Chapter 13
Hierarchical Facility Location Problems

Ivan Contreras and Camilo Ortiz-Astorquiza

Abstract *Hierarchical facility location problems* (HFLPs) are an important class of problems arising in numerous contexts such as in the design of health care, telecommunications and transportation systems. HFLPs deal with the location of interacting facilities at different levels of a hierarchical system. This chapter describes the distinguishing features and main areas of application of HFLPs and provides a comprehensive classification scheme based on several attributes. It also presents a concise overview of four classes of HFLPs that have received the most attention in the literature: multi-level facility location, median and covering hierarchical facility location, multi-echelon location-routing problems, and hierarchical hub location problems. For these classes of HFLPs, we highlight their main characteristics and point out to some of the integer programming formulations and efficient algorithms that have been developed.

13.1 Introduction

Health care, telecommunications, and transportation systems are examples where hierarchical structures arise having different types of interacting facilities that collectively provide services or products to a set of customers. There are three key features of such hierarchical systems which are crucial for their design. The first one is that the different types of facilities in the system are characterized by the services they provide. The second one is that there exists either an inherent hierarchy given by the nature of the system or a ranking mechanism that allows the complete ordering of the different types of facilities into levels, each of which contains only

I. Contreras (✉)
Concordia University and Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, QC, Canada
e-mail: icontrer@encs.concordia.ca

C. Ortiz-Astorquiza
Departamento de Matemáticas, Pontificia Universidad Javeriana, Bogotá, Colombia
e-mail: camiloortiz@javeriana.edu.co

G. Laporte et al. (eds.), *Location Science*,
https://doi.org/10.1007/978-3-030-32177-2_13

facilities of the same type. The third feature is that there is an implicit or explicit relationship between facilities among them.

This chapter focuses on a general class of discrete location problems arising in the design of hierarchical systems. Given the wide variety of applications in which a hierarchical structure is present, several variants of this class of problems have been studied, usually under different names. Depending on the application, as well as authors' discipline and fields of expertise, the terms *hierarchical*, *multi-level*, *multi-echelon*, *multi-stage*, *multi-tier*, and *multi-layer* have been used to refer to different facility location problems with an underlying hierarchical structure. Therefore, the main goal of this chapter is to present a unified view of such problems that fit under one umbrella: *hierarchical facility location*. Broadly speaking, given a set of customers that demand one or multiple services and a set of facilities of k different types associated with the various services, *hierarchical facility location problems* (HFLPs) consist of selecting a set of facilities to open so that each customer receives the requested service(s) from facilities of one or multiple types, while optimizing an objective function.

HFLPs are closely related to classical facility location problems such as p-median, uncapacitated facility location, p-center, and covering problems. In these problems, there is the implicit assumption that all customers request one and the same type of service, which is offered by any candidate facility. In contrast, HFLPs extend these problems to deal with more realistic situations in which facilities are arranged in a hierarchical structure determined by the bundle of services that each of them provides or by other natural order inherited from the considered system. In this sense, classical facility location problems are single-level variants of the more general class of HFLPs.

Applications of HFLPs are abundant. These include supply chains and production-distribution systems, where manufacturing facilities, warehouses, distribution centers, and retail stores interact to provide cost-effective production, storage and transportation services to customers. Health care systems are another common example, where patients seeking health related services travel to different facilities such as local clinics, community and regional hospitals, each of them providing a variety of services. Other examples of applications of HFLPs arise in solid waste management, education systems, emergency medical services, air freight and passenger travel, postal delivery, telecommunication networks, and urban transportation planning.

The study of hierarchical systems in location science has its origins in the area of health care planning with the works by Schultz (1970) and Dökmeci (1973, 1977), for continuous location models, and those by Calvo and Marks (1973), Schilling et al. (1979), and Moore and ReVelle (1982), for discrete models. Hierarchical discrete location models were initially studied in the context of production-distribution systems by Kaufam et al. (1977), where the so-called plant and warehouse location problem was introduced. Narula (1984) provided the first classification scheme for HFLPs based on the relationship between the different types of facilities and the flow pattern. Narula (1986) and Church and Eaton (1987) are early reviews on HFLPs focusing on median and covering-based models, respectively. Şahin and

Süral (2007) provide a more comprehensive classification scheme and review of the growing literature on HFLPs before 2007. Daskin (2013) discusses the basic notions of hierarchical facilities and applications and describes some of the fundamental median and covering HFLPs. Zanjirani Farahani et al. (2014) review applications, models and solution algorithms mainly for median and covering-based HFLPs. Ortiz-Astorquiza et al. (2018) present a comprehensive review of a special case of HFLPs, denoted multi-level facility location problems, and propose a unified framework to classify them based on the types of strategic and tactical decisions involved.

This chapter is organized as follows. In Sect. 13.2 we first discuss the fundamentals of discrete HFLPs, including their distinguishing features and key concepts. We also provide a comprehensive classification scheme based on several attributes such as nature of customer demand, service availability, flow pattern, and decisions and objectives. A review of classical and more recent applications areas is then given in Sect. 13.3. Finally, in Sect. 13.4 we provide a concise overview of four classes of HFLPs that have received the most attention in the literature: *multi-level facility location*, *median* and *covering hierarchical facility location*, *multi-echelon location-routing*, and *hierarchical hub location*. In particular, we highlight their main characteristics and point out to the some integer programming formulations and efficient algorithms to solve some variants of them. We note that our intention is neither to pose a disjoint partition of the set of HFLPs nor to offer an exhaustive survey on the topic but rather to clarify the most relevant features of HFLPs.

13.2 Fundamentals

The distinguishing features and key concepts arising in hierarchical systems need to be introduced in order to have a better understanding of the structure and inherent complexity of modeling and solving HFLPs. We next discuss the following aspects: (1) nature of customer demand, (2) service availability, (3) flow pattern and spatial configuration, and (4) decisions and objectives.

13.2.1 Nature of Demand

When referring to the nature of demand in the context of a hierarchical system we must identify who the customers are, which services they may request, and how these services will be offered by the facilities.

The most common setting consists of a set of customers represented as demand points in a graph with service given *at* or *from* a facility. Depending on the context, customers may correspond to patients, retail stores, patrons, students, geographical zones, etcetera. These *customers* can have one or more types of demand. In other words, each customer may have different service requirements to be offered by facilities. For instance, a patient that requires a specialized test offered at regional hospitals has a type of demand different from one whose requirement is an annual control visit offered at local clinics. Other patients could request both diagnostic and out-patient surgery services offered at regional hospitals. Note that in this context, service is given at facilities and thus customers have to travel to them to receive service. We denote as *single demand* and *multiple demand* customers to distinguish cases in which one or more types of demand are requested by customers, respectively.

In some other situations, customers may request several types of service as well as multiple products. For example, in the fashion industry, retail stores sell hundreds of products produced in dozens of manufacturing facilities. Customers (i.e. retail stores) continuously request production, transportation and storage services of various products such as clothes, shoes, accessories, etcetera. Note that in this context, service is provided by facilities and thus products are eventually delivered to the customers. We note that in this context, each customer demands both products and services, and these may be different from one customer to the other. For instance, one customer may request production and transportation services for clothes and shoes whereas another customer may request production, transportation, and storage services only for accessories. We denote as *single product* and *multiple product* customers to differentiate cases in which one or more types of product are requested by customers, respectively.

Another class of customers arising in hierarchical systems are those in which service demand corresponds to the movement of commodities, people, or information between an origin and a destination point. Each origin/destination (O/D) pair represents one or more customers requesting transportation or communication services between two specific points. Facilities provide such services by acting as transshipment and consolidation points in the paths of many O/D pairs. For instance, in express package delivery, a package is usually picked-up directly at its origin by a small vehicle and then moved to a branch office in order to be consolidated with other packages. It is then moved forward using larger vehicles to regional and possible central hub facilities where the package is sorted and rerouted to its destination.

Except from the hierarchical hub location models discussed in Sect. 13.4.4, in this chapter we focus on models in which customers are represented as demand points.

13.2.2 Service Availability

The nature of the customers demand is intimately related to the services that facilities can offer. This has been referred to as *service availability* or *service varieties* criterion in classifications of HFLPs (see, Narula 1986; Şahin and Süral 2007).

According to Narula (1986), a *successively inclusive facility hierarchy* is one in which a facility at level r provides all services offered by a facility at level $r - 1$ plus one or more additional services. A *successively exclusive facility hierarchy* is one in which the set of services offered by a facility at level r are not offered by any other facility at levels $q \neq r$. We note that there is a third category, denoted as *mixed facility hierarchy*, in which the set of services offered by a facility at level r has a non-empty intersection with the set of services offered by facilities at other levels $q \neq r$, without necessarily offering all services of lower-level facilities. Şahin and Süral (2007) refer to the first category as *nested* and the last two as *non-nested*.

Additional categories can be considered in more complex settings in which services offered by the facilities also depend on geographical considerations. For example, one can refer to *locally inclusive service hierarchies*. These make use of a measure such as distance to determine whether or not a service is offered to a particular customer. For instance, suppose that facility of type r located at node i offers services 1 through r to demands from node i, but does not offer services 1 to $r - 1$ to demands from nodes $j \neq i$. That is, only service r is available for demands whose origin is not i. For example, a regional hospital located in a given district may provide basic care to patients living is such district but not to other patients living in other districts. However, it may provide out-patient surgical services to any patient when needed. In contrast, *globally inclusive service hierarchies* consider that a facility at level r located at node i can offer services 1 to r to all customers requesting any of such services. Some examples of problems in which these service hierarchies arise can be found in Tien et al. (1983), Mirchandani (1987), and Daskin (2013).

13.2.3 Flow Pattern and Spatial Configuration

The *flow pattern* refers to the way in which network flows are routed through the various levels of a hierarchical system. Şahin and Süral (2007) propose two possible patterns: *single-flow* or *multi-flow*. In a single-flow pattern, flow can start at a demand point at the customers level and visit facilities in each of the levels until it reaches the highest level k. Similarly, flow can start in a facility at the highest level k and pass through all levels until it arrives to its demand point at the customer levels. On the other hand, in a multi-flow pattern, flow can start at a demand point and visit facilities in a subset of levels, possibly skipping some levels. Alternatively, the flow can start at a facility in any level r and pass through a subset of levels until arriving to its demand point. A third type of pattern, denoted as *bidirectional-flow*, arises whenever demand flow is routed in both directions. That is, in a bidirectional-flow pattern, flow starts at the customer level and visit some or all higher-levels and then is routed back to lower-levels until it arrives to a demand point at the customer level.

In some applications, regardless of the considered flow pattern, a portion of customer demand served at a given level is eventually referred for additional services to a higher level. This is denoted as *referral* systems. Alternatively, when referrals are not considered between levels, this can be denoted as *non-referral* systems.

Şahin and Süral (2007) describes another relevant concept, denoted as coherency, which relates to the *spatial configuration* of levels in the hierarchy. A *coherent* system is one in which all demand flow entering to a particular facility at a lower-level is assigned to exactly one facility at a higher-level. Some authors refer to this concept as single assignment due to its resemblance with single-sourcing in single-level discrete location problems. A *non-coherent* system is one in which facilities at lower-levels can send demand flows to more than one facility at a higher level.

13.2.4 Decisions and Objectives

We now discuss the types of decisions and objectives commonly involved in HFLPs. For this purpose, we introduce the main notation used throughout this chapter. Let $G = (V, E)$ be a graph with node set $V = I \cup J$ and edge set E. The set I corresponds to the sites of potential facilities and J to the customers. We consider facilities of types 1 to k. An HFLP involves some of the following decisions.

- **Design Decisions: Facility Location and Edge Activation** The *location decisions* determine where to open the facilities. Given an underlying network G, facilities may be located at both the nodes or the edges of the network. Here we focus on discrete location problems, where it is assumed that facilities can only be located at the nodes of G. The *network design decisions* select the edges to be activated. These edges are used to provide transportation or communication services between demand points and facilities, and between facilities of the same or different levels. We concentrate on those problems where the facility location decisions are non-trivial.
- **Tactical Decisions: Allocation and Routing** The *allocation decisions* determine which facilities will be used to serve each demand point. In FLPs, two types of allocation strategies have been considered. In single allocation, each customer is assigned to exactly one facility, whereas in multiple allocation each customer is allowed to be assigned to more than one facility, if beneficial. The *routing decisions* indicate the routes (or paths) on G that will be used to satisfy the customer demands. We use the term route to indicate the sequence of edges used to send flows between pairs of nodes. These types of decisions commonly appear in network flow problems which have been widely studied (Ahuja et al. 1993). In the general case of an HFLP the term routing could also be used in the sense of location-routing models (see Chap. 15) where tours or paths between nodes of the same level in the hierarchy are considered. Finally, observe that the network design and routing decisions are interrelated, since the edges that can be used in the paths are determined by the network design decisions.

Both of the above types of decisions are directly related to the fixed and variable costs. For example, when a node $i \in I$ is selected to locate a facility, a setup cost f_i

may be incurred. However, note that these costs in the context of HFLPs typically depend on the type of facility. Analogously, when an edge $\{a, b\} \in E$ is activated a setup cost h_{ab} may be paid. On the other hand, the tactical decisions are affected by variable costs. A common example are transportation costs which are generally related to the distances between the nodes. Assuming that customers are at level 0, transportation costs (or distances) $c_{i_r i_s}$, for r and s in $\{0, \cdots, k\}$ are variable since they typically depend on the flow passing through the corresponding edge $\{i_r, i_s\}$ which is dependent on customers demands.

In any case, we note that in an HFLP there must be location decisions involved for one or more types of facilities. Depending on the application, network design and routing decisions may be explicitly considered or not, that is, the activation of edges and flow patterns are not necessarily non-trivial decisions in this context. These types of decisions will help us define differentiating features for families of HFLPs in the following sections.

Analogously to single-level facility location, HFLPs can be classified based on the type of objective function.

- *fixed-charge models:* consider that the number of facilities to locate at each level is not known a priori, but a fixed setup cost f_i for each facility at each level is considered. The objective is to minimize the sum of facilities fixed costs and of demand-weighted distance.
- *median models:* consist of selecting a set of facilities to open, such that no more than a given number of facilities is opened with the objective of minimizing the total demand-weighted distance (or transportation cost). In this case, the maximum number of facilities to open can be given by type r, as p_r, or in total. Moreover, it can be generalized to one or multiple budget constraints that limit either the total setup cost incurred in locating facilities at each level or for all of the facilities.
- *coverage-based models:* assume that a customer is covered if its demand point is within a specified distance of facilities offering the requested service(s). *Set-covering* models assume that all service demand must be covered and the goal is to minimize the setup cost for the facilities. *Maximum covering* models consider that the number of facilities to locate is given as an input and the objective is to maximize the total number of demands of all types that are covered. Some variants of maximum covering models assume that each demand point can only be considered as being covered if all the requested services at such point are satisfied.

13.2.5 Classification Scheme

The classification scheme takes into account the attributes mentioned above, namely *flow pattern and spatial configuration*, *service availability*, *nature of customers demand* and *decisions and objective*. The classification criteria for HFLPs is

Table 13.1 Classification criteria

Criterion	Description
Nature of the demand	Single demand/multiple demand, single product/multiple product
Flow pattern	Single-flow/multi-flow/bidirectional-flow
Service availability	Nested/non-nested/locally inclusive/globally inclusive
Spatial configuration	Coherent (single assignment)/non-coherent
Decisions	Network design/tactical/network design and tactical
Objective function	Median/fixed-charge/covering

summarized in Table 13.1. This classification generalizes the scheme proposed by Şahin and Süral (2007) which in turn extends that of Narula (1986). Note that additional attributes may be considered such as capacities at facilities and edges as well as the interaction between facilities of the same type or between customers. However, we focus on those that we consider to have the most impact on the hierarchical structure.

13.3 Applications

We next review some of the most relevant areas of application of HFLPs. These range from heath care and production-distribution systems, which are arguably among the oldest and most studied hierarchical systems in location science, to telecommunication and transportation systems which have given rise to new variants of HFLPs.

13.3.1 Health Care Systems

Given the wide variety of services that health care systems must provide within a specific region, these services are naturally governed by a hierarchical structure. Although the number of levels may vary by country and region, three-level systems are commonly found. Patients in a geographical region (i.e. neighborhood, district, or county) requesting a variety of services are modeled with a demand point at level 0. Local clinics (level 1) may provide basic care and diagnostic services. Community hospitals (level 2) could offer, in addition to basic care and diagnostic services, other services such as out-patient surgery and specialized clinical tests. Regional hospitals provide a wide variety of in-patient services and may or may not provide basic care and diagnostic services. Successively inclusive facility and mixed facility hierarchies as well as multi-flow patterns are predominant in health care systems. It is also somehow common to observe systems with a locally inclusive service hierarchy, specially in countries with a public health care system (Rahman and Smith 2000; Smith et al. 2013). In this case, a regional hospital may provide basic care and diagnostic services but only to patients living in the proximity of the hospital. The interaction between facilities of different levels arises when patients are referred to community or regional hospitals after being diagnosed at a local clinic. Thus, it is important for health authorities to jointly determine the location of various health care facilities and how these will be interconnected to provide the best possible service to patients. We refer to Ahmadi-Javid et al. (2017) for a recent survey on the topic and to Chap. 23 for other facility location problems in health care.

13.3.2 Production-Distribution Systems

The design of production-distribution systems plays a central role in supply chain management. In particular, for companies that produce and deliver their goods. In such cases, a variety of products are produced in manufacturing facilities (level 3) and shipped to distributions centers (level 2), in which products are sorted, consolidated, and rerouted to regional warehouses (level 1) in order to be stored for a period of time (days, weeks, months) before being finally distributed to retail stores (level 0) for sale. This is a common example of a single-flow pattern on a multiple product environment. The interaction between facilities arises naturally due to the routing of products through the supply chain (from the highest level to its lowest level). Sometimes a multi-flow pattern may arise whenever products are directly shipped from manufacturing facilities to regional warehouses or retail stores. Both coherent and non-coherent structures have been reported in the literature (see, for instance Şahin and Süral 2007; Gendron et al. 2016). Locational and network design decision arising in supply chain management have been extensively studied in the literature. We refer to Melo et al. (2006) and Chap. 16 for an in-depth discussion of facility location problems in the context of supply chain and logistics.

13.3.3 Telecommunications Systems

Telecommunication networks are frequently built with a hierarchical structure having two or three levels. A classical example is the so-called hub-and-spoke architecture used in various distributed data networks. In those cases, service demand corresponds to electronic data transmissions between O/D pairs that are routed over a variety of links in the access-level and backbone-level networks. Hub facilities correspond to electronic equipment such as concentrators, multiplexors, and routers. An example of a three-level hub network arises in an intra-local access transport area network architecture (see, Wu et al. 1988; Yaman 2009). O/D nodes are the central offices and each central office is served by a regional hub. A group of central offices served by the same hub is referred to as a cluster. Each hub is then connected to a central hub (gateway). A group of clusters served by the same central hub is a sector. Central hubs are connected by a complete fiber network. Communication services between O/D pairs are offered by using O/D paths in which data transmissions first visit the regional hub of its cluster, then one or two central hubs, and finally the regional hub of the cluster where the destination point belongs to. That is, demand flows move from lower-levels to higher-levels and back to lower-levels of the hierarchy (i.e. a bidirectional-flow pattern). Catanzaro et al. (2011) provide another example of a two-level hub network arising in the deployment of an Internet routing protocol called Intermediate System-Intermediate System (ISIS). We refer to Chap. 12 for a description of hub location problems arising in the design of hub-and-spoke networks and to Fortz (2015) for an overview of other location problems arising in telecommunications.

13.3.4 Urban Transportation Systems

Hierarchical structures of interacting facilities have recently appeared in the area of city logistics, in which consolidation activities can take place at different levels of an urban supply chain (Mancini et al. 2014; Savelsbergh and Van Woensel 2016). Many logistics companies deliver goods destined for an urban area by using long-haul transportation vehicles that arrive at consolidation facilities (level 2). These facilities are referred to as urban consolidation centers and are usually located in the boundaries of the urban zone. Commodities are then unloaded, sorted, consolidated, and loaded into smaller vehicles which are then routed to other intermediate logistics facilities (level 1), usually referred as cross-dock satellites. From these facilities commodities are shipped to retail stores (level 0) using different vehicle fleets to avoid the presence of large vehicles in the city center. Contrary to urban consolidation centers, cross-dock satellites can be located within the urban zone, even in dense populated areas. These may correspond to basic rendezvous points such as parking lots, rail stations or bus exchanges, where commodities are transferred from one vehicle to another. Cross-dock satellites can also be small warehousing facilities with limited storage capabilities. In any case, transshipment of flows is done in a highly synchronized fashion. The use of these consolidation-distribution strategies can follow a single-flow or multi-flow pattern and all flows are clearly routed form top to bottom.

13.3.5 Air Transportation Systems

Hub-and-spoke architectures are also widely used in air freight and passenger travel systems. In the case of the airline industry, global alliances and mergers have given rise to complex global air transportation systems (Adler and Smilowitz 2007; Bernardes Real et al. 2018). Some alliances operate extensive three-level hub networks, in which local airports (level 0), regional hubs (level 1), and international gateways (level 3) interact among them to route millions of passengers each year. Regional hubs allow passengers to make connections along their routes, while gateways are necessary for connecting continents and for performing immigration, customs and security checks. Passengers traveling within the same continent or geographical region are routed via regional hubs, whereas transcontinental passengers are frequently routed via a combination of regional hubs and gateways to reach their destinations. Local airports are connected to one or more regional hubs and maybe a gateway. International gateways are connected (indirectly or directly) to all regional hubs in its geographical region. All international gateways are interconnected across continents. Similar to telecommunications systems, demand flow follows a bidirectional-flow pattern.

13.3.6 Cargo and Postal Delivery Systems

Other transportation systems, such as cargo and postal delivery, employ a hierarchical structure in their facilities and a mix of air and ground transportation services. In the case of postal services, customers deliver mail or small parcels at post boxes (level 0) in a city. At branch offices (level 1) customers can deposit mail and obtain other services such as buying stamps and envelopes, among other things. Postal flow is then routed to central post offices (level 2) to be sorted and rerouted to other central post offices and branch offices for delivery.

A similar situation arises in the case of cargo delivery systems where branch offices (level 1) collect and distribute cargoes from/to customers (level 0) directly using small trucks. Operations centers (level 2) collect and distribute cargoes on different geographical regions, which are connected with a central hub facility (level 3). This means that all flow must pass through a central hub facility. In some applications (see, for instance, Dukkanci and Kara 2017), operation centers and a central hub facility are connected with a set of tours performed by airplanes.

In both cargo and postal delivery systems the time aspect plays a major role. Thus, it is usually integrated in the design of the system with location and link activation decisions. In particular, these hierarchical networks must be designed in such as way that transportation services between O/D demand points can be performed within a predefined service time limit (i.e. same-day or next-day delivery). Operational scheduling decisions, such as release times at branch offices and operations centers need to be taken into account while designing the network to ensure demand flows can be delivered on time (Yaman et al. 2012).

13.4 Families of Hierarchical Facility Location Problems

Several variants of HFLPs have been studied under various names. We recall that when referring to different types of facilities numerous terms have been used such as level, layer, echelon, stage, tier, among others. We next discuss four classes of HFLPs: multi-level facility location, median and covering hierarchical facility location, multi-echelon location-routing, and hierarchical hub location. The first two families are those HFLPs that were first studied since the early 1970's and also the ones that have received the most attention in the literature. On the other hand, the last two families of HFLPs have emerged over the last two decades and fewer references can be found. We would like to clarify that these four classes of HFLPs do not constitute a partition of the general field of HFLPs. That is, there could also be some overlapping between them and also some HFLPs may not necessarily belong to any of these classes of problems.

13.4.1 Multi-Level Facility Location Problems

Multi-level facility location problems (MLFLPs) are typically found in the context of supply chains and production-distribution systems. There are two main distinguishing features underlying most MLFLPs: there exist an inherent hierarchy given by the nature of the system and a successively exclusive facility hierarchy is usually considered. For example, in the case of production-distribution systems, the products need to be first produced in order to be shipped to regional warehouses for temporary storage. Once the products are requested with a given due date, they are routed to retail stores. In this case, the hierarchy of the different types of facilities is implicitly given by the nature of the system, i.e one cannot store a product which has not yet been produced. Moreover, production services offered at manufacturing facilities are not available at warehouses. Similarly, warehousing services such as sorting, labeling and consolidation operations as well as storage space are usually not available at manufacturing facilities.

Another distinguishing feature of MLFLPs is that non-trivial facility location decisions are taken at every level of the hierarchy, simultaneously. Other problems involve two or more types of facilities but only in one of them is the selection of facilities considered (see for instance Sect. 13.4.3). Moreover, in a MLFLP, there is no direct interaction between customers, and no horizontal interactions between facilities of the same level. Typically, the edges between facilities of different types are defined sequentially. Thus, a sequence of exactly one opened facility at each level is required which corresponds to what we called a single-flow pattern. Nevertheless, some problems with multi-flow patterns could also be considered as MLFLPs when demand flows are allowed to skip levels in the hierarchy (i.e some services may not be requested by some customers). Most of these multi-flow pattern problems can be modeled as single-flow-patterns by simply adding dummy nodes in the corresponding missing levels, at the expense of increasing the instance size. Also, when flow directions are considered, the flow between levels of an MLFLP must go in one direction and there ought to be only one type of arc available. Some HFLPs, especially those that arise in the framework of waste management systems, consider bidirectional-flows or more than one type of arc (see, for instance Barros et al. 1998; Mitropoulos et al. 2009). In terms of the coherency criterion both cases have been studied for MLFLPs. Finally, we note that fixed-charge and median-based objective functions are more common for this class of problems.

As an example of an MIP formulation for an uncapacitated MLFLP that extends the *uncapacitated facility location problem* (UFLP) and the *p-median problem* (*p*-MP), we consider the following so-called path-based formulation (Tcha and Lee 1984; Aardal et al. 1999). Let $G = (I \cup J, E)$ be a graph with vertex set $I \cup J$ partitioned into $k+1$ levels, where J represents the set of customers, I is partitioned into $\{I_1, \cdots, I_k\}$, corresponding to the sets of potential facilities at levels 1 to k, and E is the set of edges. Let S be the set of all possible simple paths having exactly one node from each level, starting from some node $i_1 \in I_1$, finishing at some node $i_k \in I_k$. Also, consider c_{js} to be the cost associated with the allocation of customer $j \in J$ to the sequence of facilities in path $s \in S$. Now, let $p = (p_1, \cdots, p_k)$ be a vector of positive integers corresponding to the maximum number of facilities that can be opened at each level, and let f_{i_r} be the non-negative fixed cost associated with opening facility i_r at level r. We define the binary variables x_{js} equal to one if and only if customer $j \in J$ is assigned to path $s = i_1, \cdots, i_k \in S$. Also, we define the binary variables y_{i_r} equal to one if and only if facility i_r of level r is open. The formulation is the following:

$$\text{minimize} \quad \sum_{j \in J} \sum_{s \in S} c_{js} x_{js} + \sum_{r=1}^{k} \sum_{i_r \in I_r} f_{i_r} y_{i_r} \tag{13.1}$$

$$\text{subject to} \quad \sum_{s \in S} x_{js} = 1 \quad \forall\ j \in J \tag{13.2}$$

$$\sum_{s \in S: i_r \in s} x_{js} \le y_{i_r} \quad \forall\ j \in J,\ \ i_r \in I_r,\ \ r = 1, \cdots, k \tag{13.3}$$

$$\sum_{i_r \in I_r} y_{i_r} \le p_r \quad r = 1, \cdots, k \tag{13.4}$$

$$x_{js} \ge 0 \quad \forall\ j \in J,\ \ s \in S \tag{13.5}$$

$$y_{i_r} \in \{0, 1\} \quad \forall\ i_r \in I_r,\ \ r = 1, \cdots, k. \tag{13.6}$$

The objective (13.1) is to minimize the sum of the assignment costs and the setup cost for opening facilities at different levels. Constraints (13.2) ensure that exactly one path is assigned to every customer, while constraints (13.3) are the linking constraints which ensure that if a path is assigned to a customer, then all the facilities in such path must be open. Constraints (13.4) are the cardinality restrictions. Finally, note that the variables x_{js} can be relaxed from binary to continuous variables, as for the UFLP (see Chap. 4).

Some properties and characteristics of classical FLPs have been extended for the more general case of MLFLPs. For instance, Aardal et al. (1996) showed that all non-trivial facet defining inequalities for the UFLP also define facets for the two-level uncapacitated facility location problem. Aardal et al. (1999), Bumn and Kern (2001), and Zhang (2006) use ideas previously developed for the UFLP, such as dual ascent and adjustment techniques (Erlenkotter 1978), in order to develop approximation algorithms for the multi-level UFLP. In this context, it is important to note that most research efforts towards the development of algorithms for MLFLPs have focused on heuristics. In particular, we can differentiate two main research streams in this field: heuristics without a performance guarantee, and ρ-approximation algorithms i.e., polynomial-time heuristics that yield a feasible solution with an objective function value lying within a factor of ρ of the optimal value. Most of the work has focused on the latter stream. A more recent example is the work of Krishnaswamy and Sviridenko (2016) who presented inapproximability results for the multi-level UFLP and showed that in the general case, the two-level UFLP is computationally harder than the single-level UFLP.

Most of the early works on MLFLPs introduced exact algorithms for different variants of the problem. For example, Kaufam et al. (1977) presented a branch-and-bound method that extended from the single-level case. Barros and Labbé (1994a) introduced a general version of an MLFLP including design and tactical decisions and developed a branch-and-bound procedure using the corresponding upper and lower bounds obtained from different Lagrangian relaxations of two formulations, and those obtained from an extension of the greedy heuristic proposed for the UFLP. More recently (Gendron et al. 2016; Ortiz-Astorquiza et al. 2019), developed efficient exact methods for MLFLPs based on Lagrangian relaxation and Benders decomposition, respectively.

As mentioned before, most of the techniques used to solve MLFLPs are especially modified from successful algorithms developed for single-level FLPs. One very important property in discrete optimization that has led to the development of algorithms for FLPs is submodularity. This property somehow resembles convexity for continuous functions on set functions. For the single-level case (Cornuéjols et al. 1977), presented worst-case bounds for greedy and local improvement heuristics for the maximization version of an FLP which includes as special cases the UFLP and the p-MP (see Chap. 4 for details on supermodularity and supermodular reformulations for the minimization version of the UFLP). Some of the first articles discussing MLFLPs assumed that the submodularity property extends directly from the single-level cases (Ro and Tcha 1984; Tcha and Lee 1984). Later (Barros and Labbé 1994b), showed that the set function associated with the natural combinatorial representation of the multi-level UFLP does not satisfy submodularity. However, other equivalent combinatorial optimization problems modeling the multi-level UFLP have an objective function that actually satisfies submodularity, as was shown in Ortiz-Astorquiza et al. (2015). This observation allowed to provide sufficient conditions to extend the results on worst-case bounds for greedy heuristics and submodular reformulations of single-level UFLP and pMP to MLFLPs (Ortiz-Astorquiza et al. 2017).

Similarly, the case of having capacities in the facilities has also received important attention. From the early works, we note that of Aardal (1992), who presented an MILP formulation for the two-level capacitated FLP and a polyhedral study. Aardal (1998) later introduced a reformulation along with computational results. Marín and Pelegrín (1999) compared two-index and a three-index formulations for the development of an exact algorithm for the two-level capacitated FLP based on Lagrangian relaxations. As for the uncapacitated case (Bumn and Kern 2001; Ageev 2002; Du et al. 2009), developed ρ-approximation algorithms for capacitated MLFLPs with values of ρ equal to 12, 9 and $k+2+\sqrt{k^2+2k+5}+\epsilon$, respectively. Finally, multi-period (or dynamic) extensions of MLFLPs have also been studied in Hinojosa et al. (2000, 2008). For more details on classification, models, properties and solution methods for MLFLPs we refer to Ortiz-Astorquiza et al. (2018).

13.4.2 Median and Covering Hierarchical Location Problems

HFLPs that are referred to as *median-based hierarchical location problems* (MHLPs) and *covering-based hierarchical location problems* (CHLPs) in the literature are those which are frequently found in the context of health care systems, educational systems, and emergency medical services. One of the main distinguishing features of MHLPs and CHLPs is that either a successively inclusive facility hierarchy or a mixed facility hierarchy is considered. Similar to MLFPLs, there exist an inherent hierarchy given by the nature of the system. For instance, regional hospitals are clearly in a higher level of the hierarchy as compared to community hospitals and local clinics. The service availability may be successively inclusive or not but the hierarchy of the hospitals is implicitly given by the level of urgency or criticality of the offered services.

The objective function has been one of the most important factors when categorizing FLPs in general. The modeling and solution structures might change drastically when different objectives are considered (e.g. p-median, p-center, fixed charged and covering). One stream of research has focused on MHLPs in which a median objective is considered. An example of a nested multiple demand MFLPs can be formulated as follows. Let d_j^s denote the demand of service s at node j and c_{ji} denote the cost associated with the allocation of customer j to facility i. Additionally, consider the decision variables y_{ir} equal to one if and only if facility of type r is located at node i and variables x_{is}^j equal to one if and only if demand of service s at node j is satisfied with facility at node i. Then, we obtain

$$\text{minimize} \quad \sum_{i\in I}\sum_{j\in J}\sum_{s=1}^{k} d_j^s c_{ji} x_{is}^j \tag{13.7}$$

$$\text{subject to} \quad \sum_{i\in I} x_{is}^j = 1 \qquad \forall j \in J \;\; s = 1,\cdots,k \tag{13.8}$$

$$\sum_{i\in I} y_{ir} \le p_r \qquad r = 1,\cdots,k \tag{13.9}$$

$$x_{is}^j \le \textstyle\sum_{l=s}^k y_{il} \qquad \forall \;\; j \in J,\; i \in I \;\; s = 1,\cdots,k \tag{13.10}$$

$$x_{is}^{j} \in \{0, 1\} \qquad \forall i \in I \;\; j \in J, \;\; s = 1, \cdots, k \quad (13.11)$$

$$y_{ir} \in \{0, 1\} \qquad \forall \; i \in I, \;\; r = 1, \cdots, k. \quad (13.12)$$

The objective (13.7) is to minimize the total assignment cost. Constraints (13.8) ensure that the demand of each service of every node is met. Constraints (13.9) limit the number of open facilities of each type, while inequalities (13.10) are linking constraints which in this case define a successively (globally) inclusive HFLP. Note that replacing constraints (13.10) with $x_{is}^{j} \leq y_{is}$ modifies the problem into a non-nested HFLP. Then, we would have a successively exclusive formulation. Some of the early works on the subject are precisely those that identified the differences between a successively exclusive, successively inclusive or locally inclusive HFLP (Tien et al. 1983; Mirchandani 1987).

Later Weaver and Church (1991), formulated a nested MHLP with two types of facilities minimizing an objective function similar to that of Narula and Ogbu (1985). They proposed a Lagrangian procedure and a primal exchange substitution heuristic. Other examples where hierarchical p-median models have been studied are those of Galvão et al. (2002), Yasenovskiy and Hodgson (2007) and Hodgson and Jacobsen (2009). Also (Serra and ReVelle 1993; Alminyana et al. 1998), present solution methods for a nested and coherent hierarchical structure combining two p-median problems referred to as the pq-median problem.

On the other hand, when referring to covering objectives, an important notion is that of a demand point being *covered* (see Chaps. 3 and 5). In the context of HFLPs the three most common types of covering found in single-level FLPs have also been studied, namely the hierarchical extensions of the set covering location problem, the p-center problem, and the maximum covering problem (Toregas et al. 1971; Church and ReVelle 1974). Note that in these cases because we are considering different types of facilities, it is more intuitive to talk about different types of demand. However, defining a critical distance is also more challenging than in the single-level case. For example, one may be interested in covering a demand point by each type of facility. Another case would be, for instance, when each customer must be covered by a first level facility, first level facilities in turn are covered by second level facilities and so on. In the context of health care systems for example, this is referred to as bottom-up referral system (Church and Eaton 1987; Gerrard and Church 1994). In such cases the facilities are typically service-nested and thus a second level facility can also cover customers. Therefore, covering type objectives are more commonly found with multiple demand type of customers.

For example, let α_{ir}^{js} be a parameter whose value is equal to one if and only if demand at node j of service s can be covered by facility of type r from node i. Also, let z_{js} be the binary variables equal to one if and only if demand of service s

from node j is covered. A hierarchical maximum coverage location problem can be formulated as

$$\text{maximize} \quad \sum_{j \in J} \sum_{s=1}^{k} d_j^s z_{js} \tag{13.13}$$

$$\text{subject to} \quad \sum_{i \in I} y_{ir} \leq p_r \qquad r = 1, \cdots, k \tag{13.14}$$

$$z_{js} \leq \sum_{i \in I} \sum_{r=1}^{k} \alpha_{ir}^{js} y_{ir} \ \forall \ j \in J, \ \ s = 1, \cdots, k \tag{13.15}$$

$$0 \leq z_{js} \leq 1 \qquad \forall \ j \in J, \ \ s = 1, \cdots, k \tag{13.16}$$

$$y_{ir} \in \{0, 1\} \qquad \forall \ i \in I, \ \ r = 1, \cdots, k. \tag{13.17}$$

The objective (13.13) is to maximize the sum of covered demand of each type of service. Constraints (13.14) limit the number of open facilities of each type, whereas constraints (13.15) ensure that demand at each node j for each service s is considered to be covered if and only if there exist at least one open facility which can provide such service to that node. The integrality restrictions on the z_{js} variables can be relaxed due to the sense of the objective function and constraints (13.17).

Note that in this formulation we may count as covered some demand points that are only partially covered. That is, customers with multiple demands which are covered for only some services still add value to the objective function. Another case is to impose that only complete covered customers add value to the objective function of total coverage.

Given the applicability of CHLPs several different variants have been presented focusing on case studies and analysis of solutions. One of the first works on CHLPs is that of Moore and ReVelle (1982) who proposed an IP formulation for a hierarchical problem with two types of facilities. Later (Gerrard and Church 1994), discussed and compared three additional CHLPs to the one proposed by Moore and ReVelle (1982). Marianov and Serra (1998, 2001) studied a CHLP in the context of congested systems. Espejo et al. (2003) developed dual based heuristics using Lagrangian relaxation to solve instances of a CHLP with two types of facilities. More recently (Lee and Lee 2010), proposed tabu-based heuristics for a generalization of the model introduced by Moore and ReVelle (1982). For more examples, formulations and solution methods on this family of HFLPs we refer the reader to Şahin and Süral (2007), Daskin (2013) and references therein.

13.4.3 Multi-Echelon Location-Routing Problems

The term *echelon* is generally associated with distribution networks where products are transported between each pair of levels. Such pairs are called echelons (Aiken 1985; Gao and Robinson 1992). Multi-echelon FLPs are thus very similar to

MLFLPs. In fact, many studies use both terms indistinctly. However, we note two main differences. The first one is that although all of the multi-echelon problems involve a multi-level environment, not all of them require facility location decisions at every level as in an MLFLP. For example, in one of the early works on the topic Geoffrion (1974) studied two-echelon FLPs in which facilities to be opened are only selected at one of the levels. This is partially because the predominant decisions are made at the echelons, and these typically involve routing decisions. Routing decisions in both senses, i.e., the flow between types of facilities and customers as well as the routes or tours in the same level of the hierarchy. Indeed, the second differentiating feature lies precisely in these routing patterns. MLFLPs are concerned with problems where facility, and sometimes network design decisions, are predominant with no routing decisions between nodes of the same level involved. The paper of Cuda et al. (2015) on two-echelon routing problems reviews a more general class of problems in which locational decisions are optional at all levels. That is, they include problems that may have no location decisions involved.

Another term that is generally related to echelons is the word *tier*, which has mainly been used in the context of freight transportation systems and city logistics (Crainic et al. 2009; Mancini et al. 2014). These HFLPs typically involve vehicle routing decisions extending those FLPs studied in Chap. 15. The term *stage* has also been used in this context and is possibly the most elusive one when trying to associate it to something in particular. In some references (e.g. Marín 2007) the term stage is used when referring to what we denote as levels. However, in other papers it has been used in the sense of what we identified as echelons (e.g. Klose 1999).

One example of a route-based formulation (Cuda et al. 2015) for a two-echelon capacitated location-routing problem is as follows. Let T^1 be the set of routes where each $t \in T^1$ starts from a facility of level 1 (e.g. warehouses) and visits one or several customers. Similarly, define the set of routes T^2 for the second echelon starting at a facility of the second level (e.g. plants) and visiting a group of warehouses. The binary parameters α_{ti_1} and β_{tj} indicate whether facility i_1 or customer j are in route t or not. Finally, given a route $t \in T^1$, let $d_t = \sum_{j \in t} d_j$ be the total demand for customers visited. Additionally to the fixed costs for setting up facilities f_i, consider the fixed costs paid for each vehicle used in each echelon g_r and the cost per route b_t. Now, let the binary decision variables y_{i_r} equal to one if and only if facility $i_r \in I_r$ of level $r = 1, 2$ is opened. Also, let variables x_t equal to one if and only if the route $t \in T^1 \cup T^2$ is in the solution, and let w_{ti_1} be a flow variable for route $t \in T^2$ that must be delivered to facility i_1. Then we have the following MIP formulation:

$$\text{minimize} \quad \sum_{r=1}^{2} \left(\sum_{i_r \in I_r} f_{i_r} y_{i_r} + \sum_{t \in T^r} (g_r + b_t) x_t \right) \tag{13.18}$$

$$\text{subject to} \quad \sum_{i_1 \in I_1} w_{ti_1} \leq c^2 x_t \qquad \forall t \in T^2 \tag{13.19}$$

$$\sum_{t \in T^2 : i_2 \in t} \sum_{i_1 \in I_1} w_{ti_1} \leq q_2 y_{i_2} \qquad \forall i_2 \in I_2 \tag{13.20}$$

$$\sum_{t \in T^2} w_{ti_1} \leq q_1 y_{i_1} \qquad \forall i_1 \in I_1 \tag{13.21}$$

$$\sum_{t \in T^2} \alpha_{ti_1} x_t = y_{i_1} \qquad \forall i_1 \in I_1 \tag{13.22}$$

$$\sum_{t \in T^2} w_{ti_1} = \sum_{t \in T^1 : i_1 \in t} d_t x_t \qquad \forall i_1 \in I_1 \tag{13.23}$$

$$\sum_{t \in T^1} \beta_{tj} x_t = 1 \qquad \forall j \in J \tag{13.24}$$

$$y_{i_r} \in \{0, 1\} \qquad \forall \ i_r \in I_r, \quad r = 1, 2 \tag{13.25}$$

$$x_t \in \{0, 1\} \qquad \forall \ t \in T^1 \cup T^2 \tag{13.26}$$

$$w_{ti_1} \geq 0 \qquad \forall \ t \in T^2 \ \ i_1 \in I_1, \tag{13.27}$$

where c^r is the capacity of the vehicles in echelon r and q_r is the capacity of facilities in level r. Constraints (13.19) ensure that if a route $t \in T^2$ is selected, then the total load delivered to all the satellites visited in that route must not exceed the vehicle capacity. Constraints (13.20) and (13.21) correspond to the capacity limits for each opened facility at both levels. Constraints (13.22) impose that if the first level facility is opened then it must be visited by one vehicle. Constraints (13.23) represent flow balance equations for each first level facility and constraints (13.24) ensure that each customer is served by one vehicle.

One of the first papers in this family of HFLPs is the one of Jacobsen and Madsen (1980) in which the problem considers location decisions at only one level of the hierarchy. More recent articles have focused on two-echelon location-routing problems having location decisions at both levels. For example, Boccia et al. (2010) and Contardo et al. (2012) present various formulations and solution methods extended from location-routing and vehicle routing problems. In general, as in the more global view of HFLPs, most of the related papers propose heuristic algorithms. In this particular family of HFLPs, perhaps one of the few exceptions where a specialized exact solution algorithm is developed is the branch-and-cut proposed in Contardo et al. (2012).

13.4.4 Hierarchical Hub Location Problems

All previously described HFLPs consider that customers, regardless if they are single/multiple demand or single/multiple product, can be represented with demand points. We now turn our attention to a different class of problems in which service demands correspond to the routing of commodities between O/D pairs over a hierarchical network. We refer to this class of HFLPs as *hierarchical hub location problems* (HHLPs). HHLPs arise in the design of hub-and-spoke networks in telecommunications, air transportation, cargo, and postal delivery systems. Even though standard hub-and-spoke networks involving two-levels (access and backbone levels) have already a hierarchical structure, in this section we limit our study to hub networks involving two (or more) levels and in which location decisions arise. We refer to Chap. 12 for an in-depth analysis of hub-and-spoke networks involving locational decisions in one level.

An important feature of HHLPs is that, given the nature of their demand, most of these problems consider a bidirectional-flow pattern. That is, demand flow originates in an O/D node (level 0) and is routed to the highest level via one or more facilities of each intermediate level and then is routed back to an O/D node in the lowest level visiting once more one or more facilities of each level. Another feature of HHLPs is that facilities of the same level are connected and thus, additional link activation decisions are usually present (unless full interconnection between them is assumed). Yet another interesting characteristic of HHLPs is that in some situations the hierarchy of the different types of hub facilities is given by the nature of the system, such as the case of telecommunication networks in which the role of the electronic equipment determines the sequence in which the flow must be routed. However, in some other situations the hierarchy is given by a ranking mechanism that allows the ordering of the facilities into levels. For example, in the case of passenger airline networks there exist multiple levels of hubs. According to the Federal Aviation Administration (FAA), air traffic hubs are classified based on the percentage of total passengers enplaned in the area into one of four types of hubs: large hubs, medium hubs, small hubs, and non-hubs (Shaw 1993). This means that the hierarchy of facilities (hub airports), is determined by such metric.

We now provide a brief overview of the HHLPs that have been studied in the last decade. In the context of telecommunications networks (Yaman 2009), studies the problem of designing a three-level hub network, where the top layer consists of a complete network connecting the central hubs, and the second and third layers are unions of star networks connecting the remaining hubs to central hubs and the O/D nodes to hubs, respectively. The objective is to minimize the total routing cost while taking into account a cardinality constraint on the number of open hubs at each level. The author also studies an extension incorporating the same delivery time restriction for all O/D pairs. Yaman and Elloumi (2012) focus on two variants of HHLPs with covering-based objectives: the star p-hub center problem and the star p-median problem. These problems consist of locating p hubs (level 1) and connect them at central hub (level 2) via a star topology. Each O/D node is assigned to one hub in level 1. The objective of the former problem is to minimize the length of the longest path between O/D pairs. The objective of the latter is to minimize the total routing cost, while taking into consideration the service quality in terms of the length of paths between pairs of O/D nodes.

In the context of cargo delivery systems (Alumur et al. 2012), introduce a multimodal HHLP in which a three-level hub network is considered. O/D nodes are connected to ground hubs (level 1), which in turn are connected to airport hubs (level 2). A central airport hub (level 3) is connected with a star topology to all airport hubs. Depending on the O/D pairs, demand flow may visit ground hubs, or a combination of ground and airport hubs. The objective is to minimize the sum of setup costs for link activation decisions and routing costs, while taking into account service time constraints on O/D paths. Dukkanci and Kara (2017) study a variant of the previous problem in which a ring-star-star topology is assumed. In particular, the airport hubs are connected with the central airport hub with a set of rings (or routes) performed by the aircrafts. The objective is to minimize the setup cost for the activation of the links between airport hubs while taking into account service time constraints on O/D paths.

For the case of air transportation systems (Adler and Smilowitz 2007), focus on the design of global three-layer hub networks in which two types of hub facilities are considered: international gateways and regional hubs. The backbone network associated with each hub-layer is assumed to be complete. The authors develop a game theoretic approach in which merger and location decisions are considered. Bernardes Real et al. (2018) introduce a more comprehensive model in which gateways (level 2) and regional hubs (level 1) need to be located on a tree-level hub network. Backbone networks are no longer assumed to be complete. Unlike previous models, local and global flows are differentiated as the structure of OD paths associated with each type of flow is different. In particular, global flows can only leave or enter a geographic region via a gateway hub, while local flows can only use domestic hubs within their region.

Zhong et al. (2018) present a HHLPs arising in the design of public transport systems in which a three-level hub network is considered. O/D nodes correspond to traffic districts (level 0) in urban and rural areas. Hubs in central towns (level 1) provide service to rural areas, which are connected to urban public transport hubs (level 2) located inside the city or on the rural-urban boundary. These urban transportation networks are used to satisfy demand generated by urban and rural residents moving into and out of the city each day. The authors consider a fixed charge objective that minimizes the sum of setup costs for the installation of hubs at both levels and transportation costs.

13.5 Conclusions

In this chapter we presented a unified view of hierarchical facility location problems. They constitute a general class of discrete location problems arising in the design of hierarchical systems, in which different types of facilities interact to collectively provide services or products to a set of customers. We discussed the fundamentals of these problems, including their distinguishing features and key concepts. We also provided a comprehensive classification scheme that combines and extends previous schemes. We provided a review on applications areas, focusing on classical and new applications that have recently emerged. We also presented a concise overview of four classes of HFLPs that have received the most attention in the literature: multi-level facility location, median and covering hierarchical facility location, multi-echelon location-routing, and hierarchical hub location.

Although a substantial progress has been done by researches and practitioners in the area of hierarchical facility location, there is still significant work to be done. Identifying new areas of application will give rise to more realistic and complex models capable of capturing features of real-life. For instance, the recent work of Smith et al. (2017) focusing on the location of IV/AIDS test laboratories in South Africa and of Teixeira et al. (2019) dealing with the location of courts of justice in Portugal, provide good examples of the innovative applications of HFLPs. Moreover, although some recent progress has been done in the solution of HFLPs of realistic size (Ortiz-Astorquiza et al. 2019), sophisticated solution algorithms capable of exploiting the network flow structure of hierarchical models still need to be investigated to solve large-scale instances of more realistic variants. Other aspects of HFLPs that have received limited attention include the uncertainty in demand and travel times, as well as the multi-period nature of decision problems involving strategic decisions. Finally, another aspect that has been rarely discussed in HFLPs is that of incorporating into the decision making process the allocation of services to facilities and to exogenously determine the number of levels in the hierarchy (see, Narasimhan and Pirkul 1992).

References

Aardal K (1992) On the solution of one and two-level capacitated facility location problems by the cutting plane approach. Louvain-la-Neuve, Belgium: Université Catholique de Louvain PhD Thesis

Aardal K (1998) Reformulation of capacitated facility location problems: how redundant information can help. Ann Oper Res 82:289–308

Aardal K, Labbé M, Leung J, Queyranne M (1996) On the two-level uncapacitated facility location problem. INFORMS J Comput 8:289–301

Aardal K, Chudak FA, Shmoys DB (1999) A 3-approximation algorithm for the *k*-level uncapacitated facility location problem. Inf Process Lett 72:161–167

Adler N, Smilowitz K (2007) Hub-and-spoke network alliances and mergers: price-location competition in the airline industry. Transp Res B-Method 41:394–409

Ageev, A (2002). Improved approximation algorithms for multilevel facility location problems. Oper Res Let 30:327–332

Ahmadi-Javid A, Seyedi P, Siddhartha SS (2017) A survey of healthcare facility location. Comput Oper Res 79:223–263

Ahuja RK, Magnanti TL, Orlin JB (1993) Network flows: theory, algorithms, and applications. Prentice Hall, New Jersey

Aikens CH (1985) Facility location models for distribution planning. Eur J Oper Res 22:263–279

Alminyana M, Borras F, Pastor JT (1998) A new directed branching heuristic for the pq-median problem. Locat Sci 6:1–23.

Alumur SA, Yaman H, Kara BY (2012) Hierarchical multimodal hub location problem with time-definite deliveries. Transp Res E-Log 48:1107–1120

Barros AI, Labbé M (1994a) A general model for the uncapacitated facility and depot location problem. Locat Sci 2:173–191

Barros AI, Labbé M (1994b) The multi-level uncapacitated facility location problem is not submodular. Eur J Oper Res 72:607–609

Barros AI, Dekker R, Scholten V (1998) A two-level network for recycling sand : a case study. Eur J Oper Res 110:199–214
Bernardes Real L, O'Kelly M, de Miranda G, de Camargo R (2018) The gateway hub location problem. J Air Transp Manag 73:95–112
Boccia M, Crainic TG, Sforza A, Sterle C (2010) A metaheuristic for a two echelon location-routing problem. In: Festa P (ed). Experimental algorithms: lecture notes in computer science, vol 6049. Springer, Berlin, pp 288–301
Bumb A, Kern W (2001) A simple dual ascent algorithm for the multilevel facility location problem. In: Goemans M, Jansen K, Rolim JDP, Trevisan L (eds) Approximation, randomization, and combinatorial optimization: algorithms and techniques. Springer, Berlin, pp 55–63
Calvo AB, Marks DH (1973) Location of health care facilities: an analytical approach. Socio Econ Plan Sci 7:407–422
Catanzaro D, Gourdin É, Labbé M, Ozsoy FA (2011) A branch-and-cut algorithm for the partitioning-hub location-routing problem. Comput Oper Res 38:539–549
Church R, Eaton DJ (1987) Hierarchical location analysis using covering objectives. In: Gosh A, Rushton G (eds) Spatial analysis and location-allocation models. Van Nostrand Reinhold, New York, pp 163–185
Church R, ReVelle C (1974) The maximal covering location problem. Reg Sci Assoc Pap 32:101–118
Contardo C, Hemmelmayr V, Crainic TG (2012) Lower and upper bounds for the two-echelon capacitated location routing problem. Comput Oper Res 39:3215–3228
Cornuéjols G, Fisher ML, Nemhauser GL (1977) Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms. Manag Sci 23:789–810
Crainic TG, Ricciardi N, Storchi G (2009). Models for evaluating and planning city logistics systems. Transp Sci 43:432–454
Cuda R, Guastaroba G, Speranza MG (2015) A survey on two-echelon routing problems. Comput Oper Res 55:185–199
Daskin MS (2013) Network and discrete location. Models, algorithms, and applications, 2nd edn. Wiley, Hoboken
Dökmeci VF (1973) An optimization model for a hierarchical spatial system. J Reg Sci 13:439–451
Dökmeci VF (1977) A quantitative model to plan regional health facility systems. Manag Sci 24:411–419
Du, D Wang, X Xu, D (2009) An approximation algorithm for the *k*-level capacitated facility location problem. J Comb Opt 20:361–368
Dukkanci O, Kara BY (2017) Routing and scheduling decisions in the hierarchical hub location problem. Comput Oper Res 85:45–57
Erlenkotter D (1978) A dual-based procedure for uncapacitated facility location. Oper Res 26:992–1009
Espejo LGA, Galvão RD, Boffey B (2003) Dual-based heuristics for a hierarchical covering location problem. Comput Oper Res 30:165–180
Fortz B (2015) Location problems in telecommunications. In: Laporte G, Nickel S, Saldanha da Gama, F(eds) Location science. Springer, Berlin, pp 537–554
Galvão RD, Espejo LGA, Boffey B (2002) A hierarchical model for the location of perinatal facilities in the municipality of Rio de Janeiro. Eur J Oper Res 138:495–517
Gao L-L, Robinson EPJ (1992) A dual-based optimization procedure for the two-echelon uncapacitated facility location problem. Nav Res Logist 39:191–212
Gendron B, Khuong P-V, Semet F (2016). A Lagrangian-based branch-and-bound algorithm for the two-level uncapacitated facility location problem with single-assignment constraints. Transp Sci 50:1286–1299
Geoffrion, AM (1974) Multicommodity distribution system design by Benders decomposition. Manag Sci 20:822–844
Gerrard RA, Church RL (1994) A generalized approach to modeling the hierarchical maximal covering location problem with referral. Pap Reg Sci 73:425–453

Hinojosa Y, Puerto J, Fernández F (2000) A multiperiod two-echelon multi-commodity capacitated plant location problem. Eur J Oper Res 123:271–291
Hinojosa Y, Kalcsics J, Nickel S, Puerto J, Velten S (2008) Dynamic supply chain design with inventory. Comput Oper Res 35:373–391
Hodgson MJ, Jacobsen SK (2009) A hierarchical location–allocation model with travel based on expected referral distances. Ann Oper Res 167:271–286
Jacobsen SK, Madsen OBG (1980) A comparative study of heuristics for a two-level routing-location problem. Eur J Oper Res 5:378–87
Kaufman L, Eede M, Hansen P (1977) A plant and warehouse location problem. Oper Res Quart 28:547–554
Klose A (1999) An LP-based heuristic for two-stage capacitated facility location problems. J Oper Res Soc 50:157–166
Krishnaswamy R, Sviridenko M (2016) Inapproximability of the multilevel uncapacitated facility location problem. ACM Trans Algorithms 13:1–26
Lee JM, Lee YH (2010) Tabu based heuristics for the generalized hierarchical covering location problem. Comput Ind Eng 58:638–645
Mancini S, Gonzalez-Feliu J, Crainic TG (2014) Planning and optimization methods for advanced urban logistics systems at tactical level. In: Gonzalez-Feliu J, Semet F, Routhier J-L (Eds.), Sustainable urban logistics: concepts, methods and information systems. Springer, Berlin, pp. 145–164
Marianov V, Serra D (1998) Probabilistic, maximal covering location-allocation models for congested systems. J Reg Sci 38:401–424
Marianov V, Serra D (2001) Hierarchical location-allocation models for congested systems. Eur J Oper Res 135:195–208
Marín A (2007) Lower bounds for the two-stage uncapacitated facility location problem. Eur J Oper Res 179:1126–1142
Marín, A Pelegrín, B (1999) Applying Lagrangian relaxation to the resolution of two-stage location problems. Ann Oper Res 86:179–198
Melo MT, Nickel S, Saldanha-da-Gama F (2006) Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Comput Oper Res 33:181–208
Mirchandani P (1987) Generalized hierarchical facility locations. Transp Sci 21:123–125
Mitropoulos P, Giannikos I, Mitropoulos I (2009) Exact and heuristic approaches for the locational planning of an integrated solid waste management system. Operat Res 9:329–347
Moore GC, ReVelle CS (1982) The hierarchical service location problem. Manag Sci 28:775–780
Narasimhan S, Pirkul H (1992) Hierarchical concentrator location problem. Comput Commun 15:185–91
Narula SC (1984) Hierarchical location-allocation problems: a classification scheme. Eur J Oper Res 15:93–99
Narula SC (1986) Hierarchical location problems. Ann Oper Res 6:257–272
Narula SC, Ogbu UI (1985) Lagrangean relaxation and decomposition in an uncapacitated 2-hierarchical location-allocation problem. Comput Oper Res 12:169–180
Ortiz-Astorquiza C, Contreras I, Laporte G (2015) The multi-level facility location problem as the maximization of a submodular set function. Eur J Oper Res 247:1013–1016
Ortiz-Astorquiza C, Contreras I, Laporte G (2017) Formulations and approximation algorithms for multi-level uncapacitated facility location. INFORMS J Comput 29:767–779
Ortiz-Astorquiza C, Contreras I, Laporte G (2018) Multi-level facility location problems. Eur J Oper Res 267:791–805
Ortiz-Astorquiza C, Contreras I, Laporte G (2019) An exact algorithm for multilevel uncapacitated facility location. Transp Sci 53:1085–1106
Rahman SU, Smith DK (2000) Use of location-allocation models in health service development planning in developing nations. Eur J Oper Res 123:437–452
Ro H-B, Tcha D-W (1984) A branch and bound algorithm for the two-level uncapacitated facility location problem with some side constraints. Eur J Oper Res 18:349–358

Savelsbergh M, Van Woensel T (2016) City logistics: challenges and opportunities. Transp Sci 50:579–590

Şahin G, Süral H (2007) A review of hierarchical facility location models. Comput Oper Res 34:2310–2331

Schilling D, Elzinga DJ, Cohon J, Church RL, ReVelle CS (1979) The team/fleet models for simultaneous facility and equipment siting. Transp Sci 13:163–175

Schultz CP (1970) The logic of health care facility planning. Socio Econ Plan Sci 4:383–393

Serra D, ReVelle CS (1993) The *pq*-median problem: location and districting of hierarchical facilities. Locat Sci 1:299–312

Shaw SL (1993) Hub structures of major US passenger airlines. J Transp Geogr 1:47–58

Smith HK, Harper PR, Potts CN (2013) Bicriteria efficiency/equity hierarchical location models for public service application. J Oper Res Soc 64:500–512

Smith H, Cakebread D, Battarra M, Shelbourne B, Cassim N, Coetzee L (2017) Location of a hierarchy of HIV/AIDS test laboratories in an inbound hub network: case study in South Africa. J Oper Res Soc 68:1068–1081

Tcha D-W, Lee B (1984) A branch-and-bound algorithm for the multi-level uncapacitated facility location problem. Eur J Oper Res 18:35–43

Teixeira JC, Bigotte JF, Repolho HM, Antunes AP (2019) Location of courts of justice: the making of the new judiciary map of Portugal. Eur J Oper Res 272:608–620

Tien JM, El-Tell K, Simons, GR (1983) Improved formulations to the hierarchical health facility location-allocation problem. IEEE Trans Syst Man Cybern 13:1128–1132

Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19:1363–1373

Weaver JR, Church RL (1991) The nested hierarchical median facility location model. INFOR Inf Sys Oper Res 29:100–115

Wu TH, Kolar DJ, Cardwell RH (1988) Survivable network architectures for broad-band fiber optic networks: model and performance comparison. J Lightwave Technol 6:1698–1709

Yaman H (2009) The hierarchical hub median problem with single assignment. Transp Res B-Method 43:643–658

Yaman H, Elloumi S (2012) Star *p*-hub center problem and star *p*-hub median problem with bounded path lengths. Comput Oper Res 39:2725–2732

Yaman H, Karasan OE, Kara BY (2012) Release time scheduling and hub location for next-day delivery. Oper Res 60:906–917

Yasenovskiy VS, Hodgson MJ (2007) Hierarchical location-allocation with spatial choice interaction modeling. Ann Assoc Am Geogr 97:496–511

Zanjirani Farahani R, Hekmatfar M, Fahimnia B, Kazemzadeh N (2014) Hierarchical facility location problem: models, classifications, techniques, and applications. Comput Ind Eng 68:104–117

Zhang J (2006) Approximating the two-level facility location problem via a quasi-greedy approach. Math Program 108:159–176

Zhong W, Juan Z, Zong F, Su H (2018) Hierarchical hub location model and hybrid algorithm for integration of urban and rural public transport. Int J Distrib Sens Netw 14:1–14