ABSTRACT
The p-hub median problems (pHMPs) are a well-researched topic within the fields of Operations Research and Industrial Engineering. These problems have been found to have a wide range of practical applications in various areas such as logistics, retailing, and Internet computing. These applications have made pHMPs an important area of study, leading to numerous research efforts aimed at solving different variations of the problem. This paper presents a simheuristic algorithm for solving the uncapacitated version of the pHMP with stochastic travel times. The proposed approach combines simulation with biased-randomized heuristics to generate high-quality solutions quickly. The proposed method is validated by testing it on huge benchmark instances, which include stochastic travel times. The results demonstrate the efficiency of the proposed approach for this particular problem variation. The simulation-optimization approach provides a promising solution to a practical problem that arises in many real-world applications.

1 INTRODUCTION
In logistics networks, various customers or companies send goods that need to be transported to multiple recipients. However, running a complete network with direct deliveries, as in the courier market, would result in high costs. Therefore, parcel service companies or truckload providers aim to save transportation costs by consolidating shipments in networks using hubs as transshipment and switching points. Unlike warehouses, hubs have low or no inventory. Due to consolidation, the costs for inter-hub movement are typically lower than those for movement between a hub and an origin or destination.

In hub systems, movements from source to destination typically pass through one or two hubs. As long as the cost of moving goods remains constant, no movement will pass through more than two hubs, as the hubs are fully interconnected (Boysen and Fliedner 2010; Campbell 1994).

On the one hand, defining hub locations is typically a strategic-level location problem, involving long-term planning that has a beneficial influence on route planning and affects the performance of networks over years. Route planning, on the other hand, is an operational task that depends on short-term information such as demand or traffic. However, only focusing on route planning or only the definition of hubs is
suboptimal because they are not interdependent and for example, some information on the routing, such as estimated tours, is known in advance and should influence the solution to the location problem.

Optimization of logistics networks has gained increased interest due to the variety of defined problems and their applications. One of the most intriguing network design problems is the hub location problem, which involves allocating nodes (spokes) to hubs and designing the associated network (Contreras and O’Kelly 2019). The interest in studying hub location problems has grown due to their impact on environmental and financial aspects, as well as their influence on the performance of logistic networks over time. Logistics network planning is often impacted by uncertainty. For example, travel time in a network can be influenced by factors such as traffic jams or weather conditions, which can lead to increased travel time. As a result, logistics networks are frequently affected by delays that have an impact on costs, transport carbon footprints, and delivery times.

Up to this point, many hub location problems have been analyzed without considering the uncertainty of problem parameters. However, according to a review by Khaleghi and Eydi (2022), it is crucial to consider the impact of uncertainty on problem parameters in future research. This paper aims to develop a framework for solving the hub location problem from a strategic perspective. The framework considers a set of potential hub locations and selects a predetermined number of locations based on resulting network routes and stochastic delays. The solution approach is based on the defining the pHMP as a facility location problem. Such approach has been defined by Benedito and Pedrosa (2019). However, Benedito and Pedrosa (2019) used their approach to solve the deterministic version of the problem. In this paper, the stochastic uncertainty is considered in the pHMP. Uncertainty is encountered in real-world problems. To handle the uncertainty, simulation is utilized in the solution approach. A simheuristic approach is used to solve the problem. In addition, a second algorithm is used to define the facility location problem from the pHMP. The work aims to define an approach to analyze the effect of stochastic travel times to minimize the long-term traveling costs in the networks.

Section 2 defines the p-hub median problem, which is the problem considered in this paper. Section 3 provides a review of previous work on this problem. Section 4 presents the framework for solving the problem and deduces the facility location problem which is used to develop a simheuristic algorithm with low run times. Sections 5 and 6 present the results and conclusions, respectively.

2 THE p-HUB MEDIAN PROBLEM

To the best of our knowledge, hub location problems were first introduced by Goldman (1969). The general objective of these optimization problems is to select hub locations and allocate routing for given delivery tasks, such that one or two hubs are used as stopovers based on a given objective function. There are two main variants of this problem: the single-allocation variant and the multi-allocation variant. Whereas in single-hub allocation each demand point is allocated to only one of the hubs, in multi-hub allocation problems each demand point can send and receive via more than one hub. Mathematically in the single-allocation variant, each branch is connected with an edge to exactly one hub, and the allocation cannot be changed for different tours.

Let $D_1$, $D_2$, and $D_3$ be demand points, and $H_1$ and $H_2$ be hubs. According to Figure 1, in the single-allocation variant $D_1$ might be connected to $H_1$, $D_2$ to $H_1$, and $D_3$ to $H_2$. Notice that, by definition, hubs are fully interconnected. The tour from $D_1$ to $D_3$ is then routed through $D_1 \rightarrow H_1 \rightarrow H_2 \rightarrow D_3$. If both demand points are connected to the same hub, only one hub is used as in $D_1 \rightarrow H_1 \rightarrow D_2$. For instance, in the multi-allocation variant, $D_1$ could be also connected to $H_2$ and use different hubs for any tour (see Figure 2).

Campbell (1994) has proposed several variants of the hub location problem. This paper will analyze the multi-allocation version of the p-hub median problem. Studying this problem aims to generate insights that can be transferred to other problem variants, as they are often extensions of the basic problem with different objective functions or additional constraints. The p-hub median problem considers the selection
of $p$ hubs, $Y = \{y_1, y_2, \ldots, y_p\}$, while accounting for routing costs based on the given demand. It imposes the constraint that no more than $p$ hubs can be used.

Let $B$ be the set of origins of demand and $H$ the set of destinations. Let $x_{ijkm}$ be a binary variable that is equal to 1 if the routing from $i \in B$ to $j \in B$ is directed through hubs $k \in H$ and $m \in H$, and 0 otherwise. Let $y_k$ be a binary variable that is equal to 1 if a hub is located at node $k$, and 0 otherwise. For a better understanding of the problem, the formulation of the integer program as described in Campbell (1994) is provided below:

$$\min \sum_i \sum_j \sum_k \sum_m w_{ij} x_{ijkm} c_{ijkm}$$ (1)

$$\text{s.t.} \quad \sum_k y_k = p$$ (2)

$$\sum_k \sum_m x_{ijkm} = 1, \quad \forall i, j$$ (3)

$$x_{ijkm} \leq y_k, \quad \forall i, j, k, m$$ (4)

$$x_{ijkm} \leq y_m, \quad \forall i, j, k, m$$ (5)

$$x_{ijkm} \in \{0, 1\}, \quad \forall i, j, k, m$$ (6)
\[ y_k \in \{0, 1\}, \quad \forall k \]  

(7)

where \( w_{ij} \) is the volume of demand from location \( i \) to \( j \). The cost \( c_{ijkm} \) for one tour \( i \rightarrow k \rightarrow m \rightarrow j \) is defined by:

\[ c_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}, \]

where \( \alpha \) is a discount factor resulting from consolidation and reduced costs for hub interconnection. The objective function (1) minimizes transportation costs from all origins to all destinations. Constraint (2) ensures that exactly \( p \) hubs are located. Constraints (3) guarantee exactly one routing for every origin-destination pair. Constraints (4) and (5) restrict tours \( x_{ijkm} \) to use only open hubs. Finally, constraints (6) and (7) define the binary variables \( x_{ijkm} \) and \( y_j \). This formulation considers the multi-allocation variant because any tour can be routed individually by \( x \). In the single-allocation variant \( x \) would only have one branch and one hub as index.

The p-hub median problem is NP-hard. Even if the locations of the hubs are fixed, the allocation part of the problem remains NP-hard as shown by (Kara 1999).

For the stochastic version of the problem, the cost parameters \( c_{ik}, c_{km}, \) and \( c_{mj} \) include uncertainty, because of traffic or construction sites on the route. To also solve huge instances and include these uncertainties, instead of an exact formulation which is mentioned above, a simheuristic for the problem is developed.

3 RELATED WORK

In the literature on hub location problems, different problem definitions and approaches are found for single-hub allocation (O’Kelly 1987; Campbell 1994; Ernst and Krishnamoorthy 1996; O’Kelly et al. 1996; Skorin-Kapov et al. 1996) and multi-hub allocation (Campbell 1992; Skorin-Kapov et al. 1996; Ernst and Krishnamoorthy 1998). The problem parameters could consider an infinite hub capacity (uncapacitated) or specify a capacity for hubs (capacitated). Fixed or variable costs could be counted in the problem parameters by the objective function or constraints. The number of hubs could be pre-specified or a part of the decision made according to a proposed solution. In addition, researchers studied the hub location problem from different perspectives, such as the minimization of maximum travel time or the minimization of costs. Other problems consider multiple objectives, in which an economic dimension forms one of the objectives in addition to another dimension, such as environmental (Pourghader Chobar et al. 2021) or transportation time (Mohammadi et al. 2019).

The multi-allocation hub problem is one of the widely studied problems. Due to the p-hub median location problem, a node (spoke) in a network could be allocated to a hub (Abdinnour-Helm 2001), and a hub could be connected to several spokes and hubs are interconnected, this problem is challenging due to the resulting extremely large search space associated with it (Farahani et al. 2013).

Several approaches have been used to recommend solutions for pHMPs. For example, Abdinnour-Helm (2001) used the simulated annealing metaheuristics, while Grine et al. (2021) employed the artificial immune system algorithm to solve a single allocation pHMP. Campbell (1994) presents four integer programming formulations for different types of discrete hub location problems: p-hub center problems, uncapacitated hub location problems, hub covering problems, and the p-hub median problem. Other approaches have been used to solve the aforementioned variants. The most popular methods are exact algorithms, heuristics, and metaheuristics (Farahani et al. 2013). Most studies with exact algorithms only solve instances with a small number of nodes within a reasonable time. Therefore heuristic algorithms are needed to solve more realistic larger instances.

Furthermore, in recent years, there has been also an increasing interest in incorporating uncertainty and risk into the hub location problem. Recent studies address uncertainty as one of the influencing and critical factors in the designing of networks. This has led to papers of robust and stochastic optimization approaches that take into account demand fluctuations, supply chain disruptions, varying travel times, and other sources of uncertainty.
For example, Ghaffarinasab (2018) proposed simulated annealing to solve the problem with uncertain demand. They studied three cases of demand uncertainty: hose, hybrid, and budget demand uncertainty. Ghaderi and Rahmanian (2016) pointed out that different sources of uncertainty cause delays at the hub. They addressed the affecting the activity runs in hubs and considered stochastic demand and travel time in a single allocation pHMP. By Jost and Clausen (2023) an approximation algorithm for the $p$-hub center problems is established. They reduce it to the $k$-center problem.

Benedito and Pedrosa (2019) introduces a first constant-factor approximation algorithm for the single-allocation median hub location problem. The algorithms presented in this study consider several variants, depending on whether the input specifies the number of open hubs or if a single hub must be assigned to each client. They propose a 2.48-approximation algorithm for the problem using an LP-rounding technique and taking advantage of the problem’s symmetries into account. The approach involves a new formulation that enhances the accuracy of the algorithm. They reduce the problem to the facility location problem. They furthermore developed a first approximation algorithm for the multi-allocation variant, which is a basis for the simheuristic.

4 SOLUTION APPROACH

The proposed solution approach for solving the p-hub median problem is based on solving a facility location problem (FLP), as shown in Figure 3. In this approach, it is decided which $p$ hubs to open. The difference between FLP and HLP is, that the $\alpha c_{km}$ is not considered in FLP. Hence, the task is not to route between different locations but to connect customers to facilities. Therefore, the decision variable $x$ only consists of two indices in FLP, and (3), (4) are replaced by

$$x_{i,k} \leq y_k.$$ (8)

In the context of a pHMP, these facilities are hubs that need to be allocated with respect to other nodes in the network. For the formulation of the FLP, several approaches are found, such as the one defined by Benedito and Pedrosa (2019). This algorithm reduces the pHMP to a $k$-FLP. After obtaining the hubs, the algorithm solves the routing problem. The algorithm of Benedito and Pedrosa (2019) uses branches as cities and hubs as facilities. Hence, the algorithm tries to minimize the summed distance between branches and hubs, since it does not consider hub-to-hub distances. The algorithm of Jost (2023) does the same with a different interpretation of the distance function. And the algorithm takes into account that hubs in the direction of the destination should be preferred. This different interpretation of the distance shows how little adjustments in the formulation of the algorithm even for approximation algorithms affect better results in reality. Both can be proven as approximation algorithms, which has also a high value for solving complex instances in reality and also provide good results in reality. The algorithms are described as Algorithm 1.

Algorithm 1 Benedito and Pedrosa / Jost

<table>
<thead>
<tr>
<th>Interpret HLP instance as FLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solve FLP and obtain facilities due to algorithm 2</td>
</tr>
<tr>
<td>open hubs corresponding to FLP solution</td>
</tr>
<tr>
<td>Solve routing optimally</td>
</tr>
</tbody>
</table>

The facility location problem is solved in this paper using a constructive heuristic (Algorithm 2). In each iteration, a facility is added to the solution. The selection of facilities continues until all required facilities are added. Thus, the best candidate facility that minimizes the total costs is selected in each iteration.

That means, that in each iteration the values

$$c_f = \sum_{i} \min_{k \in F \cup f} x_{ik} \cdot c_{ik}$$

for different choices of $f$ are compared.
In this context, the costs are the summed deterministic travel times as in the objective of the FLP. In each iteration the available facility locations are ranked according to their cost contribution and the lowest is selected. The greedy approach selects the best facility location as the location with the lowest cost contribution.

**Algorithm 2** randomized FLP algorithm

1. Initialize an empty set `F`.
2. **while** further facility can be opened **do**
   - Compute the solution cost $c_f$ for choosing $F \cup f$ as facilities for any non-opened facility $f$.
   - Sort $c_f$.
   - Draw $i$ randomly by a geometric distribution.
   - Let $\hat{f}$ be the facility for which $c_f$ is $i$ smallest value.
   - $F \leftarrow F \cup \hat{f}$ ▶ open facility
3. **end while**
4. Return $F$ as a set of open facilities.

The result from the greedy approach is the same solution each time it is executed. The selection probability is assigned as 1 to the candidate facility at the top of the ranked candidate list of facilities. To examine different solutions following the same construction logic, biased randomization is integrated into
the constructive heuristic. This biased randomization strategy is described in Grasas et al. (2017). Thus, the facility in each location is selected randomly with a higher selection probability assigned to the facility with the lowest cost contribution. In this approach, not only does one facility get the probability of being selected, but each of them gets a selection probability depending on their position in the ranked list of candidates. The selection probability follows a skewed probability distribution, e.g., a geometric one. As a result, different solutions are defined based on this approach from each run of the constructive heuristic.

The generated solutions consist of \( p \) hubs (facilities) and vary in the hub locations. In the second step, solutions are evaluated (Figure 3). The stochastic travel time over the routes is considered. As a result, a single solution has different objective function values in each evaluation run. For evaluation purposes, Monte Carlo simulation (MCS) is used, and the objective function in each run is recorded. The variation between the objective function values results from the stochastic travel time along the routes in the network. The solutions are ranked with respect to their stochastic objective function values, and the most promising solution is identified. In addition, risk analysis could be conducted considering the variation in objective function values of solutions.

Due to the biased-randomized strategy, running this algorithm multiple times gives alternative solutions. Notice that the routing can be computed in multi-allocation for a known set of opened hubs. This can easily be performed by testing any pair of open hubs for each tour. This is only possible since in the multi-allocation variant the routings are independent. This procedure can be accelerated by creating a sorted list of open hubs for any branch, such that far-away hubs do not need to be considered.

The routing adapts if edge weights are modified. Now, some solutions of this algorithm are tested on instances that are modified by a delay. In the stochastic version, hubs should be decided in the first step. In the second step, the distance function \( d \) is influenced randomly. This can be interpreted as traffic or construction sites on the connection. In the third step, the routing is decided.

5 RESULTS AND DISCUSSION

Our experiment consists of running the algorithm of Benedito and Pedrosa (2019) and one by Jost. These algorithms form the logic base to define the FLP from the pHMP. Thus, the definition of the FLP in Figure 3 is done according to one of the algorithms. The defined FLP is solved using the randomized FLP algorithm (Algorithm 2). In the experiments, the number of potential hubs is 50, the number of hubs to be opened is 10, the number of branches is 200 with 2000 flows, and \( \alpha = 0.8 \). All instances can be downloaded from https://github.com/aleksandra-gro/Dataset.git (Jost et al. 2023).

Our contribution is using the biased randomized approach (Algorithm 2) for solving the FLP inside the algorithms. The parameter of the geometric distribution was varied as shown in Table 1.

Table 1: The median of travel time for 100 iterations of pHMP after defining the FLP based on two different algorithms.

<table>
<thead>
<tr>
<th>Geometric parameter</th>
<th>Benedito and Pedrosa (2019)</th>
<th>Jost Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>1691.78</td>
<td>1659.54</td>
</tr>
<tr>
<td>0.5</td>
<td>1689.20</td>
<td>1657.34</td>
</tr>
<tr>
<td>0.6</td>
<td>1687.86</td>
<td>1655.49</td>
</tr>
<tr>
<td>0.7</td>
<td>1690.04</td>
<td>1657.29</td>
</tr>
<tr>
<td>0.8</td>
<td>1689.05</td>
<td>1655.92</td>
</tr>
<tr>
<td>0.9</td>
<td>1688.92</td>
<td>1653.93</td>
</tr>
</tbody>
</table>

A small difference is noticed corresponding to the geometric distribution parameter changing (Table 1). This notice verifies that each problem has its best parameter set for the biased randomization approach, which varies between problems. Both algorithms differ in the facility location problem definition from the pHMP; hence, different problems are defined. The best parameter for the Benedito and Pedrosa (2019) algorithm is 0.6, while for the Jost Algorithm is 0.9.
The considered pHMP is stochastic in terms of the travel time between the nodes in the network. Thus, stochastic travel time delays are added. MCS is used to evaluate each generated solution. The median of MCS runs is displayed in Table 2. In Jost algorithm, the found solutions are much better with respect to the travel time in the network.

In order to show the impact of the solution approach, a comparison between the deterministic solution (Det) and stochastic solutions (Stoch) is required. The deterministic solutions based on Benedito and Pedrosa (2019) algorithm and Jost algorithm are 1074.79 and 1060.19, respectively. The travel times associated with these solutions, selected hubs locations, under stochastic conditions (DetStoch) are 1688.41 and 1654.35, respectively. The DetStoch represents the impact of the solution found for the deterministic version of the problem under uncertainty. In finding the deterministic solution, possible uncertainty in the problem is not considered, and the evaluation of this solution under stochastic uncertainty results in DetStoch. On the contrary, the stochastic uncertainty influences the search for the Stoch solution. These values are tabulated in Table 2 and displayed in Figure 4. It is clear that the deterministic solution results in significantly different travel times under stochastic conditions, since the stochastic solution takes into account the presented variability and uncertainty. On the contrary, the deterministic solutions assumes deterministic conditions.

Figure 4: Gaps (in % with respect to the solution provided by the Jost algorithm for the deterministic version.
Table 2: Comparison between different solutions impact, in which the gap is calculated with respect to the Jost algorithm \( \text{Det} \).

<table>
<thead>
<tr>
<th>Solution</th>
<th>Median value</th>
<th>Gap (in %) w.r.t. Jost ( \text{Det} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benedito and Pedrosa (2019) ( \text{Stoch} )</td>
<td>1687.86</td>
<td>59.20</td>
</tr>
<tr>
<td>Jost algorithm ( \text{Stoch} )</td>
<td>1653.93</td>
<td>56.00</td>
</tr>
<tr>
<td>Benedito and Pedrosa (2019) ( \text{Det} )</td>
<td>1074.79</td>
<td>1.38</td>
</tr>
<tr>
<td>Jost algorithm ( \text{Det} )</td>
<td>1060.19</td>
<td>0</td>
</tr>
<tr>
<td>Benedito and Pedrosa (2019) ( \text{Det}_{\text{Stoch}} )</td>
<td>1688.41</td>
<td>59.26</td>
</tr>
<tr>
<td>Jost algorithm ( \text{Det}_{\text{Stoch}} )</td>
<td>1654.35</td>
<td>56.04</td>
</tr>
</tbody>
</table>

6 CONCLUSION

In this paper, a novel approach have been proposed for solving the stochastic version of the pHMP, which is a challenging problem in the field of network optimization. The proposed approach combines Monte Carlo simulation with a biased-randomized algorithm. This algorithm hybridizes random sampling from a geometric probability distribution with a greedy selection strategy to identify the best hubs for the network. The algorithm have been tested on a dataset, varying the problem parameters and comparing its performance against other state-of-the-art approaches from the literature.

Our experimental results demonstrate that the proposed approach is capable of providing better solutions. Also, this paper shows how simulation can be useful to address stochastic versions of this problem. Overall, the findings suggest that the biased-randomized algorithm, when combined with Monte Carlo simulation, is a promising approach for solving stochastic versions of the pHMP, and can be used as a practical tool for optimizing network designs in various applications such as transportation, logistics, and telecommunication.

Future research directions could include extending the proposed approach to handle more complex network topology and incorporating other optimization criteria such as reliability and resilience.

REFERENCES


AUTHOR BIOGRAPHIES
NIKLAS JOST is a researcher at the Institute of Transport Logistics TU Dortmund University. He holds a M.Sc. in computer science from Bonn University. His research interests include location problems, approximation algorithms, and heuristics. His e-mail address is niklas.jost@tu-dortmund.de.
MAJSA AMMOURIOVA is a researcher at a postdoctoral position in the Computer Science Department at the Universitat Oberta de Catalunya. She earned her Ph.D. degree from TU Dortmund University in 2021. Her research interests are x-heuristics which includes simheuristics. Her email address is mammouriova@uoc.edu.
ALEKSANDRA GROCHALA is a Student Assistant at the Institute of Transport Logistics TU Dortmund University. She holds a B.Sc. in Business Mathematics from TU Dortmund University. Her research interests include facility layout, location, and knapsack problems. Her email address is aleksandra.grochala@tu-dortmund.de.
ANGEL A. JUAN is a Full Professor in the Dept. of Applied Statistics and Operations Research at the Universitat Politècnica de València, as well as an Invited Professor at University College Dublin and Euncet Business School. Dr. Juan holds a Ph.D. in Industrial Engineering and an M.Sc. in Mathematics. He completed a predoctoral internship at Harvard University and postdoctoral internships at the Massachusetts Institute of Technology and the Georgia Institute of Technology. His main research interests include applications of simheuristics and learnheuristics in computational logistics and finance. He has published over 150 articles in JCR-indexed journals and more than 325 papers indexed in Scopus. His website address is https://ajuang.upv.es and his email address is ajuanp@upv.es.
CHRISTIN SCHUMACHER is the chief engineer at the Institute of Transport Logistics at TU Dortmund University. She defended her Ph.D. in Computer Science with the topic of adaptive machine scheduling of a practice-oriented hybrid flow shop in 2022 and holds a M.Sc. in Business Mathematics. Her research interests include mathematical optimization, heuristics, and discrete event simulation for scheduling and other real-world optimization problems. Her e-mail address is christin.schumacher@tu-dortmund.de. Her website is https://itl.mib.tu-dortmund.de/en/institute/team/dr-ing-christin-schumacher/.