European Journal of Operational Research 291 (2021) 1–17

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Invited Review

Perspectives on modeling hub location problems

Sibel A. Alumur[a,*], James F. Campbell[b], Ivan Contreras[c], Bahar Y. Kara[d], Vladimir Marianov[e], Morton E. O'Kelly[f]

[a] *Department of Management Sciences, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada*
[b] *College of Business Administration, University of Missouri-St. Louis, 63121, USA*
[c] *Concordia University and Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, QC H3G 1M8, Canada*
[d] *Department of Industrial Engineering, Bilkent University, Ankara 06800, Turkey*
[e] *Department of Electrical Engineering, Pontificia Universidad Católica de Chile and Instituto Sistemas Complejos de Ingeniería, Santiago, Chile*
[f] *Department of Geography, The Ohio State University, Columbus, OH 43210, USA*

ARTICLE INFO

Article history:
Received 27 September 2019
Accepted 27 September 2020
Available online 2 October 2020

Keywords:
Location
Hub network design
Hub-and-spoke networks

ABSTRACT

The aim of this paper is to provide insights for better modeling hub location problems to help create a road map for future hub location research. We first present a taxonomy to provide a framework for the broad array of hub location models, and then seek to identify key gaps in the literature that provide opportunities for better models. We provide some new perspectives in several areas, including the historical evolution of hub location research, models for economies of scale, and relevant characteristics of different applications. We also provide a succinct summary of state-of-the-art formulation and solution approaches. We conclude with a set of themes that can be addressed in the future for better modeling hub location problems.

1. Introduction

Hub location problems (HLPs) provide a rich source of challenging models based on real-world transportation and telecommunications systems. Hubs are facilities of various types that perform switching, sorting, connecting, and consolidation/break-bulk functions for traffic (e.g., passengers, freight or information) between many origins and destinations. HLPs necessarily address the *location* of hub facilities, but also include network design decisions. Broadly speaking, HLPs consist of locating hub facilities, designing the hub network and of determining the routing of flows through the network, while optimizing a cost-based or service-based objective.

The primary advantages of hub networks stem from (i) lower movement (i.e. transportation or transmission) costs from consolidated flows that exploit economies of scale, especially between hubs, (ii) reduced costs from establishing a sparser network to connect many dispersed origin–destination (O–D) pairs, and (iii) better service from allowing more frequent connections due to the consolidated flows. Real-world examples of hub networks include passenger and freight airlines, less-than-truckload (LTL) and truckload (TL) transportation, postal operations, express shipment and cargo delivery, liner shipping, public transit, and computer and telecommunication networks.

In this paper, we identify some important gaps in the hub location literature and highlight key themes for future research. These key themes include modeling of economies of scale, including time considerations and more complex objectives, better modeling real-world details, selecting formulations and solution approaches, new application areas, and developing insights. We hope this paper provokes a deeper assessment of the choices made to model various HLPs which leads to better models that more accurately reflect real-world hub systems while retaining computational tractability.

Network hub location models include origin/destination (O/D) nodes, hub nodes (to be located), an access network to connect the non-hub O/Ds to hubs, and an inter-hub network connecting the hubs. In hub location models, one key modeling decision is whether to assume the topology of the access and inter-hub networks, or to let this emerge based on the models for the underlying costs, revenues, service, and performance. Another important modeling decision is to properly account for the different link (edge/arc) types, such as for different transportation modes or cable/transmission types, that occur in real-world hub networks.

To provide a framework for the broad array of hub location problems and models, we present a taxonomy in Fig. 1 built around the problem setting, the nature of demand, the objectives and constraints. Consideration of these elements is useful in

* Corresponding author.
E-mail addresses: sibel.alumur@uwaterloo.ca (S.A. Alumur), campbell@umsl.edu (J.F. Campbell), icontrer@encs.concordia.ca (I. Contreras), bkara@bilkent.edu.tr (B.Y. Kara), marianov@ing.puc.cl (V. Marianov), okelly.1@osu.edu (M.E. O'Kelly).

https://doi.org/10.1016/j.ejor.2020.09.039
0377-2217/

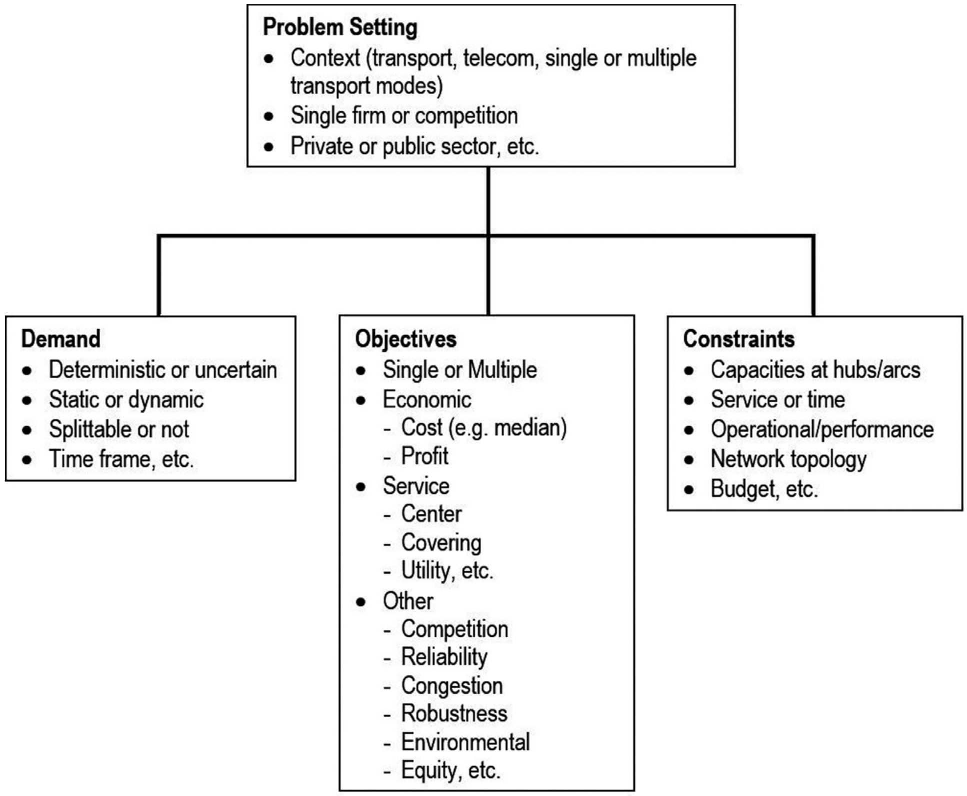

Fig. 1. Taxonomy of hub location models.

building effective models that better reflect the relevant real-world characteristics being modeled. The problem setting for HLPs is a fundamental driver for the modeling decisions. Key features of the problem setting include the context (transportation, telecom, single or multiple transportation modes, etc.), whether for a single firm or a competitive setting, whether in the private or public sector, etc. Given the wide range of applications and problem settings, building a single model that handles well all features of HLPs is unrealistic.

The nature of the demand for service is a key aspect of modeling HLPs for various settings. In HLPs, each element of the demand is a request for service for a specified *O–D pair*; e.g., transportation of a given amount of freight, passengers, or information from a specified origin to a specified destination, possibly via hubs to be located. O–D pairs are allocated to one or more of the hubs being located, where the hubs are intermediate points that provide a desirable service (e.g., consolidation or sorting of the flows) along the O–D flow paths. Demand may be categorized as deterministic or uncertain, static or dynamic, splittable (i.e. some O–D demands may be split to follow different O–D paths, as for packet switching in telecom) or not, and integer (e.g., freight may be packed in large containers transported as unit loads for all or part of an O–D trip) or not.

Objectives in hub location models address economic or service measures, with cost (or a weighted distance) minimization being the most common. However, there are models with objectives related to profit, competition (e.g., maximize market share captured), reliability, congestion, robustness, environmental performance and equity. Note that traditional service measures for center and covering problems can have multiple interpretations in HLPs, such as minimizing the longest O–D path length, or minimizing the longest arc in any path. Multi-objective models or broader measures like utility may better reflect real-world considerations, though at an increase in the complexity of modeling and computation.

Constraints are essential in hub location models to delimit the allowable activities and operations, as well as the design of the network itself. These are generally derived directly from the real-world setting, such as for vehicle or facility capacities. However, constraints also arise indirectly as with topological restrictions captured in variable definitions (e.g., single allocation access networks), or by modeling a level of service goal (e.g., seeking shorter flight times by using hop constraints to limit the number of hubs in an O–D flow path). Due to the intermediate nature of hubs in O–D paths, their capacities can be modeled as the flow moving through, entering, or leaving the hub. Arc capacities are important in some settings, and may be measured in terms of vehicles or frequencies, or simply flow units. Note that capacities will generally depend on the available equipment and infrastructure (e.g., airport gates, roads, sorting equipment, communication links, vehicles, etc.) and capacities should have time dimensions to reflect the resources being used (e.g., packages per hour that can be processed), though these are often not explicit in hub location models.

Network topology constraints (or assumptions) are commonly used to force certain topologies (e.g., single allocation of non-hubs to hubs), often as a proxy for some aspects of service or cost. The topology of the inter-hub network (e.g., complete, tree, ring, etc.) is especially important to reflect the cost and service aspects of establishing the network, as well as the reduced costs from consolidation of flows between hubs. Hub location-routing models include constraints on routing aspects to model multi-stop collection and distribution routes at hubs. Finally, budget constraints are used in a variety of ways, such as exogenous limits on the number of hubs or number of vehicles, or on various aspects of system costs.

Service in real-world hub systems can usually be viewed in terms of customer satisfaction (with dimensions such as price and quality) and/or time (e.g., delivery time(s), delays, etc.) though broader operational or performance metrics such as late arrivals, congestion and waiting times (e.g., from queueing), or reliability can also be relevant. Service is commonly modeled via constraints, (e.g., maximum or average delivery time, hop constraints for paths), but it can also be an objective, as for center and covering problems (or maximizing utility). Note that using constraints and assumptions to enforce certain features in hub location models can create unintended consequences and unreasonably limit the response of the model to varying inputs. One important example of this is how assumptions in the classic HLPs often lead to improper modeling of economies of scale. This is very important as economies of scale often provide a major reason for adopting hub-and-spoke network designs; thus, the issue of modeling economies of scale is addressed in some detail in Section 3.

In practice, hub networks are determined by, and evolve in response to, economic and service pressures, so one goal of hub location research should be to replace artificial assumptions and constraints with better models for cost and service. This will allow the fundamental cost and service performance to drive the design of the network, including location of hubs. Thus, a key question for better modeling HLPs is *how to properly model the costs and service*. Within this question are important sub-questions about: (i) How to handle time dimensions of cost and service, (ii) How to allocate costs to nodes and arcs (or vehicles) in the models, and (iii) How to model the time spent at nodes and arcs. In real-world hub systems there are capital costs and operating costs. Capital costs are incurred for facilities, equipment, network infrastructure, and vehicles, and these are usually reflected in fixed costs for establishing (locating) hubs, for establishing connections between nodes, and for acquisition of vehicles. Operating costs are incurred for labor (for handling, driving, etc.), fuel and maintenance, and these are reflected in transportation costs, and material (or passenger) handling costs. These costs may be combined in an objective or used in budget constraints.

The time dimension of demand is another area that needs more careful attention in model building, as time is an essential aspect of transportation and logistics, as well as telecommunications (though on a different scale), that is linked to both service levels and financial measures. Hub location models typically and importantly require a coordination of flows across time to allow for efficient consolidation, switching and sorting. Demand in HLPs is generally presented as a set of numbers, one for each O–D pair, without detailed explanation of the represented time frame. This implies a common time period for all O–D pairs, so HLPs are generally solved for a single time period (e.g., month or year) which is presumed to repeat. Yet, demand is necessarily a *rate* in units of flow per unit time (e.g., passengers or tons per week), and it is generally *dynamic* in practice, especially over longer time periods (e.g., air passenger travel and freight shipment patterns vary widely, though somewhat predictably, throughout the year). Furthermore, the operations in a hub network occur on a relatively small time scale (e.g., hours or seconds), as movement on arcs consumes time for transportation or transmission, and the key activities to consolidate and connect traffic at hubs necessarily involves temporal coordination of incoming and outgoing flows. This time dimension has rarely been dealt with in the literature on hub location and we address this further in Section 3.

To this date, most papers on hub location are oriented to the design of airline, cargo delivery and telecommunications networks. These applications are rich enough as to span a very large number of papers. However, new applications are appearing, in completely unrelated areas, and in our view, the models and mathematical methods developed for applications in transportation could enrich other fields, extending even to neuroscience (Esfahlani, Bertolero, Bassett, & Betzel, 2020). We review the particularities of some known applications and propose some extensions in Section 4.

Proper modeling of the economic aspects, network constraints, practical aspects of old and new applications, and service needs, require a huge effort dedicated to innovative modeling and development of solution methods. Further complexity is introduced by the practical need to model dynamics, as with changing capacities over time in response to changing demands (e.g., to reflect seasonal shipping and travel patterns), or disruptions. While impressive progress has been made in optimizing (often idealized) mathematical models of HLPs, we believe there are still great opportunities for significant contributions, especially from developing and solving better hub location models. We present a succinct summary of promising formulations and solution algorithms in Section 5.

The goal of this paper is not to provide a comprehensive survey of past hub location research, as there are several good reviews and surveys (e.g., Alumur & Kara, 2008; Campbell & O'Kelly, 2012; Contreras, 2020; Contreras & O'Kelly, 2019; Farahani, Hekmatfar, Arabani, & Nikbakhsh, 2013). While our goal is more forward looking, we do briefly note below some recent publications (since the surveys above) to show the very wide range of activity in the field.

Fernández and Sgalambro (2020) address different policies for carriers collaboration. Soylu and Katip (2019) locate hubs in airports, optimizing multiple objectives. Competitive hub location is revisited by Araújo, Roboredo, Pessoa and Pereira (2020), who use an exact method to solve the $(r|p)$ hub centroid problem. Najy and Diabat (2020) use Benders decomposition to solve a multiple-allocation HLP with congestion and flow-dependent economies of scale. Aboytes-Ojeda, Castillo-Villar, and Roni (2020) solve a two-stage stochastic biofuel hub-and-spoke network problem using a combination of exact and heuristic methods. Yang, Bostel, and Dejax (2019) explore MILP and memetic methods for the joint hub location and routing problem. Chen, Hsieh, Hung, and Klasing (2020) propose approximation algorithms for the p-hub center routing problem. Mohammadi, Jule, and Tavakkoli-Moghaddam (2019) study the reliable HLP with disruptions and Zheng, Zhang, Qi, and Wang (2020) locate hubs for a liner shipping system.

Some new applications have also been addressed in the recent literature. Dukkanci, Peker, and Kara (2019) introduce the green hub problem, Mirzapour-Kamanaj, Majidi, Zare, and Kazemzadeh (2020) address energy transmission hub networks, and bike-and-ride hub systems are addressed by Tavassoli and Tamannae (2020). Hub systems for medical applications are important including in drone delivery networks (Macias, Angeloudis, & Ochieng, 2020), to access medication treatment for opioid use disorders (Miele et al., 2020; Reif, Brolin, Stewart, Fuchs, & Mazel, 2020), and in identifying structures in functional brain networks (Esfahlani et al., 2020). This very broad range of newer applications indicates the power of HLPs and the need for more and better models.

The remainder of this article begins with some historical perspectives, and then includes discussions of modeling economies of scale, application-related issues, and state-of-the-art formulations and solution approaches for HLPs. The article concludes with a summary that highlights key areas for future research.

2. A historical perspective

Some perspectives on the origins of hub location research can help clarify key features that persist in today's models. The beginning of hub research was with planar models (O'Kelly, 1986), originally presented at ISOLDE III, 1984 (see http://isoldeconference.org/), though discrete approaches quickly became predominant. O'Kelly (1987a) designed a discrete model for a HLP based on the p-median problem, which is not surprising from a geographer familiar with the seminal paper by Re Velle and Swain (1970). By adapting the existing p-median constraints, and developing an objective function to account for the interaction between facilities, this hub location model was naturally viewed as single allocation, which is a characteristic of p-median optimal solutions. These early ideas were also influenced by the multi-facility location problem (where there are flows between facilities) and the location of cluster centroids, as well as the quadratic assignment problem (QAP) (e.g., Snickars, 1978) and quadratic knapsack problem. A key new feature of HLPs was that the optimal flows between hub facilities depend on how the O/D are connected to the hubs. While these first hub location models certainly are rather straightforward, single allocation hub location models with a complete inter-hub network continue to provide interesting challenges and extensions in the literature (perhaps because of their interpretation as a binary quadratic program).

The early stages of hub location research developed methodically with systematic extensions to hub versions of the traditional facility location problems, including fixed cost models (O'Kelly, 1992) and ideas that reflect adaptation of various heuristics (GRASP, tabu-search, genetic algorithms, etc.) to the problem. The accumulation of these various extensions led ultimately (after some years) to hub location models becoming accepted as a class of important and distinctive problems in location theory. Early reviews of hub location research include O'Kelly and Miller (1994), Campbell (1994b), Klincewicz (1998), and Campbell, Ernst, and Krishnamoorthy (2002), and the first special journal issue on hub location appeared in Location Science in 1996 (vol. 4(3)). The second special journal issue on HLPs appeared in 2009 in Computers & Operations Research (vol. 36(12)), and a special section was published in 2019 again in Computers & Operations Research (vol. 104).

The early hub location models provided some interesting conceptual insights regarding allocation and quadratic objectives from links to related problems. The *nearest facility allocation* property that aids in the solution of many location problems, unfortunately, cannot be relied upon for HLPs, as the allocation decision for an O/D node depends on all the O–D flows associated with that node. However, because *spatial interaction* tends to exhibit a distance decay effect, allocation to the nearest hub may provide a good approximation (a property used in some heuristics). The nonconvexity of the hub location (cost) objective function in quadratic formulations of hub models adds complexity, but also provides a linkage to well-studied quadratic problems, where advances in formulation and solution may help in solving large instances of hub location problems.

Multiple allocation HLPs arose with Campbell's binary linear programming model (presented at ISOLDE V in 1990, and Campbell (1992)), building off of ideas for many-to-many distribution (e.g., Daganzo, 1987). Multiple allocation hub location models allow each non-hub O/D to send and receive from multiple hubs, and from a node pair perspective this does produce the "nearest allocation" property (in an uncapacitated environment), as each O–D pair is optimally allocated to the least cost hub pair (including single node hub "pairs"). Consequently, O–D routes are shortest paths in the network, which makes multiple allocation variants easier to solve than the analogous single allocation HLPs.

The early MIP formulations were often designed more for clarity and exposition, than for computation efficiency. However, a key modeling advance was the multicommodity flow models of Ernst and Krishnamoorthy (1996, 1998a, 1998b, 1999). Another interesting early paper (perhaps ahead of its time) was the airline network design model of Jaillet, Song, and Yu (1996) that considered multiple aircraft types, and rather than explicitly deciding on hub locations allowed hubs to emerge from the optimal flows of different types of aircraft. In a different application, the review by Klincewicz (1998) collected early research on telecom and computer communication hub networks.

Planar hub location problems also contributed to early modeling insights in several ways including minimax planar location (O'Kelly, 1987b; O'Kelly & Miller, 1991) (e.g., visualizing the elliptical service "radii" of hub pairs in minimax location in the plane). The planar "round trip location problem" (Drezner & Wesolowsky, 1982) also provided a useful algorithm for the planar hub covering problems, though multi-facility planar hub covering remains challenging; see Sim, Lowe, and Thomas (2009) for an interesting perspective.

Some ideas from the earliest works have still not been fully explored, such as the linkage between the demand and how well the O–D nodes are serviced by the hubs (in a solution). This type of elastic response is typically omitted in models, though Taner and Kara (2016), Han and Zhang (2013), O'Kelly, Luna, de Camargo, and de Miranda (2015), and Lüer-Villagra and Marianov (2013) have contributions in this area. Also, single allocation formulations may benefit from more deeply exploring connections to the semi-assignment problem polytope (e.g., Saito, Fujie, Matsui, & Matuura, 2009), and properties of the distance metric (e.g., see Meier & Clausen, 2017).

3. Modeling economies of scale in hub location

Because economies of scale are a key motivation for hub networks in transportation and telecommunications – and many of the fundamental (and early) hub location models poorly capture economies of scale, this section provides details on this key modeling decision. Consolidation of flows to provide economies of scale can be a *raison d'etre* for hub networks, but *economies of density* and *economies of spatial scope* are also important. Basso and Jara-Díaz (2006a,b) and Jara-Díaz, Cortés, and Morales (2013) distinguish economies of density as the reduction in cost per traffic unit due to an increase in traffic density while the route structure of the network remains unchanged, while *economies of scale* is similar, but without the assumption of a fixed route structure. Also, *economies of spatial scope* can be defined as unit cost reductions due to the addition of new services or new O–D pairs.

For a given set of O–D demand, moving from a network in which O–D flows are sent directly between every pair of nodes (a complete network) to having flows sent through a hub network in any of its variants allows achieving economies of scale just by changing the route structure. In such networks, the consolidation of flows increases the traffic density in some (or most) route segments. For a transportation setting, this greater traffic density allows using larger and more cost efficient vehicles (e.g., aircraft or trucks) – with appropriate trip frequencies. The reduction of unit costs comes from sharing fixed costs over more units of demand (e.g., passengers), and possibly from using vehicles with lower variable costs. Additional benefits of hub networks can come from increasing the frequencies of service on links (as a result of higher traffic density), and a better traffic balance across the network. Further, hubs can concentrate administrative and technical resources, reducing the investment, operational cost and inventories. Based on empirical evidence, Johnston and Ozment (2013) found that economies of scale do exist in the USA airline industry. A drawback of hub networks is that hubs can be expensive, routes are not direct, and transshipments/connections are required.

In this section we concentrate on modeling economies of scale for transportation HLPs, a topic that had drawn repeated attention over the past two decades (e.g., Alumur, Kara, & Karasan, 2009b; Bryan, 1998; Camargo, de Miranda, & Luna, 2009; Campbell, 2013; Campbell, Ernst, & Krishnamoorthy, 2005a; Campbell, Ernst, & Krishnamoorthy, 2005b; Kimms, 2006; Martins de Sá, Contreras, & Cordeau, 2015a; Martins de Sá, Contreras, Cordeau, Camargo, & Miranda, 2015b; O'Kelly & Bryan, 1998; Tanash, Contreras, & Vidyarthi, 2017). The original hub location models considered a single firm locating hubs to provide service to all O–D demands, an inelastic demand, and only variable costs on route segments or arcs. In classic single and multiple allocation models, cost mini-

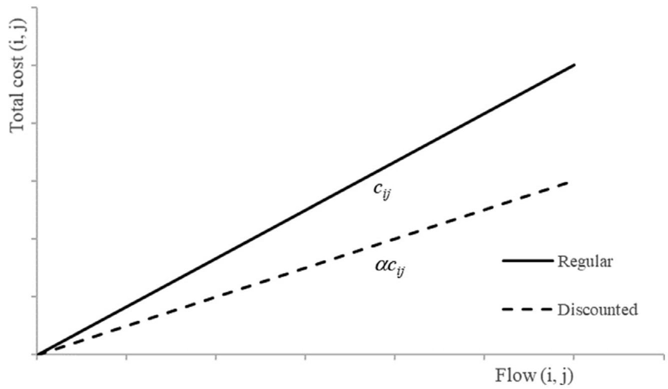

Fig. 2. Separated cost curves for inter-hub and spoke arcs. No fixed costs.

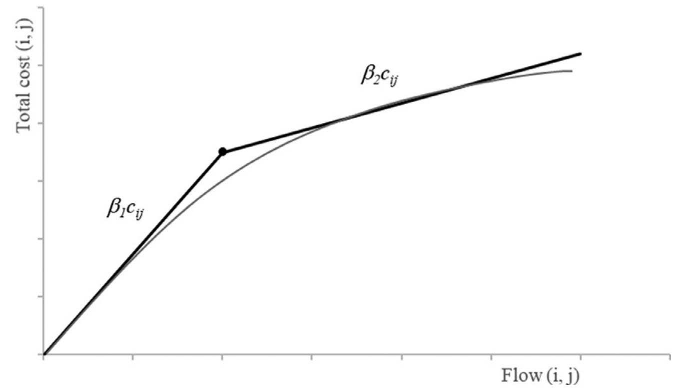

Fig. 3. Piecewise linearization (black) of a nonlinear curve (gray).

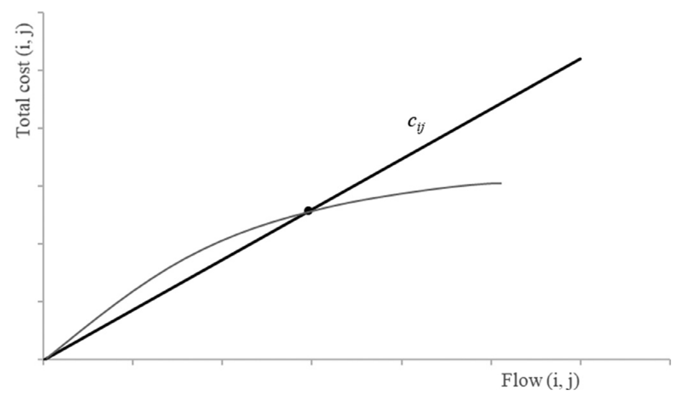

Fig. 4. Inter-hub cost (gray) and regular cost (black).

mization resulted in complete inter-hub networks. The cost model included a known "regular" unit cost of traffic corresponding to that of transporting low volumes of traffic on the *spokes* in the access network, and a discount $0 \le \alpha \le 1$ applied to the regular unit cost for all inter-hub traffic. As a result of consolidation, all *hub arcs* connecting hubs (i.e. the inter-hub network) were assumed to carry a large amount of traffic, and this assumption was reflected in presumed economies of scale via a discount on these arcs. The flow on all spokes (i.e. allocation arcs) in the access network was assumed to be low, so these unit costs were never discounted. By definition, there was no threshold defining high and low traffic from the point of view of costs (though Campbell, 1994a, did introduce minimum flow thresholds for the spokes).

This simple cost model allowed formulating tractable, but still often computationally intensive, hub location models that have been (and continue to be) the basis for much of the research in the field. Fig. 2 depicts the cost functions used in these models, for each arc, where the slopes of the two lines represent respectively the regular unit cost (e.g., cost per ton-mile of passenger-mile) for the access network (i.e. spokes) (upper line with slope c_{ij}) and the discounted cost for the inter-hub network (i.e. hub arcs) (lower line with slope αc_{ij}). A common interpretation is that these two lines result from using two sizes of vehicles, with small vehicles limited to the access network and large vehicles limited to the inter-hub network. The pairing of the two cost lines with the two network components is fixed by assumption, as there is no level of flow for an arc at which the cost line would switch. Because the unit cost in this basic model is independent of flows, the model does not reproduce economies of density.

In optimal solutions of such models, it is quite often the case that some of the inter-hub links with a discounted cost rate carry a low level of traffic, while some spokes (all presumed to not have a discounted cost rate) carry a (very) high level of flow. This simple cost model is clearly not true in practice, as it would be uneconomical to dispatch a partially filled large vehicle when a lower cost small vehicle would suffice. (Further, sending a partially full vehicle would not provide the same unit cost in practice as sending a completely full vehicle.) Similarly, sending multiple full small vehicles (on a spoke) is uneconomical compared to sending one large vehicle. This illustration highlights the importance of frequency of service and scheduling considerations for HLPs, which are absent from basic hub location models. Note that a low rate of flow could be accumulated over time to create a full large vehicle load, which would be infrequently sent, so this has clear consequences for consolidation opportunities.

In summary, the abstractions and assumptions in the classic hub location models often produce poor models of economies of scale, and also obscure important "time" issues of scheduling and vehicle use. The consequence of using these models is a mismatch of costs and flows as the optimal solution is a complete inter-hub network of discounted cost hub arcs, regardless of the flows on the arcs!

One method proposed to improve the model of economies of scale, while retaining the same two separate cost functions in Fig. 2, is the use of an incomplete inter-hub network (e.g., Alumur et al., 2009b; Campbell et al., 2005a; Contreras, Fernández, & Marín, 2010; Martins de Sá, Camargo, & Miranda, 2013). An incomplete hub network tends to reduce the number of low-traffic inter-hub arcs, though it may produce a network with more circuitous flow routings that is less suitable for passenger (and high service level) settings.

Another method proposed to improve modeling of economies of scale uses a nonlinear cost function of the flow on a link, represented as a gray curve in Fig. 3 (Bryan, 1998; O'Kelly & Bryan, 1998). The curve can be approximated by a piecewise linearization as shown by the black line segments in Fig. 3 representing two modes or technologies (e.g., small and large vehicles). This modeling approach reflects the use of large vehicles (with discount $\beta_2 < \beta_1$) only for the large traffic flows, although economies due to the increase in occupancy of small vehicles could be lost unless there is a large number of linear segments (e.g., representing different load levels of the same vehicle). The threshold between lines is a consequence of the formula for the nonlinear curve, as opposed to a threshold computed using the real costs of each type of vehicle. Note that some research uses nonlinear cost functions directly (e.g., Horner & O'Kelly, 2001). Racunica and Wynter (2005) also use a generic concave curve for inter-hub arcs for locating intermodal freight hubs. The difference is that their concave curve intersects the linear regular cost, as shown in Fig. 4. This feature aims at representing an *efficiency threshold* up to which the

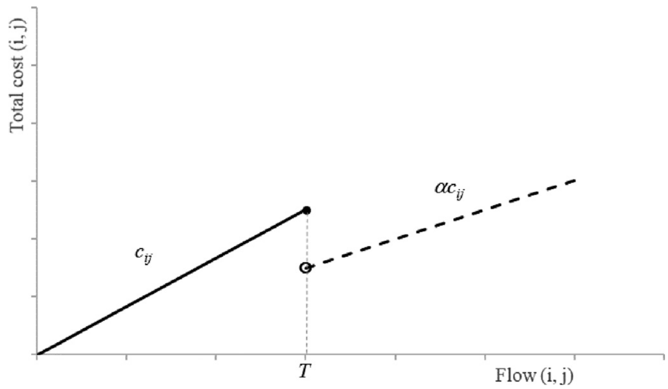

Fig. 5. Two types of vehicles, no fixed costs.

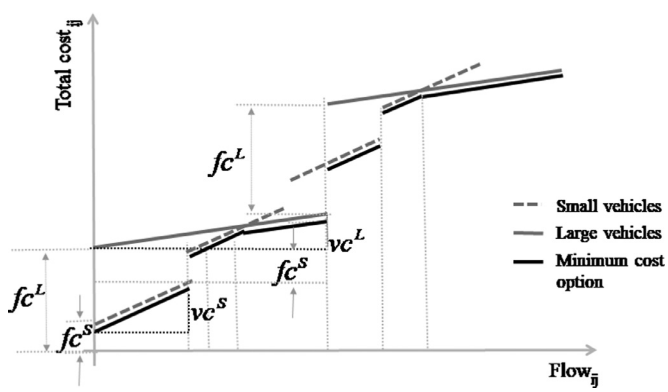

Fig. 6. A cost curve considering fixed and unit costs.

inter-hub cost is higher than the regular cost, and above which discount is applied.

This nonlinear curve shows how marginal costs decrease, as vehicle load increases. For a model with strategic purposes, however, the nonlinearity may introduce more difficulty than it does improve the match to reality. Furthermore, these curves are not the best representation of a system with different types of vehicles.

An alternative linear cost model with a threshold for switching cost lines (e.g., vehicles) is shown in Fig. 5 (see Podnar, Skorin-Kapov, & Skorin-Kapov, 2002 and Cunha & Silva, 2007). Here the fixed cost discount α applies only when traffic on the link exceeds the threshold. (Note that a shipper could artificially inflate their traffic to lower the costs, which can be a valid strategy.) This cost model allows either large or small vehicles to be used on any link, based on the level of traffic. This cost model, as well as that in Fig. 2, is suitable for the case in which a third party provides transportation service, as the absence of fixed costs as well as volume discounts are common in this case.

Kimms (2006) proposes a more detailed cost model including fixed costs, as shown in Fig. 6 when there are two types of vehicles. Large vehicles have a high fixed cost (fc^L) and low unit cost (vc^L), shown in solid gray, and small vehicles have low fixed cost (fc^S) and high unit cost (vc^S), shown in dashed gray. Generally, $fc^L > fc^S$ and $vc^L < vc^S$. Fixed vehicle costs are incurred every time a new vehicle is added to the route. In Fig. 6, for low traffic levels, one small vehicle is used, and then as the flow increases, the fixed cost of the vehicle is shared among more flow units (e.g., passengers), resulting in economies of density. When the capacity of a small vehicle is reached, a second small vehicle is added, with the corresponding fixed cost that produces the step in the small vehicle cost curve. In the setting shown in the figure, when the flow exceeds the capacity of approximately one and a half small vehicles, it is lower total cost to use one large vehicle. Above this level of flow various combinations of large and small vehicles may be best; though as the figure shows there could be higher flows for which a set of only small vehicles is best.

The cost model in Fig. 6 is readily applicable to the fleet mix problem in which a fleet of vehicles is to be chosen to carry a known amount of demand (as implemented in Serper & Alumur, 2016). However, it is less appropriate for the more practical fleet mix problem with changing levels of demand, given how the optimal mix of vehicles on a link changes with the demand, thus leading to possibly very frequent changes throughout the network. Also note that the cost model is independent of any detailed scheduling, but it effectively assumes that on a leg or route segment, there may be several vehicles departing at the same time or within short time periods. In practice, specified service frequencies and scheduling to coordinate operations at hubs are important (e.g., to provide for consolidation or reduce passenger waiting times). As earlier, this highlights how the cost model links to the issues of both scheduling and frequency of service. Masaeli, Alumur, and Bookbinder (2018) provide integrated models that include shipment scheduling to determine the number of dispatches of different types of vehicles between hubs, as well as the time period to dispatch each vehicle from a hub. Along a different line, some researchers have chosen to design transportation networks with explicit consideration of economies of scale, but without explicitly locating hubs (e.g., Jaillet et al., 1996, and Podnar et al., 2002). Furthermore, a simpler version of the costs represented in Fig. 6, namely hub networks with modular arc capacities, have been proposed by Rostami and Buchheim (2015), Hoff, Peiro, and Corberán (2017), Meier (2017), and Tanash et al. (2017).

Transportation hub location researchers have generally taken a strategic approach that effectively models a single time period and ignores the vehicles that transport the flows. While there are some multi-period hub location models (e.g. Alumur, Nickel, Saldanha-da Gama, & Secerdin, 2016; Contreras, Cordeau, & Laporte, 2011b; Gelareh, Monemi, & Nickel, 2015), these still do not address the issues of service frequency and synchronization. In developing better models for HLPs, several issues should be kept in mind.

1. Modeling of economies of scale for network design (strategic) purposes requires further research, as the simple models (e.g., Fig. 2) that are solvable in reasonable times do not represent adequately these economies, while the improved models of Figs. 3–6 are difficult to solve. A promising approach to overcome this increased difficulty is to use linear approximations of different cost models that either minimize some measurement deviation between the original (nonlinear) cost model and its linear approximation or provide a performance guarantee in the quality of the approximation. Fard and Alfandari (2019) have obtained promising results in this direction when considering specific generalized linear cost functions to approximate step-wise cost functions. Lüer-Villagra, Eiselt, and Marianov (2019) have applied metaheuristics to concave and convex cost curves of several segments.
2. Even the more sophisticated model shown in Fig. 6 disregards the fundamental issue of service frequencies. Ideally, one would choose both vehicle types and frequencies, and hub locations and network designs together (see for example Masaeli et al., 2018; see also the service network design literature). A good review can be found in Danach (2017). This is a very complicated problem, because the travel frequencies on inter-hub arcs depend on the frequencies on spokes feeding the corresponding hubs and, since the allocation and location decisions are not known beforehand, travel frequencies cannot be determined *a priori* to be used for the network design. Also, the traffic on a specific arc is usually comprised of non-stop, 1-stop and 2-stop

flows, and these flow components may deserve different treatments. For example, consider the air passenger flow originating from non-hub city Boston on the Boston-New York (spoke) arc. This may include a large non-stop flow ending at the hub at New York, a 1-stop flow bound for a second hub in Denver, and a 2-stop flow that travels via New York and Denver to its destination of Las Vegas. The large non-stop Boston-New York flow might be sent in small planes at a high frequency (e.g., hourly), while the 1-stop Boston-Denver flow is sent to the New York hub in two waves (am and pm), and the 2-stop Boston-Las Vegas flow is sent to New York in a single wave. Further, these waves need to be synchronized across all the hubs and nodes to facilitate connection and consolidation throughout the network.

3. Once a hub network is established, it is possible to add new origins and destinations at a reduced cost, since this new traffic can use many of the resources already deployed. In this case, there are *economies of scope*, since the company owning (or using) the network can open the new routes at a lower cost than its competitors. Finally, while models with fixed costs on arcs have been addressed (e.g., O'Kelly, Campbell, Camargo, & Miranda, 2014; Tanash et al., 2017), more work is needed on the problem of building a network with fixed costs per unit of distance on route segments (as for digging trenches and laying cable, constructing roads, railroads, or pipelines); see the related problem in Marianov, Gutiérrez-Jarpa, Obreque, and Cornejo (2012).

4. Applications of hub location problems

This section discusses several application areas of HLPs, first focusing on classical applications that were primarily motivated by airline, cargo delivery and telecommunications networks. We then describe additional modeling challenges for other applications in an "Extensions" subsection.

4.1. Classical application areas

HLPs were originally motivated by applications from airline, cargo delivery and telecommunications networks. Even though the concept and the need for hubbing is similar in all of these applications, the nature of the flows (people, freight, or data) creates important distinctions.

Passenger airline networks have long been a prime application area for hub location models. Because the demand in these systems is people making individual choices and who prefer not to wait, most real-life airline networks use a multiple allocation access network. Each O–D pair is then served by the "best" (e.g., fastest or lowest cost) route though the set of hubs. Even though multiple allocation is used in airline systems, a detailed examination of the networks shows that the non-hub nodes are often allocated to only a few hubs, as in *r*-allocation models (Corberan, Landete, Peiro, & Saldanha-da Gama, 2019; Yaman, 2011). Also, the design of incomplete inter-hub networks can be important (Alibeyg, Contreras, & Fernandez, 2016; 2018; Alumur et al., 2009b; Campbell et al., 2005a; 2005b). Note however that some airlines, especially low-cost carriers, provide point-to-point (non-stop) services instead of a hub-and-spoke structure (Alibeyg et al., 2016; Cook & Goodwin, 2008; Taherkhani & Alumur, 2019). In this case, only O–D pairs with sufficient traffic will have flights, so some O–D pairs may not be connected. While hubbing can occur in these networks ("hubbing by coincidence"), the scheduling of operations and facilities is not designed for easy connections (e.g., short waits and baggage transfers).

Cargo (or freight) delivery, systems transport goods (not people) and often use a single allocation structure, due to the greater time flexibility that allows longer O–D paths, economies of scale for sorting, and managerial ease. While passengers traveling between two cities prefer faster (more direct) and less expensive routes, goods are indifferent to the route they travel, so there is flexibility in selecting a route, as long as the service level is met (e.g., next day delivery by noon). This allows freight carriers to use more circuitous routes that may increase travel distance or time (versus a more direct O–D route) to allow savings from consolidation and economies of scale. While the shipper and receiver (consignee) do not care about the details of the O–D path followed by their cargo, as long as it arrives when expected, the routing details are a great concern to the cargo carrier, as those determine the cost and service/performance (and thus revenues). Using single allocation allows a specific hub to process all the incoming and outgoing cargo of each city (spoke), and also increases the flows on the access network links which can reduce costs from economies of scale. Also, note that the inter-hub networks of cargo companies tend to be rather sparse, so incomplete networks are relevant (e.g., Mokhtar, Redi, Krishnamoorthy, & Ernst, 2019; Tanash et al., 2017). To ensure cargo is not "late", hub location models for cargo applications often include constraints to ensure desired service levels (e.g., Alumur & Kara, 2009a; Tan & Kara, 2007; Yaman, Karasan, & Kara, 2012). Another distinguishing factor in some cargo applications is the use of intermediate stops in the access network connecting O/Ds to hubs, often called stopovers or feeders, (see Kuby & Gray, 1993).

Design considerations for telecommunication systems are different than for transportation, because there is no physical commodity or vehicles and for these systems hubs may be routers, switches, concentrators, etc. (Klincewicz, 1998). Furthermore, because the number of potential O–D pairs is often far too large for direct connections, some form of hubbing is required. One distinguishing factor of telecommunication systems is the dominance of the fixed infrastructure cost of links and hub facilities over the variable per unit flow costs. Thus, models need to include design decisions on both the hub nodes and the links to deploy, as well as flow decisions for using the network. Generally, telecommunication systems are studied based on the considered network topology of the hub-level and (local) access networks. For instance, tree-star hub networks arise in the design of digital service networks, where hubs are connected with fiber optic links and due to their high setup cost, tree topologies are usually considered to minimize the number of required links while providing connection services between customer sites (Contreras, Fernández, & Marín, 2009b; 2010; Lee, Lim, & Park, 1996; Martins de Sá et al., 2013). Star-star hub networks arise in the design of satellite communication networks, where homing stations (hub facilities) are used in combination with terrestrial and satellite links to connect node pairs (Helme & Magnanti, 1989; Yaman, 2008; Yaman & Elloumi, 2012). Another commonly studied network topology in telecommunications is the ring-star hub network, which addresses reliability concerns by allowing 2-connectivity between node pairs (Contreras, Tanash, & Vidyarthi, 2017; Lee, Ro, & Tcha, 1993). For recent examples with ring networks, see Rieck, Ehrenberg, and Zimmermann (2014), Rodríguez-Martín, Salazar-González, and Yaman (2016), Dukkanci and Kara (2017), and Kartal, Krishnamoorthy, and Ernst (2019). There has also been hub location research in several areas in addition to the original settings of airline, cargo delivery, and telecommunications. Hub network design is an important concern for liner shipping and Zheng et al. (2020) studied the impact of the location of hub ports on the liner shipping networks. Hub networks can arise also in public transportation planning, in particular in the design of rapid transit systems and highway networks (see, for instance,Contreras et al., 2009b; Contreras et al., 2010; Martins de Sá et al., 2015a; Martins de Sá et al., 2015b; Zhong, Juan, Zong, & Su, 2018). Network planners are often interested in extending current transport options by constructing new highways or express lanes, or new rapid transit lines, and the many-to-many nature of O–D demand often leads to a hub configuration. Another transportation extension is in "city logistics", especially with the growing interest in sustainability issues. For example, "satellite" terminal facilities around a city might be viewed as hubs, with warehouses or factories being the origins and other businesses in the city being destinations; (e.g., Morganti & Gonzalez-Feliu, 2015 provide an example for food distribution to hotels, restaurants, and retail chains in a city via food hubs). Hub networks are also relevant in humanitarian logistics, as for biomedical sample transportation (Smith, Cakebread, Battarra, Shelbourne, & Cassim, 2017) or for primary healthcare delivery (Lee, Rammohan, & Sept, 2013) in rural regions in Africa.

A novel application of HLPs is in the analysis of brain connectivity networks. Neuroscientists have shown that some regions in the brain play a hub role, acting as intermediary areas for the transmission of signals between other regions (Esfahlani et al., 2020; Oldham & Fornito, 2019). Although some studies have used analytical tools such as graph theory to identify hub regions in brain connectivity networks (Bassett & Sporns, 2017), to the best of our knowledge the work of Khaniyev, Elhedhli, and Erenay (2020) is the first to employ a hub location model of the brain.

4.2. Extensions

The classical hub location applications have been extended in a variety of ways to better tailor the models to particular application settings. Hub network design for bike and ride systems has been studied in Tavassoli and Tamannae (2020). In energy management, Aboytes-Ojeda et al. (2020) provides an example for a hub-and-spoke network design problem encountered in biofuel supply chains. Finally, HLPs have been extended to include sustainability issues, with Dukkanci et al. (2019) examining carbon emissions from routing in a hub location model.

In passenger airline hub location models, the emphasis on high service levels (e.g., a high value of time) and competitive pressures (especially from low-cost carriers), has led to models that allow direct connections between (certain) O–D pairs. This high level of service comes at a higher cost (from lower traffic levels with less consolidation via hubs), so pricing these flights becomes an important additional decision (Lüer-Villagra & Marianov, 2013; O'Kelly et al., 2015). Competition poses additional challenges for airlines (and in many applications) and game theoretical approaches have been utilized in various hub location models (e.g. Araújo et al., 2020; Gelareh, Nickel, & Pisinger, 2010; Lin & Lee, 2010; Mahmutogullari & Kara, 2016; Marianov, Serra & Re Velle, 1999). An alternative to competitive pricing is a collaborative perspective as in Fernández and Sgalambro (2020).

Cargo companies usually provide several types of services, such as same-day delivery, next-day delivery, standard delivery, etc. These different service types should be considered while designing the hub network, and a company serving a large geographical area may utilize multi-modal transportation (e.g., both air and ground) to effectively provide these services (e.g., Alumur, Kara, & Karasan, 2012a; Racunica & Wynter, 2005). Because of the multiple service levels and different transportation modes, cargo applications may require a hierarchical approach that treats each service type for an O–D pair as a commodity. Each commodity can then travel through its best route that obeys the specific service requirements for that commodity. Hierarchical networks may include several layers of nodes (e.g., O/Ds, ground hubs, air hubs), with distinctive topologies (e.g., stars, rings, etc.) connecting nodes in each layer, and connecting nodes between layers (e.g., Alumur, Yaman, & Kara, 2012b; Bernardes Real, O'Kelly, de Miranda, & de Camargo, 2018; Yaman, 2009). Multimodal hub networks might also include unmanned aerial vehicles (UAVs) (drones) where UAV range limits due to battery life provide an added challenge; see Macias et al. (2020) for a study of UAVs delivering medical supplies.

Note that changing modes in a multi-modal hub environment may require more complex unloading/loading activities, along with synchronization. Synchronization is also an issue for passenger hub networks to reduce passenger waiting times at hubs. However, such synchronization of the incoming and outgoing flows at hubs is not often included in hub modeling, and because synchronization is effectively vehicle-based, hub network modeling should consider vehicle and time aspects of operations. One approach, based on "latest arrivals", was presented by Kara and Tansel (2001), and later led to radius type formulations for hub center and covering problems (Ernst, Hamacher, Jiang, Krishnamoorthy, & Woeginger, 2009; Ernst, Jiang, Krishanmoorthy, & Baatar, 2018; Hamacher & Meyer, 2006).

Congestion and reliability issues are also important for passenger and cargo delivery applications due to the additional costs (and waiting), and this can be incorporated in the network design stage, usually through nonlinear models and decomposition methods (e.g., Alumur, Nickel, Rohrbeck, & Saldanha-da Gama, 2018; Camargo & de Miranda, 2012; Camargo, de Miranda, & Ferreira, 2011; Elhedhli & Hu, 2005; Grove & O'Kelly, 1986; Najy & Diabat, 2020). Reliability can be a major factor for both telecommunication and transportation applications (An, Zhang, & Zeng, 2015; Azizi, Chauhan, Salhi, & Vidyarthi, 2016; Kim & O'Kelly, 2009; Mohammadi et al., 2019; Rostami, Kämmerling, Buchheim, & Clausen, 2018; Tran, OHanley, & Scaparra, 2016). Considering reliability for hubs and/or hub arcs has led to several protection and interdiction models (Azizi et al., 2016; Ghaffarinasab & Atayi, 2018; Ghaffarinasab & Motallebzadeh, 2018; Lei, 2013; Ramamoorthy, Jayaswal, Sinha, & Vidyarthi, 2019).

Extending HLPs to hub location-routing problems allows multi-stop routes in the access network connecting hubs and non-hubs. Hub location-routing problems combine two challenging problems, with differing time scales, and are a subject of much research (e.g., Bernardes Real, Contreras, Cordeau, de Camargo, & de Miranda, 2020; Camargo, de Miranda, & Løkketangen, 2013; Catanzaro, Gourdin, Labbe, & Ozsoy, 2011; Kartal, Hasgul, & Ernst, 2017; Rieck et al., 2014; Rodríguez-Martín, Salazar-González, & Yaman, 2014; Yang et al., 2019). HLP models have been extended to consider sustainability issues, with Dukkanci et al. (2019) introducing the green hub problem, and Parsa, Nookabadi, Flapper, and Atan (2019) considering cost and two pollution objectives for a multi-objective "green aviation" HLP.

There are also areas of potential new applications for HLPs beyond transportation including social media networks, which can involve large scale multi-layered network design problems with the potential of re-routing and controlling the O–D traffic. The rapid expansion of wireless and mobile networks can provide new opportunities for large scale hub modeling, as do the developments with IoT (Internet of Things) networks on a smaller scale.

5. State-of-the-art formulations and solution algorithms

In this section we provide a concise overview of the most important mathematical programming formulations for HLPs. We start by describing commonly considered modeling assumptions and associated properties and discuss how these impact the formulation of HLPs. We then present several formulations for both single and multiple allocation variants, which have as key decisions the location of a set of hub facilities and the allocation of O/D nodes to these facilities. A key feature of such HLPs is that most design and routing decisions are implicitly determined by the assumptions made on the cost structure and the allocation pattern. We then focus on more general class of problems, denoted as *hub network design problems* (HNDPs), which explicitly consider network design decisions and non-trivial routing decisions.

5.1. Modeling assumptions and properties

Four assumptions that underlie classical HLPs are: (i) flows have to be routed via a set of hubs, (ii) arcs have no set-up cost, (iii) a common discount factor α is used for all inter-hub arcs and α is independent of the flows, and (iv) distances d_{ij} satisfy the triangle inequality. The consequences of these assumptions are:

- O–D paths must include at least one hub node. (Assumption (i) is actually rather mild, as it is always possible to add a dummy hub and associated flow costs to represent direct connections between non-hub nodes.)
- Hubs can be interconnected at no extra cost and the set of inter-hub arcs defines a complete subgraph on the hub nodes. Thus, hub arcs are determined by the location of the hub nodes.
- All O–D paths contain at least one and at most two hubs. Note that if Assumption (iv) is not satisfied, then paths may contain more than two hubs and more than one inter-hub arc.

The above properties simplify the network design decisions in classic HLPs, providing a *polynomial* (in size) characterization of all O–D paths. In HLPs, O–D paths include either a single hub node and no inter-hub arc, or two hub nodes and a single inter-hub arc. Moreover, because of Assumptions (ii) and (iv), each collection and distribution (spoke) leg, if present, contains only one access arc. Thus, the O–D paths are of the form (i, k, m, j), where $(k, m) \in N \times N$ is the ordered pair of hubs to which i and j are allocated, respectively, and N is the set of O/D nodes that are also potential hub locations, where $|N| = n$. Note that these paths contain at most three arcs, depending on the number of visited hubs and whether or not the O/Ds are also hubs. This characterization of O–D paths contains only $O(n^2)$ paths per O–D pair and allows the development of compact formulations with $O(n^4)$ variables that explicitly consider all these paths and do not require the use of flow conservation constraints.

In the case of more general HNDPs, O–D paths are more involved because the inter-hub network may be incomplete or include non-discounted bridge arcs (i.e. an arc connecting two hubs with a cost rate equal to the one of an access arcs). Therefore, O–D paths may consist of more than three arcs and visit more than two hub nodes, and include several bridge and hub arcs, depending on the setting. This means that a much larger number of O–D paths may exist, and as a consequence, path-based formulations for HNDP could have up to $O(n^{n-2})$ variables, and flow conservation constraints are needed when extending arc-based formulations of HLPs, which contain $O(n^4)$ variables.

5.2. Hub location problems

HLPs consider the location of hub facilities and the allocation of nodes to open hubs. Three allocation strategies have been investigated: single allocation, multiple allocation, and r-allocation. In single allocation HLPs, each node must be assigned to exactly one hub facility. As a consequence, all flow with the same origin (or destination) is routed via the same access arc. In the case of multiple allocation HLPs, nodes can be assigned to more than one hub. This provides greater flexibility in the routing of flows and allows lower flow cost solutions. However, this could increase the network design cost as a larger number of links must be activated. Finally, in r-allocation HLPs each node can be connected to at most r hubs. This strategy generalizes both single and multiple allocation models.

We next present an overview of the most promising formulations and solution algorithms for both single allocation HLPs and multiple allocation HLPs. We refer the reader to Yaman (2011) and Corberan et al. (2019) for formulations and algorithms for r-allocation HLPs.

5.2.1. Single allocation HLPs

For each node pair $i, j \in N$, the parameter W_{ij} denotes the amount of demand to be routed from node i to j. Additionally, χ and δ represent the collection and distribution cost rates, respectively. We define binary location/allocation variables z_{ik} for each pair $i, k \in N$, with z_{ik} is equal to one if node i is assigned to hub k and zero otherwise. When $i = k$, variable z_{kk} represents the establishment or not of a hub at node k. Using these variables, the *uncapacitated single allocation hub location problem* (USAHLP) can be stated as a quadratic extension of a discrete location problem, where the quadratic term corresponds to the routing cost between hubs (O'Kelly, 1987a).

The most commonly used approach to handle the quadratic term in objective is to define the variables $x_{ijkm} = z_{ik}z_{jm}$, $i, j, k, m \in N$, equal to 1 if and only if the flow originated at i and destination j transits via a first hub node k and a second hub node m. Several families of constraints have been studied to ensure a valid relationship among the x and z variables. The most promising one so far is the one proposed in Skorin-Kapov, Skorin-Kapov, and O'Kelly (1996). This formulation is usually referred to as a *path-based formulation* given that it uses path variables x_{ijkm} to characterize all O–D paths visiting either one or two hub nodes. Due to the particular structure of fully interconnected hub-level networks, it also coincides with *arc-based formulations* commonly used in network optimization. Moreover, it is known to provide tight linear programming (LP) relaxation bounds.

This formulation has been exploited by decomposition methods capable of handling the large number of variables ($O(n^4)$) and constraints ($O(n^3)$). Lagrangean relaxation (LR) in which the linking constraints between the z and the x variables are relaxed to approximately solve single allocation p-hub median problems is a common solution approach (Contreras, Diaz, & Fernández, 2009a; Elhedhli & Wu, 2010; Pirkul & Schilling, 1998). Contreras, Diaz, and Fernández (2011d) combines LR and column generation techniques within a branch-and-price framework to optimally solve single-level capacitated instances with up to 200 nodes. These studies have shown that in order to solve large-scale instances, additional algorithmic refinements are needed. In particular, initial heuristics, in combination with variable fixing and partial enumeration techniques, are particularly effective in considerably reducing the size of the formulation by permanently fixing candidate hubs to be open or closed. Note that for each potential hub facility that is permanently closed, one can eliminate $O(n^3)$ routing variables, $O(n)$ allocation variables, and $O(n^2)$ constraints. For example, if facility k is permanently closed, we can set all x_{ijkm} and x_{ijmk} for all $i, j, m \in N$ to zero, given that any O–D path having k as a first or second hub will no longer be available.

Other linearization strategies have also been used to handle the quadratic term in objective. In particular, An et al. (2015) uses a compact linear reformulation in which only a linear number of additional variables and constraints are used to linearize the multiplication of binary variables (see also, Azizi et al., 2016; Azizi, Vidyarthi, & Chauhan, 2018; Chaovalitwongse, Pardalos, & Prokopyev, 2004; Sherali & Smith, 2007). However, we do not know how the quality of the associated LP relaxation bounds compares with the LP bounds obtained from the path-based formulation.

An alternative to handle the large number of x_{ijkm} variables is to project them out from the path-based formulation. One way is by using Benders decomposition (BD) to obtain a valid formulation in the space of the original z_{ik} variables. In particular, the *Benders* *reformulation* of the USAHLP is:

$$\text{minimize} \quad \sum_{k\in N} f_k z_{kk} + \sum_{i,k\in N} c_{ij} z_{ik} + \eta$$

$$\text{subject to} \quad \sum_{k\in N} z_{ik} = 1 \qquad i \in N \tag{1}$$

$$z_{ik} \le z_{kk} \qquad i, k \in N \tag{2}$$

$$\sum_{k\in N} z_{kk} \ge 1 \tag{3}$$

$$\eta \ge \sum_{i,k\in N} a^r_{ik} z_{ik} \qquad r = 1, \dots, |P_D|, \tag{4}$$

$$z_{ik} \in \{0, 1\} \qquad i, k \in N, \tag{5}$$

where P_D is the set of extreme points of the dual subproblem associated with the linking constraints

$$\sum_{m\in N} x_{ijkm} = z_{ik} \qquad i, j, k \in N$$

$$\sum_{k\in N} x_{ijkm} = z_{jm} \qquad i, j, m \in N,$$

$$x_{ijkm} \ge 0 \qquad i, j, k, m \in N,$$

η is an artificial variable used to evaluate the transportation cost between hub nodes, and a^r_{ik} are the coefficients of the Benders optimality cuts obtained from an optimal dual solution associated with the extreme point r. Even though there is an exponential number of constraints (4), these can be efficiently separated with ad-hoc algorithms that resort to the solution of LPs and network flow problems. Camargo et al. (2011) and Camargo and de Miranda (2012) first introduced this class of Benders reformulations for solving single allocation HLPs considering hub congestion costs, using a hybrid outer-approximation/Benders decomposition algorithm for dealing with the nonlinearity caused by the functions used to represent congestion at hubs. Contreras (2020) use this Benders reformulation within a branch-and-cut framework to optimally solve uncapacitated and capacitated instances with up to 1,000 nodes. Similar to the path-based formulations, the use of heuristics together with variable fixing techniques play an important role in solving large-scale instances. Another critical aspect of the efficiency of cutting plane algorithms based on Benders reformulation relies on the generation of non-dominated Benders cuts, which can can be quickly separated by solving *parametric minimum cost flow problems*.

Another way of projecting out the x_{ijkm} variables was given in Labbé and Yaman (2004) and Labbé, Yaman, and Gourdin (2005), in which continuous flow variables associated with the arcs, together with an exponential number of constraints, are used to model inter-hub flow costs. The authors provide several classes of facet defining inequalities for these projected formulations and use them within a branch-and-cut framework for solving HLPs with single allocation and quadratic capacity constraints for instances with up to 50 nodes.

An alternative way to handle the quadratic term of objective is given in Rostami, Errico, and Lodi (2019) In particular, for each i, $j \in N$, $i \neq j$, the quadratic term can be written as

$$\sum_{k,m\in N} F_{ijkm} z_{ik} z_{jm} = \sum_{k\in N} y^k_{ij} z_{ik} + \sum_{m\in N} y^m_{ji} z_{jm},$$

where y^k_{ij} are auxiliary continuous (unrestricted) variables used to evaluate the routing cost from i to j. In particular, for each i, j, k, $m \in N$, $i \neq j$, the y variables must satisfy

$$y^k_{ij} + y^m_{ji} \ge F_{ijkm}.$$

Using these relationships, the USAHLP can be stated as the following convex mixed integer nonlinear program (MINLP):

$$\text{minimize} \quad \sum_{k\in N} f_k z_{kk} + \sum_{i,k\in N} c_{ij} z_{ik} + \eta$$

$$\text{subject to} \quad (1), (2), (5)$$

$$\eta \ge \Phi(z), \tag{6}$$

where,

$$\Phi(z) = \min \sum_{i\in N} \sum_{j\in N} \left(\sum_{k\in N} y^k_{ij} z_{ik} + \sum_{m\in N} y^m_{ji} z_{jm} \right)$$

$$\text{s.t.} \quad y^k_{ij} + y^m_{ji} \ge F_{ijkm} \qquad \forall i, j, k, m \in N, i \neq j.$$

Rostami et al. (2019) use an outer approximation method to linearize this convex MINLP, which is then solved using a branch-and-cut algorithm. The authors show that this reformulation is equivalent to applying the level-1 *reformulation linearization technique* (RLT) of Adams and Sherali (1990). Given that the path-based formulation is actually a relaxation of the RLT reformulation (see, Contreras, Zetina, Jayaswal, & Vidyarthi, 2020), this formulation provides the strongest reformulation (in terms of quality of linear programming relaxation bounds) known to date for single allocation HLPs. Rostami et al. (2019) present computational results in which instances with up to 150 nodes can be solved in reasonable computing times.

Flow-based formulations, commonly used in multi-commodity network design problems, have also been employed for formulating single allocation HLPs. They use continuous variables to compute the amount of flow that originated at a given node which is routed on a particular arc. In the case of single allocation, we only need to use one set of flow variables associated with the hub arcs, i.e., Y_{ikm}, i, j, $k \in N$, equal to the amount of flow originated at node i and passing through hub arc (k, m). The USAHLP can be formulated as follows (Ernst & Krishnamoorthy, 1996):

$$\text{minimize} \quad \sum_{k\in N} f_k z_{kk} + \sum_{i,k\in N} c_{ik} z_{ik} + \sum_{i,k,m\in N} \alpha d_{km} Y_{ikm}$$

$$\text{subject to} \quad (1), (2), (5)$$

$$\sum_{j\in N} W_{ij} z_{jk} + \sum_{m\in N} Y_{ikm} = \sum_{m\in N} Y_{imk} + O_i z_{ik} \qquad i, k \in N \tag{7}$$

$$Y_{ikm} \ge 0 \qquad i, k, m \in N. \tag{8}$$

The above formulation contains fewer variables ($O(n^3)$) and constraints ($O(n^2)$) as compared to the path-based formulation, at the expense of producing weaker LP bounds. Contreras et al. (2010) and Contreras et al. (2017) present some families of extended cut-set inequalities that can help improve the LP bounds. A positive aspect of this formulation is that it can be used directly in a general purpose solver to obtain optimal solutions for several HLP variants when considering small to medium-size instances. This is not the case for the previously described formulations, given that they either require ad hoc algorithms to be handled or are too big to be loaded directly into a solver. This formulation has been used to model many extensions of single allocation HLPs such as capacity decisions and balancing requirements (e.g., Correia, Nickel, & Saldanha-da Gama, 2010; Correia, Nickel, & Saldanha-da Gama, 2011).

5.2.2. Multiple allocation HLPs

Path-based formulations are a natural approach for multiple allocation HLPs. Given that O/D nodes can be connected to more than one hub facility, one can exploit some properties on the structure of O–D paths with preprocessing techniques that are not applicable to single allocation models. This preprocessing can significantly reduce the number of required variables to formulate multiple allocation variants. In particular, it is known that every flow uses at most one direction of a hub arc, the one with lower flow cost (see Hamacher, Labbé, Nickel, & Sonneborn, 2004). We thus define an *undirected* flow cost F_{ije} for each $e = (k, m) \in E$ and $i, j \in N$ as $F_{ije} = \min\{F_{ijkm}, F_{ijmk}\}$. The number of variables can be further reduced by defining a set of candidate hub arcs E_{ij} for each O–D pair (Contreras, Cordeau, & Laporte, 2011a). This is done by using the property that no flow will be routed through a hub arc containing two different hubs whenever it is cheaper to route it through only one of them (Boland, Krishnamoorthy, Ernst, & Ebery, 2004; Marín, 2005). We also define binary location variables Z_i, $i \in N$, equal to 1 if and only if a hub is located at node i. The *uncapacitated multiple allocation hub location problem* (UMAHLP) can be stated as follows (Hamacher et al., 2004; Marín, 2005):

$$\text{minimize} \quad \sum_{k\in N} f_k Z_k + \sum_{i,j\in N} \sum_{e\in E_{ij}} F_{ije} x_{ije}$$

$$\text{subject to} \quad \sum_{e\in E_{ij}} x_{ije} = 1 \qquad i, j \in N \tag{9}$$

$$\sum_{e\in E_{ij}: k\in e} x_{ije} \le Z_k \qquad i, j, k \in N \tag{10}$$

$$x_{ije} \ge 0 \qquad i, j \in N, e \in E_{ij} \tag{11}$$

$$Z_k \in \{0, 1\} \qquad k \in N. \tag{12}$$

This formulation has $O(n^4)$ variables and $O(n^3)$ constraints and usually provides tight LP bounds. Hamacher et al. (2004) and Marín (2005) independently prove that constraints (10) are indeed facet-defining inequalities. Marín (2005) and Corberan et al. (2019) provide other classes of inequalities which also define facets by reformulating the problem as a set packing problem in an auxiliary graph.

In Marín (2005) the above formulation is used in a relax-and-cut algorithm, whereas in Cánovas, Garcia, and Marín (2007) a dual-ascent technique is used to obtain optimal solutions with up to 120 nodes. Contreras et al. (2011b) use an extension of this path-based formulation to model dynamic (or multi-period) HLPs, and develop a branch-and-bound algorithm that relies on a LR that relaxes the linking constraints (10) to optimally solve instances with up to 100 nodes and 10 time periods.

Similar to single allocation HLPs, the x_{ije} variables can be projected out from the path-based formulation via Benders decomposition to obtain a valid formulation in the space of the binary variables Z_i. The Benders reformulation of the UMAHLP is:

$$\text{minimize} \quad \sum_{k\in N} f_k Z_k + \eta$$

$$\text{subject to} \quad (12)$$

$$\sum_{k\in N} Z_k \ge 1 \tag{13}$$

$$\eta \ge \sum_{i\in N} a_i^r Z_i \qquad r = 1, \ldots, |Q_D|, \tag{14}$$

where Q_D is the set of extreme points of the dual subproblem associated with constraints (9)–(11).

Camargo, de Miranda, and Luna (2008) presents an iterative BD algorithm, based on the above Benders reformulation (but without using the preprocessing), to solve instances of the UMAHLP with up to 200 nodes. Camargo et al. (2009) presents a Benders reformulation for a complex HLP with flow dependent discounted costs. Contreras et al. (2011a) employ the Benders reformulation given above and embed it in an iterative BD that is capable of solving instances for the UMAHLP with up to 500 nodes. Once more, the use of several algorithmic features such as a heuristic, elimination tests, and the generation of non-dominated Benders cuts plays a crucial role in the solution of large-scale instances. This BD algorithm was extended to solve multi-level capacitated instances with up to 300 nodes (Contreras, Cordeau, & Laporte, 2012), and stochastic problems dealing with uncertainties in both demand flows and transportation costs (Contreras, Cordeau, & Laporte, 2011c). Other recent and successful implementations of BD algorithm for multiple allocation HLPs include (Martins de Sá, Morabito, & Camargo, 2018; Mokhtar, Krishnamoorthy, & Ernst, 2018; Taherkhani, Alumur, & Hosseini, 2020).

Flow-based formulations can also be used to model multiple allocation HLPs. However, we now need additional flow variables for the collection and distribution legs. For $i, j, m \in N$, let X_{ijm} be equal to the amount of flow from hub m to destination j that originates at node i. Also, for each $i, k \in N$, let Z_{ik} be equal to the amount of flow from origin node i to hub k. Using these sets of decision variables, we can formulate the UMAHLP as follows (Ernst & Krishnamoorthy, 1998a):

$$\text{minimize} \quad \sum_{k\in N} f_k Z_k + \sum_{i,k\in N} \chi d_{ik} Z_{ik} + \sum_{i,k,m\in N} \alpha d_{km} Y_{ikm} + \sum_{i,j,m\in N} \delta d_{jm} X_{ijm}$$

$$\text{subject to} \quad (12)\text{–}(8)$$

$$\sum_{k\in N} Z_{ik} = O_i \qquad i \in N \tag{15}$$

$$\sum_{m\in N} X_{ijm} = W_{ij} \qquad i, j \in N \tag{16}$$

$$Z_{ik} + \sum_{m\in N} Y_{ikm} = \sum_{m\in N} Y_{imk} + \sum_{j\in N} X_{ijk} \qquad i, k \in N \tag{17}$$

$$Z_{ik}, X_{ijm} \ge 0 \qquad i, j, m \in N. \tag{18}$$

The above formulation contains $O(n^3)$ variables and $O(n^2)$ constraints. Boland et al. (2004) presents some preprocessing procedures that can be used to reduce the number of variables and constraints, and some valid inequalities to improve the LP bounds of capacitated variants.

Radius-based formulations, successfully used for modeling well-known discrete location problems (see, García, Labbé, & Marín, 2011), have also been used to formulate multiple allocation HLPs. For each O/D pair (i, j), we define continuous decision variables η_{ij} to evaluate the flow cost from i to j, and for each $e \in E$ we define binary hub arc location variables y_e, equal to 1 if and only if a hub arc is located at link $e \in E$. For each $i, j \in N$, we order the elements of E_{ij} by non-decreasing values of their coefficients F_{ije}, and we denote e_{ijr} as the r^{th} element according to that ordering. That is, $F_{ije_1} \le \cdots \le F_{ije_{|E_{ij}|}} \le F_{e_{ij|E_{ij}|+1}}$, where $F_{e_{ij|E|+1}} = F_{ije^*}$ is the cost for the fictitious edge e^* such that (i) $F_{ije^*} > \max_{e\in E} F_{ije}$, for all $i, j \in N$; and (ii) $\sum_{i,j\in N} F_{ije^*} > \max_{e\in E}(f_e + \sum_{i,j\in N} F_{ije})$. This assumption guarantees that at least one hub variable y_e is at value one in any optimal solution. The UHLPMA can be stated as the following MIP (García, Landete, & Marín, 2012):

$$\text{minimize} \quad \sum_{k\in N} f_k Z_k + \sum_{i,j\in N} \eta_{ij}$$

$$\text{subject to} \quad \eta_{ij} \ge F_{ije_r} + \sum_{e\in E} (F_{ije} - F_{ije_r})^{-} y_{km} \qquad r = 1, \ldots, |E| + 1, \;\; i, j \in N \tag{19}$$

$$y_{km} \le Z_k \quad (k, m) \in E \tag{20}$$

$$y_{km} \le Z_m \quad (k, m) \in E \tag{21}$$

$$y_{km}, Z_i \in \{0, 1\} \; (k, m) \in E, i \in N, \tag{22}$$

where $(x)^{-} = \min\{0, x\}$. This new formulation has $O(n^2)$ variables and $O(n^4)$ constraints.

We note that the above radius-based formulation coincides with the *supermodular formulation* of Contreras and Fernández (2014). The latter formulation exploits the supermodular properties of a combinatorial representation of a more general class of multiple allocation HLPs that satisfies submodularity. García et al. (2012) present some families of valid inequalities that can be used to strengthen the LP bounds of the radius-based formulation.

5.3. Hub network design problems

Formulating and solving HNDPs presents an even bigger challenge than for HLPs, due to the fact that HNDPs involve additional design decisions for link activation, as well as non-trivial routing decisions. Further, it is no longer evident to see HNDPs as quadratic extensions of FLPs but rather as extensions of *multi-commodity network design problems* (Magnanti & Wong, 1984) with additional location decisions for hubs. This class of network optimization problems are known to be significantly more difficult to solve than FLPs in practice, and one of the main reasons is that flow conservation constraints and additional design variables for link selection decisions may be needed to extend both the flow and arc-based formulations. This is known to have a negative effect on the quality of the LP relaxation bounds associated with these formulations.

Hub arc location problems (HALPs) are a class of HNDPs focusing on the establishment of hub arcs (with discounted costs), rather than purely on the location of hub nodes (Campbell et al., 2005a; 2005b). While both single and multiple allocation HALPs exist, a fundamental feature of HALPs is that the hubs (i.e. endpoints of hub arcs) need no longer be fully interconnected. Contrary to HLPs, hub arcs are established only when justified by the objective (e.g., substantial savings in the transportation cost).

Contreras and Fernández (2014) show how a general class of multiple allocation HALPs containing setup costs and cardinality constraints on hub arcs can be stated as the minimization of a supermodular function. This allows the development of effective supermodular formulations that, when embedded in a branch-and-cut framework, are capable of solving instances with up to 125 nodes. Other research on HALPs considers competition (Gelareh et al., 2010; Sasaki, Campbell, Krishnamoorthy, & Ernst, 2014), flow dependent costs (Camargo, de Miranda, O'Kelly, & Campbell, 2017; Tanash et al., 2017), solution via Benders decomposition (Gelareh et al., 2015; Gelareh & Nickel, 2011), and hub arc capacities (Rothenbcher, Drexl, & Irnich, 2016).

A variety of specific inter-hub network topologies have been examined for hub location problems. Tree-star hub networks (hubs are connected by a tree with single allocation for non-hubs) are examined by Contreras et al. (2009b, 2010) and Martins de Sá et al. (2013), and the Benders decomposition algorithm in Martins de Sá et al. (2013) optimally solves instances with up to 100 nodes. Star-star networks (hubs are directly connected to a central hub with single allocation for non-hubs) are explored by Labbé and Yaman (2008) and Yaman (2008), including a polyhedral study on two different formulations and LR algorithms. Martins de Sá et al. (2015a, 2015b) focus on designing a multiple allocation network where hubs are connected with one or more hub lines. The authors present a Benders reformulation and a branch-and-cut algorithm to solve instances with up to 100 nodes. Contreras et al. (2017) study cycle-star hub networks in which the hubs are connected with a cycle with single allocation of non-hubs, and develop a branch-and-cut algorithm that uses a flow-based formulation in combination with two families of mixed dicut inequalities to solve instances with up to 100 nodes.

Limited work has been done for designing more general hub networks that do not assume a given allocation pattern. In this case, non-trivial design decisions related to the access-level network are included as part of the decision-making process. Thomadsen and Larsen (2007) and Catanzaro et al. (2011) are two examples. Camargo et al. (2013) and Rodríguez-Martín et al. (2014) study hub location-routing problems (a type of HNDP) using Benders decomposition and a branch-and-cut framework, respectively. Bernardes Real et al. (2020) study a more general class of HNDP in which hub nodes are connected via a set of flexible routes that may contain a mix of hub and non-hub nodes. It is assumed that commodity transfers can only be performed at hubs and that transportation costs are flow-dependent. The authors develop two decomposition strategies: i) a *top down strategy* in which the hub configuration is first defined and then routes are constructed, and ii) a *bottom up strategy* in which the routes are built before the hub configuration is designed.

Alibeyg, Contreras, and Fernandez (2018) study HNDPs with profits in which additional decisions related to the activation of O/D nodes are taken into account. A branch-and-bound algorithm based on a LR of an arc-based formulation is used as a bounding procedure to solve instances with up to 100 nodes.

In summary, there has been significant progress in developing sophisticated solution approaches for a range of HLPs, including some that consider complex features (e.g., flow dependent costs). These often require combinations and customization of algorithms, but can solve very hard problems with hundreds of nodes to optimality. While research should continue to improve and refine these approaches, some other directions for future research include: (1) matheuristics that exploit the network flow structure inherent to hub location and hub network design problems, (2) alternative (strong) formulations and associated polyhedral studies for hub network design, where the challenges stem from modeling complex O–D paths, (3) Branch-and-cut-and-price algorithms for hub network design that rely on the use of path-based formulations with an exponential number of variables, and (4) directly exploiting the nonlinear nature of HLPs.

6. Summary themes

Hub location problems are notable for their appeal to transportation, location and telecom researchers, and for providing formulation and solution issues that continue to challenge researchers and professionals. Great progress has been made in solving a wide range of HLP models of realistic problem sizes (in the largest cases), though many models still incorporate some unrealistic simplifying assumptions or idealizations that depart significantly from the real networks being modeled. An overarching goal for hub location research should be to develop better hub location models, especially more "intelligent" models that decide on the optimal network topology and flows in response to the cost and service pressures, not "artificial" assumptions. But the notion of "better" raises the issue of how to measure model quality. Real-world validity from comparing model results to real-world networks and flows is one approach, but the data is often lacking for such a comparison – and getting the "right" answer does not necessarily mean the model is good (for all data sets). Certainly all models should be subject to internal consistency checks to assess for example, whether the model reflects economies of scale properly. But the appropriate metric for this and the necessary degree of agreement remain unclear and problematic. There is a clear tradeoff for HLPs between solution difficulty and model quality, as capturing real-world details (that are often nonlinear) can greatly increase the computational challenges. The classical hub location models have served their purpose well, to help develop insights and understanding and as stepping stones to more sophisticated models; but they generally fall in the lower quality and "easier" categories. We believe there is plenty of opportunity for developing higher quality models for HLPs that are now amenable to state-of-the art solution approaches.

Below we provide several key themes that we hope to see addressed in the future for better modeling HLPs:

1. *Better model economies of scale*: One issue that continues to vex researchers is modeling of economies of scale. Twenty years ago Bryan (1998) wrote: “To date, models of hub location do not adequately capture the scale economies that accrue due to the bundling of flows.” – and in spite of pockets of impressive progress, this is still a legitimate criticism of many hub location studies. The complex interdependencies in HLPs that impact economies of scale still need to be sorted out. “Good” representations of economies of scale tend to lead to nonlinearities in the models, and consequently, to longer solution times and intractable models for large instances. Therefore, the field would enormously benefit from “elegant”, simple models, that can be solved in short times and yield solutions that are correct in their modeling of economies of scale, although they may be suboptimal. In parallel, more detailed models, though harder to solve, should provide true optimal solutions, even in large instances. A key issue with more detailed and comprehensive models is the need for the modeler to know and use realistic values of the relevant parameters based on the real-world setting being modeled.
2. *Incorporate time in HLP models*: The time dimension has not received the attention it deserves in hub location research, in spite of its capital importance in the design and operation of hub systems. As discussed earlier, time and synchronization play a key role in economies of scale for transportation networks, and time is often the most relevant metric for quality of service in HLPs. Service frequencies and flow synchronization should be considered together with the hub location and network design; however, synchronization can be complex, even for simple topologies, and it also depends on the problem setting (e.g., passengers vs cargo). Research on synchronized vehicle routing problems (VRPs) may provide some fruitful ideas for modeling different types of synchronization in hub location problems (e.g., Bredstrom & Ronnqvist, 2008; Drexl, 2012). There is also a need to better integrate hub location models with service network design research to bridge the strategic and tactical decision levels. This, and other models that bridge long- and short-term decisions, require managing the time scale differences between the different decisions.
   Time is also an essential part of accurately modeling demand as real-world demands are not static. In response to dynamic demand, the node and arc capacities should change over time to reflect redeployment of resources (e.g., moving vehicles) and/or more strategic capacity management strategies that provide flexibility. Further, allowing hub locations to change dynamically (e.g., having spoke nodes become temporary hubs) may be beneficial, and research on dynamic hub location and hub location with disruptions may suggest directions for future research (e.g., Alumur et al., 2016; Contreras et al., 2011b; Gelareh et al., 2015; Torkestani, Seyedhosseini, Makui, & Shahanaghi, 2018).
3. *Consider more sophisticated objectives and multiple criteria:* Consideration of multiple criteria in the areas of cost, service and sustainability is an important need for better modeling HLPs. While practical logistics network design problems involve the balancing of multiple objectives, multi-criteria approaches for HLPs have not been studied extensively. In most studies, time-based service metrics, appear as (hard) constraints with an objective based on minimizing cost. Given the large (continental or global) scale of many hub networks, sustainability issues, in concert with traditional cost and service measures, deserve more attention. The interconnections between cost, service and sustainability require careful modeling, and each dimension may be quite different for passenger, cargo and telecom networks due to the fundamentally different natures of the traffic. As usual for transportation systems, some key issues are properly allocating costs between the system users and service providers, providing multiple levels of service with the same physical network, and properly modeling how the benefits of good service in hub networks accrue differently to the relevant parties (e.g., customers or operators). Competitive models are also important in this area as cost and service pressures are the usual bases for competition.
   There exist several applications for hub models that call for the use of more sophisticated objectives. One example of such an objective function is the ordered median function (see Puerto, Ramos, & Rodriguez-Chia, 2011; Puerto, Ramos, & Rodriguez-Chia, 2013; Puerto, Ramos, Rodriguez-Chia, & Sanchez-Gil, 2016). This can provide a unifying framework for various classes of objectives commonly used in location theory (e.g., minsum, minmax, k-centrum, and trimmed mean), and it also allows modeling of new distribution patterns induced by the different roles of the user within a collaborative hub-based supply chain network. This adds a sorting problem to the standard HLP, making the formulation and solution more challenging.
4. *Relate to real-world problems*: In general, inspiration from real-world service networks will likely prove very fruitful. It is noteworthy that the earliest hub models of O'Kelly appeared shortly after the airline deregulation act in the US – and Campbell's early work too was spurred by trucking deregulation in the US. At that time, transportation carriers were freed to expand and reconfigure their networks, and large-scale telecom networks were growing as well. Today, in many areas, hub networks are well established so that designing systems from scratch may be unlikely. Therefore, more attention is warranted on models that reflect structural changes in industries (e.g., mergers), significant changes in travel patterns, or the deployment of new technologies (e.g., drones). In Location Theory, these models are called *conditional location models*, which study the location of facilities under the condition that there are previously located facilities, either by competing firms (Dasci, 2011; Eiselt & Marianov, 2009) or non-competing entities (Wang, Lin, Lin, & Ku, 2008). A particularly important field of application for conditional location models are mobile telecommunications networks using 5G technology, which require increasing the number of antennas (that act as hubs) by orders of magnitude, while still retaining for some time the existing antennas of older technologies. Another rapidly developing application field is that of wireless networks for the IoT (Internet of Things), where hub-and-spoke topologies may prove essential. In addition, new models are needed for modifications of hub networks and development of flexible and agile networks that accommodate relocation of resources (e.g., truck terminals, runways, airport gates, etc.) and expansion or contraction of routes, nodes and/or hubs in an established network (e.g., so as to respond effectively to dramatic demand changes as from the COVID-19 pandemic).
5. *Integrate hub location with other problems*: While strategic hub location problems have offered a rich set of challenges, expanding hub location modeling to include more tactical issues and more application-specific features will provide new challenges. Especially promising areas include links to the service network design (as noted earlier) and multi-commodity network flow research, and greater detail on modeling particular transportation vehicles (including emissions). More research is also needed that integrates congestion, competition, reliability and robustness driven by real-world needs, including recovery after failures (e.g., for airlines). Integrating hub location and

routing aspects (with links to the rich literature on location-routing problems) is a very promising area for future research, especially to include automated and unmanned vehicles. Given the individual difficulty of hub location problems and vehicle routing problems, the combined hub location-routing problems promise to provide strong computational and methodological challenges for years to come.

6. *Incorporate nature of the demand*: The nature of the demand and how it moves (in trucks, planes, pipelines, ships, wires, drones, etc.) is a key aspect of better modeling HLPs for various settings. The demands for real-world hub networks are uncertain in the longer term and can be highly variable, and seasonal. The time dimension of the demand, including the rate of demand as well as the represented time horizon, is an area that needs more careful attention. Similarly, research should consider the elasticity of demand to price and quality of service (including locations of hubs), as well as opportunities to not serve all the demand (this is usually addressed in some fashion in competitive hub location and profit maximizing models).
7. *Use real data*: To date, a great many hub location studies have used a small group of rather old and in some ways artificial data sets (i.e. CAB, AP, and Turkish datasets). There are strong advantages to using common data sets to provide a common basis for comparison and understanding of results. However, with the advances in formulations and solution methods, new larger data sets based on real-world networks from varied applications (trucking, airlines, ocean shipping, telecom networks, etc.) at a variety of spatial scales (metropolitan, national, continental, global) would provide a rich test bed to "stress test" models. This should help discover some deeper insights into how the underlying characteristics of the hub networks and applications (e.g., spatial distributions of population, or varying levels of fixed and variable costs) are related to the success of various formulation and solution approaches. While artificial data sets (including random data) can be easily generated, there are some strong features in real-world data sets related to the underlying geography and human spatial interaction patterns that are quite un-random. One newer data set that has some different characteristics is the sparse ICD data set based on container shipping in Indonesia (from Mokhtar et al., 2019).
8. *Obtain insights from results*: The goal of hub location modeling is to develop a better understanding of the nature of optimal (or near-optimal) solutions of these challenging problems. To accomplish this, it is important to examine details of solutions beyond just objective function values and cpu times for a range of realistic data sets and realistic parameter settings. Documenting hub locations, allocations and flows is essential to validate models and to identify spatial arrangements that are particularly problematic. Finding metrics that are independent of problem instances (e.g., built from relevant ratios) is a useful and challenging area of research that would allow combining results from disparate instances. (Developing a compact way to visualize allocations and flows in large networks is another challenge.) We believe that the real value from solving a wide range of problem instances and data sets is to reveal the underlying principles that govern optimal or near-optimal design and behavior. Solving a particular set of problem instances efficiently is certainly important, but we see great value in better understanding the general forces that influence the interplay of geography (distance), demand and complex cost structures that is inherent in hub networks.
9. *Use the best solution approaches – and develop new ones*: Development of exact and approximate solution algorithms for solving more complex HLPs is certainly a necessity. There is good progress being made in this area, as shown by the succinct summary in Section 5, and there have been some good insights arising from efforts to elucidate the approaches that are best for single allocation and multiple allocation HLPs. But there is still a need for a deeper understanding why certain formulations and solutions approaches work best. This insight can then be leveraged for approaching related problems. Also note that planar hub location models may provide deeper geometric insights. In many cases, sophisticated algorithmic strategies that combine optimal approaches with initial heuristics, variable fixing and preprocessing techniques work well to solve large-scale problems. Results have documented the benefit of using the proper formulations – and of expending efforts to develop tailored solution algorithms. Even with commercial solvers becoming ever more sophisticated, their performance is still well below that of a carefully designed custom algorithm due to the very challenging nature of the more complex HLPs now being considered.

Finally, we acknowledge that incorporating all nine of these themes in one new model poses significant challenges, but we strongly encourage research in all nine areas noted above. Not surprisingly, several themes here are interrelated, especially the modeling of economies of scale, incorporating time, and better relating models to real-world problems. The early hub location models have served their purpose well, and we believe it is now time for a new generation of better models. Developing and solving new and better HLP models poses a (large) challenge, but we believe such efforts will provide a significant payoff to the research community, and to practice.

Acknowledgments

Alumur and Campbell thank all of their co-authors for graciously sharing their thoughts and expertise. Alumur gratefully acknowledges the support by grant RGPIN-2015-05548 of the Natural Sciences and Engineering Research Council of Canada (NSERC). Contreras gratefully acknowledges the support by grant 2018–06704 of NSERC. O'Kelly thanks Yongha Park and James Campbell for comments on an earlier version. Marianov gratefully acknowledges the support by grants FONDECYT 1190064 and CONICYT-PIA-AFB180003. Marianov is also indebted to Armin Lüer-Villagra and H.A. Eiselt, for the long and insightful discussions about the subject of economies of scale. Complementary material to what is in this paper appears in Lüer-Villagra et al. (2019).

References

Aboytes-Ojeda, M., Castillo-Villar, K. K., & Roni, M. S. (2020). A decomposition approach based on meta-heuristics and exact methods for solving a two-stage stochastic biofuel hub-and-spoke network problem. *Journal of Cleaner Production, 247*, 119176.
Adams, W. P., & Sherali, H. D. (1990). Linearization strategies for a class of zero-one mixed integer programming problems. *Operations Research, 38*(2), 217–226.
Alibeyg, A., Contreras, I., & Fernandez, E. (2016). Hub network design problems with profits. *Transportation Research Part E, 96*, 40–59.
Alibeyg, A., Contreras, I., & Fernandez, E. (2018). Exact solution of hub network design problems with profits. *European Journal of Operational Research, 266*(1), 57–71.
Alumur, S., & Kara, B. (2009a). A hub covering network design problem for cargo applications in Turkey. *Journal of the Operational Research Society, 60*(10), 1349–1359.
Alumur, S., Kara, B., & Karasan, O. (2009b). The design of single allocation incomplete hub networks. *Transportation Research Part B: Methodological, 43*(10), 936–951.
Alumur, S., Kara, B., & Karasan, O. (2012a). Multimodal hub location and hub network design. *Omega, 40*(6), 927–939.
Alumur, S., & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research, 190*(1), 1–21.
Alumur, S. A., Nickel, S., Saldanha-da Gama, F., & Secerdin, Y. (2016). Multi-period hub network design problems with modular capacities. *Annals of Operations Research, 246*(1–2), 289–312.
Alumur, S. A., Nickel, S., Rohrbeck, B., & Saldanha-da Gama, F. (2018). Modeling congestion and service time in hub location problems. *Applied Mathematical Modelling, 55*, 13–32.

Alumur, S. A., Yaman, H., & Kara, B. Y. (2012b). Hierarchical multimodal hub location problem with time-definite deliveries. *Transportation Research Part E: Logistics and Transportation Review, 48*(6), 1107–1120.

An, Y., Zhang, Y., & Zeng, B. (2015). The reliable hub-and-spoke design problem: Models and algorithms. *Transportation Research Part B: Methodological, 77*, 103–122.

Araújo, A. C. A., Roboredo, C. M., Pessoa, A. A., & Pereira, V. (2020). Exact methods for the discrete multiple allocation ($r|p$) hub-centroid problem. *Computers & Operations Research, 116*, 104870.

Azizi, N., Chauhan, S., Salhi, S., & Vidyarthi, N. (2016). The impact of hub failure in hub-and-spoke networks: Mathematical formulations and solution techniques. *Computers & Operations Research, 65*, 174–188.

Azizi, N., Vidyarthi, N., & Chauhan, S. (2018). Modelling and analysis of hub-and-spoke networks under stochastic demand and congestion. *Annals of Operations Research, 264*, 1–2.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience, 20*, 353–364.

Basso, L. J., & Jara-Díaz, S. R. (2006). Are returns to scale with variable network size adequate for transport industry structure analysis? *Transportation Science, 40*(3), 259–268.

Basso, L. J., & Jara-Díaz, S. R. (2006). Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. *Networks and Spatial Economics, 6*(2), 149–162.

Bernardes Real, L., Contreras, I., Cordeau, J.-F., de Camargo, R., & de Miranda, G. (2020). Multimodal hub network design with flexible routing. Submitted to Transportation Research Part E.

Bernardes Real, L., O'Kelly, M., de Miranda, G., & de Camargo, R. (2018). The gateway hub location problem. *Journal of Air Transportation Management, 73*, 95–112.

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

Bredstrom, D., & Ronnqvist, M. (2008). Combined vehicle routing and scheduling with temporal precedence and synchronization constraints. *European Journal of Operational Research, 191*, 19–31.

Bryan, D. L. (1998). Extensions to the hub location problem: Formulations and numerical examples. *Geographical Analysis, 30*(4), 315–330.

Camargo, R. S., & de Miranda Jr. , G. (2012). Single allocation hub location problem under congestion: Network owner and user perspectives. *Expert Systems with Applications, 39*(3), 3385–3391.

Camargo, R. S., de Miranda Jr. , G., & Ferreira, R. (2011). A hybrid outer-approximation/benders decomposition algorithm for the single allocation hub location problem under congestion. *Operations Research Letters, 39*, 329–337.

Camargo, R. S., de Miranda Jr. , G., & Løkketangen, A. (2013). A new formulation and an exact approach for the many-to-many hub location-routing problem. *Applied Mathematical Modelling, 37*(12), 7465–7480.

Camargo, R. S., de Miranda Jr. , G., & Luna, H. P. L. (2009). Benders decomposition for hub location problems with economies of scale. *Transportation Science, 43*(1), 86–97.

Camargo, R. S., de Miranda Jr. , G., O'Kelly, M., & Campbell, J. (2017). Formulations and decomposition methods for the incomplete hub location problem with and without hop-constraints. *Applied Mathematical Modelling, 51*, 274–301.

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

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

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

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

Campbell, J. F. (2013). Modeling economies of scale in transportation hub networks. In *Proceedings of the 46th Hawaii international conference on system sciences* (pp. 1154–1163). IEEE. Doi: 10.1109/HICSS.2013.411

Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2002). Hub location problems. In Z. Drezner, & H. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 373–407). Berlin: Springer.

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

Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2005). Hub arc location problems: Part II – Formulations and optimal algorithms. *Management Science, 51*(10), 1556–1571.

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

Cánovas, L., Garcia, S., & Marín, A. (2007). Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. *European Journal of Operational Research, 179*(3), 990–1007.

Catanzaro, D., Gourdin, E., Labbe, M., & Ozsoy, F. A. (2011). A branch-and-cut algorithm for the partitioning-hub location-routing problem. *Computers & Operations Research, 38*(2), 539–549.

Chaovalitwongse, W., Pardalos, P. M., & Prokopyev, O. A. (2004). A new linearization technique for multi-quadratic 0–1 programming problems. *Operations Research Letters, 32*, 517–522.

Chen, L.-H., Hsieh, S.-Y., Hung, L.-J., & Klasing, R. (2020). Approximation algorithms for the p-hub center routing problem in parameterized metric graphs. *Theoretical Computer Science, 806*, 271–280.

Contreras, I. (2020). Hub network design. In T. Crainic, M. Gendreau, & B. Gendron (Eds.), *Network design with applications to transportation and logistics*. Springer.

Contreras, I., Cordeau, J. F., & Laporte, G. (2011a). Benders decomposition for large-scale uncapacitated hub location. *Operations Research, 59*(6), 1477–1490.

Contreras, I., Cordeau, J. F., & Laporte, G. (2011b). The dynamic uncapacitated hub location problem. *Transportation Science, 45*(1), 18–32.

Contreras, I., Cordeau, J. F., & Laporte, G. (2011c). Stochastic uncapacitated hub location. *European Journal of Operational Research, 212*(3), 518–528.

Contreras, I., Cordeau, J. F., & Laporte, G. (2012). Exact solution of large-scale hub location problems with multiple capacity levels. *Transportation Science, 46*(4), 439–459.

Contreras, I., Diaz, J. A., & Fernández, E. (2009a). Lagrangean relaxation for the capacitated hub location problem with single assignment. *OR Spectrum, 31*(3), 483–505.

Contreras, I., Diaz, J. A., & Fernández, E. (2011d). Branch and price for large-scale capacitated hub location problems with single assignment. *INFORMS Journal on Computing, 23*(1), 41–55.

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

Contreras, I., Fernández, E., & Marín, A. (2009b). Tight bounds from a path based formulation for the tree of hub location problem. *Computers & Operations Research, 36*(12), 3117–3127.

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

Contreras, I., & O'Kelly, M. (2019). Hub location problems. In G. Laporte, S. Nickel, & F. Saldanha-da Gama (Eds.), *Location science* (2nd). Switzerland: Springer.

Contreras, I., Zetina, C., Jayaswal, S., & Vidyarthi, N. (2020). An exact algorithm for large-scale non-convex quadratic capacitated facility location. *Submitted to Operations Research*.

Contreras, I., Tanash, M., & Vidyarthi, N. (2017). Exact and heuristic approaches for the cycle hub location problem. *Annals of Operations Research, 258*, 655–677.

Cook, G. N., & Goodwin, J. (2008). Airline networks: A comparison of hub-and-spoke and point-to point systems. *Journal of Aviation/Aerospace Education & Research, 17*(2), 50–51.

Corberan, A., Landete, M., Peiro, J., & Saldanha-da Gama, F. (2019). Improved polyhedral descriptions and exact procedures for a broad class of uncapacitated p-hub median problems. *Transportation Research Part B: Methodological, 123*, 38–63.

Correia, I., Nickel, S., & Saldanha-da Gama, F. (2010). The capacitated single-allocation hub location problem revisited: A note on a classical formulation. *European Journal of Operational Research, 207*(1), 92–96.

Correia, I., Nickel, S., & Saldanha-da Gama, F. (2011). Hub and spoke network design with single-assignment, capacity decisions and balancing requirements. *Applied Mathematical Modelling, 35*(10), 4841–4851.

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

Daganzo, C. F. (1987). The break-bulk role of terminals in many-to-many logistics networks. *Operations Research, 35*(4), 543–555.

Danach, K. (2017). Service network design: A literature review. *Journal of Computer Engineering & Information Technology, 6*(6). https://doi.org/10.4172/2324-9307.1000188.

Dasci, A. (2011). Conditional location problems on networks and in the plane. In H. Eiselt, & V. Marianov (Eds.), *Foundations of location analysis*. In *International series in operations research & management science: 155*. New York, NY.: Springer.

Drexl, M. (2012). Synchronization in vehicle routing – A survey of VRPs with multiple synchronization constraints. *Transportation Science, 46*(3), 297–316.

Drezner, Z., & Wesolowsky, G. O. (1982). A trajectory approach to the round-trip location problem. *Transportation Science, 16*(1), 56–66.

Dukkanci, O., & Kara, B. Y. (2017). Routing and scheduling decisions in the hierarchical hub location problem. *Computers & Operations Research, 85*, 45–57.

Dukkanci, O., Peker, M., & Kara, B. Y. (2019). Green hub location problem. *Transportation Research Part E: Logistics and Transportation Review, 125(C)*, 116–139.

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

Elhedhli, S., & Hu, F. X. (2005). Hub-and-spoke network design with congestion. *Computers & Operations Research, 32*(6), 1615–1632.

Elhedhli, S., & Wu, H. (2010). A Lagrangean heuristic for hub-and-spoke system design with capacity selection and congestion. *INFORMS Journal on Computing, 22*(2), 282–296.

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

Ernst, A. T., Jiang, H., Krishanmoorthy, M., & Baatar, D. (2018). Reformulations and computational results for the uncapacitated single allocation hub covering problem. In R. Sarker, H. Abbass, S. Dunstall, P. Kilby, R. Davis, & L. Young (Eds.), *Data and decision sciences in action*. In *Lecture notes in management and industrial engineering*. Cham: Springer.

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

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

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

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

Esfahlani, F. Z., Bertolero, M. A., Bassett, D. S., & Betzel, R. F. (2020). Space-independent community and hub structure of functional brain networks. *NeuroImage, 211*, 116612.

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

Fard, M. K., & Alfandari, L. (2019). Trade-offs between the stepwise cost function and its linear approximation for the modular hub location problem. *Computers & Operations Research, 104*, 358–374.

Fernández, E., & Sgalambro, A. (2020). On carriers collaboration in hub location problems. *European Journal of Operational Research, 283*(2), 476–490.

García, S., Labbé, M., & Marín, A. (2011). Solving large *p*-median problems with a radius formulation. *INFORMS Journal on Computing, 23*(4), 493–655.

García, S., Landete, M., & Marín, A. (2012). New formulation and a branch-and-cut algorithm for the multiple allocation *p*-hub median problem. *European Journal of Operational Research, 220*(1), 48–57.

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

Gelareh, S., & Nickel, S. (2011). Hub location problems in transportation networks. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1092–1111.

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

Ghaffarinasab, N., & Atayi, R. (2018). An implicit enumeration algorithm for the hub interdiction median problem with fortification. *European Journal of Operational Research, 267*, 23–39.

Ghaffarinasab, N., & Motallebzadeh, A. (2018). Hub interdiction problem variants: Models and metaheuristic solution algorithms. *European Journal of Operational Research, 267*, 496–512.

Grove, P. G., & O'Kelly, M. E. (1986). Hub networks and simulated schedule delay. *Papers in Regional Science, 59*(1), 103–119.

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

Hamacher, H., & Meyer, T. (2006). Hub cover and hub center problems. *Technical paper nr. 98*. Department of Mathematics, University of Kaiserslautern, Germany.

Han, L., & Zhang, N. (2013). P-hub airline network design incorporating interaction between elastic demand and network structure. In W. E. Wong, & T. Ma (Eds.), *Emerging technologies for information systems, computing, and management*. In *Lecture notes in electrical engineering: 236* (pp. 89–96).

Helme, M. P., & Magnanti, T. L. (1989). Designing satellite communication networks by zero-one quadratic programming. *Networks, 19*(4), 427–450.

Hoff, A., Peiro, J., & Corberán, M. R. (2017). Heuristics for the capacitated modular hub location problem. *Computers & Operations Research, 86*, 94–109.

Horner, M. W., & O'Kelly, M. E. (2001). Embedding economies of scale concepts for hub networks design. *Journal of Transport Geography, 9*, 255–265.

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

Jara-Díaz, S. R., Cortés, C. E., & Morales, G. A. (2013). Explaining changes and trends in the airline industry: Economies of density, multiproduct scale, and spatial scope. *Transportation Research Part E: Logistics and Transportation Review, 60*, 13–26.

Johnston, A., & Ozment, J. (2013). Economies of scale in the US airline industry. *Transportation Research Part E: Logistics and Transportation Review, 51*, 95–108.

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

Kartal, Z., Hasgul, S., & Ernst, A. T. (2017). Single allocation p-hub median location and routing problem with simultaneous pick-up and delivery. *Transportation Research Part E: Logistics and Transportation Review, 108*, 141–159.

Kartal, Z., Krishnamoorthy, M., & Ernst, A. T. (2019). Heuristic algorithms for the single allocation *p*-hub center problem with routing considerations. *OR Spectrum, 41*(1), 99–145.

Khaniyev, T., Elhedhli, S., & Erenay, F. S. (2020). Spatial separability in hub location problems with an application to brain connectivity networks. *INFORMS Journal on Optimization*. https://doi.org/10.1287/ijoo.2019.0031.

Kim, H., & O'Kelly, M. (2009). Reliable *p*-hub location problems in telecommunication networks. *Geographical Analysis, 41*(3), 283–306.

Kimms, A. (2006). Economies of scale in hub & spoke network design models: We have it all wrong. In M. Morlock, C. Schwindt, N. Trautmann, & J. Zimmermann (Eds.), *Perspectives on operations research* (pp. 293–317). Weisbaden: Springer.

Klincewicz, J. (1998). Hub location in backbone/tributary network design: A review. *Location Science, 6*(1), 307–335.

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

Labbé, M., & Yaman, H. (2004). Projecting the flow variables for hub location problems. *Networks, 44*(2), 84–93.

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

Labbé, M., Yaman, H., & Gourdin, E. (2005). A branch and cut algorithm for hub location problems with single assignment. *Mathematical Programming, 102*(2), 371–405.

Lee, C.-H., Ro, H.-B., & Tcha, D.-W. (1993). Topological design of a two-level network with ring-star configuration. *Computers & Operations Research, 20*, 625–637.

Lee, H. L., Rammohan, S. V., & Sept, L. (2013). Innovative logistics in extreme conditions: The case of health care delivery in gambia. In J. H. Bookbinder (Ed.), *Handbook of global logistics*. In *International series in operations research & management science: 181*. New York: Springer.

Lee, Y., Lim, B. L., & Park, J. S. (1996). A hub location problem in designing digital data service networks: Lagrangian relaxation approach. *Location Science, 4*, 185–194.

Lei, T. L. (2013). Identifying critical facilities in hub-and-spoke networks: A hub interdiction median problem. *Geographical Analysis, 45*(2), 105–122.

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

Lüer-Villagra, A., Eiselt, H. A., & Marianov, V. (2019). A single allocation p-hub median problem with general piecewise-linear costs in arcs. *Computers & Industrial Engineering, 128*, 477–491.

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

Macias, J. E., Angeloudis, P., & Ochieng, W. (2020). Optimal hub selection for rapid medical deliveries using unmanned aerial vehicles. *Transportation Research Part C: Emerging Technologies, 110*, 56–80.

Magnanti, T. L., & Wong, R. T. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science, 18*(1), 1–55.

Mahmutogullari, A., & Kara, B. Y. (2016). Hub location under competition. *European Journal of Operational Research, 250*(1), 214–225.

Marianov, V., Gutiérrez-Jarpa, G., Obreque, C., & Cornejo, O. (2012). Lagrangean relaxation heuristics for the p-cable-trench problem. *Computers & Operations Research, 39*, 620–628.

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

Marín, A. (2005). Uncapacitated euclidean hub location: Strengthened formulation, new facets and a relax-and-cut algorithm. *Journal of Global Optimization, 33*(3), 393–422.

Martins de Sá, E. M., Camargo, R. S., & Miranda, G. (2013). An improved benders decomposition algorithm for the tree of hubs location problem. *European Journal of Operational Research, 226*(2), 185–202.

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

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

Martins de Sá, E. M., Morabito, R., & Camargo, R. S. (2018). Benders decomposition applied to a robust multiple allocation incomplete hub location problem. *Computers & Operations Research, 89*, 31–50.

Masaeli, M., Alumur, S. A., & Bookbinder, J. H. (2018). Shipment scheduling in hub location problems. *Transportation Research Part B: Methodological, 115*, 126–142.

Meier, J. F. (2017). An improved mixed integer program for single allocation hub location problems with stepwise cost function. *International Transactions in Operational Research, 24*, 983–991.

Meier, J. F., & Clausen, U. (2017). Solving single allocation hub location problems on euclidean data. *Transportation Science, 52*(5), 1035–1296.

Miele, G. M., Caton, L., Freese, T. E., Mc Govern, M., Darfler, K., Antonini, V. P., … Rawson, R. (2020). Implementation of the hub and spoke model for opioid use disorders in california: Rationale, design and anticipated impact. *Journal of Substance Abuse Treatment, 108*, 20–25.

Mirzapour-Kamanaj, A., Majidi, M., Zare, K., & Kazemzadeh, R. (2020). Optimal strategic coordination of distribution networks and interconnected energy hubs: A linear multi-follower bi-level optimization model. *International Journal of Electrical Power & Energy Systems, 119*, 105925.

Mohammadi, M., Jule, P., & Tavakkoli-Moghaddam, R. (2019). Reliable single-allocation hub location problem with disruptions. *Transportation Research Part E: Logistics and Transportation Review, 123*, 90–120.

Mokhtar, H., Krishnamoorthy, M., & Ernst, A. T. (2018). A modified benders method for the single- and multiple allocation *p*-hub median problems. *Operations Research Proceedings 2017*, 135–141.

Mokhtar, H., Redi, A. A. N. P., Krishnamoorthy, M., & Ernst, A. T. (2019). An intermodal hub location problem for container distribution in indonesia. *Computers & Operations Research, 104*, 415–432.

Morganti, E., & Gonzalez-Feliu, J. (2015). City logistics for perishable products. the case of the Parma's food hub. *Case Studies on Transport Policy, 3*(2), 120–128.

Najy, W., & Diabat, A. (2020). Benders decomposition for multiple-allocation hub-and-spoke network design with economies of scale and node congestion. *Transportation Research Part B: Methodological, 133*, 62–84.

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

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

O'Kelly, M. E. (1987). Solutions for some minimax hub location problems. In *Proceedings of the ISOLDE IV international symposium on locational decisions*. Namur, Belgium

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

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

O'Kelly, M. E., Campbell, J. F., Camargo, R. S., & Miranda, G. (2014). Multiple allocation hub location model with fixed arc costs. *Geographical Analysis, 47*(1), 73–96.
O'Kelly, M. E., Luna, H. P. L., de Camargo, R. S., & de Miranda, J. G. (2015). Hub location problems with price sensitive demands. *Networks and Spatial Economics, 15*(4), 917–945.
O'Kelly, M. E., & Miller, H. J. (1991). Solution strategies for the single facility minimax hub location problem. *Papers in Regional Science, 70*(4), 367–380.
O'Kelly, M. E., & Miller, H. J. (1994). The hub network design problem: A review and synthesis. *Journal of Transport Geography, 2*(1), 31–40.
Oldham, S., & Fornito, A. (2019). The development of brain network hubs. *Developmental Cognitive Neuroscience, 36*, 100607.
Parsa, M., Nookabadi, A. S., Flapper, S. D., & Atan, Z. (2019). Green hub-and-spoke network design for aviation industry. *Journal of Cleaner Production, 229*, 1377–1396.
Pirkul, H., & Schilling, D. A. (1998). An efficient procedure for designing single allocation hub and spoke systems. *Management Science, 44*(12), 235–242.
Podnar, H., Skorin-Kapov, J., & Skorin-Kapov, D. (2002). Network cost minimization using threshold-based discounting. *European Journal of Operational Research, 137*(2), 371–386.
Puerto, J., Ramos, A. B., & Rodriguez-Chia, A. M. (2011). Single-allocation ordered median hub location problems. *Computers & Operations Research, 38*, 559–570.
Puerto, J., Ramos, A. B., & Rodriguez-Chia, A. M. (2013). A specialized branch & bound & cut for single-allocation ordered median hub location problems. *Discrete Applied Mathematics, 161*, 2624–2646.
Puerto, J., Ramos, A. B., Rodriguez-Chia, A. M., & Sanchez-Gil, M. C. (2016). Ordered median hub location problems with capacity constraints. *Transportation Research Part C: Emerging Technologies, 70*, 142–156.
Racunica, I., & Wynter, L. (2005). Optimal location of intermodel freight hubs. *Transportation Research Part B: Methodological, 39*, 453–477.
Ramamoorthy, P., Jayaswal, S., Sinha, A., & Vidyarthi, N. (2019). Multiple allocation hub interdiction and protection problems: Model formulations and solution approaches. *European Journal of Operational Research, 1*, 230–245.
Re Velle, C. S., & Swain, R. (1970). Central facilities location. *Geographical Analysis, 2*, 30–42.
Reif, S., Brolin, M. F., Stewart, M. T., Fuchs, T. J., & Mazel, S. B. (2020). The washington state hub and spoke model to increase access to medication treatment for opioid use disorders. *Journal of Substance Abuse Treatment, 108*, 33–39.
Rieck, J., Ehrenberg, C., & Zimmermann, J. (2014). Many-to-many location-routing with inter-hub transport and multi-commodity pickup-and-delivery. *European Journal of Operational Research, 236*(1), 863–878.
Rodríguez-Martín, I., Salazar-González, J. J., & Yaman, H. (2014). A branch-and-cut algorithm for the hub location and routing problem. *Computers & Operations Research, 50*, 161–174.
Rodríguez-Martín, I., Salazar-González, J. J., & Yaman, H. (2016). The ring *k*-rings network design problem: Model and branch and cut algorithm. *Networks, 68*(2), 130–140.
Rostami, B., & Buchheim, C. (2015). The uncapacitated single allocation *p*-hub median problem with stepwise cost function. Optimization Online. Technical report. http://www.optimization-online.org/DB_FILE/2015/07/5044.pdf.
Rostami, B., Errico, F., & Lodi, A. (2019). A convex reformulation and an outer approximation for a class of binary quadratic program. Data science for real-time decision-making. Technical report, DS4DM-2-18-002. https://cerc-datascience.polymtl.ca/wp-content/uploads/2018/03/Technical-Report_DS4DM-2018-002.pdf.
Rostami, B., Kämmerling, N., Buchheim, C., & Clausen, U. (2018). Reliable single allocation hub location problem under hub breakdowns. *Computers & Operations Research, 96*, 15–29.
Rothenbcher, A.-K., Drexl, M., & Irnich, S. (2016). Branch-and-price-and-cut for a service network design and hub location problem. *European Journal of Operational Research, 255*(3), 935–947.
Saito, H., Fujie, T., Matsui, T., & Matuura, S. (2009). A study of the quadratic semi-assignment polytope. *Discrete Optimization, 6*(1), 37–50.
Sasaki, M., Campbell, J. F., Krishnamoorthy, M., & Ernst, A. (2014). A Stackelberg hub arc location model for a competitive environment. *Computers & Operations Research, 47*, 27–41.
Serper, E. Z., & Alumur, S. A. (2016). The design of capacitated intermodal hub networks with different vehicle types. *Transportation Research Part B: Methodological, 86*, 51–65.
Sherali, H. D., & Smith, J. C. (2007). An improved linearization strategy for zero-one quadratic programming problems. *Optimization Letters, 1*, 33–47.
Sim, T., Lowe, T. J., & Thomas, B. W. (2009). The stochastic *p*-hub center problem with service-level constraints. *Computers & Operations Research, 36*(12), 3166–3177.
Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated *p*-hub median problemsx. *European Journal of Operational Research, 94*(3), 582–593.
Smith, H., Cakebread, D., Battarra, M., Shelbourne, B., & Cassim, N. (2017). Location of a hierarchy of HIV/AIDS test laboratories in an inbound hub network: case study in South Africa. *Journal of the Operational Research Society, 68*, 1068–1081.
Snickars, F. (1978). Convexity and duality properties of a quadratic intraregional location model. *Regional Science and Urban Economics, 8*(1), 5–19.
Soylu, B., & Katip, H. (2019). A multiobjective hub-airport location problem for an airline network design. *European Journal of Operational Research, 277*, 412–425.
Taherkhani, G., & Alumur, S. A. (2019). Profit maximizing hub location problems. *Omega, 86*, 1–15.
Taherkhani, G., Alumur, S. A., & Hosseini, S. M. (2020). Benders decomposition for profit maximizing capacitated hub location problems with multiple demand classes. *Transportation Science*. https://doi.org/10.1287/trsc.2020.1003.
Tan, P. Z., & Kara, B. Y. (2007). A hub covering model for cargo delivery systems. *Networks, 49*(1), 28–39.
Tanash, M., Contreras, I., & Vidyarthi, N. (2017). An exact algorithm for the modular hub location problem with single assignments. *Computers & Operations Research, 85*, 32–44.
Taner, M. R., & Kara, B. Y. (2016). Endogenous effects of hubbing on flow intensities. *Networks and Spatial Economics, 16*(4), 1151–1181.
Tavassoli, K., & Tamannae, M. (2020). Hub network design for integrated bike-and-ride services: A competitive approach to reducing automobile dependence. *Journal of Cleaner Production, 248*(1), 1–24.
Thomadsen, T., & Larsen, J. (2007). A hub location problem with fully interconnected backbone and access networks. *Computers & Operations Research, 34*(8), 2520–2531.
Torkestani, S. S., Seyedhosseini, S. M., Makui, A., & Shahanaghi, K. (2018). The reliable design of a hierarchical multimodes transportation hub location problems (HMMTHLP) under dynamic network disruption (DND). *Computers & Industrial Engineering, 122*, 39–86.
Tran, T. H., O'Hanley, J. R., & Scaparra, M. P. (2016). Reliable hub network design: Formulation and solution techniques. *Transportation Science, 51*(1), 358–375.
Wang, B.-F., Lin, T.-C., Lin, C.-H., & Ku, S.-C. (2008). Finding the conditional location of a median path on a tree. *Information and Computation, 206*, 828–839.
Yaman, H. (2008). Star p-hub median problem with modular arc capacities. *Computers & Operations Research, 35*(9), 3009–3019.
Yaman, H. (2009). The hierarchical hub median problem with single assignment. *Transportation Research Part B: Methodological, 43*, 643–658.
Yaman, H. (2011). Allocation strategies in Hub Networks. *European Journal of Operational Research, 211*, 442–451.
Yaman, H., & Elloumi, S. (2012). Star *p*-hub center problem and star *p*-hub median problem with bounded path lengths. *Computers and Operations Research, 39*, 2725–2732.
Yaman, H., Karasan, O. E., & Kara, B. Y. (2012). Release time scheduling and hub location for next-day delivery. *Operations Research, 60*, 906–917.
Yang, X., Bostel, N., & Dejax, P. (2019). A MILP model and memetic algorithm for the hub location and routing problem with distinct collection and delivery tours. *Computers & Industrial Engineering, 135*, 105–119.
Zheng, J., Zhang, W., Qi, J., & Wang, S. (2020). Canal effects on a liner hub location problem. *Journal of Cleaner Production, 247*, 119176.
Zhong, W., Juan, Z., Zong, F., & Su, H. (2018). Hierarchical hub location model and hybrid algorithm for integration of urban and rural public transport. *Distributed Sensor Networks, 14*, 1–14.