Hub Network Design Problems with Profits

Armaghan Alibeyg[a], Ivan Contreras[a], Elena Fernández[b,c]

[a]*Concordia University and Interuniversity Research Centre on Enterprise Networks,
Logistics and Transportation (CIRRELT), Montreal, Canada H3G 1M8*
[b]*Statistics and Operations Research Department, Universitat Politècnica de Catalunya,
Barcelona, Spain*
[c]*Barcelona Graduate School of Mathematics*

Abstract

This paper presents a class of hub network design problems with profit-oriented objectives. Potential applications arise in the design of air and ground transportation networks, where companies need to jointly determine the location of hub facilities as well as the design of the hub network. For this, strategic network design decisions must be integrated within the decision-making process. Such decisions may include the selection of the origin/destination nodes that will be served as well as the activation of different types of edges. This class of problems considers the simultaneous optimization of the collected profit, the setup cost of the hub network and the transportation cost. Several alternative models that can be used in a variety of situations are proposed and analyzed. For each model an integer programming formulation is presented and computationally tested in terms of both, the structure of its solution network and the difficulty for solving it with a commercial solver.

Keywords: hub network design; hub location; profits; discrete location.

1. Introduction

Hub-and-spoke networks are frequently employed in transportation and telecommunication systems to efficiently route commodities between many

---

Email addresses: `a_alibey@encs.concordia.ca` (Armaghan Alibeyg),
`icontrer@encs.concordia.ca` (Ivan Contreras), `e.fernandez@upc.edu` (Elena
Fernández)

origins and destinations. One of the key features of these networks is that direct connections between origin/destination (O/D) pairs can be replaced by fewer, indirect but privileged connections by using transshipment, consolidation, or sorting points, called *hub facilities*. This reduces the total setup cost at the expense of increasing some individual transportation costs. Overall transportation costs may also decrease due to the bundling or consolidating of flows through inter-hub arcs.

Hub Location Problems (HLPs) deal with joint location and network design decisions so as to optimize a cost-based (or service-based) objective. The location decision focuses on the selection of a set of nodes to place hub facilities, whereas the network design decisions deal with the selection of the links to connect origins and destinations, possibly via hubs, as well as the routing of commodities through the network. Typically, HLPs assume that hubs must be located at the nodes of a given network, distances satisfy the triangle inequality, and there is a constant discount factor on the transportation costs of the arcs connecting hubs. In addition, classical HLPs impose that all flows are routed via the selected hubs and ignore all arc setup costs. Such problems have optimal solutions where an arc exists connecting each pair of hubs, so optimal routing paths consist of at most three arcs, two arcs connecting non-hub nodes and hub nodes, plus one intermediate arc connecting two hub nodes. This optimality condition implies that the network design decisions are mainly determined by the allocation of non-hub nodes to hubs (see, Contreras, 2015), and has been extensively exploited to develop formulations and solution algorithms for solving these classical HLPs.

Hub Arc Location Problems (HALPs) no longer assume that the above optimality condition holds, and incorporate explicit hub arc selection decisions. HALPs, in which in which a cardinality constraint on the number of opened hub arcs is considered, were introduced in Campbell et al. (2005). HALPs that incorporate setup costs for the hub nodes and hubs arcs were studied in Contreras and Fernández (2014) and Gelareh et al. (2015). Other HALPs impose particular topological structures, such as tree-star (Contreras et al., 2010), star-star (Labbé and Yaman, 2008), ring-star (Contreras et al., 2016), and hub lines (Martins de Sá et al., 015a,b).

In most hub location applications arising in the design of distribution and transportation systems, a profit is obtained for serving (i.e. routing) the demand of a given commodity. Capturing such profit may incur not only a routing cost but also additional setup costs, as the O/D nodes of the commodity may require the a priori installation of transport infrastructure.

Classical HLPs and HALPs, however, ignore such profits and associated setup costs, as reflected by the requirement that the demand of every commodity must be served. Indeed, the overall profit obtained when all the commodities must be served is constant, and it does not affect the optimization of the distribution system. Broadly speaking, this requirement expresses the implicit hypothesis that the overall costs of solution networks will be compensated by the overall profits. Of course, such hypothesis does not necessarily hold, and incorporating decisions on the O/D nodes that should be served and their associated commodities may have important implications in the strategic and operational costs.

In this paper we study a new class of problems in hub location denoted as *Hub Network Design Problems with Profits* (HNDPPs). HNLPPs integrate within the decision-making process additional strategic decisions on the nodes and the commodities that have to be served and consider a profit-oriented objective which measures the tradeoff between the profit of the commodities that are served and the overall network design and transportation costs. Broadly speaking, HNDPPs focus on the following strategic decisions: *i)* where to locate the hubs; *ii)* what edges to activate and of what type; and, *iii)* what commodities to serve (this also dictates the nodes to activate). As usual, the operational decisions determine how to route the commodities that are selected to be served. HNDPPs generalize HLPs and HALPs as they incorporate one additional level to the decision-making process. To the best of our knowledge, hub location models incorporating explicit decisions on the nodes to be served have not yet been addressed in the literature.

Potential transportation applications of HNDPPs arise in the airline and ground transportation industries. As an example, in the case of airline companies network planners have to design their transportation network when they are first entering into the market, or may have to modify already established hub-and-spoke networks through alliances, merges and acquisitions of companies. The involved decisions are to determine the cities that will be part of their network, i.e. what cities they will provide service to (served nodes) and what O/D flights to activate (served commodities) in order to offer air travel services to passengers (served demand) between city pairs. Additional decisions focus on the location of their main airports (hub facilities) and on the selection of the legs used for connecting regional airports (served nodes) with hub airports and for connecting some hub airports between them. Finally, the transportation of passengers using one or more O/D paths on their established network. The objective is to find an optimal hub

network structure that maximizes the total net profit for providing air travel services to a set of O/D flights while taking into account the (re)design cost of the network. Depending on the regulations or the company service policy, passenger air travel services could be provided: *i)* only to city pairs that are profitable, *ii)* between all city pairs that are served by the company, or *iii)* to a percentage of them (private companies with service commitment or with market penetration policies).

The main contribution of this paper is to introduce the foundations of HNDPPs and to propose alternative models of increasing complexity, which incorporate additional features. We start with a pure profit-driven model, and progressively present variations which consider alternative constraints and/or additional decisions, which, in turn, may imply additional costs. As alternative constraints we consider the possibility of forcing to serve a commodity whenever its two end-nodes are activated. As for decisions we consider the activation, with associated setup costs, of two additional types of edges. Access edges allow non-hub nodes to be connected to a hub whereas bridge edges allow connecting hub nodes without using a discount factor. We finally consider two more general models that allow serving the existing demand at different levels. The second of such models allows, in addition, to activate at different levels the various elements of solution networks. This results in a capacitated HNDPP of notable difficulty. A summary of the proposed models and their main features is given in Table 1.

Mathematical programming formulations for these models are presented and computationally tested in terms of: the structure of the solution networks it produces, its sensitivity to the input parameters, its relation to the other models, and its difficulty for being optimally solved with a commercial solver. A companion paper (Alibeyg et al., 2016) presents an exact algorithm for the pure profit-driven HNDPP and provides extensive results of computational experiments and analyses.

The remainder of the paper is organized as follows. Section 2 reviews the most relevant literature related to HNDPPs. Section 3 presents the formal definition and modeling assumptions of a primary HNDPP. It also presents a mathematical programming formulation for the problem and some variants of it. Sections 4 and 5 provide more realisitc and complex extensions of HNDPPs. Section 6 describes the computational experiments we have run. The results produced by each model are presented and analyzed. The results of the different models are compared among them. The paper ends in Section 7 with some comments and conclusions.

Main features	PO_1	PO_2	PND	POM_1	POM_2
Locational decisions					
- Hub nodes	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
- Served nodes	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
- Multiple capacity/operational levels					$\checkmark$
Link activation decisions					
- Hub edges	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
- Access edges			$\checkmark$		
- Bridge edges			$\checkmark$		
Operational decisions					
- Single demand level	$\checkmark$	$\checkmark$	$\checkmark$		
- Multiple demand levels				$\checkmark$	$\checkmark$
Demand service type					
- Only to profitable pairs of served nodes	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$
- Between all pairs of served nodes		$\checkmark$			
Objective					
- Pure profit-oriented	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$
- Profit-oriented with service commitments		$\checkmark$			

Table 1: Summary of considered HNDPPs.

2. Literature Review

HNDPPs are related to two families of HLPs: Maximal Hub Covering Problems (MHCPs), and Competitive Hub Location Problems (CHLPs). MHCPs impose that commodities between O/D pairs have to be delivered within a time limit (service level). It is implicitly assumed that a commodity is served whenever its O/D nodes are within a predefined radius of some hub node. Because MHCPs restrict the length of the arcs of O/D paths to a given coverage radius, applications of these problems frequently arise in the design of telecommunication networks, where the signal deterioration must be taken into account (Campbell and O'Kelly, 2012). Campbell (1994) introduces different MHCPs, which have also been studied and extended by other authors (see Alumur and Kara, 2008; Zanjirani Farahani et al., 2013). More recently, Hwang and Lee (2012) study the uncapacitated single allocation p-hub maximal covering problem, which maximizes the overall demand that can be covered by p facilities within a fixed coverage radius. Lowe and Sim (2012) studies a MHCP that considers jointly hubs setup costs and flow transportation costs, subject to covering constraints. Similarly to HNDPPs,

in MHCPs some commodities may remain unserved. However, in contrast to HNDPPs, MHCPs implicitly assume that the setup cost for providing service to O/D nodes is zero and thus, they do not incorporate decisions on the nodes to be served.

While most HLPs are concerned with the design of the hub network of a single firm, CHLPs consider an environment in which several firms exist in a market and compete to provide service to customers. In CHLPs each commodity chooses the competing firm that will serve its demand, based on several criteria such as travel time or service cost. The usual objective in CHLPs is to maximize the market share of some firm. Marianov et al. (1999) introduce CHLPs with two competitors in which the follower looks for the best location for a set of hubs so as to maximize the captured demand. The first model assumes that a commodity demand will be fully captured if its routing cost does not exceed the current competitor's cost. A more realistic model is also considered, in which the fraction of the commodity demand that is captured is modeled using a stepwise linear function, which is used for the comparison with the competitor's routing costs. In both models, at most one path can be used to route commodities between each O/D pair. Eiselt and Marianov (2009) extend these models to allow using more than one path to connect an O/D pair. The fraction of commodity demand that is routed on a particular path is modeled with a gravity-like attraction function that depends on both, the routing cost and the travel time.

Gelareh et al. (2010) present a model arising in liner shipping networks, where a new liner service provider designs its network to maximize its market share, using a stepwise attraction function, which depends on service times and routing costs. Lüer-Villagra and Marianov (2013) study a competitive model in which a new company wants to enter the market of an existing company. The aim is to determine the prices to charge to served commodities so as to maximize the profit of the entering company, rather than its market share. Commodities preferences for the selected firm and service route are modeled using a logit model. OKelly et al. (2015) present a model with price-sensitive demands. It considers three different service levels for routing commodities between O/D pairs that use either two-hub O/D paths, one-hub O/D paths or direct connections. The model is formulated as an economic equilibrium problem that maximizes a nonlinear concave utility function minus the routing costs and the setup cost for the location of the hubs.

CHLPs have also been studied under a game theoretic framework, such

as Stackelberg hub location models, cooperative game theoretic models with alliances and mergers, and non-cooperative game theoretic models (see Adler and Smilowitz, 2007; Lin and Lee, 2010; Asgari et al., 2013; Sasaki et al., 2014; Contreras, 2015). We note that HNDPPs can be clearly differentiated from CHLPs, as the focus of the former is to optimize the individual decision related to one single firm rather than on competition aspects. To the best of our knowledge, besides Sasaki et al. (2014) all CHLPs previously studied focus on the location of hubs and do not explicitly consider hub arc selection decisions. Moreover, none of them consider other relevant decisions such as the activation of access/bridge arcs and servicing decisions for O/D nodes.

HNDPPs are also related to other network optimization problems, aiming at maximizing the captured demand or optimizing some profit-oriented objective. Examples of the former are the maximal covering location (Church and ReVelle, 1974) or the competitive facility location problem (Aboolian et al., 2007). Examples of the latter are prize-collecting versions of problems that do not consider locational decisions: traveling salesman (Feillet et al., 2005), vehicle routing (Aras et al., 2011), rural postman (Aráoz et al., 2009), and prize-collecting Steiner tree problems (Álvarez-Miranda et al., 2013).

The above mentioned prize-collecting problems share with HNDPPs a distinguishing feature: they generalize their corresponding *classical* version by incorporating one additional level to the strategic decision-making process, so as to determine the demand customers to be served. In its turn, such decisions induce additional network design decisions. Nevertheless, according to the classification of Contreras and Fernández (2012), all mentioned problems are user-facility demand. That is, service demand relates users (nodes) and service centers (facilities). Instead, HNDPPs are user-user demand, as service demand relates pairs of users among them (O/D nodes of commodities). To the best of our knowledge this is the first time a prize-collecting version of a user-user demand general network design problem is addressed.

3. Primary HNDPPs

In this section we first introduce a primary model where the core strategical and operational decisions in HNDPPs are identified. In this model, the main criterion that guides decisions is profit. It is applicable to private companies where their ultimate goal is to maximize their net profit, independently of any other consideration. Companies would only provide service to O/D nodes that increase their profit and, among all commodities asso-

ciated with served O/D nodes, only the profitable ones would be actually routed. We next describe possible variants in which: (i) external regulations could force companies to provide transportation services to any commodity where both its origin and destination nodes are served, even if this would reduce their profit, and (ii) market penetration policies are applied to ensure a predefined presence of a company in the market by forcing to serve a minimum number of customer demands, even if this is suboptimal from a profit perspective.

3.1. Formal Definition and Modeling Assumptions

We can formally define a HNDPP as follows. Let $G = (N, A)$ be a complete directed graph, where $N = \{1, 2, \ldots, n\}$ represents the set of nodes and A represents the set of arcs. Let $H \subset N$ be the set of potential hub locations. For each $i \in N$, $c_i \geq 0$ denotes the setup cost for serving node i and for each $i \in H$, $f_i \geq 0$ is the fixed setup cost for opening a hub at node i. If a node $i \in H$ is selected to be a hub, it is assumed that it will be possible to serve commodities originated (or with destination) at i without activating node i as a servicing node. That is, there is no need to incur in the setup cost c_i for serving node i if it becomes a hub. Each node will thus be exactly one of the following: a hub node, a served node, or an unserved node. For $(i, j) \in A$, $d_{ij} \geq 0$ denotes the distance or unit transportation cost between nodes i and j, which we assume to be symmetric, i.e., $d_{ij} = d_{ji}$, and to satisfy the triangle inequality. Let $A_H \subset A$ be the subset of arcs connecting two potential hub nodes, i.e. $A_H = \{(i, j) \in A \mid i, j \in H\}$, where it is possible that the two hubs coincide, i.e., $i = j$. We also consider the following two sets of undirected edges. The set of edges connecting two potential hubs, denoted as $E_H = \{\{i, j\} \mid i, j \in H\}$, and the set of edges where at least one endnode is a potential hub, denoted by $E_B = \{\{i, j\} \mid i \in N, j \in H, i \neq j\}$. Since N and H are different sets, so are E_H and E_B. Any edge $\{i, j\} \in E_H$ is indistinctively denoted as $\{j, i\}$. Instead, when we write $\{i, j\} \in E_B$, we assume that $i \in N$, $j \in H$. The elements of E_H are called *hub edges* whereas the elements of E_B are either *access* or *bridge edges* and will be discussed in detail later in this section. In the literature hub edges are often referred to as hub arcs. Throughout this paper we prefer to maintain the distinction between edges and arcs.

Edges in E_H can be activated incurring setup costs. We denote by $r_e \geq 0$ the setup cost of hub edge $e \in E_H$. When edges in E_H are activated, their associated arcs can be used for sending flows in any of their two directions. A

hub edge $e = \{i, j\} \in E_H$ has a per unit flow cost αd_{ij}. The parameter $\alpha, (0 \leq \alpha \leq 1)$ is used as a discount factor to provide reduced unit transportation costs on hub edges to represent economies of scale. Similarly to other HALPs, in this primary HNDPP variant edges in E_B are activated without incurring any setup cost. Also, no discount factor is applied to flows sent via edges in E_B. The per unit transportation cost of the two arcs associated with edge $e = \{i, j\} \in E_B$ is d_{ij}.

Let K denote the set of commodities where each $k \in K$ is defined as $(o(k), d(k), W_k)$, where $o(k), d(k) \in N$, respectively denote its origin and its destination, also referred to as its O/D pair, and W_k denotes its service demand, i.e., the amount of flow that must be routed from $o(k)$ to $d(k)$ if commodity k is served. The effect of serving commodity k is threefold. On the one hand it forces the activation of its O/D nodes $o(k)$ and $d(k)$. On the other hand, it produces a per unit revenue $R_k \geq 0$, which is independent of the path used to send the commodity demand W_k through the solution network. Finally, serving commodity k also incurs a transportation cost which depends on W_k and on the path that is used to route it from $o(k)$ to $d(k)$.

Similarly to most HLPs, we require that all O/D paths include at least one hub node. That is, the solution network contains no direct connections between two non-hub nodes. We assume that served nodes can be assigned to more than one hub node, i.e. multiple assignments. Moreover, we require solution networks to contain at most three edges in each O/D path. While this hypothesis is common in classical hub location models it may seem restrictive as compared to general network design models. Note, however, that this hypothesis is consistent with the potential applications that we mention, mainly air transportation where paths with three legs already correspond to two intermediate transfers. On the other hand, our models are profit-oriented so they include additional decisions on the commodities to be served, increasing their difficulty with respect to cost-oriented models.

For a given commodity k let $(o(k), i, j, d(k))$ denote the path connecting $o(k)$ and $d(k)$, which uses a collection edge between $o(k)$ and hub i, a transfer edge between hubs i and j, and a distribution edge between hub j and $d(k)$. When $i \neq j$, not only both i and j are hub nodes, but also the intermediate leg, $\{i, j\}$, must be a hub edge. Note that O/D paths of the form $(o(k), o(k), d(k), d(k))$, using just one hub edge, may arise only when both $o(k)$ and $d(k)$ are hub nodes. O/D paths with $i = j$ do not use any hub edge and consist solely of the collection and distribution legs, i.e. $(o(k), i, i, d(k))$

(origin-hub-destination) with $o(k) \neq i$ and $d(k) \neq i$.

Paths using at least two edges necessarily contain a collection or a distribution leg, i.e. some edge from E_B used with no discount factor. Such edges are of one of the following two classes: *access* or *bridge* edges. The only difference between an access and a bridge edge is that the former connects a non-hub node to a hub node whereas the latter connects two hub nodes. Even if a bridge edge connects two hub nodes, it differs from a hub edge in its setup cost and its per unit (non-discounted) routing cost. In the primary HNDPP we assume that no bridge edge will be used as intermediate transfer edge in a three-leg O/D path. The reader is addressed to Campbell et al. (2005) for further details and an extensive analysis on possibilities for O/D paths in hub location.

Taking into account the above mentioned assumptions and requirements on the structure of O/D paths, we define the per unit transportation cost for routing commodity k on the path $(o(k), i, j, d(k))$ as $F_{ijk} = (\chi d_{o(k)i} + \alpha d_{ij} + \delta d_{jd(k)})$, where the parameters χ and δ reflect weight factors for collection and distribution, respectively.

The HNDPP consists of (i) selecting a set of O/D nodes to be served; (ii) locating a set of hub facilities; (iii) activating a set of hub edges; (iv) selecting a set of commodities to be served, both of whose O/D nodes have been selected in (i); and, (v) determining the flows routing the selected commodities through the solution network, with the objective of maximizing the difference between the total revenue obtained for routing the demand of the served commodities minus the sum of the setup costs for the design of the network and the transportation costs for routing the commodities. The HNDPP is clearly NP-hard given that it has as a particular case the classical uncapacitated hub location problem with multiple assignments (UHLPMA), which is known to be NP-hard (Contreras and Fernández, 2014) Indeed, the HNDPP reduces to the UHLPMA when $c_i = 0$, for $i \in N$, $r_e = 0$, for $e \in E_H$, and $R_k = \sum_{i \in N} f_i + \max\{F_{ijk} : (i, j) \in A_H\}$, for $k \in K$.

3.2. An Integer Programming Formulation

For $i \in H$, we introduce binary location variables z_i equal to 1 if and only if a hub is located at node i, and for $i \in N$ we define binary variables s_i equal to 1 if and only if node i is served (i.e. activated as a non-hub node). For $e \in E_H$, we define y_e equal to 1 if and only if hub edge e is activated. Finally, for $k \in K$, $i, j \in H$, we define routing variables x_{ijk} equal to 1 if and only if commodity k is routed via arc $(i, j) \in A_H$. When $i = j$, $x_{iik} = 1$

indicates that commodity k is routed on the path $(o(k), i, d(k))$ visiting only hub i and thus, it is not routed via a hub edge. Using these sets of variables, the HNDPP can be formulated as follows:

$$(PO_1) \quad \text{maximize} \quad \sum_{k \in K} \sum_{(i,j) \in A_H} W_k(R_k - F_{ijk})x_{ijk} - \sum_{i \in H} f_i z_i - \sum_{i \in N} c_i s_i$$

$$- \sum_{e \in E_H} r_e y_e \tag{1}$$

$$\text{subject to} \quad \sum_{(i,j) \in A_H} x_{ijk} \le s_{o(k)} + z_{o(k)} \qquad k \in K \tag{2}$$

$$\sum_{(i,j) \in A_H} x_{ijk} \le s_{d(k)} + z_{d(k)} \qquad k \in K \tag{3}$$

$$\sum_{j \in H} x_{ijk} + \sum_{j \in H: i \ne j} x_{jik} \le z_i \qquad k \in K, i \in H \tag{4}$$

$$x_{ijk} + x_{jik} \le y_e \qquad k \in K, e = \{i,j\} \in E_H \tag{5}$$

$$x_{ijk} \ge 0 \qquad (i,j) \in A_H, k \in K \tag{6}$$

$$z_i \in \{0,1\} \qquad i \in H \tag{7}$$

$$s_i \in \{0,1\} \qquad i \in N \tag{8}$$

$$y_e \in \{0,1\} \qquad e \in E_H \tag{9}$$

The first term of the objective function represents the net profit for routing the commodities. The other terms represent the total setup costs of the hubs that are chosen, the non-hub nodes that are selected to be served, and the hub edges that are used. Constraints (2) and (3) impose that the O/D nodes of each routed commodity are activated, either as hub or served nodes. Constraints (4) prevent commodities from being routed via non-hub nodes, whereas constraints (5) activate hub edges. Finally, constraints (6) to (9) define the domain for the decision variables. As usual in uncapacitated hub location models, the above formulation does not require to explicitly impose the integrality of the routing variables x. Each commodity, if routed, will use exactly one path of the solution network. Also, given that $f_i \ge 0$ and $c_i \ge 0$, in any optimal solution to PO_1 a hub node will not be activated also as a served node, that is $s_i + z_i \le 1$ for each $i \in H$.

The above formulation has a very large number of variables and constraints. However, we can exploit the following properties to reduce its size.

Property 1. *There is an optimal solution to formulation* (1)–(9) *where* $x_{ijk} = 0$, *for every* $k \in K$ *and* $(i,j) \in A_H$, *with* $P_k - F_{ijk} \leq 0$.

Property 1 is a direct consequence of the modeling assumption that only profitable commodities will be routed. According to it, for each commodity $k \in K$ all the routing variables whose cost is not strictly smaller that its revenue R_k can be eliminated, as routing them will not increase the system overall profit.

Property 2. *Let* $Q = \{(z, s, y, x) \text{ that satisfy } (2) - (9)\}$ *be the domain of feasible solutions to* PO_1. *Then,*
    For every $k \in K$ *and* $e = \{i, j\} \in E_H$, $y_e \leq z_i$ *and* $y_e \leq z_j$.

Property 2 is a direct consequence of the fact that points (z, s, y, x) that satisfy constraints (4) and (5) ensure that $y_e = 1$ if its endnodes are hubs.

3.3. HNDPPs with Service Commitments

Model PO_1, is "flexible", in the sense that, among all commodities connecting served O/D nodes, only those that are actually profitable will be routed. In PO_1 it is thus possible that a commodity is not routed even if both its origin and destination are activated. It will only be served if routing it produces an additional profit. As mentioned, such a model can be applicable, for instance, in airline and ground transportation systems. Servicing a city does not mean that connections between this city and any other servicing city in a company's network will be necessarily offered. Only connections between such city and other cities that are profitable will be offered.

A more restrictive variant of PO_1, denoted as PO_2, arises in applications where either service commitments or external regulations impose the decision maker to serve any commodity whose O/D nodes are both activated, even if this would reduce the total profit. An important consequence of this requirement is that the solution networks to PO_2 will consist of a single connected component with no isolated hub nodes. PO_2 can be formulated by adding to PO_1 the following constraint:

$$s_{o(k)} + z_{o(k)} + s_{d(k)} + z_{d(k)} \leq \sum_{(i,j) \in A_H} x_{ijk} + 1 \quad k \in K. \tag{10}$$

Constraints (10) force commodities to be routed if their O/D nodes are both activated. We note that Property 1 no longer holds for PO_2 because

of the addition of constraints (10). As it will be shown in Section 6, this additional requirement considerably increases the complexity for optimally solving PO_2 with a general purpose solver.

Previous models can be easily adapted to deal with market penetration policies that ensure a predefined presence of a company in the market by servicing a minimum number of customers demands, even if this is suboptimal from a profit perspective. This can be attained by imposing to serve fraction of the total number of commodities or to route a fraction of the total demand, for example.

A constraint that imposes that a minimum fraction $0 \leq \beta_1 \leq 1$ of the total number of commodities are served is:

$$\sum_{k \in K} \sum_{(i,j) \in A_H} x_{ijk} \geq \beta_1 |K|, \tag{11}$$

Similarly, a constraint that imposes that the overall flow that is routed through the network is at least a fraction $100\beta_1$ of the overall demand $\sum_{k \in K} W_k$ is:

$$\sum_{k \in K} \sum_{(i,j) \in A_H} W_k x_{ijk} \geq \beta_2 \sum_{k \in K} W_k. \tag{12}$$

4. HNDPPs with Setup Costs on Access/Bridge Edges

We now introduce an extension of the primary HNDPPs presented in the previous section that incorporates link activation decisions on access and bridge edges. This makes more challenging not only the design of the hub network, but also the routing of commodities, which in turn makes the problem considerably more difficult to solve. We recall that E_B denotes the set of edges which can be activated as access or bridge edges. Let q_e denote the setup cost of edge $e \in E_B$.

Contrary to previous models where bridge edges can only appear in a three-leg O/D path as collection or distribution leg, in this new model, we now consider that a bridge edge can also appear as a transfer (or intermediate) leg. Consequently, the transportation cost of commodities that are routed through bridge arcs is different from the ones that use hub arcs since there is no discount factor on the bridge arcs. We thus define the per unit transportation cost for routing commodity k on the path $(o(k), i, j, d(k))$ (where arc $(i, j) \in E_B$ is a bridge arc) as $F'_{ijk} = (\chi d_{o(k)i} + d_{ij} + \delta d_{jd(k)})$.

Given that now it is possible to route commodities between hub nodes with a bridge edge on a three-leg O/D path, the model will have to select whether to activate a link as a hub edge if enough flow is being routed, so as to compensate the higher setup cost associated with a hub edge. Otherwise, it may active the link only as a bridge edge to get a smaller profit out of a set of commodities. Note that, because of the setup costs on bridge edges, optimal solutions may concatenate two consecutive bridge arcs, despite of the triangle inequality on routing costs.

We introduce two new sets of decision variables. For $e \in E_B$, t_e equal to 1 if and only if edge e is activated as an access or bridge edge. The new set of routing variables are used to differentiate the type of routing used for each commodity. In particular, we define x'_{ijk} equal to 1 if and only if commodity $k \in K$ is routed via bridge arc $(i,j) \in A_H$. the HNDPPs with setup costs on access/bridge edges can then be formulated as:

$$(PND) \quad \text{maximize} \quad \sum_{k \in K} \sum_{(i,j) \in A_H} W_k(P_k - F_{ijk})x_{ijk} + \sum_{k \in K} \sum_{(i,j) \in A_H} W_k(P_k - F'_{ijk})x'_{ijk}$$

$$- \sum_{i \in H} f_i z_i - \sum_{i \in N} c_i s_i - \sum_{e \in E_H} r_e y_e - \sum_{e \in E_B} q_e t_e \tag{13}$$

$$\text{subject to} \quad (4) - (9)$$

$$\sum_{(i,j) \in A_H} (x_{ijk} + x'_{ijk}) \le s_{o(k)} + z_{o(k)} \qquad k \in K \tag{14}$$

$$\sum_{(i,j) \in A_H} (x_{ijk} + x'_{ijk}) \le s_{d(k)} + z_{d(k)} \qquad k \in K \tag{15}$$

$$\sum_{j \in H} x'_{ijk} + \sum_{j \in H: i \ne j} x'_{ijk} \le z_i \qquad k \in K, i \in H \tag{16}$$

$$x'_{ijk} + x'_{jik} \le t_e \qquad k \in K, e = \{i,j\} \in E_B \tag{17}$$

$$\sum_{(i,j) \in A_H} (x_{ijk} + x'_{jik}) \le 1 \qquad k \in K \tag{18}$$

$$\sum_{j \in H} (x_{ijk} + x'_{jik}) \le t_{o(k)i} \qquad k \in K, (o(k),i) \in E_B \tag{19}$$

$$\sum_{i \in H} (x_{ijk} + x'_{jik}) \le t_{d(k)j} \qquad k \in K, (d(k),j) \in E_B \tag{20}$$

$$x'_{jik} \ge 0 \qquad (i,j) \in A_H, k \in K \tag{21}$$

$$t_e \in \{0,1\} \qquad e \in E_B. \tag{22}$$

The first term of the objective function represents the net profit of routing commodities through hub edges (with discount factor) while the second term is the net profit of routing commodities through bridge edges (without discount factor). The setup costs are the same as in PO_1 with additional setup costs of the access/bridge edges. Constraints (14), (15), and (16) are equivalent to constraints (2), (3), and (4). Constraints (17) activate bridge edges. Constraints (18) indicate that commodities can be routed using either hub edges or bridge edges. Constraints (19) and (20) impose that collection and distribution edges are activated (either as access or bridge edges).

5. HNDPPs with Multiple Demand Levels

In all previous models, it is assumed that if a commodity $k \in K$ is served then all its demand W_k will be routed and a revenue R_k will be received. However, in practice, for a given O/D pair the amount of demand W_k that is actually served can be related to the price set to provide such transportation service. That is, the amount of demand that requires service associated with a commodity k will depend on the per unit revenue R_k set by the company. Therefore, an additional operational decision can be considered, which is to select for each commodity $k \in K$ the revenue level that will allow the company to capture the optimal portion of the total demand W_k.

In this section we extend the primary model PO_1 to the case with multiple demand levels, and consider profit-oriented models where the above mentioned decisions are taken into account. The amount of price-dependent demand that is captured for each commodity, is usually modeled with various nonlinear continuous functions (see, for instance Lüer-Villagra and Marianov, 2013; OKelly et al., 2015). In this paper, to keep the model tractable while maintaining the rest of the decisions already considered, we employ a discrete approximation function that considers a set of possible values for commodities demands, each of them associated with a profit. We use L as the index set of demand and revenue levels for the commodities. For each commodity $k \in K$ and level $l \in L$, let now W_k^l denote the amount of demand that is routed if commodity k is served at level l, and R_k^l the corresponding revenue. All other data remains as in the previous models.

To formulate the first profit-oriented model with multiple demand levels, denoted as POM_1, for each $l \in L, i, j \in H$ and $k \in K$, we substitute the original set of routing variables x by an extended set of continuous routing variables, x_{ijk}^l, which denote the fraction of commodity k served at demand

level l that is routed via arc $(i,j) \in A_H$. The remaining decision variables are the same as in previous primary models, since we assume that they do not depend on demand levels. The POM_1 can be formulated as follows:

$$(POM_1) \text{ maximize} \quad \sum_{l \in L} \sum_{k \in K} \sum_{(i,j) \in A_H} W_k^l (P_k^l - F_{ijk}) x_{ijk}^l - \sum_{i \in H} f_i z_i$$

$$- \sum_{i \in N} c_i s_i - \sum_{e \in E_H} r_e y_e$$

$$\text{subject to} \quad (6) - (9)$$

$$\sum_{l \in L} \sum_{(i,j) \in A_H} x_{ijk}^l \leq s_{o(k)} + z_{o(k)} \qquad k \in K \qquad (23)$$

$$\sum_{l \in L} \sum_{(i,j) \in A_H} x_{ijk}^l \leq s_{d(k)} + z_{d(k)} \qquad k \in K \qquad (24)$$

$$\sum_{l \in L} \sum_{j \in H} x_{ijk}^l + \sum_{l \in L} \sum_{j \in H : i \neq j} x_{jik}^l \leq z_i \quad k \in K, i \in H \quad (25)$$

$$x_{ijk}^l + x_{jik}^l \leq y_e \quad k \in K, e = \{i,j\} \in E, l \in L \quad (26)$$

$$x_{ijk}^l \geq 0 \qquad\qquad (i,j) \in A_H, k \in K, l \in L. \quad (27)$$

The first term of the objective function represents the net profit for routing commodities at their different demand levels. The other terms are as in previous models. Constraints (23)-(27) are the analog to (2)-(6) taking into account the possible demand levels of the commodities.

Given that POM_1 does not consider any capacity constraints on the hubs or edges, it has a very useful property which can be exploited to considerably reduce the size of the above formulation. In particular, it can be shown that there is always an optimal solution to POM_1 in which for each served commodity, exactly one demand level and one path are selected. Moreover, for each commodity $k \in K$, its optimal demand level can be identified a priori. This observation is formalized in the following result.

Proposition 1. *For each $k \in K$, let $\bar{l}_k \in \arg\max_{l \in L}\{W_k^l P_k^l\}$. Then,*

1. *There is an optimal solution to POM_1 where $x_{ijk}^l = 0$, for $l \neq \bar{l}_k$, $(i,j) \in A_H$.*
2. *An optimal solution to POM_1 can found by solving PO_1 with $W_k = W_k^{\bar{l}_k}$ and $R_k = R_k^{\bar{l}_k}$, for each $k \in K$.*

Proposition 1 is a direct consequence of the fact that in POM_1 we assume that the demand levels of the commodities have no effect on of the setup costs of the network design decisions, particularly, on the setup costs of the hubs and served nodes.

Instead, in the model that we present next, denoted as POM_2, we assume that hubs and served nodes can be activated at different operation levels, incurring setup costs, which depend on the amount of flow that is processed at the nodes. That is, POM_2 is a capacitated model which considers multiple capacity levels to limit the maximum flow processed at a hub or served node. To this end, we denote as T the index set of operation levels for the hubs and for the served nodes (for ease of notation and without loss of generality we assume they are the same). For each potential hub $i \in H$ and operation level $t \in T$, let f_i^t denote the setup cost for hub i with operation level t, which allows serving a maximum amount of flow φ_i^t. Similarly, for each $i \in N$ and $t \in T$, let c_i^t denote the setup cost for serving node i with operation level t, which allows serving a maximum amount of flow ρ_i^t. We now extend the set of decision variables for the hubs and served nodes to the following. For each $i \in H$ and $t \in T$, variable z_i^t takes the value 1 if and only if a hub is located at node i with operation level t. For $i \in N$ and $t \in T$, variable s_i^t is equal to 1 if and only if node i is served with operation level t. POM_2 can be formulated as follows:

$$(POM_2) \text{ maximize} \quad \sum_{l \in L} \sum_{k \in K} \sum_{(i,j) \in A_H} W_k^l (P_k^l - F_{ijk}) x_{ijk}^l - \sum_{i \in H} \sum_{t \in T} f_i^t z_i^t$$

$$- \sum_{i \in N} \sum_{t \in T} c_i^t s_i^t - \sum_{e \in E_H} r_e y_e$$

$$\text{subject to} \quad (6) - (9), (26) - (27)$$

$$\sum_{l \in L} \sum_{(i,j) \in A_H} x_{ijk}^l \leq \sum_{t \in T} (s_{o(k)}^t + z_{o(k)}^t) \qquad k \in K \qquad (28)$$

$$\sum_{l \in L} \sum_{(i,j) \in A_H} x_{ijk}^l \leq \sum_{t \in T} \left(s_{d(k)}^t + z_{d(k)}^t \right) \qquad k \in K \qquad (29)$$

$$\sum_{l \in L} \left(\sum_{j \in H} x_{ijk}^l + \sum_{j \in H: i \neq j} x_{jik}^l \right) \leq \sum_{t \in T} z_i^t \quad k \in K, i \in H \ (30)$$

$$\sum_{t \in T} s_i^t + \sum_{t \in T} z_i^t \leq 1 \qquad\qquad\qquad i \in H \qquad (31)$$

$$\sum_{k \in K} \sum_{l \in L} W_k^l \left(\sum_{j \in H} x_{ijk}^l + \sum_{l \in L} \sum_{j \in H: i \neq j} x_{jik}^l \right) \leq$$

$$\sum_{t \in T} \varphi_i^t z_i^t \qquad \qquad i \in H \qquad \qquad (32)$$

$$\sum_{(i,j) \in A} \sum_{l \in L} \left(\sum_{k \in K: o(k)=h} W_k^l x_{ijk}^l + \sum_{k \in K: d(k)=h} W_k^l x_{ijk}^l \right) \leq$$

$$\sum_{t \in T} \rho_h^t s_h^t + M \sum_{t \in T} z_h^t \qquad h \in N \qquad (33)$$

$$z_i^t, s_i^t \in \{0,1\} \qquad \qquad i \in N, t \in T. \qquad (34)$$

The objective function and constraints (28)-(31) have a similar interpretation to those of PO_1. Constraints (32) guarantee that the service level at which a hub is opened allows to serve all the incoming and outgoing flow that is routed through it. Constraints (33) have a similar interpretation, with respect to the served nodes. They state that the total incoming and outgoing flow at a served node must not exceed its installed operational capacity. The last term $M \sum_{t \in T} z_m^t$ on the right hand side of the constraints is used to deactivate the constraint in case node h becomes a hub node, where M stands for a sufficiently large constant.

Of course more general models could be considered where different operation levels and associated setup costs are considered also for all edges. This will indeed increase further the complexity of the models, although the modeling techniques will be quite similar to the ones we have used so far. We close this section by noting that Property 2 holds for all the considered models.

6. Computational Experiments

In this section we describe the computational experiments we have run in order to analyze the performance and various aspects of the HNDPPs we have introduced in Sections 3 and 4. We give numerical results that allow quantifying the quality of the formulations we have presented and comparing the computational difficulty of the different HNDPPs models. We also provide insight on the tradeoff of the decisions involved in our models by analyzing their optimal network structures and by evaluating the effect of the

different parameters on the characteristics of the optimal solutions obtained with each of the considered models.

This section is structured in several parts. First we describe the computational environment and the set of benchmark instances we have used. In Sections 6.1 to 6.3 we respectively give numerical results to analyze the computational performance and limitations of the formulations for the primary models PO_1 and PO_2, model PND that incorporates decisions on bridge arcs, and model POM_2, which allows multiple service levels. Section 6.4 focuses on decision-making aspects and managerial insight by analyzing the structure of the solution networks produced by the different models and their sensitivity with respect to some of the parameters.

All experiments were run on an HP station with an Intel Xeon CPU E3-1240V2 processor at 3.40 GHz and 24 GB of RAM under Windows 7 environment. All formulations were coded in C and solved using the callback library of CPLEX 12.6.3. We use a traditional (deterministic) branch-and-bound solution algorithm with all CPLEX parameters set to their default values. In all experiments the maximum computing time was set to 86,000 seconds (one day).

The benchmark instances we have used for the experiments are the well-known CAB data set of the US Civil Aeronautics Board, with additional data that we generated for the missing information. These instances were obtained from the website *(http://www.researchgate.net/publication/269396247_cab100_mok)*. The data in the CAB set refers to 100 cities in the US. It provides Euclidean distances between cities, d_{ij}, and the values of the service demand between each pair of cities, W_k, where $o(k) \neq d(k)$. We have considered instances with $n \in \{15, 20, 25, 30, 35, 40, 45, 50, 60, 70\}$ and $\alpha \in \{0.2, 0.5, 0.8\}$. The largest 70 nodes instances have only been used with the primary formulation PO_1. Since CAB instances do not provide the setup costs for opening facilities, we use as the setup cost of opening hubs, i.e. f_i, generated by de Camargo et al. (2008). The setup costs c_i, $i \in N$, for served nodes are $c_i = \nu f_i$, where $\nu = 0.1$ unless otherwise stated. The setup costs r_e, $e = \{i, j\} \in E_H$, for activating hub edges are $r_e = \tau(f_i + f_j)/2$, where $\tau \in \{0.3, 0.6, 0.4\}$ is a parameter used to model the increase (decrease) in setup costs on the hub edges when considering smaller (larger) discount factors α. The setup costs q_e, $e = \{i, j\} \in H_B$, for activating access/bridge edges are set to $q_e = \sigma(f_i + f_j)/2$, where $\sigma = 0.01$ unless otherwise stated. The revenues R_k, $k \in K$, for routing commodities are randomly generated as $R_k = \varphi \sum_{(i,j) \in A_H} F_{ijk}/|A_H|$, where φ is a continu-

ous random variable following a uniform distribution $\varphi \sim U\,[0.25, 0.35]$. The collection and distribution factors are $\chi = \delta = 1$.

6.1. Numerical Results for Primary HNDPPs

Our first series of experiments was oriented to study the computational performance of the primary HNDPPs, represented by formulations PO_1 and PO_2, whose numerical results are summarized in Tables 2 and 3.

Table 2 gives results on the computational effort needed to solve the primary profit-oriented models with a set of 24 instances with up to 70 nodes for PO_1 and the subset with the 21 instances with up to 60 nodes in the case of PO_2. The first two columns give information on the instances: α, the discount factor on hub edges, and n, the number of nodes. The first of the two 4-columns blocks corresponds to PO_1, whereas the second one corresponds to PO_2. Each column within each block gives information about the performance of the solution algorithm and its associated bounds for the corresponding model. *% LP GAP* show percentage gaps between the values of the linear programming (LP) relaxations and optimal values, computed as $100(v_{LP} - v^*)/v^*$, where v^* and v_{LP} denote the optimal and LP values, respectively. *Optimal value* give optimal solution values (v^*), *Time(sec)* the computing times (in seconds) needed to optimally solve each instance, and *Nodes* the number of nodes explored by CPLEX in the enumeration tree.

The results of Table 2 show that CPLEX can solve to optimality all considered PO_1 instances with up to 70 nodes. The computing times largely depend not only on the sizes of the instances but also on the discount factor α. While for $\alpha = 0.8$ all instances were solved in less than 5 minutes, the largest 70 node instance required almost three hours of computing time for the smallest discount factor $\alpha = 0.2$, which still can be considered small for an instance of that size. These small computing times are attributed to the effectiveness of Property 1 for eliminating a large number of x_{ijk} variables and constraints (5). In particular, for $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 0.8$, respectively, the average percentage of eliminated variables is 94% , 97% and 98%, whereas the average percentage of eliminated constraints (5) is 87%, 94%, and 96%. Note that all 24 considered instances were optimally solved at the root node, as the solutions to their LP relaxations were already optimal. From this point of view, PO_1 has a performance similar to other traditional hub location models without capacity constraints, that very often have integer LP solutions (Hamacher et al., 2004).

		PO_1				PO_2			
α	n	% LP gap	Time (sec)	Nodes	Optimal value	% LP gap	Time (sec)	Nodes	Optimal value
0.2	25	0.00	3.00	0	4940995.75	0.00	25.15	0	4794995.92
	30	0.00	9.64	0	5475562.01	0.00	130.19	0	5286060.60
	35	0.00	30.05	0	6219389.73	0.00	386.81	0	6007414.96
	40	0.00	126.12	0	5992658.91	0.00	1162.89	0	5808861.17
	45	0.00	249.46	0	6171839.75	0.00	2364.55	0	5963158.91
	50	0.00	513.20	0	6471414.64	0.00	5557.70	0	6255892.56
	60	0.00	2370.97	0	7416226.43	0.00	37065.80	0	7153438.48
	70	0.00	10460.44	0	7795411.69				
0.5	25	0.00	1.60	0	3315967.83	0.00	10.24	0	2926230.46
	30	0.00	4.55	0	3624158.73	0.00	31.81	0	3201896.50
	35	0.00	10.16	0	4048946.70	0.00	109.04	0	3488014.39
	40	0.00	21.53	0	3909497.15	0.00	216.29	0	3405507.62
	45	0.00	42.17	0	3894805.98	0.00	566.33	0	3316412.33
	50	0.00	75.90	0	4219096.87	0.00	1364.90	0	3522264.30
	60	0.00	309.06	0	4777497.83	0.00	8339.37	0	3990983.53
	70	0.00	1006.20	0	5006776.75				
0.8	25	0.00	1.36	0	2940862.05	0.00	7.06	0	2557728.67
	30	0.00	3.62	0	3125465.17	0.00	17.87	0	2701547.65
	35	0.00	7.89	0	3503413.58	0.00	42.68	0	2939186.01
	40	0.00	15.42	0	3354761.27	0.00	88.42	0	2879733.41
	45	0.00	27.38	0	3451620.79	0.00	160.15	0	2846768.07
	50	0.00	44.93	0	3679333.83	0.00	305.75	0	2991043.52
	60	0.00	121.70	0	4083218.91	0.00	805.18	0	3300799.91
	70	0.00	293.20	0	4276872.10				

Table 2: Computational experiments for PO_1 and PO_2.

The last four columns of Table 2, which summarize the results for model PO_2, allow us to quantify the effect of the constraints (10) on the difficulty for solving the basic models. Recall that these constraints force commodities to be routed if their O/D nodes are both activated. We can observe a notable increase of the computing times relative to those of PO_1, particularly for the instances with the smallest discount factor $\alpha = 0.2$ (observe the 10 hours of computing time that were needed to solve the largest such instance with $n = 60$). This is indeed due to the fact that Property 1 no longer holds

for PO_2 so it is not possible to eliminate *a priori* variables and constraints. Nevertheless, PO_2 has shown to be a tight formulation, in the sense that, similarly to PO_1, the LP relaxation of all the considered instances was already integer, so no additional enumeration was needed.

The information on the structure of optimal networks and on operational aspects of solution networks for PO_1 and PO_2 is summarized in Table 3, which contains one block with six columns for each formulation. Columns *Open n-H* and *Open H* respectively show the number of nodes activated as non-hub and as hubs, whereas *Hub edges* give the actual number of hub edges relative to its maximum possible value. Recall that the number of hub edges in a fully interconnected hub-level network is $\sum_{i \in H} z_i (\sum_{i \in H} z_i - 1)/2$. The last three columns in each block help analyzing the operational implications of the obtained solutions: *%Served nodes* indicate the percentage of nodes served (including both non-hub and hub nodes), *% Served O/D* show the percentage of commodities served in the solution network, computed as $100 \sum_{k \in K} \sum_{i,j \in H} x_{ijk}/|K|$, and *% Routed Flows* give the percentage of all the demand that is served, computed as $100 \sum_{k \in K} \sum_{i,j \in H} W_k x_{ijk} / \sum_{k \in K} W_k$.

All indicators point out the high influence of the discount factor α on the design of optimal networks for both PO_1 and PO_2. As could be expected, the value of α has an important effect on the number of hub edges in optimal networks, but its effect is also noticeable on the number of hubs opened and non-hub nodes activated. This indicates that, even if it is not explicit in the formulations, large values of the discount factor for hub edges have a *discouraging* effect on the number of nodes that are activated (as hubs or as non-hubs) in optimal networks.

In particular, for PO_1 the number of open hubs and activated non-hub nodes range in the intervals [3, 8] and [16, 62], respectively. For n fixed, both numbers decrease as α increases. For PO_2, the effect of constraints (10) on the number of activated non-hub nodes is evident for all values of α. In contrast, the effect of constraints (10) on the number of hubs opened in optimal networks is influenced by the value of α. For the smallest value $\alpha = 0.2$, this number is quite similar to that of PO_1, whereas when α increases, PO_2 produces optimal networks where the number of open hubs is smaller than in the case of PO_1.

For both models, the hub-level solution network is incomplete for all instances. Again the effect of the discount factor is relevant, as the sparsity of the hub-level solution networks clearly increases with the value of α. The reduction on the number of open hubs of PO_2 relative to PO_1 produces, in

		PO_1						PO_2				
		Open		Hub	%Served	%Served	%Routed	Open		Hub	%Served	%Served	%Routed
α	n	n-H	H	Edges	Nodes	O/D	Flows	n-H	H	Edges	Nodes	O/D	Flows
0.2	25	19	6	10/15	100.00	75.67	91.87	17	6	9/15	92.00	84.33	91.43
	30	24	6	9/15	100.00	78.51	91.48	19	5	6/10	80.00	63.45	78.91
	35	28	7	11/21	100.00	83.61	92.91	27	7	11/21	97.14	94.29	95.59
	40	33	7	11/21	100.00	80.64	91.77	31	8	12/28	97.50	95.00	95.86
	45	37	8	11/28	100.00	75.61	91.17	36	8	12/28	97.78	95.56	95.77
	50	42	8	12/28	100.00	78.12	93.00	40	8	12/28	96.00	92.08	94.92
	60	52	8	13/28	100.00	75.14	92.42	49	8	12/28	95.00	90.17	94.22
	70	62	8	13/28	100.00	75.61	92.10						
0.5	25	16	5	4/10	84.00	34.50	58.81	12	4	5/6	64.00	40.00	57.48
	30	21	5	4/10	86.67	36.21	59.30	15	4	5/6	63.33	39.31	56.94
	35	24	6	4/15	85.71	38.07	60.91	18	4	5/6	62.86	38.82	56.38
	40	28	6	4/15	85.00	37.24	59.05	21	4	5/6	62.50	38.46	56.35
	45	30	5	3/10	77.78	32.02	51.71	21	3	3/3	53.33	27.88	44.25
	50	34	6	4/15	80.00	32.94	57.79	26	4	4/6	60.00	35.51	54.04
	60	44	7	6/21	85.00	39.83	63.42	33	5	6/10	63.33	39.72	56.95
	70	50	7	5/21	81.43	35.76	61.20						
0.8	25	16	3	1/3	76.00	26.83	48.57	12	2	1/1	56.00	30.33	45.28
	30	20	4	1/6	80.00	30.80	50.28	14	3	1/3	56.67	31.26	45.52
	35	23	5	1/10	80.00	32.69	52.03	17	3	1/3	57.14	31.93	45.69
	40	27	5	1/10	80.00	32.95	50.49	20	3	1/3	57.50	32.44	45.33
	45	29	5	1/10	75.56	28.84	49.41	21	3	1/3	53.33	27.88	43.43
	50	31	5	1/10	72.00	26.04	48.42	22	3	1/3	50.00	24.49	42.12
	60	36	5	1/10	68.33	24.35	48.66	26	3	1/3	48.33	22.94	41.59
	70	41	5	1/10	65.71	22.42	47.91						

Table 3: Structure of optimal networks for PO_1 and PO_2.

its turn, an increase on the sparsity of the hub-level solution networks of PO_2, particularly for the instances with the highest value of $\alpha = 0.8$, whose optimal solutions always have just one hub edge.

Focusing on the operational aspects of solution networks, for PO_1 we can observe that the percentage of served nodes and served O/D pairs range between 65%-100% and 48%-98%, respectively. For instances of all sizes, both percentages clearly decrease as the value of α increases, although the decrease is more evident for the served O/D pairs, which for $\alpha = 0.8$ is below

33%, than for the served nodes, which for the same value of $\alpha = 0.8$ ranges in $65\% - 80\%$. The percentage of overall demand routed in optimal networks, ranges between 48%-93%, which clearly indicates that optimal networks tend to serve commodities with higher demand. The effect of α on these values is also clear and similar to that on the served nodes.

Despite the increase of the sparsity of the hub-level solution networks of PO_2, the effect of constraints (10) is not so evident on the operational indicators of its solution networks. While there is a slight decrease with respect to PO_1 on the percentage of served nodes, which ranges between 48%-98%, it is difficult to appreciate a decrease on the percentage of served O/D pairs, which ranges between 23%-95%. The percentage of overall demand routed in optimal networks, ranges in 41%-96%. While this percentage is higher for PO_2 than for PO_1 when $\alpha = 0.2$ (with the exception of the 30 nodes instance), it is smaller for PO_2 than for PO_1 when $\alpha = 0.8$. Taking into account that for each instance and value of α, the net profit obtained with PO_2 is always smaller than that of PO_1 (see the optimal values in the corresponding columns of Table 2), it seems clear that model PO_1 should be preferred to model PO_2 for larger values of α. Nevertheless for smaller values of α the comparison is not clear, as PO_2 produces solutions in which the percentage of served demand is higher than with PO_1.

6.2. Numerical Results for HNDPPs with Setup Costs on Access/Bridge Edges

Next we discuss the results we have obtained with formulation PND, which incorporates network design decisions on bridge and access arcs. For these experiments we have considered the subset of 21 instances with up to 60 nodes and values of $\alpha \in \{0.2, 0.5, 0.8\}$. The obtained results are summarized in Table 4, where columns *Access Edges* and *Bridge Edges* give the number of edges of each type in optimal solutions.

A first observation is that the computing times of PND are considerably higher than those of the most time consuming primary formulation, PO_2, particularly as the number of nodes of the instances increase. This is not surprising as PND has a larger number of both binary variables (those associated with the activation of access and bridge edges) and continuous variables (those associated to the flows routed via inter-hub bridge arcs). In any case, for PND we can again observe the influence of the discount factor α on the difficulty for solving the instances. For the largest value of $\alpha = 0.8$ the computing time for the largest 60 nodes instance is moderate, as it can be solved in less than 1.50 hours. Instead, the same instance becomes really

α	n	% LP gap	Time (sec)	Nodes	Optimal value	Open n-H	H	Access Edges	Bridge Edges	Hub Edges	Served Nodes	%Served O/D pairs	%Routed Flows
	25	2.09	15.77	0	4850184.96	19	6	28/114	1/15	10/15	100.00	74.67	91.32
	30	2.08	87.97	0	5362669.76	22	5	29/110	1/10	7/10	90.00	61.84	79.81
	35	2.01	325.68	0	6086281.06	28	7	41/196	1/21	11/21	100.00	76.89	91.02
0.2	40	3.31	1568.98	0	5831812.30	33	7	49/231	1/21	11/21	100.00	74.36	90.19
	45	3.20	4127.66	0	5969894.50	38	7	55/266	1/21	11/21	100.00	70.30	88.52
	50	2.86	9856.98	0	6239525.42	41	7	58/287	1/21	11/21	96.00	66.82	87.98
	60	2.80	66650.76	3	7105816.66	51	8	88/408	5/28	13/28	98.33	70.85	91.22
	25	9.62	12.48	0	3246177.24	16	5	22/80	1/10	4/10	84.00	34.33	58.61
	30	8.14	54.49	0	3530787.94	21	5	29/105	0/10	4/10	86.67	35.75	58.66
	35	9.64	158.94	4	3928183.15	24	6	37/144	1/15	4/15	85.71	37.23	60.21
0.5	40	9.73	421.83	0	3773879.34	26	6	42/156	2/15	4/15	80.00	34.04	56.92
	45	10.69	951.06	6	3752423.62	28	5	46/140	1/10	3/10	73.33	30.00	50.24
	50	8.85	1675.92	4	4051369.73	34	6	54/204	1/15	4/15	80.00	31.80	57.38
	60	10.79	21927.88	13	4521280.50	39	6	62/234	2/15	5/15	75.00	31.16	59.15
	25	14.91	6.94	0	2875551.32	16	3	23/48	0/3	1/3	76.00	26.83	48.57
	30	15.87	39.66	7	3011761.00	20	4	33/80	1/6	1/6	80.00	30.57	49.91
	35	16.73	154.96	26	3363665.77	23	5	41/115	2/10	1/10	80.00	32.27	51.72
0.8	40	17.03	244.24	2	3210437.26	25	5	46/125	2/10	1/10	75.00	30.00	48.53
	45	17.02	413.09	7	3283658.34	27	5	51/135	2/10	1/10	71.11	26.57	47.67
	50	15.78	624.69	2	3507899.84	30	5	54/150	2/10	1/10	70.00	24.16	47.54
	60	15.50	4450.06	8	3864127.25	36	5	70/180	3/10	1/10	68.33	24.18	48.63

Table 4: Computational experiments for PND.

challenging when $\alpha = 0.2$, as the computational effort needed to solve it rises to more than 18 hours.

From a computational point of view, a clear difference of PND with respect of PO_1 and PO_2 is that, for all instances and values of α, the LP relaxation of PND produced non-integral solutions with strictly positive percentage gaps. These gaps are quite small for $\alpha = 0.2$ (smaller 3.5%), and except for the largest instance with $n = 60$, they can be closed already at the root node by the CPLEX cuts added by default. As α increases the values of *% LP gap* get larger and range in 15%-17% for $\alpha = 0.8$. Still, the computational effort required to close such gaps is moderate.

Looking at the structure of optimal networks produced by PND we can observe that, similarly to the previous formulations, the hub-level solution networks are incomplete in all instances and their sparsity increases with the value of α. Furthermore, we can appreciate that adding decisions on access/bridge edges does not seem to have an important effect on the number of hub nodes that are opened, which is quite similar to that of PO_1. It can

be also observed that, for most of the instances, the number of hub edges in optimal solutions is the same as in PO_1 and decreases slightly in only three instances. There is usually a small number of bridge edges, which seems independent of the value of α. For the smaller value of $\alpha = 0.2$ the number of hub edges is always higher than that of bridge edges, although this relation tends to change as α increases. The explanation is clear: hub edges are activated only if the discount factor α produces enough reduction in the routing costs since their setup cost is higher than that of bridge edges; otherwise profitable commodities are routed via bridge edges.

When analyzing the operational indicators of solution networks, it can be seen that the percentage of served nodes and served O/D pairs ranges in 68%-100% and 24%-75%, respectively. These values are very similar to those of PO_1, although some decreases can be appreciated, mainly in the instances where the number of hub edges does not coincide. Something similar can be observed with the percentage of routed flows, which ranges in 48%-92%, and are always slightly smaller than those of PO_1 except for the instances where there is a reduction on the number of hub edges, where the decrease on the flows that are routed may reach 11%. Similarly to the previous models, the difference on the percentage of served O/D pairs and routed flows indicates that optimal networks tend to serve commodities with higher demand.

6.3. Numerical Results for HNDPPs with Multiple Demand Levels

We have run a last series of computational experiments to evaluate the more general model POM_2, in which hubs and served nodes can be activated at different operation levels, incurring setup costs, which depend on the amount of flow that is being processed at the nodes. Now we have only considered the instances with up to 35 nodes as the computing times of larger instances become prohibitive. Moreover, for larger instances, in most cases no feasible integer solution was known at termination.

For these experiments we have adapted the CAB instances used in the previous sections to incorporate demand and revenue levels for the commodities and multiple capacity levels for the hub facilities in the following way. For each $k \in K$, we set W_k^1 and P_k^1 to W_k and P_k, respectively. Data for the other levels are generated by decreasing demand and increasing revenue. That is, we defined $W_k^l = 0.3\, W_k^{l-1}$ and $1.2\, P_k^l = P_k^{l-1}$ for $l = 2, \ldots, |L|$. In addition, for each $i \in H$, we set $f_i^1 = f_i$ and $f_i^t = 0.9 f_i^{t-1}$, $t = 2, ..., |T|$. We generated in the same way different levels of setup costs for served nodes. That is, for each $i \in N$, $c_i^1 = c_i$ and $c_i^t = 0.9\, c_i^{t-1}$, $t = 2, ..., |T|$. We have

also generated different levels of capacities for the hub and served nodes. For each $i \in H$, we set $\varphi_i^1 = \lambda \sum_{i \in H} O_i / \sum_{i \in H} z_i^*$, where λ is a continuous random variable following a uniform distribution $\lambda \sim U[0.9, 1.1]$, and O_i is the total flow passing through hub i at the optimal solution of POM_1 (denoted as z^*). For other capacity levels of hub nodes, $\varphi_i^t = 0.7 \varphi_i^{t-1}$, $t = 2, ..., |T|$. Capacities of the served nodes are generated as a fraction of the capacities for the hubs, i.e. $\rho_i^t = \gamma \varphi_i^t$, $t = 1, ..., |T|$, and $\gamma = 0.5$. Finally, we have considered $|L| = |T| = 5$.

The obtained numerical results are summarized in Table 5, which highlights the computational difficulty of POM_2. Since optimality of the best-known solution could not be proven in all cases, column *%gap end* gives the percentage optimality gap at termination and *%LP gap* the percentage deviation of the LP bound with respect to the best-known solution at termination. The value of such solution is given in column *Best-known value*. A value 0.01 in column *%gap end* indicates that such value corresponds to an optimal solution.

α	n	% LP gap	% gap end	Time (sec)	Nodes	Best-known value	Best bound	Open n-H	Hub H	Edges	%Served Nodes	%Served O/D pairs	%Routed Flows
	25	3.82	0.00	17379.39	3306	3571879.69	3571537.82	16	9	4/36	100.00	96.83	68.11
0.2	30	4.45	2.90	86404.12	1435	3485278.93	3387018.15	19	11	3/55	100.00	97.47	65.13
	35	6.49	5.52	86407.77	495	3372157.52	3195736.83	23	12	2/66	100.00	86.64	56.42
	25	4.93	0.00	2367.17	1967	2947676.67	2947461.66	17	7	1/21	96.00	83.17	53.99
0.5	30	4.87	0.00	43392.63	13324	2893186.37	2892897.33	22	8	1/28	100.00	86.55	51.77
	35	5.91	0.74	86410.32	3655	2964770.23	2943064.70	24	10	1/45	97.14	82.27	50.71
	25	5.25	0.00	595.45	1981	2839814.46	2839554.07	18	6	0/15	96.00	81.17	48.84
0.8	30	5.50	0.00	2398.61	3489	2822441.80	2822176.98	22	7	0/21	96.67	80.34	49.55
	35	5.87	0.00	21198.95	15736	2936660.82	2936367.74	25	9	0/36	97.14	80.76	50.01

Table 5: Computational experiments for POM_2.

From the obtained results it can be seen that formulation POM_2 is quite tight, producing rather small % LP gaps at the root node, which do not exceed 6.5%, even for the instances that could not be solved to optimality. Unfortunately, these small gaps are very difficult to close, as indicated by the high number of nodes that are explored in the search trees and by the high computing times, which, for the small size instances, are affordable when $\alpha = 0.8$, but become prohibitive as α decreases or as the size of the instances increases.

Anyway, Table 5 allows to appreciate that in the optimal/best-known solution networks produced by POM_2 the total number of activated nodes

is roughly the same as in all previous models for instances of the same size and value of α, although the number of open hubs is slightly higher (so the number of activated non-hubs is slightly lower). At the hub-level network, however, we can observe very small values for the ratio of the number of hub edges relative to the number of hub nodes, resulting in highly sparse hub-level networks. As can be seen, only the instances with $\alpha = 0.2$ produced solution networks with more than one (but very few) hub edges, whereas the hub-level networks for the $\alpha = 0.5$ instances have just one hub edge, and no hub edge is activated in the solution networks to the instances with $\alpha = 0.8$.

In all cases, the percentages of served nodes are very high, and follow a similar trend as in previous models, with full node service for instances with $\alpha = 0.2$ and decreasing slightly when $\alpha = 0.5, 0.8$. Nevertheless, even for the instances with a higher value of $\alpha = 0.8$ this percentage is never below 96% which is considerably higher than the percentage of nodes served with the other models for the same value of α. Something similar happens with the percentage of served O/D pairs, which is never below 80%, independently of the value of α. On the contrary, the percentages of routed flows are noticeably smaller than in previous models. This is indeed a sign of the selective nature of POM_2 where activated nodes and hubs can operate at different service levels and demand flows partially routed. This can be better appreciated in Table 6 that shows the service levels of the solution networks of POM_2.

α	n	Open Hubs level l					Served Nodes level l					Routed Flows level l				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	25	7	0	1	0	1	2	2	5	1	6	224	75	97	95	105
0.2	30	8	1	1	0	1	1	4	6	3	5	331	127	159	160	94
	35	9	1	0	0	2	2	4	8	1	8	348	141	143	176	250
	25	3	2	1	0	1	2	2	4	3	6	170	61	72	78	126
0.5	30	5	1	1	0	1	3	2	5	4	8	242	119	103	117	188
	35	7	1	0	1	1	3	3	7	6	5	344	119	137	161	242
	25	2	2	1	0	1	1	4	4	3	6	154	68	64	86	127
0.8	30	4	2	0	0	1	3	2	6	3	8	228	89	99	117	184
	35	7	0	1	0	1	3	4	7	5	6	327	138	139	161	229

Table 6: Service levels for solution networks to POM_2.

As can be seen, most of the hub nodes are activated at the lowest service level, and this trend is more evident for the lowest value $\alpha = 0.2$. In contrast, served nodes tend to be activated at higher service levels. In particular, the percentage of served nodes activated at the highest service level ranges in

30% – 60%. Routed flows are served at all service levels, although higher frequencies correspond either to the lowest or highest service levels.

6.4. Comparison and Tradeoff of Proposed Models

Since all the models we have proposed are profit-oriented, a natural question that may arise is how they compare to traditional *cost-oriented* models. The comparison we make below aims at appreciating the advantage of integrating within the decision-making process additional strategic decisions on the nodes and the commodities that have to be served. In particular, in Figure 1 we compare the optimal networks obtained with some of the profit-oriented models we propose and with some classical cost-oriented counterparts for an instance with $n = 25$ and $\alpha = 0.6$. The profit-oriented models we have considered for this comparison are the primary HNDPP formulated via PO_1 and the HNDPP with setup costs on access/bridge edges, formulated with PND. The *cost-oriented* counterparts have been modeled as a hub arc location models that impose activating all demand nodes and routing all commodities, and aim at minimizing the total setup cost for the hub nodes and hub edges and the transportation cost for routing commodities (see, for instance, Contreras and Fernández, 2014). In Figure 1 the corresponding cost-oriented counterparts are denoted by PO_1-HALP and $PND-HALP$, respectively.

To make the comparison as fair as possible, we first solve formulations PO_1 and PND with $c_i = 0$, for $i \in N$, and $P_k = \sum_{i \in N} f_i + \sum_{e \in E_H} r_e + \max\{F_{ijk} : (i, j) \in A_H\}$, for $k \in K$. For PND we also set $q_e = 0$, for $e \in E$. The optimal PO_1 and PND solutions obtained with this data consist, in each case, of sets of served nodes, open hub nodes and hub edges, and routed commodities. These solutions are compared to the solutions to the $PO_1 - HALP$ and $PND - HALP$ stated on the graphs induced by the set of nodes served in the optimal solution to the corresponding profit-oriented problem.

Figures 1a and 1c show the optimal networks of PO_1 and PND, respectively, whereas Figures 1b and 1d show the optimal networks $PO_1 - HALP$ and $PND-HALP$, respectively. Triangles represent hubs, full circles served nodes, and empty circles unserved nodes. Black lines represent hub edges while gray lines represent access and bridge edges.

Given that cost-based hub models impose that all commodities are served, they imply a larger number of hub nodes and hub edges. As can be seen, in both cases the profit-oriented model which incorporates decisions on the
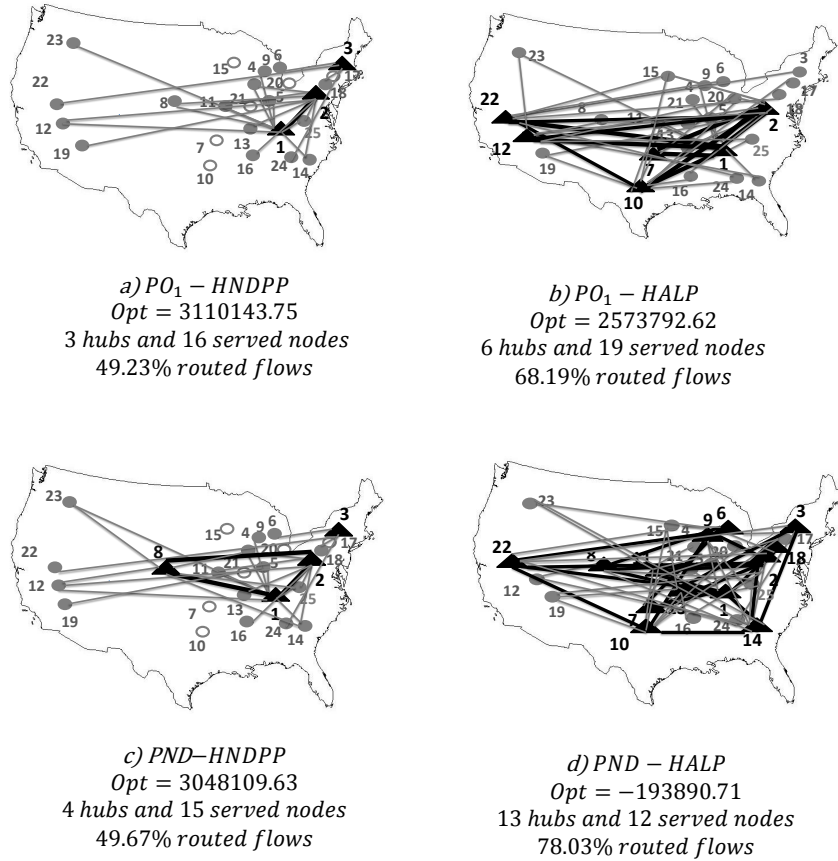
Figure 1: Optimal networks for PO_1 and PND with $n = 25$ and $\alpha = 0.6$.

nodes and demand that must be served produces a considerably better solution than the one obtained with the cost-oriented counterpart. This is particularly true in the case of the formulations that incorporate decisions on access/bridge edges, which produces a negative total profit. This creates an increase in the setup cost of the network and, as a result, solutions networks with a notable decrease in the total profits when compared to their profit-oriented counterpart. Figure 1 also allows to compare the optimal networks produced by the profit-oriented formulations PO_1 and PND, so we can analyze the effect of incorporating decisions on the use of access/bridge edges, inducing additional setup costs. As can be observed the difference on the number of served nodes and served commodities is very small. Neverthe-

less, the total profit obtained with the solution produced by PO_1 is about 2% higher than the one obtained with PND.

Taking into account that all our proposed models are profit-oriented ones, a more systematic alternative to the comparison among them is to analyze the profit each of them produces per O/D pair served and per unit of flow routed. Below we analyze the tradeoff among the different models. For this we use Figures 2 and 3, which respectively depict the profit per O/D pair served and the profit per unit of flow routed for the primary models, represented by formulations PO_1, PO_2, the model with access/bridge edge decisions, represented by PND, and the capacitated multiple level model, represented by POM_2. For a better visualization, each figure is separated in three parts, one for each tested value of α.
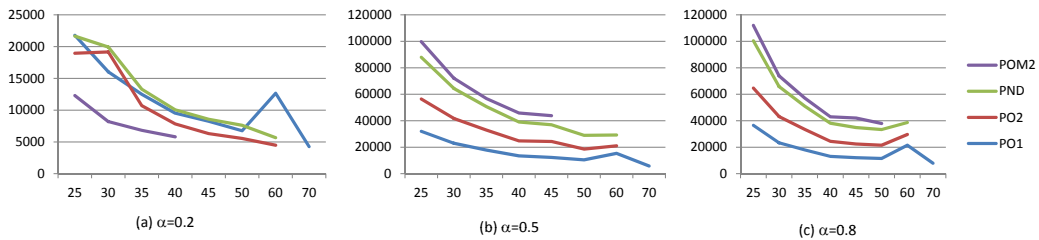

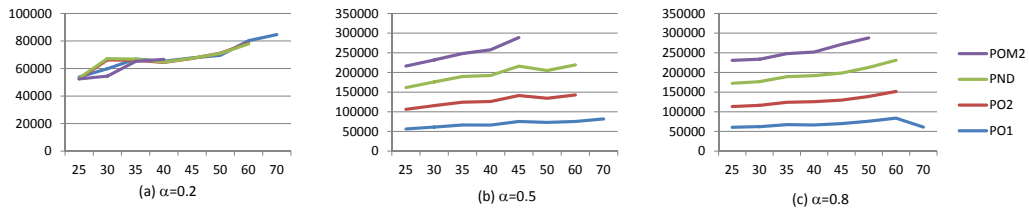Figure 2: Comparison of Models: Profit per served O/D pair.

Figure 3: Comparison of Models: Profit per routed unit of flow.

Both figures clearly illustrate the superiority of the capacitated multiple level model with respect to the other models when $\alpha = 0.5, 0.8$ both in terms of the profit per served O/D pair and per unit of routed flow. For these values of α, the *quality* of the models, measured in terms of their ability of producing solutions with a better tradeoff between their profit and the service level attained, is proportional to their sophistication. Thus the multiple level model is followed by the model with access/bridge edge decisions, which

outperforms the primary model that forces to serve any commodity whose O/D nodes are both activated, which, in turn, outperforms the purely profit-oriented primary HNDP. In contrast, the situation changes for the small discount factor $\alpha = 0.2$. On the one hand, Figure 3 shows that, in terms of the profit obtained per unit of flow routed all models are very similar (note that their lines nearly overlap). In contrast, in terms of the the profit obtained per O/D pair served, the capacitated multiple service model is outperformed by any of the other models, which are quite similar among them. In our opinion, both figures could be very useful to help a decision maker chose among the presented models, taking into account the potential context and priorities.

6.5. Sensitivity Analysis

We conclude this section with a sensitivity analysis of the presented models with respect of some of their input data. Figure 4 compares the optimal hub networks produced by primary formulation PO_1 for the CAB instance with $n = 25$ and $\alpha = 0.5$ when setting c_i as 0%, 15% and 40%, of the setup cost f_i.

a) $c_i = 0.00 \, f_i$
Opt: 8,811,009.83
6 hubs and 15 served nodes
54.38% routed flows

b) $c_i = 0.15 \, f_i$
Opt: 7,388,362.45
6 hubs and 10 served nodes
43.57% routed flows

c) $c_i = 0.40 \, f_i$
Opt: 5,636,245.08
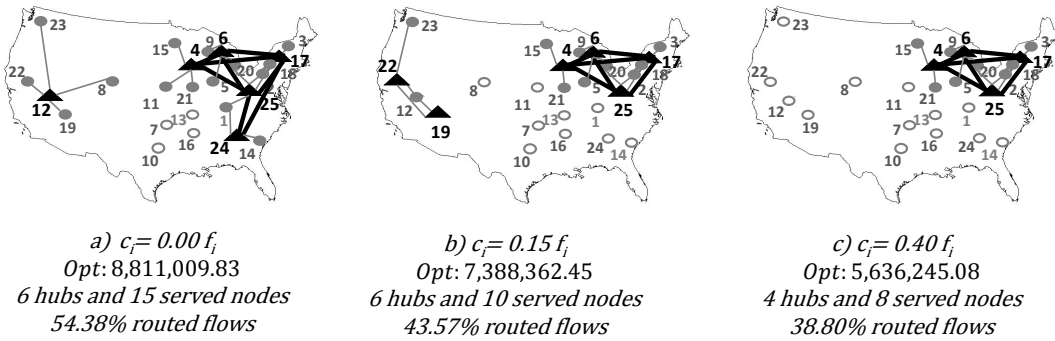4 hubs and 8 served nodes
38.80% routed flows

Figure 4: Optimal network for PO_1 with different setup costs c_i with $n = 25$ and $\alpha = 0.5$.

Figure (4)a depicts the optimal solution network with no setup costs for serviced nodes. It consists of two disconnected components with five interconnected hubs, one isolated hub, and 15 served nodes. Even if there are no setup costs for activating served nodes, four nodes remain unserved. Figures 4b and 4c show that, as could be expected, increasing the setup costs for serving nodes reduces the number of served nodes. In particular,

using setup costs $c_i = 0.15 f_i$ (Figure 4b), reduces to 10 the number of served nodes. Moreover, the topology of the hub-level network also changes, even if the number of hubs has not changed. The overall profit is reduced by 16.14%. When setup costs are further increased to $c_i = 0.40 f_i$ (Figure 4c), the optimal solution network consists of a single connected component with four fully interconnected hubs. Now the number of served nodes has decreased to eight and the total profit is reduced by 36.03% with respect to the case where $c_i = 0$.

Figure 5 allows to compare the effect of the discount factor α in solution networks. It shows the optimal networks produced by the primary formulation PO_1 for the CAB instance with $n = 25$ and three different values of the discount factor α. The optimal network for $\alpha = 0.2$ (Figure 5a) consists of a single connected component with six hubs, seven hub edges, and six unserved nodes. When increasing the discount factor to $\alpha = 0.5$ (Figure 5b), the solution network consists of two disconnected components but one hub node less and and nine unserved nodes. This causes a considerable reduction in both the number of served O/D pairs and routed flows. Figure 5c shows the solution network for the highest value $\alpha = 0.8$. Now the number of hub nodes has further decreased to three. Even if the number of served nodes remains the same as with $\alpha = 0.5$ there is a further decrease on the served O/D pairs and the total routed flow.
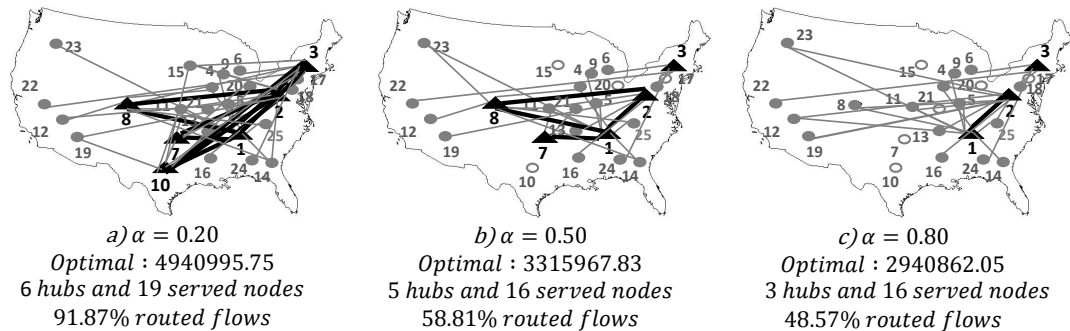

a) $\alpha = 0.20$
Optimal : 4940995.75
6 *hubs and* 19 *served nodes*
91.87% *routed flows*

b) $\alpha = 0.50$
Optimal : 3315967.83
5 *hubs and* 16 *served nodes*
58.81% *routed flows*

c) $\alpha = 0.80$
Optimal : 2940862.05
3 *hubs and* 16 *served nodes*
48.57% *routed flows*

Figure 5: Optimal network for PO_1 with different discount factors α with $n = 25$ and $\nu = 0.1$.

7. Conclusions

In this paper we introduced a class of hub network design problems with a profit-oriented objective. These problems integrate several locational and network design decisions such as the selection of origin/destination nodes, a set of commodities to serve, and a set of access, bridge and hub edges. They consider the simultaneous optimization of the collected profit, the setup costs of the hub network and the total transportation cost. We introduced the foundations of such problems and proposed alternative models of increasing complexity. We first studied a primary model which considers a purely profit-driven objective and discussed possible extensions to incorporate service commitments. We then introduced models where additional link activation decisions on access and bridge edges were considered. The last two of the considered models allowed serving demand at different levels. Each model was analyzed and a mathematical programming formulation was computationally tested using a general purpose solver. Given the inherent difficulty of the considered models, CPLEX was only able to solve small to medium-size problems. The companion paper of Alibeyg et al. (2016) presents and exact solution algorithm that is capable of solving more realistic, large-scale instances for the primary model. We are currently working on more challenging extensions of these problems in which O/D paths can use several hub and bridge arcs.

8. Acknowledgments

References

Aboolian, R., O. Berman, and D. Krass (2007). Competitive facility location and design problem. *European Journal of Operational Research 182*(1), 40–62.

Adler, N. and K. Smilowitz (2007). Hub-and-spoke network alliances and mergers: Price-location competition in the airline industry. *Transportation Research Part B 41*(4), 394–409.

Alibeyg, A., I. Contreras, and E. Fernández (2016). An exact algorithm for hub network design problems with profits. *in preparation*.

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

Álvarez-Miranda, E., I. Ljubić, and P. Toth (2013). Exact approaches for solving robust prize-collecting steiner tree problems. *European Journal of Operational Research 229*(3), 599–612.

Aráoz, J., E. Fernández, and O. Meza (2009). An LP based algorithm for the privatized rural postman problem. *European Journal of Operational Research 196*(5), 866–896.

Aras, N., D. Aksen, and M. Tuğrul Tekin (2011). Selective multi-depot vehicle routing problem with pricing. *Transportation Research Part C 19*(5), 866–884.

Asgari, N., R. Z. Farahani, and M. Goh (2013). Network design approach for hub ports-shipping companies competition and cooperation. *Transportation Research Part A 48*, 1–18.

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

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

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

Church, R. and C. R. ReVelle (1974). The maximal covering location problem. *Papers in Regional Science 32*(1), 101–118.

Contreras, I. (2015). Hub location problems. In G. Laporte, F. Saldanha da Gama, and S. Nickel (Eds.), *Location Science*, pp. 311–344. Springer.

Contreras, I. and E. Fernández (2012). General network design: A unified view of combined location and network design problems. *European Journal of Operational Research 219*(3), 680–697.

Contreras, I. and E. Fernández (2014). Hub location as the minimization of a supermodular set function. *Operations Research 62*(3), 557–570.

Contreras, I., E. Fernández, and A. Marín (2010). The tree of hubs location problem. *European Journal of Operational Research 202*(2), 390–400.

Contreras, I., M. Tanash, and N. Vidyarthi (2016). Exact and heuristic approaches for the cycle hub location problem. *Annals of Operations Research*. DOI 10.1007/s10479-015-2091-2.

de Camargo, R. S., G. Miranda, and H. P. Luna (2008). Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers and Operations Research 35*(4), 1047–1064.

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

Feillet, D., P. Dejax, and M. Gendreau (2005). Traveling salesman problems with profits. *Transportation Science 39*(2), 188–205.

Gelareh, S., R. N. Monemi, and S. Nickel (2015). Multi-period hub location problems in transportation. *Transportation Research Part E: Logistics and Transportation Review 75*, 67 – 94.

Gelareh, S., S. Nickel, and D. Pisinger (2010). Liner shipping hub network design in a competitive environment. *Transportation Research Part E 46*(6), 991–1004.

Hamacher, H. W., M. Labbé, S. Nickel, and T. Sonneborn (2004). Adapting polyhedral properties from facility to hub location problems. *Discrete Applied Mathematics 145*(1), 104 – 116.

Hwang, Y. H. and Y. H. Lee (2012). Uncapacitated single allocation p-hub maximal covering problem. *Computers and Industrial Engineering 63*(2), 382–389.

Labbé, M. and H. Yaman (2008). Solving the hub location problem in a star–star network. *Networks 51*(1), 19–33.

Lin, C.-C. and S.-C. Lee (2010). The competition game on hub network design. *Transportation Research Part B 44*(4), 618–629.

Lowe, T. and T. Sim (2012). The hub covering flow problem. *Journal of the Operational Research Society 64*(7), 973–981.

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

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

Martins de Sá, E., I. Contreras, and J.-F. Cordeau (2015a). Exact and heuristic algorithms for the design of hub networks with multiple lines. *European Journal of Operational Research 246*(1), 186–198.

Martins de Sá, E., I. Contreras, J.-F. Cordeau, R. Saraiva de Camargo, and G. de Miranda (2015b). The hub line location problem. *Transportation Science 49*(3), 500–518.

OKelly, M. E., H. P. L. Luna, R. S. De Camargo, and G. de Miranda Jr (2015). Hub location problems with price sensitive demands. *Networks and Spatial Economics 15*(4), 917–945.

Sasaki, M., J. F. Campbell, A. Ernst, and M. Krishnamoorthy (2014). A Stackelberg hub arc location model for a competitive environment. *Computers and Operations Research 47*, 27–41.

Zanjirani Farahani, R., M. Hekmatfar, A. B. Arabani, and E. Nikbakhsh (2013). Hub location problems: A review of models, classification, solution techniques, and applications. *Computers & Industrial Engineering 64*(4), 1096–1109.