Hub location problems: A review of models, classification, solution techniques, and applications

Reza Zanjirani Farahani, Masoud Hekmatfar, Alireza Boloori Arabani, Ehsan Nikbakhsh

ABSTRACT

Hub location problem (HLP) is a relatively new extension of classical facility location problems. Hubs are facilities that work as consolidation, connecting, and switching points for flows between stipulated origins and destinations. While there are few review papers on hub location problems, the most recent one (Alumur and Kara, 2008. Network hub location problems: The state of the art. European Journal of Operational Research, 190, 1–21) considers solely studies on network-type hub location models prior to early 2007. Therefore, this paper focuses on reviewing the most recent advances in HLP from 2007 up to now. In this paper, a review of all variants of HLPs (i.e., network, continuous, and discrete HLPs) is provided. In particular, mathematical models, solution methods, main specifications, and applications of HLPs are discussed. Furthermore, some case studies illustrating real-world applications of HLPs are briefly introduced. At the end, future research directions and trends will be presented.

1. Introduction

Hub location problem (HLP) is one of the novel and thriving research areas in location theory. In order to satisfy a demand, HLP involves the movement of people, commodities, or information between required origin–destination pairs. Hubs are applied to decrease the number of transportation links between origin and destination nodes. For example, a fully connected network with $k$ nodes and with no hub node has $k(k - 1)$ origin–destination links. However, if a hub node is selected to connect all other nodes (non-hub nodes, also known as spokes) with each other, there will be only $2(k - 1)$ connections to serve all origin–destination pairs. This idea can be extended to a network with more than a hub node, called a multiple-hub network. Thus, by using fewer resources, demand pairs can be served more efficiently with a hub network than with a fully connected structure.

Costs of a hub network depend on its structure. The total distance of arcs connecting the whole network might be less in a hub network, but the total travel distance may be greater since there is no guarantee that the number of people, merchandise, or information moving on the hub-to-hub connections be greater than those moving between hub and spokes. Therefore, it will be very complicated to determine location of hub facilities as well as allocation scheme of clients to them (Campbell, Lowe, & Zhang, 2007).

It seems that the telecommunication industry is originally one of the oldest users of hub network concept. However, in logistical systems, airline industry and postal companies are one of the main users of this concept. Today, there are many other areas that can take advantage of hub concept like maritime industry, freight transportation companies, public transit, and message delivery networks.

Originally, Hakimi (1964) published the first paper in the area of node optimality which was motivated by similar concepts to HLP. Later, Toh and Higgins (1985) discussed application of the hub location networks for airlines and aviation industry as the first pertinent research. The first papers in the HLP field, presenting the first mathematical formulation and solution method, were due to efforts of O’Kelly (1986a, 1986b). Since then, many papers have been published over recent years with a significant increasing trend. Regarding this trend, the focus of papers was on modeling in the late 1980s, on optimizing and modeling in the 1990s and, eventually, on advanced models and solution methods in recent years.

Concerning the most important preliminary studies in the area of HLPs, O’Kelly had a key position to develop the first quadratic mathematical formulation for HLP (O’Kelly, 1987, 1992). Later, Campbell proposed multiple mathematical formulations for HLP to consider similar objective functions as several classical facility location problems (Campbell, 1994b, 1996). Moreover, Aykin and Klincewicz had also important roles in advancing the field (Aykin, 1994, 1995a, 1995b; Klincewicz, 1991, 1992). Recently, Campbell and O’Kelly (2012) have recently discussed the origins and motivations of HLP as well as some of the shortcomings in this field.
The first surveys of HLP were offered by Campbell (1994a), O’Kelly and Miller (1994). In addition, Klincewicz (1998) and Bryan and O’Kelly (1999) reviewed applications of HLP for the telecommunication and air transportation industries, respectively. More recently, Alumur and Kara (2008) surveyed and classified papers on HLPs published until 2007. In addition to these reviews, one can refer to Campbell, Ernst, and Krishnamoorthy (2002) and Farahani and Hekmatfar (2009) for fundamental definitions, classification, mathematical models, and solution methods of HLPs. Even though these studies have dealt with HLPs from various viewpoints, there have been some motivations for this review paper as follows:

- Over the past 5 years, many researchers have modeled and solved numerous variants of hub location problems in their research.
- Alumur and Kara (2008) put their emphasis on network hub location models and not the HLP from a general aspect.
- More particularly, Alumur and Kara (2008) did not consider any studies addressing continuous HLPs, in which a network will be replaced by a plane.
- Some new recent trends need to be taken into account such as the implementation of reliability, sustainability, global logistics, and multi-criteria decision making (MCDM) in HLPs.

Therefore, the significance of this review is to not only consider previously analyzed papers, but also shed light on some uncovered aspects of HLPs that have been published after 2007.

2. Mathematical models and classification of HLPs

In order to classify HLPs, some definitions and classifications will be given as follows:

- **Solution domain**: Network (the domain of candidate hub nodes is all of the network nodes), discrete (the domain of candidate hub nodes is a series of particular nodes) and continuous (the domain of hub nodes is a plane or a sphere).
- **Criterion**: Mini-Max (the maximum transportation cost from origin nodes to destination nodes is minimized) and Mini-Sum (the total cost incurred by locating hub nodes and allocation of non-hub nodes to hub nodes is minimized).
- **Source determining the number of hubs to locate**: Exogenous (the number of hubs to locate is primarily specified) and endogenous (the number of hubs to locate is not pre-specified but is determined as part of the solution).
- **The number of hub nodes**: single hub and multiple hubs.
- **Hub capacity**: uncapacitated (unlimited) and capacitated (limited).
- **The cost of locating hub nodes**: No cost, fixed cost, and variable cost.
- **The allocation of a non-hub node to hub nodes**: To one hub (single allocation) and to more than one hub (multiple allocation).
- **The cost of connecting non-hub nodes to hub nodes**: No cost, fixed cost, and variable cost.

By considering this classification, various mathematical models for hub network design can be formulated. The most common formulations, which have been widely applied by the literature, are introduced in this section.

2.1. Single-HLP

O’Kelly (1987) represented this problem in which the criterion is Mini-Sum, the solution domain is network, the non-hub nodes are connected to the hub node, the number of hub nodes to locate is defined exogenously and is equal to one, there is no cost for establishing the hub facility, the hub facility to locate has unlimited capacity, and the problem is a single allocation as only one hub is to be located.

The inputs of the problem are as follows: $h_{ij}$ is the amount of flow between nodes $i$ and $j$, $C_{ij}$ is the unit cost of transferring from non-hub node $i$ to hub node $j$. Furthermore, the outputs of the model are binary, in which $Y_{ij}$ is equal to 1 if node $i$ is allocated to a hub located at node $j$ (and 0, otherwise). If $Y_{ij}$ is equal to one, it means that node $j$ is allocated to itself and in fact it is a hub node. Considering this notation, the mathematical formulation of single-HLP is as follows:

$$\text{min} \sum_i \sum_j \sum_k h_{ik} (C_{ij} + C_{jki}) Y_{ij} Y_{kj}$$

Subject to

$$\sum_j Y_{ij} = 1$$

$$Y_{ij} - Y_{jj} \leq 0 \quad \forall i, j$$

$$Y_{ij} \in \{0, 1\} \quad \forall i, j$$

Eq. (1) minimizes the total transfer cost via the hub. Eq. (2) stipulates that there is only one hub. Eq. (3) stipulates that node $i$ can only be linked to a hub node at $j$. Eq. (4) defines decision variable type to be binary.

To linearize the objective function, if non-hub node $i$ is allocated to a hub at node $j$, we will have only one hub and all other non-hub nodes must be allocated to that hub as well. Therefore, Eq. (1) can be rewritten as follows:

$$\sum_i \sum_j \sum_k h_{ik} (C_{ij} + C_{jki}) Y_{ij} Y_{kj}$$

$$\quad \quad = \sum_i \sum_j C_{ij} Y_{ij} \left( \sum_k h_{ik} \right) + \sum_j \sum_i C_{jki} Y_{jj} \left( \sum_k h_{jk} \right)$$

$$\quad \quad = \sum_i \sum_j C_{ij} Y_{ij} (O_i + D_i)$$

where $O_i$ is the entire outgoing flow from node $i$, and $D_i$ is the entire incoming flow to node $i$. Furthermore, after replacing Eq. (1) with Eq. (5), the optimal single-HLP is presented, in which the Mini-Sum criterion should be minimized; meanwhile, the number of decision variables is of the order of $O(N^2)$.

2.2. p-HLP

In this problem, each non-hub node must be allocated to just one hub node and hence is considered as a single allocation $p$-hub location problem as, originally proposed by O’Kelly (1987). In this model, the criterion is Mini-Sum, the solution domain is network, the hub nodes are completely linked together, and every non-hub node is linked to a single hub node. The number of hubs to locate is defined exogenously and denoted by $p$, and at least one or at most two hub nodes have to be traversed for traveling between two non-hub nodes. Furthermore, there is no cost for establishing hubs, the hubs are uncapacitated, and the model outputs are binary.

In addition to all inputs introduced for the single-HLP, $x$ is the discount factor denoting economies of scale for transferring between hub nodes ($0 \leq x < 1$). To compute the transfer cost between hub nodes $C_{ij}$ is multiplied by $x$. This is due to the fact that the transportation cost between hub nodes is smaller than the transportation cost between hubs and non-hub nodes. Moreover, the output of the model is similar to that of the previous
model. Therefore, the criterion and model constraints for p-HLP are defined as follows:

$$\min \sum_{i} \sum_{k} C_{ik} Y_{ik} \left( \sum_{j} h_{ij} \right) + \sum_{k} \sum_{i} C_{ik} Y_{ik} \left( \sum_{j} h_{ij} \right) + \alpha \sum_{i} \sum_{j} \sum_{m} \sum_{k} h_{ij} C_{km} Y_{ik} Y_{jm}$$

(6)

Subject to

$$\sum_{j} Y_{ij} = 1 \quad \forall \ i$$

(7)

$$\sum_{j} Y_{ij} = P$$

(8)

$$Y_{ij} - Y_{jj} \leq 0 \quad \forall \ i,j$$

(9)

$$Y_{ij} \in \{0, 1\} \quad \forall \ i,j$$

(10)

Eq. (6) minimizes the total transportation cost between network nodes. The first term in Eq. (6) is the connection cost of all transportations from non-hub node i to hub node k when i is allocated to k. Moreover, the second term is the connection cost of all travels predetermined for non-hub node i to hub node k when non-hub node i is allocated to hub node k. In addition, the third term is the transportation cost of inter-hub flows. Furthermore, Eq. (7) stipulates that non-hub node i is allocated to precisely one hub node. Eq. (8) stipulates that exactly p hub nodes are selected. Eq. (9) enforces that node i is allocated to a hub node at j only if a hub is located at node j. Finally, Eq. (10) defines decision variables to be binary.

2.3. p-Hub median location problem (multiple allocation p-HLP)

While the criterion of the p-HLP is quadratic, Campbell (1991) proposed a linear mathematical formulation. The formulation of this problem is similar to the p-median formulation and is named p-hub median location problem. Because every non-hub node could be allocated to one hub node or more in p-hub median location problems, this model is named multiple allocation p-HLP. This model has similar assumptions to those of the p-HLP except that the allocation variables, here denoted by $Z_{km}^{lm}$, are assumed to be non-negative ($Z_{km}^{lm} \geq 0$), and non-hub nodes could be allocated to several hub nodes.

All inputs in p-HLP are also used here. Moreover, $C_{ik}^{lm}$ is defined as the unit transportation cost between start node i, end node j, and hub nodes k and m (note that this is in order i → k → m → j).

$$C_{ij}^{km} = C_{ik} + \alpha C_{km} + C_{mj}$$

(11)

The outputs of the problem are as follows: $X_{ik}$ which is 1 when a hub facility is located at node j (and 0, otherwise), and $Z_{km}^{lm}$ which is the flow from the origin node i to the destination node j via hub facilities located at nodes k and m. Therefore, the mathematical formulation of the multiple allocation p-Hub median location problem is defined as follows:

$$\min \sum_{i} \sum_{j} \sum_{m} \sum_{k} C_{km} h_{ij} Z_{km}^{lm}$$

(12)

Subject to

$$\sum_{k} X_{ik} = P$$

(13)

$$\sum_{k} \sum_{m} Z_{km}^{lm} = 1 \quad \forall \ i,j$$

(14)

$$Z_{km}^{lm} \leq X_{km} \quad \forall \ i,j,k,m$$

(15)

$$Z_{km}^{lm} \geq 0 \quad \forall \ i,j,k,m$$

(17)

$$X_{ik} \in \{0, 1\} \quad \forall \ k$$

(18)

Regarding the objective function, Eq. (12) minimizes the total transportation cost. Eq. (13) ensures that exactly p hubs are selected. Eq. (14) stipulates that each origin–destination pair (i, j) is allocated to one pair of hub nodes (k, m). Note that the origin–destination pair (i, j) could be allocated to a single hub facility as indices k and m could be the same. Eqs. (15) and (16) ensure that demand from origin node i to destination node j cannot be allocated to a hub pair (k, m) unless both nodes (k, m) are selected as hub facilities. Furthermore, Eqs. (17) and (18) define the decision variable types. Daskin (1995) reconsidered this formulation assuming single allocation non-hub nodes to hub nodes, known as the single allocation p-hub median location problem.

One of the main difficulties with this formulation is that the number of allocation decision variables ($Z_{km}^{lm}$) can be extremely large. As a matter of fact, in network hub location problems with every origin and destination node as a candidate hub node, there are variables of size $O(N^6)$. The size of such model grows rapidly as the number of network nodes is increasing, unless that the set of candidate hub nodes are reduced beforehand as in discrete location models (Hekmatfar & Pishvaee, 2009). The use of the quadratic formulation decreases the number of decision variables remarkably, however it does not always ensure a fast solution of the problem. Ernst and Krishnamoorthy (1998) introduced a 3-index formulation which has been used for developing other advanced models. Hamacher, Labbé, Nickel, and Sonneborn (2004) represented a tighter formulation for polyhedral multiple allocation p-HLP to allow larger problems to be solved in most applications of HLP, Marín, Cánovas, and Landete (2006) introduced new formulations for this problem that generalized the basic models with providing tighter LP bounds. Note that in case of using decomposition techniques (e.g., Lagrangian relaxation and benders’ decomposition), formulation resulting from Eqs. (12)–(18) would be a better choice as it provides tighter linear programming relaxation bounds.

2.4. p-HLP with fixed link cost

As non-hub nodes have to be allocated to a hub, a fixed cost can be considered for all connections of non-hub nodes to hub nodes. However, to connect a non-hub node to a hub node, a fixed cost must be considered. Campbell (1994b) suggested that the basic models can be extended with a fixed cost for connecting non-hub nodes to hub nodes. Problem inputs and outputs are similar to those of the p-hub median location problem model. Moreover, $g_{ik}$ is the fixed cost of connecting non-hub node i to a hub facility located at node k. In addition, $W_{ik}$ is binary variable denoting selection of link (i, k) if it is equal to one. The criterion of this problem is similar to the criterion of the p-hub median location problem as well as the cost term provided below:

$$\sum_{i} \sum_{k} g_{ik} W_{ik}$$

(19)

2.5. Minimum-value flow on links model

Instead of imposing that each non-hub node should be allocated to a single hub node, it is sometimes more applied to specify that the flow between any non-hub node and hub node must be greater than or equal to some minimum flow threshold value. Campbell (1994b) proposed this problem that is similar to p-HLP. This
problem can be analogously formulated as p-HLP with a minimum threshold for flow on each non-hub node to hub node link. The inputs of the problem are similar to those of the p-hub median location problem. Furthermore, $L_{q}$ is the minimum threshold value for flow between non-hub node $i$ and hub node $k$. The outputs of the problem are similar to those of the p-hub median location problem.

Additionally, the following equations should be added to the criterion and constraints of the p-hub median location problem:

$$Y_{ik} + Y_{jm} - 2z^{km}_{ij} \geq 0 \quad \forall \, i, j, k, m$$

(20)

$$\sum_{m} \sum_{j} h_{iy}z^{km}_{ij} + \sum_{s} \sum_{j} h_{sj}z^{sk}_{ik} \geq L_{q}Y_{ik} \quad \forall \, i, k$$

(21)

In Eq. (20), the demand from the origin node $i$ to the destination node $j$ could only be routed via hub facilities located at nodes $k$ and $m$ if: (1) non-hub node $i$ is allocated to a hub facility located at node $k$ and (2) non-hub node $j$ is allocated to a hub located at node $m$. On the other hand, Eq. (21) ensures the minimum flow thresholds on non-hub node to hub node connection links. Its first term computes total flow on a non-hub to a hub node link. The second term calculates total flow from every origin–hub node pair destined for a specific non-hub node through a specific hub node. The sum of these two terms should be more than the specified threshold if the link connecting the non-hub node to the hub node is established.

2.6. p-HLP with limited capacity

Campbell (1994b) proposed a problem in which capacities of hub facilities limit the flows in the network. In fact, the incoming and outgoing flows for each hub facility must be smaller than or equal to that hub facility capacity. The problem was formulated similar to p-hub median location problem with similar inputs and outputs to those of p-hub median location problem. In addition, $h_{iy}$ denotes the capacity of a hub facility located at node $k$.

Moreover, the following constraint should be considered along the original criterion and constraints of the p-hub median location problem:

$$\sum_{m} \sum_{i} \sum_{j} h_{iy}z^{km}_{ij} + \sum_{s} \sum_{i} \sum_{j} h_{sj}z^{sk}_{ik} \leq h_{iy}X_{k} \quad \forall \, k$$

(22)

In this constraint, the first and second terms denote the incoming and outgoing flows for the considered hub node, respectively.

2.7. Continuous p-HLP

HLPs are usually modeled as either a network or a discrete facility location problem. However, there are some studies considering a continuous solution domain. In continuous HLPs, the domain of hub nodes is not a series of particular nodes on a graph, but a plane or a sphere (Aykin, 1988, 1995b; Aykin & Brown, 1992; O’Kelly, 1986a, 1986b). In this problem, every non-hub node could be allocated to only one hub as in single allocation p-HLP. Special cases of this problem involving only one or two hub nodes were first considered by O’Kelly (1986a, 1986b). Afterwards, Aykin and Brown (1992) extended this model to consider $p$ hub nodes. In this model, the criterion is Mini-Sum, the solution domain is a plane and is continuous, the hub nodes are completely linked together, and every non-hub node is linked to only one hub facility. The number of hubs to locate is exogenous, and at least one or at most two hub nodes have to be traversed for traveling between two non-hub nodes. In addition, the fixed cost of opening hub facilities is not considered, the hubs are uncapacitated, and the decision variables are binary.

The inputs of the problem are similar to p-HLP, while $N_{i}$ the vector location of non-hub node $i$. Moreover, the outputs of the model are similar to p-HLP while decision variable $P_{k}$ denotes the vector location of hub node $k$ ($k = 1, \ldots, p$). Considering $d(a, b)$ as the Euclidian distance between two nodes $a$ and $b$, the criterion and model constraints are as follows:

$$\min \sum_{i} \sum_{j} \sum_{k} h_{iy}Y_{jm}(d(N_{i}, P_{k}) + \alpha d(P_{k}, P_{m}) + d(N_{j}, P_{m}))$$

(23)

Subject to

$$\sum_{j} Y_{ij} = 1 \quad \forall \, i$$

(24)

$$P_{k} = (a_{k}, b_{k}) \quad k = 1, \ldots, p$$

(25)

$$Y_{ij} \in \{0, 1\} \quad \forall \, i, j$$

(26)

Eq. (23) minimizes the total transportation cost in the hub network. The first term is the connection cost of all transportations originating from non-hub nodes to their respective hub nodes. Moreover, the second term is the transportation cost between hub nodes, while the third term is the transportation cost of all travels between hub nodes to non-hub node. Eq. (24) stipulates that every non-hub node $i$ is allocated to precisely one hub node. Eventually, Eqs. (25) and (26) define types of decision variables. In this model, the interaction between hub nodes are specified by non-hub nodes allocated to them. Hence, without the second cost term of Eq. (23), the problem reduces to a location–allocation (LA) problem.

2.8. Multi-objective p-HLP

Costa, Captivo, and Climaco (2009) proposed a multi-objective HLP in which the first objective minimizes the total transportation cost, while the second one minimizes the maximum time that the hub nodes take to process the flow (i.e., minimizes the maximum service time of the hub nodes). In this problem, each non-hub node is allocated to only one hub node like as in p-HLP. In this model, the criteria are Mini-Sum and Mini-Max, the solution domain is network, the hub nodes are completely linked together, and every non-hub node is linked to a single hub facility. The number of hubs to locate is exogenous, and at least one or at most two hub nodes have to be traversed for traveling between two non-hub nodes. In addition, the fixed costs to initiate service at hub nodes are not considered, the hubs are uncapacitated, and the decision variables are binary.

The inputs of the problem are similar to p-HLP in addition to $T_{ik}$, which is the time that hub node $k$ takes to process one unit of flow. Moreover, the outputs of the model are similar to p-HLP. The criterion and model constraints are as follows:

$$\min \sum_{i} \sum_{j} \sum_{k} h_{iy}Y_{jm}(C_{ik} + \alpha C_{km} + C_{jm})$$

(27)

$$\min \max_{k} \left\{ T_{ik} \sum_{j} h_{iy}Y_{jm} + \sum_{i} \sum_{j} h_{ij}Y_{jm}Y_{ik} \right\}$$

(28)

Subject to

$$\sum_{k} Y_{ik} = 1 \quad \forall \, i$$

(29)

$$\sum_{k} Y_{ik} = p$$

(30)

$$Y_{ik} - Y_{ik} \leq 0 \quad \forall \, i, k$$

(31)
\[
Y_k \in \{0, 1\} \quad \forall \ i, k
\]  
Eq. (27) considers the total transportation cost in the hub network. Eq. (28) minimizes the maximum service time that hub node \( k \) takes to process the total incoming and outgoing flow. Eq. (29) stipulates that each non-hub node \( i \) should be allocated to precisely one hub node. Eq. (30) enforces that \( p \) hub nodes are selected. Eq. (31) ensures that no non-hub node can be linked to a hub node until that hub node is selected. Moreover, Eq. (32) defines the decision variable type to be binary.

2.9. \( p \)-Hub center location problem

One of the most important variations of facility location problems is the \( p \)-center location problem. For example, this problem is useful for emergency facility location problem. For a problem similar to \( p \)-center, one can refer to the \( p \)-hub center location problem. Assume that there are origin–destination pairs in HLP, which could be represented as demand nodes in the \( p \)-center problem. A \( p \)-center hub location problem is defined based on the Mini-Max criterion in which the maximum cost of origin–destination pairs is minimized. Originally proposed by Campbell (1994b), \( p \)-Hub center location problem is applicable to situations involving perishable commodities in transportation networks. Notation of this problem is similar to \( p \)-hub median location problem except that the criterion is Mini-Max. Moreover, the criterion and model constraints are as below:

\[
\min \max_{i, k, m} \left\{ c_{ij} x_{ij} z_{km} \right\} 
\]  
Subject to
\[
\sum_{k} X_k = p \tag{34}
\]
\[
\sum_{k} \sum_{m} z_{ij}^{km} = 1 \quad \forall \ i, j \tag{35}
\]
\[
z_{ij}^{km} \leq X_k \quad \forall \ i, j, k, m \tag{36}
\]
\[
z_{ij}^{km} \leq X_m \quad \forall \ i, j, k, m \tag{37}
\]
\[
z_{ij}^{km} \geq 0 \quad \forall \ i, j, k, m \tag{38}
\]
\[
X_k \in \{0, 1\} \quad \forall \ k \tag{39}
\]

In Eq. (33), the maximum transportation cost from origin nodes to destination nodes is minimized. In addition, Eqs. (34)–(39) are alike those of the \( p \)-hub median location problem. It should be noted that a novel improvement of \( p \)-hub center location problems was proposed by Campbell et al. (2007). Ernst, Hamacher, Jiang, Krishna-moorthy, and Woeginger (2009) developed new mathematical formulations for \( p \)-hub center location problem and demonstrated the superiority of this formulation to the one by Kara and Tansel (2000) for single allocation \( p \)-hub center location problems. Finally, Yaman and Elloumi (2012) also proposed new formulations for star \( p \)-hub center location problem to minimize the length of the longest origin–destination path.

2.10. \( p \)-Hub covering location problem

\( p \)-Hub covering location problems are extensions of to the classical covering location problems. These problems tries to locate hub facilities in a way that the origin–destination pair of two non-hub nodes is covered by a pair of hub nodes. In fact, origin–destination pairs are covered only if there are hub facilities in pre-specified distances from their links. For this sake, the cost of transportation from start nodes to end nodes through the selected hub nodes must be greater than or equal to a pre-specified value.

\[
c_{ij}^{km} \leq \gamma_{ij} \tag{40}
\]

Originally proposed by Campbell (1994b), this problem could be further developed into two other HLPs, namely hub set covering location and hub maximal covering location problems, which would be discussed next. Kara and Tansel (2003) and Wagner (2008b) proposed new mathematical formulations for the single allocation \( p \)-hub covering location problem and the \( p \)-hub covering location problem with bounded path lengths, respectively.

2.11. Hub set covering location problem

Hub set covering location problem is a particular extension of hub covering location problems, which is formulated similar to the \( p \)-hub median location problem with an exceptions. Before solving the problem, the number of hub nodes to locate is not known. Hence, a fixed cost of establishing hub facilities is considered. The inputs of this problem are \( F_k \), the fixed cost of establishing a hub facility in node \( k \); \( c_{ij}^{km} \), transfer cost from the origin node \( i \) to the destination node \( j \) via hub nodes located at nodes \( k \) and \( m \); \( \gamma_{ij} \), maximum cost for covering links connecting demand nodes \( i \) to \( j \), and \( \eta_{ij}^{km} \) that is equal to one if hubs located at nodes \( k \) and \( m \) could cover demand pair \( (i, j) \). The outputs of the model are similar to those of the \( p \)-hub median location problem. Furthermore, the criterion and model constraints of hub set covering location problem are as below:

\[
\min \sum_{k} F_k X_k \tag{41}
\]

Subject to
\[
\sum_{k} \sum_{m} \sum_{i, j} V_{ij}^{km} = 1 \quad \forall \ i, j \tag{42}
\]
\[
Z_{ij}^{km} \leq X_k \quad \forall \ i, j, k, m \tag{43}
\]
\[
Z_{ij}^{km} \leq X_m \quad \forall \ i, j, k, m \tag{44}
\]
\[
Z_{ij}^{km} \geq 0 \quad \forall \ i, j, k, m \tag{45}
\]
\[
X_k \in \{0, 1\} \quad \forall \ k \tag{46}
\]

The total cost of opening new hub facilities is minimized by Eq. (41). Moreover, Eq. (42) stipulates that each demand pair is covered at least one time by a hub pair. The other constraints are similar to those of the previous problems.

2.12. \( p \)-Hub maximal covering location problem

Hub maximal covering location problem is a particular case of hub covering location problem. The proposed formulation for this problem is similar to the one for \( p \)-hub median location problem as the number of hub nodes to locate is determined exogenously. In addition, the fixed cost of establishing hub facilities is not regarded. The criterion and model constraints of this problem are as follows:

\[
\max \sum_{i} \sum_{j} \sum_{k} \sum_{m} h_{ij} V_{ij}^{km} \eta_{ij}^{km} \tag{47}
\]

Subject to
\[
\sum_{k} X_k = p \tag{48}
\]
\[ \sum_k \sum_m Z_{ij}^{km} = 1 \quad \forall \, i,j \] (49)

\[ Z_{ij}^{km} \leq X_k \quad \forall \, i,j,k,m \] (50)

\[ Z_{ij}^{km} \leq X_m \quad \forall \, i,j,k,m \] (51)

\[ Z_{ij}^{km} \geq 0 \quad \forall \, i,j,k,m \] (52)

\[ X_k \in \{0,1\} \quad \forall k \] (53)

Eq. (47) maximizes the amount of transportation demand covered. The other constraints of the problem are alike those of the p-hub median location problem.

### 2.13. HLP with star network structure

Suppose we have a set of nodes with a central hub node, and we intend to select \( p \) hub nodes such that hub nodes are connected to the central hub node with direct links. In addition, each non-hub node should be connected to a hub node. The resulting hub network structure will be a star/star network. Yaman (2008) proposed new formulations for star p-hub median location problem in which the total cost for selecting capacitated links was minimized. Yaman and Elloumi (2012) also developed formulations for the star p-hub median location problem with bounded path lengths in which the total routing cost was minimized subject to some upper bound constraints on the path lengths.

### 3. Solution approaches and algorithms for HLPs

A large variety of solution algorithms have been proposed to solve different types of HLPs during the past two decades. In this part, the relevant research is reviewed, and some of represented algorithms and solution approaches are introduced. Most of the HLPs are modeled as network location problems, but some studies have been done in discrete and continuous domains. Solution methods for this kind of problems are also represented in this section.

To introduce various kinds of HLPs properly, some notations are listed in Table 1. For example, when the hub nodes are capacitated, every non-hub node can be assigned to several hub nodes, and the problem is set covering, the notation will be C-MA-SC-HLP. If one of the categories of Table 1 is not applicable to any case, this part will be eliminated from the notation.

#### 3.1. Application of exact algorithms in HLPs

In this section, we represent studies in which exact optimization methods have been used for solving the problems. Table 2 shows this category.

According to Table 2, most of the papers on HLPs have considered uncapacitated cases, and also all articles with limited capacity of hub nodes are published in recent years. It reveals that solving capacitated HLPs by exact solution algorithms have attracted more consideration among researchers recently.

#### 3.2. Application of heuristic and meta-heuristic algorithms in HLPs

Even though integer programming optimization approaches are applied to solve small hub problems, larger instances of HLPs need to be solved by heuristic procedures or meta-heuristic procedures. As a matter of fact, while large-size instances can be dealt with specialized exact methods (e.g., benders decomposition and branch and price methods), development of meta-heuristics has helped many real-world applications, in which optimal/near-optimal solutions can even be obtained in less computational time. Some of relevant studies are summarized in Table 3.

Based on Table 3, we can see the trend of heuristic and meta-heuristic algorithms in HLPs similar to the exact solution algorithms. First, the majority of studies have dealt with uncapacitated cases of HLPs. Second, most of the capacitated HLPs have been investigated in recent years. Another observation is that hub median location problems have attracted more attention than hub covering location problems and hub center location problems in the HLP literature. In fact, most of the studies on hub covering location problems and hub center location problems are published in recent 3 years. Finally, this table shows that all investigations of continuous HLPs have assumed uncapacitated hub facilities with single allocation. In addition, these problems have barely attracted attention in recent years.

### 4. Applications and real-life case studies

In this section, we go through the HLP literature in terms of its applications as well as the relevant case studies. Table 4 classifies the literature based on applications, in which the industrial context of a specific area (for which the HLP concept has been applied) will be pointed out. Meanwhile, the literature will be categorized based on case studies, for which some real-world instances of HLPs will be mentioned along with their related applications.

According to Table 4, the most studied application areas in HLP are airlines and airport industries and transportation systems. The difference between these two areas is that the former has been primarily used in the literature, but the latter has attracted more attention in the recent 5 years (76% of studies on transportation networks that have been reviewed are published after 2007). Another observation from Table 4 is that supply chain management and logistics application area is a novel domain in HLPs that has been considered in recent years. Finally, we can find out from this table that most of the real-world case studies surveyed in the literature have been done in the recent 7 years.

### 5. Conclusions and future trends

In this paper, we shed light on basic classifications and important mathematical models and formulations for different variants of hub location problems. Then, we delivered a categorization of solution approaches and algorithms including exact methods as
5.1. Gaps in HLP trends of hub location problems. To open new insights for further research considering gaps and tions and case studies of HLPs were reviewed. This section intends well as heuristics and meta-heuristics. Afterwards, some applications and case studies of HLPs were reviewed. This section intends to open new insights for further research considering gaps and trends of hub location problems.

5.1. Gaps in HLP

- When a hub is installed, it may have some influences in the traffic that is generated, known as hubbing effect. Without considering effect of hubs on the generated traffic, there is no reason to take advantages of economies of scale. For instance, in airline networks, the effect of hubbing on passenger welfare is a critical issue for decision makers. As there is no research in this subject in the previous articles, this could be an important issue for decision makers. As there is no research in this subject in the previous articles, this could be an important issue for decision makers. As there is no research in this subject in the previous articles, this could be an important issue for decision makers. As there is no research in this subject in the previous articles, this could be an important issue for decision makers. As there is no research in this subject.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Article</th>
<th>Solution algorithm</th>
<th>Efficiency (# of nodes)</th>
<th># of hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-MA-p-HLP</td>
<td>Marín (2005)</td>
<td>Integer Linear Programming</td>
<td>40</td>
<td>–</td>
</tr>
<tr>
<td>C-SA-p-HLP</td>
<td>Costa et al. (2008)</td>
<td>Bi-criteria Linear Programming</td>
<td>40</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Correia, Nickel, and Saldanha-da-Gama (2010b)</td>
<td>Linear Programming</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Kratica, Milanovic, Stanimirovic, and Totic (2011)</td>
<td>Mixed Integer Programming</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>de Camargo and Miranda (2012)</td>
<td>Generalized Benders decompositio method</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Taghipourian et al. (2012)</td>
<td>Fuzzy Integer Linear Programming</td>
<td>20</td>
<td>4–7</td>
</tr>
<tr>
<td></td>
<td>Gelareh and Nickel (2007)</td>
<td>Bender decomposition method</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>de Camargo, Miranda, and Luna (2009)</td>
<td>Benders decomposition</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Contreras, Cordeau, and Laporte (2011c)</td>
<td>Enhanced Bender decomposition method</td>
<td>500</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Gelareh and Nickel (2011)</td>
<td>Bender decomposition method</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Vasconcelos, Nasis, and Lopes (2011)</td>
<td>Integer Programming</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Vidovic et al. (2011)</td>
<td>Mixed Integer Programming</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Lin (2010)</td>
<td>Integer Linear Programming</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Campbell (1994a, 1994b)</td>
<td>Integer Programming</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Campbell (2009)</td>
<td>Mixed Integer Programming</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>García, Landete, and Marín (2012)</td>
<td>Integer Programming-Branch and Cut</td>
<td>200</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>He, Chen, Chaowalitwongse, and Liu (2009)</td>
<td>Quadratic Programming</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Yaman (2009)</td>
<td>Mixed Integer Programming</td>
<td>81</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Puerto, Ramos, and Rodriguez-Chia (2011)</td>
<td>Mixed Integer Programming</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Yaman and Elloumi (2012)</td>
<td>Mixed Integer Programming</td>
<td>70</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Ernst et al. (2009)</td>
<td>Mixed Integer Programming</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

well as heuristics and meta-heuristics. Afterwards, some applications and case studies of HLPs were reviewed. This section intends to open new insights for further research considering gaps and trends of hub location problems.

Hub location with MCDM (e.g., Costa et al., 2008) has not been considered extensively in the previous studies. Most of the current literature has focused on maximizing profit or minimizing cost (or time and travel distance) criteria as traditional objective functions. However, in some applications like airline industries, we are dealing with other objective functions such as like maximization of market share. By considering more conflicting objectives, more real-world problems can be analyzed more efficiently.

In addition to finding optimal location and allocation schemes, there are other decisions that can be potentially considered in the hub network design phase to model a more realistic problem. Some examples include but not limited to the following:

- When liner shipping companies try to locate their favorite hub ports, at the same time, they need to consider competition between alternative hub ports in the region. Competitive HLP with application in maritime industries can be an example of considering other decision levels. An example of considering HLP in a competitive environment can be seen in Wagner (2008a).
<table>
<thead>
<tr>
<th>Problem</th>
<th>Article</th>
<th>Solution algorithm</th>
<th>Efficiency</th>
<th># of hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-SA-p-HLP</td>
<td>Ernst and Krishnamoorthy (1999)</td>
<td>Simulated annealing</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Yaman, Kara, and Tansel (2007)</td>
<td>Tabu search (with greedy algorithm) – branch and cut</td>
<td>49</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Chen (2008)</td>
<td>Heuristic method</td>
<td>200</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Randall (2008)</td>
<td>Ant colony optimization (ACO)</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Randall, Hendtlass, and Lewis (2009)</td>
<td>Extremal optimization (a meta-heuristic algorithm)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Lin and Lee (2010)</td>
<td>Lagrangian relaxation</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Contreras, Daz, and Fernandez (2011)</td>
<td>Branch and price – Lagrangian relaxation</td>
<td>200</td>
<td>–</td>
</tr>
<tr>
<td>C-MA-M-p-HLP</td>
<td>Lin, Lin, and Chen (2012)</td>
<td>Genetic algorithm</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C-SA-V-HLP</td>
<td>Mohammadi, Jolai, and Rostami (2011)</td>
<td>Imperialist competitive algorithm and genetic algorithm</td>
<td>70</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Klincewicz (1991)</td>
<td>Clustering heuristic</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Klincewicz (1992)</td>
<td>Tabu search and greedy random algorithm</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Klincewicz (1996)</td>
<td>Dual ascent and dual adjustment techniques within a branch and bound scheme</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Pamuk and Sepil (2001)</td>
<td>Tabu search (with greedy local algorithm)</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Marianov and Serra (2003)</td>
<td>Tabu search</td>
<td>50</td>
<td>4–5</td>
</tr>
<tr>
<td></td>
<td>Thomadsen and Larsen (2007)</td>
<td>Branch and price (combination of column generation and branch and bound)</td>
<td>25</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Wagner (2007)</td>
<td>Clustering (genetic algorithm and Tabu search)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Contreras et al. (2011a)</td>
<td>Monte Carlo simulation based algorithm/benders decomposition method</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Aykin and Brown (1992)</td>
<td>Location-allocation heuristics</td>
<td>Plane-Sphere</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>O’Kelly (1992)</td>
<td>Heuristic algorithm</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Aykin (1995a)</td>
<td>Simulated annealing (greedy interchange)/branch and bound</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Aykin (1995b)</td>
<td>Location-routing heuristic algorithms</td>
<td>Plane</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Abdinnour-Helm (1998)</td>
<td>Genetic algorithm and Tabu search (GATS)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Labbe, Yaman, and Gourdin (2005)</td>
<td>Branch and cut</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Topcuoglu, Corut, Ermis, and Yilmaz (2005)</td>
<td>Genetic algorithm</td>
<td>200</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Chen (2007)</td>
<td>Tabu search and simulated annealing</td>
<td>200</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Contreras, Fernandez, and Marin (2009)</td>
<td>Lagrangian relaxation</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Iwasa, Saito, and Matsui (2009)</td>
<td>Randomized approximation algorithms based on LP relaxation (nearest neighbor algorithms)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Han (2010)</td>
<td>Integer programming – Tabu search</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Catanzaro, Gourdin, Labbe, and Oszoy (2011)</td>
<td>Branch and cut</td>
<td>20</td>
<td>–</td>
</tr>
<tr>
<td>U-SA-2-HLP</td>
<td>Ostresh (1975)</td>
<td>Location-allocation (shortest route)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>O’Kelly (1986a, 1986b)</td>
<td>Nearest distance heuristic</td>
<td>Plane</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ernst and Krishnamoorthy (1998)</td>
<td>Heuristic methods</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Problem</th>
<th>Article</th>
<th>Solution algorithm</th>
<th>Efficiency</th>
<th># of hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-SA-M-p-HLP</td>
<td>Aykin (1994)</td>
<td>Branch and bound (Lagrangian relaxation)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U-MA-M-V-HLP</td>
<td>Qu and Weng (2009)</td>
<td>Evolutionary approach based on path relinking</td>
<td>82</td>
<td>4</td>
</tr>
<tr>
<td>U-MA-V-p-HLP</td>
<td>Karimi and Bashiri (2011)</td>
<td>Heuristic algorithms</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td>U-SA-V-p-HLP</td>
<td>Calik, Alumur, Kara, and Karasan (2009)</td>
<td>TABU search</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U-MA-T-p-HLP</td>
<td>Ernst, Hamacher, Jiang, Krishnamoorthy, and Woeginger (2002)</td>
<td>Innovative heuristic methods</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U-SA-T-p-HLP</td>
<td>Ernst et al. (2002)</td>
<td>Innovative heuristic methods</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Calik et al. (2009)</td>
<td>TABU search based heuristic</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Gavriliouk (2009)</td>
<td>Heuristic based on aggregation</td>
<td>1000</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Meyer, Ernst, and Krishnamoorthy (2009)</td>
<td>Branch and bound algorithm and heuristic algorithm based on ant colony optimization (ACO)</td>
<td>40</td>
<td>5</td>
</tr>
</tbody>
</table>
Considering dynamic aspect of lower level decisions (tactical and operational planning levels) in addition to HLP in airline industries in postal companies (for example, dynamic pricing and revenue management) can be another potential research area.
5.2. Future trends in HLP

Considering facility location problem with reliability in order to resist against unforeseen disasters (manmade or natural disasters) is a new trend. This can also be applied on HLP considering possible disasters and disruptions on hubs or on spokes. For modeling reliability in HLPs, one can refer to Kim and O’Kelly (2009) in which flows transmit between origins and destinations without failing for a particular time period. In addition, one can refer to Carello, Della Croce, Ghirardi, and Tadel (2004) and An, Zhang, and Zeng (2011) because of considering backup hub nodes.

Modeling sustainability and paying special consideration to social and environmental impacts in addition to classic economic aspect of HLP can be an applicable research area. Application of such problems in hazardous material handling and contamination of logistics systems (air, ground, marine and urban logistics) is evident. However, a major challenge in this area is quantifying social and environmental impacts in models.

Concerning reviewing the literature, it should be noted that studies have done on a wide range of hub location modeling and solution methods. However, some of the novel formulations and their solution techniques are not pertinent to real-world problems and are not really applicable formulations. For instance, dynamic hub location is a more realistic potential area as various studies show that after several years, the located hubs are no longer optimal locations due to various changes in initial data over time (Alder & Hashai, 2005). In dynamic hub location area, we can refer to two recent articles: Contreras, Cordeau, and Laporte (2011b) and Taghipourian, Mahdavi, Mahdavi-Amiri, and Makui (2012).

Considering aspect of global logistics and supply chain (like tariff, exchange rate, etc.) in global hub logistic network design is another application that can be an appropriate area for more investigation in the near future. For instance, we can refer the readers to Wang and Cheng (2010) and Ishfaq and Sox (2012).

Another new subject in the hub location area is the multimodal HLPs in which the HLP is considered from a network design perspective. In these types of networks not only the usual location–allocation problems such as hierarchical location problem and obnoxious facility location problem. To our best knowledge, the only two papers dealing with hierarchical hub networks are Yaman (2009) and Lin (2010).

Most of the models applied in HLPs literature are considering deterministic parameters, while stochastic and uncertain cases are more realistic. In fact, many attributes of HLPs such as demand and setup costs for the hubs have inherent uncertainty in real-world applications. The work of Alumur, Nickel, and Saldanha-da-Gama (2012) is a good instance for this category. Another papers dealing with uncertainty in HLP are Marianov and Serra (2003); Yang (2009); Sim, Lowe, and Thomas (2009), and Contreras, Cordeau, and Laporte (2011a). Thus, focusing on developing stochastic programming and robust optimization models and solution methods for HLPs is an important trend for future research.

Even though we discussed the necessity of heuristics and meta-heuristics when it came into solution methods, they might not be able to completely deal with large-size problem. In other words, there is a lack of solution approaches for solving large problems; therefore, researchers can use hybrid methods, especially hybrid meta-heuristics, to solve HLPs with large number of hubs. Moreover, exact methods like Benders’ decomposition, column generation, branch and cut, etc. are still important. Furthermore, HLP has applications in very large-scale systems that require large amount of investments. Hence, ignoring gaps of even 1% by using imprecise methods is not really acceptable.

Even if exact or heuristic optimization methods cannot solve large-size instances properly, some bounding techniques can be implemented to better analyze these methods. Therefore, development of some relaxation methods in integer programming (e.g. Lagrangian relaxation) in HLPs can still attract more interest.

References


