HUB LOCATION UNDER COMPETITION

A THESIS
SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

by
Ali İrfan Mahmutoğulları
July 2013

I certify that I have read this thesis and that in my opinion it is full adequate, in scope and in quality, as a dissertation for the degree of Master of Science.

Assoc. Prof. Bahar Yetiş Kara (Advisor)

I certify that I have read this thesis and that in my opinion it is full adequate, in scope and in quality, as a dissertation for the degree of Master of Science.

Prof. M. Selim Aktürk

I certify that I have read this thesis and that in my opinion it is full adequate, in scope and in quality, as a dissertation for the degree of Master of Science.

Assoc. Prof. Sinan Gürel

Approved for the Graduate School of Engineering and Science

Prof. Dr. Levent Onural
Director of the Graduate School of Engineering and Science

ABSTRACT

HUB LOCATION UNDER COMPETITION

Ali İrfan Mahmutoğulları

M.S. in Industrial Engineering

Supervisor: Assoc. Prof. Bahar Yetiş Kara

July 2013

Hubs are consolidation and dissemination points in many-to-many flow networks. The 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 thesis, we propose a competitive hub location problem where the market is assumed to be a duopoly. Two decision-makers (or firms) sequentially decide the locations of their hubs and then customers choose the firm according to provided service levels. Each decision-maker aims to maximize his/her market share. Having investigated the existing studies in the field of economy, retail location and operation research, 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. After defining them as combinatorial optimization problems, the problems are proved to be NP-hard. Linear programming models are presented for these problems as well as exact solution algorithms for the *$(r|p)$ hub-centroid* problem that outperform the linear model in terms of memory requirement and CPU time. The performance of models and algorithms are tested by the computational analysis conducted on two well-known data sets from the hub location literature.

Keywords: Hub location, competition models, competitive location.

ÖZET

REKABET ORTAMINDA ADÜ YER SEÇİMİ PROBLEMİ

Ali İrfan Mahmutoğulları

Endüstri Mühendisliği Yüksek Lisans

Tez Yöneticisi: Doç. Dr. Bahar Yetiş Kara

Temmuz 2013

Ana dağıtım üsleri (ADÜ) her noktadan diğer her noktaya akışın olduğu ağlarda toplama ve dağıtma noktalarıdır. ADÜ yer seçimi problemi, ADÜ'lerin yerlerinin belirlenmesi ve ADÜ olmayan noktaların bu ADÜ'lere atanması olarak tanımlanmaktadır.ADÜ'lerin kullanıldığı sektörlerde çok sayıda firma rekabet halinde olsa da ana akım ADÜ yer seçimi çalışmaları tek karar vericinin amaç fonksiyonları üzerinde yoğunlamıştır. Bu tezde iki karar vericinin olduğu bir ADÜ yer seçimi problemi incelenmiştir. Karar vericiler sırayla ADÜ yerlerini seçmekte ve müşteriler sağlanan hizmet seviyelerine göre bunlardan birini tercih etmektedir. Karar vericiler kendi pazar paylarını enbüyüklemeye çalışmaktadır. Ekonomi, perakende yer seçimi ve yöneylem araştırması alanlarındaki çalışmalarının incelenmesinin ardından lider (ilk karar verici) ve takipçi (sonraki karar verici) için iki farklı problem tanımlanmıştır. Problemler kombinatoriyal eniyileme problmeri olarak tanımlanmış ve karmaşıklık sınıflarının NP-zor olduğu ispatlanmıştır. Bu problemler için doğrusal modeller sunulmuştur. Ayrıca takipçinin problemi için doğrusal modelden daha az bilgisayar hafızası ve çalışma süresine ihtiyaç duyan kesin çözüm algoritmalar geliştirilmiştir. Modeller ve algoritmaların performansı ADÜ çalışmalarında sıkça kullanılan iki veri kümesi üzerindeki sayısal çalışmayla incelenmiştir.

Anahtar sözcükler: ADÜ yer seçimi, rekabet modelleri, rekabet ortamında yer seçimi.

Annem'e…

ACKNOWLEDGEMENT

I would like to express my gratitude to Assoc. Prof. Bahar Yetiş Kara for her guidance and support during my graduate study. Without her supervision, this thesis would not be possible.

I am also grateful to Prof. M. Selim Aktürk and Assoc. Prof. Sinan Gürel for accepting to read and review this thesis. Their comments and suggestions have been invaluable.

I also would like to thank my mother Zehra and sister Özlem for their endless love and support every day of my life. My father passed after two months I started Bilkent. I wish he was able to see this thesis.

Without existence of some friends, I could not succeed. Haşim Özlü, Nur Timurlenk, Okan Dükkancı and Cemal İlhan have helped me whenever I need since the very first day I moved to Ankara. I have always felt their unconditional support and trust to me.

Feyza Güliz Şahainyazan, Görkem Özdemir and Fırat Kılcı have always been *so close no matter how far* they physically are right now. Their help and friendship have been precious since the first day of my graduate study.

I also would like to thank dear friends and officemates Bengisu Sert, Başak Yazar, Gizem Özbaygın, Meltem Peker and Hatice Çalık. We have so many enjoyable moments together.

I would like to acknowledge financial support of The Scientific and Technological Research Council of Turkey (TUBİTAK) for the Graduate Study Scholarship Program.

Finally, I would like to thank Halenur who carries her beauty and goodness wherever she touches. Every single line of this thesis became possible with her love and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Hubs are consolidation and dissemination points in many-to-many flow network systems. Consolidation generates economies of scale and thus unit transportation cost is reduced between hubs. Hubbing also reduces the number of required links to ensure that each flow is routed to its desired destination. Many applications benefit from hub networks such as airline, cargo and telecom industries. Hub location problem is deciding the location of hubs and allocation of non-hub nodes to the hubs with respect to a given allocation structure and an objective such as minimizing the system-wide operating cost.

The facility location literature can be categorized into three groups regarding decision space of the problem. Planar models assume that demand points (or customers) are spread over a plane. The facilities can be located anywhere on this plane. In network models, demand points are regarded as nodes of a graph and facilities can be located on nodes or edges. The third category is discrete models where both demand points and available sites for facilities are nodes of a graph. The majority of hub location literature falls into discrete facility location category due to practical reasons in cargo, air transportation and telecommunication.

Today, many industries are ruled by a few numbers of competing firms. Such a market is called as an oligopoly (from Greek words *oligoi*: few and *polein*: to sell). Hence, market share and profit of a firm is affected by the decision made by itself and other competitors in the market. Also, customer behavior is another concern in oligopolistic markets. Market share is affected by the criteria that customers prefer one firm among others. Competition in oligopolies has been studied by economists to observe optimal decisions (including location) of each competing firm. However, studies considering competition in hub networks are rare in the literature.

Widely speaking, the hub location problem is to determine the location of hubs with respect to a given objective (or at least two objectives in existence of multi-objective optimization problems). A single decision-maker can determine the locations of hubs depending on the parameters: amount of flow and cost of distance between each pairs of nodes, interhub transportation discount factor, allocation strategy (single- or multi-allocation and structure of the network (incomplete, star network etc.). However, in a competitive environment a decision-maker should also consider the decisions made by his/her rivals and the preference of customers. In this study, we consider a *duopolistic market* -a special case of oligopoly- where the number of operating firms is two. The one who makes the location decision formerly is called as *the leader* and the other one is *the follower*.

Then by combining retail location from marketing, spatial competition in economics and location theory in operations research, in this thesis, we propose a discrete Stackelberg hub location problem where each firms makes decisions sequentially. Each decision-maker (or firm) decides the location of hubs and allocation of non-hub nodes to the hubs considering market share maximization.

Chapter 2 presents competitive location and hubbing in the literature. In chapter 3, we propose *(r|*X_p*) hub-medianoid* and *(r|p) hub-centroid* problems as combinatorial optimization problems. In chapters 4 and 5, mathematical models, complexity results, solution techniques and computational experiments for *(r|*X_p*) hub-medianoid* and *(r|p) hub-centroid* are presented, respectively. Finally, a general discussion and possible future researches related with these competitive hub location problems are discussed in Chapter 6.

Chapter 2

Competitive Location and Hubbing in the Literature

In this chapter, we present the literature of competition in economics, competitive location models and hub location problem. Then, we investigate hub location studies in which competition is considered.

2.1 Competition in Economy & Competitive Location Models

Competitive models in economy date back to 19^{th} century. The book *Recherches sur les principes mathématiques de la théorie des richesse (Researches into the Mathematical Principles of the Theory of Wealth)* published in 1838 by Cournot is the pioneering study in the competition in economics [1]. Cournot, a French economist, considers two competing firms operating in the same market. The firms decide the amount of production of a single product. The demand of the product is not known *a priori* and depends on the total amount of production. Hence, the profit of a firm depends on the amount of production made by itself and the competitor firm. Following Cournot's study, another French economist Bertrand considers a duopoly model where the competitors decide the price of a single product in *Theorie mathematique de la richesse sociale (Mathematical theory of social wealth)* published in 1883 [2]. In Bertrand's model the total demand is known before the decisions and each of the firms aims to maximize its market share or equivalently its total revenue. He considers that each customer prefers the firm that offers lower price to the product.

Hotelling presents first competitive model that includes location decisions in 1929 [3]. He considers the location and price decisions of two ice cream vendors operating on a beach. Each customer prefers the vendor that offers lower cost. Cost function includes the price of the ice cream and a linear function of transportation. The demand is assumed to be uniformly distributed on a line segment.

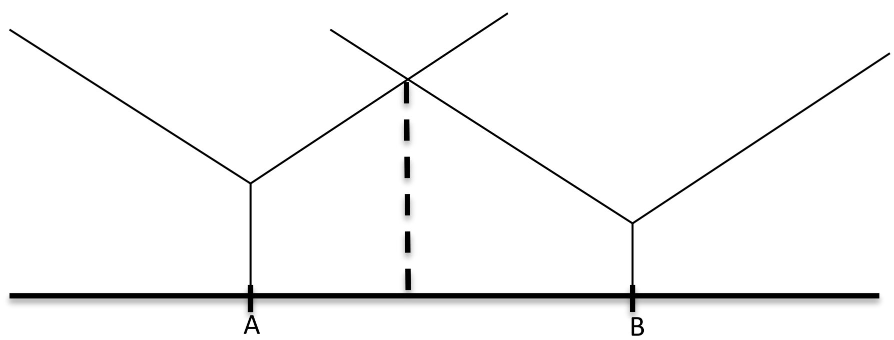


Figure 2-1: Hotelling's model on a line segment

Figure 2-1 is an illustration of Hotelling's model. The "Y-shaped" functions indicate the cost of buying the ice cream form the vendors *A* and *B*. The region between the locations of *A* and *B* is called as *competitive region*. The segment that is located on the LHS of vendor *A* is called *A's hinterland*. Similarly, the segment located on the RHS of the location of *B* is called as *B's hinterland*. The dashed line indicates the point where customers are indifferent between both vendors. The customers located in the hinterland of *A* and LHS of the dashed line in the competitive region prefers the vendor *A*. Moving *A*'s position towards to *B* will increase its market share since *A's hinterland* increases.

Similarly, *B* has incentive to move towards *A*. The solution of Hotelling's model is found as location of both vendors clusters at the center of the market with equal prices. At this equilibrium point, each vendor captures half of the market.

Hotelling's pioneering work attracted many researchers. Later, Lerner and Singer consider the same model with more than two competitors [4], Smithies considers different customer behaviors [5], and Eaton and Lipsey investigate the equilibrium point on a plane rather than a line segment [6].

In duopoly models presented by Cournot, Bertrand and Hotelling, the decisions of the two firms are made simultaneously. Solutions to these kinds of simultaneous decisions are called as *Nash equilibrium* (sometimes called as *Cournot-Nash equilibrium*), after John F. Nash's great contributions to Game Theory [7]. A *Nash equilibrium* is a decision vector of all decision-makers where no one can achieve a better objective by changing his/her decision given that other decision-makers do not deviate from their current decisions.

Another streamline research in competitive models deals with not simultaneous but sequential decision making process. The preliminary work of sequential decision making of location is first proposed in Stackelberg's book *Grundlagen der theoretischen Volkswirtschaftslehre (The Theory of Market Economy)* published in 1943 [8]. Since sequential decision making results in an asymmetry between decision makers, we need to differentiate the identities of decision makers. Stackelberg considers a duopoly where the firm that makes the initial decision is called *the leader* and the other one as *the follower*. Stackelberg's model has three major assumptions:

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

If leader's decisions are given, the follower's decisions are made while optimizing his/her own objective. These decisions are called as *reaction function* of the follower. Since both parties have the complete information of the system, the leader observes the reaction function of the follower. Hence, leader gives the 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 as an input to the leader's decisions. Bard [9] and Dempe [10] give detailed discussion on bilevel programming models and solution techniques.

Teitz is the first non-economies scholar who studies sequential location on a line segment in 1968 [11]. His findings are similar to Hotelling's observations. Moreover, Teitz considers the extension of Hotelling's model by allowing that each decision maker locates more than one facility. First, Tietz proposes a sequential location model, where one firm, say *A*, locates two facilities, but the other firm, say *B*, locates only one facility. The decisions are made by based on short-term maximum gain (referred as "*conservative maximization"* by Tietz) and continue until equilibrium point is found. *A* moves first and relocates one of his/her facilities, and then *B* moves and relocates his/her facility. Later, *A* relocates, then *B* and so on *ad infinitum*. Tietz claims that in such a model the equilibrium point is clustering at the center where at each turn the order of facilities change (for example *AAB*, *ABA*, *BAA*, *ABA*). At this *dancing equilibrium point*, *A* gets ¾ of the market where *B*'s share is ¼. Tietz generalizes his results for the case where *A* has *n* facility and *B* has only one. Then, in resulting equilibrium, *A* gets *(2n-1)/2n* of the market.

Although sequential location models have been studied by economists until 1980s, this topic also attracted OR specialists attention. Some predate works were proposed by Wendell and Thorson [12]; Slater [13]; Wendell and McKelvey [14]; and Hansen and Thisse [15].

Drezner [16] and Hakimi [17] independently propose sequential location problems with an OR point of view and attracted the community's attention in 1982 and 1983, respectively. They both consider the same competitive environment but decision space is the only difference in their studies. While Drezner considers the decision space as a plane, Hakimi deals with network models. Their problem includes a number of customers with *inelastic demand*, that is, the amount of demand of each customer is known *a priori* and does not affected by the decisions of leader and follower. The customers prefer the closest facility to buy a homogenous product. The decision-makers act sequentially, first leader locates p facilities and then the follower locates r facilities.

In order to describe Drezner and Hakimi's contributions, following conventions are necessary. Assume that n customers (or demand points) are located on points $V=\{v_1,v_2,\dots,v_n\}$. The demand of customer i is $w(v_i)$. Let $D(v,z) = min\{d(v,z) : z \in Z\}$ where $d(v,z)$ is the distance between v and z for any subset of points $Z \subseteq V$. The distance between two points is Euclidean distance in 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 $X_p=\{x_1,x_2,\dots,x_p\}$ and $Y_r=\{y_1,y_2,\dots,y_r\}$ respectively. A customer v_i prefers the follower if $D(v_i,Y_r) < D(v_i,X_p)$. Then, the demand captured by the follower can be defined 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 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 the realistic assumption that the follower will respond by *$(r|X_p)$ medianoid*.

Drezner initially considers a Stackelberg location model where both leader and follower locate one facility each, that is $p = 1$ and $r = 1$. Figure 2-2 is an example of Drezner's model. The demand points are marked with dots on the plane and the leader and the follower located their facilities on point X and Y, respectively. Draw the hyperplane that is perpendicular to the line segment connecting X and Y at the center of the segment. Since customers prefer the closest facility, customers that are on the same half-plane with X prefer the leader and the remaining ones prefer the follower. Then, if X is given, the follower prefers a location Y that is as close as possible to X so as to capture more customers. Hence, the optimal location for the follower's facility can be searched at the points that are infinitesimally close to X.

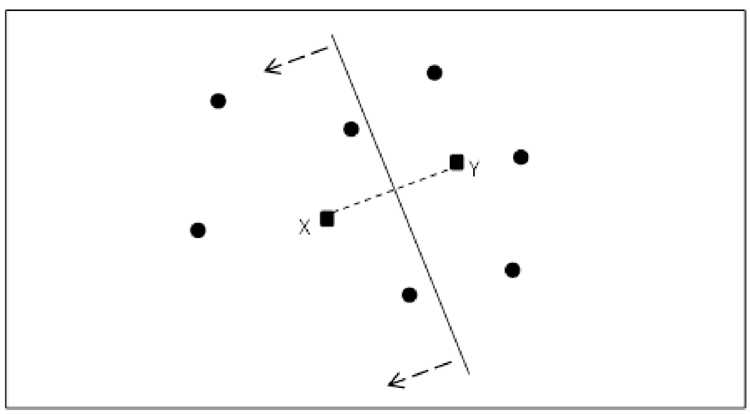


Figure 2-2: An example of Drezner's model when $p = 1$ and $r = 1$

Since the demand points are finite, Drezner presents an algorithm that find the $(1|X_1)$ medianoid in $O(n \log n)$ time. His algorithm is based on sorting the total amount of demand captured by the follower rotating the position of Y along a circle with center X and an infinitesimal radius. The (1|1) centroid problem is more challenging. Realizing the follower's optimal response the leader positions X so as to maximize his/her market share. Drezner provided an $O(n^4 \log n)$ algorithm for the leader's problem that utilizes the intersections of hyperplanes for all pairs of demand nodes. Drezner did not study the cases where $p > 1$ and $r > 1$. He provided that when $p = 1$ and $r > 1$, the (r|p) centroid is on the node with highest demand since the follower can sandwich the leader's location. He also proposed an $O(n^2 \log n)$ algorithm to solve for (r|p) centroid when $p > 1$ and $r = 1$.

Hakimi proposes medianoid and centroid problems on a network rather than in a plane [17]. He first tries to find general properties of medianoid and centroids by working on some illustrative examples. Hakimi realizes that the centroid problem can be considered as a minimax problem since the leader aims to minimize the amount of demand captured by the follower. However, center and centroid do not necessarily coincide. In Figure 2-3, each node has same demand value. The point A is *1-center* of the tree, but not *(1|1) centroid*. Moreover, the point B is a *(1|1) centroid*, but not a *1-center*.

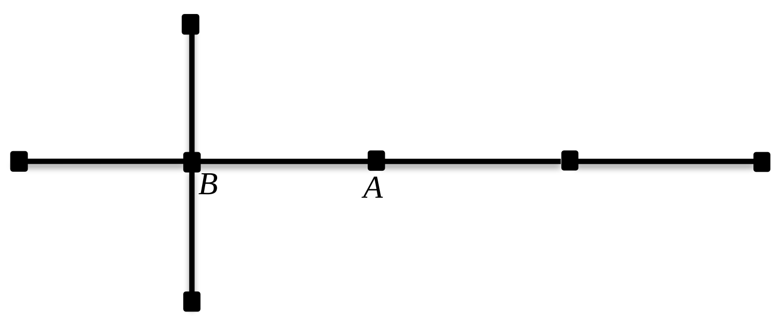


Figure 2-3: An example where 1-center and 1-centroid do not coincide

Hakimi also investigates the existence of node optimality of medianoid. His findings also reveal that a medianoid is not necessarily to be a node on the network. In figure 2-4, the leader has already located a facility at one of the edges of an equilateral triangle where the total demand is equally distributed over the vertices. Then, the follower can capture the two-third of the total demand by locating a facility at the center of the side that is opposite to X_1.

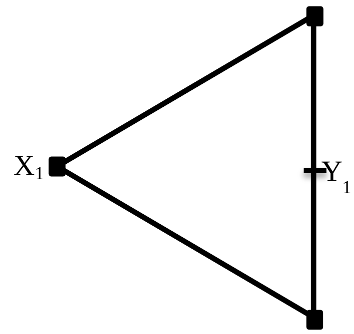


Figure 2-4: An example where a 1-medianoid is not an vertex

Hakimi cannot find special characteristic of centroids and medianoids. He only proposed that a 1-centroid of a tree network coincides with 1-median. Later, Hakimi proves for the proportional demand case where the leader and the follower proportionally capture the demand of a customer with respect to distance, node optimality exists [18].

Hakimi later proves that both centroid and medianoid problems are NP-hard. His proofs are based on reduction of *the dominating set problem* and *the vertex cover problem* to finding $(r|X_1)$ *medianoid* and $(1|p)$ *centroid problems*, respectively.

Drezner and Hakimi's influential works attracted many academicians attention in the last three decades. However, most of the studies focus on the medianoid problems since centroid problem is still challenging. ReVelle propose an integer programming formulation for the discrete medianoid problem [19]. His formulation, namely MAXCAP, is based on the maximization of the total demand captured by the follower. In his model, the follower choses p facilities among the set of possible sites J. The follower captures the whole demand a_i of node $i \in I$ if he/she can provide a strictly better service level than the follower. If both decision-makers have same service level for a customer demand is shared equally between both firms. The decision variables used by ReVelle:

y_i = 1 if some server is closer to i than its previous closest server for $i \in I$, and 0 otherwise;

z_i = 1 if node i is captured by a server within K_i, that is , at the currently closest server to i or at a site whose distance from i is equal to the distance from i to its currently closest server for $i \in I$, and 0 otherwise;

x_j = 1 if a facility is sited at $j \in J$, and 0 otherwise;

The MAXCAP model is as follows:

$$\textit{maximize} \quad \sum_i a_i y_i + \sum_i (a_i/2) z_i \tag{2.1}$$

$$\textit{subject to} \quad y_i \leq \sum_{j \in N_i} x_j \qquad \forall i, \tag{2.2}$$

$$z_i \leq \sum_{j \in K_i} x_j \qquad \forall i, \tag{2.3}$$

$$y_i + z_i \leq 1 \qquad \forall i, \tag{2.4}$$

$$\sum_j x_j = p \tag{2.5}$$

$$x_j, y_i, z_i \in \{0,1\} \qquad \forall\, i \text{ and } j \tag{2.6}$$

where N_i is the subset of possible sites that are strictly closer to the demand point i than the closest facility of the leader. Similarly, K_i is the subset of possible sites that are equally close to the demand point i with the closest facility of the leader. The objective (2.1) maximizes the captured demand where constraints (2.2), (2.3) and (2.4) determine whether the demand is totally captured, partially captured or lost. Constraints (2.5) limit the number of opened facility to p. Constraints (2.6) are domain constraints.

Later, Eiselt and Laporte extend the MAXCAP formulation by introducing attraction functions [20]. Serra et al. [21], and Serra and Colome [22] solve the MAXCAP model for partial demand preferences. Later, Benati addresses the sub-modularity of the objective function of MAXCAP [23], and Benati and Hansen demonstrates that the problem can be modeled as a *p-median* type problem [24].

For centroid models that are more challenging than medianoid not so many results are obtained so far. Most remarkable work is proposed by Hansen and Labbe [25]. They propose an algorithm that solves (1|1) centroid problem in polynomial time. The algorithm runs in $O(n^2m^2 \log mn \log D)$ time where n is the number of nodes, m is the number of edges and D is the total demand on the network. For $p,r > 1$ no algorithm is available that runs efficiently. Serra and ReVelle propose two heuristic methods based on the response of the follower for every action of the leader [26].

An interested reader may refer to surveys by Eiselt and Laporte [27] and Daşçı [28] for a detailed discussion for competitive location problems.

2.2 Hub Location Problem

Hubs are special kinds of facilities that are consolidation and dissemination points in many-to-many flow network systems. The flow originating from a point visits one or two hubs before arrival its destination. Since the links between hubs carry high volume of flows, economies of scale is generated on these hubs links and transportation cost (or time) attribute is discounted by a factor α ($0 \leq \alpha \leq 1$).

The hub location problem is to decide locations of hubs and allocations of non-hub nodes to the hubs to optimize a given objective. Two different allocation strategies are considered in the hub location literature. In single-allocation models, the whole incoming and outgoing flow of a node is transferred via a single hub. In multi-allocation case different hubs can be used for transferring the flow of a node.

Figure 2-5 gives an example of cost structure in a hub network where squares indicate hubs. Flow from node i to node j first visits the hub to which node i is allocated, say k.

Then, the flow is sent to destination's hub, say *m* and finally to the destination which is node *j*. The total cost of sending one unit of flow from origin to destination consists of collection, transfer and distribution costs where interhub transfer cost is discounted. In the example depicted in Figure 2-5 cost of sending one unit of flow from node *i* to node *j* is equal to $c_{ik} + \alpha c_{km} + c_{mj}$.

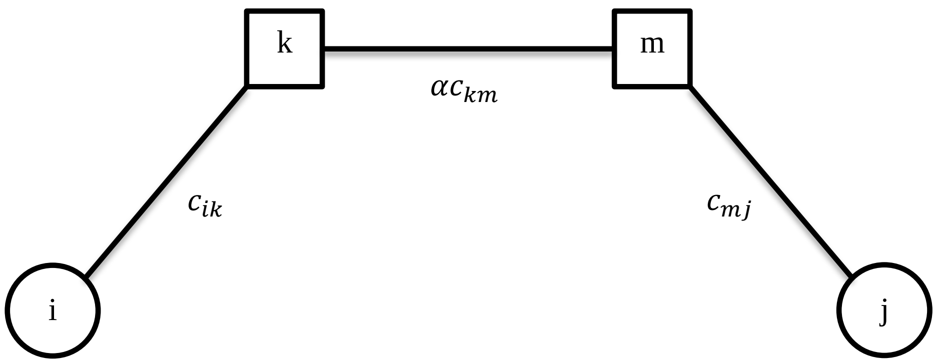


Figure 2-5: Cost structure in a hub network

O'Kelly presents the hub location problem where the system-wide transportation cost of the network is minimized by locating *p* hubs in a single-allocation structure (This problem is later referred as *single allocation p-hub median problem*) [29,30].

O'Kelly also proposed the first mathematic formulation of the single-allocation p-hub median problem [31]. Define x_{ik} as *1* if node is allocated to hub *k* and *0*, otherwise. If a hub is located at node *k*, then $x_{kk} = 1$. With these parameters and decision variables, O'Kelly proposes his model as follows:

$$\textit{minimize} \quad \sum_i \sum_j w_{ij} \left(\sum_k x_{ik} c_{ik} + \sum_m x_{jm} c_{jm} + \alpha \sum_k \sum_m x_{ik} x_{jm} c_{km} \right) \tag{2.7}$$

$$\textit{subject to} \quad (n - p + 1) x_{kk} - \sum_i x_{ik} \geq 0 \qquad \forall k, \tag{2.8}$$

$$\sum_{k} x_{ik} = 1 \qquad \forall i, \tag{2.9}$$

$$\sum_{k} x_{kk} = p\ , \tag{2.10}$$

$$x_{ik} \in \{0,1\} \qquad \forall i \text{ and } k. \tag{2.11}$$

The quadratic objective (2.7) minimizes the total collection, distribution and transfer cost of the system. Constraint (2.8) ensures that a node is not allocated to a non-hub node. Constraints (2.9) guaranties that each node is allocated to a single hub. The total number of hub to be opened is *p*, as in constraint (2.10). Constraint (2.11) is the binary constraint. Constraint (2.8) can be replaced with the following corresponding constraint.

$$x_{ij} \leq x_{jj} \qquad \forall i \text{ and } j \tag{2.12}$$

Campbell proposes the first linear formulation for the *single allocation p-hub median problem* with n^4+n^2+n variables of which n^2+n are binary and n^4+2n^2+n+1 constraints [32]. Later Skorin-Kapov et al. provides a new linear model with n^4+n^2 variables of which n^2 are binary and $2n^3+n^2+n+1$ constraints [33]. Their model is as follows:

$$\textit{minimize} \qquad \sum_{i}\sum_{j}\sum_{k}\sum_{m} w_{ij} x_{ijkm}\left(c_{ik} + \alpha c_{km} + c_{mj}\right) \tag{2.13}$$

subject to (2.9)-(2.12)

$$\sum_{m} x_{ijkm} = x_{ik} \qquad \forall i, j \text{ and } k, \tag{2.14}$$

$$\sum_{k} x_{ijkm} = x_{jm} \qquad \forall i, j \text{ and } m, \tag{2.15}$$

$$x_{ijkm} \geq 0 \qquad \forall i, j, k \text{ and } m. \tag{2.16}$$

where x_{ijkm} is the fraction of flow from node *i* to node *j* that is transferred via hubs *k* and *m* in that order.

Ernst and Krishnamoorthy model the *single allocation p-hub median problem* as a multi-commodity flow problem [34]. Their formulations require n^3+n^2 variables of which n^2 are binary and $2n^2+n+1$ constraints. The formulation is as follows:

$$\textit{minimize} \quad \sum_i \sum_k c_{ik} x_{ik} (\chi O_i + \delta D_i) + \sum_i \sum_k \sum_l \alpha c_{kl} Y_{kl}^i \tag{2.17}$$

subject to (2.9)-(2.12)

$$\sum_l Y_{kl}^i - \sum_l Y_{lk}^i = O_i x_{ik} - \sum_j w_{ij} x_{jk} \qquad \forall i \text{ and } k, \tag{2.18}$$

$$Y_{kl}^i \geq 0 \qquad \forall i, k \text{ and } l. \tag{2.19}$$

where Y_{kl}^i is the amount of flow originating from node i and visits hubs k and l in that order. χ and δ are unit collections and distribution costs, respectively. Constraint (2.18) is the flow balance constraint that ensures that each flow is transferred to its destination via one or two hubs.

The *single allocation p-hub median problem* is NP-hard. Kara proves that even if the hub locations are given, the remaining allocation decisions are still cannot be solvable with an algorithm that runs in polynomial time [35]. Since the problem is NP-hard, heuristics are widely used to come up with a promising solution to the *single allocation p-hub median problem* such as [31].

The *multi-allocation p-hub median problem* has also attracted attention. Campbell presents the first multi allocation hub model [36]. His formulation is as follows:

minimize (2.13)

subject to (2.10),(2.11) and (2.16)

$$\sum_k \sum_m x_{ijkm} = 1 \qquad \forall i \text{ and } j, \tag{2.20}$$

$$x_{ijkm} \leq x_{kk} \qquad \forall i, j, k \text{ and } m, \tag{2.21}$$

$$x_{ijkm} \leq x_{mm} \quad \forall i, j, k \text{ and } m. \tag{2.22}$$

Campbell also states that in absence of capacity each x_{ijkm} has a value of 0 or 1 since each flow travels the least cost path through opened hubs. Later, Skorin-Kapov develop a linear model with n^4+n variables of which n are binary and $2n^3+n^2+1$ constraints by aggregating constraints (2.21) and (2.22) [33]. Ernst and Krishnamoorthy model the *multi-allocation p-hub median problem* based on the idea that they use for the single-allocation version of the problem [37]. Their model requires $2n^3+n^2+n$ variables of which n are binary and $4n^3+n+1$ constraints.

Some heuristic models are also improved to solve *multi-allocation p-hub median problem.* Some examples are can be found in the studies proposed by Campbell [38], Ernst and Krishnamoorthy [37], and Boland et al. [39].

Although hub location problem under median objective constitutes the main streamline of the literature, other types of objectives are also investigated by the researchers. Another channel of research on hub location problem is the *hub location problem with fixed costs*. In the structure of the problem the number of hubs to be opened is exogenous. The constraint on the number of hubs - constraint (2.10) - is removed and a fixed cost of opening a hub at node k, say f_k, is included in the objective function. O'Kelly [40] and Campbell [36] propose mathematical models for the *hub location problem with fixed costs* where capacities of hubs are ignored. The capacity constraints can be included in *hub location problem with fixed costs* to ensure that the total flow throughout a hub does not exceed a threshold value. Campbell [36] and Aykin [41] present mathematical models for the capacitated version of the problem.

In some applications of hub networks, for example in cargo applications where the cargo should be delivered within a 24-hours period, not only the cost but also service levels are considered. The *p-hub center problem* is to locate p hub on a network to ensure that the distance or cost between the most disadvantageous pair of nodes does not exceed a given cover radius.

Campbell proposes the first linear model for the hub location problems with center-type objectives [32]. Kara and Tansel prove that the *p-hub center problem* is NP-hard by using reduction from the *dominating set problem* [42]. They also propose different mathematical models for the model. Later, Ernst et al. provide 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 [43].

Hub covering problem is another version of the hub location problem. There are two types of hub covering problem: *Hub set covering problem* and *maximal hub covering problem*. *Hub set covering problem* is to minimize the number of hubs to be located by ensuring that the distance or cost between each O-D pair does not exceed a given threshold value. On the other hand, *maximal hub covering problem* is to maximize the total demand that are covered by a given number of hubs.

Campbell is the first researcher who presents mathematical models for different types of hub covering problem [32]. After his contribution, Kara and Tansel study *single allocation hub set covering problem* and propose three different linearizations of the problem [44]. They also worked on the complexity of the problem and conclude that the problem is NP-hard. Later, Ernst et al. provide new formulations of the problem based on the idea that they use for the *p-hub center problem* [45].

An interested reader may refer to surveys by Campbell et al. [46], Alumur and Kara [47] and Kara and Taner [48] for a detailed discussion of hub location problems.

2.3 Hub Location with Competition

Although the competition in location decisions has been studied in detail, competitive hub location studies in the literature are rare. The first hub location problem with competition is proposed by Marianov et al. [49]. They propose mathematical models for the follower's problem where the leader has already been operating the market with existing hubs. First, they assume that the follower will capture the whole demand between nodes i and j if he/she can provide a better or equal service level than the leader. The idea is based on defining a capture set $N_{ij} = \{(k,l) : c_{ik} + \alpha c_{km} + c_{lj} \leq C_{ij}\}$ for all pair of nodes i and j where C_{ij} is the current service level provided by the leader. The number of hubs to be opened by the follower is restricted by p. However, they relax this assumption by redefining the objective function as the total profit made by captured flow and the fixed cost of opening a hub.

They also consider proportional capture levels instead of all-or-nothing type capture. For example, they assume that the leader capture half of the flow between nodes i and j if his/her service levels is between $0.9C_{ij}$ and $1.1C_{ij}$, three-fourth of the flow if his/her service levels is between $0.7C_{ij}$ and $0.9C_{ij}$ and captures the whole flow if his/her service level is less than $0.7C_{ij}$. Then, the capture sets are redefined as N_{ij}^{50}, N_{ij}^{75} and N_{ij}^{100} for the capture levels 50%, 75% and 100%. The mathematical model is provided for the proportion capture case by triplicating the capture variables and constraints. However, due to large number of constraints and variables it is hard to get an optimal solution within reasonable time.

The authors propose a meta-heuristic to solve the problem on AP data set. The heuristic consists of three steps. First, an initial solution is generated by opening hubs based on the marginal improvements obtained by opening a hub at a specific node. Later, a heuristic is used to improve the objective by relocating one hub at each iteration. Finally, to prevent the trap on a local optimal a tabu-search heuristic is used. The efficiency of the heuristic is tested by the randomly generated instances and AP data set with 20, 25, 40 and 50 nodes. They point out that the heuristic yields an optimal solution in most of the instances within seconds. It is also stated that, the LP relaxation of the model and the branch and bound technique does not yield a fast solution even if the number of nodes is 20.

Wagner criticizes the study by Marianov et al. on his note [50]. Wagner states that the capture sets should be redefined since the follower captures the whole demand in case of a tie when all-or-nothing type capture is considered. When the number of hubs to be opened is equal to the number of existing hubs the follower can get the whole market by location the hubs at the location of existing hubs. Then, Wagner proposes a new capture set where the follower gets nothing in case of equal service levels. Wagner was also able to solve the model optimally up to 50 nodes by eliminating some redundant routes that visit two hubs.

Sasaki and Fukushima propose a new kind of competitive hub location model where the decision space is a plane [51]. The 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 state that the problem has a Stackelberg model due to its sequential decision structure.

Sasaki and Fukushima uses logit functions for customer preferences to express the proportional capture in their model. They initially model the problem as a bilevel program and use *sequential quadratic programming* approach that updates the Hessian at each iteration to solve the problem. They conduct computational experiments on CAB data set and conclude that the big firm gets the highest market share with the advantage of first move.

Sasaki applies the same idea in the study by Sasaki and Fukushima to a discrete environment with some modifications [52]. Her model includes two decision-makers: one leader and one follower. The leader and the follower locate p and q hubs on the network, respectively. The capture rule of the customer is similar to their previous study and each route contains one hub again. In her new problem environment, Sasaki also considers a threshold value of the captures amount of flow. Her solution methods are complete enumeration and a greedy heuristic that does not perform very well in terms of CPU time when $p < q$.

Eiselt and Marianov propose another hub location model with competition where an airline transportation company enters a market [53]. It is assumed that some other companies already operate the market. The entrant firm aims to capture as much customer as possible. Customers' preferences are based on the basic attractiveness of the firms (such as safety record, personal space, quality of the foods etc.), the number of stopover on the trip, cost of the route and time required by the flight. These factors are converted to an attraction function by using a Huff-like model. So, the fractional capture is allowed. They propose a nonlinear mathematical model of the problem which is solved with a two phase meta-heuristics. The first phase, set of available sites is restricted to a smaller set, called as *concentration set*, and an initial solution is obtained. This initial solution is improved in the second phase by relocation the hubs of the follower.

They test the meta-heuristics with the 25-node version of the AP data set. The computational analysis reveals that follower has a great advantage in the most of the instances and is able to capture 70% of the total customers. Eiselt and Marianov later used a 50-node version of the same data set; however they were not able to solve the problem within reasonable CPU times.

Another hubbing problem with Stackelberg competition is studied by Sasaki et al. [54]. In their problem environment, the decision-makers do not locate hubs but they locate 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 the flow between any O-D pair based on the cost and the travel time of the trip and the remaining customers prefer the follower. They propose a bilevel program of the model and use a smart complete enumeration scheme that does not perform quick solutions for the instances with large q^a and q^b to solve the problem. CAB data set is used to test the efficiency of the solution technique. The authors conclude that the geography plays an important role on the location of hubs in the competitive environment.

Although existing studies contribute to hub location and competition literature, both theoretical aspect of the problem and application in industry required much more effort. Therefore, in this thesis, we formally define *hub-medianoid* and *hub-centroid* problems by following the terminology used by Hakimi [17] for the analogous competitive location problems in order to motivate the studies in this area. Moreover, we prove that both problems are NP-hard.

The following table summarizes the paper mentioned above with their contribution to the literature. The last row indicates the contributions made by our study.

Table 2-1: Summary of competitive hub location literature

Paper	Decision Space	Market Type	Decisions	Capture Type	Conputational Study	Solution Techniques	Contribution
Marianov et al. [49]	Network	Duoploy	Follower's hubs	Partial Discrete	Random & AP(20,25,40,45)	Heuristic	Competitive hub location model
Wagner [50]	Network	Duoploy	Follower's hubs	Partial Discrete	AP(50)	MIP	First exact solution for moderate size instances
Sasaki and Fukushima [51]	Plane	Oligoploy	One hub for leader and each follower	Partial Continuous	CAB(25)	SQP	Oligopolistic market where followers cooperate, first Stackelberg competition
Sasaki [52]	Network	Duoploy	Leader and Follower's hubs	Partial Continuous	CAB(25)	Enumeration & Heuristic	Application of [51] to a network
Eiselt and Marianov [53]	Network	Oligoploy	Follower's hubs	Partial Continuous	AP(25)	Heuristic	Various customer preferences considered
Sasaki et al. [54]	Network	Duoploy	Leader and Follower's hubs	Partial Continuous	CAB(25)	Enumeration	Bilevel model, exact solutions for small size instances
Mahmutoğulları and Kara	Network	Duoploy	Leader and Follower's hubs	Binary Discrete	CAB(25) & TR(81)	MIP & Enumeration	Formal definition, complexity results, exact solutions for large size instances

Chapter 3

Problem Definition

Given a network $G=(N,E)$ where N is the set of nodes and E is the set of edges, let w_{ij} be the flow between nodes i and j for all $i,j \in N$ and c_{ij} be the transportation cost of a unit flow from node i to node j for all $i,j \in N$. The interhub transportation cost is discounted by a factor α, $0 \leq \alpha \leq 1$. (We use $<G=(N,E), w_{ij}, c_{ij}, \alpha>$ nomenclature to refer this many-to-many flow network in the remainder of the thesis.) The leader and follower want to enter a market with prespecified number of hubs. Say p and r be the number hubs to be opened by the leader and follower, respectively. We assume that both p and r are grater or equal to 2 since otherwise there is no interhub link and economies of scale is not generated. Let $H \in N$ be the subset of nodes that are available to locate a hub. The 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 less 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 has already operating with the leader when the follower enters the market and the customer has no incentive to deviate from the current position.

First, assume that the leader has already 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 originated from node i visits one or two hubs before arrival to its destination node j. Therefore, we can easily compute the service level, say β_{ij}, provided by the leader for the flow between nodes i and j.

$$\beta_{ij} = \min_{k,m\in X_p}\{c_{ik} + \alpha c_{km} + c_{mj}\} \tag{3.1}$$

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, follower's service levels, say γ_{ij}, for all node pairs i and j can be calculated as:

$$\gamma_{ij} = \min_{k,m\in Y_r}\{c_{ik} + \alpha c_{km} + c_{mj}\} \tag{3.2}$$

The flow w_{ij} is captured by the follower if $\gamma_{ij} < \beta_{ij}$. Given that the leader and follower's hubs are located on the subset of nodes X_p and Y_r, respectively, the total flow captured by the follower can be expressed by a function $f : \mathscr{P}_p(H) \times \mathscr{P}_r(H) \rightarrow [0,W]$ such that

$$f(X_p, Y_r) = \sum_{i,j\in N:\gamma_{ij}<\beta_{ij}} w_{ij} \tag{3.3}$$

where $\mathscr{P}_p(H)$ is collection of subsets of H whose cardinalities are 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 wants to find the set Y_r that maximizes $f(X_p, Y_r)$ assuming the follower will respond (or act) rationally. Rational behavior means that the leader wants to capture more demand as more as he/she can.

We define set Y_r^* as *$(r|X_p)$ hub-medianoid* if $f(X_p, Y_r^*) \geq f(X_p, Y_r)$, $\forall Y_r \in \mathscr{P}_r(H)$. In plain words, *$(r|X_p)$ hub-medianoid* is the subset of nodes with r elements to locate hubs that maximizes the demand captured by the follower given the hub set of the leader.

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 $f(X_p^*, Y_r^*(X_p^*)) \leq f(X_p, Y_r^*(X_p))$, $\forall X_p \in \mathscr{P}_p(H)$ where $Y_r^*(X_p)$ is the *$(r|X_p)$ hub-medianoid* given X_p. To simplify, we can say that *$(r|p)$ hub-centroid* is the best choice of the leader's hub locations so that in the remaining scenario the follower can capture the least possible demand.

Chapter 4

$(r|X_p)$ Hub-medianoid Problem

In Chapter 3, we define the *$(r|X_p)$ hub-medianoid* problem as a combinatorial optimization problem from viewpoint of the follower. In this chapter, we provide linearization of the problem and prove that the problem is NP-hard by reduction from *clique* problem. Also, we present numerical analysis conducted to observe the efficiencies of the linear model.

4.1 Linearization of $(r|X_p)$ Hub-medianoid Problem

Let $<G=(N,E), w_{ij}, c_{ij}, \alpha>$ be a many-to-many flow network. At the time the follower makes the decision, the leader has already located his/her hubs and locations of these hubs are correctly observed by the follower. Assume that the leader have already located p hubs on the set $X_p \subseteq H$. 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 found as

$$\beta_{ij} = \min_{k,m \in X_p}\{c_{ik} + \alpha c_{km} + c_{mj}\} \tag{4.1}$$

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

h_k = 1 if the follower locates a hub on node $k \in H$, and 0 otherwise;

u_{ijk} = 1 if the flow from node $i \in N$ to node $j \in N$ visits hub $k \in H$ as the first hub, and 0 otherwise;

o_{ijm} = 1 if the flow from node $i \in N$ to node $j \in N$ visits hub $m \in H$ as the second hub, and 0 otherwise;

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

a_{ij} = 1 if the flow form node $i \in N$ to $i \in N$ is captured by the follower, and 0 otherwise;

The following mixed integer problem **H-MED0** correctly linearizes the *$(r|X_p)$ hub-medianoid* problem:

H-MED0

maximize $$\sum_i \sum_j a_{ij} w_{ij} \tag{4.2}$$

subject to $$\sum_k h_k = r, \tag{4.3}$$

$$\sum_k u_{ijk} = 1 \quad \forall\, i \text{ and } j \in N, \tag{4.4}$$

$$\sum_m o_{ijm} = 1 \quad \forall\, i \text{ and } j \in N, \tag{4.5}$$

$$u_{ijk} \leq h_k \quad \forall\, i,j \in N \text{and } k \in H, \tag{4.6}$$

$$o_{ijm} \leq h_m \quad \forall\, i,j \in N \text{ and } m \in H, \tag{4.7}$$

$$\gamma_{ij} \geq \sum_k u_{ijk}(c_{ik} + \alpha c_{km}) + c_{mj} - (1 - o_{ijm})M$$

$$\forall\ i, j\ \in N \; and \; m\ \in H, \tag{4.8}$$

$$\gamma_{ij} - \beta_{ij} + \varepsilon \leq (1 - a_{ij})M \quad \forall\, i \; and \; j\ \in N, \tag{4.9}$$

$$h_k, u_{ijk}, o_{ijm}, a_{ij} \in \{0,1\} \; and \; \beta_{ij} \geq 0$$

$$\forall\, i \; and \; j\ \in N, k \; and \; m\ \in H. \tag{4.10}$$

The objective (4.2) maximizes the amount of flow captured by the follower. Constraint (4.3) ensures that follower locates *r* hubs on the set of available nodes. Constraints (4.4), (4.5), (4.6) and (4.7) 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 (4.8) correctly calculate the service levels of the follower in the following manner: if $o_{ijm} = 0$, the constraint becomes redundant. However, if $o_{ijm} = 1$ the RHS of the constraint becomes the service level for flow from node $i \in N$ to $j \in N$. M is a large positive value but $M = (2 + \alpha) \max_{i,j} c_{ij}$ value is large enough since the RHS can be at most $(2 + \alpha) \max_{i,j} c_{ij}$ Let ε be very small positive number used to break ties in favor of the leader. Constraints (4.9) 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 equal to or worse than service level provided by the leader, the RHS of the constraint must be positive and $a_{ij} = 0$. Otherwise, the constraint becomes redundant. Constraints (4.10) are domain constraints.

We can eliminate decision variable γ_{ij} by combining constraints (4.8) and (4.9). Moreover, aggregating allocation variable o_{ijm} and capture variable a_{ij} we define a binary variable v_{ijm} such that

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

Then, the following mixed integer problem **H-MED** correctly linearizes the *$(r|X_p)$ hub-medianoid* problem with fewer variables and constraints than **H-MED0**:

H-MED

maximize $$\sum_i \sum_j \sum_m w_{ij} v_{ijm} \quad (4.11)$$

subject to $$\sum_m v_{ijm} \leq 1 \quad \forall\, i \; and \; j \in N, \quad (4.12)$$

$$v_{ijm} \leq h_m \quad \forall\, i, j \in N \; and \; m \in H, \quad (4.13)$$

$$\sum_k u_{ijk}(c_{ik} + \alpha c_{km}) + c_{mj} - \beta_{ij} + \varepsilon$$

$$\leq (1 - v_{ijm})M \quad \forall\; i, j \in N \; and \; m \in H, \quad (4.14)$$

$$h_k, u_{ijk}, v_{ijm}, \in \{0,1\} \quad \forall\, i, j \in N \; and \; k, m \in H. \quad (4.15)$$

(4.3), (4.4) and (4.6)

The objective function (4.11) maximizes captured flow by the follower. Constraints (4.12) ensure that flows from node $i \in N$ to node $j \in N$ can be captured by the follower at most once. Constraints (4.13) do not allow that the flow from node $i \in N$ to node $j \in N$ is captured via hub $m \in H$ unless m is a hub node. Constraints (4.14) determine the captured flows in the following manner: if the LHS of the constraint is non-negative, the corresponding variable v_{ijm} is forced to be 0; otherwise there is no restriction on v_{ijm} and together with the objective function its value is assigned to 1 which means that the follower can provide a strictly better service level than the follower for the flow from node $i \in N$ to node $j \in N$. Constraints (4.15) indicate that all variables can take binary values.

We can easily argue that **H-MED** correctly linearizes the *$(r|X_p)$ hub-medianoid* problem with fewer variables and constraints than **H-MED0**. The following table depicts the number of variables and constraints of both models where n is the number of nodes and m is the number of available nodes to locate hub in the network, that is $|N| = n$ and $|H| = m$.

Table 4-1: Comparison of H-MED0 and H-MED in terms of size of the models

Model	Number of Constraints	Number of Variables	
		Continuous	Binary
H-MED0	$3mn^2+3n^2+1$	n^2	$2mn^2+n^2+m$
H-MED	$3mn^2+2n^2+1$	--	$2mn^2+m$

4.2 Problem Complexity

We prove that the problem of finding a *$(r|X_p)$ hub-medianoid* is NP-hard and the corresponding decision problem is NP by using reduction from *clique* problem, an NP-complete problem by Karp [56].

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: *$(r|X_p)$ hub-medianoid* problem is clearly in NP since given the set of leaders and followers hubs for each pair of nodes $i,j \in N$, we can solve the shortest path problem and determine if the flow w_{ij} is captured or not. This process can be done in polynomial time.

Given an instance of *clique* problem, we construct a network $G'=(N',E')$ where $N' = N \cup X_p$, where $X_p = \{x_1,x_2,...,x_p\}$ 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 $i,j \in N$ is set to 1. Clearly $\beta_{ij} = 1$ 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 $f\big(X_p, Y_r(X_p)\big) \geq C(r,2) = (r^2-r)/2$ if and only if there exists an r-clique on G where $C(r,2)$ is 2-combination of a set with cardinality r.

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, $f\big(X_p, Y_r(X_p)\big) \geq (r^2-r)/2$.

On the other hand, suppose Y_r in G' is such that $f\big(X_p, Y_r(X_p)\big) \geq (r^2-r)/2$. 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 spoke link of the follower which contradicts with the assumption.

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

4.3 Computational Study

Performance of **H-MED** is investigated by the computational experiments conducted on two different data sets: CAB and TR. The following table summarizes properties of CAB and TR data sets.

Table 4-2: Summary of properties of CAB and TR data sets

	CAB	**TR**
Proposed by	O'Kelly [31]	Tan and Kara [55]
$\lvert N\rvert$	25	81
$\lvert H\rvert$	25	22
Symmetric flow matrix	Yes	No
Symmetric distance matrix	Yes	Yes

α values are chosen as either 0.6 or 0.8. Also, for TR data set results for $\alpha = 0.9$ are obtained since Tan and Kara propose that this value is obtained from cargo companies of Turkey [55]. 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. The maps showing the spatial locations of nodes and potential hubs of CAB and TR data sets can be found in Appendix 1 and 2, respectively. All instances are solved with CPLEX 12.4.0.0 and a 4 x AMD Opteron Interlagos 16C 6282SE 2.6G 16M 6400MT computer running under Linux operating system.

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 leader's hub set in advance. Therefore, we consider that the leader locates his/her hubs on a set of nodes according to his/her optimal choices of well-studied multi-allocation hub location problems: *uncapacitated multi-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 leader's hub set for this data set. Table 4-3 and 4-4 show the optimal solutions of **UMApHM** and **UMApHC** problems that are used in the computational study.

Table 4-3:Optimal solutions of UMApHM and UMApHC in CAB data set

p	*α*	*UMApHM*	*UMApHC*
2	0.6	12,20	8,21
	0.8	12,20	8,21
3	0.6	4,12,17	8,18,24
	0.8	4,12,17	8,17,24
4	0.6	1,4,12,17	1,12,17,23
	0.8	1,4,12,17	3,6,8,24
5	0.6	4,7,12,14,17	1,18,19,22,23
	0.8	4,7,12,17,24	17,19,22,23,24

Table 4-4 Optimal solutions of UMApHM in TR data set

p	*α*	*UMApHM*
6	0.6	1,6,21,34,35,55
	0.8	1,6,21,34,35,55
	0.9	1,6,21,34,35,55
8	0.6	1,3,6,21,25,34,35,55
	0.8	1,3,6,21,25,34,35,55
	0.9	1,3,6,21,25,34,35,55
10	0.6	1,3,6,16,21,25,34,35,38,55
	0.8	1,3,6,16,21,25,34,35,38,55
	0.9	1,3,6,16,21,25,34,35,38,55
12	0.6	1,6,7,16,21,25,27,34,35,38,42,55
	0.8	1,6,7,16,21,25,27,34,35,38,42,55
	0.9	1,6,7,16,21,25,27,34,35,38,42,55
14	0.6	1,3,6,7,16,21,25,27,34,35,38,42,55,61
	0.8	1,3,6,7,16,21,25,27,34,35,38,42,55,81
	0.9	1,3,6,7,16,21,25,27,34,35,38,42,44,55

The distance matrices of both data sets are symmetric. Therefore, if from node *i* to node *j* 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 (4.4), (4.6) and (4.12)-(4.15) of **H-MED** are imposed for only $i < j$ and the objective (4.11) is replaced with $\sum_{i<j}\sum_m (w_{ij} + w_{ji}) v_{ijm}$ for computational studies.

The following table summarizes all of the totaling up to 139 instances used in the computational study of $(r|X_p)$ *hub-medianoid* problem:

Table 4-5: Summary of the instances used in the computational study

Data Set	CAB	TR
Hub set of the leader	UMApHM & UMApHC	UMApHM
$\boldsymbol{p}$	2,3,4 and 5	6,8,10,12 and 14
$\boldsymbol{r}$	2,3,4 and 5	6,8,10,12 and 14
$\boldsymbol{\alpha}$	0.6 and 0.8	0.6,0.8 and 0.9

Table 4-6 summarizes the CPU time, the market share and hub sets of the follower in the optimal solution of $(r|X_p)$ *hub-medianoid* problem where the follower has already located his/her hubs on the optimum solution of **UMApHM** and **UMApHC** on CAB data set with α = 0.6.

Table 4-6: Results of the ($r|X_p$) hub-medianoid problem on CAB where X_p = UMApHM with α = 0.6

p	Leader's hubs = UMApHM		r = 2	r = 3	r = 4	r = 5
2	{12,20}	CPU	*6.15*	*5.59*	*7.95*	*12.49*
		Share	**65.62%**	**78.25%**	**87.08%**	**92.26%**
		Hubs	{2,6}	{2,6,12}	{2,6,12,19}	{2,5,12,19,20}
3	{4,12,17}	CPU	*11.16*	*9.05*	*14.15*	*10.97*
		Share	**30.49%**	**45.13%**	**53.69%**	**62.02%**
		Hubs	{17,25}	{17,21,25}	{9,17,18,21}	{9,17,18,21,22}
4	{1,4,12,17}	CPU	*23.44*	*17.93*	*20.79*	*24.61*
		Share	**17.91%**	**28.39%**	**37.73%**	**46.18%**
		Hubs	{2,21}	{17,21,25}	{14,17,18,21}	{9,14,17,18,21}
5	{4,7,12,14,17}	CPU	*11.28*	*9.32*	*12.63*	*10.55*
		Share	**18.64%**	**28.14%**	**35.04%**	**42.32%**
		Hubs	{17,25}	{9,17,18}	{9,17,18,21}	{9,17,18,21,22}
p	Leader's hubs = UMApHC		r = 2	r = 3	r = 4	r = 5
2	{8,21}	CPU	*2.86*	*4.32*	*4.74*	*3.69*
		Share	**75.86%**	**85.20%**	**90.98%**	**94.74%**
		Hubs	{5,19}	{4,13,19}	{4,8,12,13}	{4,8,12,13,21}
3	{8,18,24}	CPU	*6.45*	*4.63*	*15.68*	*18.3*
		Share	**51.81%**	**70.25%**	**79.08%**	**85.23%**
		Hubs	{4,17}	{5,17,19}	{5,14,17,19}	{6,14,17,19,21}
4	{1,12,17,23}	CPU	*21.72*	*22.67*	*20.94*	*19.91*
		Share	**36.56%**	**47.39%**	**57.38%**	**66.93%**
		Hubs	{18,20}	{18,20,24}	{13,18,20,24}	{13,18,19,20,24}
5	{1,18,19,22,23}	CPU	*6.75*	*13.39*	*16.08*	*10.96*
		Share	**45.62%**	**57.27%**	**69.34%**	**76.75%**
		Hubs	{4,17}	{1,5,17}	{1,5,12,17}	{6,12,13,17,24}

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 β_{ij}-ε 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 seconds. Appendix 3 summarizes the results of rest of the computational experiment conducted on CAB data set in terms of solution time, follower's optimal hub set and market share. Table 4-7 depicts the percentage of the market that is captured by the follower in the optimal solution of *$(r|X_p)$ hub-medianoid* problem on TR data set.

Table 4-7: Market share captured by the follower in the optimal solution of H-MED for TR data set where hub set of the leader is UMApHM

α	**p\r**	**6**	**8**	**10**	**12**	**14**
0.6	**6**	39.31%	49.19%	56.94%	64.02%	68.91%
	8	28.58%	37.09%	44.37%	51.77%	57.97%
	10	19.91%	27.13%	34.10%	40.48%	45.73%
	12	15.83%	21.79%	27.06%	31.37%	35.48%
	14	13.04%	17.87%	22.25%	26.00%	28.42%
0.8	**6**	37.97%	48.24%	55.70%	61.84%	66.97%
	8	29.37%	37.08%	44.35%	50.71%	56.33%
	10	20.12%	27.03%	33.84%	40.74%	46.84%
	12	16.93%	23.41%	28.62%	32.81%	35.85%
	14	13.02%	18.57%	22.52%	25.20%	27.40%
0.9	**6**	40.86%	49.44%	56.06%	61.54%	66.45%
	8	31.11%	38.69%	44,83%	50.49%	55.77%
	10	20.74%	27.77%	33.86%	39.89%	44.90%
	12	18.45%	24.59%	29.08%	32.98%	36.18%
	14	13.66%	18.81%	22.50%	25.60%	28.18%

The above table also reveals that in case of equal number of hubs, that is $p = r$, the follower captures more than half of the market. The follower should open at least 2 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.

Table 4-8 shows CPU time to solve the **H-MED** on TR data set. As seen in Table 4-8 all instances are solved within 8 minutes even for TR data set. Another observation is that as values of both p and r increase, the amount of time required solving the problem decreases.

Table 4-8: CPU times of H-MED for TR data set where hub set of the leader is UMApHM

α	p\r	6	8	10	12	14
0.6	**6**	467.27	358.34	266.14	80.93	20.77
	8	326.31	286.47	213.4	76.28	21.11
	10	302	190.99	144.75	68.05	18.38
	12	168.81	125.01	61.06	28.81	13.45
	14	141.97	108.2	22.23	10.43	9.76
0.8	**6**	449.76	330.15	158.63	85.34	35.69
	8	393.24	263.35	174.8	84.56	39.49
	10	385.03	222.86	200.72	72.85	26.06
	12	232.28	104.74	107.7	13.03	15.83
	14	109.78	13.44	13.66	34.18	25.14
0.9	**6**	354.78	339.55	213.72	177.01	74.73
	8	182.35	197.03	158.88	129.43	76.34
	10	287.35	121.56	128.64	90.58	32.29
	12	127.1	42.89	36.44	42.05	31.13
	14	66.42	29.76	30.4	31.8	30.64

Hub sets of the follower in the optimum solution of *$(r|X_p)$ hub-medianoid* problem on TR data set are presented in Appendix 4.

As p value gets closer to $|H|$, using the advantageous of being former decision-maker, the leader prevents that the follower can capture at least half of the market even for the case $r > p$. As seen in Table 4-7, for the instances $p \geq 10$, the leader locates his/her hubs on strategic locations and prevents good choices for the follower. Then, for these instances the follower is not able to capture 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, in simple words, the firms have incentive of competing to be the leader. For example, even if the leader choose his/her hubs according to the optimal solution of **UMApHM** for α = 0.6, p = 10 and r = 14 he/she can capture more flow than follower even without having the information about competition. However, this may always not be the case. In CAB instances, p is relatively small than $|H|$ so after the leader makes his/her decision, the follower still has a big action space and being the latter decision-maker is more advantageous if the former one does not have information about the competition.

Chapter 5

(r|p) Hub-centroid Problem

In Chapter 3, we define the *(r|p) hub-centroid* problem as a combinatorial optimization problem from viewpoint of the leader. In this chapter, we provide a bilevel linear model for the problem and provide linearization of this bilevel model. Then, we prove that problem is NP-hard by reduction from *vertex cover* problem. Since bilevel model and its linearization are hard to solve, we propose enumeration-based algorithms for the *(r|p) hub-centroid* problem. Finally, we present numerical studies conducted to observe the performance of linearization of bilevel model as well as proposed algorithms.

5.1 Linearization of (r|p) Hub-centroid Problem

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 (choosing X_p as his/her set of hubs), he/she has the knowledge 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, the leader's problem has a bilevel structure.

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

H_k = 1 if the leader locates a hub on node $k \in H$, and 0 otherwise;

U_{ijk} = 1 if the flow from node $i \in N$ to node $j \in N$ visits hub $k \in H$ as the first hub, and 0 otherwise;

V_{ijm} = 1 if the flow from node $i \in N$ to node $j \in N$ visits hub $m \in H$ as the second hub, and 0 otherwise;

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

a_{ij} = 1 if the flow form node $i \in N$ to $j \in N$ is captured by the follower, and 0 otherwise;

$(h_k, u_{ijk}, v_{ijm}, \gamma_{ij}) = \xi(H_k, U_{ijk}, V_{ijk}, \beta_{ij})$ = the values of decision variables h_k, u_{ijk}, v_{ijm} are provided from the optimal solution of *$(r|X_p)$ hub-medianoid* given X_p. γ_{ij} are the induced values of service levels provided by the follower according to his/her optimal solution. Observe that the capital letter decision variables H_k, U_{ijk}, V_{ijk} of the follower are anaologous ones to their lowercase versions defined in Chapter 4.1.

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

$$\textit{minimize} \quad \sum_i \sum_j a_{ij} w_{ij} \tag{5.1}$$

$$\textit{subject to} \quad \sum_k H_k = p, \tag{5.2}$$

$$\sum_{k} U_{ijk} = 1 \quad \forall\, i \; and \; j \in N, \tag{5.3}$$

$$\sum_{m} V_{ijm} = 1 \quad \forall\; i \; and \; j \; \in N, \tag{5.4}$$

$$U_{ijk} \leq H_k \quad \forall\; i,j \in N \; and \; k \in H, \tag{5.5}$$

$$V_{ijm} \leq H_m \quad \forall\; i,j \in N \; and \; m \in H, \tag{5.6}$$

$$\beta_{ij} \geq \sum_{k} U_{ijk}(c_{ik} + \; \alpha c_{km}) + V_{ijm} c_{mj} \;\; \forall\, i,j \; \in N \; and \; m \; \in H, \tag{5.7}$$

$$\beta_{ij} - \; \gamma_{ij} \leq a_{ij} M \quad \forall\; i,j \; \in N \tag{5.8}$$

$$\left(h_k, u_{ijk}, v_{ijm}, \gamma_{ij}\right) = \xi\left(H_k, U_{ijk}, V_{ijk}, \beta_{ij}\right) \tag{5.9}$$

$$H_k, U_{ijk}, V_{ijm}, \in \; \{0{,}1\} \; and \; \beta_{ij} \geq 0 \; \forall \;\; i,j \; \in N \; and \; k,m \in H. \tag{5.10}$$

The objective (5.1) minimizes the amount of flow captured by the follower which is equivalent to maximizing the amount of flow captured by the leader. Constraint (5.2) ensures the leader locates *p* hubs on the set of available nodes. Constraints (5.3), (5.4), (5.5) and (5.6) 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 (5.7) 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 (5.8) 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 (5.9) guarantees that the follower respond optimally after observing the hub set of the leader. Constraints (5.10) are the domain constraints.

As stated by Bard [9] and Dempe [10] bilevel models are hard to solve even for small number of decision variables. Therefore, 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} = the service level for node pair $i,j \in N$ provided by the follower if he/she choose $S \subseteq H$ 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:

$a_{ij}^{S} = 1$ if the flow form node $i \in N$ to $i \in N$ is captured by the follower when he/she choose $S \subseteq H$ as hub set, and 0 otherwise;

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

minimize $\quad Z$ (5.11)

subject to $\quad Z \geq \sum_{i} \sum_{j} a_{ij}^{S} w_{ij} \quad \forall\, i,j \in N \text{ and } S \subseteq H\ |S| = r\,,$ (5.12)

$\beta_{ij} - \gamma_{ij}^{S} \leq a_{ij}^{S} M \quad \forall\, i,j \in N \text{ and } S \subseteq H\ |S| = r,$ (5.13)

$H_k, U_{ijk}, V_{ijm}, a_{ij}^{S} \in \{0,1\} \text{ and } \beta_{ij} \geq 0$

$\forall\, i,j \in N\ k,m \in H \text{ and } S \subseteq H\ |S| = r.$ (5.14)

(5.2)-(5.7)

Objective function (5.11) and constraints (5.12) together minimize the highest possible captured flow value by the follower in the remaining scenario. Constraints (5.13) correctly calculate whether a flow is captured with a hub set $S \subseteq H$ 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 service level provided by the leader, the RHS of the constraint must be positive and $a_{ij}^S = 1$. Otherwise, the constraint becomes redundant. Constraints (5.14) are domain constraints.

The mixed integer problem **H-CEN** has $3n^2m + 2n^2 + 2n^2C(m,r) + 1$ constraints and $2n^2m + n^2C(m,r) + n^2 + m + 1$ variables of which $2n^2m + n^2C(m,r) + m$ are binary where $|N| = n$, $|H| = m$ and $C(m,r)$ is r-combination of the set H.

5.2 Problem Complexity

We prove that the problem of finding a *(r|p) hub-centroid* is NP-hard by using reduction from *vertex cover problem*, an NP-complete problem by Karp [56]. However, decision version of the *(r|p) hub-centroid* does not belong to complexity class NP.

Decision Version of Vertex Cover Problem: Given an undirected graph $G=(N,E)$ and an integer p, determine if G has a vertex cover C, 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 a *(r|p) hub-centroid* is NP-hard even if $\alpha = 1$.

Proof: *(r|p) hub-centroid* problem is not in NP since given the set of leader's hub set we need to solve *$(r|X_p)$ hub-medianoid* problem to observe the amount of flow captured by the follower. Since *$(r|X_p)$ hub-medianoid* cannot be solved in polynomial time, we can conclude that decision version of *(r|p) hub-centroid* problem does not belong to complexity class NP.

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, w_{ij}, 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 $f\left(X_p, Y_r^*(X_p)\right) = 0$ 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 for each 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, $f\left(X_p, Y_r^*(X_p)\right) = 0$.

On the other hand, suppose X_p in G' is such that $f\left(X_p, Y_r^*(X_p)\right) = 0$. 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, $f\left(X_p, Y_r^*(X_p)\right) \geq 1$ 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-hard. □

5.3 Computational Performance of H-CEN

We used CAB data and computer set presented in Chapter 4.3 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 2 which yield $O(n^4)$ variables and constraints. Table 5-1 summarizes the results of the computational study for these instances within a time limit of 7200 second (= 2 hours). First column of the table (n) indicates the number of nodes in the instance, the second row (Follower's capture (%)) is the percentage of total flow that the follower captures in the optimal solution of the *(r|p) hub-centroid* problem, the third row (Solution Time (sec)) is the required CPU time of the optimal solution if it found within the time limit, finally the fourth row (Gap %) shows the optimality gap if the optimal solution is not found within 2 hours.

Table 5-1: Summary of numerical experiments for H-CEN model

n	Follower's capture (%)	Solution Time (sec)	Gap %	*n*	Follower's capture	Solution Time (sec)	Gap %
5	**41.39%**	*1.04*	--	15	**43.75%**	--	**55.92%**
6	**40.16%**	*3.83*	--	16	**43.15%**	--	**76.74%**
7	**40.59%**	*13.30*	--	17	**58.49%**	--	**83.53%**
8	**36.36%**	*18.34*	--	18	**61.16%**	--	**86.10%**
9	**34.31%**	*109.31*	--	19	**100.00%**	--	**91.41%**
10	**39.72%**	*475.02*	--	20	**100.00%**	--	**92.33%**
11	**41.03%**	*325.55*	--	21	**58.18%**	--	**88.02%**
12	**40.55%**	--	**4.50%**	22	**98.36%**	--	**92.19%**
13	**39.55%**	--	**20.62%**	23	**57.65%**	--	**87.82%**
14	**46.18%**	--	**17.16%**	24	**100.00%**	--	**93.02%**
				25	**100.00%**	--	**93.33%**

The conducted computational study revealed that the **H-CEN** model can only be solvable within 2 hours for $n \leq 11$. Moreover, for values $n \geq 15$, the optimality gap is greater than 50%. Therefore, for even very small instances, exact solution of **H-CEN** model 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 get optimal solutions of *(r|p) hub-centroid* problem for the 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:

---

algorithm *complete enumeration*

begin

initialize all X_p in $\mathscr{P}_p(H)$ as unmarked

total_flow := $\sum_{i,j \in N} w_{ij}$;

leader_objective := 0,

leader_hubset := {};

while $\mathscr{P}_p(H)$ contains an unmarked element X_p

    initialize all Y_r in $\mathscr{P}_r(H)$ as unmarked

    current_leader_objective := 0;

    follower_objective := 0,

    follower_hubset :={};

    while $\mathscr{P}_r(H)$ contains an unmarked element Y_r

        current_follower_objective := 0;

        for each $i \in N$ to $j \in N$ **do**

            $\beta_{ij} = \min_{k,m \in X_p}\{c_{ik} + \alpha c_{km} + c_{mj}\}$

            $\gamma_{ij} = \min_{k,m \in Y_r}\{c_{ik} + \alpha c_{km} + c_{mj}\}$

            if $\gamma_{ij} < \beta_{ij}$ **then**

                current_follower_objective

                := current_follower_objective + w_{ij};

        if current_follower_objective > follower_objective **then**

            follower_objective := current_follower_objective;

            follower_hubset := Y_r;

            current_leader_objective

            := total_flow - follower_objective ;

```
            mark Y_r;
        if current_leader_objective > leader_objective then
            leader_objective := current_leader_objective;
            leader_hubset := X_p;
        mark X_p;
end;
```

The flow chart of the *complete enumeration* algorithm is presented in Appendix 5.

The above *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, we can say that the running time of the algorithm is proportional to $n^2|\mathscr{P}_p(H)||\mathscr{P}_r(H)|$.

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 exists X_p' and Y_r' with $f\left(X_p, Y_r^*\left(X_p\right)\right) < f\left(X_p', Y_r'\right)$ then X_p' cannot be an optimal solution to *(r|p) hub-centroid* problem.

Proof: $f\left(X_p', Y_r'\right) \leq f\left(X_p', Y_r^*(X_p')\right)$ 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 $f\left(X_p, Y_r^*\left(X_p\right)\right) < f\left(X_p', Y_r'\right)$ and $f\left(X_p', Y_r'\right) \leq f\left(X_p', Y_r^*(X_p')\right)$ together imply that $f\left(X_p, Y_r^*\left(X_p\right)\right) < f\left(X_p', Y_r^*(X_p')\right)$. 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 optimum solution to *(r|p) hub-centroid* problem. Then, we propose *smart enumeration* algorithm by inserting following lines between lines 22 and 23 of *complete enumeration* algorithm.

---

if current_follower_objective > total_flow - leader_objective **then**

    mark all Y_r in $\mathscr{P}_r(H)$ and continue with another X_p

---

The flow chart of the *smart enumeration* algorithm is presented in Appendix 6.

Even if the worst case running time of the *smart enumeration* algorithm is equal to the *smart enumeration*, in practice solution times are reasonably short when compared to its worst case performance as will be seen in Section 5.4.

We can still decrease the 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 $f\left({X_p}^*, {Y_r}^*({X_p}^*)\right) < \frac{W}{2}$ 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 $f\left({X_p}^*, {Y_r}^*({X_p}^*)\right) \geq \frac{W}{2}$. Then, at least 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 half of the total flow by setting his/her hub set ${X_p}' = {Y_r}^*({X_p}^*)$. Then, $f\left({X_p}', {Y_r}^*({X_p}^*)\right) = 0$ 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 and $j \in H \subseteq N$ 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, $f\left(X_p{}', Y_r{}'\right) > 0$. 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 $f\left(X_p{}^*, Y_r{}^*(X_p{}^*)\right) \geq f\left(X_p{}^*, Y_r{}'\right)$.

Hence, under the conditions $p \geq r$, $p < |N|-2$, $r \geq 2$, 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 $f\left(X_p{}^*, Y_r{}^*(X_p{}^*)\right) < \frac{W}{2}$ □

Utilizing Theorem 4, we can further improve the 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 exists X_p and Y_r with $f\left(X_p, Y_r\right) > \frac{W}{2}$ 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.

Then, we propose *smart enumeration with 50%-bound* for the instances with $p \geq r$ algorithm by inserting following lines between lines 22 and 23 of *complete enumeration* algorithm:

---

if current_follower_objective > total_flow - leader_objective **then**

  mark all Y_r in $\mathscr{P}_r(H)$ and continue with another X_p

if current_follower_objective > total_flow /2 **then**

  mark all Y_r in $\mathscr{P}_r(H)$ and continue with another X_p

---

The flow chart of the *smart enumeration with 50%-bound* algorithm is presented in Appendix 7.

5.5 Computational Study

The following example of first 5 nodes of CAB data set with p = r = 2 and α = 0.6 depicts how *complete enumeration*, *smart enumeration* and *smart enumeration with 50%-bound* work. The rows and columns of Table 5-2 indicate feasible hub set choices of the leader and follower, respectively. The entity in each shell shows the percentage of total flow captured by the follower. For example, if $X_p = \{1,2\}$ and $Y_r = \{1,3\}$, then the market share of the follower is 37.76% as indicated in the intersection of first row and second column in Table 5-2. The bolded entities in the table are the maximizers of that row.

Table 5-2: The percentage of captured flow by the follower for first 5 nodes of CAB data set with p = r = 2 and α = 0.6 calculated by complete enumeration algorithm

$X_p \backslash Y_r$	*{1,2}*	*{1,3}*	*{1,4}*	*{1,5}*	*{2,3}*	*{2,4}*	*{2,5}*	*{3,4}*	*{3,5}*	*{4,5}*
{1,2}	0.00	37.76	57.56	38.58	41.92	52.64	59.83	46.65	46.65	**72.79**
{1,3}	18.20	0.00	43.52	47.15	55.11	55.22	32.60	55.22	**59.83**	43.52
{1,4}	28.19	20.60	0.00	20.90	50.41	**55.11**	27.88	41.92	50.41	42.68
{1,5}	**58.85**	43.22	53.37	0.00	58.45	47.92	55.11	47.92	47.84	52.64
{2,3}	44.89	39.88	33.96	33.96	0.00	**71.81**	51.77	61.19	65.81	44.58
{2,4}	14.56	14.16	25.08	29.69	20.60	0.00	10.83	37.76	**41.39**	38.58
{2,5}	26.46	48.67	48.67	25.08	43.22	53.37	0.00	48.67	37.76	**57.56**
{3,4}	17.14	14.56	24.18	26.75	28.19	18.20	26.45	0.00	21.75	**52.16**
{3,5}	37.07	26.46	31.16	24.18	31.61	31.16	18.20	**53.37**	0.00	40.94
{4,5}	21.00	21.00	26.46	30.09	52.85	**58.85**	38.81	43.22	47.84	0.00

If X_p is given, then the follower choices the hub set which maximizes his/her market share since we assume that the follower acts rationally. For example, if $X_p = \{1,2\}$ is given, then the follower responds with $Y_r^*(X_p) = \{4,5\}$ which is the maximum value of the first row. Therefore, we can argue that an optimal solution of *(r|p) hub-centroid* problem for this instance needs to be searched among row maximums. Since we also assume that the leader acts rationally, he/she aims to minimize the market share of the follower. Therefore, the leader choices the hub set so that the row maximum gets the smallest value among all other row maximums. So, $X_p^* = \{2,4\}$ is the optimal solution where the follower responds with $Y_r^*(X_p^*) = \{3,5\}$ and follower's capture is 41.39%. The *complete enumeration* algorithm calculates all the entities of the table and finds the minimum of the maximizers of each row as the optimal solution.

As mentioned before, *smart enumeration* algorithm is an improved version of *complete enumeration* algorithm using the results of Theorem 3. Here, we give an example of how the improvement is obtained. Assume that we calculated all values in the first three rows of Table 5-2. Then, the current objective values is 55.11% with $X_p = \{1,4\}$ and $Y_r^*(X_p) = \{2,4\}$. Again assume that we continued the search and find the first entity of the fourth row whose value is 58.85%. This value is already greater that the current objective, 55.11%. Hence, maximum value of the fourth row is greater than the current objective. We can conclude that the optimal solution of *(r|p) hub-centroid* problem cannot be found on the fourth row and calculating other values in the fourth row is redundant. In Table 5-3 indicated entities calculated by *smart enumeration* algorithm where redundant entities are removed.

Table 5-3: The percentage of captured flow by the follower for first 5 nodes of CAB data set with p = r = 2 and α = 0.6 calculated by smart enumeration algorithm

$X_p \backslash Y_r$	*{1,2}*	*{1,3}*	*{1,4}*	*{1,5}*	*{2,3}*	*{2,4}*	*{2,5}*	*{3,4}*	*{3,5}*	*{4,5}*
{1,2}	0.00	37.76	57.56	38.58	41.92	52.64	59.83	46.65	46.65	**72.79**
{1,3}	18.20	0.00	43.52	47.15	55.11	55.22	32.60	55.22	**59.83**	43.52
{1,4}	28.19	20.60	0.00	20.90	50.41	**55.11**	27.88	41.92	50.41	42.68
{1,5}	**58.85**									
{2,3}	44.89	39.88	33.96	33.96	0.00	**71.81**				
{2,4}	14.56	14.16	25.08	29.69	20.60	0.00	10.83	37.76	**41.39**	38.58
{2,5}	26.46	48.67								
{3,4}	17.14	14.56	24.18	26.75	28.19	18.20	26.45	0.00	21.75	**52.16**
{3,5}	37.07	26.46	31.16	24.18	31.61	31.16	18.20	**53.37**		
{4,5}	21.00	21.00	26.46	30.09	52.85					

As seen in Table 5-3, only 72 entities are calculated by *smart enumeration* algorithm instead of 100.

Smart enumeration with 50%-bound algorithm uses the results of Theorem 4 in a similar manner. For an instance of the problem with $p \geq r$, Theorem 4 implies that total captured flow by the follower in optimal solution of the *(r|p) hub-centroid* problem is less than 50%. Therefore, if we found an entity with a value of greater than or equal to 50%, searching optimal value over the remainder of the values in this row is redundant. Table 5-4 shows the entities calculated by *smart enumeration with 50%-bound* algorithm. For this example instance only 55 cells need to be calculated.

Table 5-4: The percentage of captured flow by the follower for first 5 nodes of CAB data set with p = r = 2 and α = 0.6 calculated by smart enumeration with 50%-bound algorithm

$X_p \backslash Y_r$	*{1,2}*	*{1,3}*	*{1,4}*	*{1,5}*	*{2,3}*	*{2,4}*	*{2,5}*	*{3,4}*	*{3,5}*	*{4,5}*
{1,2}	0.00	37.76	57.56							
{1,3}	18.20	0.00	43.52	47.15	55.11					
{1,4}	28.19	20.60	0.00	20.90	50.41					
{1,5}	**58.85**									
{2,3}	44.89	39.88	33.96	33.96	0.00	**71.81**				
{2,4}	14.56	14.16	25.08	29.69	20.60	0.00	10.83	37.76	**41.39**	38.58
{2,5}	26.46	48.67								
{3,4}	17.14	14.56	24.18	26.75	28.19	18.20	26.45	0.00	21.75	**52.16**
{3,5}	37.07	26.46	31.16	24.18	31.61	31.16	18.20	**53.37**		
{4,5}	21.00	21.00	26.46	30.09	52.85					

All algorithms are coded in Java 1.6.0_23 on the computer presented in Section 4.3. The following table summarizes all of the totaling up to 80 instances used in the computational study of smart *enumeration* and *smart enumeration with 50%-bound* algorithms:

Table 5-5: Summary of the instances used in the computational study

Data Set	CAB	TR
p	2,3,4 and 5	2,3,4 and 5
r	2,3,4 and 5	2,3,4 and 5
α	0.6 and 0.8	0.6,0.8 and 0.9

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 $n^2|\mathscr{P}_p(H)||\mathscr{P}_r(H)|$, in practice *smart enumeration* and *smart enumeration with 50%-bound* algorithms outperforms *complete enumeration* dramatically especially for large instances. Table 5-6 illustrates the running times of the algorithms for CAB data set for relatively small values of p and r.

Table 5-6: Running times of three algorithms for CAB data with p,r $\in$ {2,3} (in CPU seconds)

	complete enumeration		*smart enumeration*		*smart enumeration with 50%-bound*	
p\r	*2*	*3*	*2*	*3*	*2*	*3*
2	19.82	325.48	1.52	12.71	0.93	--
3	191.96	2513.34	5.81	19.94	5.61	11.46

The above example reveals that the *complete enumeration* algorithm takes much longer times compared to other algorithms even for $p,r \in \{2,3\}$. Therefore, computational analysis is conducted for *smart enumeration* and *smart enumeration with 50%-bound* algorithms only.

Tables 5-7 and 5-8 depict the solution times of *smart enumeration* and *smart enumeration with 50%-bound* algorithms for CAB data set with p and r values range from 2 to 5 with α = 0.6 and 0.8, respectively. Since *smart enumeration with 50%-bound* algorithm cannot be applied for $p < r$ corresponding cells are marked with "--". The optimal locations of hubs and objective value in the optimal solution are presented in Appendix 8 for these instances.

Table 5-7: Solution times of smart enumeration and smart enumeration with 50%-bound algorithms for CAB data with α = 0.6 (in seconds)

	smart enumeration					*smart enumeration with 50%-bound*			
p\r	**2**	**3**	**4**	**5**		**2**	**3**	**4**	**5**
2	*1.52*	*12.71*	*70.32*	*320.78*		*0.93*	--	--	--
3	*5.81*	*19.94*	*88.02*	*557.16*		*5.61*	*11.46*	--	--
4	*19.27*	*36.62*	*141.6*	*631.1*		*17.24*	*33.27*	*77.38*	--
5	*70.35*	*117.4*	*371.14*	*1498.94*		*70.09*	*117.15*	*341.28*	*1272.24*

Table 5-8: Solution times of smart enumeration and smart enumeration with 50%-bound algorithms for CAB data with α = 0.8 (in seconds)

	smart enumeration					*smart enumeration with 50%-bound*			
p\r	**2**	**3**	**4**	**5**		**2**	**3**	**4**	**5**
2	*1.35*	*11.78*	*100.37*	*535.15*		*0.72*	--	--	--
3	*4.31*	*23.13*	*142.68*	*791.55*		*4.13*	*14.65*	--	--
4	*17.96*	*30.7*	*212.61*	*1015.49*		*18.56*	*25.58*	*155.75*	--
5	*74.25*	*139.05*	*382.09*	*1335.03*		*72.77*	*135.53*	*360.54*	*1043.81*

As can be inferred in Tables 5-7 and 5-8, the solution times of these algorithms seem satisfactory and (*r*|*p*) *hub-centroid* problem on a 25 node can be solved in reasonable CPU times. All instances of CAB data set are solved within half an hour. Therefore, it can be concluded that *smart enumeration* and *smart enumeration with 50%-bound* algorithms are efficient tools for the solution of (*r*|*p*) *hub-centroid* problem for moderate size networks.

Although *smart enumeration* and *smart enumeration with 50%-bound* algorithms can solve the *(r|p) hub-centroid* problem on CAB within reasonable times, TR data set require more CPU time. Table 5-9 depicts the running time of *smart enumeration with 50%-bound* algorithm for the instances of TR data set.

Table 5-9: Solution times of smart enumeration with 50%-bound algorithms for TR data set (in CPU seconds)

α	**p\r**	**2**	**3**	**4**	**5**
0.6	**2**	7.20	45.69	257.80	1409.52
	3	24.60	77.19	400.15	1630.12
	4	75.35	161.12	724.81	2166.95
	5	415.39	551.04	1098.69	4911.17
0.8	**2**	8.81	39.23	280.44	1326.98
	3	21.65	73.32	351.76	1354.41
	4	80.30	154.12	549.42	2087.01
	5	440.62	534.12	997.63	4450.71
0.9	**2**	8.39	38.84	265.09	1775.52
	3	26.90	68.93	365.13	2002.92
	4	77.32	176.41	901.68	3022.19
	5	455.12	583.39	1706.91	6634.97

Even for $p,r \in \{2,3,4,5\}$ running time of the algorithm worsens dramatically since TR has 81 nodes and 81x81 = 6561 flows in comparison to CAB data set where only 25x25 = 625 flow attribute exist. Optimal hub sets of the leader and percentage of total flow captured by the follower in the optimal solution of *(r|p) hub-centroid* problem are presented in Appendix 9 and 10, 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 without the knowledge of that another firm will enter the same market, 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 enter the market and captures some of customers that belong to the leader previously. In Table 5-9, we compare the 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.

Table 5-10: Market share of the follower for CAB data set with α = 0.6 in the optimal solutions of (r|p) hub-centroid, p-hub median and p-hub center

p	**r**	*(r|p)* hub-centroid	*$(r\|X_p)$* hub-medianoid X_p = p-hub median	**Difference with optimal**	*$(r\|X_p)$* hub-medianoid X_p = p-hub center	**Difference with optimal**
2	**2**	46.14%	65.62%	**19.48%**	75.86%	**29.72%**
	3	64.37%	78.25%	**13.88%**	85.20%	**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%	**0.10%**	51.81%	**21.42%**
	3	45.13%	45.13%	**0.00%**	70.25%	**25.12%**
	4	53.69%	53.69%	**0.00%**	79.08%	**25.39%**
	5	62.02%	62.02%	**0.00%**	85.23%	**23.21%**
4	**2**	17.91%	17.91%	**0.00%**	36.56%	**18.65%**
	3	28.39%	28.39%	**0.00%**	47.39%	**19.00%**
	4	37.73%	37.73%	**0.00%**	57.38%	**19.65%**
	5	46.18%	46.18%	**0.00%**	66.93%	**20.75%**
5	**2**	14.30%	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%**

For example, if $p = r = 2$ 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 lost 29.72%.

As seen in Table 5-9, 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 the difference in definition of the problems. While *p-hub median* problem minimizes the weighted sum of the service levels of each node pair where the weights are flow between node pairs, *p-hub center* problem aims to minimize the service level of most disadvantageous node pair. *p-hub center* problem ignores the flows between node pairs and focuses only the distance between them. On the other hand, *p-hub median* problem locates the 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* is reasonably small and for 7 of the 16 instances the optimum 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 provides genuine bound is already obtained, the running time of algorithm can be improved. For the instances reported in Table 5-10, 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* ∈ *{2,3,4,5}* and α = 0.6 by including the bound obtained by the optimal solution of *p-hub median* problem. Table 5-11 depicts CPU time of this experiment

Table 5-11: CPU time of smart enumeration algorithm with p-hub median bound (in seconds)

p\r	**2**	**3**	**4**	**5**
2	1.52	10.94	52.84	269.13
3	3.21	8.90	20.76	105.55
4	14.64	20.04	45.73	145.72
5	68.51	105.97	300.05	1078.97

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, the higher improvement is obtained.

Chapter 6

Conclusion & Research Extensions

In this thesis, we propose a duopoly model where two competitors sequentially choose hub locations and aim to maximize their own market share under Stackelberg competition rules. Although competitive location has attracted the attention of economists and OR practitioners, hub location considering competition studies are rare. Therefore, some 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. Perfect information means each player can observe the system correctly and each player knows that his/her competitor can observe the environment correctly. It is also assumed that both players are rational which means that they aim to maximize their market share. The market share of the firms is determined by the flows (or customers). Although choice of the customers depends on many attributes, we assume that the customers prefer the firm which offers a better service level.

First, from the view of the follower *$(r|X_p)$ hub-medianoid* problem is defined. It is also proved that finding *$(r|X_p)$ hub-medianoid* is NP-hard. At the time the follower makes his/her decisions, the hub set of the leader has already chosen. Therefore, for the follower the problem is a "maximum capture" or a "maximum cover" problem rather that a competition. Hence, both formulation and solution of *$(r|X_p)$ hub-medianoid* problem require less effort.

On the other hand, the competition issue becomes important from the viewpoint of the leader. After the leader makes his/her decisions, the follower takes action and then the markets shares are determined. Therefore, the *$(r|p)$ hub-centroid* problem has a bilevel nature. We then propose a bilevel model and its linearization as well as the computational complexity of the problem. The linearization of the bilevel model can only be solvable for very small instances where solving the bilevel mode is even harder. However, proposed enumeration-based algorithms can solve the problem for relatively bigger instances even though the worst-case complexity tends to complete enumeration.

Conducted computational analysis revealed that the leader can increase his/her market share by choosing hub set based on *$(r|p)$ hub-centroid* problem rather than optimal solutions of classical hub location problems such as *p-hub median* and *p-hub center* problems. The leader can increase his/market share by being aware of competition up to 37.43% as seen in the computational analysis conducted on CAB data set.

We hope that this thesis will motivate researchers to study various possible extensions of hub location problem under competition. Some possible extensions could be following:

Although we assume that both firms use same parameter values, allowing firm-specific parameters will increase the applicability of the results. Since different firms use different technology, vehicles and operational strategies, the service levels which depend on cost and interhub discount values may be specific to the firms.

In duopoly, the number of competing firms is assumed to be two. However, one may extend the problem with three, four and even more firms. As the number of firm increases, the complexity of the models and solution techniques will increase. Even if this extension may be challenging, models and solution algorithms can be adopted easily. The bilevel model can be extended as a multi-level programming model and enumeration-based algorithms can be expanded for three or more firms.

Although objectives of the firms are 100% conflicting, that is one aims to maximize the amount of captured flow by the follower while the other tries to minimize the same objective, in reality cooperation may be profitable for both parties. For example, a passenger may prefer to fly with firm A from his/her origin to origin's hub and firm B from destination's hub to his/her destination with firm B where interhub transportation is operated in the coordination of both firms. Such coordination will bring benefit to both parties by providing better service levels to the customers. Especially in air transportation, customers seek for the cheapest way to fly their destinations which means that they are willing to buy ticket from different firms for different legs of their trips.

Other possible extensions may be competitive hub location problem with *elastic demand* and/or *partial capture*. Despite of the fact that the flow between node pairs are known *a priori*, the demand may be related to the service levels and some demand may be lost. Also, some fraction of the flow between a pair of nodes may be captured by one firm instead of *all-or-nothing* type capture. This partial capture function may even be a distribution function that represents the stochastic nature of customer decisions.

Finding an exact solution to *(r|p) hub-centroid* problem requires big amounts of memory requirement and CPU time. Therefore, some heuristic approaches may be developed to find a near optimal solution to this problem in reasonable CPU time.

BIBLIOGRAPHY

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

[2] Bertrand, J. (1883). Théorie des richesses: revue de ‘Théorie mathématique de la richesse sociale’par Léon Walras et ‘Recherches sur les principes mathématiques de la richesse’par Augustin Cournot. *J. des Savants*.

[3] Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39, 41-57

[4] Lerner, A. P., & Singer, H. W. (1937). Some notes on duopoly and spatial competition. *The Journal of Political Economy*, *45*(2), 145-186.

[5] Smithies, A. (1941). Optimum location in spatial competition. *The Journal of Political Economy*, *49*(3), 423-439.

[6] Eaton, B. C., & Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies*, *42*(1), 27-49.

[7] Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, *36*(1), 48-49.

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

[9] Bard, J. (1999). *Practicle bilevel optimization*. Kluwer Academicians, Dordrecht.

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

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

[12] Wendell, R. E., & Thorson, S. J. (1974). Some generalizations of social decisions under majority rule. *Econometrica: Journal of the Econometric Society*, 893-912.

[13] Slater, P. J. (1975). Maximin facility location. *Journal of National Bureau of Standards B*, *79*, 107-115.

[14] Wendell, R. E., & McKelvey, R. D. (1981). New perspectives in competitive location theory. *European Journal of Operational Research*, *6*(2), 174-182.

[15] Hansen, P., & Thisse, J. F. (1981). Outcomes of voting and planning: Condorcet, Weber and Rawls locations. *Journal of Public Economics*, *16*(1), 1-15.

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

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

[18] Hakimi, S. L. (1986). *p*-Median theorems for competitive locations. *Annals of operations research*, *6(4)*, 75-98.

[19] ReVelle, C. (1986). The Maximum Capture or “Sphere of Influence” Location Problem: Hotelling Revisited on a Network. *Journal of Regional Science*, *26*(2), 343-358.

[20] Eiselt, H. A., Laporte, G., & Thisse, J. F. (1993). Competitive location models: A framework and bibliography. *Transportation Science*, *27*(1), 44-54.

[21] Serra, D., & ReVelle, C. (1999). Competitive location and pricing on networks. *Geographical analysis*, *31*(2), 109-129.

[22] Serra, D., & Colomé, R. (2001). Consumer choice and optimal locations models: formulations and heuristics. *Papers in Regional Science*, *80*(4), 439-464.

[23] Benati, S., & Laporte, G. (1994). Tabu Search Algorithms for the (r/Xp)-medianoid and (r/p)-centroid Problems. *Centre de recherche sur les transports publicatıon*, (941).

[24] Benati, S., & Hansen, P. (2002). The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, *143*(3), 518-530.

[25] Hansen, P., & Labbé, M. (1988). Algorithms for voting and competitive location on a network. *Transportation Science*, *22*(4), 278-288.

[26] Serra, D., & ReVelle, C. (1994). Market capture by two competitors: the preemptive location problem. *Journal of regional Science*, 34(4), 549-561.

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

[28] Daşçı, A. (2011). Conditional Location Problems on Networks and in the Plane. In *Foundations of Location Analysis* (pp. 179-206). Springer US.

[29] O'Kelly, M. E. (1986). The location of interacting hub facilities. *Transportation Science*, *20*(2), 92-106.

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

[31] 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.

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

[33] 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.

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

[35] Kara, B.Y. (1999). Modelling and analysis of issues in hub location problem. Ph.D. Thesis, Bilkent University Industrial Engineering Department, 06800 Bilkent, Ankara, Turkey.

[36] Campbell, J. F. (1992). Location and allocation for distribution systems with transsiphments and transportion economies of scale. *Annals of Operations Research*, 40(1), 77-99.

[37] 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.

[38] Campbell, J. F. (1996). Hub location and the p-hub median problem. *Operations Research*, *44*(6), 923-935.

[39] Boland, N., Krishnamoorthy, M., Ernst, A. T., & Ebery, J. (2004). Preprocessing and cutting for multiple allocation hub location problems. European Journal of Operational Research, 155(3), 638-653.

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

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

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

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

[44] 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.

[45] 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*.

[46] Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2002). Hub location problems. *Facility location: applications and theory*, 1, 373-407.

[47] Alumur, S., & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research*, *190*(1), 1-21.

[48] Kara, B. Y., & Taner, M. R. (2011). Hub location problems: The location of interacting facilities. In *Foundations of Location Analysis* (pp. 273-288). Springer US.

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

[50] Wagner, B. (2008). A note on “location of hubs in a competitive environment”. *European Journal of Operational Research*, *184*(1), 57-62.

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

[52] 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.

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

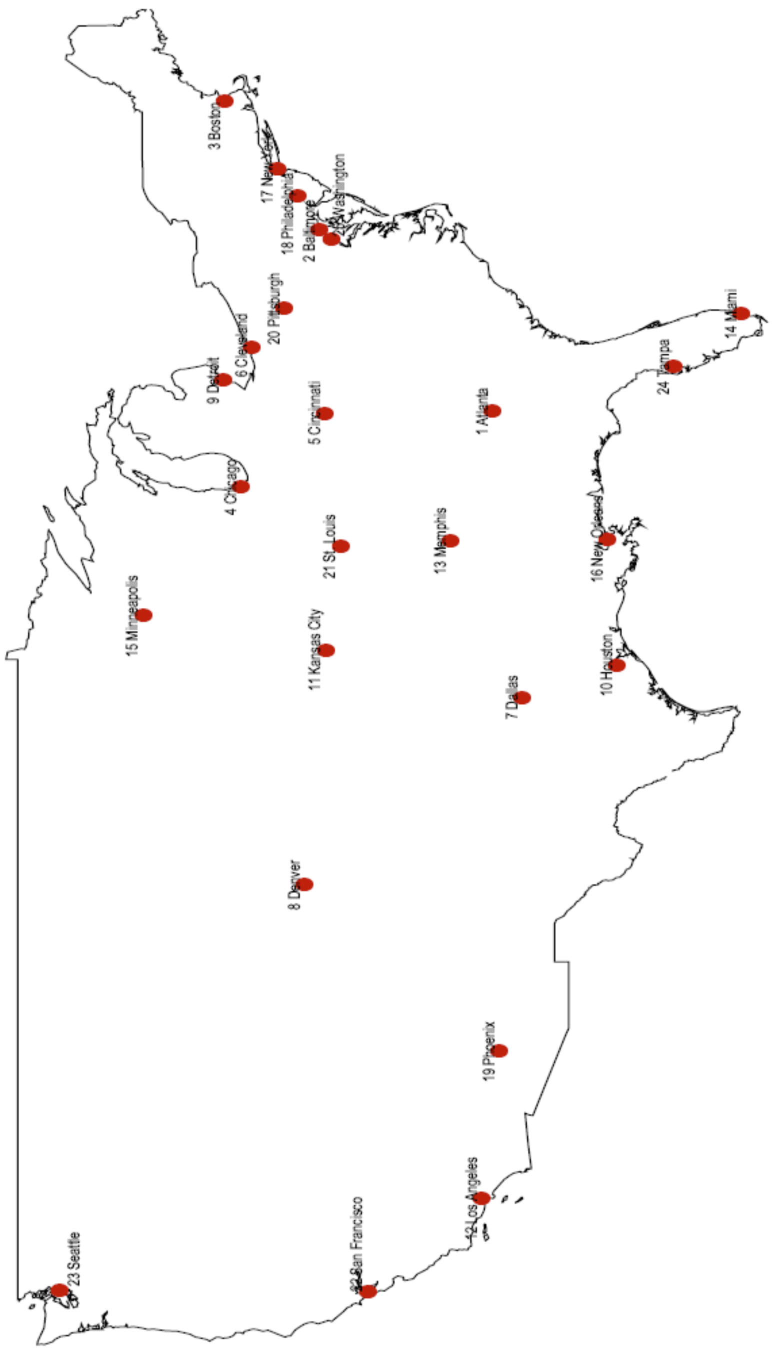
[54] Sasaki, M., Campbell, J. F., Krishnamoorthy, M., & Ernst, A. T. (2009, February). Designing Hub and Spoke Transportation Systems for a Competitive Environment. In *International Conference on Industrial Engineering and Systems Management* (pp. 13-15).

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

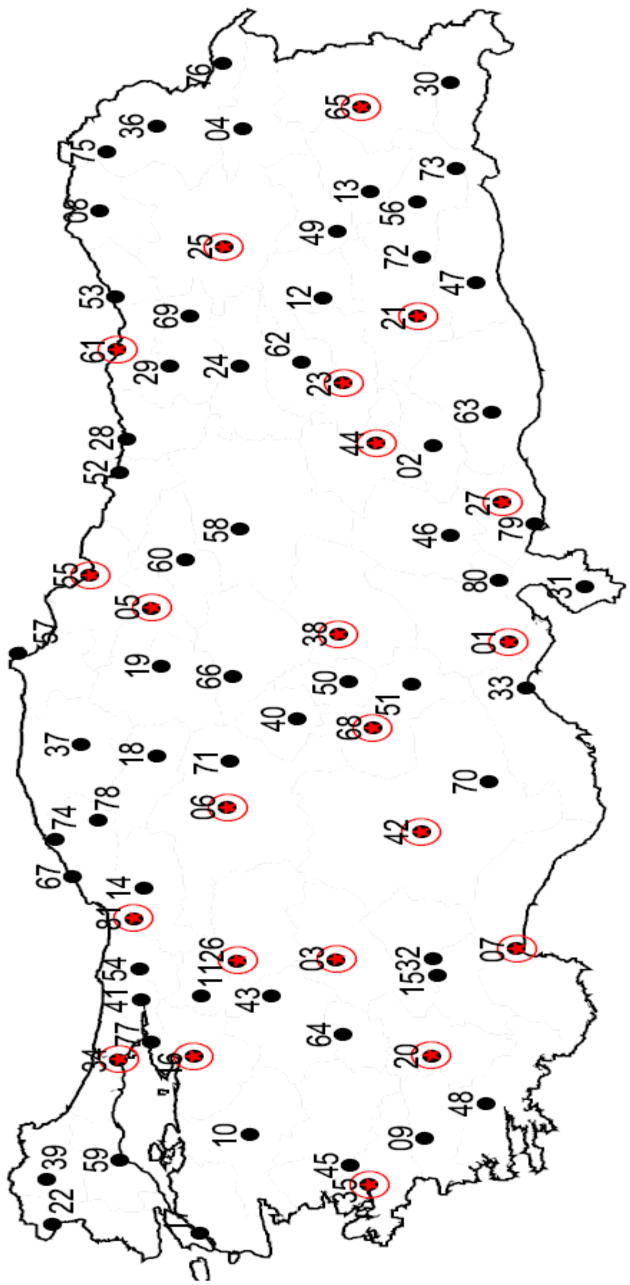
[56] Karp, R. M. (1972). *Reducibility among combinatorial problems* (pp. 85-103). Springer US.

APPENDIX

Appendix 1: Map showing the nodes of CAB data set



Appendix 2: Map showing the nodes of TR data set where nodes with circles are possible hub locations



Appendix 3: Summary of computational experiment conducted on CAB data set in terms of solution time, follower's optimal hub set and market share

$\alpha = 0.8$	UMApHC		r = 2	r = 3	r = 4	r = 5
		CPU time	*2.75*	*4.65*	*9.47*	*11.31*
p = 2	{8,21}	Follower's share	**73.04%**	**82.43%**	**89.68%**	**92.32%**
		Follower's hubs	{5,12}	{4,12,13}	{4,8,13,19}	{4,8,13,19,21}
		CPU time	*11.6*	*12.3*	*13.01*	*7.32*
p = 3	{8,17,24}	Follower's share	**42.37%**	**55.89%**	**65.90%**	**75.00%**
		Follower's hubs	{12,20}	{12,17,20}	{12,17,18,20}	{12,14,17,18,20}
		CPU time	*4.17*	*3.76*	*10.52*	*12.55*
p = 4	{3,6,8,24}	Follower's share	**44.19%**	**58.80%**	**65.64%**	**73.18%**
		Follower's hubs	{2,4}	{4,12,25}	{3,4,12,17}	{3,4,12,17,25}
		CPU time	*9.29*	*10*	*7.92*	*9.67*
p = 5	{17,19,22,23,24}	Follower's share	**42.19%**	**52.65%**	**62.66%**	**71.62%**
		Follower's hubs	{17,20}	{12,17,20}	{12,17,18,20}	{12,14,17,18,20}

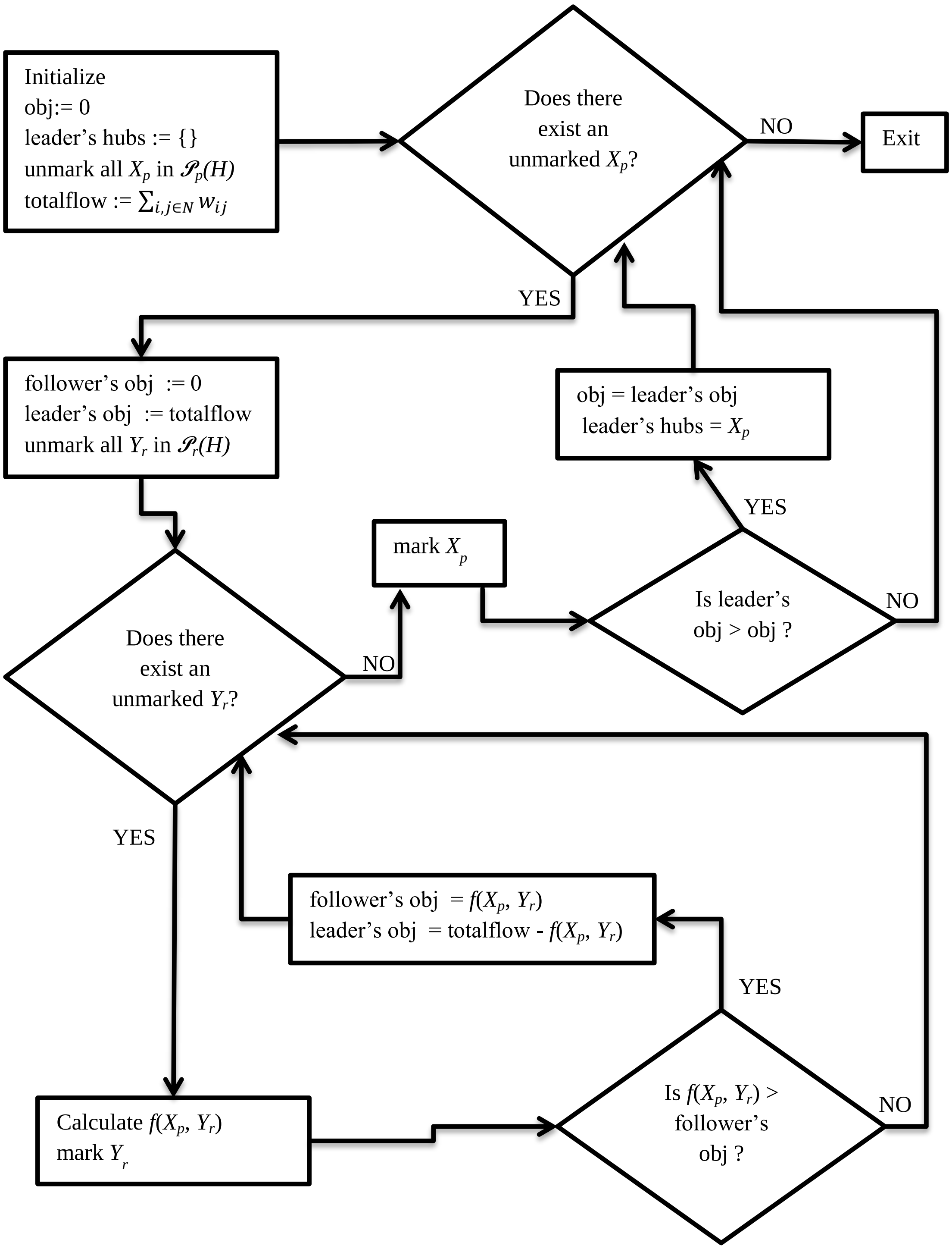
$\alpha = 0.8$	UMApHM		r = 2	r = 3	r = 4	r = 5
		CPU time	*2.84*	*11.89*	*15.82*	*13.63*
p = 2	{12,20}	Follower's share	**65.84%**	**74.19%**	**80.69%**	**87.14%**
		Follower's hubs	{6,25}	{6,8,25}	{6,8,12,25}	{6,12,21,22,25}
		CPU time	*11.5*	*6.7*	*10.3*	*10.39*
p = 3	{4,12,17}	Follower's share	**29.04%**	**42.92%**	**52.83%**	**60.14%**
		Follower's hubs	{17,25}	{17,21,25}	{11,17,18,20}	{17,18,20,21,22}
		CPU time	*9.82*	*10.47*	*8.91*	*19.02*
p = 4	{1,4,12,17}	Follower's share	**21.06%**	**32.69%**	**42.10%**	**48.60%**
		Follower's hubs	{17,25}	{17,18,20}	{11,17,18,20}	{11,17,18,20,22}
		CPU time	*12.19*	*7.5*	*9.31*	*8.78*
p = 5	{4,7,12,17,24}	Follower's share	**18.19%**	**29.12%**	**36.93%**	**44.24%**
		Follower's hubs	{2,18}	{17,18,20}	{6,17,18,20}	{8,14,17,18,20}

Appendix 4: Hub locations of the leader (*UMApHM*) and the best response of the follower given the hub set of the leader

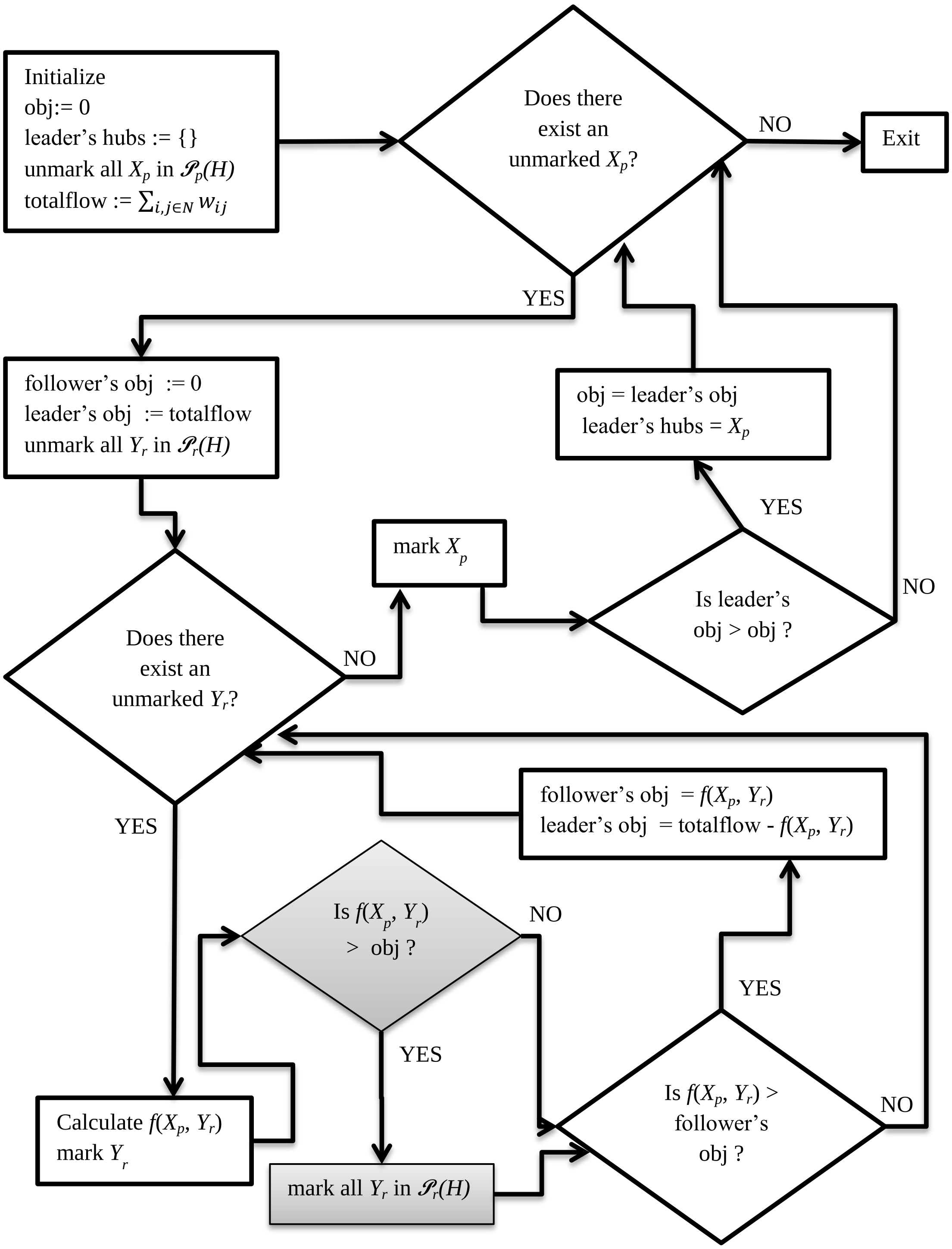
α	***p***	***Xp***	***r = 6***	***r = 8***	***r = 10***
0.6	***6***	{1,6,21,34,35,55}	{3,16,25,27,34,38}	{3,5,6,16,23,27,34,68}	{3,5,6,16,23,27,34,35,61,68}
	8	{1,3,6,21,25,34,35,55}	{6,16,27,34,38,42}	{6,16,20,27,34,38,42,61}	{6,16,20,23,27,34,35,38,42,61}
	10	{1,3,6,16,21,25,34,35,38,55}	{5,20,27,34,42,81}	{5,6,20,23,27,34,42,81}	{5,6,20,23,26,27,34,42,61,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{3,5,34,44,61,81}	{1,3,5,20,23,34,61,81}	{1,3,5,6,20,23,34,61,65,81}
	14	{1,3,6,7,16,21,25,27,34,35,38,42,55,61}	{1,5,20,23,34,81}	{1,5,6,20,23,26,34,81}	{1,5,6,20,26,34,44,65,68,81}
0.8	***6***	{1,6,21,34,35,55}	{3,5,16,27,34,68}	{3,5,6,16,23,27,34,68}	{3,5,6,16,23,27,34,38,68,81}
	8	{1,3,6,21,25,34,35,55}	{16,27,34,38,42,81}	{6,7,16,27,34,38,42,81}	{6,16,20,27,34,35,38,42,61,81}
	10	{1,3,6,16,21,25,34,35,38,55}	{5,7,27,34,42,81}	{5,6,7,27,34,42,44,81}	{5,6,20,27,34,35,42,44,68,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{3,6,34,44,68,81}	{3,5,6,34,44,61,68,81}	{1,3,5,6,20,34,44,61,68,81}
	14	{1,6,7,16,21,25,27,34,35,38,42,55}	{6,20,34,44,61,68}	{5,6,20,26,34,44,61,68}	{1,5,6,20,26,34,44,61,65,68}
0.9	***6***	{1,6,21,34,35,55}	{3,6,16,34,38,81}	{3,6,16,23,34,38,68,81}	{1,3,6,16,23,34,38,42,61,81}
	8	{1,3,6,21,25,34,35,55}	{16,27,34,38,42,81}	{6,16,20,23,34,38,42,81}	{1,6,7,16,23,34,38,42,61,81}
	10	{1,3,6,16,21,25,34,35,38,55}	{27,34,38,42,44,81}	{6,20,27,34,42,44,68,81}	{6,20,27,34,35,42,44,61,68,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{3,6,34,44,68,81}	{1,3,6,34,44,61,68,81}	{1,3,6,20,26,34,44,61,68,81}
	14	{1,3,6,7,16,21,25,27,34,35,38,42,44,55}	{6,23,34,61,68,81}	{5,6,20,23,34,61,68,81}	{1,5,6,20,23,26,34,61,68,81}

α	***p***	***Xp***	***r = 12***	***r = 14***
0.6	***6***	{1,6,21,34,35,55}	{1,3,5,6,16,23,27,34,35,61,68,81}	{1,5,6,16,20,21,25,27,34,35,38,42,61,81}
	8	{1,3,6,21,25,34,35,55}	{1,5,6,16,20,21,27,34,38,42,61,81}	{1,5,6,16,20,21,25,27,34,35,38,42,61,81}
	10	{1,3,6,16,21,25,34,35,38,55}	{1,5,6,20,23,26,27,34,35,42,61,81}	{1,5,6,7,20,23,26,27,34,35,42,61,65,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{1,3,5,6,20,23,26,34,61,65,68,81}	{1,3,5,6,20,23,26,34,35,44,61,65,68,81}
	14	{1,3,6,7,16,21,25,27,34,35,38,42,55,61}	{1,5,6,20,23,26,34,35,44,65,68,81}	{1,5,6,20,23,25,26,34,35,42,44,65,68,81}
0.8	***6***	{1,6,21,34,35,55}	{1,3,5,6,16,23,27,34,38,61,68,81}	{1,3,5,6,16,23,27,34,35,38,42,61,68,81}
	8	{1,3,6,21,25,34,35,55}	{1,6,16,20,23,27,34,35,38,42,61,81}	{1,5,6,16,20,23,27,34,35,38,42,61,68,81}
	10	{1,3,6,16,21,25,34,35,38,55}	{1,6,20,26,27,34,35,42,44,61,68,81}	{1,5,6,7,20,26,27,34,35,42,44,61,68,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{1,3,5,6,20,26,34,44,61,65,68,81}	{1,3,5,6,20,21,26,34,35,44,61,65,68,81}
	14	{1,6,7,16,21,25,27,34,35,38,42,55}	{1,5,6,20,26,34,35,44,61,65,68,81}	{1,5,6,7,20,23,26,34,35,44,61,65,68,81}
0.9	***6***	{1,6,21,34,35,55}	{1,3,6,16,23,27,34,35,38,61,68,81}	{1,3,5,6,16,20,23,27,34,35,38,61,68,81}
	8	{1,3,6,21,25,34,35,55}	{1,6,16,20,27,34,35,38,42,61,68,81}	{1,6,7,16,20,23,27,34,35,38,42,61,68,81}
	10	{1,3,6,16,21,25,34,35,38,55}	{1,6,20,26,27,34,35,42,44,61,68,81}	{1,5,6,7,20,26,27,34,35,42,44,61,68,81}
	12	{1,6,7,16,21,25,27,34,35,38,42,55}	{1,3,5,6,20,21,26,34,44,61,68,81}	{1,3,5,6,20,21,26,34,44,55,61,65,68,81}
	14	{1,3,6,7,16,21,25,27,34,35,38,42,44,55}	{1,5,6,20,23,26,34,35,61,65,68,81}	{1,5,6,20,21,23,26,34,35,55,61,65,68,81}

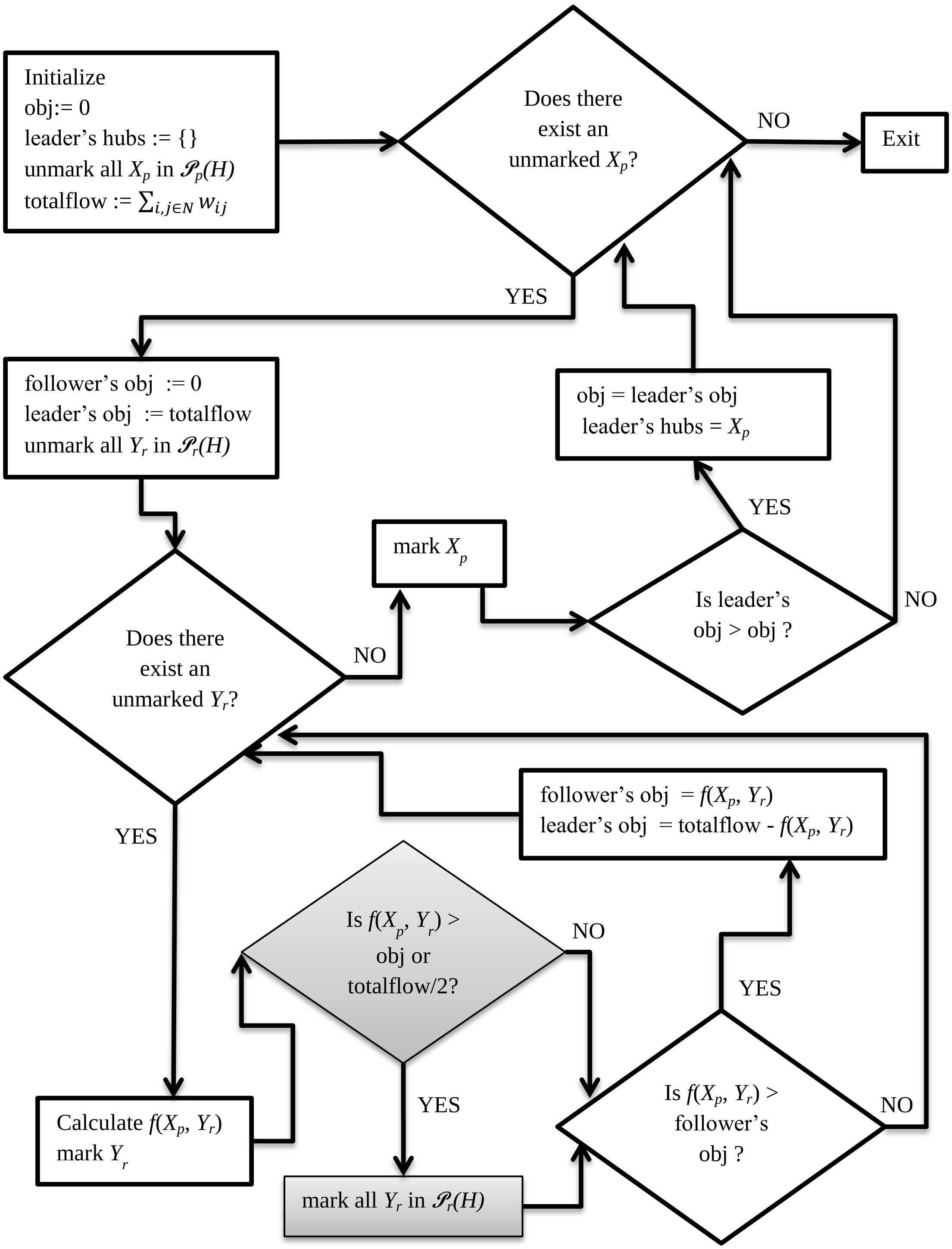
Appendix 5: Flow Chart of *complete enumeration* Algorithm



Appendix 6: Flow Chart of *smart enumeration* Algorithm



Appendix 7: Flow Chart of *smart enumeration with 50%-bound* Algorithm



Appendix 8: Hub locations and percentages of total flow captured by the follower in the optimal solution of *(r|p) hub-centroid* problem on CAB data set

α	**p\r**	**2**	**3**	**4**	**5**
0.6	**2**	{4,17}	{17,21}	{4,17}	{4,17}
	3	{4,17,19}	{4,12,17}	{4,12,17}	{4,12,17}
	4	{1,4,12,17}	{1,4,12,17}	{1,4,12,17}	{1,4,12,17}
	5	{1,4,12,17,25}	{1,4,12,17,20}	{1,4,12,17,20}	{1,4,12,17,20}
0.8	**2**	{4,17}	{4,17}	{4,17}	{4,17}
	3	{4,12,17}	{4,8,17}	{4,12,17}	{4,12,17}
	4	{1,4,12,17}	{4,12,17,25}	{4,12,17,25}	{4,12,17,25}
	5	{2,4,12,13,17}	{4,12,13,17,25}	{4,12,13,17,25}	{4,12,13,17,25}

α	**p\r**	**2**	**3**	**4**	**5**
0.6	**2**	46.14%	64.37%	74.75%	83.52%
	3	30.39%	45.13%	53.69%	62.02%
	4	17.91%	28.39%	37.73%	46.18%
	5	14.30%	23.73%	31.91%	39.58%
0.8	**2**	43.68%	59.59%	70.75%	78.74%
	3	29.18%	42.87%	52.84%	60.14%
	4	21.06%	30.70%	38.39%	45.24%
	5	15.30%	23.24%	31.78%	38.57%

Appendix 9: Hub locations in the optimal solution of *(r|p) hub-centroid* problem on TR data set

α	**p\r**	**2**	**3**	**4**	**5**
0.6	**2**	{6,44}	{1,6}	{6,44}	{16,38}
	3	{6,34,44}	{6,34,44}	{6,27,34}	{6,27,34}
	4	{3,6,27,34}	{3,6,34,44}	{3,6,34,44}	{3,6,27,34}
	5	{1, 3, 6, 23,34}	{1, 3, 6, 25, 34}	{1, 3, 6, 23,34}	{1, 3, 6, 23,34}
0.8	**2**	{38,81}	{6,34}	{6,34}	{6,34}
	3	{6,34,44}	{6,34,44}	{6,34,44}	{6,34,44}
	4	{3,6,34,44}	{3,6,34,44}	{3,6,34,44}	{3,6,34,44}
	5	{1, 3, 6, 23,34}	{1, 3, 6, 23, 34}	{1, 3, 6, 23,34}	{1, 3, 6, 23,34}
0.9	**2**	{38,81}	{6,34}	{6,34}	{34,38}
	3	{6,27,34}	{1,6,34}	{6,27,34}	{6,27,34}
	4	{3,6,34,44}	{3,6,23,34}	{3,6,23,34}	{3,6,23,34}
	5	{1, 3, 6, 34,44}	{1, 3, 6, 34,44}	{1, 3, 6, 34,44}	{3, 6, 23, 27,34}

Appendix 10: Percentages of total flow captured by the follower in the optimal solution of *(r|p) hub-centroid* problem on TR data set

α	p\r	2	3	4	5
0.6	**2**	49.44%	64.65%	74.97%	84.72%
	3	30.49%	40.82%	56.18%	65.58%
	4	20.07%	30.57%	42.15%	51.89%
	5	14.32%	23.61%	32.34%	40.05%
0.8	**2**	46.84%	60.05%	70.03%	77.97%
	3	30.68%	40.81%	51.43%	60.66%
	4	20.33%	30.19%	39.41%	48.57%
	5	14.82%	22.12%	29.28%	37.44%
0.9	**2**	44.12%	58.74%	67.98%	75.45%
	3	30.35%	39.90%	50.03%	58.18%
	4	20.38%	29.55%	38.11%	46.83%
	5	14.27%	22.87%	31.76%	38.91%