THE CAPACITATED DIRECTED CYCLE HUB LOCATION AND ROUTING PROBLEM UNDER CONGESTION

Cihan Bütün[(1)], Sanja Petrovic[(2)*], Luc Muyldermans[(2

[(1)]Brunel Business School, Brunel University London, Uxbridge, Middlesex UB8 3PH, United Kingdom

[(2)]Nottingham University Business School, Jubilee Campus, Nottingham, NG81BB, United Kingdom

cihan.butun@brunel.ac.uk

sanja.petrovic@nottingham.ac.uk

luc.muyldermans@nottingham.ac.uk

ABSTRACT

This paper deals with hub-and-spoke network design in the liner shipping sector. It introduces a capacitated directed cycle hub location and cargo routing problem under congestion. The problem involves four decisions: location of hub ports; allocation of non-hub ports to hub ports; construction of a directed cyclic route at the hub port network level; and the routing of cargo between all origin-destination demand pairs in the network. The objective is to minimize the cost which includes fixed hub opening, feeder collection and distribution, inter-hub transportation, cargo handling, and non-linear hub port congestion costs. We present a mixed integer linear programming model in which the non-linear congestion costs at the hub ports are approximated through a (semi-continuous) piecewise linear function and use this model to calculate lower bounds on the objective function. We also develop a Tabu Search algorithm, which employs a hierarchical approach for the different decisions in the hub-and-spoke network design problem, with customized procedures for the generation of the initial solution and the selection of the search moves. The neighborhood search is diversified by randomly changing the locations of hubs based on their location frequency history in previous solutions. Computational experiments, using instances from the literature and problems based on real-world data, demonstrate that the algorithm finds high quality solutions in a reasonable time. The experiments show that the network design can be highly influenced by scale economies in mainline vs. feeder transportation costs, the port locations and hinterland flows, and congestion at the hub ports.

KEYWORDS: Location, Directed cycle hub location and routing problem, Network design under congestion, Tabu Search

* corresponding author

1. INTRODUCTION

Global shipping lines implement various strategies to cope with the increasing demand for containerized maritime transport, to remain competitive in the global environment, and to increase their market share (Cariou, 2008). A core strategy in liner shipping involves bundling containerized cargo flows through transshipment hub ports. Hub-and-spoke (HS) networks are widespread in liner shipping and provide several benefits, including scale economies in shipping costs, higher service quality, and additional transshipment volumes. Two important drawbacks of liner HS networks are the cost of extra cargo handling at transshipment points and potential congestion created at hubs due to concentration of flows.

The hub location problem is a network design problem concerned with four decisions: finding locations of the hub facilities, assigning non-hub nodes to hubs, establishing links between hubs, and routing the flows within the network. A widely accepted assumption in the literature is that the hub-level network forms a complete graph, with arcs connecting each pair of hubs. More recent studies relax this assumption because incomplete networks are more realistic and less costly in many transportation settings. Campbell et al. (2005a; 2005b), Calik et al. (2009), Alumur et al. (2009), Gelareh and Nickel (2011), Contreras and Fernandez (2014), De Camargo et al. (2017), and Martins de Sa et al. (2018) address problems where the hub-level network is not limited to any specific topology. Other authors impose a specific topology on the network. Examples include: a star structure (Yaman, 2008; Labbé and Yaman, 2008; Yaman, 2009; Yaman and Elloumi, 2012), a tree configuration (Contreras et al., 2009; 2010; Martins de Sa et al., 2013), a line structure (Martins de Sa et al., 2015a; 2015b), and cycle topologies (Lee et al., 1993; Gelareh and Pisinger, 2011; Contreras et al., 2016).

Despite the widespread use of HS networks in maritime transport, their application in liner shipping has been overlooked in the OR literature for many years. Research interest has surged more recently, with several publications focusing on different aspects of the industry (Table 1). Aversa et al. (2005), for instance, define a *p*-hub median problem to locate hubs at South American ports and consider the option of road transport between the ports. Takano and Arai (2008) develop a Genetic Algorithm and solve a *p*-hub median problem to locate hubs among 18 major global container ports. Gelareh et al. (2010) study HS network design in a competitive environment. They define a problem in which an entrant shipping company aims to maximize its market share in a region where an existing company has already established its HS network. Chou (2010) argue that shipping lines consider not only quantitative but also qualitative factors when choosing hub ports. He identifies five groups of factors and develops a fuzzy multiple criteria decision-making model. Gelareh and Pisinger (2011) consider profit maximization and their study is the only liner shipping application in which a directed cyclic topology is imposed on the hub-level network. Gelareh and Nickel (2011) argue that complete hub-level networks are rarely applicable in maritime transportation and also relax the complete hub-level network assumption. The impact of maritime cabotage on the liner shipping networks with transit time constraints is analyzed by Zheng et al. (2014). The authors propose a two-phase mathematical programming model in which hub locations and feeder allocations are determined first, and ship routes and fleet deployment in the second phase. A similar decomposition of the problem is proposed by Zheng et al. (2015) to design a HS network, route ships, and assign container flows. Because many studies of liner shipping network design focus on existing ports and network structures, Sun and Zheng (2016) claim that their results have limited applicability. To imitate the actual planning process of a liner service design, they model potential routes and hub locations on Arctic waterways. Finally, Zheng et al. (2018) study hub ports in a global shipping network. They propose a two-stage optimization method which first partitions the ports into communities, and then finds optimal hub locations and node allocations in each community.

Although HS networks are often preferred by the leading shipping companies, research on HS network design in liner shipping has attracted less attention from the hub location community compared with other industries such as aviation or postal delivery. Furthermore, hub congestion is an important disadvantage in HS networks. While consolidation of flows on the hub-level network brings scale economies, over-accumulating traffic in hubs leads to congestion. Networks designed by ignoring congestion are likely to be unrealistic and inefficient, but most existing studies have not addressed congestion at the hub ports

Table 1:Liner shipping applications in the hub location literature

Study	Main features of the research	Solution methodology
Aversa et al. (2005)	Transshipment and port costs Alternative road transport option	Mixed integer linear programming
Takano and Arai (2008)	Transshipment cost	Genetic algorithm
Gelareh et al. (2010)	Competitive environment Origin-destination paths include more than one hub-level arc Direct non-hub port connections	Lagrangean relaxation
Chou (2010)	Qualitative port choice criteria: port location, hinterland economy, physical factors, port efficiency, cost, and others.	Fuzzy multiple criteria decision making
Gelareh and Pisinger (2011)	Network design and fleet deployment Transshipment cost Directed cyclic hub-level network	Benders decomposition
Gelareh and Nickel (2011)	Incomplete hub-level network Fixed cost of establishing hub-level arcs	Benders decomposition Greedy neighborhood search
Zheng et al. (2014)	Maritime cabotage Two-phase model of hub location and ship route design and fleet deployment Transit time constraints	Lagrangean relaxation
Zheng et al. (2015)	Hub location, fleet deployment, and ship route design Transit time constraints Direct non-hub port connections Transshipment cost	Genetic algorithm
Sun and Zheng (2016)	Potential hubs on Arctic waterways	Branch-and-bound
Zheng et al. (2018)	Two-stage optimization Communities of hub and feeder ports	Community detection algorithm
This study	Hub port congestion cost Congestion cost approximation Directed cyclic hub-level network Transshipment cost	Tabu Search with local search

In the hub location literature, there are two prevalent approaches to modelling hub congestion. The first is to formulate congestion cost based on a power-law cost function in which the congestion cost of a hub changes proportionally with the total flow according to a constant exponent. An example of this approach is the study by Elhedhli and Hu (2005), who linearize the cost function with piecewise linear tangent hyperplanes and solve the resulting MILP model with a Lagrangean heuristic. In (De Camargo et al. 2011) and (De Camargo and Miranda 2012), the power-law function is used if the capacity utilization ratio exceeds a specified threshold. De Camargo et al. (2011) apply a hybrid of outer approximation and Benders decomposition methods, and De Camargo and Miranda (2012) use generalised Benders decomposition. Kian and Kargar (2016) incorporate congestion cost into the single allocation *p*-hub median problem and apply conic quadratic programming., Özgün-Kibiroglu et al. (2019) reformulate three

multiple allocation hub location models in which the congestion cost is represented with a penalty cost and develop a particle swarm optimization approach to solve the problem. The second approach to modeling congestion costs is by considering the hubs and OD flows as a queuing system. Hubs represent servers and OD flows represent customer demand streams. Elhedhli and Wu (2010) argue that congestion is more related to the relative difference between hub flow and hub capacity than to hub flow alone. They calculate the hub congestion cost based on the ratio of the total flow and the hub's surplus capacity. An alternative way to derive the congestion cost is by viewing hubs as M/M/1 queues and by calculating the system wide average waiting time. The authors propose a mixed integer nonlinear programming model; linearize the objective function using a piecewise linear function and apply a Lagrangean heuristic. De Camargo et al. (2011) replace the same congestion cost function with a linear function if the capacity utilisation exceeds a specified threshold. In (Rodriguez et al. 2007), the waiting times at hubs in a cargo transportation network are calculated by considering each hub as an M/M/1 queuing system. The arrival and service times refer to the number of trucks assigned to the hub and the maximum number of trucks that can be processed by the hub, respectively. Mohammadi et al. (2019) analyze reliable hub-and-spoke networks in a French cargo and passenger transportation network by considering road, rail, and air travel. Other studies in the literature rely on other queuing model results, including M/M/c (Mohammadi et al., 2011, 2016, 2017; Zhalechian et al., 2017; Kahag et al., 2019; Ghodratnama et al., 2019; Khodemani-Yazdi et al., 2019), M/M/c/K (Rahimi et al., 2016), M/D/c (Marianov and Serra, 2003), and GI/G/1 (Ishfaq and Sox, 2012). In addition, a few other studies followed different approaches. Köksalan and Soylu (2010) used the ratio of the demand to the hub capacity as the main factor of delay in service time, and Alumur et al. (2018) discretized congestion by defining sets of handling times at each hub for corresponding congestion factors.

Shipping companies often route their mainline ships on simple directed cyclic networks (Fig.1). In a simple directed cycle, ships visit each port in a tour. Empirical data reported by Song and Dong (2013) indicate that more than 40% of the global liner routes were simple directed cycles in 2009, which corresponds to 31% of containers carried on these routes. In 2012, 41% of the routes of Maersk shipping line were simple cycles (Reinhardt and Pisinger, 2012). Despite its practical relevance in liner shipping, very few papers in the literature have investigated HS networks with imposed cycle structure at the hub level.

The objective in this paper is to address the following gaps in the literature. We introduce the capacitated directed cycle hub location and routing problem under congestion (DCHC), which can be viewed as a prototype HS decision problem, where the hub level route is cyclic and hub congestion important (as is the case in many liner shipping network problems). Our problem involves the strategic planning of liner shipping service network design. As such, the tactical and operational decisions such as ship speed and schedules are outside the scope of the problem. The decision maker wants to exploit the scale economies offered by HS operations and to design a cost-optimal network. However, he/she is also aware that concentrating too much flow to a few hubs can create bottlenecks in the network and lead to congestion. Therefore, the hub port congestion costs are considered at the design stage, in which the only available information are origin-destination (OD) demands and cargo handling capacities of hubs.

Our study is the first to address and quantify hub port congestion in a hub location problem with liner shipping focus. It includes other features motivated by the practices in the industry such as the cost of cargo handling at transshipment. We develop a mixed integer nonlinear programming model and propose a procedure to approximate the nonlinear congestion cost with a semi-continuous piecewise linear function based on capacity utilizations. The resulting mixed integer linear programming model is used to compute lower bounds for small problem instances. We then design a Tabu Search algorithm to solve the problem, with customized components for the different decisions in our problem. We use five sets of instances: set 1 involves small randomly generated problems, sets 2 and 3 are based on real world data for liner shipping operations in the Mediterranean Sea, and sets 4 and 5 involve benchmark instances from the literature for other hub location application areas, but adapted for the liner shipping industry. Computational experiments demonstrate that the Tabu Search algorithm finds high quality solutions in a reasonable time for benchmark instances (the largest instance consisting of 81 nodes). Experiments also show that the network design can

be highly influenced by factors such as scale economies in mainline vs. feeder transportation costs, the port locations and hinterland flows, and congestion at the hub ports.

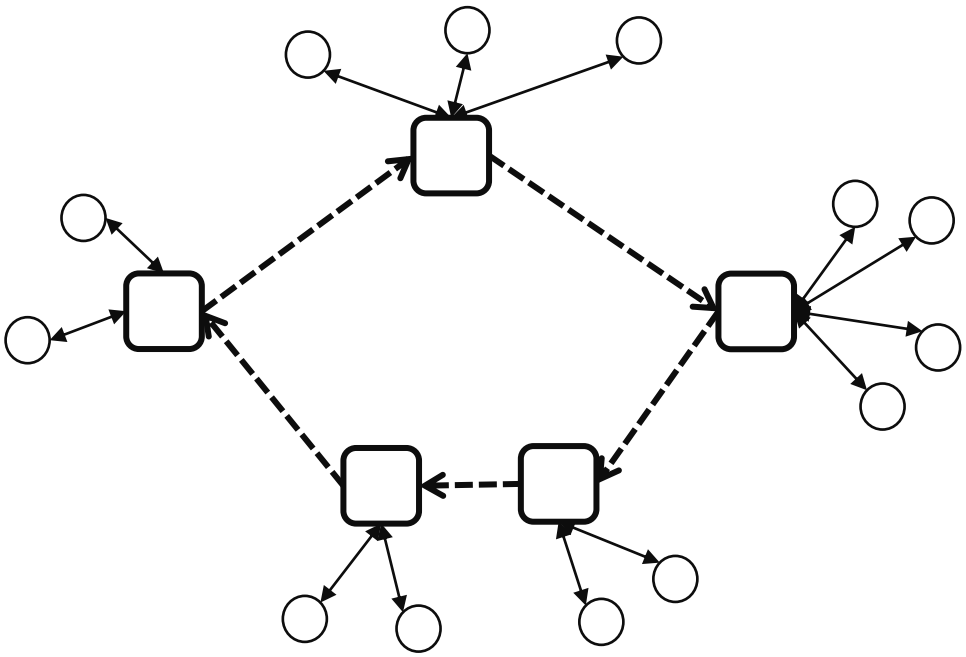

Figure 1: An example of a directed cyclic hub-and-spoke network (squares (circles) represent (non-)hub ports)

Our problem bears some similarities to other network design and location problems in the literature with respect to its hub-level network topology and routing aspect. For example, the ring star problem (Labbé et al., 2004) seeks to locate a simple cycle through a subset of vertices of a graph and assign the remaining vertices to them so as to minimize the sum of cycle setup and assignment costs. However, this problem does not consider flow routing costs. Other relevant problems include the median cycle problem (Labbé et al., 2005) and the capacitated *m*-ring star problem (Baldacci et al., 2007). Rapid transit, telecommunication and hierarchical logistics networks (Simonetti et al., 2011) are application areas of these problems. The many-to-many (hub) location-routing problem is another closely related problem type, in which the delivery and collection of cargoes between hubs and non-hub nodes are done via local multi-stop vehicle routes (Drexl and Schneider, 2015), which may include vehicle capacity or route length constraints. The hub-level network can be fully interconnected or subject to a specified network topology with an inter-hub route (e.g. Lopes et al., 2013). In addition, collection and delivery of cargoes can be done through combined (e.g. Rodriguez-Martin et al., 2014) or separate routes (e.g. Rieck et al., 2014). With respect to its flow routing aspect, our problem is similar to the minimum flow cost Hamiltonian cycle problem introduced by Ortiz-Astorquiza et al. (2015), which aims at finding, for a given set of nodes with pairwise flows, an undirected Hamiltonian cycle that minimizes the flow costs on the cycle.

The remainder of the paper is organized as follows: Section 2 describes the characteristics and assumptions of the capacitated directed cycle hub location and routing problem under congestion. The model formulations – as a non-linear mixed integer programming model and an approximate mixed integer linear programming model are presented in Section 3. The Tabu Search algorithm is explained in Section 4. Section 5 presents benchmark instances and discusses computational experiments. Managerial insights are presented in Section 6. Finally, concluding remarks and future research opportunities are provided in Section 7.

2. PROBLEM DEFINITION

The problem under investigation is defined by a set of container ports/terminals that have containerized cargo demands between each other, also called origin-destination (OD) pairs. Each port has a known cargo handling capacity. The network is to be designed as a HS structure such that all ports in the network have a path between each other. Some ports (or nodes) are to be designated

as hub ports, and the remaining non-hub ports are allocated to these hubs. Hubs ports perform the transshipment function to route the cargo flows in the network. Because shipping companies often route their cargoes in mainline ships on simple directed cyclic networks (Song and Dong (2013), Reinhardt and Pisinger, (2012)), a directed cycle is imposed on the hub-level network level only. In Figure 1, the squares represent hub ports, the circles denote non-hub ports, the dashed arrows represent inter-hub arcs, and the double arrows represent access arcs. In this network, containerized cargo from an origin port to a destination port is transported as follows: first, cargo from the origin port is collected by a feeder ship and transported to its allocated hub port on access arcs that connect non-hub and hub nodes. Here, the cargo is transshipped onto a mainline ship and sent on the inter-hub arcs in the cyclic hub-level network to the hub to which the destination port is allocated. Finally, the cargo is transshipped onto another feeder ship at the discharge hub port to be sent to the destination port. If origin and/or destination is a hub port, then collection and/or distribution by feeder ships is unnecessary. If both feeder ports are assigned to the same hub, then collection and distribution is done by feeder ships only. The mainline ship typically has a larger carriage capacity than the feeder ship, which leads to a lower unit transportation cost.

In the capacitated directed cycle hub location problem under congestion, the task is to design a liner shipping HS network which contains a directed cyclic hub-level network connecting all ports in the network so as to minimize the total network cost. Four decisions are to be made to solve the problem: the number and locations of hub ports, allocation of non-hub ports to hubs, establishing the inter-hub arcs (or hub-level route) to construct the hub-level network, and routing of the cargo flows in the HS network. The total network cost comprises five components: the cost of the container cargo collection and distribution between non-hub ports and hub ports by the feeder ships, the inter-hub cargo transfer by the mainline ships, cargo handling at transshipment, hub port congestion costs, and fixed hub opening costs. To represent the cost of a single handling of a container, terminal handling charges (THC) reported by the port/terminal operators are used. THC are the fees charged by the port for lifting the containers from the port quay onto the container ship's board and vice versa.

The assumptions underlying this problem are listed below:

1. Flows must be routed through the hubs.
2. Arcs in the network do not have a setup cost.
3. A subset of the ports in the network is considered as the candidate hub set.
4. The cargo handling capacity of the hubs is limited.
5. Non-hub ports can be allocated to a single hub only.
6. The structure of the hub-level network is a directed cycle.
7. THC at any hub port are the same for feeder or mainline ship handling. They are assumed to be the same for incoming and outgoing cargo at a port. Other port costs (e.g. pilotage fees) are omitted.
8. All containerized cargo transportation is made in standard twenty feet equivalent (TEU) containers. Unit cost and flow parameters are determined on this basis.
9. There are no physical restrictions apart from the port's handling capacity at potential hub ports to handle mainline ships.
10. A single planning horizon, for example one year, is considered in the problem.

Two common assumptions in the hub location literature, namely that the paths between any OD pair visit at most two hubs and that the hub-level network forms a complete graph, are not retained in our problem statement. Whereas a direct connection between any two pairs of hubs is possible in other transport networks such as airlines or postal delivery, ships in a liner service are set to follow a predefined route consisting of a sequence of ports. Therefore, the path of a cargo flow may include more than two hubs depending on the hub-level route. To illustrate the OD path of a cargo flow in a cyclic liner route, we refer to Figure 2. The containers originating from non-hub port 6 and destined to non-hub port 7 are first collected by a feeder ship from non-hub port 6 to hub port 1. At hub port 1, the containers are transshipped from the feeder to a mainline ship. The mainline ship transfers

the containers on the inter-hub arcs (1,2), (2,3), (3,4), and (4,5). The second transshipment takes place at hub port 5 from the mainline to another feeder ship. Finally, the containers are delivered to non-hub port 7 by the feeder ship. Thus, the cargo flow from 6 to 7 (w_{67}) is carried on arcs (6,1), (1,2), (2,3), (3,4), (4,5), and (5,7) where all arcs apart from the first and last belong to the hub-level route. Had there been a direct connection between 1 and 5, then w_{67} could be carried on the inter-hub arc (1,5). Therefore, the routing of the cargo flows and the cost of inter-hub cargo transfer are affected by how the hub ports sequenced on the hub-level cycle.

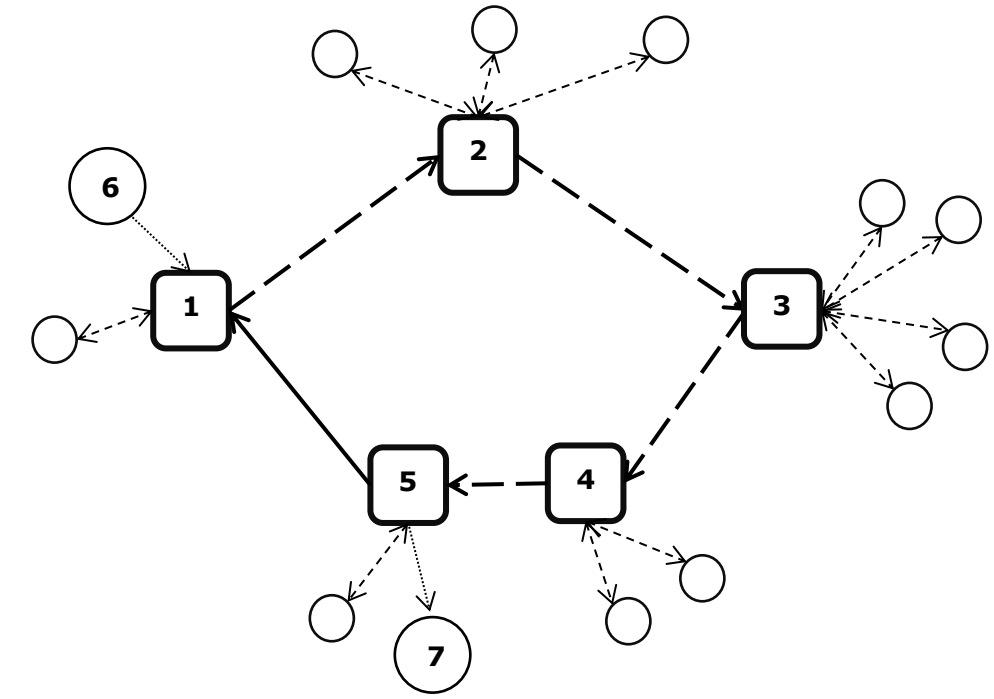

Figure 2: The OD path of the cargo flow from node 6 to 7

3. MODEL FORMULATIONS

We present two flow-based formulations to model our problem, using the following notation:

Sets	
N	Set of ports, $N = \{1,2, \dots, n\}$
H	Set of hub candidate ports, $H = \{1,2, \dots, h\}, H \subseteq N$
Parameters	
n	Number of ports
h	Number of hub candidate ports, $h \leq n$
w_{ij}	Flow from origin port *i* to destination port *j* in TEU, $i, j \in N$
o_i	Total flow emanating from port *i* in TEU, $o_i = \sum_{j \in N} w_{ij}$, $i, j \in N$
d_i	Total flow destined to port *i* in TEU, $d_i = \sum_{j \in N} w_{ji}$, $i, j \in N$
sd_{ik}	Sea distance between non-hub port *i* and hub port *k* in nautical miles (nm), $i \in N, k \in H$
cap_k	Cargo handling capacity of hub candidate port *k* in TEU, $k \in H$
tcf	Unit feeder ship transportation cost in USD/TEU-nm,
tcm	Unit mainline ship transportation cost in USD/TEU-nm,

pcf	Feeder ship port cost in USD for the given planning horizon (e.g. annual)
pcm	Mainline ship port cost in USD for the given planning horizon (e.g. annual)
thc_k	Unit terminal handling charges at hub candidate port *k* in USD/TEU, $k \in H$
fc_k	Hub opening cost at hub candidate port *k* in USD, $k \in H$
con_k	Congestion cost at hub *k*, USD, $k \in H$
ρ_k	Capacity utilization of hub *k*, $k \in H$, $\rho_k \in [0,1]$
α	Discount factor on inter-hub arcs, $\alpha = tcm/tcf$

Decision variables

Z_{ik}	1 if port *i* is allocated to hub port *k*, 0 otherwise, $i \in N, k \in H$
Z_{kk}	1 if a hub is located at *k*, 0 otherwise, $k \in H$
T_{ijk}	1 if ports *i* and *j* are allocated to hub port *k*, 0 otherwise, $i, j \in N, k \in H$
X_{kl}	1 if hub *l* is visited after hub *k* on the hub-level route, 0 otherwise, $k, l \in H$
F_k	Total flow through hub node *k* in TEU, $k \in H$.
Y_{kl}^i	Amount of cargo flow from *i* that is routed through hub-level arc (*k,l*), TEU, $i \in N, k, l \in H$

The model formulation is based on the 3-index hub location model of Ernst and Krishnamoorthy (1999). The following decision variables are defined. Binary variables Z_{ik} indicating whether or not port *i* is allocated to hub port *k*; binary variables Z_{kk} denoting whether or not a hub is located at port *k*; binary variables T_{ijk} indicating if two ports are allocated to the same hub (Mohammadi et al., 2011); binary variables X_{kl} indicating whether or not a hub arc exists between hubs *k* and *l*. In addition, F_k are continuous variables defined as the total flow at hub *k*; and Y^i_{kl} are continuous variables to measure the total flow originated from *i* and routed on inter-hub arc (*k,l*).

An important aspect in hub location problems is the scale economy achieved by consolidating OD flows through the hubs (Contreras, 2015). We model scale economies through a constant discount factor $\alpha \in [0,1]$, corresponding to the ratio of the mainline ship transportation unit cost and the feeder ship transportation unit cost:

$$\alpha = {}^{tcm}/_{tcf} \tag{1}$$

The discount factor α shows how much the transportation cost on inter-hub arcs is reduced compared to the transportation cost on arcs between non-hub nodes and hubs. The closer α is to zero, the more scale economies exist on inter-hub arcs.

Besides scale economies, it is also important to consider the trade-off between the benefits gained from flow consolidation and the cost of hub congestion. Not considering congestion may cause models to concentrate too much flow on a small number of hubs and hub arcs (De Camargo and Miranda, 2012). Rather than considering detailed port operations and queuing aspects, we look at ports from a macro perspective and we follow the approach of Elhedhli and Wu (2010) to model congestion cost. Elhedhli and Wu (2010) argue that congestion is closely related to the relative difference between hub capacity and the flows originating from the hub, and calculate the congestion cost of a hub *k* as follows:

$$con_k = t_0 \frac{F_k}{cap_k - F_k} = t_0 \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Z_{ik}}{cap_k - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Z_{ik}} \tag{2}$$

In Eq. (2), the term $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} Z_{ik}$ denotes the total flow emanating from hub *k*, cap_k is the hub capacity and t_0 is the unit cost of congestion. Elhedhli and Wu (2010) assume that only the flows that are sent from a hub create congestion. In many real-world ports, feeder and mainline ships share cargo handling and loading equipment (e.g. quay cranes) at the hub port and both loading and discharging operations are done at the same berth. Accordingly, all the cargo flows that are loaded and discharged at a hub

port are assumed to contribute to congestion. (In a M/M/1 queue perspective, the hub port is a single server and all in- and outgoing cargoes are assumed to join a single queue.) The total cargo flow to a hub *k* is the sum of four flow streams:

1. Incoming cargoes from non-hub ports that are allocated to hub *k* (*Feeder discharge*): This stream encompasses all cargo flows originating from the non-hub ports of *k*. They are carried by feeder ships and discharged at the hub to be loaded onto feeder or mainline ships depending on their destination.
2. Incoming cargoes from other hubs (*Mainline discharge*). These are the cargo flows that originate from all other ports in the network and destined to hub *k* and its non-hub ports. They are carried by mainline ships and discharged at the hub port to be distributed to their destinations by feeder ships.
3. Outgoing cargoes from hub *k* that are destined to the non-hub ports allocated to *k* (*Feeder load*). This stream refers to all cargo flows destined to the non-hub ports of *k*. They are loaded onto feeder ships and distributed to their destinations.
4. Outgoing cargoes from hub *k* that are destined to other ports in the network (*Mainline load*). The cargo flows originating from hub *k* and its non-hub ports and destined to other ports in the network are loaded onto the mainline ships and carried on the hub-level network.

As an illustrative example, we refer to the HS network presented in Figure 3 where there are two hubs with flows in both directions. There are six ports and two of them, 3 and 4, are designated as hubs. Suppose that cargo flow between each port is as shown in the flow matrix below:

Nodes	1	2	3	4	5	6	o_i
1	0	300	700	400	1,000	500	**2,900**
2	1,000	0	200	400	900	1,000	**3,500**
3	900	100	0	900	200	100	**2,200**
4	400	400	800	0	900	600	**3,100**
5	100	700	900	600	0	700	**3,000**
6	1,000	600	100	500	200	0	**2,400**
d_i	**3,400**	**2,100**	**2,700**	**2,800**	**3,200**	**2,900**	

The cargo flows carried by feeder ships on access arcs (dotted arrows) and by mainline ships on inter-hub arcs (dashed arrows) are shown on the network. From Figure 3, the total flow F_3 carried by feeder and mainline ships to be loaded or discharged at hub 3 is calculated as follows:

$$F_3 = 3{,}400 + 2{,}100 + 2{,}900 + 3{,}500 + 5{,}400 + 5{,}000 = 22{,}300$$

The breakdown of the four flow streams at hub 3 are shown below:

$$F_3 = Feeder\ load + Feeder\ discharge + Mainline\ load + Mainline\ discharge$$

$$Feeder\ load = \overbrace{w_{21} + w_{31} + w_{41} + w_{51} + w_{61} + w_{12}}^{\text{Flows destined to port 1}} + \overbrace{w_{32} + w_{42} + w_{52} + w_{62}}^{\text{Flows destined to port 2}}$$

$$Feeder\ discharge = \overbrace{w_{12} + w_{13} + w_{14} + w_{15} + w_{16} + w_{21}}^{\text{Flows sent from port 1}} + \overbrace{w_{23} + w_{24} + w_{25} + w_{26}}^{\text{Flows sent from port 2}}$$

$$Mainline\ load = \overbrace{w_{14} + w_{15} + w_{16} + w_{24}}^{\text{Mainline flows sent from port 1}} + \overbrace{w_{25} + w_{26} + w_{34}}^{\text{Mainline flows sent from port 2}} + \overbrace{w_{35} + w_{36}}^{\text{Mainline flows sent from hub port 3}}$$

$$Mainline\ discharge = \overbrace{w_{41} + w_{51} + w_{61} + w_{42}}^{\text{Mainline flows destined to port 1}} + \overbrace{w_{52} + w_{62} + w_{43}}^{\text{Mainline flows destined to port 2}} + \overbrace{w_{53} + w_{63}}^{\text{Mainline flows destined to hub 3}}$$

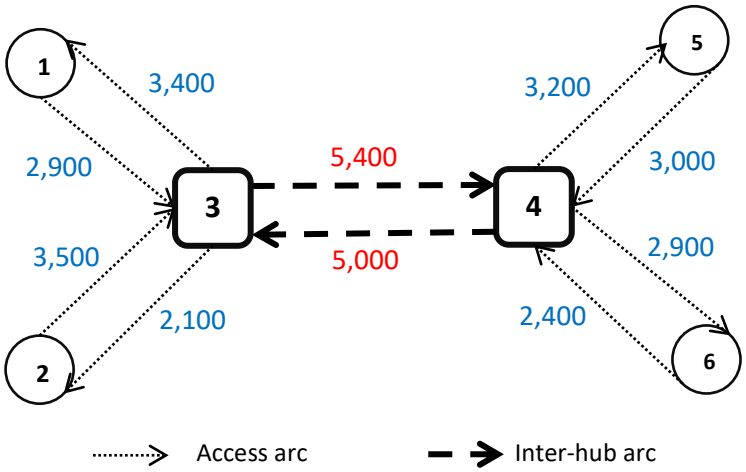

Figure 3: An example of a hub-and-spoke network with two hubs

The flows loaded onto and discharged from feeder ships include all flows originating from and destined to non-hub ports 1 and 2. On the other hand, the mainline load and discharge include flows between ports 1, 2, hub port 3 and hub port 4, ports 5, 6. The total flow at hub *k*, F_k, can be calculated as:

$$F_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} (o_i + d_i) Z_{ik} + \sum_{i=1}^{n} \left((o_i + d_i) Z_{ik} - \sum_{\substack{j=1 \\ j \neq i}}^{n} (w_{ij} + w_{ji}) T_{ijk} \right) k \in H \quad (3)$$

In Eq. (3), the first term captures the total feeder flows of loading and discharging at hub *k*. The second term is the mainline flow, calculated by subtracting the flows between the hub port and its non-hub ports from the sum of flows of all ports associated with *k*. Eq.(3) can be easily verified for the flow F_3 in Figure 3, where $Z_{13} = Z_{23} = 1$ in the first term and $Z_{13} = Z_{23} = Z_{33} = 1$ and $T_{123} = T_{133} = T_{213} = T_{313} = T_{233} = T_{323} = 1$ in the second term:

$$F_3 = (o_1 + d_1 + o_2 + d_2) + \big(o_1 + d_1 + o_2 + d_2 + o_3 + d_3 - (w_{12} + w_{21} + w_{13} + w_{31} + w_{21} + w_{12} + w_{23} + w_{32} + w_{31} + w_{13} + w_{32} + w_{23})\big)$$

$$F_3 = (2{,}900 + 3{,}400 + 3{,}500 + 2{,}100) + \big(2{,}900 + 3{,}400 + 3{,}500 + 2{,}100 + 2{,}200 + 2{,}700 - (300 + 1{,}000 + 700 + 900 + 1{,}000 + 300 + 200 + 100 + 900 + 700 + 100 + 200)\big)$$

$$F_3 = (2{,}900 + 3{,}400 + 3{,}500 + 2{,}100) + (2{,}900 + 3{,}400 + 3{,}500 + 2{,}100 + 2{,}200 + 2{,}700 - 6{,}400)$$

$$F_3 = 22{,}300$$

Because port costs of feeder and mainline ships differ from each other due to different ship characteristics, feeder and mainline flows are treated separately when calculating the congestion cost. Hence, we use the following formula to calculate the total congestion cost at hubs:

$$con_k = \frac{pcf \left(\sum_{\substack{i=1 \\ i \neq k}}^{n} (o_i + d_i) Z_{ik} \right) + pcm \left(\sum_{i=1}^{n} \left((o_i + d_i) Z_{ik} - \sum_{\substack{j=1 \\ j \neq i}}^{n} (w_{ij} + w_{ji}) T_{ijk} \right) \right)}{cap_k - F_k} \quad (4)$$

In Eq. (4), the total cost of the congestion created by feeder and mainline ships is calculated by multiplying the respective port costs with the total flows and dividing the sum by the difference between the hub capacity and the total flow through the hub. The cargo handling cost due to transshipment at hubs is calculated similarly to the congestion cost except that the cargo flows originating from and destined to the hub itself are omitted as they do not count as extra cargo handling. In the HS network shown in Figure 3, the total cargo handled at hub 3 is calculated as:

$Cargo\ loaded\ onto\ feeder\ vessel = w_{21} + w_{41} + w_{51} + w_{61} + w_{12} + w_{42} + w_{52} + w_{62}$

$Cargo\ loaded\ onto\ mainline\ vessel = w_{14} + w_{15} + w_{16} + w_{24} + w_{25} + w_{26}$

$Cargo\ discharged\ from\ feeder\ vessel = w_{12} + w_{14} + w_{15} + w_{16} + w_{21} + w_{24} + w_{25} + w_{26}$

$Cargo\ discharged\ from\ mainline\ vessel = w_{41} + w_{51} + w_{61} + w_{42} + w_{52} + w_{62}$

The cargo flows between non-hub ports that are allocated to the same hub as well as flows between ports that are allocated to different hubs are handled twice. For example, the cargo flow from non-hub port 1 to 2 (w_{12}) is handled once during discharge from feeder ship at hub 3, and then again for loading to the feeder ship sailing from hub 3 to non-hub port 2. Accordingly, the total cost of cargo handling due to transshipment in the network is calculated by multiplying the cargo handled at each hub with the corresponding THC: $\sum_{k=1}^{h} thc_k \sum_{\substack{i=1 \\ i\neq k}}^{n} \left[(2(o_i + d_i - w_{ik} - w_{ki})Z_{ik}) - \left(\sum_{\substack{j=1 \\ j\neq k,i}}^{n} (w_{ij} + w_{ji}) T_{ijk} \right) \right]$. With the details of congestion and cargo handling cost calculations explained, the model formulation for DCHC is as follows:

Mixed integer nonlinear programming model for DCHC

$$
\begin{aligned}
Minimize\ tcf & \left(\sum_{i=1}^{n} \sum_{k=1}^{h} sd_{ik}(o_i + d_i) Z_{ik} \right) + tcm \left(\sum_{i=1}^{n} \sum_{k=1}^{h} \sum_{l=1}^{h} sd_{kl} Y_{kl}^{i} \right) + \sum_{k=1}^{h} fc_k Z_{kk} \\
& + \sum_{k=1}^{h} thc_k \sum_{\substack{i=1 \\ i\neq k}}^{n} \left[(2(o_i + d_i - w_{ik} - w_{ki})Z_{ik}) - \sum_{\substack{j=1 \\ j\neq k,i}}^{n} (w_{ij} + w_{ji}) T_{ijk} \right] \\
& + \sum_{k=1}^{h} \frac{pcf \left(\sum_{\substack{i=1 \\ i\neq k}}^{n} (o_i + d_i)\, Z_{ik} \right) + pcm \left(\sum_{i=1}^{n} \left((o_i + d_i) Z_{ik} - \sum_{j\neq i}^{n} (w_{ij} + w_{ji})\, T_{ijk} \right) \right)}{cap_k - \left(\sum_{\substack{i=1 \\ i\neq k}}^{n} (o_i + d_i) Z_{ik} + \sum_{i=1}^{n} \left((o_i + d_i) Z_{ik} - \sum_{j\neq i}^{n} (w_{ij} + w_{ji})\, T_{ijk} \right) \right)}
\end{aligned}
\tag{5}
$$

subject to

$$\sum_{k=1}^{h} Z_{ik} = 1 \quad \forall\, i \in N \tag{6}$$

$$Z_{ik} \leq Z_{kk} \quad \forall\, i \in N, k \in H \tag{7}$$

$$\sum_{\substack{l=1 \\ l\neq k}}^{h} X_{kl} = Z_{kk} \quad \forall\, k \in H \tag{8}$$

$$\sum_{\substack{k=1 \\ k\neq l}}^{h} X_{kl} = Z_{ll} \quad \forall\, l \in H \tag{9}$$

$$X_{kk} = 0 \quad \forall\, k \in H \tag{10}$$

$$\sum_{\substack{i=1 \\ i\neq k}}^{n}(o_i + d_i)Z_{ik} + \sum_{i=1}^{n}\left((o_i + d_i)Z_{ik} - \sum_{\substack{j=1 \\ j\neq i}}^{n}(w_{ij} + w_{ji})T_{ijk}\right) \leq cap_k \quad \forall\, k \in H \tag{11}$$

$$\sum_{l=1}^{h}(Y_{kl}^{i} - Y_{lk}^{i}) = o_i Z_{ik} - \sum_{j=1}^{n} w_{ij} Z_{jk} \quad \forall\, i \in N, k \in H, k \neq l \tag{12}$$

$$Y_{kl}^{i} \leq o_i X_{kl} \quad \forall\, i \in N, k, l \in H \tag{13}$$

$$T_{ijk} \leq Z_{ik} \;\forall\, i, j \in N, k \in H \tag{14}$$

$$T_{ijk} \leq Z_{jk} \;\forall\, i, j \in N, k \in H \tag{15}$$

$$Z_{ik} + Z_{jk} \leq T_{ijk} + 1 \;\forall\, i, j \in N, k \in H \tag{16}$$

$$Y_{kl}^{i} \geq 0, \quad \forall\, i \in N, k, l \in H \tag{17}$$

$$X_{kl}, Z_{ik}, T_{ijk} \in \{0,1\}, \forall\, i, j\; k, l \in H \tag{18}$$

The objective function given in Eq. (5) minimizes the sum of the following costs: the feeder collection and distribution costs, inter-hub transfer costs, hub opening costs, cargo handling costs at transshipments, and hub port congestion costs. Constraints (6) are the single-allocation constraints. Ports are allocated to hubs only through constraints (7). Constraints (8) and (9) are the directed cyclic route constraints and ensure that both tails and heads of inter-hub arcs (X_{kl}) are hubs. Constraints (10) prevent the formation of arcs that emanate from and terminate in the same hub. Constraints (11) are the capacity constraints for each hub *k*. The capacity of a hub corresponds to the maximum number of containers that can be loaded or discharged at the hub by also considering double handling of containers due to transshipment according to Eq. (3). Hence, we consider both the incoming and outgoing flows in our capacity modelling, which is different from most of the capacitated HLPs studied in the literature where only the incoming flows were taken into account (e.g. Ernst and Krishnamoorthy, 1999, Boland et al., 2004, Correia et al., 2010). Constraints (12) are flow conservation constraints through hubs, which ensure that all hub-level flows are correct. Constraints (13) force that the flows are routed only on inter-hub arcs. Constraints (14) to (16) ensure that T_{ijk} is 1 if and only if both Z_{ik} and Z_{jk} are 1, that is, if both node *i* and *j* are allocated to hub *k*. Finally, constraints (17) and (18) are integrality and non-negativity constraints. In this model, if none of the OD flows in the network are equal to zero, then constraints (8), (9), (12), and (13) ensure that none of the inter-hub arcs form a sub-tour (Ortiz-Astorquiza et al., 2015; Contreras et al., 2016). The flow conservation constraints (12) are not violated only if the hub-level network is a single directed cycle without sub-tours. Therefore, with the assumption that the hub-level network is required to form a single directed cycle, no additional sub-tour breaking constraints are required in the model. DCHC is a difficult problem with a non-linear congestion cost component in the objective function. To ease the computational burden, we linearize the objective function and approximate the congestion cost. More specifically, we model the congestion at each hub as a semi-continuous piecewise linear function. Let ρ_k denote the capacity utilization of hub *k*, that is, the ratio of the total flow through the hub and its capacity:

$$\rho_k = {}^{F_k}\!/_{cap_k}, \rho_k \in [0,1] \tag{19}$$

The congestion cost in Eq. (4) can be expressed in terms of ρ_k as follows:

$$con_k = \left(\frac{1}{1-\rho_k}\right)\left[\frac{pcf\left(\sum_{\substack{i=1 \\ i\neq k}}^{n}(o_i + d_i)\, Z_{ik}\right) + pcm\left(\sum_{i=1}^{n}\left((o_i + d_i)Z_{ik} - \sum_{j\neq i}^{n}(w_{ij} + w_{ji})T_{ijk}\right)\right)}{cap_k}\right] \tag{20}$$

The first term in Eq. (20), $f(\rho_k) = 1/(1-\rho_k)$, is non-linear; the second term in Eq. (20) is linear. In Figure 4, we approximate $f(\rho_k)$ in interval $[a', b']$ with $0 \leq a' < b' < 1$ (a' and b' are given lower and upper limits of ρ_k) by a semi-continuous piecewise linear function shown in dashed line segments. Each segment operates between two (pre-defined) capacity utilization levels $[x_l, x_{l+1}]$. Each part of the approximation is tangent to the $f(\rho_k)$ curve at the first capacity utilization level x_l, i.e., $f'(x_l)(\rho_k - x_l) + f(x_l)$. Because $f(\rho_k)$ is convex, the approximation underestimates $f(\rho_k)$. In Appendix I, we propose a non-linear optimization model to find the best approximation for a given number of segments by optimizing the position of breakpoints x_l.

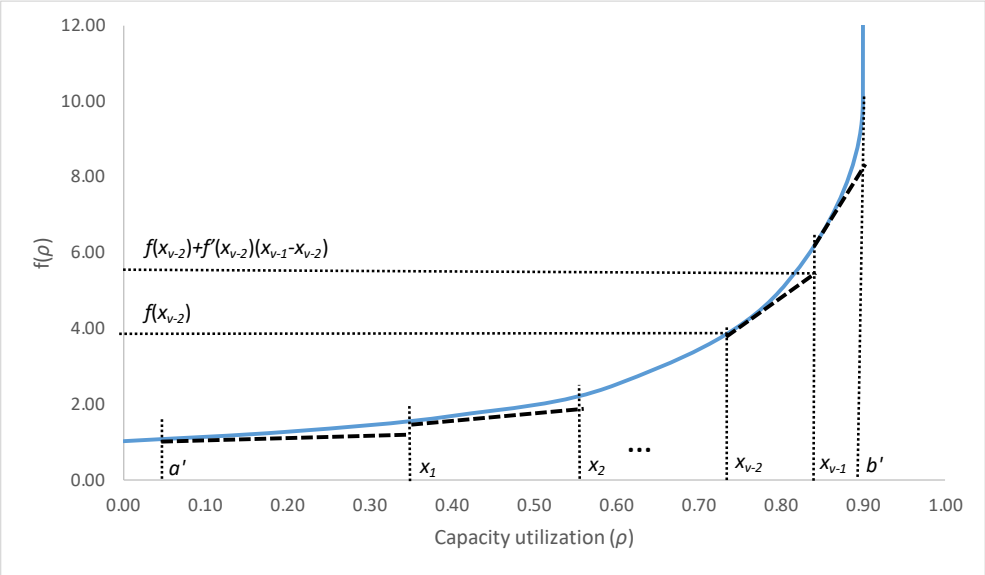

Figure 4: An illustration of a semi-continuous piecewise linear approximation

Let $V = \{1,2,3, \dots v\}$ denote the set of capacity utilization intervals for the approximation of $f(\rho_k)$ with $0 \leq a' \leq \rho_k \leq b' < 1$. Due to the discontinuities at the breakpoints, we consider in total $p = 2v$ points in $[a', b']$ and we define $c_m \in P = \{1,2,3, \dots p\}$ as the start (when index *m* is odd) or end capacity utilization level (when index *m* is even) of each interval. We also define u_m as the approximation value for $f(c_m)$, i.e., $u_m = f(c_m)$ when *m* is odd and $u_m = f'(c_{m-1})(c_m - c_{m-1}) + f(c_{m-1})$ when *m* is even. We introduce auxiliary variables to determine the capacity utilization interval that each ρ_k belongs to and its position:

G_{kq}	1 if ρ_k belongs to capacity utilization interval *q*, 0 otherwise, $k \in H, q \in V$
λ_{km}	determines the position of ρ_k in a utilization interval, $\lambda_{km} \in [0,1], k \in H, m \in P$

The capacity utilization ρ_k can be calculated as $\sum_{m=1}^{p} c_m \lambda_{km}$ and the approximation value for $f(\rho_k)$ as $\sum_{m=1}^{p} u_m \lambda_{km}$, with additional constraints to restrict λ_{km} to the start and end points of a single capacity utilization interval. The λ_{km} variables and u_m parameters are incorporated into the congestion cost function in Eq. (21).

$$\sum_{k=1}^{h} \frac{pcf\left(\sum_{\substack{i=1 \\ i\neq k}}^{n}\sum_{m=1}^{p}(o_i+d_i)u_m\lambda_{km}Z_{ik}\right)+pcm\left(\sum_{i=1}^{n}\left(\sum_{m=1}^{p}(o_i+d_i)u_m\lambda_{km}Z_{ik}-\sum_{\substack{j=1 \\ j\neq i}}^{n}\sum_{m=1}^{p}(w_{ij}+w_{ji})u_m\lambda_{km}T_{ijk}\right)\right)}{cap_k} \quad (21)$$

Eq. (21) is quadratic due to the product terms $\lambda_{km}Z_{ik}$ and $\lambda_{km}T_{ijk}$. We replace each product term $\lambda_{km}Z_{ik}$ and $\lambda_{km}T_{ijk}$ by continuous variables Z^{λ}_{ikm} and T^{λ}_{ijkm} on which additional constraints are imposed (Bisschop, 2020). The resulting mixed integer linear model with approximate congestion cost and additional constraints is presented below.

Mixed integer linear programming model for DCHC with congestion cost approximation

$$Minimize\ tcf\left(\sum_{i=1}^{n}\sum_{k=1}^{h} sd_{ik}(o_i+d_i)Z_{ik}\right)+tcm\left(\sum_{i=1}^{n}\sum_{k=1}^{h}\sum_{l=1}^{h} sd_{kl}Y^{i}_{kl}\right)+\sum_{k=1}^{h} fc_k Z_{kk}$$
$$+\sum_{k=1}^{h} thc_k\sum_{\substack{i=1\\ i\neq k}}^{n}\left[(2(o_i+d_i-w_{ik}-w_{ki})Z_{ik})-\sum_{\substack{j=1\\ j\neq k,i}}^{n}(w_{ij}+w_{ji})T_{ijk}\right] \quad (22)$$
$$+\sum_{k=1}^{h}\frac{pcf\left(\sum_{\substack{i=1\\ i\neq k}}^{n}\sum_{m=1}^{p}(o_i+d_i)u_m Z^{\lambda}_{ikm}\right)+pcm\left(\sum_{i=1}^{n}\left(\sum_{m=1}^{p}(o_i+d_i)u_m Z^{\lambda}_{ikm}-\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{p}(w_{ij}+w_{ji})\,u_m\,T^{\lambda}_{ijkm}\right)\right)}{cap_k}$$

subject to

(6)-(17)

$$\sum_{q=1}^{v} G_{kq}=1 \quad \forall\, k\in H \quad (23)$$

$$\sum_{m=1}^{p}\lambda_{km}=1 \quad \forall\, k\in H \quad (24)$$

$$\lambda_{k(2q-1)}+\lambda_{k(2q)}\leq G_{kq} \quad \forall\, k\in H, q\in V \quad (25)$$

$$\sum_{m=1}^{p} c_m\lambda_{km}=\frac{\sum_{\substack{i=1\\ i\neq k}}^{n}(o_i+d_i)Z_{ik}+\sum_{i=1}^{n}\left((o_i+d_i)Z_{ik}-\sum_{\substack{j=1\\ j\neq i}}^{n}(w_{ij}+w_{ji})T_{ijk}\right)}{cap_k} \quad \forall\, k\in H \quad (26)$$

$$Z^{\lambda}_{ikm}\leq Z_{ik} \quad \forall\, i\in N, k\in H, m\in P \quad (27)$$

$$Z^{\lambda}_{ikm}\leq \lambda_{km} \quad \forall\, i\in N, k\in H, m\in P \quad (28)$$

$$Z^{\lambda}_{ikm}+1\geq Z_{ik}+\lambda_{km} \quad \forall\, i\in N, k\in H, m\in P \quad (29)$$

$$T^{\lambda}_{ijkm}\leq T_{ijk} \quad \forall\, i,j\in N, k\in H, m\in P \quad (30)$$

$$T^{\lambda}_{ijkm}\leq \lambda_{km} \quad \forall\, i,j\in N, k\in H, m\in P \quad (31)$$

$$T^{\lambda}_{ijkm}+1\geq T_{ijk}+\lambda_{km} \quad \forall\, i,j\in N, k\in H, m\in P \quad (32)$$

$$T^{\lambda}_{ijkm}\leq Z^{\lambda}_{ikm} \quad \forall\, i,j\in N, k\in H, m\in P \quad (33)$$

$$T^{\lambda}_{ijkm}\leq Z^{\lambda}_{jkm} \quad \forall\, i,j\in N, k\in H, m\in P \quad (34)$$

$$Z_{ik}, X_{kl}, G_{kq}\in\{0,1\} \quad \forall\, i\in N, k,l\in H, q\in V \quad (35)$$

$$T_{ijk}, \lambda_{km}, Z^{\lambda}_{ikm}, T^{\lambda}_{ijkm}\in[0,1] \quad \forall\, i,j\in N, k\in H, m\in P \quad (36)$$

The objective function Eq. (22) is linear and calculates the sum of all cost components in which the hub port congestion cost is approximated by a semi-continuous piecewise linear function with v segments. The constraints (23)-(36) are interpreted in the context of the capacity utilizations of the hubs and the additional decision variables. Each hub candidate's capacity utilization belongs to a single capacity utilization interval through constraints (23). Constraints (24) ensure that the sum of λ_{km} variables for

each hub candidate is equal to one. Constraints (25) guarantee that only the λ_{km} variables that refer to the start and end points of the correct capacity utilization interval can take values larger than zero. Constraints (27) - (29) force the value of each Z_{ikm}^{λ} variable to λ_{km} if and only if Z_{ik} is 1 and to 0 otherwise. Similarly, constraints (30) – (34) ensure that T_{ijkm}^{λ} is equal to λ_{km} if and only if T_{ijk} is 1 and to 0 otherwise. Additionally, they ensure that the value of T_{ijkm}^{λ} does not exceed Z_{ikm}^{λ} and Z_{jkm}^{λ}. Finally, (35) and (36) are integrality and non-negativity constraints.

The congestion cost approximation leaves some ambiguity in the $f(\rho_k)$ approximation value when the hub capacity utilization is exactly at a breakpoint value. Each breakpoint value is both at the end of an interval and at the start of the next a capacity utilization interval, and both have different u_m values. The cost minimization model will favor the lowest of these values and hence set ρ_k at the end of a capacity utilization interval (rather than the start of the next interval) in case ρ_k is exactly at a breakpoint value.

The linearized mixed integer programming model can be solved to optimality with commercial software such as CPLEX for small problem instances only. The optimal solutions from the approximate model provide lower bounds to solutions to the original problem. In Section 5 we report on computational experiments with this formulation and we also compare the solution costs with those obtained by applying a Tabu Search procedure. Our Tabu Search procedure is described in the next section.

4. TABU SEARCH ALGORITHM

We developed a Tabu Search algorithm (TSHLP) which employs a hierarchical approach to the decisions in the hub-and-spoke network design problem: the number and locations of hubs, the allocations of non-hub ports to hubs, and hub-level routing. The algorithm starts with the generation of a feasible initial solution. Tabu Search is used to determine hub locations and non-hub port allocations only and involves three neighborhood moves: open a new hub, close an existing hub, and shift the allocation of a non-hub port from one hub to another. The procedure includes a probabilistic diversification routine that randomly changes the locations of hubs depending on the characteristics of previous solutions. A directed cyclic hub-level network is constructed, and local search is applied to improve the tour. Solutions are evaluated with the non-linear objective function (Eq. (5)).

4.1. Initial solution

The construction of an initial solution starts with locating hubs based on a hub candidate index similar to Chen (2007). In Chen's formula, hub candidate nodes for which the difference between their relative total flow and relative distance from other nodes is large have a greater index value. Motivated by the preliminary computational tests, we have modified Chen's hub candidate index as follows:

$$I_k = \left(\frac{(o_k + d_k)}{\sum_{k=1}^{h}(o_k + d_k)}\right)\left(\frac{\sum_{k=1}^{h}\sum_{i=1}^{n} sd_{ki}}{\sum_{i=1}^{n} sd_{ki}}\right) \quad (37)$$

In Eq. (37), the index is calculated by multiplying relative flows with relative distances (rather than taking the difference as in Chen (2007)), and hence the hub candidates that have relatively larger total flow and shorter distance to other ports are favored.

4.1.1. Initial hub locations

The hub location procedure ranks the candidate hub ports in non-increasing order of I_k. Starting with an empty set, each iteration adds the new hub with the highest index. The procedure stops when the total capacity of the selected hubs is larger than or equal to the total flow in the network multiplied by *volume_factor*. The *volume_factor* is a constant input parameter larger than 1. For example, if *volume_factor* is set to 1.8, then the selected hubs in the initial solution must have enough capacity to handle at least

180% of the total flow. As explained in Section 3, the flows carried by feeder and mainline ships to and from the hub ports are handled twice. A larger *volume_factor* ensures that the located hubs have enough capacity to handle the cargo flows and also prevents the congestion cost of the initial solution from being too high.

4.1.2. Non-hub port allocations

The remaining non-hub ports are allocated to the nearest hub using a proximity measure inspired by the work of Abyazi-Sani and Ghanbari (2016). While Abyazi-Sani and Ghanbari use only distances to sort the candidate nodes, we apply a modified distance measure, which accounts for the flow between a port and the candidate hub port. This modified distance favors ports which have more flow interaction with a hub candidate.

$$wd_{ik} = sd_{ik}\left(1 - \frac{w_{ik} + w_{ki}}{o_i + d_i}\right), \quad \forall\, i \in N, k \in H \tag{38}$$

4.1.3. Hub-level network design

The last part of the initial solution generation is the hub-level network design. The directed cyclic hub-level network is obtained by applying the Nearest-Neighbor heuristic to the selected hubs. The first hub located during the hub location procedure is the starting port of the cycle. Then, at each iteration, the hub that has the shortest distance from the last hub is selected to be the next hub in the hub-level route. The procedure stops when all selected hubs are sequenced.

4.2. Neighborhood search moves

TSHLP starts the neighborhood search based on three move operators. Two of these moves, *open_hub* and *close_hub* change the set of hub locations, and the third move, *shift_allocation* makes changes in non-hub port allocations. For all move operators, only the moves that do not violate the capacity constraints are considered. The move operators are explained below:

Open_hub: A new hub is located at each non-hub port iteratively. In the hub-level network, the new hub is inserted in the cycle after the hub to which it was originally allocated.

Close_hub: Each of the currently open hubs is closed iteratively. The ports in the closed hub cluster are allocated to the nearest open hub as per the modified distances in Eq. (38).

Shift_allocation: The allocation of a non-hub port is changed from one open hub to another. Executing this move may take considerably more computational time especially for large size problems. Besides, shifting the allocation of a port to hubs that are very far from the port hardly ever yields good solutions. Therefore, we set a limit to the neighborhood for each port. This limit, denoted by the integer parameter *range*, is a number less than or equal to the number of candidate hubs. For instance, if *range* is set to five, instead of checking all possible allocation shifts of a non-hub port, only the five nearest (according to the modified distance) hub candidate ports are considered. Out of these five candidates, the *shift_allocation* move is applied to the open hubs only.

Each move has its own Tabu list and Tabu tenure. The Tabu tenures of the moves are static and deterministic. If a solution improves the incumbent solution, it is accepted. This is defined as the aspiration criterion of the algorithm.

4.3. Diversification

The frequency of the hub locations throughout the Tabu Search is recorded in the long-term memory. The frequency of each hub candidate node refers to the number of times a hub was located at this node in the previously visited solutions. Diversification is triggered when the maximum of the hub location frequencies, denoted by *max_freq*, reaches a predefined limit, denoted by *freq_lim*. For example, if *freq_lim* is set to 25, diversification is triggered whenever the frequency reaches 25, 50, 75, and so on.

At each diversification moment, Tabu Search starts a new search phase. The logic in the diversification is to close a number of hubs among the nodes with higher frequencies (which we refer to as the high frequency region) and to open a number of new hubs at nodes with lower frequencies (the low frequency region). The selection of hubs to be closed and opened is done randomly within the respective region, using the following parameters:

high_region: The percentage of *max_freq* that is used to determine the high frequency region. For instance, if this parameter is set to 75% and *max_freq* is 100, hubs to be closed are selected among the nodes which have frequency higher than or equal to 75.

low_region: The percentage of *max_freq* that is used to determine the low frequency region. For instance, if this parameter is set to 25%, and *max_freq* is 100, new hubs to be opened are selected among the nodes which have frequency less than or equal to 25.

close_percentage: indicates what portion of the hubs in the high frequency region specified by *high_region* is to be closed. For instance, if *close_percentage* is set to 50%, half of the current hubs in the high frequency region are selected randomly and closed.

open_percentage: indicates how many hubs are to be opened at the nodes in the low frequency region specified by *low_region*. If there are 20 candidate nodes in low frequency region and *open_percentage* is 25%, then five of them are chosen randomly to serve as hubs.

High_region and low_region are constant input parameters and do not change in the algorithm. On the other hand, *open_percentage and close_percentage* are dynamic and can be modified at the next trigger of the diversification phase, depending on the best solution obtained in the search phase that follows the previous diversification. If the best solution in the current search phase of the algorithm (S_{best}) is worse than the global best solution (S_{global}), the algorithm makes one of the following changes depending on the number of hubs located in the two solutions:

1. If the number of hubs located in S_{best} is greater than or equal to that in S_{global}: Decrease *open_percentage* and increase *close_percentage* by a fixed amount of percentage points, for example 5.
2. If the number of hubs located in S_{best} is less than that in S_{global}: Increase *open_percentage* and decrease *close_percentage* by a fixed amount of percentage points, for example 5.

The non-hub ports in the closed hub clusters are allocated to the nearest open hubs without violating the capacity constraints. If some of the ports remain unallocated due to insufficient capacities at the open hubs, new hubs are opened at each of these unallocated ports. The solution obtained by diversification is taken as the initial solution and the S_{best} is reset to the initial solution cost. TSHLP terminates after a predefined number of consecutive iterations without improvement, and S_{global} is reported. The pseudocode of the algorithm is provided in Appendix II.

4.4. Local search for hub-level cyclic network

We attempt to improve the hub-level cycle in each iteration of the Tabu Search algorithm if the inter-hub transfer cost of the current solution is larger than that of the previous solution. The local improvement algorithm (LISUB) takes the hub-level network of the current solution as the initial solution, and the inter-hub transfer cost formula as the evaluation function. A single move operator, 2-opt exchange, is defined in the algorithm. The 2-opt exchange move is a well-known heuristic for the travelling salesman problem and involves removing two arcs from a tour and reconnecting the two partial routes by two other arcs (Talbi, 2009). In LISUB, the inter-hub transfer cost is calculated for all single 2-opt exchange moves based on the current solution, and the cycle that yields the lowest cost is recorded as the improved hub-level network. The current solution of TSHLP is updated with the new hub-level network.

5. COMPUTATIONAL EXPERIMENTS

TSHLP was programmed in C/C++, and all experiments ran on a desktop computer with 8 GB memory and 3.40 GHz processor. We first introduce the data sets that are used in the experiments and then present and discuss the results. We benchmark the best results obtained by TSHLP for the original non-linear problem against the results obtained by CPLEX for the linear model with approximate congestion costs. In addition, we analyze the performance of the algorithm and evaluate the effect of the diversification. We recall that the diversification phase of Tabu Search contains random elements, and that the algorithm is nondeterministic and may produce different results at each execution. To obtain meaningful results, the algorithm is run 100 times for each problem instance, and both the best and average solutions are reported.

5.1. Problem instances

Five data sets are used in computational experiments to assess the effectiveness of the proposed algorithm and to analyze the factors that affect the network design. The data sets are labelled RAND10, MED15, MED20, CAB25, and TR81. Five instances with different values for the discount factor, α, are generated in each set. Table 3 summarizes the sources and parameter values for the different sets.

The RAND10 set contains the smallest problem instances. It is a 10-node set with randomly generated parameter values. In order to keep the variation in the parameter values limited and to avoid any port being too advantageous for flow concentration or cost benefits, we use discrete random uniform distributions for flow, distance, capacity, and THC parameters. A single parameter value is used for each hub opening cost, unit feeder transportation cost, and unit feeder/mainline ship port costs.

The MED15 and MED20 sets involve 15 and 20 major ports in the Mediterranean Sea, respectively (Table 2). The parameter values of these data sets are motivated by real world data. The distances, OD flows, and terminal handling charges were obtained from external sources. The remaining parameters were either approximated based on real data or they were generated when exact or estimated data could not be found.

Table 2: Ports included in MED15 and MED20 experiment sets.
Ports with asterisk are only used in MED20

Country	Ports
France	Marseille
Greece	Piraeus, Thessaloniki
Italy	Cagliari, Genova, Gioia Tauro, La Spezia, Livorno*, Trieste, Venezia
Slovenia	Koper*
Spain	Algeciras, Barcelona*, Las Palmas*, Valencia
Turkey	Ambarli, Gemlik*, Izmir, Izmit, Mersin

The CAB25 and TR81 sets have the larger problem instances, and they are based on the CAB and Turkish postal network data sets respectively. CAB is a data set of airline passenger flows and distances between 25 major cities in the United States. It was introduced in the hub location literature by O'Kelly (1987) and is accessible from the OR-Library (2018). The Turkish data set involves postal delivery data of 81 cities in Turkey (Bilkent.edu.tr, 2018). Both data sets are compatible with conventional hub location problems only (e.g. *p*-hub median problem, hub covering problem), and require additional parameters for our problem setting such as terminal handling charges and hub opening costs. For the CAB25 instances we generated values for the missing parameters, and for the TR81 instances we modified some of the existing parameters and generated the missing parameters values. The purpose of these modifications was to make the distance, cost, and flow parameters realistic from the perspective of liner shipping. For example, the OD flows in the Turkish postal network data set refer to the annual postal deliveries between the cities in Turkey, and these values are extremely large when compared to global containerized cargo flows.

Table 3: Input parameters of the experiment sets

Parameter	RAND10	MED15	MED20	CAB25	TR81
Number of nodes (N)	10	15	20	25	81
Distances, nautical mile (sd_{ik})	10 to 100	Real sea-distances compiled from https://sea-distances.org/		Compiled from the original dataset	The distances of the original data set are multiplied by 10.
OD flows, TEU (w_{ij})	10 to 100	Annual containerized cargo movements between ports according to customs declarations in 2016. Compiled and approximated from EUROSTAT database (http://ec.europa.eu/eurostat/data/database)		Compiled from the original data set	The flows of the original data set are divided by 5.
Hub capacities, TEU per day (cap_k)	2,500 to 3,500	Port's hinterland flow ($o_k + d_k$) is multiplied by 4.		Random assuming the hub candidate port's capacity is normally distributed with mean 3 (TR81) or 4 (CAB25) and standard deviation 0.5 times its hinterland flow ($o_k + d_k$)	
Terminal handling charges, in USD per TEU (thc_k)	5 to 25	Extracted from the websites of terminal operators and shipping companies		Discrete random values with a range of 150 to 250	
Hub opening cost, USD (fc_k)	250,000	Discrete random values ranging from 30 to 70 million	60 million for each hub candidate	75, 100, or 125 million depending on the hub candidate port's capacity	
Unit feeder transportation cost, USD/TEU-nm (tcf)	10	0.0839 (Calculated based on the running cost estimation of a 600 TEU feeder ship by Baird (2006))			
Unit mainline transportation cost, USD/TEU-nm (tcm)	2, 4, 6, 8, or 10, depending on the problem instance	Calculated based on the running and capital costs of an 18,000 TEU ship. 0.0168, 0.0336, 0.0503, 0.0671, or 0.0839 depending on the problem instance			
Unit feeder ship port cost, USD (pcf)	365	9,233 (Calculated based on the running cost estimation of a 600 TEU feeder ship by Baird (2006))			
Unit mainline ship port cost in USD (pcm)	1,825	83,890 (Calculated based on the running and capital costs of an 18000 TEU ship)			

For each problem set, five instances are generated with α taking values of 0.20, 0.40, 0.60, 0.80, and 1.00. The names of the instances are formed by combining the experiment set name with the corresponding α value. For example, MED20-0.20 denotes the instance of the MED20 set with $\alpha = 0.20$. Recall from Eq. (1) in Section 3 that α refers to the ratio of unit mainline ship transportation cost, tcm, and the unit feeder ship transportation cost, tcf. Assuming that tcf remains fixed, α changes with tcm. The more containers a mainline ship carries, the less it costs to transport a single container. Therefore, the economies of scale gained on the transportation between hubs deteriorates when α increases or when the mainline ship capacity utilization

decreases. Through experiments with different problem instances we can investigate the effect of the mainline ship capacity utilization on the network design and cost.

5.2. Experiments and results

Table 4 shows the values of the TSHLP parameters that we used in the experiments. These parameter values were tuned in preliminary experiments.

Table 4: TSHLP input parameter values

Input parameter	Value
$volume\_factor$	ranges between 2-3.5
max_iter	100
Tabu tenure (number of iterations)	5 (RAND10, MED15/20, and CAB25), 10 (TR81)
$freq\_lim$	25-50
$high\_region$	50
low_region	50
$close\_percentage$*	25
$open\_percentage$*	25
$range$	5-25
$incr$	5

**initial values only, the parameters are modified by the algorithm*

For each problem instance, the initial solution cost, the average solution cost from 100 runs, and the overall best solution cost (i.e., the minimum of the 100 runs) obtained with TSHLP are reported in Table 5. For RAND10, MED15, and MED20 instances, the optimal solution costs obtained with CPLEX and the MILP with congestion cost approximation and the percentage gap between the best TSHLP solution cost and CPLEX solution cost are also presented. Larger instances could not be solved by CPLEX. The computation times in seconds with TSHLP (average of 100 runs and for the overall best solution) and with CPLEX are also reported. The results presented in Table 5 demonstrate that TSHLP finds good solutions in reasonable computation time. The percentage gap between the optimal solution costs for the approximation model and the best solution costs found by TSHLP is below 0.20% for all benchmark instances. Furthermore, the asterisk for 15 instances indicates that both TSHLP and CPLEX found the same solution, which means the same network design: the number and locations of hub ports, the allocation of the non-hub ports to hubs, and the hub port sequence on the cycle. For all benchmark instances, the same network designs were obtained by CPLEX and TSHLP. We note that although the networks obtained with TSHLP and CPLEX may be the same, the network costs are always different because an approximated congestion cost formula is used for CPLEX. The % gap is calculated based on the difference in the objective function values only (and ignoring the solution network structure).

Although the instances of the CAB25 and TR81 sets could not be included in the comparison, the results obtained by TSHLP show that the algorithm is efficient concerning the computational effort. While it took CPLEX more than 17,000 seconds to find the optimal solution for some smaller instances, TSHLP found solutions within 1 second for the instances with up to 25 nodes, and in less than 4 seconds for the instances of TR81. The maximum computation time was 6 seconds for the TR81-0.60 instance. We analyze the effect of diversification on the solution quality. It is important to find out whether implementing diversification (at the expense of more computational effort) yields indeed better solutions. For this purpose, TSHLP was run without diversification, and the solution costs were compared to measure the improvement due to diversification. The average percent improvements in instances of each set are shown in Figure 5. Diversification clearly improved the best solutions found in all sets. The largest average improvement is observed in CAB25 instances with 10.67%, and the smallest improvement is recorded for MED15 instances with 1.21%. Clearly, TSHLP yields better solutions when diversification is applied.

Table 5: Results and statistics

	Solution costs (in million USD)					Computation time (in seconds)		
Experiment instance	TSHLP (Initial solution)	TSHLP (Average)	TSHLP (Best)	CPLEX	% GAP	TSHLP (Average)	TSHLP (Best)	CPLEX
RAND10-0.20*	3.64	3.29	3.18	3.17	0.17%	0.0133	0.0156	255.03
RAND10-0.40*	4.64	3.77	3.66	3.65	0.15%	0.0211	0.0157	89.39
RAND10-0.60*	5.65	4.21	4.04	4.03	0.13%	0.0136	0.0157	61.22
RAND10-0.80*	5.53	4.61	4.41	4.41	0.12%	0.0119	0.0157	65.75
RAND10-1.00*	7.66	4.99	4.79	4.78	0.11%	0.0122	0.0157	56.16
MED15-0.20*	950.67	774.47	769.92	769.64	0.04%	0.0614	0.0781	17,531.28
MED15-0.40*	968.65	825.41	810.47	810.19	0.03%	0.0533	0.0313	5,539.53
MED15-0.60*	1,032.02	849.55	849.48	849.27	0.02%	0.0336	0.0469	6,989.73
MED15-0.80*	904.24	892.18	886.61	886.37	0.03%	0.0364	0.0625	4,073.33
MED15-1.00*	992.97	908.71	905.17	904.84	0.04%	0.0194	0.0313	2,911.22
MED20-0.20*	1,474.06	1,060.42	1,047.88	1,047.69	0.02%	0.1042	0.1094	11,143.24
MED20-0.40*	1,462.11	1,113.83	1,076.04	1,075.86	0.02%	0.1063	0.0781	6,688.56
MED20-0.60*	1,449.36	1,143.32	1,099.74	1,099.57	0.02%	0.0698	0.0781	6,243.53
MED20-0.80*	1,151.18	1,123.59	1,123.59	1,123.41	0.02%	0.0369	0.0469	5,638.23
MED20-1.00*	1,280.82	1,147.44	1,147.44	1,147.26	0.02%	0.0355	0.0313	4,044.58
CAB25-0.20	3,531.85	3,224.94	3,173.21	-	-	0.7078	0.8594	-
CAB25-0.40	4,187.53	3,653.20	3,400.84	-	-	0.4147	0.2969	-
CAB25-0.60	5,184.46	3,927.99	3,538.87	-	-	0.2388	0.2500	-
CAB25-0.80	5,350.70	3,994.18	3,652.47	-	-	0.1381	0.2500	-
CAB25-1.00	5,440.65	4,089.69	3,744.65	-	-	0.1198	0.1094	-
TR81-0.20	10,179.30	8,351.30	7,901.48	-	-	3.7527	5.4196	-
TR81-0.40	13,852.60	10,437.50	10,073.00	-	-	3.3739	4.8282	-
TR81-0.60	17,157.80	12,439.80	11,244.30	-	-	2.8203	6.0001	-
TR81-0.80	15,186.60	13,460.80	12,414.40	-	-	2.2724	3.1095	-
TR81-1.00	15,255.90	13,606.70	13,213.80	-	-	1.3942	3.2031	-

**Same solutions (i.e. network design) were obtained with CPLEX and TSHLP*

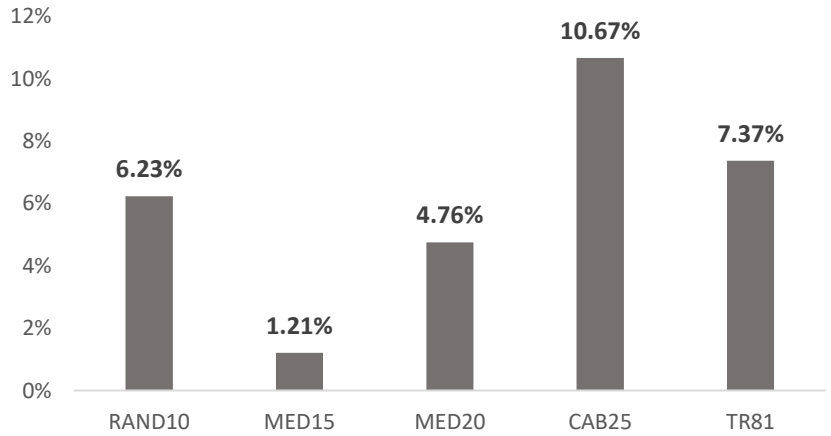

Figure 5: Average percent improvement by diversification

6. MANAGERIAL INSIGHTS

One of the objectives of our hub location model is to provide insights into the factors that affect liner shipping HS networks. In this section, we analyze the effects of the discount factor α, the ports' location and hinterland flows, and the hub port congestion on the HS network design and costs.

Firstly, the number of hubs decreases when α increases (Fig. 6). This is in line with expectations because when the unit transport cost on inter-hub arcs is closer to that of feeder transportation, locating more hubs (at the expense of a longer cyclic hub-level network and higher hub opening cost) does not yield any cost advantage. Referring to the relationship between the mainline ship capacity utilization and the economies of scale, we can state that a decrease in mainline ship capacity utilization (or a higher α) results in locating fewer hubs.

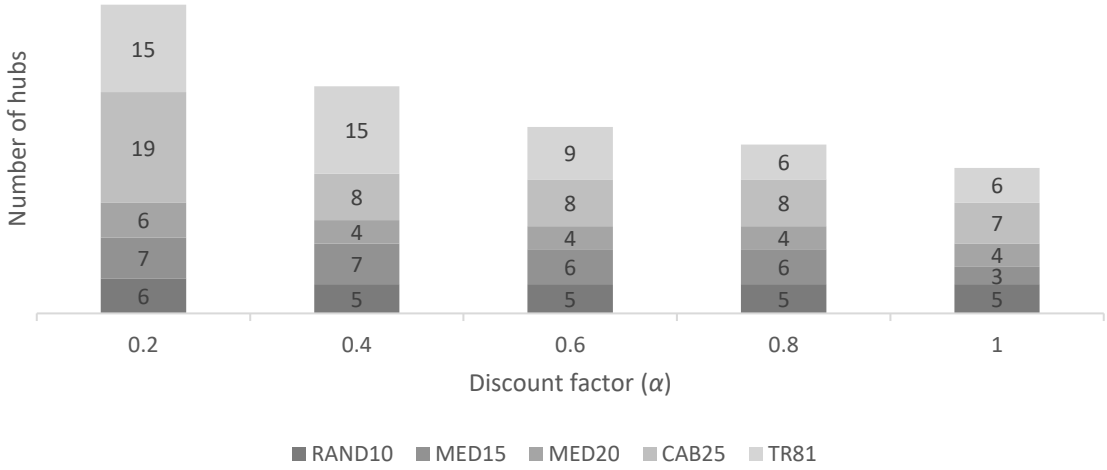

Figure 6: Number of hubs located in problem instances for different α.

The decreasing trend in the number of hubs with deteriorating scale economies is visualized for the networks of the MED15-0.20 and MED15-1.00 instances in Figures 7 and 8. In both figures the squares represent hub ports; circles denote non-hub ports; arrows represent the hub-level arcs, and the dotted lines show the node-hub allocations. When the scale economies due to flow consolidation disappear, fewer hubs are located.

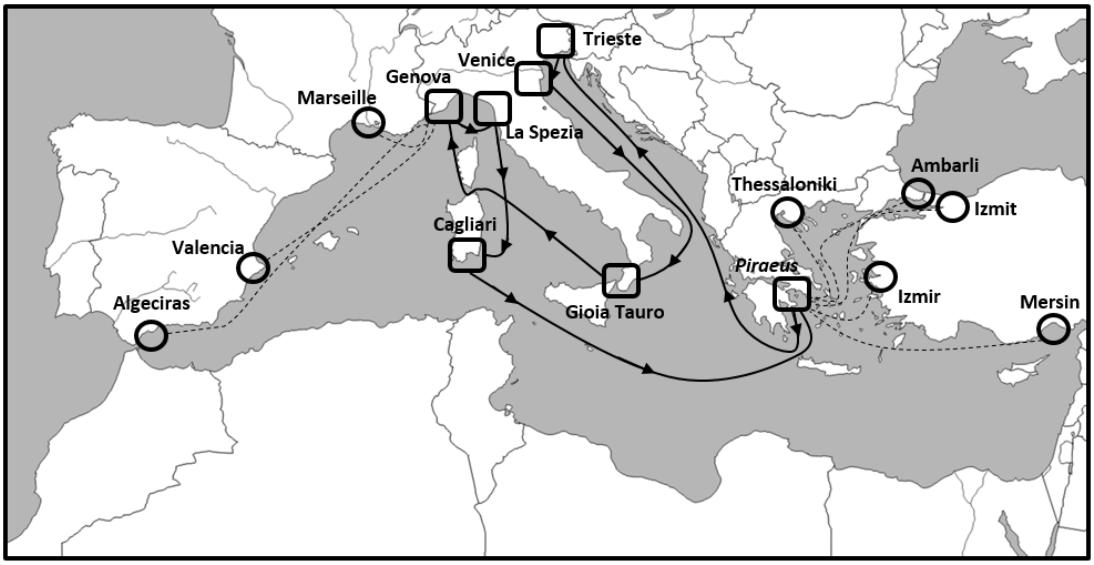

Figure 7: The HS network of the solution for the MED15-0.20 instance
(map template source: d-maps.com, https://d-maps.com/carte.php?num_car=3128&lang=en)

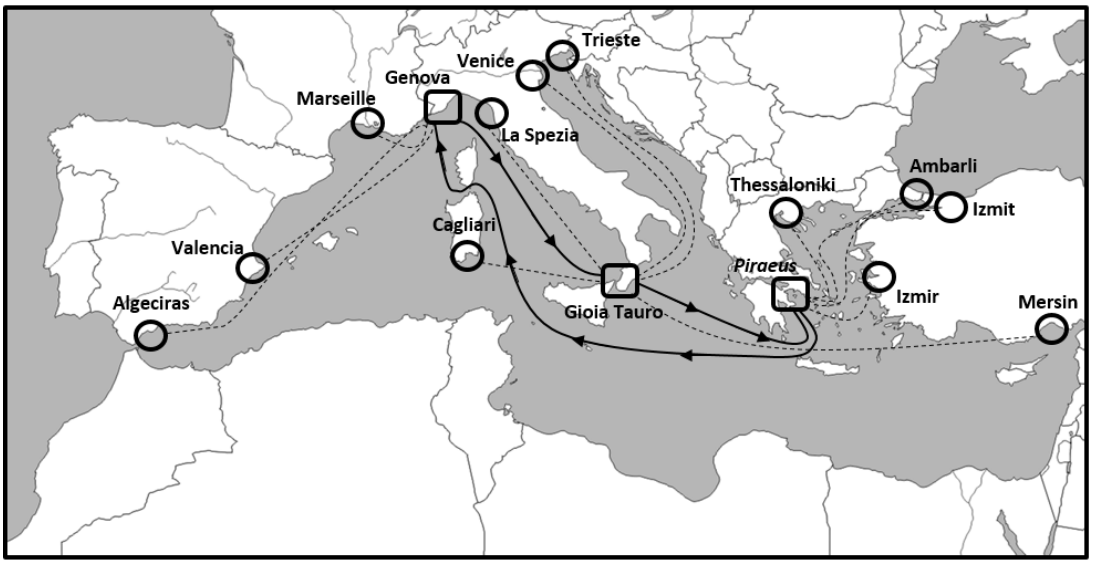

Figure 8: The HS network of the solution for the MED15-1.00 instance
(map template source: d-maps.com, https://d-maps.com/carte.php?num_car=3128&lang=en)

Two other factors worth analyzing are related to the locational attributes of the ports. Centrality and intermediacy are two common attributes used to describe locations of ports and cities (Fleming and Hayuth, 1994). In our analysis, we measure the centrality of a port as the hinterland flow originating from and destined to that port. Thus, hinterland flow of port *k* is equal to the sum of o_k and d_k.

Intermediate locations, on the other hand, are between important origin and destination places, and therefore they are suitable for transshipment operations (Fleming and Hayuth, 1994; Rodrigue et al., 2013). Our interpretation of intermediacy refers to the closeness of a port to other ports, and it is measured by the sum of the distances from that port to all other ports. Hence, the lower the total distance of port *i* has, the higher its intermediacy.

The results show that while there is a tendency of locating hubs at ports that have higher hinterland flows and/or are closer to other ports, this should not be considered as a rule. Peripheral ports also have a chance of becoming hubs depending on other factors. For example, the majority of the hub ports in MED15 and MED20 have a high degree of centrality and/or intermediacy (Tables 6 and 7). Piraeus has a total hinterland flow of almost 1 million TEU which constitutes about 17% of the total OD flows among the 20 ports of MED20. A hub was located at this port in all instances of MED15 and MED20. The other ports such as Gioia Tauro, Genova, and La Spezia where a hub was located in majority of instances also control a considerable portion of hinterland flows (Fig. 9).

The intermediary position of the hubs is also noticeable in Figures 7 and 8. The hubs that are common in both instances lie between peripheral ports and are not too far from the other ports. Gioia Tauro, for example, has the lowest total distance to the other ports, and it was a hub in all instances of MED15 and MED20. On the other hand, the geographical advantage of a port or its hinterland flow intensity are not unique determinants of hub locations. For instance, the port of Venice does not control a very high hinterland flow compared to other ports in the region, nor is it located near to other ports. Nevertheless, a hub was opened at this port in MED15-0.20 and MED20-0.20 instances. Similarly, in the instances of CAB25 and TR81, some of the hubs were located at remote ports which do not have a high degree of centrality and intermediacy.

Finally, we investigate the impact of omitting congestion in the process of network design and optimization. The problem instances of MED15, MED20, and TR81 were selected for this analysis, and the resulting networks and network costs of the original problem instances were compared to the instances which were optimized without taking congestion cost into account. First, the solutions for each problem instance were obtained by TSHLP by excluding the congestion cost component from the evaluation function of the algorithm. Then, the cost of congestion was calculated for the resulting network and added to the solution cost to obtain the

total network cost, denoted by Θ^*. The total network cost of the corresponding original problem instance is denoted by Θ. Based on these costs, we calculated the percent increase in the total network cost due to omission of the congestion cost, denoted by Φ, as follows: $\Phi = \left(\frac{\Theta^* - \Theta}{\Theta}\right) 100\%$.

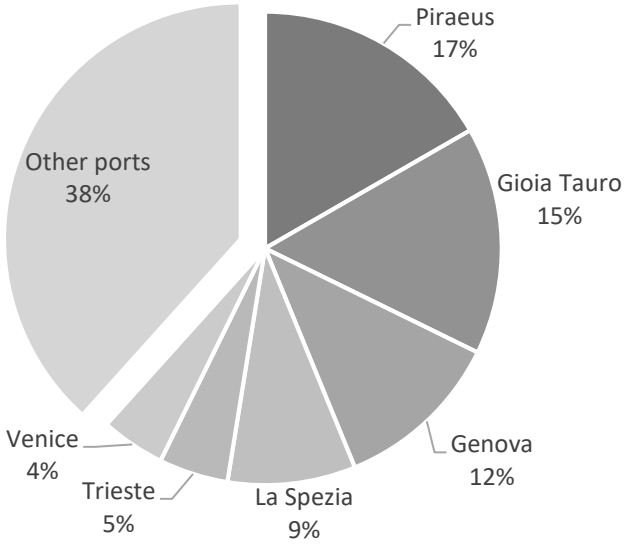

Figure 9: Percent share of hinterland flows by the hub ports located in MED15 and MED20 instances

Table 6: Centrality and intermediacy of ports in MED20 set (ranked in decreasing order of hinterland flows)

Port	Hinterland flow (TEU)	Total distance (nm)
Piraeus* (5)	948,618	15,906
Gioia Tauro* (5)	884,304	13,388
Genova* (5)	657,032	15,701
La Spezia* (5)	496,930	15,479
Valencia	269,467	18,448
Trieste* (1)	269,322	20,166
Thessaloniki	255,344	18,630
Venice* (1)	248,735	20,175
Ambarli	226,677	19,318
Mersin	219,856	24,044
Cagliari	178,886	14,157
Marseille	165,654	16,355
Algeciras	152,122	22,740
Izmir	126,729	17,533
Izmit	126,027	19,733
Livorno	122,898	15,330
Barcelona	120,989	17,308
Gemlik	105,036	19,445
Las Palmas	60,834	34,508
Koper	48,732	20,148

***** *A hub was located at this port in (x) out of 5 instances*

Table 7: Centrality and intermediacy of ports in MED15 set (ranked in decreasing order of hinterland flows)

Port	Hinterland flow (TEU)	Total distance (nm)
Piraeus* (5)	862,208	10,476
Gioia Tauro* (5)	810,654	9,050
Genova* (5)	609,846	11,246
La Spezia* (4)	465,100	11,040
Thessaloniki	244,666	12,317
Trieste* (4)	241,703	14,091
Venice* (4)	234,253	14,037
Valencia	233,642	13,624
Mersin	212,576	16,159
Ambarli	205,618	12,927
Cagliari* (2)	166,233	9,999
Marseille	137,385	11,786
Algeciras	134,467	17,079
Izmir	123,052	11,517
Izmit	118,727	13,224

*****A hub was located at this port in (x) out of 5 instances*

The percent increase in the network cost gives an idea about the importance of taking congestion into account for the liner shipping HS network design. The results of the analysis are shown in Table 8. In addition to Φ, the number of hubs located, the

percent contribution of the congestion cost to the total cost, the average capacity utilization rate of the hubs, and the number of hubs with utilization rate exceeding 95% were compared.

Table 8: Comparison of the results with and without considering congestion in optimization

MED15	**0.20**		**0.40**		**0.60**		**0.80**		**1.00**	
Φ (%)	1.45		0.00		0.15		0.01		4.85	
Congestion included?	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
% contribution of congestion to the total network cost	10	12	10	10	9	9	8	8	7	13
Average hub capacity utilization (%)	31	28	31	31	33	31	33	33	42	65
Number of hubs with utilization above 95%	0	0	0	0	0	0	0	0	0	0
Number of hubs	7	9	7	7	6	7	6	6	3	2
MED20	**0.20**		**0.40**		**0.60**		**0.80**		**1.00**	
Φ (%)	0.20		2.31		0.00		65.12		62.21	
Congestion included?	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
% contribution of congestion to the total network cost	8	9	6	9	6	6	6	43	6	43
Average hub capacity utilization (%)	37	38	46	38	46	46	46	73	46	73
Number of hubs with utilization above 95%	0	0	0	0	0	0	0	1	0	1
Number of hubs	6	6	4	6	4	4	4	2	4	2
TR81	**0.20**		**0.40**		**0.60**		**0.80**		**1.00**	
Φ (%)	53.33		98.16		95.82		16.86		96.18	
Congestion included?	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
% contribution of congestion to the total network cost	6	40	5	53	5	52	4	21	5	53
Average hub capacity utilization (%)	51	66	63	89	73	97	71	93	84	92
Number of hubs with utilization above 95%	0	4	0	4	0	6	0	4	0	2
Number of hubs	17	15	12	7	9	6	6	5	6	5

Table 8 demonstrates that, in general, omitting hub port congestion in network optimization has a negative impact on costs. Although the effect of congestion is modest for MED15 and some of the MED20 instances, the cost increase is high in MED20-0.80, MED20-1.00, and all TR81 instances where Φ is considerably large. For the instances of TR81-0.40, TR81-0.60, TR81-1.00, omitting congestion cost nearly doubled the total cost of the resulting networks. Likewise, the network costs increased by 65% and 62% in MED20-0.80 and MED20-1.00, respectively. Furthermore, when the hub port congestion cost was left out of the optimization, the network was designed and optimized in such a way that, first, fewer hubs were located, and second, the capacities of the located hub ports were utilized almost fully. This resulted in a very high cost of congestion and increased the total network costs. In the TR81-0.60 instance for example, all the six hub ports' capacity utilizations exceed 95%, and the average hub capacity utilization is 97%. Consequently, the percent contribution of the cost of hub port congestion to the total network cost was larger than 50%. Moreover, in all instances of TR81, as well as in MED15-1.00, MED20-0.80, and MED20-1.00, fewer hubs were located when the congestion cost was not considered because locating fewer hubs reduces the hub opening cost and may yield a lower cost of inter-hub cargo transfer. These results demonstrate that ignoring hub port congestion cost in the process of liner shipping network design may result in costly and inefficient designs.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we studied the design and optimization of liner shipping HS networks. In line with practices of the liner shipping industry, we introduced the capacitated directed cycle hub location and routing problem under congestion (DCHC), which features hub port congestion, a cyclic hub-level network, and cargo transshipment cost characteristics. The problem was formulated as a mixed integer programming model with a nonlinear congestion cost component in the objective function. We formulated a semi-

continuous piecewise linear approximation for the congestion cost to linearize the objective function and designed a probabilistic Tabu Search algorithm. Problem instances with 10 to 81 nodes were generated from five data sets and used in computational experiments. Analysis of the solutions obtained through the Tabu Search algorithm and lower bound solutions from the approximate congestion cost model solved with CPLEX demonstrates the effectiveness of the Tabu Search algorithm. Finally, we derived managerial insights into the effect of the discount factor, centrality and geographical intermediacy of ports, and congestion on the HS network design and costs.

Although DCHC is defined in the context of the liner shipping network design, the model can be applied to other applications. Potential areas include the design of public transportation, multimodal transportation, or telecommunications networks, where establishment of a fully connected hub-level network is expensive due to infrastructure and maintenance costs or impractical due to technological and practical considerations. As mentioned by Contreras et al. (2016), in the case of public transportation networks the hub-level network may refer to a circular rapid transit line such as a metro or high-speed train, hubs to metro/train stations, access arcs to transportation links by bus or taxi, and non-hub nodes to bus stops or remote districts. Similarly, in an intermodal transportation network the hub-level network can be regarded as a high-capacity transport mode (e.g. rail). Hubs refer to logistics centers where the modal transfer from a low-capacity transport mode (e.g. trucks) to the high-capacity mode is done. Non-hub nodes correspond to regional centers or customer areas.

Our model could be extended in future research as follows. We assumed that there are no setup costs for inter-hub or access arcs, but there are usually administrative and other costs involved in establishing links between ports. The problem is strictly cost-oriented; however, the service quality of the network could be analyzed through the incorporation of the hub-level network length or total OD transit time constraints.

There are several other exciting research opportunities worth investigating. Uncertainty is common in many transportation networks including liner shipping and may be caused by a wide range of sources such as OD flows, hub opening costs, and other network design parameters. Data imprecision is one type of uncertainty encountered in hub location and other network design problems and refers to the ambiguity in parameter values due to the decision maker's lack of knowledge. Uncertainty caused by ambiguity can be explored using fuzzy measures in possibilistic programming approaches. Another potential research area relates to extending the planning horizon from single to multiple periods. The decisions concerned with HS network design are long-term and the design parameters such as demand or unit transportation costs may vary over time. Designing the network with the inputs of a single planning period may prove to be short-sighted. Considering multiple periods and taking variations in demand and other parameters in the problem can improve the network design in this respect. Finally, the objective function of our problem is cost minimization and ignores other requirements of the industry stakeholders such as schedule reliability, delivery flexibility, and environmental concerns. These requirements are often in conflict and yet shipping lines must pay attention to all of them to attract more customers and comply with maritime regulations. Defining multiple objectives can emphasize the priorities of other stakeholders in network design.

REFERENCES

Abyazi-Sani, R., & Ghanbari, R. (2016). An efficient tabu search for solving the uncapacitated single allocation hub location problem. *Computers & Industrial Engineering*, 93, 99-109.

Alumur, S. A., Kara, B. Y., & Karasan, O. E. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B: Methodological*, 43(10), 936-951.

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.

Aversa, R., Botter, R. C., Haralambides, H. E., & Yoshizaki, H. T. Y. (2005). A Mixed Integer Programming Model on the Location of a Hub Port in the East Coast of South America. *Maritime Economics & Logistics*, 7(1), 1-18.

Baird, A.J. (2006). Optimising the container transshipment hub location in northern Europe. *Journal of Transport Geography*, 14, 195-214.

Baldacci, R., Dell'Amico, M., & González, J. S. (2007). The capacitated *m*-ring-star problem. *Operations Research*, 55(6), 1147-1162.

Bazaraa, M.S., Sherali, H.D., & Shetty, C.M. (2006). *Nonlinear programming: Theory and algorithms* (3rd ed.). New Jersey: Wiley.

Bilkent.edu.tr (2018). *HUB LOCATION*. http://bkara.bilkent.edu.tr/hubloc.htm. Accessed 1 February 2018.

Bisschop, J. (2020). Chapter 7 Integer Linear Programming Tricks. In Bisschop, J., *AIMMS Optimization Modeling* https://documentation.aimms.com/_downloads/AIMMS_modeling.pdf. Accessed 2 July 2020.

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

Calik, H., Alumur, S. A., Kara, B. Y., & Karasan, O. E. (2009). A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Computers & Operations Research*, 36(12), 3088-3096.

Campbell, J.F., Ernst, A.T., & Krishnamoorthy, M. (2005a). Hub Arc Location Problems: Part I-Introduction and Results. *Management Science*, 51(10), 1540-1555.

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

Cariou, P. (2008). Liner shipping strategies: an overview. Int. J. *Ocean Systems Management*, 1(1), 1-12.

Chen, J.-F. (2007). A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega*, 35(2), 211-220.

Chou, C.-C. (2010). Application of FMCDM model to selecting the hub location in the marine transportation: A case study in southeastern Asia. *Mathematical and Computer Modelling*, 51(5-6), 791-801.

Contreras, I. (2015). Hub Location Problems. In G. Laporte, S. Nickel, F. Saldanha da Gama, *Location Science* (pp. 311-339). Cham: Springer International Publishing.

Contreras, I., Fernández, E., & Marín, A. (2009). 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., & Fernández, E. (2014). Hub Location as the Minimization of a Supermodular Set Function. *Operations Research*, 62(3), 557-570.

Contreras, I., Tanash, M., & Vidyarthi, N. (2016). Exact and heuristic approaches for the cycle hub location problem. *Annals of Operations Research*, 258(2), 655-677.

Correia, I., Nickel, S., & Saldanha-da-Gama, F. (2010). Single-assignment hub location problems with multiple capacity levels. *Transportation Research Part B: Methodological*, 44(8-9), 1047-1066.

de Camargo, R. S., de Miranda, G., & Ferreira, R. P. M. (2011). A hybrid Outer-Approximation/Benders Decomposition algorithm for the single allocation hub location problem under congestion. *Operations Research Letters*, 39(5), 329-337.

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

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

Drexl, M., & Schneider, M. (2015). A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, 241(2), 283-308.

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

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, 2019, 358-374

Fleming, D.K. & Hayuth, Y. (1994). Spatial characteristics of transportation hubs: centrality and intermediacy. *Journal of Transport Geography*, 2(1), 3-18.

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.

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., & Pisinger, D. (2011). Fleet deployment, network design and hub location of liner shipping companies. Transportation Research Part E: *Logistics and Transportation Review*, 47(6), 947-964.

Ghodratnama, A., Arbabi, H. R., & Azaron, A. (2019). Production planning in industrial townships modeled as hub location–allocation problems considering congestion in manufacturing plants. *Computers & Industrial Engineering*, 129, 479-501.

Ishfaq, R., & Sox, C. R. (2012). Design of intermodal logistics networks with hub delays. *European Journal of Operational Research*, 220(3), 629-641.

Kahag, M. R., Niaki, S. T. A., Seifbarghy, M., & Zabihi, S. (2019). Bi-objective optimization of multi-server intermodal hub-location-allocation problem in congested systems: modeling and solution. *Journal of Industrial Engineering International*, 15(2), 221-248.

Khodemani-Yazdi, M., Tavakkoli-Moghaddam, R., Bashiri, M., & Rahimi, Y. (2019). Solving a new bi-objective hierarchical hub location problem with an M/M/c queuing framework. *Engineering Applications of Artificial Intelligence*, 78, 53-70.

Kian, R., & Kargar, K. (2016). Comparison of the formulations for a hub-and-spoke network design problem under congestion. *Computers & Industrial Engineering*, 101, 504-512.

Köksalan, M., & Soylu, B. (2010). Bicriteria *p*-Hub Location Problems and Evolutionary Algorithms. *INFORMS Journal on Computing*, 22(4), 528-542.

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

Labbé, M., Laporte, G., Rodriguez-Martin, I., & Gonzalez, J.J.Z. (2004). The Ring Star Problem: Polyhedral Analysis and Exact Algorithm. *Networks*. 43(3). 177-189.

Labbé, M., Laporte, G., Rodriguez-Martin, I., & Gonzalez, J.J.Z. (2005). Locating median cycles in networks. *European Journal of Operational Research. 160*. 457-470.

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(6), 625-637.

Lopes, R. B., Ferreira, C., Santos, B. S., & Barreto, S. (2013). A taxonomical analysis, current methods and objectives on location-routing problems. *International Transactions in Operational Research*, 20(6), 795-822.

Marianov, V. & Serra, D. (2003). Location models for airline hubs behaving as M/D/c queues. *Computers & Operations Research*, 30, 983-1003.

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

Martins de Sá, E., Contreras, I.,Cordeau, J.-F., de Camargo, R.S., de Miranda, G. (2015b). The Hub Line Location Problem. *Transportation Science*, 49(3), 433-719.

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

Mohammadi, M., Jolai, F., & Rostami, H. (2011). An M/M/c queue model for hub covering location problem. *Mathematical and Computer Modelling*, 54(11-12), 2623-2638.

Mohammadi, M., Tavakkoli-Moghaddam, R., Siadat, A., & Rahimi, Y. (2016). A game-based meta-heuristic for a fuzzy bi-objective reliable hub location problem. *Engineering Applications of Artificial Intelligence*, 50, 1-19.

Mohammadi, M., Jula, P., & Tavakkoli-Moghaddam, R. (2017). Design of a reliable multi-modal multi-commodity model for hazardous materials transportation under uncertainty. *European Journal of Operational Research*, 257(3), 792-809.

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

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

Ortiz-Astorquiza, C., Contreras, I., & Laporte, G. (2015). The minimum flow cost Hamiltonian cycle problem: A comparison of formulations. *Discrete Applied Mathematics*, 187, 140-154.

OR-Library (2018). OR-Library. http://people.brunel.ac.uk/~mastjjb/jeb/info.html. Accessed 1 February 2018.

Rahimi, Y., Tavakkoli-Moghaddam, R., Mohammadi, M., & Sadeghi, M. (2016). Multi-objective hub network design under uncertainty considering congestion: An M/M/c/K queue system. *Applied Mathematical Modelling*, 40(5-6), 4179-4198.

Reinhardt, L. R. & Pisinger, D. (2012). A branch and cut algorithm for the container shipping. Flexible Services and Manufacturing Journal, 24, 349-374.

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(3), 863-878.

Rodrigue, J.P., Comtois, C., & Slack, B. (2013). *The Geography of Transport Systems* (3rd Edition). Oxon: Routledge.

Rodriguez, V., Alvarez, M. J., & Barcos, L. (2007). Hub location under capacity constraints. *Transportation Research Part E: Logistics and Transportation Review*, 43(5), 495-505.

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.

Simonetti, L., Frota, Y., & de Souza, C. C. (2011). The ring-star problem: a new integer programming formulation and a branch-and-cut algorithm. *Discrete applied mathematics*. 159(16), 1901-1914.

Song, D.-P., & Dong, J.-X. (2013). Long-haul liner service route design with ship deployment and empty container repositioning. *Transportation Research Part B: Methodological*, 55, 188-211.

Sun, Z., & Zheng, J. (2016). Finding potential hub locations for liner shipping. *Transportation Research Part B: Methodological*, 93, 750-761.

Takano, K., & Arai, M. (2008). A genetic algorithm for the hub-and-spoke problem applied to containerized cargo transport. *Journal of Marine Science and Technology*, 14(2), 256-274.

Talbi, E. (2009). Metaheuristics from design to implementation. New Jersey: John Wiley & Sons.

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(6), 643-658.

Yaman, H., & Elloumi, S. (2012). Star *p*-hub center problem and star *p*-hub median problem with bounded path lengths. *Computers & Operations Research*, 39(11), 2725-2732.

Zhalechian, M., Tavakkoli-Moghaddam, R., & Rahimi, Y. (2017). A self-adaptive evolutionary algorithm for a fuzzy multi-objective hub location problem: An integration of responsiveness and social responsibility. *Engineering Applications of Artificial Intelligence*, 62, 1-16.

Zheng, J., Meng, Q., & Sun, Z. (2014). Impact analysis of maritime cabotage legislations on liner hub-and-spoke shipping network design. *European Journal of Operational Research*, 234, 874–884.

Zheng, J., Meng, Q., & Sun, Z. (2015). Liner hub-and-spoke shipping network design. *Transportation Research Part E: Logistics and Transportation Review*, 75, 32-48.

Zheng, J., Qi, J., Sun, Z., & Li, F. (2018). Community structure based global hub location problem in liner shipping. *Transportation Research Part E: Logistics and Transportation Review*, 118, 1-19.

Özgün-Kibiroğlu, Ç., Serarslan, M. N., & Topcu, Y. İ. (2019). Particle swarm optimization for uncapacitated multiple allocation hub location problem under congestion. *Expert Systems with Applications*, 119, 1-19.

APPENDICES

APPENDIX I: SEMI-CONTINUOUS PIECEWISE LINEAR APPROXIMATION FOR THE CONGESTION COST

For the approximation of the hub congestion function $f(\rho_k) = 1/(1-\rho_k)$ in the interval $[a', b']$ with $0 \leq a' < b' < 1$, and a' (b') a lower (upper) limit on the utilization ρ_k, we consider an equivalent curve $f(x) = 1/x$ in $[a, b]$ via the transformation: $x = 1 - \rho_k$; $a = 1 - b'$; $b = 1 - a'$. Figure 10 shows $f(x)$ in $[a = 0.1, b = 0.9]$ and a semi-continuous piecewise linear approximation with four segments. The approximation improves when more segments are defined. Our aim is to find the best (semi-continuous) piecewise linear approximation for a given number of segments v, i.e., the approximation that minimizes the total area between $f(x)$ and the approximation, or that maximizes the area of the v trapezoids in interval $[a, b]$.

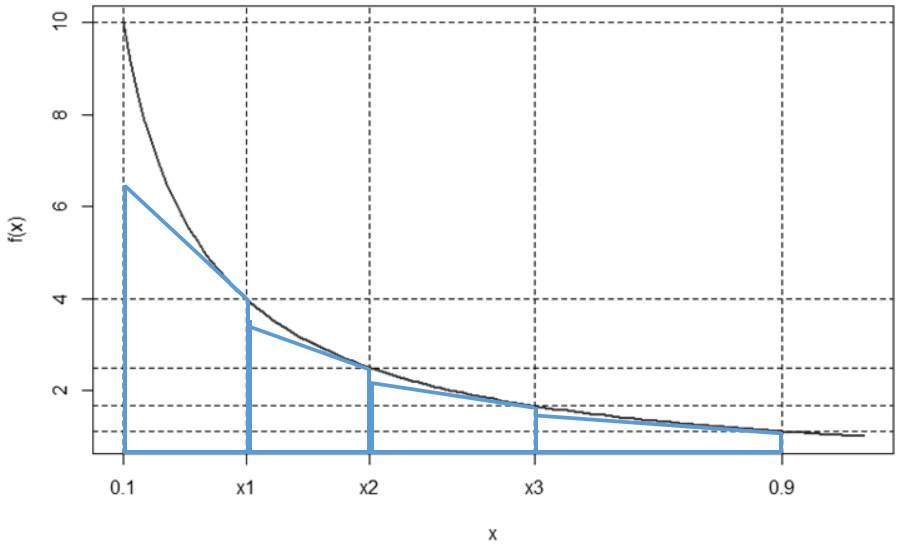

Figure 10: An illustration of the semi-continuous piecewise linear approximation

The best approximation through a semi-continuous piecewise linear function with v segments is determined by the position of the breakpoints $x_1, x_2, \ldots, x_{v-1}$, such that the area of the trapezoids, denoted by $G(x_1, \ldots, x_{v-1}) = \left[(x_1 - a)f(x_1) - \left(\frac{1}{2}\right)(x_1 - a)^2 f'(x_1)\right] + \left[(x_2 - x_1)f(x_2) - \left(\frac{1}{2}\right)(x_2 - x_1)^2 f'(x_2)\right] + \cdots + \left[(x_{v-1} - x_{v-2})f(x_{v-1}) - \left(\frac{1}{2}\right)(x_{v-1} - x_{v-2})^2 f'(x_{v-1})\right] + \left[(x_b - x_{v-1})f(b) - \left(\frac{1}{2}\right)(x_b - x_{v-1})^2 f'(b)\right]$ is maximized, and subject to the constraints: $a \leq x_1; \ldots; x_l \leq x_{l+1}; \ldots; x_{v-1} \leq b$. By substituting $f(x_l) = 1/x_l$ and $f'(x_l) = -1/{x_l}^2$, we can rewrite the objective function as $G(x_1, \ldots, x_{v-1}) = 3v/2 - 2(a/x_1 + x_1/x_2 + \ldots + x_{v-1}/b) + \frac{1}{2}(a^2/{x_1}^2 + {x_1}^2/{x_2}^2 + \ldots + {x_{v-1}}^2/b^2)$. This is a constrained non-linear optimization problem (Bazaraa et al., 2006). The feasible region is bounded (limits a and b) and convex (linear constraints). Furthermore, the optimal solution cannot be on the boundary of the feasible region. A boundary solution has at least one binding constraint, which means that the positions of at least two breakpoints coincide (and hence the approximation being worse compared to when all breakpoints have different positions). Therefore, the optimum solution must be an interior point in the feasible region.

The first order conditions for $G(x_1, \ldots, x_{v-1})$ lead to expressions: $dG/dx_1 = 2a/{x_1}^2 - a^2/{x_1}^3 - 2/x_2 + x_1/{x_2}^2, \ldots, dG/dx_l = 2x_{l-1}/{x_l}^2 - {x_{l-1}}^2/{x_l}^3 - 2/x_{l+1} + x_l/{x_{l+1}}^2, \ldots, dG/dx_{v-1} = 2x_{v-2}/{x_{v-1}}^2 - {x_{v-2}}^2/{x_{v-1}}^3 - 2/b + x_{v-1}/b^2$. Equating the partial derivatives to zero, leads to quartic polynomial equations: ${x_1}^4 - 2x_2{x_1}^3 + 2a{x_2}^2x_1 - a^2{x_2}^2 = 0, \ldots, {x_l}^4 -$

$2x_{l+1}{x_l}^3 + 2x_{l-1}{x_{l+1}}^2 x_l - {x_{l-1}}^2{x_{l+1}}^2 = 0, \ldots, {x_{v-1}}^4 - 2b{x_{v-1}}^3 + 2x_{v-2}b^2x_{v-1} - {x_{v-2}}^2b^2 = 0.$ Each equation can be factorized, for example, the quartic polynomial equation,

${x_l}^4 - 2x_{l+1}{x_l}^3 + 2x_{l-1}{x_{l+1}}^2x_l - {x_{l-1}}^2{x_{l+1}}^2 = 0$ leads to $\left(x_l - (x_{l-1}x_{l+1})^{1/2}\right)\left(x_l + (x_{l-1}x_{l+1})^{1/2}\right)\left(x_l - x_{l+1} - ({x_{l+1}}^2 - x_{l-1}x_{l+1})^{1/2}\right)\left(x_l - x_{l+1} + ({x_{l+1}}^2 - x_{l-1}x_{l+1})^{1/2}\right) = 0$. It is easy to verify that only the first root, $x_l = (x_{l-1}x_{l+1})^{1/2}$, is feasible (the second root is negative, the third root violates $x_l \leq x_{l+1}$ and the fourth root violates $x_{l-1} \leq x_l$). Therefore, there is only one critical point $x^*({x_1}^*, \ldots, {x_{v-1}}^*)$, which is also an interior point in the feasible region, with ${x_1}^* = (ax_2)^{1/2}$, ${x_2}^* = (x_1x_3)^{1/2}$, ${x_l}^* = (x_{l-1}x_{l+1})^{1/2}$, ${x_{v-1}}^* = (x_{v-2}b)^{1/2}$. After some algebra, it can be shown that: ${x_l}^* = (a^{v-l}b^l)^{1/v}$, for $l = 1,2, \ldots, v-1$.

To establish that x^* is a local maximum of G, we consider the Hessian matrix of $-G$ and show that all principal minors are positive in x^* (i.e., x^* is a local minimum of $-G$). Local convexity at the unique, critical point x^* together with the fact that the optimum cannot occur on the boundary, is sufficient to claim that x^* is a global minimizer of $-G$ (or that x^* is a global maximizer of G).

The Hessian of $-G$ is:

$$H = \begin{bmatrix}
\frac{4a}{x_1^3} - \frac{3a^2}{x_1^4} - \frac{1}{x_2^2} & -\frac{2}{x_2^2} + \frac{2x_1}{x_2^3} & 0 & 0 & \ldots & 0 & 0 & 0 \\
-\frac{2}{x_2^2} + \frac{2x_1}{x_2^3} & \frac{4x_1}{x_2^3} - \frac{3x_1^2}{x_2^4} - \frac{1}{x_3^2} & -\frac{2}{x_3^2} + \frac{2x_2}{x_3^3} & 0 & & 0 & 0 & 0 \\
0 & -\frac{2}{x_3^2} + \frac{2x_2}{x_3^3} & \frac{4x_2}{x_3^3} - \frac{3x_2^2}{x_3^4} - \frac{1}{x_4^2} & -\frac{2}{x_4^2} + \frac{2x_3}{x_4^3} & & 0 & 0 & 0 \\
0 & 0 & -\frac{2}{x_4^2} + \frac{2x_3}{x_4^3} & \frac{4x_3}{x_4^3} - \frac{3x_3^2}{x_4^4} - \frac{1}{x_5^2} & & 0 & 0 & 0 \\
\ldots & & & & \ldots & & & \ldots \\
0 & 0 & 0 & & & \frac{4x_{v-4}}{x_{v-3}^3} - \frac{3x_{v-4}^2}{x_{v-3}^4} - \frac{1}{x_{v-2}^2} & -\frac{2}{x_{v-2}^2} + \frac{2x_{v-3}}{x_{v-2}^3} & 0 \\
0 & 0 & 0 & & & -\frac{2}{x_{v-2}^2} + \frac{2x_{v-3}}{x_{v-2}^3} & \frac{4x_{v-3}}{x_{v-2}^3} - \frac{3x_{v-2}^2}{x_{v-2}^4} - \frac{1}{x_{v-1}^2} & -\frac{2}{x_{v-1}^2} + \frac{2x_{v-2}}{x_{v-1}^3} \\
0 & 0 & 0 & 0 & \ldots & & -\frac{2}{x_{v-1}^2} + \frac{2x_{v-2}}{x_{v-1}^3} & \frac{4x_{v-2}}{x_{v-1}^3} - \frac{3x_{v-2}^2}{x_{v-1}^4} - \frac{1}{b^2}
\end{bmatrix}$$

The first principal minors are the diagonal elements of H. Evaluated in x^*, these are all positive. For example, the l^{th} diagonal element, $\frac{4x_{l-1}}{x_l^3} - \frac{3x_{l-1}^2}{x_l^4} - \frac{1}{x_{l+1}^2}$, evaluated with ${x_{l-1}}^* = (a^{v-l+1}b^{l-1})^{1/v}, {x_l}^* = (a^{v-l}b^l)^{1/v}$ and ${x_{l+1}}^* = (a^{v-l-1}b^{l+1})^{1/v}$, yields $\frac{4\left(\left(\frac{b}{a}\right)^{\frac{1}{v}} - 1\right)}{(a^{v-l-1}b^{l+1})^{\frac{2}{v}}}$, which is positive because $b > a$. We note that the other (non-zero) elements of H (evaluated in x^*) are all negative.

For example, evaluating the element in row l, column l+1, $-\frac{2}{x_{l+1}^2} + \frac{2x_l}{x_{l+1}^3}$ in x^* results in: $\frac{-2\left(1 - \left(\frac{a}{b}\right)^{\frac{1}{v}}\right)}{(a^{v-l-1}b^{l+1})^{\frac{2}{v}}}$, which is negative because $b > a$.

There are two different types of second principal minors: those whose matrix ($M_{2,1}$) has only diagonal elements (retaining two rows and corresponding columns which were not adjacent in H), and those with matrix structure $M_{2,2}$ (retaining two adjacent rows l and l+1 and corresponding columns in H):

$$M_{2,2} = \begin{bmatrix}
\frac{4x_{l-1}}{x_l^3} - \frac{3x_{l-1}^2}{x_l^4} - \frac{1}{x_{l+1}^2} & -\frac{2}{x_{l+1}^2} + \frac{2x_l}{x_{l+1}^3} \\
-\frac{2}{x_{l+1}^2} + \frac{2x_l}{x_{l+1}^3} & \frac{4x_l}{x_{l+1}^3} - \frac{3x_l^2}{x_{l+1}^4} - \frac{1}{x_{l+2}^2}
\end{bmatrix}$$

The determinants of the matrices $M_{2,1}$ with diagonal elements only, evaluated in x^*, are positive. A positive determinant in x^* for the $M_{2,2}$ matrices corresponds to testing if $\left(\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2}\right)\left(\frac{4x_l}{x_{l+1}^3}-\frac{3x_l^2}{x_{l+1}^4}-\frac{1}{x_{l+2}^2}\right)-\left(-\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3}\right)^2>0$ in x^*. It is easily verified that this leads to $\frac{12\left(\left(\frac{b}{a}\right)^{\frac{1}{v}}-1\right)^2}{\left(a^{2v-2l-3}b^{2l+3}\right)^{\frac{2}{v}}}$, which is positive because $b>a$, so that all second minors are positive.

The matrices for the third principal minors come in three different structures: $M_{3,1}$ with only diagonal elements, $M_{3,2}$ by retaining two adjacent rows and corresponding columns in H; and $M_{3,3}$ matrices by retaining three adjacent rows and corresponding columns in H.

$$M_{3,2}=\begin{bmatrix}\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2} & -\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3} & 0\\ -\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3} & \frac{4x_l}{x_{l+1}^3}-\frac{3x_l^2}{x_{l+1}^4}-\frac{1}{x_{l+2}^2} & 0\\ 0 & 0 & \frac{4x_p}{x_{p+1}^3}-\frac{3x_p^2}{x_{p+1}^4}-\frac{1}{x_{p+2}^2}\end{bmatrix}$$

$$M_{3,3}=\begin{bmatrix}\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2} & -\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3} & 0\\ -\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3} & \frac{4x_l}{x_{l+1}^3}-\frac{3x_l^2}{x_{l+1}^4}-\frac{1}{x_{l+2}^2} & -\frac{2}{x_{l+2}^2}+\frac{2x_{l+1}}{x_{l+2}^3}\\ 0 & -\frac{2}{x_{l+2}^2}+\frac{2x_{l+1}}{x_{l+2}^3} & \frac{4x_{l+1}}{x_{l+2}^3}-\frac{3x_{l+1}^2}{x_{l+2}^4}-\frac{1}{x_{l+3}^2}\end{bmatrix}$$

The determinants of the $M_{3,1}$ matrices are positive in x^*. Expanding the last row of $M_{3,2}$ shows that $detM_{3,2} = detM_{2,2}*(\frac{4x_p}{x_{p+1}^3}-\frac{3x_p^2}{x_{p+1}^4}-\frac{1}{x_{p+2}^2})$, and both $detM_{2,2}$ and the coefficient $(\frac{4x_p}{x_{p+1}^3}-\frac{3x_p^2}{x_{p+1}^4}-\frac{1}{x_{p+2}^2})$ are positive in x*. All $M_{3,3}$ determinants are also positive: expanding the first row yields $detM_{3,3} = detM_{2,2}*(\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2})$- $detM_{2,1}*(-\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3})$. Both $detM_{2,2}$ and $detM_{2,1}$ are positive in in x* with a positive (diagonal element of H) and negative coefficient (non-zero, non-diagonal element of H) respectively.

Likewise, all higher principal minors are positive in x*: expanding, for example, the first row of an $M_{k,k}$ matrix (i.e., the matrix by retaining k adjacent rows and corresponding columns in H), yields:

$detM_{k,k} = detM_{k\text{-}1,k\text{-}1}*(\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2})$ - $detM_{k\text{-}1,k\text{-}2}*(-\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3})$. Because the lower principal minors and $(\frac{4x_{l-1}}{x_l^3}-\frac{3x_{l-1}^2}{x_l^4}-\frac{1}{x_{l+1}^2})$ are positive in x^*, and $(-\frac{2}{x_{l+1}^2}+\frac{2x_l}{x_{l+1}^3})$ is negative in x^*, $detM_{p,p}$ is also positive in x^*.

All principal minors are positive in critical point x^*. Because the optimum cannot be a boundary solution, point x^* is also a global minimizer of $-G$ (or x^* is a global maximizer of G).

We can measure the quality of the approximation through the percentage difference between the area under $f(x)$ and the area under the approximation in interval $[a,b]$ (i.e. $G(x^*)$):

$$\% \, error = \left(1 - \frac{v\left(\left(\frac{3}{2}\right) - 2\left(\frac{a}{b}\right)^{\frac{1}{v}} + \frac{1}{2}\left(\frac{a}{b}\right)^{\frac{2}{v}} \right)}{\ln\left(\frac{b}{a}\right)} \right) * 100\% \tag{39}$$

Table 9 shows the approximation error for values of v between 5 and 50 and demonstrates that the approximation improves with an increasing number of segments. In our experiments we used $v = 25$ segments, $a = 0.05$ and $b = 0.90$, which corresponds to minimum and maximum utilization levels ρ_k between 0.1 and 0.95, and a percentage error of 0.41%. (In the experiments with the mixed integer programming model, we allowed hubs to operate with utilization rates outside interval [0.1, 0.95]. In case a hub had a very low utilization ($\rho_k < 10$), $f(\rho_k)$ was approximated with a line segment that is tangent to $f(\rho_k)$ at $\rho_k = 0$; in case a hub had a very high utilization ($\rho_k > 0.95$), $f(\rho_k)$ was approximated with a line segment that is tangent to $f(\rho_k)$ at $\rho_k = 0.95$.)

Table 9: Percentage error of the approximation (a=0.05, b=0.90)

v	5	10	15	20	25	30	35	40	45	50
% error	7.38	2.25	1.07	0.63	0.41	0.29	0.21	0.16	0.13	0.11

To assess the quality of the lower bound solutions using the MILP with congestion cost approximation we solved the MED15-0.20 instance with CPLEX and varied the number of segments between 5 and 50. The minimum and maximum utilization levels were kept at 0.10 and 0.95. The results are presented in Figure 11 where the dots show the MILP lower bound values and the straight line represents the best solution cost obtained with TSHLP. A congestion cost approximation with more segments clearly tends to produce better bounds although the improvement is not necessarily monotone and some fluctuations may exist. For this instance, the lower bound value with *v* = 30 is slightly worse than the bounds obtained with *v* = 25 and *v* = 35, while the actual solution structures (i.e., the DCHC network and flows) are exactly the same. This is due to the nature of the approximation and the jumps at the utilization breakpoints: the approximation is better (worse) when the hub capacity utilization in a solution is slightly above (below) a breakpoint value. Changing *v*, re-positions the breakpoints and the same solution may produce a slightly better or worse lower bound value in the MILP. These fluctuations are small and diminish when *v* grows larger.

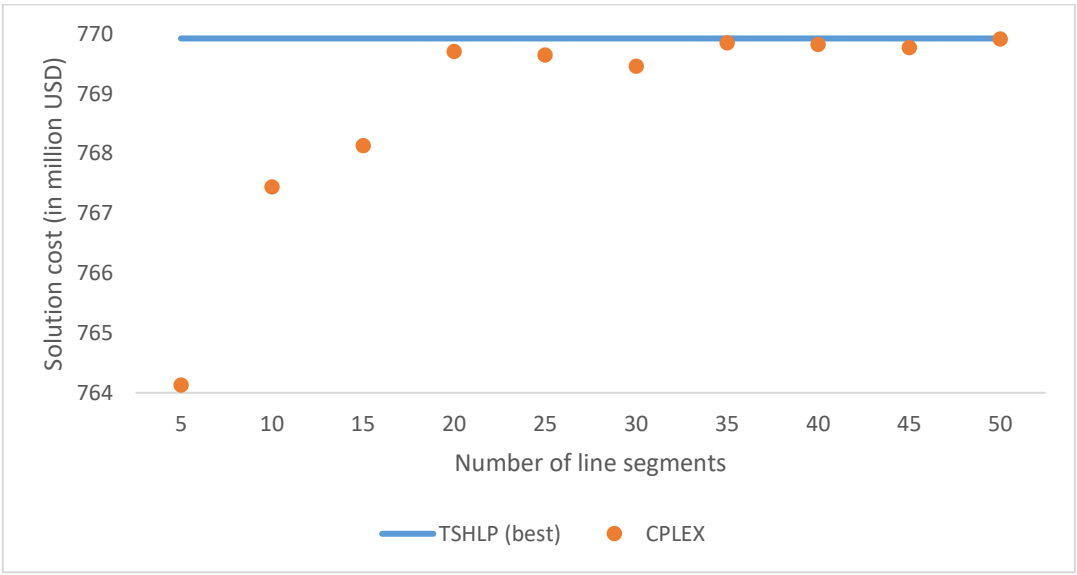

Figure 11: Improving lower bound solutions with increasing number of line segments in the approximation (MED15-0.20 instance)

APPENDIX II: PSEUDOCODE OF TSHLP

Sets

L	Set of located hubs in the initial solution
E	Set of unallocated non-hub nodes in the initial solution

Parameters

tf	Total flows routed in the network, in TEU, $tf = \sum_{i=1}^{n}(o_i + d_i)$
$tcap$	Total capacity of hub candidates in TEU, $tcap = \sum_{k=1}^{h} cap_k$
I_k	Hub candidate index of $k \in H$
wd_{ik}	The modified distance between node $i \in N$ and hub candidate node $k \in H$
$volume\_factor$	determines the minimum total capacity the hubs located in the initial solution must have, $1 < volume\_factor \leq tcap/tf$
$termination\_iter$	Termination iteration counter of TSHLP
max_iter	TSHLP terminates when $termination\_iter$ reaches max_iter
f_open	List of hub location frequencies
max_freq	The maximum frequency of f_open
$freq\_lim$	Diversification is triggered when max_freq reaches $freq\_lim$
S_0	The initial feasible solution
S_{iter}	The solution selected in the current iteration
S_{best}	The best solution of the current run of the algorithm
S_{global}	The global best solution obtained throughout TSHLP

// Initial solution procedure
Input: Problem and initial solution generation parameters
Step 1: Read the input data.
Step 2: Set E to N, S_0 to $\emptyset$.
Step 3: Calculate I_k for each hub candidate, $k \in H$.
Step 4: Rank the hub candidates in non-increasing order of I_k.
Step 5: Calculate the modified distances wd_{ik} for all $i \in N$.
Step 6: Obtain the initial hub locations
Start with an empty set of located hubs, $L = \{\emptyset\}$.
while $\sum_{k \in L} cap_k < volume\_factor * tf$ ***do***
Open a new hub from the top of the index list and add to L
Remove the hub node from E
end
Step 7: Obtain the initial node allocations
while $E \neq \{\emptyset\}$ ***do***
Allocate i to the nearest available hub according to $wd_{ik} \forall\ i \in E, k \in H$
Remove i from E
end
Step 8: Update S_0 with the hub locations and node allocations.
Step 9: Apply nearest neighbour heuristic to the hubs in S_0 to obtain a cyclic hub-level network
Step 10: Return S_0.

//TSHLP

Input: Problem instance and TSHLP parameters

Step 1: Run initial solution procedure, obtain S_0.

Step 2: Set $termination\_iter$ to 0.

Step 3: Set $S_{iter}, S_{best}, S_{global}$ to S_0.

while $termination\_iter < max\_iter$ **do**

Step 4: Do neighbourhood search and update S_{iter}

Step 5: Check if the conditions to run local improvement for hub-level network (**LISUB)** holds

if $(Interhub\ transfer\ cost\ of\ S_{iter} > S_{best})$ ***then***

run LISUB;

$Update\ S_{iter}$;

end

Step 6: Update S_{best}.

if $(S_{iter} < S_{best})$ ***then***

$Set\ S_{best}\ to\ S_{iter}$;

end

Step 7: Update S_{global}.

if $\left(S_{best} < S_{global}\right)$ ***then***

$Set\ S_{global}\ to\ S_{best}$;

else

$termination\_iter = termination\_iter + 1$;

end

Step 8: Update f_open and Tabu lists.

Step 9: Check if the conditions for diversification hold.

if $(max\_freq \geq freq\_lim)$ ***then***

run Diversification

end

end

Step10: Return S_{global}.