij}^S \leq a_{ij}^S M \; \forall i, j \in N \; \forall S \subseteq H \text{ with } |S| = r, \tag{24}$$

$$H_k, U_{ijk}, V_{ijk}, a_{ij}^S \in \{0, 1\} \text{ and } \beta_{ij} \geq 0 \; \forall i, j \in N \; \forall S \subseteq H \text{ with } |S| = r, \tag{25}$$

Objective function (22) and constraints (23) together minimize the highest possible captured flow value by the follower in the remaining scenario. Constraints (24) correctly calculate whether a flow is captured with a hub set $S \subseteq H$ by the follower or not, similar to (18). Constraints (25) are domain constraints.

The mixed integer program **H-CEN** has $3n^2m + 2n^2 + 2n^2\binom{m}{r} + 1$ constraints and $2n^2m + n^2\binom{m}{r} + n^2 + m + 1$ variables of which $2n^2m + n^2\binom{m}{r} + m$ are binary where $|N| = n$, $|H| = m$.

5.3. Computational performance of H-CEN

We used CAB data set to observe the performance of **H-CEN** model via CPLEX. Since **H-CEN** model contains exponential number of variables and constraints, the experiment is conducted for first n nodes of the data set where n ranges from 5 to 25 for the $\alpha = 0.6$ value. Moreover, values of problem parameters p and r are set to two which yield $O(n^4)$ variables and constraints. Table 5 summarizes the results of the computational study for these instances within a time limit of 7200 s ($= 2$ h).

The conducted computational study revealed that the **H-CEN** model can only be solvable within 2 h for $n \leq 11$. Moreover, for values $n \geq 15$, the optimality gap is greater than 50%. Therefore, for even very small instances, an optimal solution of **H-CEN** cannot be obtained via CPLEX. Thus, we develop enumeration-based solution algorithms presented in the next section.

5.4. Enumeration-based solution algorithms

Since **H-CEN-B** is a bilevel model and **H-CEN** contains exponential number of constraints, they are inefficient to solve $(r|p)hub - centroid$ problem for even small and medium size networks. Therefore, we propose enumeration-based algorithms to obtain optimal solutions of $(r|p)hub - centroid$ problem for problem instances with reasonable sizes.

The first idea is observing all possible choices of leader's hub sets and the response that the follower gives to these possible solutions. This leads us to complete enumeration algorithm for $(r|p)hub - centroid$ problem where for all possible hub set choices of the leader and follower, service level provided to each flow is calculated.

The *complete enumeration algorithm* enumerates all the possible choices of hub sets of the leader and follower, then for all node pairs $i, j \in N$ determines if the flow w_{ij} is captured by the follower or not. Therefore, running time of the algorithm is proportional to $n^2|\mathcal{P}_p(H)||\mathcal{P}_r(H)|$.

However, the following theorem states that enumerating all of the remaining feasible solutions is redundant if a feasible solution to $(r|p)hub - centroid$ problem is observed.

Theorem 3. *Let X_p be a feasible solution to $(r|p)hub - centroid$ problem. If there exist X'_p and Y'_r with $f(X_p, Y^*_r(X_p)) < f(X'_p, Y'_r)$ then X'_p cannot be an optimal solution to $(r|p)hub - centroid$ problem.*

Proof. $f(X'_p, Y'_r) \leq f(X'_p, Y^*_r(X'_p))$ where $Y^*_r(X_p)$ is the optimal solution to $(r|X_p)hub - medianoid$ problem given that the hub set of the leader is X_p. Then, $f(X_p, Y^*_r(X_p)) < f(X'_p, Y'_r)$ and $f(X'_p, Y'_r) \leq f(X'_p, Y^*_r(X'_p))$ together imply that $f(X_p, Y^*_r(X_p)) < f(X'_p, Y^*_r(X'_p))$. Therefore, X'_p cannot be an optimal solution to $(r|p)hub - centroid$ problem. □

By using Theorem 3, we can improve the solution time of complete enumeration algorithm by skipping the search of the follower's reaction to the choices of the leader which cannot be an optimal solution to $(r|p)hub - centroid$ problem. We call this modified version of the algorithm as *smart enumeration* algorithm.

We can still decrease running time of smart enumeration algorithm if another bound on the amount of the flow captured by the leader is obtained. For the special case $p \geq r$, we can improve the efficiency of the algorithm by skipping some feasible solutions that cannot be optimal.

Theorem 4. *If $p \geq r, p < |H| - 2, r \geq 2$, all flow values $w_{ij} > 0$ for all $i \neq j$ and the cost matrix satisfies triangular inequality, then the optimal solution of $(r|p)hub - centroid$ problem X^*_p satisfies $f(X^*_p, Y^*_r(X^*_p)) < W/2$ where W is the total flow over the network.*

Proof. Assume that X^*_p is an optimal solution of $(r|p)hub - centroid$ problem which satisfies $f(X^*_p, Y^*_r(X^*_p)) \geq W/2$. Then, at least one half of the total flow on the network is captured by the follower. Equivalently, we can say $\gamma_{ij} < \beta_{ij}$ hold for at least half of the total flow where γ_{ij} and β_{ij} values are implied by X^*_p and $Y^*_r(X^*_p)$, respectively. Then, the follower can provide a better service level (viz. can provide a better β_{ij} value) for at least one half of the total flow by setting his/her hub set $X'_r = Y^*_r(X^*_p)$. Then, $f(X'_r, Y^*_r(X^*_p)) = 0$ since both the leader and follower provide same service levels for all flows and in case of equity the follower captures the flow. Since $p < |H| - 2$ then there are two nodes $i, j \in H\subseteq N$ but not in X'_p. The follower can move two of his/her hubs to i and j ,and captures the flow w_{ij} due to triangular inequality. Let Y'_r this new hub set. Then, $f(X'_p, Y'_r) > 0$. So, we can say that the service levels induced by Y'_r dominate the service levels implied by $Y^*_r(X^*_p)$ contradicting with the optimality condition $f(X^*_p, Y^*_r(X^*_p)) \geq f(X^*_p, Y'_r)$.

Hence, under the conditions $p \geq r, p < |H| - 2, r \geq 2$, all flow values $w_{ij} > 0$ for all $i \neq j$ and the cost matrix satisfies triangular inequality, an optimal solution of $(r|p)hub - centroid$ problem X'_p satisfies $f(X^*_p, Y^*_r(X^*_p)) < W/2$ □

Utilizing Theorem 4, we can further improve running time of the algorithm. The bound states that in an optimal solution the leader should get at least 50% of the total flow, so if there exist X_p and Y_r with $f(X_p, Y_r) > W/2$ where W is the total flow on the network with $p \geq r$ then we can say that X_p is not an optimal solution to $(r|p)hub -$

centroid problem. We call this improved version of the algorithm as *smart enumeration with 50%-bound*.

5.5. Computational analysis of enumeration-based solution algorithms

All algorithms are coded in Java 1.6.0_23 using the same computer. Table 6 summarizes all 80 instances used in the computational study of *smart enumeration* and *smart enumeration with 50%-bound* algorithms:

For $(r|p)hub - centroid$ problem, TR instances are generated for relatively smaller values of number of hubs to be located, that is $p, r \in \{2, 3, 4, 5\}$, unlike the instances for $(r|X_p)hub - medianoid$ problem due to memory requirements and long CPU times. Although worst case running times of all three algorithms are proportional to $n^2|\mathcal{P}_p(H)||\mathcal{P}_r(H)|$, in practice *smart enumeration* and *smart enumeration with 50%-bound* algorithms outperforms complete enumeration dramatically in large instances.

Tables 7 and 8 summarize conducted experiments for $(r|p)hub - centroid$ problem for CAB and TR data sets respectively.

Computational analysis also revealed that the leader can increase his/her market share by acting rationally in case of competition. If the leader makes his/her decision ignoring competition, his/her decision will be based on the solutions of some classic models, such as p-hub median and p-hub center. However, the leader may lose some of his/her market in case of another firm entering the market and capturing some of the customers that previously belonged to the leader. In Table 9, we compare percentage of captured flow by the follower if the leader locates his/her hubs on the optimal locations of $(r|p)hub - centroid$ or the leader locates his/her hubs on p-hub median and p-hub center (without considering competition) and the follower responds based on $(r|X_p)hub - medianoid$ problem.

For example, if $p = r = 2, \alpha = 0.6$ and the leader locates his/her hubs by being aware of competition, then the follower can only capture 46.14% of the market. However, if the leader locates his/her hubs according to the optimal solution of p-hub median problem, the follower can capture 65.62% of the market and leader loses 19.48% of the market to the follower. Likewise, optimal solution of p-hub center problem is a worse choice and the follower can capture 75.86% of the market which means that the leader losts 29.72%. Fig. 1 depicts the optimal hub locations of $(r|p)hub - centroid$, p-hub median and p-hub center problems where $p = r = 2$ and $\alpha = 0.6$.

As seen above, optimal solution of p-hub median is preferable to optimal solution of the p-hub center problem in all instances. This result is a direct consequence of difference in definition of these two problems. While p-hub median problem minimizes weighted sum of service levels of each node pair, p-hub center problem minimizes service level of the most disadvantageous node pair. p-hub center problem ignores the flows between node pairs and focuses only on the distance between them. On the other hand, p-hub median problem locates hubs on a set of node so that the node pairs with higher flow are given more consideration.

Also observe that the p-hub median optimal solution can be regarded as a promising solution to $(r|p)hub - centroid$ problem. Especially for larger values of p, the difference in the market share between the optimal solution of $(r|p)hub - centroid$ and p-hub median

Table 7
Summary of experiment results for $(r|p)hub - centroid$ problem for CAB data set.

α	p	r	Follower's Capture	CPU sec smart	CPU sec smart-%50	α	p	r	Follower's Capture	CPU sec smart	CPU sec smart-%50
0.6	2	2	46.14%	1.52	0.93	0.8	2	2	43.68%	1.35	0.72
		3	64.37%	12.71	–			3	59.59%	11.78	–
		4	74.75%	70.32	–			4	70.75%	100.37	–
		5	83.52%	320.78	–			5	78.74%	535.15	–
	3	2	30.39%	5.81	5.61		3	2	29.18%	4.31	4.13
		3	45.13%	19.94	11.46			3	42.87%	23.13	14.65
		4	53.69%	88.02	–			4	52.84%	142.68	–
		5	62.02%	557.16	–			5	60.14%	791.55	–
	4	2	17.91%	19.27	17.24		4	2	21.06%	17.96	18.56
		3	28.39%	36.62	33.27			3	30.70%	30.7	25.58
		4	37.73%	141.6	77.38			4	38.39%	212.61	155.75
		5	46.18%	631.1	–			5	45.24%	1015.49	–
	5	2	14.30%	70.35	70.09		5	2	15.30%	74.25	72.77
		3	23.73%	117.4	117.15			3	23.24%	139.05	135.53
		4	31.91%	371.14	341.28			4	31.78%	382.09	360.54
		5	39.58%	1498.94	1272.24			5	38.57%	1335.03	1043.81

Table 8
Summary of experiment results for $(r|p)hub - centroid$ problem for TR data set.

α	p	r	Follower's capture	CPU sec smart-%50	α	p	r	Follower's capture	CPU sec smart-%50	α	p	r	Follower's capture	CPU sec smart-%50
0.6	2	2	49.44%	7.2	0.8	2	2	46.84%	8.81	0.9	2	2	44.12%	8.39
		3	64.65%	45.69			3	60.05%	39.23			3	58.74%	38.84
		4	74.97%	257.8			4	70.03%	280.44			4	67.98%	265.09
		5	84.72%	1409.52			5	77.97%	1326.98			5	75.45%	1775.52
	3	2	30.49%	24.6		3	2	30.68%	21.65		3	2	30.35%	26.9
		3	40.82%	77.19			3	40.81%	73.32			3	39.90%	68.93
		4	56.18%	400.15			4	51.43%	351.76			4	50.03%	365.13
		5	65.58%	1630.12			5	60.66%	1354.41			5	58.18%	2002.92
	4	2	20.07%	75.35		4	2	20.33%	80.3		4	2	20.38%	77.32
		3	30.57%	161.12			3	30.19%	154.12			3	29.55%	176.41
		4	42.15%	724.81			4	39.41%	549.42			4	38.11%	901.68
		5	51.89%	2166.95			5	48.57%	2087.01			5	46.83%	3022.19
	5	2	14.32%	415.39		5	2	14.82%	440.62		5	2	14.27%	455.12
		3	23.61%	551.04			3	22.12%	534.12			3	22.87%	583.39
		4	32.34%	1098.69			4	29.28%	997.63			4	31.76%	1706.91
		5	40.05%	4911.17			5	37.44%	4450.71			5	38.91%	6634.97

Table 9
Comparison market share of the follower in the optimal solution of $(r|X_p)hub - medianoid$ with the classical model for CAB data set with $\alpha = 0.6$.

|		$(r|p)hub-$ centroid	$(r|X_p)hub-$ medianoid	Difference with centroid	$(r|X_p)hub-$ medianoid	Difference with centroid
p	r		$X_p = p - hub$ median		$X_p = p - hub$ center	
2	2	46.14%	65.62%	19.48%	75.86%	29.72%
	3	64.37%	78.25%	13.88%	85.2%	20.83%
	4	74.75%	87.08%	12.33%	90.98%	16.23%
	5	83.52%	92.26%	8.74%	94.74%	11.22%
3	2	30.39%	30.49%	0.1%	51.81%	21.42%
	3	45.13%	45.13%	0%	70.25%	25.12%
	4	53.69%	53.69%	0%	79.08%	25.39%
	5	62.02%	62.02%	0%	85.23%	23.21%
4	2	17.91%	17.91%	0%	36.56%	18.65%
	3	28.39%	28.39%	0%	47.39%	19.00%
	4	37.73%	37.73%	0%	57.38%	19.65%
	5	46.18%	46.18%	0%	66.93%	20.75%
5	2	14.3%	18.64%	4.34%	45.62%	31.32%
	3	23.73%	28.14%	4.41%	57.27%	33.54%
	4	31.91%	35.04%	3.13%	69.34%	37.43%
	5	39.58%	42.32%	2.74%	76.75%	37.17%
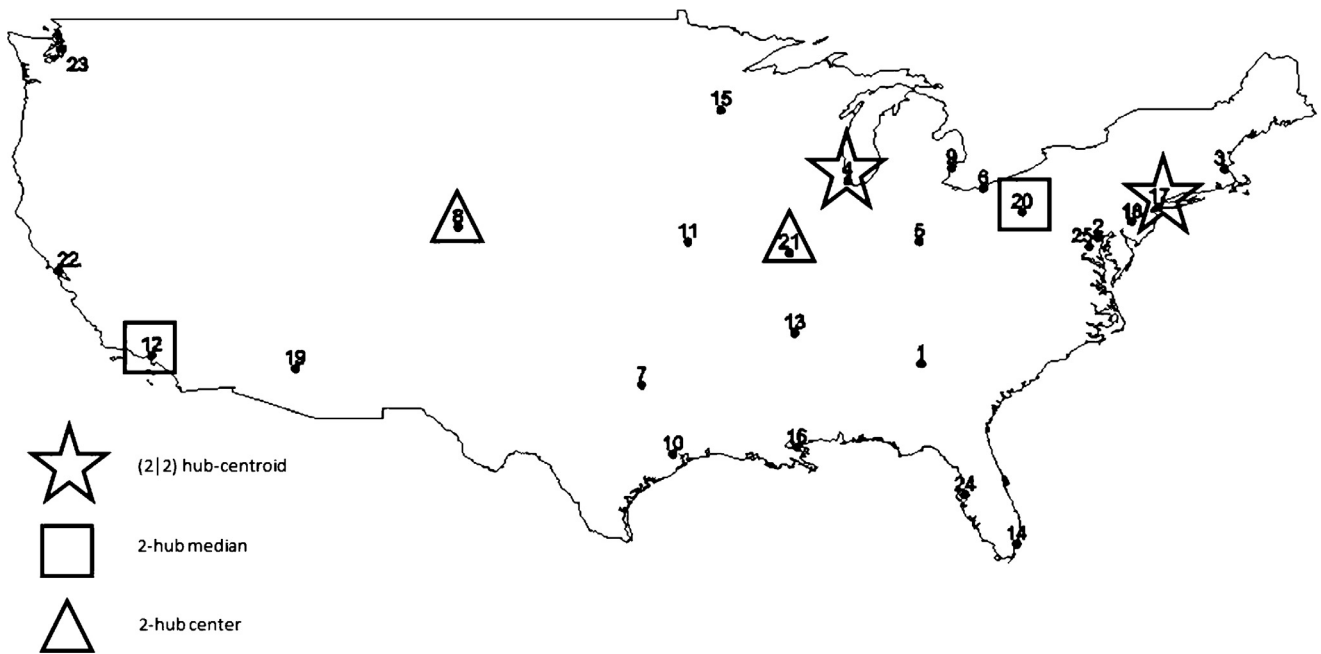
Fig. 1. Optimal hub locations of $(r|p)hub - centroid$, p-hub median and p-hub center problems where $p = r = 2$ and $\alpha = 0.6$.

Table 10
CPU times of the *smart enumeration* where *p*-hub median problem solution is initial incumbent.

p / r	2	3	4	5
2	1.52	10.94	52.84	269.13
3	3.21	8.9	20.76	105.55
4	14.64	20.04	45.73	145.72
5	68.51	105.97	300.05	1078.97

Table 11
Demand loss of the leader by choosing **UMApHM** as his/her hub set.

α	p\r	2	3	4	5
0.6	2	1.17%	4.08%	5.16%	5.26%
	3	0.00%	0.00%	0.22%	0.85%
	4	2.07%	3.13%	2.65%	3.81%
	5	0.70%	0.27%	1.64%	2.15%
0.8	2	3.12%	2.43%	2.45%	6.91%
	3	0.00%	0.00%	0.00%	0.00%
	4	0.00%	0.00%	0.00%	0.00%
	5	0.90%	2.12%	3.42%	2.77%
0.9	2	6.55%	8.35%	9.54%	9.83%
	3	4.24%	3.82%	5.22%	4.21%
	4	0.00%	1.01%	0.36%	0.58%
	5	2.20%	1.07%	1.28%	2.10%

is reasonably small and for seven of the 16 instances the optimal hub sets and optimal values of these problems coincide.

Required CPU time for *smart enumeration algorithms* directly depends on the order of enumeration of leader's hub set choices. Currently, the algorithm enumerates sets lexicographically. For example, if $p = 3$, first algorithm starts with $X_p = \{1, 2, 3\}$, then goes on with $\{1, 2, 4\}$, $\{1, 2, 5\}$ and so on. However, as stated in Theorem 3, if a feasible solution which provided genuine bound is already obtained, running time of the algorithm can be improved.

For the instances reported above, the optimal solution of p-hub median problem diverges 4.32% on average from the optimal solution of $(r|p)hub - centroid$ problem. Then, another computational experiment is conducted for smart enumeration algorithm on CAB data set with $p, r \in \{2, 3, 4, 5\}$ and $\alpha = 0.6$ by including the bound obtained by the optimal solution of p-hub median problem. Table 10 depicts CPU times of this experiment.

The experiment revealed that the running time of smart enumeration algorithm has improved up to 81% (37% on average) for these instances when the optimal solution of p-hub median problem is used a bound on the optimal value of $(r|p)hub - centroid$ problem. Also, as the difference between optimal solutions of p-hub median and $(r|p)hub - centroid$ problems get smaller, higher improvement is obtained.

5.6. Discussion on larger p and r values

Since running times of proposed enumeration based algorithms increase exponentially as p or r increase, for large p and r values

these algorithms may not be very efficient. However, as discussed in Table 9, follower's captures are compared for the cases the leader either chooses **UMApHM** or $(r|p)hub - centroid$ as his/her hub set. Similar comparison is summarized in Table 11 for TR data set.

As in Table 11, the average demand lost by the leader is 2.45% on average and for the instances where $p = r$ the average lost is 2.06%. Therefore, additional computational experiments are conducted for the case where the leader chooses **UMApHM** for TR data set where both the leader and follower locate equal number of hubs.

Since, we do not have chance to compare **UMApHM** and $(r|p)hub - centroid$ results for large p and r values, a fair measure of performance **UMApHM** in a competitive environment could be the case where both firms have equal capabilities. As seen on Table 12, if both player locate equal number of hubs, the leader can capture high amount of flows even ignoring competition and locating hubs according to **UMApHM**. As number of hubs increases, solution times get reasonably short as well. Hence, it can be concluded that for large values of p and r, **UMApHM** is a preferable solution for the leader.

Table 12
Percentage of lost market share of the follower by choosing **UMApHM** as his/her hub set rather than optimal $(r|p)hub - centroid$.

α	$p = r$	follower's capture	CPU time	α	$p = r$	follower's capture	CPU time	α	$p = r$	follower's capture	CPU time
0.6	6	39.31%	688.48	0.8	6	37.97%	532.49	0.9	6	40.86%	388.77
	7	37.26%	408.19		7	36.91%	474.21		7	38.35%	404.49
	8	37.09%	515.36		8	37.08%	301.86		8	38.69%	260.56
	9	35.18%	243.62		9	36.21%	186.07		9	36.99%	168.50
	10	34.10%	199.44		10	33.84%	172.06		10	33.86%	202.53
	11	33.42%	97.47		11	34.13%	133.85		11	34.38%	102.63
	12	31.37%	104.39		12	32.81%	38.10		12	32.98%	51.08
	13	30.61%	41.46		13	31.00%	36.35		13	31.65%	38.85
	14	28.42%	34.01		14	27.40%	37.11		14	28.18%	31.82
	15	24.19%	31.96		15	25.18%	32.90		15	24.80%	31.55

6. Conclusions

In this paper, we propose a duopoly model where two competitors sequentially choose hub locations and aim to maximize their own market share under Stackelberg competition. Although competitive location has attracted attention of economists and OR practitioners, hub location studies considering competition are rare in the literature. Therefore, formal definitions of terminology and problem were deficiencies in both competitive location and hub location literature. It is assumed that both players have perfect information of the environment. It is also assumed that both players are rational which means that they aim to maximize their market share. The market share of firms is determined by captured flow. Although choice of customers depend on many attributes, we assume that customers prefer the firm which offers a better service level.

From the view of the follower the problem is a maximum cover problem rather than a competitive model. On the other hand, the competition issue becomes important from perspective of the leader. After the leader makes his/her decisions, the follower takes action and then the markets shares are determined. Therefore, the leader's problem has a bilevel nature. We then propose a bilevel model and an equivalent mathematical model as well as computational complexity results for the problems. Mathematical model of the leader can only be solvable for very small instances and solving the bilevel model is much harder. However, proposed enumeration-based algorithms can solve the problem for relatively bigger instances even though the worst-case complexity tends to complete enumeration. On the other hand, **UMApHM** can be a promising solution for the leader when solving $(r|p)hub - centroid$ requires higher amount of CPU time and memory.

Solving the $(r|p)hub - centroid$ problem exactly or near-optimally is an open research direction. Exact methods such as column generation or branch-and-price can be thought of as possible ideas. Also, some meta-heuristic approaches can be employed to solve the problem within an acceptable optimality gap.

Acknowledgment

References

Alumur, S., & Kara, B. (2008). Network hub location problems: the state of the art. *European Journal of Operational Research, 190*(1), 1–21.

Aykin, T. (1994). Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research, 79*(3), 501–523.

Bard, J. (1999). *Practicle bilevel optimization*. Dordrecht: Kluwer Academicians.

Bertrand, J. (1883). Théorie des richesses: revue de théorie mathématique de la richesse socialepar léon walras et recherches sur les principes mathématiques de la richessepar augustin cournot. *J. des Savants*.

Campbell, J. F. (1992). Location and allocation for distribution systems with transshipments and transportation economies of scale. *Annals of Operations Research, 40*(1), 77–99.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research, 72*(2), 387–405.

Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2002). Hub location problems. *Facility location: Applications and Theory, 1*, 373–407.

Campbell, J. F., & O'Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science, 46*(2), 153–169.

Cournot, A. A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. chez L. Hachette.

Dasci, A. (2011). Conditional location problems on networks and in the plane. In *Foundations of location analysis* (pp. 179–206). Springer.

Dempe, S. (2002). *Foundations of bilevel programming*. Springer.

Drezner, Z. (1982). Competitive location strategies for two facilities. *Regional Science and Urban Economics, 12*(4), 485–493.

Eiselt, H. A., & Laporte, G. (1997). Sequential location problems. *European Journal of Operational Research, 96*(2), 217–231.

Eiselt, H. A., & Marianov, V. (2009). A conditional p-hub location problem with attraction functions. *Computers and Operations Research, 36*(12), 3128–3135.

Ernst, A. T., Hamacher, H., Jiang, H., Krishnamoorthy, M., & Woeginger, G. (2009). Uncapacitated single and multiple allocation p-hub center problems. *Computers and Operations Research, 36*(7), 2230–2241.

Ernst, A. T., Jiang, H. Krishnamoorthy, M. (2005). Reformulations and computational results for uncapacitated single and multiple allocation hub covering problems. Unpublished report, CSIRO mathematical and information sciences, Australia.

Ernst, A. T., & Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location science, 4*(3), 139–154.

Ernst, A. T., & Krishnamoorthy, M. (1998). Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research, 104*(1), 100–112.

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

Hotelling, H. (1929). *Stability in competition'economic journal, 39*, 41–57. Reprinted in Dean RD, In Leahy WH and McKee DL (Eds.), Spatial economic theory, (Free Press, New York), (pp. 103–18)

ILOG, IBM. (2012). Cplex 12.4.

Kara, B. Y. (1999). *Modelling and analysis of issues in hub location problem*. Bilkent University Ph.D. thesis.

Kara, B. Y., & Taner, M. R. (2011). Hub location problems: the location of interacting facilities. In *Foundations of location analysis* (pp. 273–288). Springer.

Kara, B. Y., & Tansel, B. C. (2000). On the single-assignment p-hub center problem. *European Journal of Operational Research, 125*(3), 648–655.

Kara, B. Y., & Tansel, B. C. (2003). The single-assignment hub covering problem: Models and linearizations. *Journal of the Operational Research Society, 54*(1), 59–64.

Kara, B. Y., & Tansel, B. C. (2001). The latest arrival hub location problem. *Management Science, 47*(10), 1408–1420.

Karp, R. M. (1972). *Reducibility among combinatorial problems*. Springer.

Kuby, M. J., & Gray, R. G. (1993). The hub network design problem with stopovers and feeders: the case of federal express. *Transportation Research Part A: Policy and Practice, 27*(1), 1–12.

Lüer-Villagra, A., & Marianov, V. (2013). A competitive hub location and pricing problem. *European Journal of Operational Research, 231*(3), 734–744.

Marianov, V., Serra, D., & ReVelle, C. (1999). Location of hubs in a competitive environment. *European Journal of Operational Research, 114*(2), 363–371.

O'Kelly, M. E. (1986a). Activity levels at hub facilities in interacting networks. *Geographical Analysis, 18*(4), 343–356.

O'Kelly, M. E. (1986b). The location of interacting hub facilities. *Transportation Science, 20*(2), 92–106.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research, 32*(3), 393–404.

O'Kelly, M. E. (1992). Hub facility location with fixed costs. *Papers in Regional Science, 71*(3), 293–306.

Puerto, J., Ramos, A. B., & Rodríguez-Chía, A. M. (2011). Single-allocation ordered median hub location problems. *Computers & Operations Research, 38*(2), 559–570.

Puerto, J., Ramos, A. B., & Rodríguez-Chía, A. M. (2013). A specialized branch & bound & cut for single-allocation ordered median hub location problems. *Discrete Applied Mathematics, 161*(16), 2624–2646.

Sasaki, M. (2005). Hub network design model in a competitive environment with flow threshold. *Journal of the Operations Research Society of Japan-Keiei Kagaku, 48*(2), 158–171.

Sasaki, M., Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2014). Hub arc location with competition. *Computers and Operations Research, 47*, 27–41.

Sasaki, M., & Fukushima, M. (2001). Stackelberg hub location problem. *Journal of the Operations Research Society of Japan, 44*(4), 390–405.

Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated *p*-hub median problems. *European Journal of Operational Research, 94*(3), 582–593.

Tan, P. Z., & Kara, B. Y. (2007). A hub covering model for cargo delivery systems. *Networks, 49*(1), 28–39.

Teitz, M. B., & Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations research, 16*(5), 955–961.

Von Stackelberg, H. (1951). *Grundlagen der theoretischen Volkswirtschaftslehre*. Mohr Siebeck.

Wagner, B. (2008). A note on location of hubs in a competitive environment. *European Journal of Operational Research, 184*(1), 57–62.

Yaman, H. (2009). The hierarchical hub median problem with single assignment. *Transportation Research Part B: Methodological, 43*(6), 643–658.

Yaman, H., Kara, B. Y., & Tansel, B. C. (2007). The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B: Methodological, 41*(8), 906–919.