



Transportation Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Hub Location, Routing, and Route Dimensioning: Strategic and Tactical Intermodal Transportation Hub Network Design

Bariş Yıldız, Hande Yaman, Oya Ekin Karaşan

To cite this article:

Bariş Yıldız, Hande Yaman, Oya Ekin Karaşan (2021) Hub Location, Routing, and Route Dimensioning: Strategic and Tactical Intermodal Transportation Hub Network Design. Transportation Science 55(6):1351-1369. <https://doi.org/10.1287/trsc.2021.1070>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages




With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Hub Location, Routing, and Route Dimensioning: Strategic and Tactical Intermodal Transportation Hub Network Design

Bariş Yıldız,^a Hande Yaman,^b Oya Ekin Karaşan^c

^aDepartment of Industrial Engineering, College of Engineering, Koç University, 34450 Istanbul, Turkey; ^bResearch Centre for Operations Research and Statistics (ORSTAT), Faculty of Economics and Business, KU Leuven, 3000 Leuven, Belgium; ^cDepartment of Industrial Engineering, Bilkent University, Bilkent, 06800 Ankara, Turkey

Contact: byildiz@ku.edu.tr,  <https://orcid.org/0000-0002-3839-8371> (BY); hande.yaman@kuleuven.be (HY); karasan@bilkent.edu.tr (OEK)

Received: February 26, 2020

Revised: November 23, 2020; April 7, 2021

Accepted: April 13, 2021

Published Online in Articles in Advance:
October 5, 2021

<https://doi.org/10.1287/trsc.2021.1070>

Copyright: © 2021 INFORMS

Abstract. We propose a novel hub location model that jointly eliminates some of the traditional assumptions on the structure of the network and on the discount as a result of economies of scale in an effort to better reflect real-world logistics and transportation systems. Our model extends the hub literature in various facets: instead of connecting nonhub nodes directly to hub nodes, we consider routes with stopovers; instead of connecting pairs of hubs directly, we design routes that can visit several hub nodes; rather than dimensioning pairwise connections, we dimension routes of vehicles; and rather than working with a homogeneous fleet, we use intermodal transportation. Decisions pertinent to strategic and tactical hub location and transportation network design are concurrently made through the proposed optimization scheme. An effective branch-and-cut algorithm is developed to solve realistically sized problem instances and to provide managerial insights.

Funding: The research of the first author has been supported from Bilim Akademisi (the Science Academy, Turkey) under the BAGEP program.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/trsc.2021.1070>.

Keywords: hub location • network design • route dimensioning • multimodal transport • next-day delivery • branch and cut

1. Introduction

Campbell and O’Kelly (2012) provide a synthesis of the 25 years of the vast hub location research following the seminal works of O’Kelly (1986a, b; 1987). Although the will to solve instances of realistic sizes has geared the literature toward models with strong assumptions, the real challenges have been in defining and solving realistic problems. Integrating the strategic-level hub location decisions with the transportation network design has been identified as a crucial yet rather challenging research direction. We take up this challenge with this paper.

On par with the academic interest, the practical motivation for the hub network design is on the rise. The explosive growth in the business-to-consumer form of e-commerce has made logistics a prominent determinant in gaining the competitive edge in this market. Logistics service providers have to distinguish themselves both in prices and in service quality, which is mainly identified with short delivery times. Carriers, in general, employ hub-and-spoke networks in their many-to-many distribution systems to meet customers’ high-service-level expectations (i.e., faster delivery, larger coverage, and higher availability) without additional costs (monetary and environmental) that customers are quite reluctant to face (Joeress et al. 2016). Strategically located operations centers can

drastically increase the efficiency of the transportation network by providing opportunities to consolidate freight from different originating branch offices, disseminate it according to destined branch offices, and enable a seamless intermodal integration. With these critical capabilities, hub networks are at the core of the emerging revolutionary logistic concepts, such as the *physical internet* (Montreuil 2011; Crainic and Montreuil 2016) and *mobility as a service* (MaaS) (Hietanen 2014; Maheo, Kilby, and Van Hentenryck 2017).

Motivated by these novelties, we investigate the next-day delivery network design problem where hubs do not solely function as consolidation points but are also required to facilitate mode shifts (transfers) in a multimodal transportation network, and the classical assumption of cost reduction in interhub transfer may or may not be valid depending on the vehicles used, the settings, and the assumptions made. In our setting, as is customary in the operations of the cargo sector, each branch office (demand center) will be assigned to an operations center (hub), where packages can be transferred between the vehicles. To benefit from economies of scale, stopovers will be permitted in both the spoke and hub networks. Service quality will be ensured by limiting both the spoke and hub segments in terms of time. The resulting network will be intermodal, utilizing various ground and air

transportation means. To this end, a parcel that is picked up from its originating branch office will (i) first travel on some form of transportation, perhaps visiting other demand centers on the way, until reaching its designated hub, (ii) then get transferred to another vehicle (possibly plane) that is either bypassing through or originating from the designated hub, (iii) traverse the hub network, perhaps visiting some stopover hubs on the way, until reaching its destination's designated hub, where (iv) it will get transferred to another vehicle and (v) travel until its final destination, possibly visiting other demand centers on the way. The parcel will travel on at most three different vehicles, and there will be a transfer of vehicles only at the hub nodes upon entering and leaving the hub network. Routes between demand centers and their designated hubs are called *access routes*, and routes traversing the hub network are called *hub routes*. Choosing the access and hub routes to operate and determining the types and numbers of vehicles to assign to those paths (to build enough capacity) are the decisions for the transportation network design and dimensioning. On the other hand, locations of the hubs, as strategic-level decisions, need to be determined to support the mentioned intermodal structure of the transportation network. Clearly, transportation mode switches (change of vehicles) can only be performed in "hub" nodes with proper sorting and packaging capabilities as well as the facilities that can conduct intermodal transfer between involved vehicles (i.e., planes, ships, trains, trucks). As such, the hub deployment decisions are strongly connected with the transportation network design decisions and need to be considered jointly, as we do with our model. In particular, our model provides the answers for the following questions:

- What are the locations of hubs?
- Which is the designated hub of each demand center?
- How should the demand centers of a particular hub be partitioned to a collection of access routes?
- What should be the route of each vehicle used on access and hub routes?
- How many vehicles of a certain type (with a given capacity) should be used on each route?

The classical hub location problem has some intrinsic as well as extrinsic assumptions that challenge its match to realistic transportation settings. Since the seminal paper of O'Kelly (1987), many variants of the hub location problem have been studied with an effort to relax some of the assumptions of the basic problem (Campbell 1996; Alumur et al. 2021) to make it more realistic. This is also our aim in this study. In particular, we jointly eliminate the following assumptions/restrictions that have also been relaxed in various studies.

Complete Hub Network: In the classical version, hubs are interconnected via direct links. An efficiently designed incomplete hub network may be cost beneficial without sacrificing service quality. Various research works in the literature have relaxed this assumption and tackled the design problem of the hub network (O'Kelly and Miller 1994; Nickel, Schöbel, and Sonneborn 2001; Yoon and Current 2008; Alumur, Kara, and Karasan 2009; Calik et al. 2009; Contreras, Fernández, and Marín 2010; de Camargo et al. 2017). Most of these studies focus on designing a connected topology in the strategic level without detailing the underlying transportation network (i.e., the actual routes). Some studies such as Gelareh and Pisinger (2011) handle specific hub network designs such as rings together with fleet deployment. We, on the other hand, design routes that visit a set of hub nodes and carry the flow originating from hubs visited earlier on the route destined to hubs that are visited later. Hub routes are served by capacitated vehicles, and consequently, it may be necessary to allocate more than one vehicle on a hub route depending on the traffic. These routes are also bounded in length and/or in hops. The resulting hub network does not have any particular property other than being connected. Note that even our connectivity assumption is relaxed in some studies such as Campbell, Ernst, and Krishnamoorthy (2005a, b).

Star Access Network: The access network is typically star shaped in the classical hub location problem. In the single allocation variant, each nonhub node is allocated to exactly one hub node, and all its traffic transits through this hub. In the multiple allocation variant (Ebery et al. 2000), a node's traffic can transit through all hubs. In both cases, the traffic from a node to another node passes through at least one hub and no other nonhub nodes. Consistent with the literature utilizing more realistic access networks with stopovers (Yaman, Kara, and Tansel 2007) or ring topologies (Nagy and Salhi 1998; Wasner and Zäpfel 2004; Rodríguez-Martín, Salazar-González, and Yaman 2014; Lopes et al. 2016; Kartal, Hasgul, and Ernst 2017; Kartal, Krishnamoorthy, and Ernst 2019), we also benefit from the consolidation of flows on our access routes carrying traffic of several nonhub nodes to their designated hub and back. These access routes may start and end at the hub as in the hub location and routing problem (Çetiner, Sepil, and Süral 2010; de Camargo, de Miranda, and Løkketangen 2013) or can start at a nonhub node and end at a hub node as in the hub location problem with stopovers. Each access route is served by one vehicle, and consequently, the total traffic that can be picked up or delivered on it cannot exceed the capacity of the

vehicle. In addition, the length of an access route is bounded to ensure quality of service. This bound can be in terms of the distance, travel time, and/or the number of nonhub nodes on the route (hops). The cost of an access route is the cost of allocating a vehicle to this route. If routes start at nonhub nodes, the access network of a particular hub is the union of paths that cover a subset of nonhub nodes and end at the hub. If routes start and end at the hub node, then each access network is a union of cycles. Though we relax the star access network topology assumption, we still have the single assignment assumption. We opted for single assignment, as we are interested in the case where some nonhub nodes do not have sufficient demand to justify the allocation of a dedicated vehicle, and so vehicles travelling on access routes visit several nonhub nodes on their way to the hub and back. With an arbitrary allocation strategy, several vehicles may need to visit a nonhub node to pick up the demand going toward different hubs, and such a network can be very difficult to handle from an operational point of view.

Constant Discount Factor: In the classical hub location problem, the unit shipment cost on the hub network is discounted by a constant factor independent of the actual flow. Campbell (2013) observes that in optimal solutions of the basic hub location problems, it happens quite frequently that the connections between nonhub nodes and hub nodes carry more flow than hub-to-hub connections. In other words, the assumption that there is more traffic flow between hubs is usually not verified, and modelling the economies of scale by using a constant discount factor for inter-hub flows might not be representative in most cases. Several studies in the literature challenged the validity of this fundamental assumption and suggested different means to relax it. Podnar, Skorin-Kapov, and Skorin-Kapov (2002) discount the transportation cost on a link if the flow on this link exceeds a threshold. O’Kelly and Bryan (1998), Horner and O’Kelly (2001), and de Camargo, de Miranda, and Luna (2009) model economies of scale as a function of flow on the inter-hub links. Campbell, Ernst, and Krishnamoorthy (2005a, b), to this end, study the problem of locating a given number of hub arcs with discounted costs rather than locating hubs. Our models design and dimension the transportation network by assigning a fleet of vehicles to routes and include as costs the vehicle-route cost information acquired from the underlying setting.

Link-Based Dimensioning: Be it an incomplete or a complete hub network design, the details on the transportation end as to how the flow will actually traverse this network are typically left out in the literature. Studies such as Yaman and Carello (2005), Yaman

(2005), Corberán et al. (2016), Serper and Alumur (2016), and Tanash, Contreras, and Vidyarthi (2017) dimension the links; that is, they decide on a fleet of vehicles enough in capacity to carry the associated link flow. With our model we determine (i) which possibly multistep access routes to use, (ii) which possibly multistop hub routes to use, and (iii) the number of vehicles of different types to use on these routes. In other words, our dimensioning is route based, and as such, the demand on a hub route will be simply transferred without any intermediate handling on the stop-over hubs.

Means of Transportation: Most of the existing hub location literature do not specify the transportation mode. Within the existing hub location literature operating with multimodes, the networks of different modes are typically assumed to be disjoint (Alumur, Yaman, and Kara 2012) or hubs are reserved for a single mode (Alumur, Kara, and Karasan 2012). The studies such as Serper and Alumur (2016) that dimension links with different modal vehicles leave out the routing issue of such vehicles. Network-loading-type studies such as Jaillet, Song, and Yu (1996) that distribute flows among hubs into paths might result in transportation networks allowing for any number of exchanges at stopover hubs, which we deliberately escape from in our designs. We simultaneously route and dimension routes with an appropriate fleet and as such detail the transportation network. Because vehicles or a sequence of vehicles is routed between demand pairs, we assume our resulting networks are connected. We also assume that the vehicles traverse simple paths (not repeating vertices), which might even be relaxed to utilize the vehicle capacities to their fullest extent.

Single Objective: Models appearing in the literature typically design hub networks solely with cost or service-level objectives. Studies on time-definite deliveries such as Alumur, Yaman, and Kara (2012) and Campbell (2009) combine both cost- and service-level dimensions. Peiró et al. (2018) consider a multiobjective hub location problem in a multiallocation setting. To ensure the quality of service, our access and hub routes are bounded. Our models allow for this bound to be in terms of the distance, travel time, and/or the number of hops on the route. With the underlying realistic application in mind, we choose to put time limits on the access and hub routes in order to meet the desired next-day delivery service level. The cost of an access or a hub route is simply the cost of the vehicle assigned to it. To this end, both cost- and service-level objectives reflect the real-life operating characteristics.

There is a large amount of literature on hub location problems. We refer the reader to Alumur and Kara

(2008), Alumur (2019), Campbell and O’Kelly (2012), Contreras (2015), Contreras and O’Kelly (2019), and Farahani et al. (2013) for overviews of the existing hub location literature. Here, we focus on the few representative studies on variants relaxing the listed assumptions.

Though we manage to relax the above-mentioned restrictions/assumptions simultaneously, we adopt some assumptions. In particular, we assume that each access node is assigned to a single hub, the hub network is connected, demand carried between any pair of hubs traverses the hub network on a single vehicle, and all the access and hub paths are simple. We manage to face the challenges of relaxing some of the simplifying assumptions through our novel modeling approach. Even though we work with incomplete designs, the demand between an origin-destination pair is carried on at most one interhub and two access vehicles: one for collecting from the origin and the other for distributing to the destination. As such, the material handling burden is the same as in the basic complete hub network designs, and the transfer between origin and destination hubs is virtually direct. Even though we model multimodal transportation, we still can keep track of the particular hub pair demand at the level of partition to different vehicles on the hub network.

We impose upper bounds on the lengths/travel times of access routes and hub routes separately. We do this by dividing the allotted time horizon into three parts: one for collection on access routes, one for transfer between hubs, and the last one for delivery on access routes. In addition to ensuring a desired level of service quality, this helps to resolve any issues concerning synchronization: the vehicles on hub routes start their trips only after all vehicles on access routes have arrived at their designated hubs. Similarly, the vehicles that travel on access routes from hubs toward demand centers start their trips after all vehicles travelling on hub routes have arrived at their destinations. Even though we exert a hub covering type of constraint to provide a service quality, we manage to identify the actual transportation cost that this service will incur and, to this end, manage to integrate the cost and service bifurcation in the hub location models.

In light of the preceding, we contribute the hub location, routing, and route dimensioning (HLRD) problem to the literature. Figure 1 contrasts the proposed network structure with that of a classical hub network. In both figures, circles and squares correspond to demand and hub locations, respectively. In Figure 1(a), we see a classical design where the hub network is complete and each access network is a star. Figure 1(b), on the other hand, depicts a typical

HLRD design. The access networks consist of stop-over paths visiting several demand locations. The hub network may be incomplete. Several vehicles with different capacities may carry the demand in the resulting network. For example, the red and green paths might be the routes of two different types of trucks, the purple path might correspond to a two-segment plane route, and the blue arc might correspond to rented aircraft belly capacity. The hub locations, allocations, and the hub routes are determined based on deterministic demand, which could be the mean or the worst-case demand based on past data. However, further operational planning can reoptimize the access routes based on changing demand.

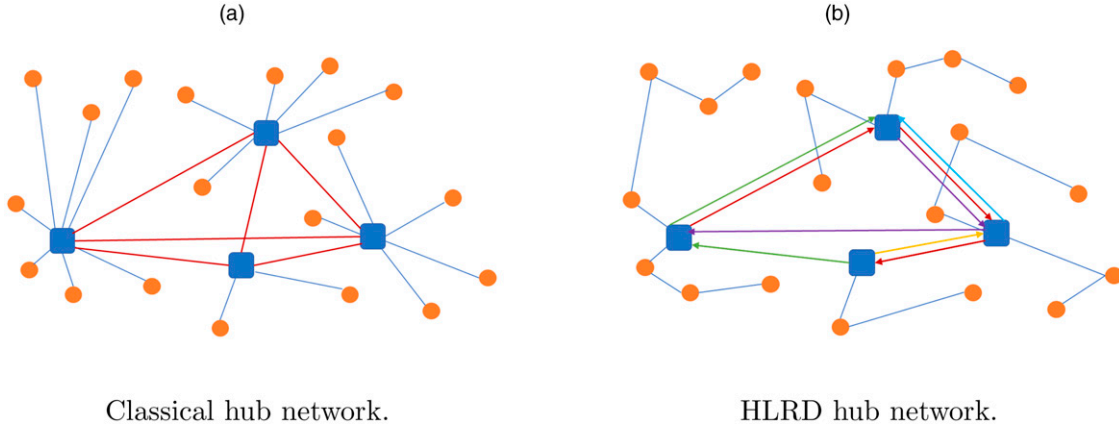
The remainder of this paper is organized as follows. In Section 2, we formally define the problem and introduce our compact mathematical model to solve it. In Section 3, we discuss the projection of the continuous variables. The branch-and-cut algorithm we propose to solve the resulting model is discussed in Section 4. In Section 5, we present and analyze the results of an extensive computational study. In Section 6, we conclude with some final remarks.

2. Preliminaries and HLRD Model

In this section we formally describe HLRD. We are given a set N of nodes with pairwise traffic demand. We define the set of distinct pairs of nodes, $D = \{(i, j) : i \in N, j \in N \setminus \{i\}\}$, and denote by w_{ij} the demand for pair $(i, j) \in D$. Our designs will select p nodes from the set N as hub locations, and the remaining will be access nodes. As depicted in Figure 1(b), these two sets of nodes will be connected to each other. We shall seek the effective routing and dimensioning of different vehicle types traversing these networks and carrying the demand with a promised quality level of service. In particular:

Access Networks: We assume each access node is assigned to exactly one hub and its traffic transits through this hub. We do not impose a direct connection from each nonhub node to its hub; instead, we design routes that connect several nonhub nodes to their hub. An access route may be a simple path that starts at a nonhub node and ends at a hub node. In this case, a vehicle starts its trip at the first nonhub node, visits other nonhub nodes—each time picking up their outgoing traffic demand—and brings this load to the hub. Then it picks up the incoming traffic demand of the nonhub nodes from the hub and starts its return trip toward the nonhub nodes. Alternatively, an access route may be a directed cycle visiting a set of nonhub nodes and one hub. In both cases, we assume that each nonhub node is served by exactly

Figure 1. (Color online) Classical vs. HLRD Hub Network Designs



one access route. To ensure the quality of service and coordinate access and hub route operations (as mentioned in Section 1), we limit the travel time on an access route to T_a .

We denote by \mathcal{P} the set of feasible access routes and by \mathcal{P}_k the set of feasible access routes ending at node $k \in N$. Note here that for a given access route $\rho \in \mathcal{P}$, we know the order in which nodes are visited and the required capacity of a vehicle that can serve this route. Consequently, the lowest-cost vehicle type that meets the travel time requirement and that has sufficient capacity can be chosen a priori. We denote the corresponding cost by c'_ρ .

Hub Network: We do not assume direct connections between hubs; we rather design and dimension the hub routes. A hub route is a simple directed path that visits a subset of hub nodes. As we did for access routes, to ensure the quality of service and coordinate with access routes, we impose a travel time limit of T_h on the hub routes. Different from the case of access routes, we do not know a priori the amount of traffic that is carried on a hub route, as this depends on the assignment of nodes to hubs. Consequently, the choice of vehicles for hub routes is part of the decision to be made by the model.

For hub routes, we use the following notation: We define Π to be the set of feasible hub routes, Π_{kl} to be the set of feasible hub routes in which node $k \in N$ comes before node $l \in N \setminus \{k\}$, and π_{kl} to be the subpath from k to l in route $\pi \in \Pi_{kl}$. We have a set of vehicles, each with a capacity, a fixed and a variable cost, and a speed. For a hub route $\pi \in \Pi$, we define V_π to be the set of vehicles that can traverse route π in the allowed amount of travel time. For $v \in V_\pi$, let c_π^v be the cost of using one vehicle of type v on route π , and let Q_v be the capacity of this vehicle. Our model designs hub

routes and decides how many of each type of vehicle to use on each route.

We use the following decision variables: x_{ik} is 1 if node $i \in N$ is assigned to hub node $k \in N$ and 0 otherwise. If node k is a hub, then x_{kk} is 1. For an access route $\rho \in \mathcal{P}$, z_ρ is 1 if this route is used and is 0 otherwise. The continuous variable y_{ijkl} is the amount of flow of pair $(i, j) \in D$ that enters the hub network at hub $k \in N$ and leaves it at hub $l \in N$, and f_π^{kl} is the amount of flow from hub $k \in N$ to hub $l \in N \setminus \{k\}$ routed on hub route $\pi \in \Pi_{kl}$. Finally, u_π^v is the number of vehicles of type $v \in V_\pi$ used on hub route $\pi \in \Pi$.

Using these variables, we model HLRD with a mixed integer program IP_{HLRD} , which we define as follows:

$$\min \sum_{\pi \in \Pi} \sum_{v \in V_\pi} c_\pi^v u_\pi^v + \sum_{\rho \in \mathcal{P}} c'_\rho z_\rho \quad (1)$$

$$\text{s.t. } \sum_{k \in N} x_{ik} = 1 \quad \forall i \in N, \quad (2)$$

$$x_{ik} \leq x_{kk} \quad \forall i, k \in N, i \neq k, \quad (3)$$

$$\sum_{k \in N} x_{kk} = p, \quad (4)$$

$$x_{ik} = \sum_{\rho \in \mathcal{P}_k: i \in \rho} z_\rho \quad \forall i, k \in N : i \neq k, \quad (5)$$

$$\sum_{l \in N} y_{ijkl} \geq w_{ij} x_{ik} \quad \forall (i, j) \in D, k \in N, \quad (6)$$

$$\sum_{k \in N} y_{ijkl} \leq w_{ij} x_{jl} \quad \forall (i, j) \in D, l \in N, \quad (7)$$

$$\sum_{\pi \in \Pi_{kl}} f_\pi^{kl} \geq \sum_{(i, j) \in D} y_{ijkl} \quad \forall k, l \in N, k \neq l, \quad (8)$$

$$\sum_{k, l \in N: k \neq l, a \in \pi_{kl}} f_\pi^{kl} \leq \sum_{v \in V_\pi} Q_v u_\pi^v \quad \forall \pi \in \Pi, a \in \pi, \quad (9)$$

$$u_\pi^v \leq M x_{kk} \quad \forall \pi \in \Pi, k \in \pi, v \in V_\pi, \quad (10)$$

$$x_{ik} \in \{0, 1\} \quad \forall i, k \in N, \quad (11)$$

$$z_p \in \{0,1\} \quad \forall p \in \mathcal{P}, \quad (12)$$

$$u_\pi^v \geq 0 \text{ and integer} \quad \forall \pi \in \Pi, v \in V_\pi, \quad (13)$$

$$f_\pi^{kl} \geq 0 \quad \forall k, l \in N, k \neq l, \pi \in \Pi_{kl}, \quad (14)$$

$$y_{ijkl} \geq 0 \quad \forall (i,j) \in D, k, l \in N. \quad (15)$$

The objective function (1) minimizes the total cost (fixed and variable) of the vehicle fleet. Constraints (2)–(4) ensure that each node either becomes a hub or is assigned to exactly one hub node and p hubs are chosen. Constraints (5) relate the assignments and the access routes: if node i is assigned to hub k , then one access route that visits node i and ends at hub k is used. Otherwise, no such route can be part of the solution. For binary x , Constraints (6) and (7) imply that $y_{ijkl} = w_{ij}x_{ik}x_{jl}$. The amount $\sum_{(i,j) \in D} y_{ijkl}$ is the amount of traffic that enters the hub network at hub k and that leaves it at hub l . Using Constraints (8), we allocate this traffic to hub routes that visit k before l . Note that Constraints (6)–(8) can also be written as equalities without changing the optimal value. For a hub route $\pi \in \Pi$, the required capacity is the maximum traffic carried on any of its arcs. The traffic on arc $a \in \pi$ is the sum of the traffic from hubs k to hubs l so that a is on the subpath of π from k to l . For example, if the hub route is 123, then arc (1, 2) carries the traffic from hub 1 to hubs 2 to 3, whereas arc (2, 3) carries the traffic from hubs 1 and 2 to hub 3. These capacity restrictions are imposed through Constraints (9), where the left-hand side is the traffic on arc a and the right-hand side is the total capacity of vehicles allocated to route π . Constraints (10) ensure that hub routes only go through hubs. In our case, as the costs satisfy the triangle inequality, we drop these constraints. The remaining constraints are domain restrictions for the variables.

To better illustrate the hub routes and what they can carry as traffic, we refer the reader to Figure 2. In this figure, we have four hubs and five hub routes, each depicted with a different color. The red route $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ can carry the traffic from hub 1 to hubs 2, 3, and 4 on arc (1, 2). The traffic from hub 1 to hub 2 is unloaded at hub 2, and the traffic from hub 2 to hubs 3 and 4 is loaded on the vehicle. Hence on arc (2, 3), the load on the vehicle is the traffic from hubs 1 and 2 to hubs 3 and 4. At node 3, the traffic from hubs 1 and 2 to hub 3 is unloaded, and the traffic from 3 to 4 is loaded. These three commodities are routed to hub 4, which is the final hub of this route.

The total traffic that needs to travel from hub 1 to hub 2 can use the red route and the green route. However, the traffic from hub 4 to hub 3 cannot first use the green route ($4 \rightarrow 1 \rightarrow 2$) and then the (2 \rightarrow 3) segment of the red route because this would require an unloading and loading operation at hub 2. For ease of operations and synchronization, we do not allow such

actions and enforce that each demand is carried in one vehicle through the hub network (i.e., through the yellow line for this particular traffic).

From the technical point of view, there are two potential difficulties when solving IP_{HLRD} : (1) the large number of paths (access and hub routes) one might need to consider in the model and (2) the need for a large number of flow variables (i.e., f and y) to properly determine the required capacity in the interhub transport network. For the former, one can directly employ a column generation approach to iteratively include the path-based variables in the model as they are needed. However, in many real-world applications, the operational requirements, service quality considerations, government regulations (driving hours, flight routes, etc.), and the conditions of the existing road network significantly limit the alternative number of routes and make it a computationally viable option to include all possible paths directly in the model, which is what we do in our numerical studies. However, a similar approach fails to address the computational difficulty that arises as a result of the high number of flow variables, which directly depends on the number of origin-destination pairs and alternative hub locations, as well as the number of hub paths. We next discuss the projection approach we propose in order to handle this issue.

3. Projection of Continuous Variables

As our model contains a large number of variables, in this section, we project out the continuous variables—namely, variables y and f . This gives us a formulation in the space of x , z , and u variables. However, this formulation is not compact.

Proposition 1. A solution (x, z, u) that satisfies (2)–(5) and (10)–(13) is feasible for HLRD if and only if the projection inequality

$$\sum_{\pi \in \Pi} \sum_{a \in \pi} \delta_{\pi a} \sum_{v \in V_\pi} Q_v u_\pi^v \geq \sum_{(i,j) \in D} w_{ij} \left(\sum_{k \in N} \alpha_{ijk} x_{ik} - \sum_{l \in N} \beta_{ijl} x_{jl} \right) \quad (16)$$

is satisfied for all $(\alpha, \beta, \gamma, \delta)$ such that

$$\alpha_{ijk} - \beta_{ijl} - \gamma_{kl} \leq 0 \quad \forall (i,j) \in D, k, l \in N : k \neq l, \quad (17)$$

$$\alpha_{ijk} - \beta_{ijl} \leq 0 \quad \forall (i,j) \in D, k \in N, \quad (18)$$

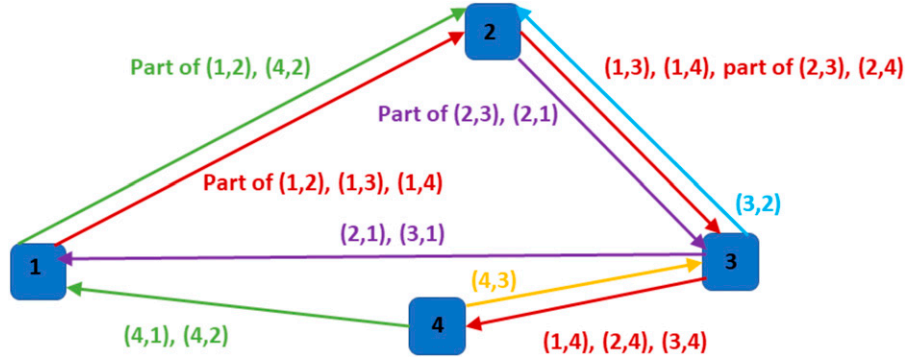
$$\gamma_{kl} - \sum_{a \in \pi_{kl}} \delta_{\pi a} \leq 0 \quad \forall k, l \in N : k \neq l, \pi \in \Pi_{kl}, \quad (19)$$

$$\alpha, \beta, \gamma, \delta \geq 0. \quad (20)$$

Proof. A solution (x, z, u) that satisfies (2)–(5) and (10)–(13) is feasible if and only if there exist y and f such that

$$\sum_{l \in N} y_{ijkl} \geq w_{ij} x_{ik} \quad \forall (i,j) \in D, k \in N, \quad (21)$$

Figure 2. (Color online) Hub Transportation Network



$$-\sum_{k \in N} y_{ijkl} \geq -w_{ij}x_{jl} \quad \forall (i, j) \in D, l \in N, \quad (22)$$

$$\sum_{\pi \in \Pi_{kl}} f_{\pi}^{kl} - \sum_{(i, j) \in D} y_{ijkl} \geq 0 \quad \forall k, l \in N, k \neq l, \quad (23)$$

$$-\sum_{\substack{k, l \in N : k \neq l, \\ a \in \pi_{kl}}} f_{\pi}^{kl} \geq -\sum_{v \in V_{\pi}} Q_v u_{\pi}^v \quad \forall \pi \in \Pi, a \in \pi, \quad (24)$$

$$y_{ijkl} \geq 0 \quad \forall (i, j) \in D, k, l \in N, \quad (25)$$

$$f_{\pi}^{kl} \geq 0 \quad \forall k, l \in N : k \neq l, \pi \in \Pi_{kl}. \quad (26)$$

We associate dual variables α_{ijk} , β_{ijl} , γ_{kl} , and $\delta_{\pi a}$ to Constraints (21), (22), (23), and (24), respectively. The feasibility problem (21)–(26) has a solution if and only if the dual problem

$$\begin{aligned} \max \quad & \sum_{(i, j) \in D} w_{ij} \left(\sum_{k \in N} \alpha_{ijk} x_{ik} - \sum_{l \in N} \beta_{ijl} x_{jl} \right) \\ & - \sum_{\pi \in \Pi} \sum_{a \in \pi} \delta_{\pi a} \sum_{v \in V_{\pi}} Q_v u_{\pi}^v \\ \text{s.t.} \quad & (17)–(20) \end{aligned}$$

is bounded. As the feasible set of the dual problem is a cone, Inequalities (16) impose that no feasible direction is an improving direction. \square

Projection inequalities (16) can be separated by solving the linear program (LP)

$$\begin{aligned} \theta = \max \quad & \sum_{(i, j) \in D} w_{ij} \left(\sum_{k \in N} \alpha_{ijk} x_{ik} - \sum_{l \in N} \beta_{ijl} x_{jl} \right) - \sum_{\pi \in \Pi} \sum_{a \in \pi} \delta_{\pi a} \sum_{v \in V_{\pi}} Q_v u_{\pi}^v \\ \text{s.t.} \quad & (17)–(19), \\ & 0 \leq \alpha, \beta, \gamma, \delta \leq 1. \end{aligned}$$

where the variables are bounded above by 1 for normalization. If $\theta \leq 0$, then all projection inequalities are satisfied. Otherwise, an optimal solution gives a most violated projection inequality.

Next we provide a set of vectors $(\alpha, \beta, \gamma, \delta)$ that satisfy (17)–(20).

Lemma 1. Let $D' \subseteq D$, $S, T \subseteq N$ with $S \cap T = \emptyset$. Consider the following vector $(\alpha, \beta, \gamma, \delta)$: $\alpha_{ijk} = 1$ for $(i, j) \in D'$ and $k \in S$, $\beta_{ijl} = 1$ for $(i, j) \in D'$ and $l \in N \setminus T$, $\gamma_{kl} = 1$ for $k \in S$ and $l \in T$. We compute δ as follows: Let $\pi \in \Pi$, and let r be the number of arcs on path π . Unmark all arcs on π . For $s = 1, \dots, r$, do the following: For each subpath π_{kl} of π with length s , if $k \in S$ and $l \in T$ and if no arc on subpath π_{kl} is marked, then set $\delta_{\pi a} = 1/s$ for all arcs $a \in \pi_{kl}$ and mark all these arcs. Set all remaining entries of $(\alpha, \beta, \gamma, \delta)$ to 0. This vector $(\alpha, \beta, \gamma, \delta)$ satisfies (17)–(20).

Proof. It is easy to check that $(\alpha, \beta, \gamma, \delta)$ satisfies (17), (18), and (20). Suppose, to the contrary, that it does not satisfy (19). Let $\pi \in \Pi$ be such that π has a subpath π_{kl} with $k \in S$, $l \in T$, $\sum_{a \in \pi_{kl}} \delta_{\pi a} < 1$ and the number of arcs on π_{kl} is the smallest among all such subpaths of π . Because $\sum_{a \in \pi_{kl}} \delta_{\pi a} < 1$, when we processed subpath π_{kl} , at least one of the arcs on this subpath was marked previously. Then path π contains another subpath π_{mn} that has a common arc with π_{kl} , $m \in S$, and $n \in T$ and π_{mn} does not contain more arcs than π_{kl} . This implies that either π_{ml} or π_{kn} is a subpath of π_{kl} and has fewer arcs than π_{kl} . Suppose that it is subpath π_{ml} . Because $\sum_{a \in \pi_{kl}} \delta_{\pi a} < 1$, then $\sum_{a \in \pi_{ml}} \delta_{\pi a} < 1$. This contradicts the assumption that π_{kl} is a subpath of π with the smallest number of arcs such that $k \in S$, $l \in T$, and $\sum_{a \in \pi_{kl}} \delta_{\pi a} < 1$.

Consider a small example with $N = \{1, 2, 3, 4\}$, $S = \{1, 2\}$, and $T = \{3\}$. Suppose that only hub routes with at most two legs are allowed. Then $\sum_{a \in \pi_{kl}} \delta_{\pi a} \geq 1$ for all paths $\pi \in \Pi_{13} \cup \Pi_{23}$. The above-mentioned approach produces the following vector δ : $\delta_{1 \rightarrow 3, (1,3)} = \delta_{2 \rightarrow 3, (2,3)} = \delta_{1 \rightarrow 3 \rightarrow 2, (1,3)} = \delta_{1 \rightarrow 2 \rightarrow 3, (2,3)} = \delta_{2 \rightarrow 3 \rightarrow 1, (2,3)} = \delta_{2 \rightarrow 1 \rightarrow 3, (1,3)} = \delta_{1 \rightarrow 3 \rightarrow 4, (1,3)} = \delta_{4 \rightarrow 1 \rightarrow 3, (1,3)} = \delta_{2 \rightarrow 3 \rightarrow 4, (2,3)} = \delta_{4 \rightarrow 2 \rightarrow 3, (2,3)} = 1$, $\delta_{1 \rightarrow 4 \rightarrow 3, (1,4)} = \delta_{1 \rightarrow 4 \rightarrow 3, (4,3)} = \delta_{2 \rightarrow 4 \rightarrow 3, (2,4)} = \delta_{2 \rightarrow 4 \rightarrow 3, (4,3)} = 1/2$, and others are 0.

Now suppose that $N = \{1, 2, 3, 4\}$, $S = \{1, 2\}$, and $T = \{3, 4\}$, and hub routes with at most three legs are allowed. We do not give the full vector δ , as the

number of paths is rather large. However, we look at two paths as examples. The first path is $1 \rightarrow 3 \rightarrow 2 \rightarrow 4$. For this path, we have $\delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (1,3)} = \delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (2,4)} = 1$ and $\delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (3,2)} = 0$. The second path is $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ and $\delta_{1 \rightarrow 2 \rightarrow 3 \rightarrow 4, (2,3)} = 1$, and $\delta_{1 \rightarrow 2 \rightarrow 3 \rightarrow 4, (1,2)} = \delta_{1 \rightarrow 2 \rightarrow 3 \rightarrow 4, (3,4)} = 0$. The variables associated with the first path appear with a coefficient of 2 ($\delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (1,3)} + \delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (2,4)} + \delta_{1 \rightarrow 3 \rightarrow 2 \rightarrow 4, (3,2)} = 2$) on the left-hand side of Inequality (16), as this path can carry the traffic from 1 to 3 and 2 to 4 but on two different legs. On the other hand, the variables associated with the second path appear with a coefficient of 1 because this path can carry traffic from 1 to 3, 1 to 4, 2 to 3, and 2 to 4 but all this traffic uses the common leg (2, 3).

3.1. Hub Routes with a Single Leg

In an attempt to gain insight into the projection inequalities, we analyze the special cases of short hub routes. If each hub route consists of a single leg, then Π_{kl} contains only one path that corresponds to arc (k, l) . For that reason, the f variables in (23) and (24) can be eliminated, and these constraints can be combined into

$$-\sum_{(i,j) \in D} y_{ijkl} \geq -\sum_{v \in V_{kl}} Q_v u_{kl}^v \quad \forall k \in N, l \in N \setminus \{k\}.$$

In a similar manner, associating dual variables α_{ijk} and β_{ijl} to Constraints (21) and (22), respectively, and γ_{kl} to the above-mentioned inequalities, we can show that a solution (x, z, u) that satisfies (2)–(5) and (10)–(13) is feasible if and only if

$$\sum_{k \in N} \sum_{l \in N \setminus \{k\}} \gamma_{kl} \sum_{v \in V_{kl}} Q_v u_{kl}^v \geq \sum_{(i,j) \in D} w_{ij} \left(\sum_{k \in N} \alpha_{ijk} x_{ik} - \sum_{l \in N} \beta_{ijl} x_{jl} \right) \quad (27)$$

is satisfied for all $(\alpha, \beta, \gamma) \geq 0$ such that

$$\alpha_{ijk} - \beta_{ijl} - \gamma_{kl} \leq 0 \quad \forall (i,j) \in D, k, l \in N : k \neq l, \quad (28)$$

$$\alpha_{ijk} - \beta_{ijk} \leq 0 \quad \forall (i,j) \in D, k \in N. \quad (29)$$

This projection has been studied in Labbé and Yaman (2004) and Labbé, Yaman, and Gourdin (2005) for a single type of vehicle. From the results of these studies, we can see easily that, in the case of hub routes of single arcs (i.e., the classical complete hub networks), the projection inequalities

$$\sum_{v \in V_{kl}} Q_v u_{kl}^v \geq \sum_{(i,j) \in D'} w_{ij} (x_{ik} + x_{jl} - 1) \quad (30)$$

for all $D' \subseteq D, k \in N$ and $l \in N \setminus \{k\}$ are sufficient to have a valid formulation of the problem. Inequalities (30) can be separated by enumerating all pairs k and l and choosing $D' = \{(i,j) \in D : x_{ik} + x_{jl} > 1\}$. The more general class of Inequalities (27) can be separated in polynomial time by solving a linear program; however, we are not aware of any polynomial-time combinatorial algorithm for their separation.

3.2. Hub Routes with at Most Two Legs

Now we analyze the projection inequalities for the special case with hub routes of at most two legs. Observe that in this case we have hub routes of the form $k \rightarrow l$ and $k \rightarrow l \rightarrow m$. The traffic from hub k to hub l can be carried on paths $k \rightarrow l, k \rightarrow l \rightarrow m, m \rightarrow k \rightarrow l$, and $k \rightarrow m \rightarrow l$ for $m \in N \setminus \{k, l\}$.

Proposition 2. Suppose that hub routes can contain at most two arcs. Let $D' \subseteq D, S, T \subseteq N$ with $S \cap T = \emptyset$. The projection inequalities

$$\begin{aligned} \sum_{k \in S} \sum_{l \in T} \left(\sum_{v \in V_{kl}} Q_v u_{kl}^v + \sum_{m \in N \setminus \{k, l\}} \left(\sum_{v \in V_{klm}} Q_v u_{klm}^v + \sum_{v \in V_{mkl}} Q_v u_{mkl}^v \right) \right. \\ \left. + \sum_{m \in N \setminus (S \cup T)} \sum_{v \in V_{kml}} Q_v u_{kml}^v \right) \geq \sum_{(i,j) \in D'} w_{ij} \left(\sum_{k \in S} x_{ik} + \sum_{l \in T} x_{jl} - 1 \right) \end{aligned}$$

are valid inequalities.

Proof. We use the vectors $(\alpha, \beta, \gamma, \delta)$ defined in Lemma 1 to derive these projection inequalities. For the case of at most two arcs, these vectors take the following form:

- $\alpha_{ijk} = 1$ for $(i,j) \in D'$ and $k \in S$,
- $\beta_{ijl} = 1$ for $(i,j) \in D'$ and $l \in N \setminus T$,
- $\gamma_{kl} = 1$ and $\delta_{kl, (k,l)} = 1$ for $k \in S$ and $l \in T$,
- $\delta_{klm, (k,l)} = \delta_{mkl, (k,l)} = 1$ for $k \in S, l \in T$ and $m \in N \setminus \{k, l\}$,
- $\delta_{kml, (k,m)} = \delta_{kml, (m,l)} = 1/2$ for $k \in S, l \in T$ and $m \in N \setminus (S \cup T)$, and
- other entries are 0.

Next, we derive the projection inequality for this dual solution. The right-hand side is equal to

$$\sum_{(i,j) \in D} w_{ij} \left(\sum_{k \in N} \alpha_{ijk} x_{ik} - \sum_{l \in N} \beta_{ijl} x_{jl} \right) = \sum_{(i,j) \in D'} w_{ij} \left(\sum_{k \in S} x_{ik} - \sum_{l \in N \setminus T} x_{jl} \right).$$

Using Constraint (2), we can substitute $\sum_{l \in N \setminus T} x_{jl} = 1 - \sum_{l \in T} x_{jl}$ to see that

$$\sum_{(i,j) \in D'} w_{ij} \left(\sum_{k \in S} x_{ik} - \sum_{l \in N \setminus T} x_{jl} \right) = \sum_{(i,j) \in D'} w_{ij} \left(\sum_{k \in S} x_{ik} + \sum_{l \in T} x_{jl} - 1 \right).$$

Finally, we need to show that

$$\begin{aligned} \sum_{\pi \in \Pi} \sum_{a \in \pi} \delta_{\pi a} \sum_{v \in V_{kl}} Q_v u_{kl}^v = \sum_{k \in S} \sum_{l \in T} \left(\sum_{v \in V_{kl}} Q_v u_{kl}^v + \sum_{m \in N \setminus \{k, l\}} \left(\sum_{v \in V_{klm}} Q_v u_{klm}^v \right. \right. \\ \left. \left. + \sum_{v \in V_{mkl}} Q_v u_{mkl}^v \right) + \sum_{m \in N \setminus (S \cup T)} \sum_{v \in V_{kml}} Q_v u_{kml}^v \right). \end{aligned}$$

For a path $\pi \in \Pi$ that consists of one arc—say, $\pi = k \rightarrow l$, $\delta_{kl, (k,l)} = 1$ if $k \in S$ and $l \in T$.

For a path $\pi \in \Pi$ that consists of two arcs, $\pi = k \rightarrow l \rightarrow m$, $\sum_{a \in \pi} \delta_{\pi a} = \delta_{klm, (k,l)} + \delta_{klm, (l,m)}$. If $k \in S$ and $l \in T$, then $\delta_{klm, (k,l)} = 1$ and $\delta_{klm, (l,m)} = 0$, and if $l \in S$ and $m \in T$, then $\delta_{klm, (l,m)} = 1$ and $\delta_{klm, (k,l)} = 0$. Also, if $k \in S, m \in T$ and $l \in N \setminus (S \cup T)$, then $\delta_{klm, (k,l)} = \delta_{klm, (l,m)} = 1/2$. In all other cases, $\delta_{klm, (k,l)} = \delta_{klm, (l,m)} = 0$. As these events are mutually exclusive, we can conclude that for path $\pi = k \rightarrow l \rightarrow m$, $\sum_{a \in \pi} \delta_{\pi a} = 1$ if and only if one of the following is true:

- i. $k \in S$ and $l \in T$,
- ii. $l \in S$ and $m \in T$, and
- iii. $k \in S$, $m \in T$ and $l \in N \setminus (S \cup T)$;
- otherwise, $\sum_{a \in \pi} \delta_{\pi a} = 0$.

We do not know an efficient separation algorithm for Inequalities (31). For given sets S and T , the best set D' can be computed as $D' = \{(i, j) \in D : \sum_{k \in S} x_{ik} + \sum_{l \in T} x_{jl} > 1\}$.

It is interesting to note that unlike the case of hub routes with single arcs, when hub routes can contain two legs, including all of the inequalities (31) does not yield a valid formulation. To see this, consider the following very simple example where we have three hubs, a , b , and c . Each hub wants to send one unit of traffic to every other hub. There is a single type of vehicle, and its capacity is 1. For this problem instance, the solution where one vehicle is allocated to each route abc , bca , and cab satisfies Inequalities (31) for all subsets S and T . However, this solution is not feasible. Each route, indeed, requires two vehicles.

4. Branch-and-Cut Algorithm

The model (1)–(15) contains a large number of variables for two reasons. First, we enumerate all feasible access routes and hub routes. Second, we have a large number of continuous variables y_{ijkl} and f_{π}^{kl} . Our preliminary analysis with the Turkish data showed that the biggest problem was due to the large number of continuous variables. For this reason, we developed a branch-and-cut algorithm, referred to as BC, based on the projection of continuous variables. To enhance the computational effectiveness of BC, we consider a heuristic separation procedure, two repair heuristics to improve the upper and lower bounds during the branch-and-bound search, and valid inequalities to strengthen our formulation.

4.1. Heuristic Separation

Let $(\bar{x}, \bar{z}, \bar{u})$ be a solution that satisfies (2)–(5) and (10)–(13). We define the candidate set $C = \{k : \bar{x}_{kk} > 0\}$ and for all distinct $k, l \in C$ we define $D_{kl} = \{(i, j) \in D : \bar{x}_{ik} + \bar{x}_{jl} - 1 > 0\}$ and check whether the following projection inequality,

$$\sum_{\pi \in \Pi_{kl}} \sum_{v \in V_{\pi}} Q_v u_{\pi}^v \geq \sum_{(i, j) \in D_{kl}} w_{ij} (x_{ik} + x_{jl} - 1), \quad (32)$$

is violated. Note that (32) is a straightforward extension of the projection inequalities (31) for singletons (i.e., when $|S|$ and $|T|$ are 1), when the hub paths are not limited to two legs. Our preliminary results showed that applying the heuristic separation at fractional solutions does not improve the solution performance. This is also true for the exact separation. So we add projection inequalities as lazy constraints.

4.2. Repair Heuristics

To further increase the computational efficiency, we also consider two alternative methods (H1 and H2) as repair heuristics to process integer solutions for which the heuristic separation method cannot find any violations. We use the repair heuristics not only to find high-quality feasible solutions to update the upper bound and the incumbent but also to cut off infeasible solutions in a computationally efficient way. In H1, we compute the optimal cost for the given set of hubs in this solution—say, N' —by solving Model (1)–(15). Note that the resulting model is much smaller when the hub locations are fixed and we can simply disregard all the hub paths that contain stops that are not hubs and access paths that end at a nonhub node (i.e., remove respective f and u variables from the model). Because the number of hubs to open (p) is typically much smaller compared with the size of the candidate locations, such a reduction can, indeed, reveal formulations that can be directly solved in small run times with small memory requirements, for the large-sized problem instances we consider in our study. We compare the optimal cost of the reduced problem with the best upper bound and update the latter and the incumbent, if necessary. Then we also add the following cut in order not to consider the same set of hubs again:

$$\sum_{k \in N'} z_{kk} \leq p - 1. \quad (33)$$

Note that if the integer solution we test is not capacity feasible (i.e., violates any of the projection inequalities) H1 would find the best possible solution one would get from the suggested hub locations and effectively cut off this solution by adding (33). Similarly, in H2, we compute the optimal cost for the given set of assignments—say, A —by solving Model (1)–(15) after fixing the values for the assignment variables x as

$$x_{ij} = \begin{cases} 1 & \text{if } (i, j) \in A \\ 0 & \text{otherwise} \end{cases} \quad \forall (i, j) \in N \times N.$$

As before, we compare the optimal cost with the best upper bound and update the latter and the incumbent if necessary. In this case, we add the following cut in order not to consider the same set of assignments again:

$$\sum_{(i, j) \in A} x_{ij} \leq |N| - 1. \quad (34)$$

Clearly, H2 requires to solve a simpler (more restricted) problem, which demands less computational effort. However, in this case the resulting integer-feasible solutions can be worse, and the no-good cuts (Codato and Fischetti 2006) are much weaker, compared with those obtained with H1. So one needs to

carefully consider the trade-off between the level of computational effort to solve the restricted problems and the quality of the integer-feasible solutions as well as the strengths of the no-good cuts. In Section 5.2 we explain the details of our implementation to enhance the computational performance of the BC algorithm for the problem instances we study in this paper.

4.3. Valid Inequalities

We also consider valid inequalities to strengthen our formulation to improve the computational efficiency of BC. Note that, with the removal of continuous variables y_{ijkl} and f_{π}^{kl} , the cost of interhub transfers can simply be avoided by the optimizer, by setting all u_{π}^v to 0 at the root node at the initial relaxations, giving a quite loose LP relaxation bound. As such, we use the following valid inequalities to strengthen our model without the mentioned continuous variables:

$$\sum_{\pi \in \Pi_{kl}} \sum_{v \in V_{\pi}} u_{\pi}^v \geq \left\lceil \frac{w_{kl}}{\max_{v \in \bigcup_{\pi \in \Pi_{kl}} V_{\pi}} Q_v} \right\rceil (x_{kk} + x_{ll} - 1) \quad (35)$$

$\forall (k, l) \in D.$

As a logical cut, Inequalities (35) simply enforce that if there is a positive amount of traffic between a pair of hubs, then a set of hub routes, with a total capacity that is no less than the traffic between the two cities, should be activated. Our numerical studies have shown that the inclusion of (35) can drastically increase the LP relaxation bound and speed up the convergence of the branch-and-cut algorithm.

Note that the valid inequalities (35) are based on the hub location decisions. However, one can also consider similar cuts by considering the assignment decisions that are either directly indicated by the assignment variables or implied by the access path variables. Respective families of valid inequalities can be written as follows:

$$\sum_{\pi \in \Pi_{kl}} \sum_{v \in V_{\pi}} u_{\pi}^v \geq \left\lceil \frac{w_{kl} + w_{kj} + w_{il} + w_{ij}}{\max_{v \in \bigcup_{\pi \in \Pi_{kl}} V_{\pi}} Q_v} \right\rceil (x_{ik} + x_{jl} - 1) \quad (36)$$

for all distinct $i, j, k, l \in N$, and

$$\sum_{\pi \in \Pi_{kl}} \sum_{v \in V_{\pi}} u_{\pi}^v \geq \left\lceil \frac{\sum_{i \in \rho} \sum_{j \in \rho'} w_{ij}}{\max_{v \in \bigcup_{\pi \in \Pi_{kl}} V_{\pi}} Q_v} \right\rceil (z_{\rho} + z_{\rho'} - 1) \quad (37)$$

for all $\rho \in P_k$ and $\rho' \in P_l$ and distinct k and l in N .

Clearly, Inequalities (36) and (37) can potentially cut off integer-infeasible solutions (with inadequate interhub transfer capacities) that do not violate (35). However, their numbers are also large, and when the

capacities of the vehicles that can be used in the interhub transfers are large (as we have in our numerical experiments), then the coefficient on the right-hand side turns out to be 1 for most of these inequalities. In this case, (36) and (37) are implied by (35) because $x_{ik} \leq x_{kk}$ for all $i, k \in N$ and $z_{\rho} \leq x_{kk}$ for all $k \in N, \rho \in P_k$.

4.4. Remarks

Note that the projection and the resulting branch-and-cut algorithm can be also considered as a Benders decomposition approach where continuous variables are handled in a subproblem that provides feasibility cuts. In our preliminary studies we observed that because of long run times and large memory needed for solving the respective Benders subproblem, which is a large LP model, the classical Benders decomposition approach does not work well for the problem instances we consider in this study. Different from the classical approach, instead of solving the subproblem as a linear program, we first use a heuristic to find violated feasibility cuts, and we solve the subproblem in an exact manner only when the heuristic fails.

Because of various service quality considerations, in many real-world applications, the number of stopovers on access and hub paths and the travel times on these paths are limited. Hence, the number of feasible access and hub paths does not grow excessively, and the BC algorithm can be used to solve large-sized problem instances, as we show in our numerical experiments in the next section. Obviously, as the service quality restrictions become tighter (i.e., guaranteed delivery lead times are shortened), the number of access and hub paths is reduced, and BC can solve the resulting problems faster. On the other hand, when the service quality restrictions are loose, the number of hub and access paths can grow prohibitively large. In such settings, one can consider combining the BC algorithm with a column generation approach to generate access and hub paths as needed, or heuristic approaches can be developed to pick a subset of the feasible access and hub paths in the model. Clearly, the entailed computational difficulties for successful implementation of these ideas pose new research questions in their own right. Keeping our focus on the practical applications with tight service quality restrictions (i.e., next-day delivery), we leave it to future studies to explore such extensions.

5. Computational Study

In order to investigate the potential applications and the computational efficacy of our solution methodology, we have conducted comprehensive numerical experiments, composed of problem instances from two widely used data sets in the literature: Turkish and

CAB data sets. In the following sections, we discuss the goals of the study, the instances used, the analysis of the results, and the insights obtained.

5.1. Experimental Design

In this section we focus on a time-constrained package delivery network design problem faced by the package carriers to offer next-day/next-morning delivery services. As a test bed we consider the problem instances in Turkey where multiple transportation modes (vans, trucks, planes, rented aircraft belly capacity, etc.) with different costs, speeds, and capacities need to be used in coordination to meet a stringent service time requirement over a large geographical area in a cost-effective way (Yaman, Karasan, and Kara 2012). We also study problem instances from the CAB data set, to have a more comprehensive computational effectiveness analysis of the methodology.

With the apparent cost advantages, vans and trucks are preferred for short-/medium-distance access and hub routes, whereas cargo planes and rented aircraft belly capacities are needed to execute the long-distance hub routes to meet the delivery time constraints. In order to achieve load consolidation and facilitate mode switches, strategically located hubs are needed to be employed in the network. Clearly, imposing time limits on the access and hub routes is essential for meeting the service time requirements, achieving sustainability, and abiding by the regulations (Barnhart and Schneur 1996; Armacost, Barnhart, and Ware 2002; Yildiz and Savelsbergh 2019). We next present the details of the experimental setting that aims to answer, among other things, the following questions:

- What are the basic characteristics of the (optimal) networks?
- What are the benefits of allowing stopovers and multiple legs in the access and hub routes?
- What is the gain of using rented aircraft belly capacity?
- What is the computational efficacy of the proposed methodology?

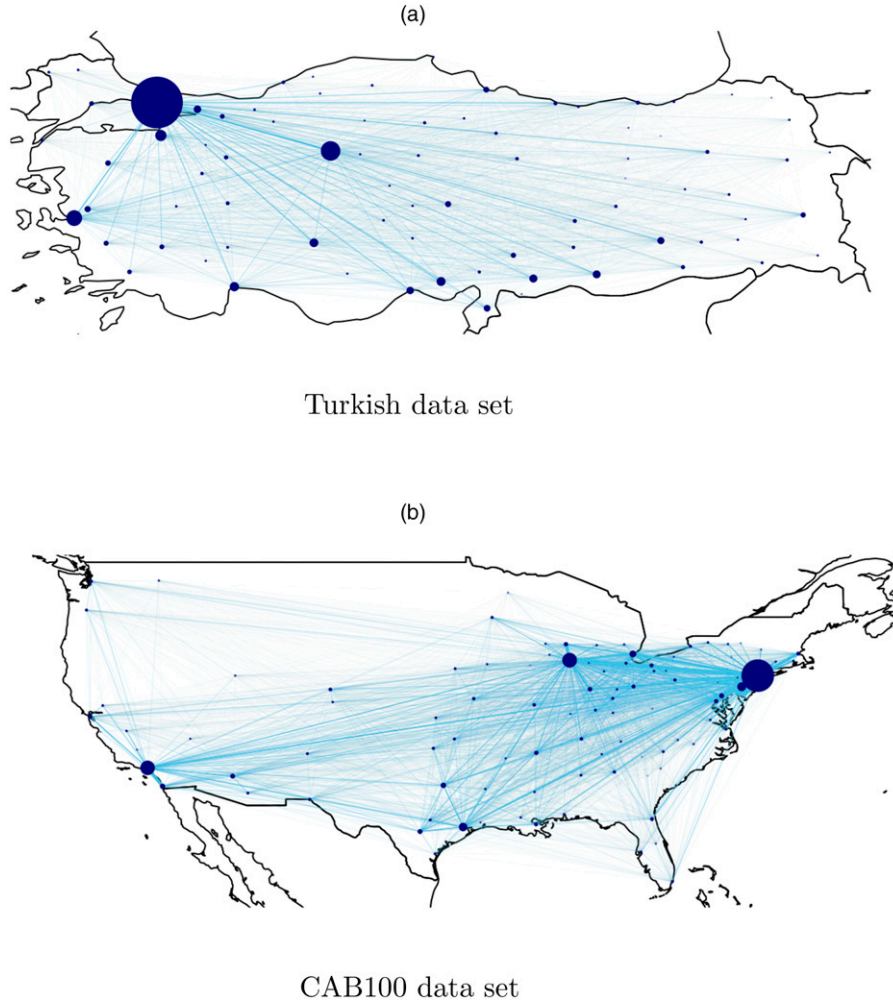
To answer such questions, we consider 39 different problem instances from the Turkish data set and 27 problem instances from the CAB data set that are generated by varying the basic problem parameters as we discuss in the following.

5.1.1. Turkish Data Set Instances. All instances from the Turkish data set consider cargo demand between 81 cities of Turkey. For the distances between the cities and the weights for the flow between them (normalized next day delivery demand volume), we use the Turkish network data (Kara 2009). In Figure 3(a)

we illustrate the locations of the cities (dots with sizes drawn proportional to the total in and outflow) and the normalized weight of next-day delivery demand among them (edge thicknesses are proportional to the normalized flow weights). All 81 cities are considered to be candidate hub locations.

Using the cost structure suggested in Serper and Alumur (2016), we consider three types of vehicles for the access routes: vans, trucks, and trailers with capacities of 3.5, 15, and 25 tons; fixed costs of 250, 309, and 364; and variable costs (per kilometer) of 0.5, 1, and 1.2. The capacities for the owned cargo planes are assumed to be 18 tons, with a fixed operating cost of 5,000 and variable cost of 6.4 (per kilometer). The per-kilometer cost of renting 5 tons of belly capacity is also assumed to be 6.4, which makes using belly capacity more advantageous for relatively small load amounts. Up to two slots of belly capacity (total of 10 tons) can be rented by the cargo company. We consider aircraft belly capacity only for the inbound and outbound flights to/from Istanbul, Ankara, and Izmir airports (the largest airports in Turkey with direct commercial flights to all other airports). For vans, trucks, and trailers, the vehicle speeds are assumed to be 80 km/hour, and for the passenger and cargo planes, we consider the ground speed of 750 km/hour. For each access route stop we add 45 minutes to the total travel time of the path (because of inner city traffic and loading/unloading operations). We only impose a total time limit on the access routes and do not directly limit the number of stops on them. For the hub route stops we also use 45 minutes for the needed loading/unloading, sorting, and repackaging operations. The flight times are calculated by dividing the great-circle distance between the origin and destination cities (found by the haversine formula) to the ground speed of the plane and adding 40 minutes to account for the time spent during the taxi and waiting for takeoff and landing permissions. Deducting six hours for inner city pickup, delivery, sorting, and packaging operations, we consider the time limit of $T = 18$ hours to complete all transfers.

5.1.2. CAB Data Set Instances. All instances from the CAB data set consider cargo demand between the 100 U.S. cities listed in O’Kelly (2017). All cities are assumed to be candidates to be chosen as hubs, and we use the great-circle distances between them to determine the distances. The weights of the flow between the cities are calculated with a gravity model in which we multiply the populations of the cities (as indicated in the data set) and divide it by the distance. In Figure 3(b) we illustrate the locations of the cities (dots with sizes are drawn proportional to the total in- and outflow) and

Figure 3. (Color online) The Spatial Distribution of Demand in Turkish and CAB Data Set Instances

the normalized weight of next-day delivery demand among them (edge thicknesses are proportional to the normalized flow weights). The rest of the problem parameters are used as defined for the aforementioned Turkish data set instances.

To generate different problem instances, we consider the following variations of the respective problem parameters:

- For the number of hubs, we consider $p \in \{3, 4, 5, 6\}$ for the Turkish data set instances and $p \in \{10, 11, 12, 13\}$ for the CAB data set instances.
- For the total daily next-day delivery demand (in tons), we consider $W \in \{100, 125, 150\}$.
- For the time allocations of the access and hub routes, we consider $(T_a, T_h) \in \{(7.5, 3), (7, 4), (6.5, 5), (6, 6)\}$ and disregard those problem instances when the number of hubs is too small to ensure feasibility for the given access and hub route time limits.

5.2. Implementation Details

In our preliminary studies we have observed that when the assignment and/or hub location decisions are fixed, the resulting *restricted* problem can be solved in a short time. To take advantage of this to speed up the whole solution process, we follow the following steps in our implementation.

We replace (1) with minimize Z and add the following constraint to our model:

$$Z \geq \sum_{\pi \in \Pi} \sum_{v \in V_\pi} c_\pi^v u_\pi^v + \sum_{\rho \in \mathcal{P}} c'_\rho z_\rho. \quad (38)$$

For each solution candidate (found by CPLEX during the branch-and-cut procedure), we first use the repair heuristic H2 and H1, in sequence (with a time limit of 150 seconds for H1), and add the respective no-good cuts to the model, if found. If H1 stops without finding the optimal solution for the restricted problem

(because of the time limit), we still use the best lower bound value L (found in H1 after 150 seconds) and the set of opened hubs N' in the solution candidate to generate the following cut:

$$Z \geq L \left(\sum_{h \in N'} z_{hh} - (p-1) \right). \quad (39)$$

Our algorithm is implemented in Java using IBM ILOG Concert technology with CPLEX 12.10 solver with its default settings. All experiments are run on a Linux workstation with 96 GB RAM and Intel Xeon Gold 6134 processor at 3.20 GHz. The time limit is set to two hours.

5.3. Analysis

In this subsection we present the results of our numerical experiments and discuss the computational and managerial insights we derive from them.

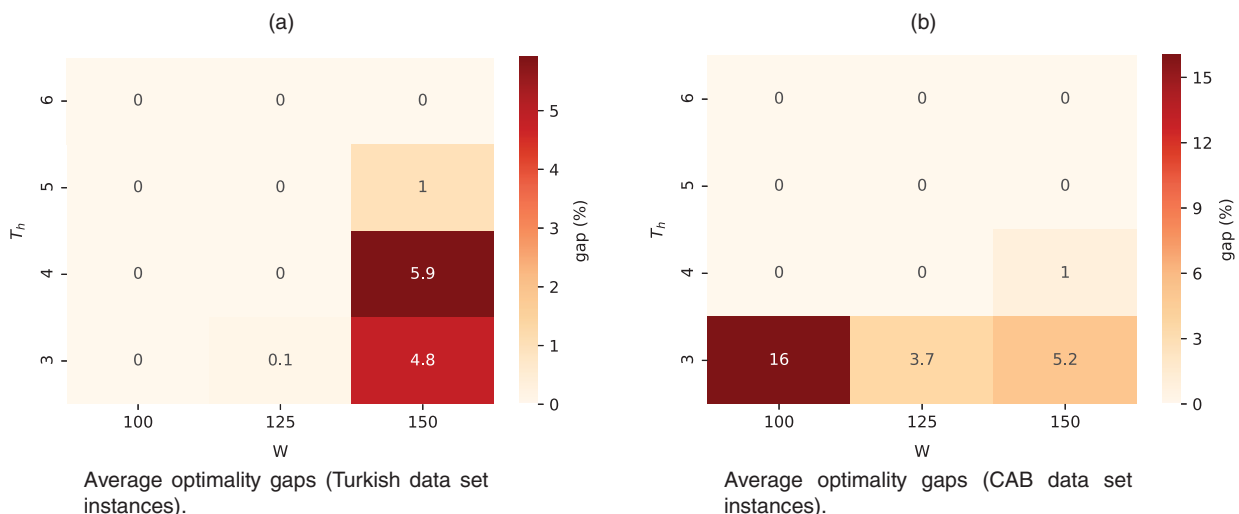
The detailed results about the computational performance of the BC algorithm in the Turkish and CAB data set instances are presented in Tables 1 and 2 in Online Appendix A, and the solution statistics are provided in Tables 3 and 4 in Online Appendix B. Details about the table headers are provided in the respective appendixes. In the following we summarize and discuss the critical insights.

We first focus on the computational performance of the BC algorithm. Our results in Tables 1 and 2 (in the online appendix) clearly indicate the computational effectiveness of the algorithm in solving problem instances from both the Turkish and CAB data sets. We see that for more than 79% of the Turkish data set instances (31 of the 39 instances), the BC algorithm could find an optimal solution within the given time limit, and for those instances where the algorithm stopped reaching the time limit, the average optimality gap is less than 5.8%. We see a similar performance

in the CAB instances, where more than 74% of the instances (20 of the 27 instances) are solved to optimality within the given time limit. As we see in Figure 4, which illustrates the distribution of optimality gaps averaged over all p -values, for different levels of demand volume and time allocated for interhub transfers, problem instances with small T_h values are hardest to solve for both network topologies. However, the impact of the total demand volume on the problem difficulty is not the same in the two networks. For the Turkish data set instances, the problem difficulty increases with the total demand volume, whereas the hardest instances from the CAB data set are the ones with $W = 100$.

Investigating Table 1 in the online appendix, we can see that the high computational performance of the BC algorithm can be explained by two basic factors: a relatively small number of (separation) cuts added during the execution of the branch-and-cut method and the high efficiency of BC to solve the separation problem with the suggested two-phase approach. As we can observe from Table 1, the total number of cuts added during the algorithm (2,878, on the average) is quite small (relative to the problem sizes), and most of the time the heuristic approach can solve the separation problem (on average, 74.8% of the cuts are found by the heuristic separation approach, which eliminates the need for resorting to computationally (more) expensive exact separation—including the no-good cuts generated by H1 and H2). A similar conclusion can be drawn from Table 2 in the online appendix, which shows that more than 85% of the cuts are found by the heuristic separation approach in CAB data set instances. In Figure 5 we plot the number of violated cuts, averaged over all p -values, and what percentage of those cuts are found by the heuristic separation. Inspecting Figures 4 and 5, we clearly see that the

Figure 4. (Color online) Average Gap Values

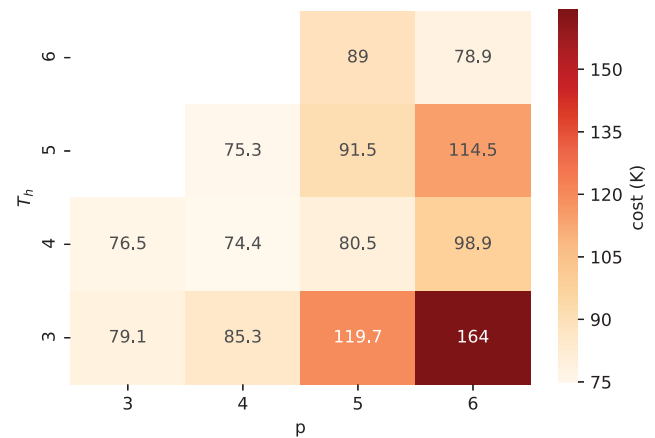


number of cuts needed to be added by the BC algorithm is one of the most important determinants of the problem difficulty, and the use of heuristic separation provides significant computational advantages to solve those hard problems. It is also interesting to see that the number of cuts needed to be added to the model is higher in Turkish data set instances. The reason for such a result is due to the differences in sizes of the geographical areas. Because the CAB data set covers a much larger geography, a larger number of hubs is needed to have feasible solutions. As such, Inequality (35) becomes much stronger in CAB instances.

We next examine the structure of the optimal solutions obtained by our model. In Figure 6 we present the optimal solution values for the TR0–TR12 instances (see Table 3 in the online appendix for respective problem parameters). As we can see from the graph, the total cost is lowest when the interhub transfer time is four hours ($T_h = 4$) and four hubs are opened ($p = 4$). On the other hand, $T_h = 3$ and $p = 6$ has the highest cost (more than two times larger than the one for $T_h = 4$ and $p = 4$). These results clearly show that the number of hubs and the interhub transfer times should be chosen carefully so that multileg interhub transfers can be effectively used to compensate for the reduced consolidation opportunities as the number of hubs are increased (to reduce the total costs of the access paths).

To take a closer look at the structure of the optimal solutions, in Figure 7 we illustrate the particular designs for TR0, TR5, TR8, and TR12 instances, which have the smallest total costs for $T_h = 3, 4, 5$, and 6 among the instances with $W = 100$ (i.e., TR0–TR12). For the access routes, the line thicknesses are proportional to the total capacity of the vehicles assigned to

Figure 6. (Color online) Optimal Costs for TR0–TR12 Instances



the respective paths. For the hub routes, the straight lines indicate the paths that use trucks, and dot-dashed (thicker) curves illustrate the routes run with owned cargo planes.

Inspecting Figure 7, we observe that the optimal designs may favor hub locations with smaller interhub path distances, even though this may necessitate longer access routes. For example, in Figure 7(c) we see that instead of Eskişehir (the hub in the west), an alternate location further west could have shortened most of the incoming access paths. But then the multileg hub paths from this hub would become infeasible, and more expensive interhub paths would need to be used. In fact, in Figure 7(e) we see that this actually happens, when T_h is increased to five hours (from four in Figure 7(c)). In our setting, when land transportation is considered, interhub transportation can still be less costly because larger vehicles with higher utilization levels can be used in hub routes. However,

Figure 5. (Color online) Average Number of Cuts Added During the Branch and Cut and the Percentage of Cuts Found by the Heuristic Separation Procedure

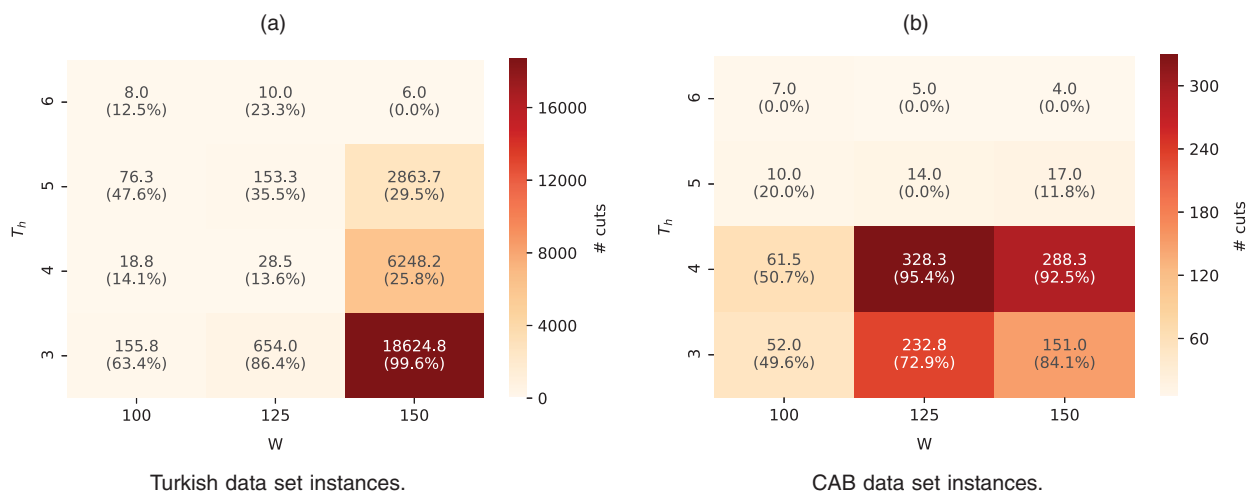
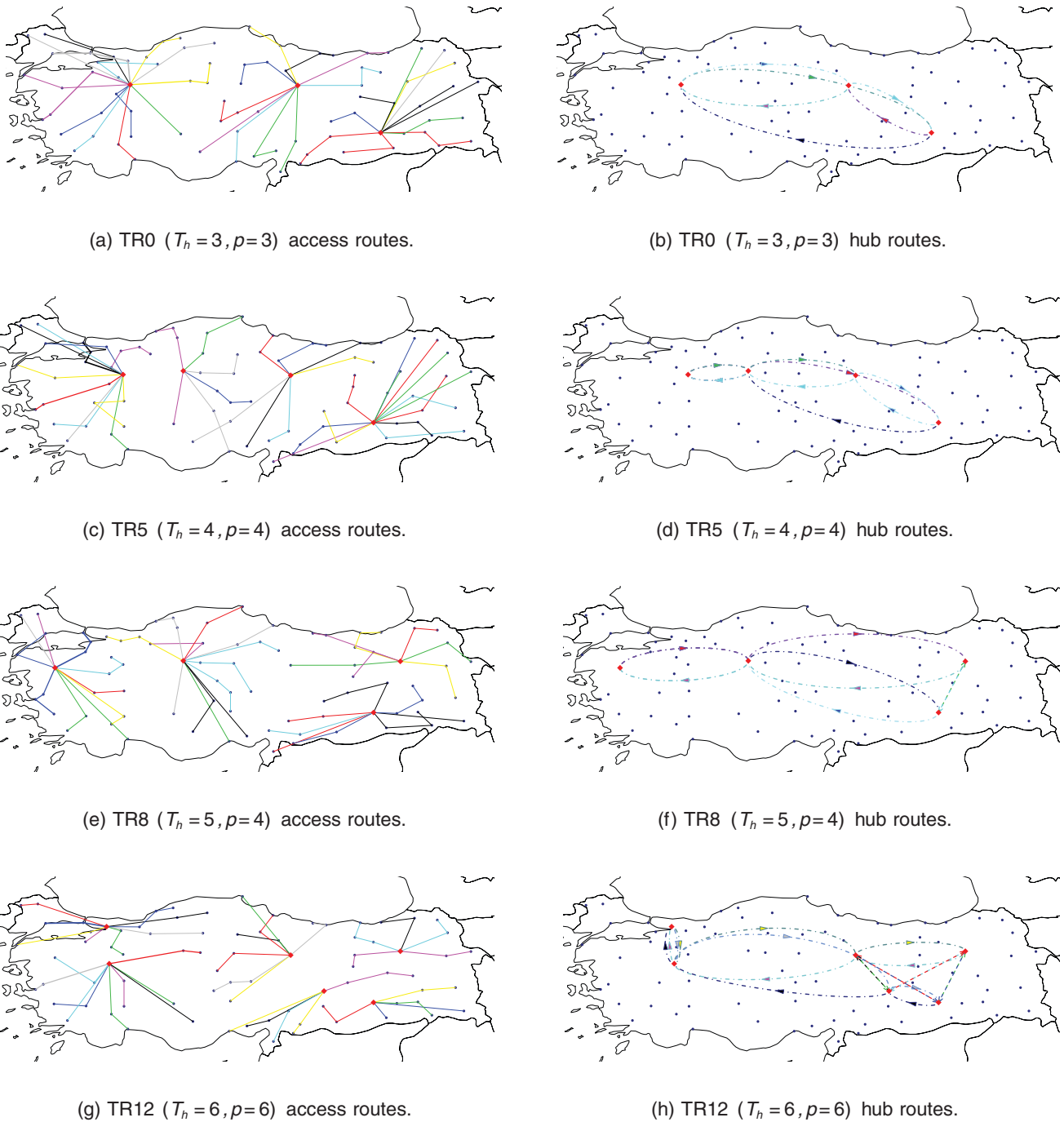


Figure 7. (Color online) Illustration of the Solutions for TR0, TR5, TR8, and TR12 Problem Instances



the time restriction does not allow using trucks between many interhub trips; hence more expensive air transportation (provided by rented aircraft belly capacity or owned cargo aircraft) needs to be used. Moreover, because we allow stopovers, access routes already provide opportunities for load consolidation and better use of vehicle capacities. For these reasons, the interhub transportation becomes more expensive

(in general), and network designs with closely located hubs are favored. Related to this observation, we also see in Table 3 (in the online appendix) that $p = 3$ and $p = 4$ typically give lower-cost solutions compared with higher p -values. With a smaller number of hubs, higher levels of load consolidation become possible in interhub transportation and utilization rates of the vehicles increase. Considering the fact that many hub

routes use air transportation with high costs, the importance of high-capacity utilization in those paths is obvious. We can clearly see this relation in Figure 8, which presents the cost and vehicle utilization percentages for access and hub routes in the optimal solutions, for different p -values considered in problem instances TR4–TR7, where we have $W = 100$ and $T_h = 4$. As we see in Figure 8(a) the cost of interhub transportation increases dramatically beyond $p = 4$, whereas the reduction in the access route costs is rather limited compared with this increase. We also see that the vehicle utilization levels drop significantly (from 56% to 14%) in the hub routes as p increases from 3 to 6, where the access route vehicle utilizations stay almost the same. As we see in Table 3 the same conclusion remains valid for different values of total demand volume W . More than that, we see that the optimal hub locations stay the same for the different W values when p and T_h are fixed. Related to this observation, we observe that, on average, the total costs increase less than 7% when the demand volume is increased from 100 tons to 150 tons. The reason for such a result is the availability of redundant capacity on both the access and interhub routes (as can be observed from the capacity utilization levels), which can be used to support higher levels of demand without increasing the cost.

Different from classical hub network design models, our approach considers multiple-stop access and hub routes as well as a multimodal structure enhanced by the inclusion of the rented aircraft belly capacity. We now investigate the impact of having these flexibilities on the total cost. In Table 5 in Online Appendix C, we present the detailed solution statistics for the instances TR0–TR12, when we consider (1) single-leg hub routes, (2) single-leg access routes

(no stopovers), and (3) no-belly capacities cases separately. However, the critical information is captured in Figure 9, where we illustrate the cost increases (compared with the base case with no restrictions, as reported in Table 3 in the online appendix) for each restriction type (single-leg access routes, single-leg hub routes, and no-belly capacity), for each problem instance.

The first thing one notices looking at Figure 9 is the drastic cost increase when access or hub routes are restricted to one-leg trips and $T_h \geq 4$. For the hub routes, this result is not very surprising, because the fixed costs of the cargo planes are quite high, and even for the relatively small T_h values, the benefit of using the same aircraft in multiple legs is obvious. As an interesting insight, a closer look at Table 5 (in the online appendix) reveals that restricting interhub transfers to single-leg transfers makes belly capacity more attractive, and the optimal solution starts to use belly capacity heavily (almost three times more than the base case with no restrictions). We also see that restricting the access routes to single stops has a significant impact on the total cost, which supports our previous discussions that with stopovers, access routes provide a significant load consolidation opportunity to reduce the costs (up to 25% in the considered instances). As expected, the cost increase with single-leg access paths is higher for smaller p -values, because the distances between the hubs and the spokes (i.e., average length of the access paths) increase, and consolidation becomes more critical in those cases.

Focusing on the role of the belly capacity, we see from Figure 9 that it can provide solid cost benefits (up to 30% savings) when the interhub transfer time is low and number of hubs is high.

Figure 8. (Color online) Access and Hub Route Costs and Vehicle Utilization Levels for Different p -Values

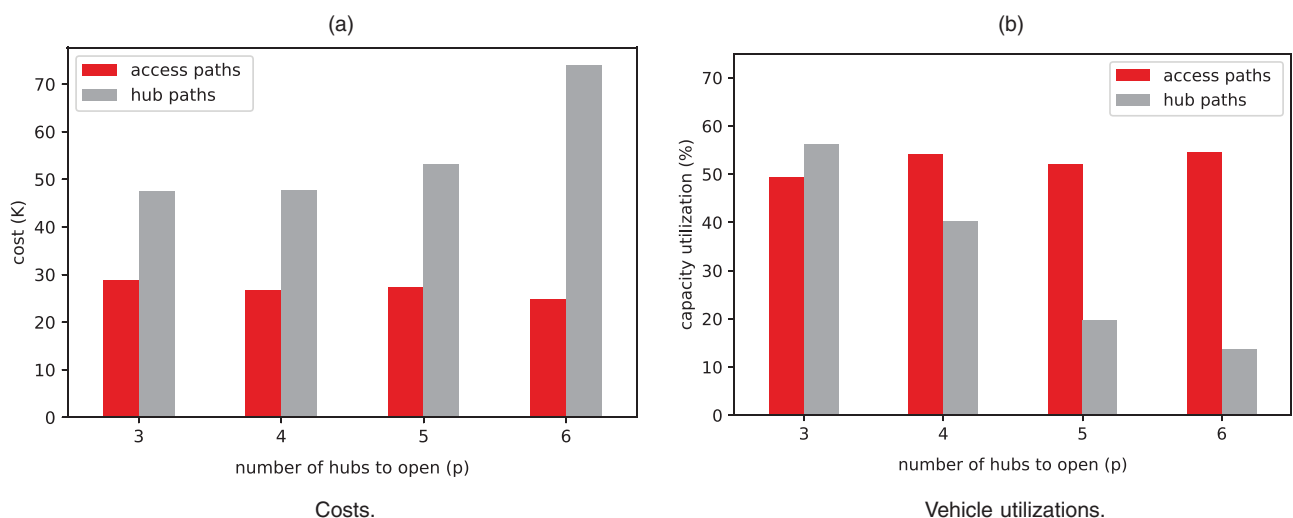
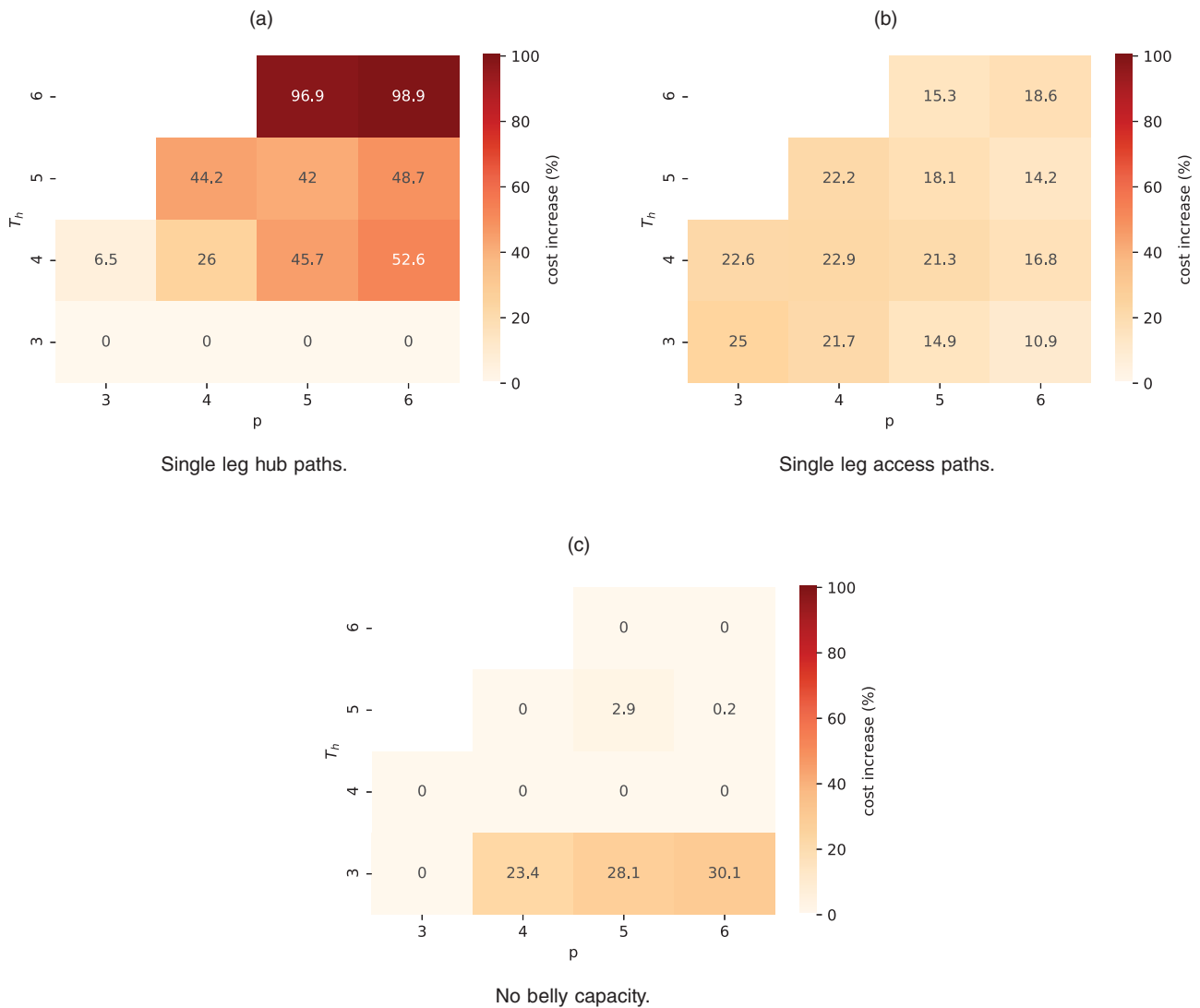


Figure 9. (Color online) Cost Increases with Traditional Hub Location Assumptions



6. Final Remarks

In this study we develop a new hub network design model that makes few assumptions on the structure of the network and relaxes the assumption of a constant cost reduction factor between hubs. The novelty of the approach lies in its linkage of the strategic and tactical-level decisions related to hub layout and transportation network design to jointly determine the locations of the hubs, hub assignments, types and number of vehicles to use, and routes to operate. We show in our numerical experiments that our model brings important practical benefits in utilizing the vehicle capacities, improving load consolidation in both access and hub routes and taking better advantage of multimodal transportation capabilities. Our results also show that hubs can play critical roles not only by facilitating load consolidation but also by functioning as transfer points to enable end-to-end intermodal

transportation. Besides the managerial insights, these experiments also show that with the recent advances in integer programming methodologies, it is possible to solve realistically sized problem instances without resorting to two-stage solution heuristics or restrictive assumptions on the solutions. This suggests that developing new hub network design models that can answer the specific needs of the novel logistics and mobility concepts such as the physical internet and mobility as a service provides rich opportunities for potential high-impact research in hub network design.

Acknowledgments

The authors acknowledge that the research of the second author was partly done while she was at Bilkent University. The authors thank the associate editor and the referees whose suggestions helped in improving the authors' contribution.

References

- Alumur SA (2019) Hub location and related models. Eiselt HA, Marianov V, eds. *Contributions to Location Analysis: In Honor of Zvi Drezner's 75th Birthday* (Springer International Publishing, Cham, Switzerland), 237–252.
- Alumur S, Kara BY (2008) Network hub location problems: The state of the art. *Eur. J. Oper. Res.* 190(1):1–21.
- Alumur SA, Kara BY, Karasan OE (2009) The design of single allocation incomplete hub networks. *Transportation Res. Part B: Methodological* 43(10):936–951.
- Alumur SA, Kara BY, Karasan OE (2012) Multimodal hub location and hub network design. *Omega* 40(6):927–939.
- Alumur SA, Yaman H, Kara BY (2012) Hierarchical multimodal hub location problem with time-definite deliveries. *Transportation Res. Part E: Logist. Transportation Rev.* 48(6):1107–1120.
- Alumur SA, Campbell JF, Contreras I, Kara BY, Marianov V, O'Kelly ME (2021) Perspectives on modeling hub location problems. *Eur. J. Oper. Res.* 291(1):1–17.
- Armacost AP, Barnhart C, Ware KA (2002) Composite variable formulations for express shipment service network design. *Transportation Sci.* 36(1):1–20.
- Barnhart C, Schneur RR (1996) Air network design for express shipment service. *Oper. Res.* 44(6):852–863.
- Calik H, Alumur SA, Kara BY, Karasan OE (2009) A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Comput. Oper. Res.* 36(12):3088–3096.
- Campbell JF (1996) Hub location and the p -hub median problem. *Oper. Res.* 44(6):923–935.
- Campbell JF (2009) Hub location for time definite transportation. *Comput. Oper. Res.* 36(12):3107–3116.
- Campbell JF (2013) Modeling economies of scale in transportation hub networks. 2013 46th Hawaii Internat. Conf. System Sci., vol. 1 (IEEE, Piscataway, NJ), 1154–1163.
- Campbell JF, O'Kelly ME (2012) Twenty-five years of hub location research. *Transportation Sci.* 46(2):153–169.
- Campbell JF, Ernst AT, Krishnamoorthy M (2005a) Hub arc location problems: Part I—Introduction and results. *Management Sci.* 51(10):1540–1555.
- Campbell JF, Ernst AT, Krishnamoorthy M (2005b) Hub arc location problems: Part II—Formulations and optimal algorithms. *Management Sci.* 51(10):1556–1571.
- Çetiner S, Sepil C, Süral H (2010) Hubbing and routing in postal delivery systems. *Ann. Oper. Res.* 181(1):109–124.
- Codato G, Fischetti M (2006) Combinatorial Benders' cuts for mixed-integer linear programming. *Oper. Res.* 54(4):756–766.
- Contreras I (2015) Hub location problems. Laporte G, Nickel S, Saldanha da Gama F, eds. *Location Science* (Springer International Publishing, Cham, Switzerland), 311–344.
- Contreras I, O'Kelly M (2019) Hub location problems. Laporte G, Nickel S, Saldanha da Gama F, eds. *Location Science*, 2nd ed. (Springer International Publishing, Cham, Switzerland), 327–363.
- Contreras I, Fernández E, Marín A (2010) The tree of hubs location problem. *Eur. J. Oper. Res.* 202(2):390–400.
- Corberán Á, Peiró J, Campos V, Glover F, Martí R (2016) Strategic oscillation for the capacitated hub location problem with modular links. *J. Heuristics* 22(2):221–244.
- Crainic TG, Montreuil B (2016) Physical internet enabled hyperconnected city logistics. *Transportation Res. Proc.* 12:383–398.
- de Camargo RS, de Miranda G, Løkketangen A (2013) A new formulation and an exact approach for the many-to-many hub location-routing problem. *Appl. Math. Model.* 37(12–13):7465–7480.
- de Camargo RS, de Miranda G Jr, Luna HPL (2009) Benders decomposition for hub location problems with economies of scale. *Transportation Sci.* 43(1):86–97.
- de Camargo RS, de Miranda G Jr, O'Kelly ME, Campbell JF (2017) Formulations and decomposition methods for the incomplete hub location network design problem with and without hop-constraints. *Appl. Math. Model.* 51(November):274–301.
- Ebery J, Krishnamoorthy M, Ernst A, Boland N (2000) The capacitated multiple allocation hub location problem: Formulations and algorithms. *Eur. J. Oper. Res.* 120(3):614–631.
- Farahani RZ, Hekmatfar M, Arabani AB, Nikbakhtsh E (2013) Hub location problems: A review of models, classification, solution techniques, and applications. *Comput. Indust. Engrg.* 64(4):1096–1109.
- Gelareh S, Pisinger D (2011) Fleet deployment, network design and hub location of liner shipping companies. *Transportation Res. Part E: Logist. Transportation Rev.* 47(6):947–964.
- Hietanen S (2014) “Mobility as a service”—The new transport model? *Eurotransport* 12(2):2–4.
- Horner MW, O'Kelly ME (2001) Embedding economies of scale concepts for hub network design. *J. Transport Geography* 9(4):255–265.
- Jaillet P, Song G, Yu G (1996) Airline network design and hub location problems. *Location Sci.* 4(3):195–212.
- Joerss M, Schröder J, Neuhaus F, Klink C, Mann F (2016) Parcel delivery: The future of last mile. Report, McKinsey & Company, New York.
- Kara BY (2009) Turkish network. Accessed January 30, 2019, <https://ie.bilkent.edu.tr/~bkara/dataset.php>.
- Kartal Z, Hasgul S, Ernst AT (2017) Single allocation p -hub median location and routing problem with simultaneous pick-up and delivery. *Transportation Res. Part E: Logist. Transportation Rev.* 108(December):141–159.
- Kartal Z, Krishnamoorthy M, Ernst AT (2019) Heuristic algorithms for the single allocation p -hub center problem with routing considerations. *OR Spectrum* 41(1):99–145.
- Labbé M, Yaman H (2004) Projecting the flow variables for hub location problems. *Networks* 44(2):84–93.
- Labbé M, Yaman H, Gourdin E (2005) A branch and cut algorithm for hub location problems with single assignment. *Math. Programming* 102(2):371–405.
- Lopes MC, de Andrade CE, de Queiroz TA, Resende MGC, Miyazawa FK (2016) Heuristics for a hub location-routing problem. *Networks* 68(1):54–90.
- Maheo A, Kilby P, Van Hentenryck P (2017) Benders decomposition for the design of a hub and shuttle public transit system. *Transportation Sci.* 53(1):77–88.
- Montreuil B (2011) Toward a physical internet: Meeting the global logistics sustainability grand challenge. *Logist. Res.* 3(2–3):71–87.
- Nagy G, Salhi S (1998) The many-to-many location-routing problem. *TOP* 6(2):261–275.
- Nickel S, Schöbel A, Sonneborn T (2001) Hub location problems in urban traffic networks. Pursula M, Niittymäki J, eds. *Mathematical Methods on Optimization in Transportation Systems* (Springer, Boston), 95–107.
- O'Kelly ME (1986a) Activity levels at hub facilities in interacting networks. *Geographical Anal.* 18(4):343–356.
- O'Kelly ME (1986b) The location of interacting hub facilities. *Transportation Sci.* 20(2):92–106.
- O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. *Eur. J. Oper. Res.* 32(3):393–404.
- O'Kelly ME (2017) citylist: For the CAB 100 data. Working paper, The Ohio State University, Columbus, Ohio.
- O'Kelly ME, Bryan D (1998) Hub location with flow economies of scale. *Transportation Res. Part B: Methodological* 32(8):605–616.
- O'Kelly ME, Miller HJ (1994) The hub network design problem: A review and synthesis. *J. Transport Geography* 2(1):31–40.
- Peiró J, Corberán Á, Laguna M, Martí R (2018) Models and solution methods for the uncapacitated p -allocation p -hub equitable center problem. *Internat. Trans. Oper. Res.* 25(4):1241–1267.

- Podnar H, Skorin-Kapov J, Skorin-Kapov D (2002) Network cost minimization using threshold-based discounting. *Eur. J. Oper. Res.* 137(2):371–386.
- Rodríguez-Martín I, Salazar-González J-J, Yaman H (2014) A branch-and-cut algorithm for the hub location and routing problem. *Comput. Oper. Res.* 50(October):161–174.
- Serper EZ, Alumur SA (2016) The design of capacitated intermodal hub networks with different vehicle types. *Transportation Res. Part B: Methodological* 86(April):51–65.
- Tanash M, Contreras I, Vidyarthi N (2017) An exact algorithm for the modular hub location problem with single assignments. *Comput. Oper. Res.* 85(September):32–44.
- Wasner M, Zäpfel G (2004) An integrated multi-depot hub-location vehicle routing model for network planning of parcel service. *Internat. J. Production Econom.* 90(3):403–419.
- Yaman H (2005) Polyhedral analysis for the uncapacitated hub location problem with modular arc capacities. *SIAM J. Discrete Math.* 19(2):501–522.
- Yaman H, Carello G (2005) Solving the hub location problem with modular link capacities. *Comput. Oper. Res.* 32(12):3227–3245.
- Yaman H, Kara BY, Tansel BC (2007) The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Res. Part B: Methodological* 41(8):906–919.
- Yaman H, Karasan OE, Kara BY (2012) Release time scheduling and hub location for next-day delivery. *Oper. Res.* 60(4):906–917.
- Yildiz B, Savelsbergh M (2019) Optimizing package express operations in China. Preprint, submitted September 17, http://www.optimization-online.org/DB_FILE/2018/09/6799.pdf.
- Yoon M-G, Current J (2008) The hub location and network design problem with fixed and variable arc costs: Formulation and dual-based solution heuristic. *J. Oper. Res. Soc.* 59(1):80–89.