A Lagrangean Relaxation Approach for the Modular Hub Location Problem

Seyed Babak Hosseini

A Thesis in the

Department of Mechanical and Industrial Engineering

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science in Industrial Engineering

Concordia University

Montreal, Quebec, Canada

October 2013

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the Thesis prepared,

By: **Seyed Babak Hosseini**

Entitled: **"A Lagrangean Relaxation Approach for the Modular Hub Location Problem"**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Industrial Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________ Chair
Dr. M. Medraj

______________________ Examiner
Dr. M. Kazemi-Zanjani

______________________ Examiner External
Dr. N. Vidyarthi
Decision Science, Management Information System

______________________ Supervisor
Dr. I. Contreras

______________________ Supervisor
Dr. S. Chauhan
Decision Science, Management Information System

Approved by: ______________________

Dr. S. Narayanswamy, MASc Program Director
Department of Mechanical and Industrial Engineering

____

Dr. C.W. Trueman,
Interim Dean
Faculty of Engineering & Computer Science

Date: ______________

Abstract

Hub location problems deal with the location of hub facilities and the allocation of the demand nodes to hub facilities so as to effectively route the demand between origin–destination pairs. Transportation systems such as mail, freight, passenger and even telecommunication systems most often employ hub and spoke networks to provide a strong balance between high service quality and low costs resulting in an economically competitive operation. In this study the *Modular Hub Location Problem* (*Multiple assignments without direct connections*) (MHLP-MA) is introduced. A Lagrangean relaxation method is used to approximately solve large scale instances. It relaxes a set of complicating constraints to efficiently obtain lower and upper bounds on the optimal solution of the problem. Computational experiments are performed in order to evaluate the effectiveness and limitations of the proposed model and solution method.

Acknowledgement

I must express my sincere gratitude to my supervisor, Prof. Ivan Contreras and my co-supervisor, Prof. Satyaveer Chauhan for the patient guidance, encouragement and advices which they have provided throughout my time as their student. I have been extremely lucky to have supervisors who cared so much about my work, and who responded to my questions and queries so promptly.

I would like to take this opportunity to express my deepest gratitude to my lovely parents, sister and friends for their continued support and encouragement throughout my educational and personal life.

Table of Content

List of Figures

List of Tables

Chapter 1: Introduction

Nowadays, transportation is an important science and one of the challenging areas in our daily lives. A huge amount of products or services are being transported all over the world every day. Trucking is one of the most important types of land freight transportation all over the world. For example, 81% of the freight bill is accounted by motor carriers in the United States. Commodities are routed more than 430 billion miles. It costs 372 billion dollars per year (Campbell, 2005). Optimization of transportation networks has been widely studied by researchers and practitioners. Many companies which deal with transportation in different applications try to design the best network to increase their productivity and reduce the total cost of transportation.

Hub Location has become an important area of location theory since 25 years ago and it has attracted many researchers in different fields of study such as geography, operation research, economy and transportation (Taaffe et al, 1996). Hub-and-spoke networks provide an efficient service by establishing hub facilities between origins and destinations (O/D). Hub-and-spoke networks use a set of hub facilities and a reduced number of links to connect a large number of (O/D) nodes.

Hub facilities help to transmit and switch flows which come from non-hub points in a huge network. They are used to take advantage of the economies of scale on inter-hub links. They have changed the way many industries do business (Pirkul, 1998). Hub location problems (HLPs) deal with the design of hub-and-spoke networks and arise when passengers, mails, cargos, and/or data must be transported between every pair of (O/D), but it could cost a lot of money to transport products or services from every single point to another point directly. Using these types of transportation networks reduces the total cost of transportation, but it might increase distance between every pair of (O/D). For example, in computer networks, fiber cables are only used to connect hubs and it is not economic to use them to join any two non-hub nodes. Finding the best locations to open hub facilities is the most challenging and time consuming part for companies. They have to spend a lot of time and it will cost a lot of money to find the best locations and build hub facilities to reduce the total cost of transportation.

As it can be seen in Figure 1, without using hub facilities, origins and destinations are connected to each other directly, and then we have a large number of links between origins and destinations. Suppose that there are *N* nodes in a fully interconnected network and each node might be either an origin or a destination, then there will be $N\ (N-1)$ pairs of nodes in a network between origins and destinations (Daskin, 1995). If trucking companies want to satisfy their customers' needs and deliver the flows, they have to use a truck for each pair of (O/D) and it would be a huge number of trucks to satisfy customers' demands. Consider that hub facilities are used in the network, and then the number of links will be reduced.

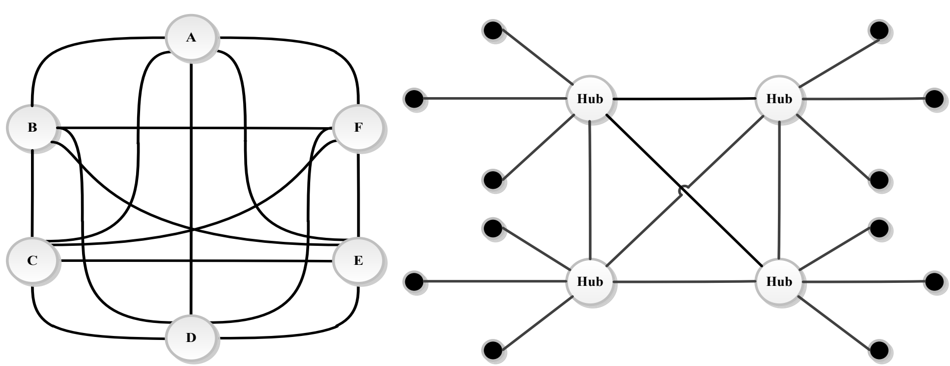

Figure 1: Fully connected network (Daskin, 1995) vs. Hub Networks with 4 hubs

If one node is selected as a hub node and it is connected to all of the other nodes, which are presented as spoke, there will be *2(N-1)* connections to connect all origins and destinations (Daskin, 1995). In this case, trucking companies do not need to use many trucks and they can take advantages of economies of scale. For instance, a general scheme of hub network and routing commodities in truck transportation is shown in Figure 2.

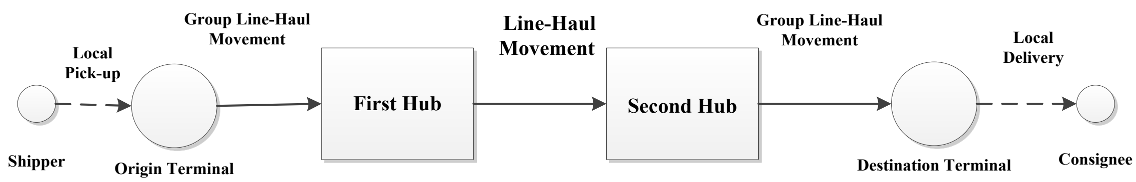

Figure 2: Shipment route through the hub-and-spoke

Similar to trucking companies, airline and postal delivery companies use hub facilities in their networks to take advantage of their benefits. These companies have more traffic in their hub facilities. Many passengers or mails are coming from other cities or facilities to a hub and they are transported from the hub to final destinations or other hub facilities. As it is shown in Figure 1, the number of arcs between origins and destinations are less than in fully interconnected networks. Nodes with large circles are hub nodes and nodes with small circles are non-hub nodes.

Hub facilities might receive flows from non-hub nodes and then transport them to other hub facilities or non-hub nodes. On the other hand, hub facilities could collect flows from other nodes in the network and route them to other hub nodes or non-hub nodes. These are two main functions of hub nodes in the network.

Usually demand is specified as flows of freights or passengers between pairs of (O/D). These flows are transported via trucks, airplanes, ships and fiber cables. In the area of telecommunications, demand is data and information (such as video, voice, etc.). Telecommunication applications include phone networks, computer networks and video teleconferences. Information is transported through a diversity of media such as fiber cables and phone lines (Campbell et al, 2002).

There are several types of hub location problems in the literature. There are a number of differences between them such as the number of hub facilities to locate and the way in which hubs are connected, the allocation of non-hub nodes to hub nodes, and capacity constraints on the hubs or arcs. However, there are four assumptions which most classical hub location problems have in common.

The first assumption is that there is no installation cost for hub arcs, then hub facilities can be fully interconnected in the network. Commodities can be routed via inter-hub arcs, and a discount factor α ($0<\alpha<1$) can be applied to cost of transportation on inter-hub links between every pair of hub facilities. The second assumption is that discount factor is flow independent and it is the same for every amount of flow on all inter-hub links between every pair of hub nodes. The third assumption is that flows must be consolidated by hub facilities. Therefore, the routes between O/D pairs have to contain at least one hub facility. The forth assumption states that the distances between pair of nodes are assumed to be symmetric and to satisfy the triangular inequality.

Hub location problems (HLPs) have been widely studied in the OR community, but the above mentioned assumptions may cause unrealistic results. Suppose that hub facilities are fully interconnected and it might give results that the amount of flow which are routed via inter-hub links are less than the amount of flows which are routed via access links, but discount factor is applied only on inter-hub links. The amount of demand which is transported via inter-hub links could be different, but the same discount factor is applied for all inter-hub arcs in classical hub location models. The assumptions which have been used in classical hub location problems cause few miscalculations in the total transportation costs and the structure of the optimal network. The results might cause to select a non-optimal set of hub facilities. It cloud also lead to assign non-hob nodes to hub facilities in the wrong way.

There are several works have relaxed these assumptions (see for instance, O'Kelly and Bryan, 1998; Kimms, 2006). A model that allows discount factors on hub arcs to be a function of flows is proposed in O'Kelly and Bryan (1998) and it has been further studied in Bryan (1998). The transportation cost in a hub arc as a function of its flow is measured by a nonlinear cost function. The relaxation of fully interconnected assumption reduces the limitations of flow dependent costs (see O'Kelly and Miller, 1994; Campbell et al., 2005 a,b). Consolidation of flow at hubs might also be unrealistic in some applications. Generally, hub facilities are used for consolidation and/or sorting of flows but, in some applications like freight transportation, hub nodes are used only for consolidation proposes. Therefore, both in terms of efficiency (low costs) and effectiveness (high levels of service) it could be better to have a direct connection between two non-hub nodes to route the flow. Some papers have considered the design of the networks based on direct connections between non-hub nodes (see for instance, Aykin, 1994 and 1995; Sung and Jin, 2001).

Mirzaghafour (2013) recently presented a new class of hub location problems, referred to Modular Hub Location Problems (MHLP). These problems overcome the above mentioned disadvantages of classical HLPs. MHLPs do not assume that hub facilities are full interconnected. They use modular costs on every links in the network to calculate the total transportation cost and it is no longer based on flow independency. The total cost of transportation is calculated directly based on the number of facility links selected in the solution network. Moreover, the presented models do not use nonlinear functions when dealing with flow dependent discounted costs. The proposed model is suited to design freight transportation and airline networks. Mirzaghafour (2013) introduced mixed integer programming formulations for four different versions of the MHLP to solve them using a general purpose solver. MHLPs turned out to be much more difficult optimization problems, as instances with only 10 nodes can be optimally solved with CPLEX.

The main contribution of this thesis is to present a Lagrangean relaxation approach that uses a path-based formulation to obtain lower and upper bounds on the optimal solution of the problem. The proposed method relaxes a set of complicating constraints that link the location, design, and routing decisions to obtain a Lagrangean function that can be decomposed into four families of independent sub-problems. Three of these families of sub-problems are knapsack problems, whereas the last ones are simple problems that can be efficiently solved by inspection. Given that the Lagrangean relaxation does not have the integrality property. In general, the obtained lower bounds will be better than the linear programing relaxation (LP) bounds. Moreover, we propose a simple heuristic algorithm to obtain upper bounds. We use the classical subgradient optimization method to solve the Lagrangean dual problem to obtain the best possible lower bounds. We have run a set of computational experiments with several benchmark instances that correspond to real life application coming from a postal delivery network in Australia. The results indicate that the Lagrangean dual problem generates lower bounds that improve on the lower bound associated with the LP relaxation with computational times that are small considering the size and difficulty of the instances. The state-of-the-art optimization software CPLEX can only optimally solve instances with up to 10 nodes, whereas the proposed solution method is able to obtain approximate solutions for instances with up to 50 nodes.

The reminder of this thesis is structured as follows. Chapter 2 presents a review on hub location and basic foundations of solution methodologies used in this document. Chapter 3 introduces the problem definition and mathematical formulations of UHLPs and MHLPs. The Lagrangean relaxation algorithm and the primal heuristic are presented in Chapter 4. Chapter 5 provides the computational experiments and an analysis of the obtained results. Finally, conclusions and directions for future research are given in Chapter 6.

Chapter 2: Preliminaries

In this chapter, network components and characteristics of hub location problems are presented and some real applications are explained. Transportation applications specially trucking transportations is the main focus of this study. Finally, solution methodologies which have been used are explained in detail.

2.1) Literature Review

Hub location has been studied by many researchers for many years. It is one of the main classes of facility location problems. Much research has been done to design hub and spoke networks for different applications in transportation and telecommunications. Goldman (1969) is the first study on hub and spoke networks. The first mathematical formulation of hub and spoke networks, as a quadratic integer programming, has been presented in O'Kelly (1987). Early surveys in this field are in Campbell (1994a) and O’Kelly and Miller (1994). Klincewicz (1998) presents a survey on the location of hubs and the design of hub networks in telecommunication applications, whereas a survey in the area of air transportation have been studied in Bryan and O’Kelly (1999). The facility location problem has been studied widely in Operations Research since the early 1960’s. The goal is to make decisions on the placement of facilities such as factories, warehouses to serve customers efficiently at minimum cost. For an overview of previous work on facility location see Cornuejols et al. (1990). Classical facility location problems and hub location problems have some features in common. They also have few significant differences. Flows are routed via intermediate facilities between pair of nodes in hub location problems. Hub nodes act as consolidation and sorting centers and they need to be linked to each other in order to connect origins to destinations. An example of hub-and-spoke network is shown in Figure 3.

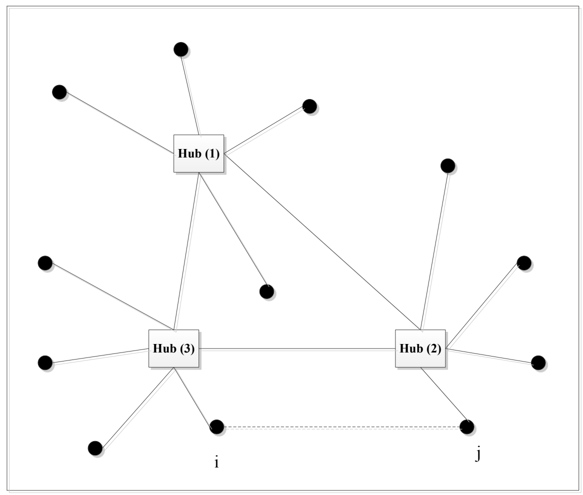

Figure 3: A typical hub location network

On the other hand, demands are sent and received to/from facilities in classical facility location problems and there is no need to connect facilities to each other in the network (see Figure 4).

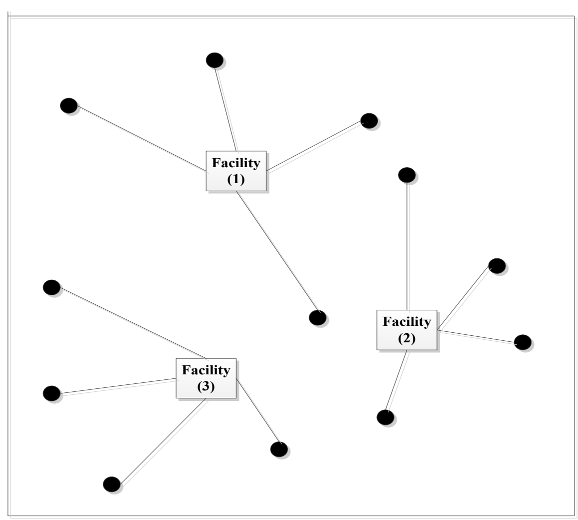

Figure 4: A typical classical facility location network

2.1.1) Hub-and-spoke networks

Hub-and-spoke systems have been used in various industrial applications. Hub-and-spoke networks assist carriers to transport commodities between many pairs of (O/D) at high frequencies and low costs. Hub facilities consolidate commodities before routing them to final destinations and it causes to reduce the number of connections and transportation costs by applying economies of scale between hub facilities. HLPs emphasis on the determination of the location of hub facilities and on the routing of flows through the network so as to minimize the total set-up and transportation cost.

2.1.2) Characteristics of Hub and Spoke Networks

Similar to other systems and structures, hub and spoke (H&S) networks have some advantages and disadvantages.

Advantages

The most important advantages of hub and spoke networks are:

Economies of Scale: The reduction of transportation cost per unit of commodities or passengers caused by the consolidation of demands on larger connections (inter-hub links). Whenever size of service or amount of flows increases, flow cost per unit of commodity decreases.

Economies of Scope: The cost of performing multiple jobs simultaneously is more efficient than performing every job separately. Therefore, hub facilities are susceptible to perform three different roles which are merging, switching and distribution at the same time.

Disadvantages

Hub and spoke (H&S) networks also have some disadvantages:

- Longer travel times and high costs of some routes.
- Capacity overload.
- Higher risk of accident.
- Congestion phenomena.
- Missing connecting facilities due to the unforeseen delay (interrupt) at some parts of the network.

2.1.3) Network Components

In every hub network, some points are selected as non-hub nodes which could be origins and destinations and some points might be selected as hub centers. All origins and destinations are connected to each other by two different types of arcs which are access arcs or inter-hub arcs.

Hub nodes:

Hub nodes are selected among a set of nodes in the network. A hub node might consecutively have three functionalities (see Figure 5):

1) Merging of flows that are received by a hub node, in order to have a larger amount of flows and letting economies of scale to be exploited.

2) Switching (transfer) which allows the flows to be readdressed at the hub node.

3) Distribution (decomposition) of large flows into smaller ones.

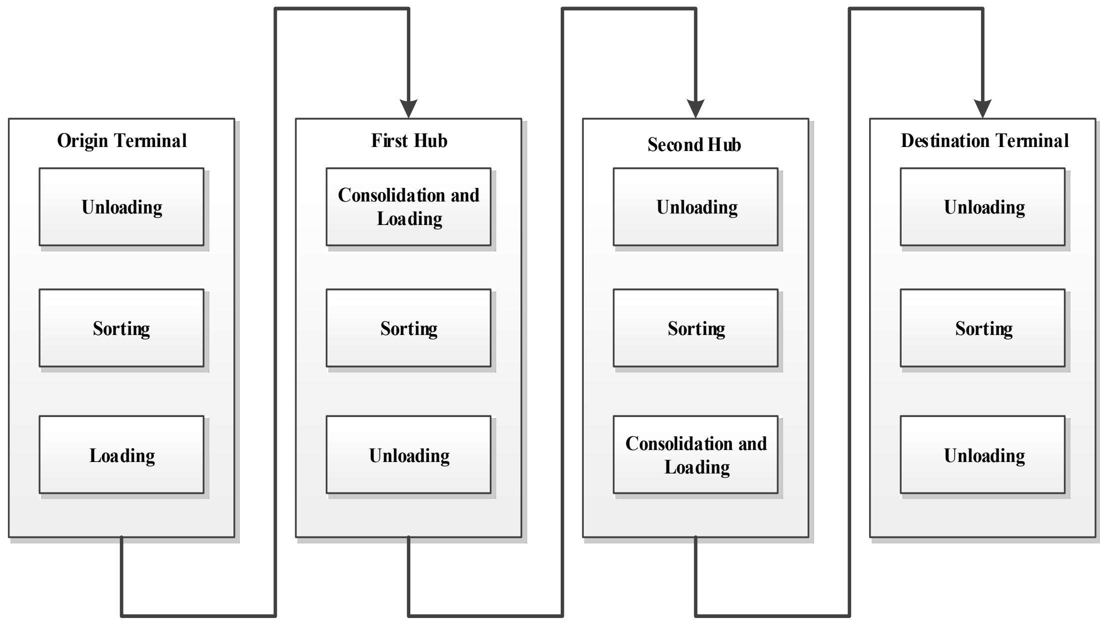

Figure 5: Flow of demands from the senders to receivers

Non-hub nodes: The nodes which do not act as hub centers are non-hub nodes and these types of points can be connected to hub facilities and non-hub nodes in several ways. In some hub location problems, non-hub nodes can be connected to just one facility and in some cases they can be connected to more than one. They can also be connected to other non-hub nodes if it is profitable, but in some cases they are not allowed to have direct connections to other non-hub nodes.

Arcs: Demands are routed from origins to destinations. Origins and destinations are connected to each other by links which are called arcs. As mentioned, every link could have a transportation rate (Campbell, 1998). Links are weighted by discount factors to present collection, transportation, and distribution costs for every unit of flow.

Generally, arcs can be divided to four categories as below:

1- Inter-hub arcs: inter-hub arcs connect hub facilities to each other and they have a discount factor α for the flows which are routed in inter-hub links.

2- Access Arcs (1): these kinds of links are used to connect non-hub nodes to hub facilities. Generally, non-hub nodes which are linked to hub nodes are origins in the networks.

3- Access Arcs (2): these access arcs are applied to connect hub facilities to non-hub nodes which are mostly destinations.

4- Arcs between non-hub nodes: In some models non-hub nodes are allowed to have direct connections if it is necessary and profitable (Aykin, 1994, 1995).

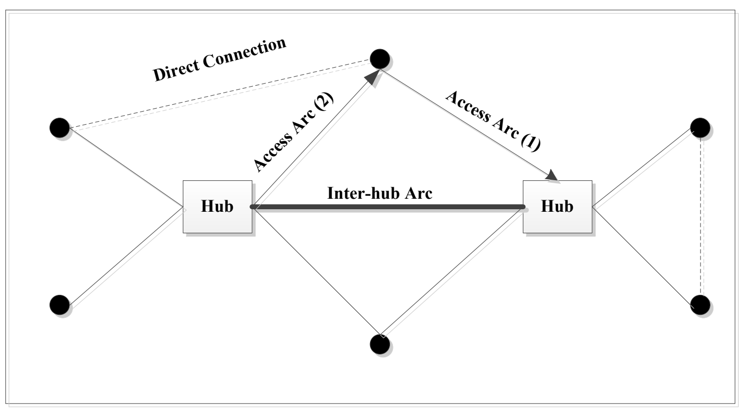

Figure 6: Different types of arcs

Flows: Flows represent products and services that are transported from origins to destinations. Types of flow might be different and they are considered as inputs for the model. Considering the influence of competition rather than assuming a fixed given demand, makes the model more realistic. Amount of flows between origins and destinations is depended on the total cost of travel between each pair of nodes in some models (Marianov et al, 1999). Mails or parcels are demands in postal delivery application. Data or information is denoted as a flow in telecommunication applications. Airline applications have various classes of flows such as passengers, cargos and mails. In other real applications flows or demands are commodities or services that need to be transported.

Constraints

Similar to other problems, hub location problems also have constraints. Some constraints are explained as follows:

Capacity constraints on nodes: Every company which is dealing with transportation has capacity constraints in its demand centers and hub facilities. For example, in a hub facility of a trucking company, many trucks are coming and leaving to load or unload commodities. There are specific numbers of docks and they cannot serve unlimited trucks to load or unload. The same situation happens for other applications such as postal delivery and airline applications. A postal delivery company might be able to sort a limited number of mails in its hub facilities. It sorts a maximum numbers of mails which is possible to do in a hub facility. Passengers or commodities are arriving or departing from many other cities or countries to a hub in airline applications. There is a limited number of terminals to serve aircrafts in the airport and it is not possible to load or unload an unlimited number of passengers or cargos.

Capacity constraints on arcs: This type of capacity defines the amount of flow that can be routed on arcs. From another point of view, it presents an upper bound on the amount of demand which can be transported on every arc. The relationship between capacities of inter-hub links has been studied in Bryan (1998).

Performance constraints: Performance constraints are important to handle demands in the system. They are applied to be sure that hub network works efficiently and it is possible to control the traffic. These are mostly used in telecommunication systems, including restrictions on the percentage of calls blocked because of not having enough capacity. Klincewicz (1998) considers performance constraints in transportation applications and Marianov and Serra (2000) proposes a model in airline application that considers a constraint on the length of the queue of aircrafts waiting for a runway at a hub.

2.1.3) Models and classification of Hub Location Problems

To solve more realistic problems, several authors have studied different aspects of the classical hub location problems. Various kinds of problems have been analyzed such as capacitated or uncapacitated problems, single allocation or multiple allocations, and models which non-hub nodes are allowed to have direct connections between each other. These problems can be classified based on the type of objective they consider.

Objective

Most HLPs have Cost oriented and/or service oriented objective functions. Minimizing the total cost is one of the important goals for most HLPs which are considered in literature. The different types of costs are considered for various applications. For instance, shipment of the right amount of demands to reduce the total cost of transportation is what all transportation enterprises are dealing with, but in telecommunication applications, fixed costs to construct the hubs and connections are concerned. Moreover, service oriented objective functions such as traveling time and coverage measures have been studied in literature, For instance, minimizing the maximum travel time between all O/D pairs in the network can be an objective function to reduce service time.

Variants of Hub Location Problems

Some of the most important classes of HLPs are:

1- ρ-Hub Median Problems (ρHMP)

Given a fixed number of hubs (ρ), the objective is to find the best location for ρ hub facilities so as to minimize the total transportation cost. p-hub median problems are studied in two different subgroups:

- Single allocation
- Multiple allocations.

Single allocation: Every non-hub node is connected to just one hub facility in single allocation model. The first linear integer programming formulation for the single allocation p-hub median problem has been introduced in Campbell (1994b). The most computationally efficient exact solution procedure is the shortest path based branch-and-bound algorithm presented in Ernst and Krishnamoorthy (1998b).

Multiple Allocations: Every demand center is allowed to send and receive flows to/from more than one hub facility in the multiple allocation problems. The first work to formulate the multiple allocation p-hub median problems as a linear integer program is Campbell (1992). Several authors have worked on p-hub median problems with multiple assignments such as Skorin-Kapov et al. (1996), Ernst and Krishnamoorthy (1998a, 1998b), and Boland et al. (2004).

Figure 7 and Figure 8 shows hub networks with single assignment and multiple assignments.

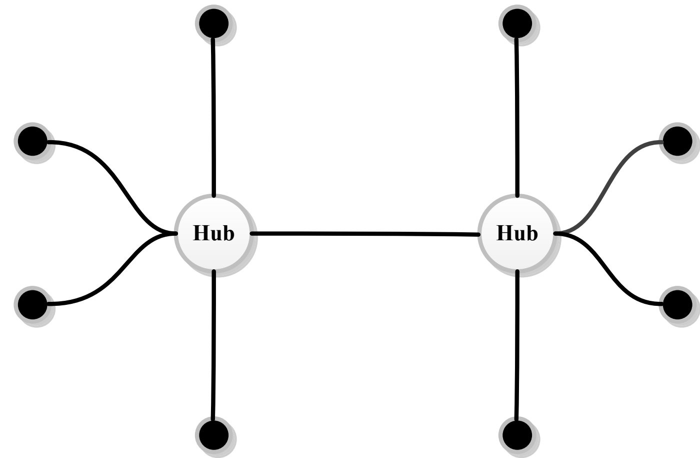

Figure 7: Hub networks with a single allocation

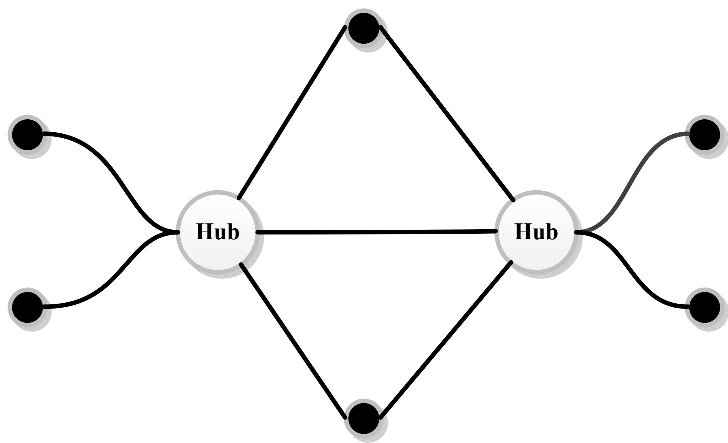

Figure 8: Hub networks with multiple allocations

2- The Hub Location Problems with Fixed Costs

The number of hub facilities is unknown in hub location problems with fixed costs. Fixed costs and variable costs are considered and these cause to reduce the total cost of transportation. Whenever the number of hub facilities increases, the total cost of opening of hub facilities increases, but because of short distances between hub nodes and non-hub nodes the total transportation cost of demands decreases. For more references refer to

Campbell (1994b), Abdinnour Helm and Venkataramanan (1998), Topcuoglu et al. (2005), Cunha and Silva (2007), and Chen (2007).

3- The p-hub Center Problem

The *p*-hub center problem is stated as a mini/max type problem. Generally, the main goal of the *p*-hub center problem is to place *p* hub facilities and to assign all non-hub nodes to the located hub facilities in order to minimize the maximum cost (time, distance) between any pair of origin and destination nodes. The first formulation of the *p*-hub center problem has been studied in Campbell (1994b). Single and multiple allocations of p-hub center problem have been studied in Kara and Tansel (2000), Ernst et al. (2009), and Meyer et al. (2009).

4- Hub Covering Problems

Demand nodes are considered to be covered if they are located within a specified distance of a hub facility in hub covering problems. Every pair of (O/D) is covered by hubs *k* and *m* if the cost of transportation from origin to destination via hubs *k* and *m* does not exceed a specified value. The first mixed integer model for the hub covering problem was introduced in Campbell (1994b). Kara and Tansel (2003), Ernst et al. (2005) present the single allocation hub covering problem and new models for both single and multiple assignments hub covering problems have been studied in Wagner (2008).

2.1.4) Potential Applications

As mentioned, hub network systems are widely used in different areas of application. Two of the most important and well known areas of hub location problems are telecommunication and transportation. The main objective for almost all hub networks is to reduce transportation cost and improve frequency of service. The most significant differences between transportation and telecommunication areas are their flows and costs. In transportation applications, such as public transportation, air passenger, air freight, express shipment, trucking, postal delivery and rapid transit, the demands are physical flows and they are in form of passengers or goods. They are transported by many different transportation vehicles such as buses, trucks, trains, taxis and planes. In telecommunication applications the flow which is routed on links is data or information. Data and information is transferred from origins to destinations via wires or optic fibers. In transportation applications the main issue is to reduce the total cost of distribution of products or services, but reducing the total expenses of building the network is the main concern in telecommunication applications.

Some of the most important applications are explained briefly as follows:

Trucking: Trucking application is one of the most important applications in hub location problems and it has been studied a lot by many authors. Trucking has two types of transportation which are Less-than-truckload (LTL) transportation and Full Truckload (FTL or TL) transportation. The goal of the transportation methods is to transfer shipments from origins (O) to destinations (D) in an effective way. The difference between LTL transportation and TL transportation is significant. Generally, truckload transportation is used for a large load to a destination or some destinations which are very close. Because truckload freights are quite large, there is no chance for the consolidation of freights from several origins because of vehicle storage space and weight limitations. Normally, truckload companies use the largest possible vehicles to ship larger orders at one time. Less-than-truckload transportation has considerable differences from an operational and physical point of view from truckload. Hub facilities might not be used for consolidation and it could be used to reduce trucks' traveling distances (Hunt, 1998). LTL enterprises do business with all small and large businesses. Transporting as much small freight as possible from several origins to destinations is the objective of LTL. The LTL carriers might pick up freights from origins and transport them to hub facilities. At hub facilities, shipments from several origins sort and consolidate into vehicles which are larger. The large trucks will transport the shipments to another hub facility, where each shipment will be categorized and sent to its respective destination. Trucks which are used between hub facilities on inter-hub links in LTL transportation are similar to trucks which are used for LT and they have similar weight and volume constraints. In many cases, LTL companies may apply TL strategies on inter-hub links which they have long distance routes. Several studies have been done about trucking applications in hub location. Taha et al. (1996) shows that several hub facilities provides better results than having just one hub facility in the network or transporting flows from origins to destinations directly. Other studies in trucking application which can be mentioned are Taha and Taylor (1994); Taha et al. (1996), Taylor et al. (1995), Taylor et al. (1999), Powel (1986), Powell and Sheffi (1983).

Air transportation: Airline applications are also one of the most important areas in transportation. These can be separated in two groups which are passenger airline and freights airline. Passengers expect to have comfortable trip and experiencing the convenient trip for passengers is the most important issue that airline companies are facing with, but the total cost and performance are the significant issues for freight airlines. Some of the most important differences between passenger airlines and freight airlines have been mentioned in O'Kelly (1998). Pricing is an issue for passenger airlines. Airline companies usually compete with each other on the price of travel. Marinov et al. (1999) proposes some ideas on pricing both for passengers and freights. Discount for large volume of demands can be achieved by using larger aircrafts in inter-hub connections and some authors presented different models and all inter-hub connections are discounted in their models. Jaillet et al. (1996) uses different types of aircrafts for different types of arcs. Larger aircrafts with larger capacity are used in inter-hub connections and smaller ones are used in access arcs.

Rapid transit: Mathematical formulations and solution methods have been proposed in Gelareh and Nickel (2007) and Nickel et al. (2001) for rapid transit systems. Gelareh (2008) presents many variants of the hub location problems with a diversity of hub level structures in specific addressing rapid transit planning. The first multi-period hub location problem for the rapid transit application has been proposed in Gelareh and Nickel (2008). Exact and heuristic solution approaches were developed for both single and multi-period models, and results show that both are very efficient.

Postal network: Postal delivery applications are similar to other applications specially airline application but with some differences. Mails and parcels are sent from several origins and they are sorted and consolidated at hub facilities and finally, they are routed to different destinations. Australia Post has been discussed in Ernst and Krishnamoorthy (1996, 1999). Australia Post provided set of data to analyze hub location models. There are few differences between postal network and airline network. For instance, origin nodes could send mails and parcels to themselves in postal applications. First of all, demands are transported to hubs where flows are sorted. Finally, they are returned to the origin nodes. Additionally, Capacities represent as the whole collection of demands which are sorted at hub facilities in postal applications.

Telecommunication: Hub networks are widely used in telecommunication applications. In telecommunications the most significant issue which is concerned is the establishment cost more than the communication cost of the networks. Minimizing the total fixed cost of designing the network is the main objective. Unlike other applications, in telecommunication applications, flows are not tangible. The flow corresponds to data and information that is transported via wires or optic fibers. Access nodes which represent the tributary network denote as origins and destinations must be transported through transit or backbone networks which are transit nodes. All traffic which is departure from an access node should be passed through transit nodes on the way to its destinations. Every access node has to route traffic to one or two transit nodes that transport the traffic to several destinations. There are two types of costs. Fixed costs which are the costs of opening a transit node, and connection costs which are the costs of installing on each edge the capacity required to transport the flow on the edge itself. The problem is to decide the number and location of the transit nodes and assigning access nodes to the right transit nodes to minimize the total cost of the network.

Many articles have been published on similar problems. Gavish (1991) proposes a telecommunication application which is about configuration of distributed computer systems. A model of large scale data for communication network has been designed in Chung et al. (1992). Many different and interesting telecommunication networks can be found in Yoon et al. (1998a, 1998b), and Yoon and Tcha (1996). A survey on backbone and tributary network design problems has been done in Klincewicz (1998).

2.2) Solution Methodology

One of the best known solution methods to solve large scale networks is the Lagrangean relaxation method. It leads to achieve better results by eliminating a set of difficult constraints. In the following section, the proposed solution methodology is explained in detail.

2.2.1) Lagrangean Relaxation

Lagrangean relaxation (LR) method is one of the most useful techniques to solve large-scale optimization problems, mainly nonlinear programming and integer programming problems. One of the key features of LR is that problem can be usually decomposed in many independent sub-problems. Solving each sub-problem is easier than solving the whole problem with many constraints (Guignard, 2003). LR method was originally introduced by Held and Karp (1970, 1971). The main idea behind LR method is to remove a set of difficult constraints and add them to the objective function. The results presented in this section are mainly derived from Held and Karp (1970, 1971), Wolfe and Crowder (1974), Geoffrion (1974), Fisher (1981), and Guignard (2003).

Consider the Linear Mixed-Integer Programming problem as:

$$\begin{aligned} z^* = min \quad & cx + dy \\ subject\ to \quad & Ax + Dy \geq b, \\ & x \in Z_+^n \\ & y \in R_+^p \end{aligned}$$

Where Z_+^n is the set of nonnegative integer n-dimensional vectors, R_+^p is the set of nonnegative real p-dimensional vectors, $x = (x_1 \ldots x_n)$ is a vector of integer variables, and $y = (y_1 \ldots y_p)$ is a vector of real variables.

The feasible region of MIP is given by the set $S = \{(x,y) \in Z_+^n \times R_+^p : Ax + Dy \geq b\}$, and any $(x,y) \in S$ is called a *feasible solution* of the problem. A given instance is feasible if $S \neq \emptyset$.

The function z is called the objective function.

$$z = cx + dy$$

An optimal value (x^*, y^*) is a feasible point for which the objective function value is as less as possible, that is

$$cx^* + dy^* \leq cx + dy \qquad \forall (x,y) \in S$$

and $z^* = cx^* + dy^*$ is the optimal value of the solution.

The linear programing problem (LP) is a special case of MIP. All integer variables are relaxed in LP and it does not have any integer variables.

$$\begin{aligned} z^* = min \quad & dy \\ \textit{subject to} \quad & Dy \geq b, \\ & y \in R_+^p \end{aligned}$$

The aims of solving IP problems by algorithms, is to find a lower bound $z \leq z^*$ and an upper bound $z \geq z^*$. Algorithms' goal is to find an increasing sequence of lower bounds and a decreasing sequence of upper bounds. Algorithms stop when the difference between the lower bound and the upper bound is within a threshold value. Therefore, our objective is to find ways of obtaining such bounds.

In the case of upper bounds, every feasible solution $x_i \in S$ provides an upper bound $z = c(x_i) \geq z^*$. Finding feasible solutions is simple for some IP problems, and the major issue is to find a way to obtain good feasible solutions which provides the best (lowest) upper bound.

In the case of lower bounds, relaxation and duality are two popular ways of finding them. The main idea behind the relaxation method is to replace a difficult IP problem by an easier optimization problem whose optimal value is at least as small as z^*.

Definition 1: Problem (*RP*) $z^R = \min \{g(x) \mid x \in W \subseteq R^n\}$ is a relaxation of problem (*P*) $z^* = \min \{f(x) \mid x \in S \subseteq R^n\}$, with the same decision variables *x*, if and only if:

1. $S \in W$, and
2. $g(x) \leq f(x)$, $\forall\, x \in S$.

Based on the conditions (1) and (2) from Definition 1, the following proposition can be established:

Proposition 1: if *RP* is a relaxation of *P*, $z^R \leq z^*$.

There are several approximation methods which can be successfully applied to integer programing problems. They provide approximate solutions since optimality of the obtained solutions cannot be proved. Some types of solution methods, such as heuristics and metaheuristic, focus on finding feasible solutions. They are used to find upper bounds on the optimal solution value of the problem. There are some other methods, such as relaxations or decomposition methods. They focus on obtaining lower bounds on the optimal solution value.

One of the most useful and natural relaxations of IP is the linear programming relaxation, where integrality constraints are not considered from the model any more.

There are other relaxations for IP problems such as Lagrangean Relaxation which is explained as follows:

Consider the problem IP, which is called the integer programing problem:

$$
\begin{aligned}
\text{(IP)}\quad z^* = min \quad & cx \\
\textit{subject to} \quad & Ax \geq b \\
& x \in X = \{x \in Z_+^n : Dx \geq d\}
\end{aligned}
$$

(A,b) and (D,d) are $m \times (n+1)$ and $r \times (n+1)$ matrices, and x is an n-vector of non-negative integers ($x \in Z_+^n$). X is a set of discrete points in a polyhedron. The problem IP is called the primal problem and its solution is called a primal solution. Consider that the constraints $Ax \geq b$ are complicated constraints to solve, and problem IP would be solved easier without them. A common method to solve IP is to solve its Lagrangean dual problem obtained via LR. In the LR method, constraints which are complicated to solve ($Ax \geq b$) are relaxed by presenting a vector $u \in R_+^m$ and it is called Lagrangean multiplier and $L(x,u)$ is Lagrangean function.

$$L(x, u) = cx + u(b - Ax)$$

The LR problem is then to solve the following:

$$
\begin{aligned}
LR(u)\quad \varphi(u) = min \quad & L(x,u) \\
\textit{subject to} \quad & x \in X
\end{aligned}
$$

It is easy to prove that for any $x \in X$, $u \in R_+^m$ and any optimal solution x^* to IP it holds that:

$$\varphi(u) \leq L(x,u) \leq cx \quad and$$

$$\varphi(u) \leq L(x^*, u) \leq cx^* = z^*$$

Efficient solution methods to solve sub-problem $LR(u)$ and the fact that $\varphi(u) \leq z^*$ allows $LR(u)$ to be used to provide lower bounds for IP. Different lower bounds $\varphi(u)$ on the optimal value z^* are provided by different values of u. Any value of u which provides the greatest lower bound of z^* is an optimal solution to the Lagrangean dual (LD) problem:

$$(LD)\ \ \varphi^* = max\ \{\varphi(u) : u \geq 0\}$$

$\varphi(u)$ is an implicit function of *u:*

$$\begin{aligned} \varphi(u) = min \quad & cx + u(b - Ax) \\ subject\ to \quad & x \in X \end{aligned}$$

LD is a problem in the dual space of the Lagrangean multipliers u, whereas $LR(u)$ is a problem in the x space. Those complicated constraints which are equality constraints the multipliers u are not restricted in sign ($u \in R^m$).

Definition 2: *LR* has the integrality property if:

$$Co\{x \in Z_+^n : Dx \geq d\} = \{x \in R_+^p : Dx \geq d\}$$

The consequence of this property is that when the integrality property holds, the LR scheme cannot produce a bound stronger than the LP bound. However, this is useful because sometimes the LP bound can be computed more efficiently using a LR scheme than the traditional linear programming methods such as primal simplex, dual simplex, and interior point methods.

The consequence of mentioned property is as stated in the following corollaries:

Corollary 1: If $Co\{x \in Z_+^n : Dx \geq d\} = \{x \in R_+^p : Dx \geq d\}$, then

$$g^* = \varphi^* \leq z^*$$

g^* is the optimal value of the LP relaxation.

Corollary 2: If $Co\{x \in Z_+^n : Dx \geq d\} \subset \{x \in R_+^p : Dx \geq d\}$, then

$$g^* \leq \varphi^* \leq z^*$$

It may happen that the LR bound is strictly better than the LP bound. Unless LR does not have the Integrality Property, it will not yield a stronger bound than the LP relaxation.

The most challenging part of using LR is to optimize efficiently the LD function. There are several primal and dual methods to solve LD either exactly or approximately. Some of these methods are subgradient method, Outer approximation method and Bundle method (see, for instance, Guignard, 2003). Subgradient type methods utilize a subgradient of φ to find a direction of movement.

Consider that u^* is the best (greatest) value of LR and with $UB^* = z(u^*)$, and let u^{k+1} be the projection of u^k . The step direction from a given point u^k is just the subgradient of the objective function.

$$\gamma(u^k) = (b - Ax^k)$$

The scalar t_k is a step size specifying how far we move from the current solution and it is positive. The step size which is generally used in practice is:

$$t_k = \frac{\lambda^k(UB - z(u^k))}{\|\gamma(u^k)\|^2}$$

λ^k is a scalar and $0 \leq \lambda^k \leq 2$.

So that:

$$u^{k+1} = u^k + t_k \ \gamma(u^k)$$

Finally, u^{k+1} should be non-negative.

$$u^{k+1} = max \ (0, u^{k+1})$$

The formula uses the unknown *UB*. An estimated value of *UB* can be too small or large. If the estimated value of *UB* is too small, then steps could be too small and convergence would be slow. If the large *UB* is used, then it is projecting on a hyperplane which is too far away from u^k and it might be beyond the u^k. If the values of objective function do not improve for a large number of iterations, then the upper bound might has been underestimated. Therefore, the difference $UB - \varphi(u^k)$, should be reduced by multiplying it by a factor λ^k less than 1.

$u^{k+1} = u^k + \gamma(u^k). \frac{\lambda^k(UB - Z_D (u^k))^2}{\|\gamma(u^k)\|^2}$ where λ^k is reduced when there is no improvement for many iterations. For more references refer to Lemarechal (1974).

A scheme of the subgradient algorithm

Iteration 0

- Initialize $\varphi(u) = -\infty;\ u^0 = 0;\ \lambda^k = 2.$
- Imagine that UB is a known upper bound over the optimal value.

Iteration *k*

- Solve the lagrangean function $L(u)$.
- If $(L(u^k) > \varphi(u)$) then:

  $\varphi(u) = L(u^k)$.

- End if
  - Calculate the subgradient $\gamma(u^k)$.
  - Compute the step size $t_k = \frac{\lambda^k(\ UB - L(u))}{\|\gamma(u^k)\|^2}$.
  - $(u^{k+1}) = (u^k) + t_k\, \gamma(u^k)$.
  - *k = k +1*

It was pointed that λ^k is reduced if there is no improvement for the particular number of iterations.

2.2.2) Heuristic

A method which is based on a role or a set of rules, and is used to construct a feasible solution is called a heuristic. Greedy and local search procedures are the most simple heuristics. A greedy heuristic aims to construct an initial feasible solution, but local search method improves some initial solutions. A lot of research has been done to develop heuristic methods that overcome local optimality. A metaheuristic is an algorithmic framework that provides a set of strategies to develop heuristic optimization algorithms. The aims of metaheuristic algorithms are to find the best (feasible) solution out of all possible solutions of an optimization problem. Several metaheuristic have been introduced and developed by authors in the area of operation research such as tabu search (Glover, 1989 ; 1990), genetic and evolutive algorithms (Holland, 1975; Michalewicz, 1992), simulated annealing (Kirpatrick et al., 1993), scatter search (Laguna and Martik, 2003), and greedy randomized adaptive search procedures (Feo and Resende, 1995). Several different Lagrangean heuristics also have been presented in the literature. Whenever LR method is used to solve problems, metaheuristic and heuristics are developed to construct feasible solutions based on information that LR provides. In this study primal heuristic is used to construct feasible solution. It provides upper bounds on the optimal values, and the best (lowest) upper bound will be selected. The gap between lower bound and upper bound is calculated and it determines that how far the results are from optimal values. Primal heuristic and algorithm to construct a feasible solution is explained in Chapter 4.

Chapter 3: Problem Definition and Formulations

Uncapacitated Hub Location Problems (UHLP) have been widely studied in the literature and are one of the most important types of HLPs. The amount of capacity is not considered in uncapacitated version of HLPs.

3.1) Uncapacitated Hub Location Problem

Let $G = (N,A)$ be a complete graph, where $N = \{1,2,3,\ldots n\}$ represents the set of nodes which can be origins, destinations and potential hub facilities in the network. A is a set of arcs in the network. Demands are routed between origins and destinations (i,j) and W_{ij} denote the amount of flow. f_i is a fixed cost of installation of a hub facility. Distances between origins and destinations are shown by d_{ij}. Distances satisfy the triangular inequality and transportation cost for each unit of flow is related to the distance between every pair of nodes. Hub facilities are assumed to be fully interconnected and a discount factor ($0<\alpha<1$) is applied to calculate transportation cost between hub facilities on inter-hub links and it causes the economies of scales in the network. The total cost of transportation in inter-hub links is less than the total cost of transportation between hub nodes and non-hub nodes because of the discount factor. Selecting hub facilities among a set of nodes and assigning non-hub nodes to the right hub facilities lead to minimize setup cost and transportation cost. As mentioned, hub facilities are fully interconnected and it is assumed that every non-hub node has to be connected to at least one hub node. Each route has to include at least one hub node and at most two hub facilities. First of all, flows have to be routed to a hub node from a non-hub node which is an origin (*O/i*) and then they are routed from a hub node (k) to another hub facility (m) if it is necessary.

Finally, flows have to be driven from the last hub node to a non-hub node which is a destination (*D/j*).

If flows pass through the path of *i-k-m-j*, then the total transportation cost of routing flows from an origin (*i*) to a destination (*j*) is as follows:

$$F_{ikmj} = W_{ij}\left(d_{ik} + \alpha d_{km} + d_{mj}\right).$$

We define the following sets of decision variables:

$X_{ijkm} =$ flows between nodes *i* and *j* which is routed via inter-hub arc *k* and *m*

$$Y_k = \begin{cases} 1, & \text{if a hub facility is located at node } k \\ 0, & \text{otherwise} \end{cases}$$

The UHLP can be formulated as:

min $\sum_i \sum_j \sum_k \sum_m F_{ikmj} X_{ijkm} + \sum_k f_k Y_k$

subject to

$$\sum_k \sum_m X_{ijkm} = 1 \qquad \forall i, j \in N \qquad (1)$$

$$X_{ijkm} \leq Y_k \qquad \forall i, j, k, m \in N \qquad (2)$$

$$X_{ijkm} \leq Y_m \qquad \forall i, j, k, m \in N \qquad (3)$$

$$Y_k \in \{0,1\} \qquad \forall k, \in N \qquad (4)$$

$$X_{ijkm} \geq 0 \qquad \forall i, j, k, m \in N \qquad (5)$$

The objective function minimizes the installation cost of potential hub facilities and the total transportation cost. Constraints (1) guarantee that there is a unique route for routing the demands between every pair of origin and destination. Constraints (2) and (3) prohibit the flow to be routed via a node which is not a hub. Therefore, an optimal value have $X_{ijkm} \geq 0$ since the total demand for every pair of origin and destination must be passed via the least cost hub pair. Constraints (4) and (5) are the classical integrality and non-negativity constraints.

3.2) Modular Hub Location Problems

As mentioned in Chapter 1, the main objective of MHLPs is to design hub-and-spoke network by estimating more accurately the total transportation cost. Difference between classical UHLPs and MHLPs is the connection of hub nodes. In classical HLPs, it is assumed that hub nodes are fully interconnected at no costs in the network, but we do not consider a fully interconnected hub network at no cost in MHLPs. MHLPs also considers installation costs for the both access links and inter-hub links. In modular hub location problems, flow dependent modular cost is applied on every link in the network instead of using fixed discount factor for every hub arc in the network. These new modeling features lead to calculate more accurate and reliable transportation costs.

Mirzaghafour (2013) introduced four different variants of the MHLP, which differ according to the way O/D nodes are connected to hub facilities and whether it is allowed to directly route flows from their origin to their destination. In this study, we consider the MHLP with multiple assignments without direct connections between non-hub nodes. As mentioned, modular hub location problems are more practical and accurate than capacitated or uncapacitated classical HLPs. Flows are routed via one or more than one hub facility, and non-hub nodes could be connected to more than one hub node, but non-hub nodes are not allowed to have direct connections to each other.

Let $G = (N,A)$ be a complete graph where $N = \{1,2,3,\ldots,n\}$ denotes the set of nodes in the network and A is a set of arcs which they connect origins to destinations. Hub facilities are not fully interconnected and distances between origins and destinations are not assumed to satisfy the triangular inequality property. Non-hub nodes are not allowed to have direct links to each other. The total transportation cost will be calculated based on the amount of flow which is routed via both access and inter-hub links. The amount of flow determines the number of facility links in the network. W_{ij} represent the amount of flow which is transported from the origin i to the destination j, and f_i denote the fixed setup cost of a hub. d_{ij} represent the distance between origin i and destination j. Demands are routed by facility links and there are two different type of them. Large facility links are used to transport flows between hub facilities. Whereas, smaller facility links are used to transport flows between origins and hub facilities or between hub facilities and destinations. Transportation cost of each facility link between hub nodes k and m is calculated by:

$$c_{km} = l_c + bd_{km},$$

Where l_c represents the fixed cost of buying or leasing a truck and b represents variable cost which can be labor and fuel costs. The capacity of every large truck is B.

Transportation cost for small trucks between nodes k and m in access arcs is also similar to large ones which is:

$$q_{km} = l_q + pd_{km},$$

Where l_q and p are fixed and variable costs of small trucks with capacity of H.

Variable and fixed cost of larger trucks is more than small trucks. Capacity of trucks on inter-hub arcs is more than capacity of trucks on access arcs ($B > H$, $b > p$, $l_c > l_q$) then, transportation cost of per unit of demands on the inter-hub arcs is less than transportation cost of per unit on the access arcs. $\frac{c_{km}}{B} < \frac{q_{km}}{H}$.

The maximum number of large facility links which can be routed on inter-hub arcs is represented by Q_1 .

Q_2 and Q_3 illustrate the maximum number of small facility links which can be routed on access arcs.

First of all, for modeling the MHLP, some decision variables will be defined as:

$$Z_k = \begin{cases} 1, & \text{if a hub facility is located at node } k \\ 0, & \text{otherwise} \end{cases}$$

a_{ijk} = flows between nodes i and j use access arc of (i,k)

s_{ijk} = flows between nodes i and j use access arc of (k,j)

X_{ijkm} = flows between nodes i and j use inter-hub arc of (k,m)

y_{km} = number of trucks between hub nodes k and m

v_{ik}^{1} = number of trucks between non hub node i and hub node k

v_{kj}^{2}= number of trucks between hub node k and none hub node j

MHLP with multiple assignments without direct connections can be formulated as follows:

$$min \ \ \sum_{k\in N} f_k\, z_k + \sum_{i\in N}\sum_{k\in N} q_{ik} v_{ik}^1 + \sum_{m\in N}\sum_{j\in N} q_{kj} v_{kj}^2 + \sum_{k\in N}\sum_{m\in N} c_{km} y_{km}$$

subject to

$$\sum_{k\in N} a_{ijk} = 1 \qquad \forall i,j \in N; i \neq j \quad (1)$$

$$\sum_{k\in N} S_{ijk} = 1 \qquad \forall i,j \in N; i \neq j \quad (2)$$

$$y_{km} \leq Q_1 z_k \qquad \forall k,m \in N \quad (3)$$

$$y_{mk} \leq Q_1 z_k \qquad \forall k,m \in N \quad (4)$$

$$\sum_{i\in N}\sum_{j\in N} W_{ij} x_{ijkm} \leq B y_{km} \qquad \forall k,m \in N \quad (5)$$

$$\sum_{j\in N} W_{ij}\, a_{ijk} \leq H v_{ik}^1 \qquad \forall i,k \in N \quad (6)$$

$$\sum_{i\in N} W_{ij}\, S_{ijk} \leq H v_{kj}^2 \qquad \forall k,j \in N \quad (7)$$

$$v_{ik}^1 \leq Q_2 z_k \qquad \forall i,k \in N \quad (8)$$

$$v_{kj}^2 \leq Q_3 z_k \qquad \forall k,j \in N \quad (9)$$

$$a_{ijk} + \sum_{m\in N} x_{ijmk} - \sum_{m\in N} x_{ijkm} - S_{ijk} = 0 \qquad \forall i,j,k \in N; i \neq j \quad (10)$$

$$z_k \in \{0,1\} \qquad \forall k \in N \quad (11)$$

$$v_{ik}^1, v_{kj}^2, y_{km}, y_{mk} \in \mathbb{Z}^+ \qquad \forall k,m,i,j \in N \quad (12)$$

$$0 \leq a_{ijk}, X_{ijkm}, s_{ijk} \leq 1 \qquad \forall i,j,k,m \in N; i \neq j \quad (13)$$

The objective function minimizes the total cost of transportation and installation cost of hub facilities. Constraints (1) and (2) show that to route the flows from an origin to a destination one or more than one hub facility should be used. Constraints (3) and (4) state that there is an inter-hub arc if and only if both of connected nodes are hub facilities. Constraint (5) is a capacity constraint for flows which are routed via inter-hub arcs between two hub facilities. Constraint (6) is also a capacity constraint for flows which are routed via access arcs between non-hub nodes and hub nodes. Constraint (7) states the capacity constraint for flows which are routed via access arcs between hub nodes and non-hub nodes. Constraints (8) and (9) guarantee that for every access arc one of the starting or ending point should be a hub facility. It could be an access arc from an origin to a hub node or from a hub node to a destination. Constraint (10) are the well-known flow conservation constraints, and model the condition that the variables x, a, and s define the paths between origin and destination nodes. It ensures that the total number of arcs exiting every node is equal to the total number of arcs entering it. Constraints (11), (12), and (13) are the classical integrity and non-negativity constraints.

Figure 11 illustrates an example of modular hub location network with multiple assignments without direct connections.

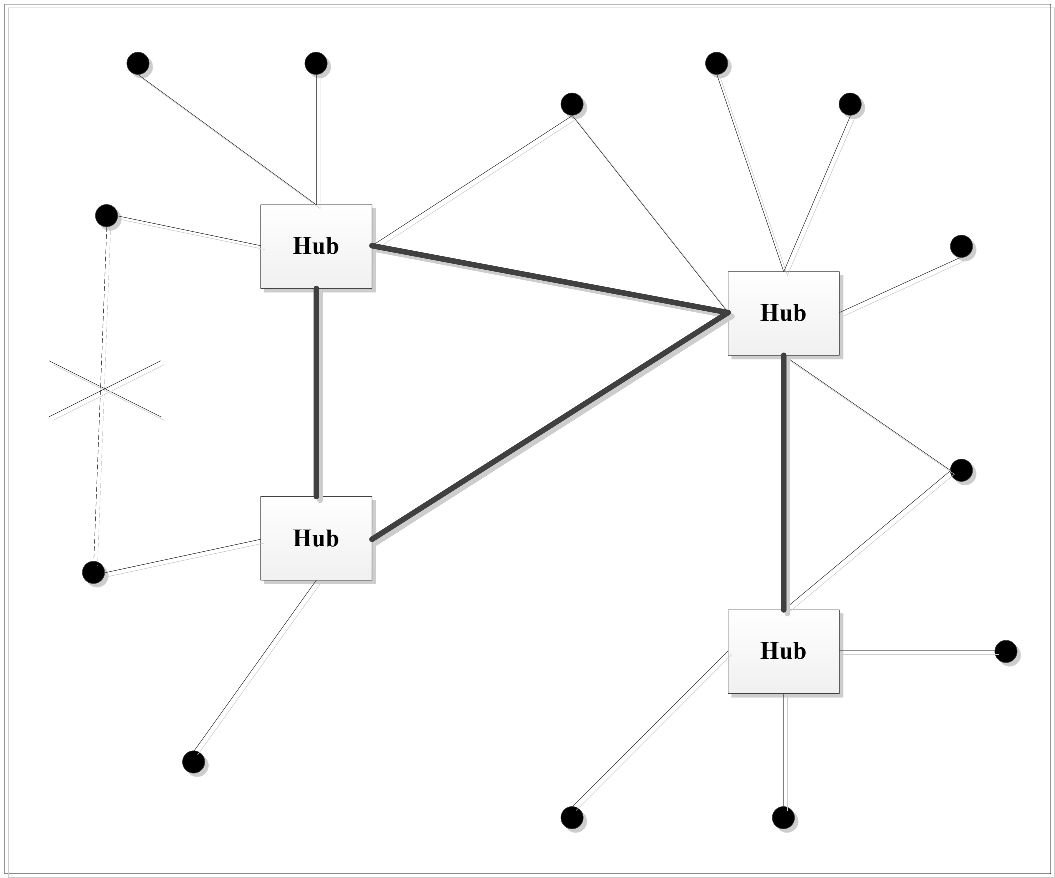

Figure 9: Network structure of MHLP-MA (no direct connections)

Chapter 4: Lagrangean relaxation

4.1) Lagrangean relaxation

Lagrangean Relaxation (LR) is a well-known method to solve large scale combinatorial optimization problems. It exploits the inherent structure of the problems to compute lower bounds on the value of the optimal solution. In the case of MHLP, if we relax constraints (1) - (4), and (8) - (10) in a Lagrangean fashion, weighting their violations with multiplier vectors $(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$ of appropriate dimension, we obtain the following Lagrangean function:

$$
\begin{aligned}
L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta) = \min \textstyle\sum_{k \in N} f_k z_k + \sum_{i \in N} \sum_{k \in N} q_{ik} V_{ik}^1 + \sum_{m \in N} \sum_{j \in N} q_{kj} V_{kj}^2 + {} \\
\textstyle\sum_{k \in N} \sum_{m \in N} c_{km} y_{km} + \sum_{i \in N} \sum_{j \in N} \alpha_{ij} \left(\sum_{k \in N} a_{ijk} - 1 \right) + {} \\
\textstyle\sum_{i \in N} \sum_{j \in N} \beta_{ij} \left(\sum_{k \in N} s_{ijk} - 1 \right) + {} \\
\textstyle\sum_{m \in N} \sum_{k \in N} \delta_{mk} \left(y_{mk} - Q_1 z_k \right) + {} \\
\textstyle\sum_{k \in N} \sum_{m \in N} \gamma_{km} \left(y_{km} - Q_1 z_k \right) + {} \\
\textstyle\sum_{i \in N} \sum_{k \in N} \varepsilon_{ik} \left(V_{ik}^1 - Q_2 z_k \right) + {} \\
\textstyle\sum_{k \in N} \sum_{j \in N} \zeta_{kj} \left(V_{kj}^2 - Q_3 z_k \right) + {} \\
\textstyle\sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \eta_{ijk} \left(a_{ijk} + \sum_{m \in N} x_{ijmk} - \sum_{m \in N} x_{ijkm} - s_{ijk} \right)
\end{aligned}
$$

subject to

$$\sum_{i \in N} \sum_{j \in N} W_{ij} x_{ijkm} \leq B y_{km} \qquad \forall\, k, m \in N; \quad i \neq j \quad (5)$$

$$\sum_{j \in N} W_{ij} a_{ijk} \leq H V_{ik}^1 \qquad \forall\, i, k \in N; \quad i \neq j \quad (6)$$

$$\sum_{i \in N} W_{ij} s_{ijk} \leq H V_{kj}^2 \qquad \forall\, k, j \in N; \quad i \neq j \quad (7)$$

$$z_k \in \{0,1\} \qquad \forall\, k \in N \quad (11)$$

$$V_{ik}^1, V_{kj}^2, y_{km} \in Z^+ \qquad \forall i,j,k,m \in N \quad (12)$$

$$0 \le a_{ijk}, X_{ijkm}, s_{ijk} \le 1 \qquad \forall i,j,k,m \in N; \;\; i \ne j \quad (13)$$

Observe that $L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$ is separable into four sub-problems: (1) a problem in the space of the (*x,y*) variables, (2) a problem in the space of the (a, V^1) variables, (3) a problem in the space of the (s, V^2) variables, and (4) a problem in the space of the (z) variables. After some algebra, the sub-problem in the space of the (*x,y*) can be expressed as:

$$L_{x,y}^{k,m}(\gamma, \delta, \eta) = min \sum_{k \in N} \sum_{m \in N} y_{km} \, (c_{km} + \gamma_{km} + \delta_{km}) + \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \sum_{m \in N} x_{ijkm} \, (\eta_{ijm} - \eta_{ijk})$$

subject to

$$\sum_{i \in N} \sum_{j \in N} W_{ij} \, x_{ijkm} \le B y_{km} \qquad \forall \, k,m \in N; \quad i \ne j$$

$$y_{km} \in z^+ \qquad \forall \, k,m \in N$$

$$0 \le X_{ijkm} \le 1 \qquad \forall i,j,k,m \in N; \, i \ne j$$

The sub-problem in the space of (a, V^1) is:

$$L_{a,V_{ik}^1}(\varepsilon, \alpha, \eta) = min \sum_{i \in N} \sum_{k \in N} V_{ik}^1 \, (q_{ik} + \varepsilon_{ik}) + \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} a_{ijk} \, (\alpha_{ij} + \eta_{ijk})$$

subject to

$$\sum_{j \in N} W_{ij} \, a_{ijk} \le H V_{ik}^1 \qquad \forall \, i,k \in N; \, i \ne j$$

$$V_{ik}^1 \in z^+ \qquad \forall\, i,k \in N$$

$$0 \leq a_{ijk} \leq 1 \qquad \forall i,j,k \in N;\; i \neq j$$

The sub-problem in the space of (s, V^2) is:

$$L_{s,V_{kj}^2}(\zeta,\beta,\eta) = min \sum_{k\in N}\sum_{j\in N} V_{kj}^2 \;(\, q_{kj} + \zeta_{kj}\,) + \sum_{i\in N}\sum_{j\in N}\sum_{k\in N} s_{ijk}\;(\,\beta_{ij} - \eta_{ijk}\,)$$

subject to

$$\sum_{i\in N} W_{ij}\, s_{ijk} \leq HV_{kj}^2 \qquad \forall k,j \in N;\; i \neq j$$

$$V_{kj}^2 \in z^+ \qquad \forall\, k,j \in N$$

$$0 \leq s_{ijk} \leq 1 \qquad \forall i,j,k \in N;\; i \neq j$$

The sub-problem in the space of *z* is:

$$L_z(\zeta,\beta,\eta,\varepsilon) = min \sum_{k\in N}(f_k + Q_1(\sum_{m\in N}(-\delta_{mk} - \gamma_{km})) +$$
$$Q_3 \sum_{j\in N} -\zeta_{kj} \;\; Q_2 \sum_{i\in N} -\varepsilon_{ik}))z_k$$

subject to

$$z_k \in \{0,1\} \qquad \forall\, k \in N$$

Note that each of the four sub-problems in which we decompose L$(\alpha,\beta,\gamma,\delta,\varepsilon,\zeta,\eta)$ captures one of the inherent structures of MHLP. The above analysis can be summarized in the following result:

Proposition 1.

$$\mathrm{L}(\alpha,\beta,\gamma,\delta,\varepsilon,\zeta,\eta) = L_{x,y}(\gamma,\delta,\eta) + L_{a,V_{ik}^1}(\varepsilon,\alpha,\eta) + L_{s,V_{kj}^2}(\zeta,\beta,\eta) + L_z(\zeta,\beta,\eta,\varepsilon) -$$
$$\sum_{i\in N}\sum_{j\in N}\alpha_{ij} - \sum_{i\in N}\sum_{j\in N}\beta_{ij}$$

Solution to Sub-problem $L_{x,y}(\gamma,\delta,\eta)$

Given that each of the y_{km} variables appear only in one constraint, we can further decompose the $L_{x,y}(\gamma,\delta,\eta)$ sub-problem in to independent sub-problems, one for each (k,m) pair, of the form:

$$L_{x,y}^{k,m}(\gamma,\delta,\eta) = \min \quad (c_{km} + \gamma_{km} + \delta_{km})\, y_{km} + \textstyle\sum_{i\in N}\sum_{j\in N}(\eta_{ijm} - \eta_{ijk})x_{ijkm}$$

subject to

$$\textstyle\sum_{i\in N}\sum_{j\in N} W_{ij}\, x_{ijkm} \le B y_{km}$$

$$y_{km} \in Z^+$$

$$0 \le x_{ijkm} \le 1 \qquad \forall i,j \in N;\; i \ne j$$

We can efficiently solve these problems by iteratively evaluating different values of the y_{km} variables and finding the optimal value for the remaining x_{ijkm} variables. That is, if we fix y_{km} to a particular value, the remaining problem reduces to a continuous knapsack problem, which can be optimally solved with the greedy knapsack algorithm (Lawler, 1979). This algorithm works by ordering the x_{ijkm} variables so that

$$\frac{(\eta_{(s)m} - \eta_{(s)k})}{W_{(s)}} \le \frac{(\eta_{(s+1)m} - \eta_{(s+1)k})}{W_{(s+1)}},$$

for $s = 1, \ldots, n^2\text{-}n$. In particular, $W_{(s)}$ denotes the weight of the s^{th} ordered pair of nodes (i,j). The greedy algorithm adds the ordered items (i.e., sets $x_{(s)km} = 1$) one at a time to the knapsack, starting from $x_{(1)km}$, and continues as long as adding an item does not exceed the capacity constraint By_{km}. The algorithm stops when residual capacity is equal to zero.

To find the optimal value of the y_{km}variable we start from $y_{km} = 1$ and evaluate the objective value by solving the correspondent continuous knapsack problem. We then increase the value of y_{km} by one, increasing the capacity of the knapsack and allowing more $x_{(s)km}$ variables to take a positive value. We keep increasing y_{km} until the capacity increases to a point that all $x_{(s)km}$ variables are set to one. Finally, we obtain the optimal solution of $L_{x,y}^{k,m}(\gamma, \delta, \eta)$ checking the value of y_{km} that provides the minimum objective function. The outline of the overall algorithm is depicted in Algorithm 1.

Algorithm 1. Solving sub-problem $L_{x,y}^{k,m}(\gamma, \delta, \eta)$

$best\_value = 0$

$y_{km} = 0$

do

  $value = 0$

  $y_{km} = y_{km} + 1$

  $s = 1$

  $capacity = By_{km}$

  while $(capacity \neq 0$ ***and*** $s < n^2 - n)$ **do**

    if $(capacity - W_{(s)} > 0)$ **then**

      $X_{(s)km} = 1$

      $capacity = capacity - W_{(s)}$

      $value = value + (\eta_{(s)m} - \eta_{(s)k})$

    else

      $X_{(s)km} = capacity / W_{(s)}$

      $capacity = 0$

      $value = value + (\eta_{(s)m} - \eta_{(s)k})X_{(s)km}$

    end-if

    $s = s + 1$

  end-while

  if $(value < best\_value)$ **then**

    $best\_value = value$

  end-if

while $(s < n^2 - n)$

Solution to Sub-problem $L_{a,V_{ik}^1}(\varepsilon, \alpha, \eta)$

Similar to the previous sub-problem, because the V_{ik}^1 variables appear only in one constraint, we can further decompose the $L_{a,V_{ik}^1}(\varepsilon, \alpha, \eta)$ sub-problem in to independent sub-problems, one for each (i, k) pair, of the form:

$$L_{a,V_{ik}^1}^{i,k}(\varepsilon, \alpha, \eta) = \ min \sum_{i \in N} \sum_{k \in N} V_{ik}^1 \ (\ q_{ik} + \ \varepsilon_{ik} \ ) + \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} a_{ijk} \ (\ \alpha_{ij} + \ \eta_{ijk} \ )$$

subject to

$$\sum_{j \in N} W_{ij} \, a_{ijk} \ \leq HV_{ik}^1$$

$$V_{ik}^1 \ \in \ z^+$$

$$0 \leq \ a_{ijk} \leq 1 \qquad \forall j \in N; \quad i \neq j$$

These problems can also be solved efficiently by iteratively evaluating different values of the V_{ik}^1 variables and finding the optimal value for the remaining a_{ijk}variables. Therefore, if V_{ik}^1 is fixed to a particular value, the remaining problem reduces to a continuous knapsack problem, which can be solved with the greedy knapsack algorithm optimally. This algorithm works by ordering the a_{ijk} variables so that

$$\frac{\left(\alpha_{i(s)} + \ \eta_{i(s)k} \ \right)}{W_{i(s)}} \leq \frac{\left(\alpha_{i(s+1)} + \ \eta_{i(s+1)k} \ \right)}{W_{i(s+1)}},$$

Similar to the previous sub-problem, for *s = 1, …,* $n - 1$. In particular, $W_{i(s)}$ denotes the weight of the s^{th} ordered pair of nodes (i, j). The greedy algorithm adds the ordered items (i.e., sets $a_{i(s)k} = 1$) one at a time to the knapsack, starting from $a_{i(1)k}$, and continues as long as adding an item does not exceed the capacity constraint HV_{ik}^{1}. The algorithm stops when residual capacity is equal to zero.

To find the optimal value of the y_{km}variable we start from $V_{ik}^{1} = 1$ and evaluate the objective value by solving the correspondent continuous knapsack problem. We then increase the value of V_{ik}^{1} by one, increasing the capacity of the knapsack and allowing more $a_{i(s)k}$variables to take a positive value. We keep increasing V_{ik}^{1} until the capacity increases to a point that all $a_{i(s)k}$variables are set to one. Finally, we obtain the optimal solution of $L_{a,V_{ik}^{1}}^{i,k}(\varepsilon, \alpha, \eta)$ checking the value of V_{ik}^{1} that provides the minimum objective function. The outline of the overall algorithm is depicted in Algorithm 2.

Algorithm 2. Solving sub-problem $L^{i,k}_{a,V^1_{ik}}(\varepsilon, \alpha, \eta)$

$best\_value = 0$

$V^1_{ik} = 0$

do

$value = 0$

$V^1_{ik} = V^1_{ik} + 1$

$s = 1$

$capacity = HV^1_{ik}$

while $(capacity \neq\ 0\ \boldsymbol{and}\ s < n - 1)$ **do**

if $(capacity - W_{i(s)} >\ 0)$ **then**

$a_{i(s)k} = 1$

$capacity := capacity - W_{i(s)}$

$value = value + (\alpha_{i(s)} +\ \eta_{i(s)k}\ )$

else

$a_{i(s)k} =\ capacity/\ W_{i(s)}$

$capacity = 0$

$value = value + (\alpha_{i(s)} +\ \eta_{i(s)k}\ )\ a_{i(s)k}$

end-if

$s = s + 1$

end-while

if $(value < best\_value)$ **then**

$best\_value = value$

end-if

while $(s < n - 1)$

Solution to Sub-problem $L_{s,V_{kj}^2}(\zeta,\beta,\eta)$

Similar to the previous sub-problems, because the V_{kj}^2 variables appear only in one constraint, we can further decompose the $L_{s,V_{kj}^2}(\zeta,\beta,\eta)$ sub-problem in to independent sub-problems, one for each (k,j) pair, of the form:

$$L_{S,V_{kj}^2}^{k,j}(\zeta,\beta,\eta) = \ min \sum_{k\in N}\sum_{j\in N} V_{kj}^2 \ (\ q_{kj} + \ \zeta_{kj}\ ) + \sum_{i\in N}\sum_{j\in N}\sum_{k\in N} s_{ijk}\ \ (\ \beta_{ij} - \ \eta_{ijk}\ )$$

subject to

$$\sum_{i\in N} W_{ij}\, s_{ijk} \ \leq HV_{kj}^2$$

$$V_{kj}^2 \in \ z^+$$

$$0 \leq s_{ijk} \leq 1 \qquad \forall i \in N;\ i \neq j$$

These problems can also be solved efficiently by iteratively evaluating different values of the V_{kj}^2 variables and finding the optimal value for the remaining s_{ijk} variables. That is, if V_{kj}^2 is fixed to a particular value, the remaining problem reduces to a continuous knapsack problem, which can be solved with the greedy knapsack algorithm optimally. This algorithm works by ordering the s_{ijk} variables so that

$$\frac{\left(\beta_{(s)j} - \ \eta_{(s)jk}\ \right)}{W_{(s)j}} \leq \frac{\left(\beta_{(s+1)j} - \ \eta_{(s+1)jk}\ \right)}{W_{(s+1)j}}$$

for *s* = *1,* ..., $n-1$. In particular, $W_{(s)j}$ denotes the weight of the s^{th} ordered pair of nodes (i,j). The greedy algorithm adds the ordered items (i.e., sets $s_{(s)jk} = 1$) one at a time to the knapsack, starting from $s_{(1)jk}$, and continues as long as adding an item does not exceed the capacity constraint HV_{kj}^2. The algorithm stops when residual capacity is equal to zero.

To find the optimal value of the V_{kj}^2variable we start from $V_{kj}^2 = 1$ and evaluate the objective value by solving the correspondent continuous knapsack problem. We then increase the value of V_{kj}^2 by one, increasing the capacity of the knapsack and allowing more $s_{(s)jk}$ variables to take a positive value. We keep increasing V_{kj}^2 until the capacity increases to a point that all $s_{(s)jk}$ variables are set to one. Finally, we obtain the optimal solution of $L_{S,V_{kj}^2}^{k,j}(\zeta, \beta, \eta)$ checking the value of V_{kj}^2 that provides the minimum objective function. The outline of the overall algorithm is depicted in Algorithm 3.

Algorithm 3. Solving sub-problem $L^{k,j}_{s,V^2_{kj}}(\zeta,\beta,\eta)$

$best\_value = 0$

$V^2_{kj} = 0$

do

    $value = 0$

    $V^2_{kj} = V^2_{kj} + 1$

    $s = 1$

    $capacity = HV^2_{kj}$

    while ($capacity \neq 0$ ***and*** $s < n-1$) **do**

        if ($capacity - W_{(s)j} > 0$) **then**

            $s_{(s)jk} = 1$

            $capacity = capacity - W_{(s)j}$

            $value = value + (\beta_{(s)j} - \eta_{(s)jk})$

        else

            $s_{(s)jk} = capacity / W_{(s)j}$

            $capacity = 0$

            $value = value + (\beta_{(s)j} - \eta_{(s)jk})s_{(s)jk}$

        end-if

        $s = s + 1$

    end-while

    if ($value < best\_value$) **then**

        $best\_value = value$

    end-if

while ($s < n-1$)

Solution to Sub-problem $L_z(\zeta, \beta, \eta, \varepsilon)$

The last sub-problem (4) is a simple optimization problem that can be efficiently solved by inspection. Given that the integrality conditions on the *z* variables are the only constraints in this sub-problem, we can obtain the optimal solution value as

$$L_z(\zeta, \beta, \eta, \varepsilon) = \sum_{k \in N} \min \left\{ 0, \left(f_k - \sum_{m \in N} Q1(\delta_{mk} + \gamma_{km}) - \sum_{j \in N} Q3\zeta_{kj} - \sum_{i \in N} Q2\varepsilon_{ik} \right) \right\},$$

by setting the variable $z_k = 1$ if its coefficient of the objective function is negative and $z_k = 0$ otherwise.

The solution of the Lagrangean Dual

In order to obtain the best lower bound one must solve the Lagrangean dual of MHLP, which is given by

$$z_D = \max_{\gamma, \delta, \varepsilon, \zeta \geq 0} L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$$

We apply the subgradient optimization method to solve this problem. For a given vector of dual multipliers $(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$, let $x_{ijkm}(\gamma, \delta, \eta)$, $y_{km}(\gamma, \delta, \eta)$, $V_{ik}^1(\varepsilon, \alpha, \eta)$, $V_{kj}^2(\zeta, \beta, \eta)$, $a_{ijk}(\varepsilon, \alpha, \eta)$, $s_{ijk}(\zeta, \beta, \eta)$ and $z_k(\zeta, \beta, \eta, \varepsilon)$ denote the optimal solution to $L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$. Therefore, the subgradient of $L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$ is

$$s(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta) =$$

$$\Big(\big(\textstyle\sum_{k \in N} a_{ijk}(\varepsilon, \alpha, \eta) - 1 \big)_{ijk}, \big(\textstyle\sum_{k \in N} s_{ijk}(\zeta, \beta, \eta) - 1 \big)_{ijk}, \big(y_{km}(\gamma, \delta, \eta) -$$

$$Q_1 \, z_k(\zeta, \beta, \eta, \varepsilon) \big)_{km}, \big(y_{mk}(\gamma, \delta, \eta) - Q_1 \, z_k(\zeta, \beta, \eta, \varepsilon) \big)_{mk}, \big(V_{ik}^1(\varepsilon, \alpha, \eta) -$$

$Q_2 z_k(\zeta, \beta, \eta, \varepsilon) \big)_{ik}, \big(V_{kj}^2(\zeta, \beta, \eta) - Q_3\, z_k(\zeta, \beta, \eta, \varepsilon) \big)_{kj}, \big(a_{ijk}(\varepsilon, \alpha, \eta) +$

$\sum_{m \in N} x_{ijmk}\,(\gamma, \delta, \eta) - \sum_{m \in N} x_{ijkm}\,(\gamma, \delta, \eta) - s_{ijk}(\zeta, \beta, \eta) \big)_{ijk} \big)$

The outline of the subgradient algorithm is depicted in Algorithm 4. The output of the algorithm is a lower bound z_D and UB denotes a known upper bound on the optimal value of the original problem.

Algorithm 4: Subgradient Optimization Method

Iteration 0

$LB = -\infty$;

$\alpha^0 = \beta^0 = \gamma^0 = \delta^0 = \varepsilon^0 = \zeta^0 = \eta^0 = 0; \lambda^k = 2.$

Let *UB* be a known upper bound on the optimal value.

Iteration *k*

Solve the lagrangean function $L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$.

if $(L(\alpha^k, \beta^k, \gamma^k, \delta^k, \varepsilon^k, \zeta^k, \eta^k) > LB)$ **then**

$LB = L(\alpha^k, \beta^k, \gamma^k, \delta^k, \varepsilon^k, \zeta^k, \eta^k)$

end-if

Evaluate the subgradient $\gamma(\alpha^k, \beta^k, \gamma^k, \delta^k, \varepsilon^k, \zeta^k, \eta^k)$.

Compute the step size $t_k = \frac{\lambda^k(UB - L(\alpha,\beta,\gamma,\delta,\varepsilon,\zeta,\eta))}{\|\gamma(\alpha^k,\beta^k,\gamma^k,\delta^k,\varepsilon^k,\zeta^k,\eta^k)\|^2}$.

$(\alpha^{k+1}, \beta^{k+1}, \gamma^{k+1}, \delta^{k+1}, \varepsilon^{k+1}, \zeta^{k+1}, \eta^{k+1}) = (\alpha^k, \beta^k, \gamma^k, \delta^k, \varepsilon^k, \zeta^k, \eta^k) + t_k\, \gamma(\alpha^k, \beta^k, \gamma^k, \delta^k, \varepsilon^k, \zeta^k, \eta^k)$.

k = k+1

In the subgradient algorithm the factor λ^k is cut off after 35 consecutive iterations without improving the lower bound and it is reset to the value 2 every 300 iterations.

4.2) Primal Heuristic

We can exploit the primal information obtained from the Lagrangean function to construct feasible solutions and thus, upper bounds on the optimal solution value of the problem. We next present a primal heuristic which is applied at some iterations of the subgradient optimization that exploits such information.

Every solution network that has enough capacity on the links to route all demands from their origins to their destinations is a feasible solution network. Every feasible solution provides an upper bound on the optimal solution value of the MHLP, but we cannot guarantee the optimality of the solution.

At iteration k of the subgradient optimization method, the Lagrangean solution may not be feasible for the original problem MHLP, because some constraints were relaxed. Let $S^k = \{ i: z_i(\ldots) = 1, i \in N\}$ be the current set of open hub facilities associated with the Lagrangean solution at iteration k, which is obtained by solving $L(\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta)$. When the location of the hub facilities is known, we note that the routing problem for the commodities is still a challenging NP-hard optimization problem, given that it can be transformed to a network loading problem (see, Magnanti et al, 1995). The main difficulty arises from the fact that we cannot know in advance how many access arcs and hub arcs are going to be used in the path for each commodity, as the links of the network have to be selected first, together with their capacities, to find such paths.

A simple way to construct a feasible solution is to ensure that there exist at least one and at most two hub nodes in each route between O/D pairs. To do so, we first open the facilities contained in set S^k and assign the remaining non-hub nodes $j \notin S^k$ to their closest open hub $i \in S^k$. Temporarily assuming full interconnection between hub nodes, the path between each pair of nodes is then computed as the shortest path on the current hub network. Next, in order to obtain the number facility links on every access arc and hub arc, the amount of flow W_{ij}, which is routed on each link of the network, must be calculated to determine the minimum number of facility links that are needed to route such flow. The outline of the primal heuristic algorithm is depicted in Algorithm 5.

Algorithm 5: A primal heuristic algorithm

$UB := 0$

for all $(i \in S^k)$ **do**

$\quad UB := UB + f_i$

end-for

for all $(i \in N \setminus S^k)$ **do**

$\quad k(i) := \arg\min\{d_{ij} : j \in S^k\}$

$\quad N_{k(i)} := N_{k(i)} \cup \{i\}$

$\quad flow_{ik(i)} := \sum_{j \in N} W_{ij}$

$\quad flow_{k(i)i} := \sum_{i \in N} W_{ij}$

$\quad UB := UB + q_{ik(i)} \times ceil\left(\frac{flow_{ik(i)}}{H}\right) + q_{k(i)i} \times ceil\left(\frac{flow_{k(i)i}}{H}\right)$

end-for

for all $((k, m) \in S^k \times S^k)$ **do**

$\quad flow_{km} := \sum_{i \in N_k} \sum_{j \in N_m} W_{ij}$

$\quad UB := UB + c_{km} \times ceil\left(\frac{flow_{km}}{B}\right)$

end-for

Chapter 5: Experimental Results and Analysis

We present the results of computational experiments performed to evaluate and analyze the proposed Lagrangean relaxation method. To assess the quality of the lower bounds obtained from the LR, we compare our results with the LP bounds of the MIP formulation. All algorithms were coded in C++ using the Visual Studio 2010 platform, and in order to solve the LPs, the MIP formulation was modeled with OPL and solved using CPLEX 12.2© Optimization Studio. The experiments have been run on a HP PC with 4.00 GB of RAM memory with a processor Dual-Core CPU 2.8 GHz, and under Windows 7 environment (64-bit Operating System).

The computational experiments were performed using the well-known Australian Post (AP) set of instances. This data set is the most commonly used in the hub location literature and can be downloaded from mscmga.ms.ic.ac.uk/jeb/orlib/phubinfo.html. It consists of the Euclidean distances d_{ij} between 200 cities in Australia and W_{ij} represents the postal flows between every pair of nodes. In this study, we have considered a set of instances containing small to medium size instances with up to 50 nodes. This set contains 20 instances, having four instances for each size N = 10, 20, 25, 40, and 50. For every problem size the four instances correspond to different combinations of characteristics for the fixed installation costs and the capacities for the hubs (parameters B, H, b, and p). The configuration of the parameters has been selected based in such a way that equivalent discount factors for the inter-hub arcs of α = {0.2, 0.6} are obtained when using fully loaded trucks.

5.1) A comparison of bounds

Our preliminary computational results focus on the comparison of lower bounds obtained between a Lagrangean relaxation and an LP relaxation solved with CPLEX. As stated in Chapter 4, the subgradient optimization algorithm (SOA) is used to solve LR. This algorithm terminates when one of the following four criteria is met:

i) All the components of the subgradient are zero. In this case the current solution is proven to be optimal.

ii) The difference between the upper bound and the lower bound is less than a threshold value.

$$|Upper\ bound - Lower\ bound| \leq \varepsilon$$

iii) The improvement of the lower bound after k consecutive iterations is less than a threshold value χ.

iv) The maximum number of iterations k_{max} is reached.

After some tuning, we set the following parameter values:

$$\varepsilon = 10^{-6}, \chi = 0.05\%, k = 1000\ , k_{max} = 10000.$$

The results of the comparison between the lower bounds obtained with the LR method and the LP relaxation of the MIP formulation are given in Table 1. The first two columns give the number of nodes and the discount factor of each instance, respectively. As mentioned in Chapter 3, *B* represents the capacity of the large facility links between the hub nodes and *H* represents the capacity of the small facility links to transport flows from non-hub nodes to hub nodes or from hub nodes to non-hub nodes. b and p represent the variable cost per unit traveled distance for the large and small facility links, respectively.

The next two columns under the heading LR depict: *i)* the best lower bound obtained with LR, and *ii)* the CPU Time in seconds needed for LR to terminate.

The next columns under the heading CPLEX provide: *i)* the linear programming relaxation bounds (LP), and *ii)* the CPU time in seconds needed for CPLEX to solve the LP relaxation.

The last column illustrates the difference between the percent deviations of LB_{LP} and LB_{LR}. The percent deviation of the LP lower bound and the best known upper bound, i.e.

$$Dev_{LP} = \frac{UB - LB_{LP}}{UB} * 100,$$

The percent deviation of the best lower bound of LR and the best known upper bound, i.e.

$$Dev_{LR} = \frac{UB - LB_{LR}}{UB} * 100$$

Finally, the difference between the percent deviations of LB_{LP} and LB_{LR} is:

$$Dev = Dev_{LP} - Dev_{LR}$$

It represents the efficiency of the presented solution method to improve the LP bounds. For the 10 node instances, the best known upper bound corresponds to the optimal solution obtained with CPLEX. However, given that CPLEX is not able to solve larger size instances with 20 nodes or more to optimality in one day of CPU time, we use the best upper bound found by the primal heuristic. Whenever CPLEX is not able to solve the LP relaxation because of time limit, we write n.a. in the corresponding entry of the table.

Table 1: Comparison of lower bounds between LR and LP relaxation

	Instances					LR		CPLEX		Dev_{LP} - Dev_{LR} (%)
N	α	B	H	b	p	LB	CPU Time(s)	LB	CPU Time(s)	
10	0.2	750	100	300	200	166431.26	3.77	161726.04	5.50	2.49
	0.2	750	100	600	400	264791.04	3.81	243229.59	6.03	7.15
	0.6	200	100	500	400	283192.15	4.40	275184.31	6.37	2.38
	0.6	300	150	500	400	213496.64	3.98	200413.68	5.89	5.03
20	0.2	750	100	300	200	178667.04	43.73	174491.99	66.22	1.62
	0.2	750	100	600	400	300098.25	45.07	278175.86	87.86	4.47
	0.6	200	100	500	400	328213.23	43.13	301969.26	52.93	5.36
	0.6	300	150	500	400	262713.69	40.14	215667.18	47.76	12.13
25	0.2	750	100	300	200	190812.47	215.38	177085.49	271.53	4.77
	0.2	750	100	600	400	320358.29	222.28	280092.51	290.03	7.30
	0.6	200	100	500	400	351137.15	220.82	304200.88	268.49	8.51
	0.6	300	150	500	400	307158.22	209.25	218209.82	243.19	21.65
40	0.2	750	100	300	200	225652.64	4137.19	177666.79	6235.76	14.85
	0.2	750	100	600	400	399941.06	4567.75	287264.24	7201.81	18.06
	0.6	200	100	500	400	451894.55	4678.85	310072.86	5350.21	22.73
	0.6	300	150	500	400	414172.35	4411.39	217219.54	3638.76	36.80
50	0.2	750	100	300	200	273228.09	23781.08	n.a.	n.a.	n.a.
	0.2	750	100	600	400	461738.30	24197.63	n.a.	n.a.	n.a.
	0.6	200	100	500	400	518675.90	23964.13	n.a.	n.a.	n.a.
	0.6	300	150	500	400	480950.06	23135.30	n.a.	n.a.	n.a

As it can be seen in Table1, the proposed LR is able to consistently obtain better lower bounds that the ones obtained with the LP relaxation of the MIP formulation. This is one of the positive results that were expected, as the proposed LR dos not have the integrality property. Moreover, the percent improvement becomes larger as the size of the instances increase. In the smaller instances the improvement is between 2% to 6% whereas in the larger instances the improvement is between 20% to 37%.

In addition, the results of this table indicate that the LR requires less CPU time than CPLEX to obtain these improved bounds in all but one of the considered instances. As the number of nodes increases, the CPU time also increase for both the LR and the LP relaxation. LR is able to converge in less than one minute for the 20 node instances, and 3.5 minutes for the 25 node instances. The running time of the LR for 40 node instances is 1.5 hours, but it takes more than 2 hours to obtain LP bounds with CPLEX. For instances with more than 40 nodes, CPLEX ran out of memory after a few hours. Therefore, it was not possible to obtain LP bounds for more than 40 node instances with 4 GB of RAM memory.

5.2) Analysis of LR

The goal of the computational experiments presented next is to analyze the capabilities and limitations of the proposed Lagrangean relaxation for obtaining approximate solutions of the MHLP. The results are given in Tables 2 and 3. For a specific configuration of *B, H, b*, and *p* the best lower bound and upper bound obtained in the LR and CPU time needed to obtain them are given. Finally, the gap between the best lower bound and the best upper bound is given, i.e.

$$Gap = \frac{UB - LB_{LR}}{UB} * 100$$

The optimal solution values of 10 node instances are given to evaluate the upper bounds obtained with the primal heuristic in Table 2. The last column illustrates the percent deviation between the best upper bound with the heuristic and the optimal solution value for 10 node instances, i.e.

$$Dev_{UB-OPT} = \frac{UB - OPT}{UB} * 100$$

Table 2: Analysis of LR for 10 node instances

	Instances					**LR**				**Optimal solution**	**% Deviation**
N	**α**	B	H	b	P	LB	UB	CPU Time(s)	Gap (%)		
10	0.2	750	100	300	200	166431.26	219705.01	3.77	24.17	188659.88	14.13
	0.2	750	100	600	400	264791.04	405243.67	3.81	34.65	301382.97	25.62
	0.6	200	100	500	400	283192.15	405243.67	4.40	30.11	335792.88	17.13
	0.6	300	150	500	400	213496.64	296201.80	3.98	27.92	260634.68	12.00

CPLEX is not able to obtain the optimal solutions for instances with more than 10 nodes. Therefore, it is not possible to further evaluate how far our upper bounds are from the optimal values. The results for analysis of LR for 20, 25, 40, and 50 node instances are given in Table 3.

Table 3: Analysis of LR for 20, 25, 40, and 50 node instances

	Instances					**LR**			
N	**α**	B	H	b	p	LB	UB	CPU Time(s)	Gap (%)
20	0.2	750	100	300	200	178667.04	257614.96	43.73	30.64
	0.2	750	100	600	400	300098.25	489814.37	45.07	38.73
	0.6	200	100	500	400	328213.23	489814.37	43.13	32.99
	0.6	300	150	500	400	262713.69	381533.07	40.14	31.14
25	0.2	750	100	300	200	190812.47	287763.88	215.38	33.69
	0.2	750	100	600	400	320358.29	551716.62	222.28	41.93
	0.6	200	100	500	400	351137.15	551716.62	220.82	36.35
	0.6	300	150	500	400	307158.22	410845.93	209.25	25.23
40	0.2	750	100	300	200	225652.64	322976.46	4137.12	30.133
	0.2	750	100	600	400	399941.06	623888.32	4567.75	35.89
	0.6	200	100	500	400	451894.55	623888.32	4678.85	27.56
	0.6	300	150	500	400	414172.35	535292.71	4411.39	22.62
50	0.2	750	100	300	200	273228.09	356280.47	23781.08	23.31
	0.2	750	100	600	400	461738.30	688119.44	24197.63	32.89
	0.6	200	100	500	400	518675.90	688119.44	23964.13	24.62
	0.6	300	150	500	400	480950.06	647100.72	23135.30	25.67

It can be seen in Tables 2 and 3, the primal heuristic is able to obtain feasible solutions for all considered instances. The obtained lower and upper bounds with LR are able to provide a percent optimality gap that ranges between 23% to 42%. For the 10 node instances that we know the optimal solution value, we can observe that the solutions obtained with our primal heuristic are not very close the optimal solution values.

Even though this results are not very good, this is already a larger improvement with respect to what a general purpose solve, such as CPLEX, can do for this problem. For instance, the obtained lower bounds could be used in a branch and bound method to optimally solve the problem. Given that the bound are always better than the LP bound and the times are smaller, we expect to obtain the optimal solution faster than CPLEX. However, a specialized implementation of this method is required in order to be competitive.

Chapter 6: Conclusion

This thesis studied a challenging class of hub location problems known as Modular Hub Location Problems (MHLP). These problems are able to create more realistic networks than other classical hub location problems. By relaxing common assumptions frequently considered in classical hub location problems, MHLPs are able to overcome several modeling weaknesses and creating more realistic networks. Hub facilities are not assumed to be fully interconnected anymore, and transportation costs are considered to be flow dependent. These costs are modeled by using step-wise (modular) functions on every link of the network. Moreover, distance between every pair of (O/D) nodes is not assumed to be symmetric or to satisfy the triangular inequality. Creating more realistic models makes MHLP much more difficult to solve, as compared with classical hub location problems.

The main contribution of this thesis was to propose an approximate solution, based on Lagrangean relaxation (LR), to obtain lower and upper bounds on the optimal solution value of the MHLP. To construct feasible solutions, a simple primal heuristic was also proposed in this study.

Based on the computational experiments, the following results can be concluded:

- Given that the proposed LR does not have the integrality property, it was capable of obtaining better lower bounds than the linear programming relaxation of the MIP formulation for all the considered instances.
- Instances containing 20, 25, 40, and 50 nodes were approximately solved by the presented solution method. In contrast, a general purpose solver such as CPLEX was not able to solve the same problems in one day of CPU time.

- The CPU times of the LR to obtain better lower bounds were much less than the CPU times required by CPLEX to solve just the LP relaxations of the MIP formulation.

A possible future research direction could be the integration of the proposed LR into a branch and bound framework to obtain optimal solutions for the MHLP. In addition, the development of sophisticated heuristic algorithms is highly relevant to obtain high quality solutions to this challenging optimization problem. Other research directions could be the incorporation of capacity constraints at the hub facilities or service level constraints to limit the structure of O/D paths to provide even more realistic hub location models.

References

1. Abdinnour-Helm, S. (1998). A hybrid heuristic for the uncapacitated hub location problem. *European Journal of Operational Research* 106(2-3) 489–499.

2. Abdinnour-Helm, S., Venkataramanan, M.A. (1998). Solution approaches to hub location problems. *Annals of Operations Research* 78 31–50.

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

4. Aykin, T. (1995). Network policies for hub-and-spoke systems with application to the air transportation system. *Transportation Science* 29(3) 201-221.

5. 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.

6. Bryan, D. (1998). Extension to the hub location problem: formulations and numerical examples. *Geographical Analysis* 30(4) 315-330.

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

8. Campbell, J.F. (1994a). A survey of network hub location. *Studies in Locational Analysis* 6 31–49.

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

10. Campbell, J.F. (1998). Hub location and network design. Unpublished.

11. Campbell, J.F. (2005). Strategic network design for motor carriers. In: *A. Langevin & D. Riopel, eds. Logistic systems: design and optimization*. New York: Springer 245-278.

12. Campbell, J.F., Ernst, A.T., Krishnamoorthy, M. (2002). Hub location problems. In: Drezner, Z., Hamacher, H. (Eds.), *Facility Location: Applications and Theory*. Springer-Verlag, Berlin.

13. Campbell, J.F., Ernst, A.T., Krishnamoorthy, M. (2005a). Hub arc location problems: part I--introduction and results. *Management Science* 51(10) 1540–1555.

14. Campbell, J.F., Ernst, A.T., Krishnamoorthy, M. (2005b). Hub arc location problems: part II--formulations and optimal algorithms. *Management Science* 51(10) 1556–1571.

15. Chen, J.F. (2007). A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega* 35 211–220.

16. Chung, S.H., Myung, Y.S., Tcha, D.W. (1992). Optimal design of a distributed network with a two-level hierarchical structure. *European Journal of Operational Research* 62 105-115.

17. Cornuejols, G., Nemhauser, G.L, and Wolsey, L.A. (1990). The uncapacitated facility location problem, in *Discrete Location Theory*, P. Mirchandani and R. Francis, eds., John Wiley and Sons, New York 119–171.

18. Cunha, C.B., Silva, M.R. (2007). A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil. *European Journal of Operational Research* 179 747–758.

19. Daskin, M.S. (1995). *Network and discrete location: models, algorithms, and applications*. Wiley, New York.

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

21. 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.

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

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

24. Ernst, A.T., Krishnamoorthy, M. (1998b). An exact solution approach based on shortest-paths for *p*-hub median problems. *Informs Journal on Computing* 10(2) 149–162.

25. Ernst, A.T., Krishnamoorthy, M. (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86 141–159.

26. Feo, T.A., Resende, M.G.C. (1995). Greedy randomized adaptive search procedures. *Journal of Global Optimization* 2 1–27.

27. Fisher, M.L. (1981). The lagrangean relaxation method for solving integer programming problems. *Management Science* 27 1–18.

28. Gavish, B. (1991). Topological design of telecommunication networks – local access design methods. *Annals of Operation Research* 33 17-71.

29. Gelareh, S. (2008). Hub location models in public transport planning. Ph.D. thesis. Technical University of Kaiserslautern, Germany.

30. Gelareh, S., Nickel, S. (2007). A benders decomposition algorithm for single allocation hub location problem. In: *Proceeding of* GOR.

31. Gelareh, S., Nickel, S. (2008). New approaches to hub location problems in public transport planning. Tech. Rep. Department of Optimization, Fraunhofer Institute for Industrial Mathematics (ITWM), D 67663 Kaiserslautern, Germany.

32. Geoffrion, A.M. (1974). Lagrangean relaxation for integer programming. *Mathematical Programming Study* 2 82-114.

33. Glover, F. (1989). Tabu search: part I. *ORSA Journal of Computing* 1 190–206.

34. Glover, F. (1990). Tabu search: part II. *ORSA Journal of Computing* 2 4–32.

35. Goldman, A.J. (1969). Optimal location for centers in a network. *Transportation Science* 3 352-360.

36. Guignard, M. (2003). Lagrangean relaxation. *TOP* 11 151–228

37. Held, M., Karp, R.M. (1970). The travelling salesman problem and minimum spanning trees. *Operations Research* 18 1138-1162.

38. Held, M., Karp, R.M. (1971). The travelling salesman problem and minimum spanning trees: part II. *Mathematical Programming* 1 6-25.

39. Holland, J.H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

40. Hunt, G.W. (1998). Transport relay network design. PHD thesis, Georgia Institute of Technology.

41. Jaillet, P., Song, G., Yu, G. (1996). Airline network design and hub location problems. *Location Science* 4(3) 195-212.

42. Kara, B.Y. (1999). Modeling and analysis of issues in hub location problems. Ph.D. Thesis, Bilkent University Industrial Engineering Department, 06800 Bilkent, Ankara, Turkey.

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

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. Kimms, A. (2006). Economies of scale in hub and spoke network design: we have it all wrong, in: M. Morlock, C.Schwindt, N.Trautmann, J.Zimmermann (eds.). *Perspectives on operations Research. Weisbaden, Germany* 293-317.

46. Kirpatrick, S., Gelatt, C.D., Vecchi, M.P. (1993). Optimization by simulated annealing. *Science* 220 671–680.

47. Klincewicz. (1998). Hub location in backbone/tributary network design: A review. *Location Science* 6(1-4) 307-335.

48. Laguna, M., Marti, R. (2003). *Scatter Search*. Springer.

49. Lawler, E.L. (1979). Fast approximation algorithms for knapsack problems. *Mathematics of Operations Research* 4(4) 339–356.

50. Lemaréchal, C. (1974). An algorithm for minimizing convex functions. *Proceedings IFIP'74 Congress* (North Holland, Amsterdam) 552-556.

51. Magnanti, L., Mirchandani, P., Vachani, R. (1995). Modeling and solving the capacitated network loading problem. *Operations Research* 43 142-157.

52. Marianov, V., Serra, D. (2000). Location models for airline hubs behaving as m/d/c queues. *Computers & Operations Research* 30(7) 983–1003.

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

54. Meyer, T., Ernst, A.T., Krishnamoorthy, M. (2009). A 2-phase algorithm for solving the single allocation *p*-hub center problem. *Computers and Operations Research* 36(12) 3143–3151.

55. Michalewicz, Z. (1992). *Genetic algorithms + data structures = evolution programs.* Springer.

56. Mirzaghafour, F. (2013). The modular hub location problem. Master Thesis, Concordia University, Canada.

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

58. O'Kelly, M.E. (1998). A geographer's analysis of hub-and-spoke networks. *Journal of Transport Geography 6*(3) 171–186.

59. O'Kelly, M.E., Bryan, D. (1998). Hub location with flow economies of scale. *Transportation Research Part B: Methodological* 32(8) 605-616.

60. O'Kelly, M.E., Bryan, D. (1999). Hub-and-spoke networks in air transportation: an analytical review. *Regional Science* 39(2) 275-295.

61. O'Kelly, M.E., Miller. (1994). the hub network design problem. *Journal of Transport Geography 2*(1) 31-40.

62. Pirkul, H., Schilling, D.A. (1998). An efficient procedure for designing single allocation of hub and spoke systems. *Management Science* 44(12) 235-242.

63. Powell,W.B. (1986). A local improvement heuristic for the design of less-than-truckload motor carrier networks. *Transportation Scienc,* 20(4) 246-257.

64. Powell,W.B., Sheffi,Y. (1983). The load-planning problem of motor carriers: problem description and proposed solution approach. *Transportation Research Part A: General* 17(6) 471-480.

65. Skorin-Kapov, D., Skorin-Kapov, J. (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research* 73 502–509.

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

67. Sung,C.S., Jin, H.W. (2001). Dual-based approach for a hub network design problem under non-restrictive policy. *European Journal of Operational Research* 132(1) 88-105.

68. Taaffe, E.J., Gauthier, H.L., O'Kelly, M.E. (1996). *Geography of transportation*, Second Edition, Upper Saddle River, NJ: Prentice Hall.

69. Taha, T.T., Taylor, G.D. (1994). An integrated modeling framework for evaluating hub-and-spoke networks in truckload trucking. *Logistics and Transportation Review* 30(2) 141-166.

70. Taha, T.T., Taylor, G.D., Taha, H.A. (1996). A simulation-based software system for evaluation of hub-and-spoke transportation networks. *Simulation Practice & Theory* 3(6) 327-346.

71. Taylor, G.D., Harit, S., English, J.R., Whicker, G. (1995). Hub and spoke networks in truckload trucking: configuration, testing and operational concerns. *Logistics & Transportation Review* 31(3) 209-238.

72. Taylor, G.D., Meinert,T.S., Killian, R.C., Whicker, G.L. (1999). Development and analysis of alternative dispatching methods in truckload trucking. *Transportation Research Part E: Logistics and Transportation Review* 35(3) 191-205.

73. Topcuoglu, H., Corut, F., Ermis, M., Yilmaz, G. (2005). Solving the uncapacitated hub location problem using genetic algorithms. *Computers and Operations Research*, 32(4) 967–984.

74. Wagner, B. (2008). Model formulations for hub covering problems. *Journal of the Operational Research Society* 59(7) 932–938.

75. Wolfe, P., Crowder, H.D. (1974). Validation of subgradient optimization. *Mathematical Programming* 1 62-88.

76. Yoon, M.-G., Back, Y.-H., Tcha, D.-W. (1998a). On a large-scale distributed communication network design problem with fixed-charge hub facilities. Working paper.

77. Yoon, M.-G., Back, Y.-H., Tcha, D.-W. (1998b). Design of a distributed fiber transport network with hubbing topology. *European Journal of Operational Research* 104 510-520.

78. Yoon, M.-G., Tcha, D.-W. (1996). Topological design of a two-level centralized network with service survivability. Working paper.