A Submodular Representation for Hub Network Design Problems with Profits and Single Assignments

MARYAM HEMMATI
Gina Cody School of Engineering and Computer Science, Concordia University

[Under Review June 2024]

Hub Location Problems (HLPs) lie at the heart of network design planning in transportation and telecommunication systems. They constitute a challenging class of optimization problems that focus on where to open hub facilities and the designing of the hub networks. In this work, we study a class of hub location problems, named hub location problem with single allocation, which allows each node to be assigned only to one hub facility. This paper presents the use of combinatorial optimization and greedy heuristics algorithms to solve hub location problems with single allocation. By proposing a submodular objective function, we present an approximation algorithm working in a polynomial amount of time. We prove the output of the algorithm shows the worst-case bounds on the quality of the approximations.

Keywords: Submodularity; Hub; Combinatorial Optimization; Single Assignment

1. Introduction

Hub-and-spoke networks optimize flow routing between multiple origins and destinations by using hub facilities for transshipment, consolidation, and sorting. This approach reduces setup costs and improves efficiency, leading to economies of scale. HLPs, being $\mathcal{NP}$-hard, involve two key decisions: locating hubs and designing the network. This work focuses on discrete location decisions, assuming a predefined set of potential hub sites typically modeled using graphs. The primary features of HLPs include demand associated with flows between origin/destination (O/D) pairs, hub facilities acting as transshipment points, cost or service-based objectives, and cost benefits from economies of scale represented by a discount factor α $(0 \leq \alpha \leq 1)$.

Interest in studying Hub Location Problems (HLPs) has grown since the seminal work of (1) on continuous space models. Subsequent significant contributions include (2) with a quadratic integer model, and (3) with the first mixed-integer linear programming formulation for HLPs.

Most literature on HLPs focuses on discrete space, with models such as the p-hub median problem (4; 5; 6), the p-hub center problem (3; 7; 8), and hub covering problems (9; 10). These problems are further classified into capacitated and uncapacitated problems, considering capacity limitations on hub nodes (11) or hub arcs (12; 13; 14).

There are two main allocation strategies for assigning non-hub nodes to hubs. ”Multiple Assignments” select hubs first and then determine the shortest path for each O/D pair, which can simplify routing but increase network design costs due to numerous access links. This strategy suits scenarios with no setup costs for access links, such as air freight. ”Single Assignments” allocate each non-hub node to exactly one hub, which can be more efficient for applications like telecommunications.

Here we are interested in studying the Hub Network Design Problem with Profits and Single Assignments (HNDPSA). The objective of HNDPSA is to maximize total profit by measuring the trade-off between the revenue from serving a set of commodities and the total cost of designing the network. The strategic decisions consist of (i) determining the nodes to open as hubs and which hub arcs should be activated, and (ii) how to assign the nodes to the opened hub nodes with a single allocation strategy (i.e., which path should be chosen to route a commodity) to maximize the objective function.

© .

The uncapacitated single allocation hub location problem (USAHLP) was first modelled by (15) as a quadratic integer programming formulation. (3) introduced the first mixed-integer linear program for the single allocation p-hub median problem (SAPHMP), aiming to minimize total flow costs with a given number of hubs. Path-based mixed-integer linear programming formulations for USAHLP were also proposed by (16) and (17). (11) studied the capacitated single allocation hub location problem (CSAHLP).

Due to the combinatorial nature of single allocation HLPs (SAHLPs), heuristic algorithms are widely used to solve larger instances. (18) developed a genetic algorithm for the single allocation p-hub center problem (SAPHCP), while (19) applied scatter search to the SAPHMP. (20) developed a tabu search algorithm for SAPHMP.

We are looking for a combinatorial optimization model with a submodular objective function for modelling the HNDPSA. The concept of submodularity is essential for several optimization problems, including HNDPSA.

In what follows, we denote the discrete derivative via $\rho_j(S) = z(S \cup \{j\}) - z(S)$ for $j \in N$, which is also known as the incremental value for a submodular function. nemhauser1978analysis.

Definition 1 *(see 21) Let N be a finite set and $z : 2^N \to \mathbb{R}$ be a real-valued function.*
(a) z is submodular if $\rho_i(S) = z(S \cup \{i\}) - z(S) \geq z(T \cup \{i\}) - z(T) = \rho_i(T), \quad \forall S \subseteq T \subseteq N,\ i \in N \setminus T$
(b) z is decreasing if $\rho_i(S) \geq 0, \quad \forall S \subseteq N,\ i \in N$

Part (a) shows that the increments of adding a single element to a smaller subset are more than adding the same element to a larger set. (22) provided a polynomial-time algorithm to minimize a submodular function, whereas maximizing a submodular function is in the class of $\mathcal{NP}$-hard problems.

Maximizing submodular functions gained attention with the work of (21), which utilized greedy heuristics and LP-relaxations to approximate an optimal solution for the k-median location problem. They demonstrated a worst-case performance result for a specific greedy algorithm for a nondecreasing submodular function z, achieving a sharper bound when $z(\emptyset) = 0$. For a set of facilities N and customers I, with a non-negative matrix $C_{|I|\times|N|} = (c_{ij})$, the objective function is $z(S) = \sum_{i\in I} \max_{j\in S} c_{ij}$ for $S \subseteq N$, and the goal is to maximize $z(S)$ subject to $|S| \leq K$. The proposed greedy heuristic iterates to approximate the value of z and provides a bound, ensuring that the value of the greedy approximation is at least $1 - [(K-1)/K]^K$ times the optimal value. This bound can be achieved for each K and has a limiting value of $(e-1)/e$, where e is the base of the natural logarithm.

In the context of Hub Location Problems (HLPs), (23) and (24) developed methods addressing submodular function constraints. (25) and (26) provided combinatorial representations and exact algorithms for multi-level facility location problems satisfying submodularity. Notably, the submodularity of the Multi-level Uncapacitated Facility Location Problem (MUFLP) was clarified by (27).

In this research, we propose a submodular representation for modelling the HNDPSA. We then propose a greedy heuristic algorithm to approximate the optimal solution value. We focus on the mathematical insights and prove the proposed algorithm runs in polynomial time. We show the quality of the approximation achieved by the designed algorithm using theoretical aspects.

2. Problem Representation

In this chapter, we first present fundamental definitions and assumptions that we use throughout our study of HNDPSA. Then, we propose three different combinatorial representations for HNDPSAs. For the first two representations, we provide counterexamples to disprove the submodularity of the objective functions. For the last one, we directly prove that the objective function is submodular.

2.1. *Fundamentals*

Let $G = (N,E)$ be a complete undirected graph without loops, where N is the set of nodes and E is the set of edges. We define K as a set of commodities, where each commodity $k \in K$ is defined as a pair of nodes (i.e., k = $\{i, j\}$), and a positive amount of flow W_k that needs to be routed between the endpoints of commodity k. Also, let $d_{ij} \geq 0$ be the distance or transportation cost, from node i to node j. From here on, we will assume that d is a distance function, and therefore is non-negative, symmetric ($d_{ij} = d_{ji}\ \forall i, j \in N$), and satisfies the triangle inequality.

Moreover, for each node $i \in N$, we define f_i to be a fixed set-up cost that is paid for installing a hub at node i. Similarly, for each $e \in E$, g_e denotes the fixed set-up cost of activating a hub edge. Hub edges $e = \{m,n\} \in E$ have a unit flow cost of αd_{mn}. Recall that $0 \leq \alpha \leq 1$ represents a discount factor to account for economies of scale. We should note that flows have directions, but we work over an undirected graph.

For a given path p of G, let l_p be the number of nodes in p. From now on, we refer to l_p as the length of path p. Also, the r-length path is a path of length r. For each commodity $k \in K$, let $p_k := i \longrightarrow m_1 \longrightarrow ... \longrightarrow m_{l_{p_k}-2} \longrightarrow j$ be a directed path with origin i and destination j and passing through hub nodes $m_1, m_2, ...m_{l_p-2}$. In this study, we focus on a set of paths, denoted as P, that have the following characteristics. Each path is divided into three legs: a collection leg, a transfer leg and a distribution leg. The collection and transfer legs may contain either no edges or exactly one access edge. The transfer leg is composed only of hub nodes and may contain a combination of hub and bridge edges with at most $l_p - 2$ edges between the first and the last hub nodes.

Figure 1 shows all possible path configurations for routing commodity k. Black squares represent the set of hub nodes and black circles correspond to the set of non-hub nodes.

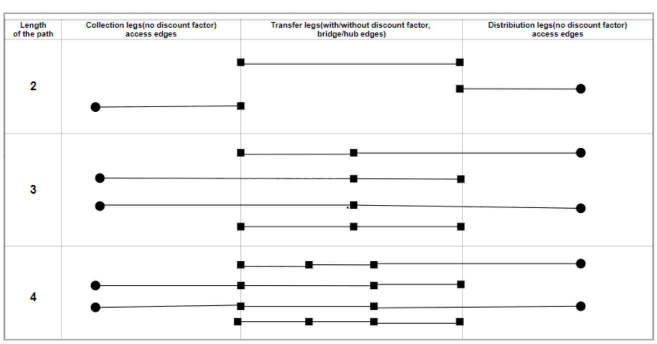


FIG. 1. The set of possible paths for routing a commodity

In a feasible solution of the HNDPSA, besides the set of hub nodes, we have a set of access edges A, a set of hub edges H, and a set of bridge edges B. Recall that a bridge edge is an edge between two hub nodes without benefiting from a discount factor α. For commodity k, let C_{p_k} be the unit transportation cost for routing k through p_k, and T_k be the total revenue received from serving the demand of commodity k. For routing commodity k through p_k, we define the profit as $h_{p_k} = \left(T_k - W_k C_{p_k}\right)^+$, where $(x)^+ = \max\{x, 0\}$. We assume that commodities cannot be split through different paths. Therefore, not every commodity must be served.

Now, we propose a combinatorial representation for the HNDPSA which is called path-based representation. We prove the submodularity of the corresponding objective function using Definition 2 and Lemma 1. The set of constraints includes the single assignment rule and opening at most p hub nodes and q hub edges ($q \leq p(p-1)/2$).

2.2. *A path-based representation*

In this section, we propose path-based representation for the HNDPSA. Consider the graph $G = (N, E)$. Let P be the set of feasible paths in G, excluding the paths with edges between non-hub nodes. Let $S \subseteq N$ be a set of hub nodes, $S' \subseteq N$ be a set of activated non-hub nodes, $H \subseteq E$ be a set of hub edges, $A \subseteq E$ be a set of access edges, and $R \subseteq P$ be a set of paths. Also, let $U = N \cup E \cup P$ be a finite set that defines the ground set such that any subset of U is of the form of $(S, S', H, A, R) \subseteq U$. Note that here since we serve just the most profitable commodities, there could be some nodes in the original graph that are neither in the set S, nor in the S'.

Let $R_k \subseteq P$ be the subset of paths for commodity $k \in K$ such that $R = \bigcup_{k=1}^{|K|} R_k$. We denote the set of nodes of the paths in R by $N(R)$, and the set of edges of R by $E(R)$. In addition, let $N_s(R)$ be the set of hub nodes and $N_{s'}(R)$ be the set of non-hub nodes in the paths of R. Likewise, let $E_h(R)$ be the set of hub edges, and $E_a(R)$ be the set of access edges of paths in R.

We define f and g as the set-up cost functions of the set of selected nodes and the set of selected edges, respectively as follows:

$$f(S, S') = -\sum_{j \in S} f_j - \sum_{j \in Q} f_j^{'}, \tag{2.1}$$

$$g(H, A) = -\sum_{e \in H} g_e - \sum_{e \in A} g_e^{'}. \tag{2.2}$$

Also, we define h as the the profit function associated with serving commodities $k \in K$ as:

$$h_k = \max_{p_k \in R_k} \left\{h_{p_k}\right\} \tag{2.3}$$

$$h(R) = \sum_{k \in K} h_k(R) \tag{2.4}$$

Moreover, we define the total profit as

$$z(S, S', H, A, R) = h(R) + g(H, A) + f(S, S'), \tag{2.5}$$

where we consider that $z(\emptyset) = 0$.

The HNDPSA can now be stated as the problem of selecting a set of nodes, edges and paths $(S,H,A,R) \subseteq U$ such that $z(S,H,A,R)$ is maximum, i.e.,

$$\begin{aligned} \max_{(S,S',H,A,R)\subseteq U} \{z(S,S',H,A,R) : |S| \le p, |H| \le q, N_s(R) = S, N_{s'}(R) = S', \\ E_h(R) = H, E_a(R) = A, S \cap S' = \emptyset, H \cap A = \emptyset, |\delta(\{i\})| = 1, \forall i \in S'\}. \end{aligned} \tag{2.6}$$

This is the path-based combinatorial representation for the HNDPSA. The first two constraints ensure that the cardinality constraints on the set of hub nodes and hub edges. The third, fourth, fifth and sixth constraints guarantee that the set of selected hub nodes, non-hub nodes, hub edges, and access edges should be exactly the same with the nodes and edges of the chosen paths R, respectively. The last set of constraints guarantees that for each non-hub node $i \in S'$, we select exactly one access edge incident to i. We note that these constraints only affect the selected non-hub nodes since we are maximizing the profit and we are not forced to serve all commodities.

2.2.1. Submodular properties of HNDPSA for the path-based representation

Now, we prove the submodularity of the objective functions considering the (21) definitions and following lemmas (see (21; 27));

Lemma 1 *Let d_j be the weight of $j \in N$. Then, the linear set function $f(S) = -\sum_{j\in S} d_j$ is a submodular function.*

Lemma 2 *A positive linear combination of submodular functions is submodular.*

Proposition 1 *(Submodularity of the objective function)*
(a) $h(R)$ is submodular and non-decreasing.
(b) $z(S,S',H,A,R)$ is submodular.

Proof: (a) Let $(S,S',H,A,R) \subseteq (T,T',L,B,Q) \subseteq U$, with $S \subseteq T \subseteq N$, $S' \subseteq T' \subseteq N$, $H \subseteq L \subseteq E$, $A \subseteq B \subseteq E$, $R \subseteq Q \subseteq T$. Without loss of generality, we assume that $q \in T \setminus Q$ is a path with endpoints of an arbitrary commodity $k = \{i, j\}$. For $R_k \neq \emptyset$:

$$\begin{aligned} h_k(R_k \cup \{q\}) - h_k(R_k) &= \max_{p_k \in R_k \cup \{q\}} h_{p_k} - \max_{p_k \in R_k} h_{p_k} \\ &= \max\{0, h_q - \max_{p_k \in R_k} h_{p_k}\} \\ &\ge \max\{0, h_q - \max_{p_k \in Q_k} h_{p_k}\} \\ &= \max_{p_k \in Q_k \cup \{q\}} h_{p_k} - \max_{p_k \in Q_k} h_{p_k} \\ &= h_k(Q_k \cup \{q\}) - h_k(Q_k) \end{aligned} \tag{2.7}$$

where the inequality follows from $\max_{p_k\in R_k} h_p \le \max_{p_k\in Q_k} h_p$. Moreover, the inequality $\max_{p\in Q} h_p - \max_{p\in R} h_p \ge 0 \quad \forall Q \subseteq T$ shows that h_k is non-decreasing.

As a consequence of Lemma 2, by taking the summation of overall commodities, we obtain that $h(R)$ is a submodular and non-decreasing function.

(b) According to Lemma 1 f, g, f' and g' satisfy submodularity over $(S,S',H,A,R) \subseteq (T,T',L,B,Q) \subseteq U$. Then, we have zero difference in the value of the function if we add a single element from the sets T or E, and finally for the function g, we can have a difference in the value of the g by just adding an element from set E. As a result, by lemma 2, z is a submodular function over the set U. □

3. Worst-Case Bounds for a Greedy Heuristic

In this section, we propose a greedy heuristic for the HNDPSA based on the representation 2.6 considering all the set-up cost functions as zero. Moreover, we provide the time complexity of the greedy algorithm and present the worst-case bounds.

3.1. *A Greedy Heuristic for the Path-Based Representation of HNDPSA*

Let $q \in P$ and $M_q = q \cup N_s(q) \cup E_h(q) \cup N_{s'}(q) \cup E_a(q) \subseteq U$. The following proposition states that we can evaluate a submodular function's decremental value by adding a subset to the solution set rather than a single element (see 25).

Proposition 2 *For* $(S,S',H,A,R) \subseteq (T,T',L,B,Q) \subseteq U$ *and any subset* $M \subseteq U \setminus (T,T',L,B,Q)$, $\rho_M(S,S',H,A,R) \geq \rho_M(T,T',L,B,Q)$.

Let (S^t,S'^t,H^t,A^t,R^t) be the solution at iteration t such that S^t is a set of hub nodes, S'^t is a set of non-hub nodes, H^t is a set of hub edges, A^t is a set of access edges, and R^t is a set of paths.

The idea of Algorithm 1 is to select a subset of commodities to be served that gives the most profit. To do so, we choose a feasible path q and its corresponding subset M_q that gives the maximum improvement at each iteration. We then add M_q to the solution set. Note that we must respect the capacity of the network (at most p hub nodes and q hub edges) at each iteration.

Moreover, if a node is selected as a hub (non-hub), we keep it as a hub (non-hub) during all the remaining iterations. Also, we handle the single assignment assumption such that for the non-hub nodes of the chosen path q^t at iteration t, we remove all their adjacent edges in the original graph G except $E_a(q^t)$. We continue iterating until either we cannot further improve the solution or we serve all commodities. After choosing the best path at each iteration, we update the hub nodes, non-hub nodes, hub edges, and access edges at each iteration. Our greedy algorithm is summarized in Algorithm 1.

Proposition 3 *The complexity of Algorithm 1 is* $O(|K|^2|N|^2)$.

Proof: The set ${M_q}^t$ at iteration t can be found efficiently by solving a series of longest-path problems on an auxiliary acyclic directed graph for each commodity $k = \{i,j\} \in K$ that is not served yet. We define auxiliary directed acyclic graphs $G_k(V_k,E_k)$ as a multi-layer graph for $U^{t-1} = (N^{t-1},E^{t-1},P^{t-1})$ at iteration t for commodity k as follows.

For constructing $G_k(V_k,E_k)$, we define the first layer as node i and the last layer as node j. If any node of the original graph G is not in the set S'^{t-1}, we copy it twice between i and j as two intermediate layers of hub nodes since we might pass through more than one hub node for routing a commodity (see Figure 2).

Assume that the amount of each commodity is equal to one (i.e., $W_k = 1$). For each $e \in E_k$ we define the weight as follows: from i to the node $\{h_{1m}\}_{m=1}^n$ of the first layer, we have $w_{ih_{1m}} = -d_{ih_{1m}}$. From

Algorithm 1 Greedy Heuristic for the HNDPSA

Set $(S,S',H,A,R)^0 \leftarrow \emptyset$, ${M_q}^0 \leftarrow \emptyset$, $U^0 = (N^0,E^0,P^0) \leftarrow U$, $S' \leftarrow \emptyset$, $A \leftarrow \emptyset$, $\rho_0 \leftarrow 0$, and $t \leftarrow 1$
while $t \leq |K|$ **do**
  Select $q^t \subseteq P^{t-1}$ such that $\rho_{M_q{}^t} = \max_{M_q \in U^{t-1}} \rho_{M_q}(S,S',H,A,R)^{t-1}$, $N_s(q^t) \notin S'$, and $E_h(q^t) \notin A$
  Set $\rho_{t-1} \leftarrow \rho_{M_q{}^t}$
  if $\rho_{t-1} \leq 0$ **then**
    Stop with $(S,S',H,A,R)^{t-1}$ as the greedy solution
  else
    Set $S' \leftarrow N_{s'}(q^t)$, $A \leftarrow E_a(q^t)$, $S \leftarrow N_s(q^t)$, and $H \leftarrow E_h(q^t)$
    $(S,S',H,A,R)^t \leftarrow (S,S',H,A,R)^{t-1} \cup {M_q}^t$
    $U^t \leftarrow U^{t-1} \setminus \{q^t\}$
  end if
  for $i \in N(q^t)$ such that $i \in S'$ **do**
    for $e \in E$ such that $i \in e$ **do**
      if $e \notin E_a(q^t)$ **then**
        $U^t \leftarrow U^t \setminus e$
      end if
    end for
  end for
  $t \leftarrow t+1$
end while
Set $(S,S',H,A,R)^t$ as the greedy solution.
STOP

$\{h_{1m}\}_{m=1}^n$ of the first layer to $\{h_{2m'}\}_{m'=1}^n$ of the second layer, we have $w_{h_{1m}h_{2m'}} = -\alpha d_{h_{1m}h_{2m'}}$. And finally, from $\{h_{2m}\}_{m=1}^n$ of the second layer to j, we have $w_{h_{2m}j} = -d_{h_{2m}j}$. We should note that having an arc between two nodes of the same layer is forbidden. The output of $G_k(V_k,E_k)$ is shown in Figure 2.

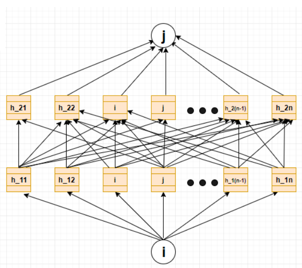


FIG. 2. $G_k(V_k,E_k)$ designed for serving commodity $k = \{i,j\}$

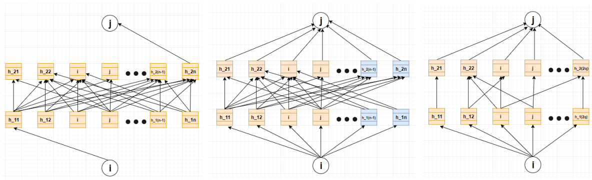
We consider the following restrictions in designing $G_k(V_k,E_k)$:

1. If either i or j are already in the set S'^{t-1}, we draw an arc just between them and their associated hubs (Figure 2, part (a)).
2. If the cardinality of the set of opened hubs has reached $p-1$, we draw arcs from the first intermediate layer of hubs just to the ones in the second layer, which are already in the set S (Figure 2, part (b)).
3. If the cardinality of the set of opened hub arcs has reached q between two intermediate layers, we will draw nodes that are already in S and hub arcs between them in H (Figure 2, part (c)).

Note that cases II and III are considered to avoid violating cardinality constraints on the hub nodes and hub edges. This operation takes $O(|K||N|^2)$ for each commodity since for each iteration we must solve a series of longest-path problems for at most $|N|$ commodities not served yet.

Thus, the time complexity of the greedy heuristic for the path-based representation is $O(|K|^2|N|^2)$. □

Figure 3 shows graph $G_k(V_k,E_k)$ for three different cases I, II and III that can happen for commodity $k=\{i,j\}$. Note that in the case I, the origin and the destination nodes are already in the set S'. In the case II, we show the intermediate hub layers that are already in S by orange and the ones that can be opened in the current iteration with blue. We do not draw any arc from the blue nodes to the blue nodes (i.e., we cannot open two new hubs). For the case III, we should not open a new hub arc, and we use the ones that are already opened and we just draw their associated hub nodes.



(a) I (b) II (c) III

FIG. 3. $G_k(V_k,E_k)$ for commodity k={i,j} with restrictions.

3.2. *A worst-case bound for the greedy heuristic*

Here we present a worst-case bound for Algorithm 1 using the following propositions. (see 21).

Let z^G denote the value of the solution constructed by the greedy heuristic.

$$z^G = z(\emptyset)+\rho_0+\ldots+\rho_{t-1} \qquad t \le |K|. \tag{3.1}$$

Let z^* be the value of an optimal solution to problem 2.6. We may assume $t \ge 1$ to exclude the trivial problem with $z^* = z^G = z(\emptyset)$.

Let $C(\theta)$ be the class of submodular set functions satisfying $\rho_{M_q}(S,S',H,A,R) \ge -\theta$, for all $(S,S',H,A,R) \subset U$ and $M_q \in U \setminus (S,S',H,A,R)$.

Proposition 4 *For all $(S,S',H,A,R),(T,T',L,B,Q) \subseteq U$ with $N_s(R)=S$, $N_{s'}(R)=S'$, $E_a(R)=A$, $E_h(R)=H$, and $N_s(Q)=P$, $N_{s'}(Q)=T'$, $E_a(Q)=B$, $E_h(Q)=L$. Then*

$$z(T,T',L,B,Q) \le z(S,S',H,A,R)+\sum_{q\in Q\setminus R}\rho_{M_q}(S,S',H,A,R)+|Q\setminus R|\theta. \tag{3.2}$$

Proof: Let $(S,S',H,A,R),(T,T',L,B,Q) \subseteq U$ with $|R\setminus Q|=\alpha$ and $|Q\setminus R|=\beta$ such that $N_s(R)=S$, $N_{s'}(R)=S'$, $E_a(R)=A$ and $E_h(R)=H$, and also $N_s(Q)=P$, $N_{s'}(Q)=T'$, $E_a(Q)=B$ and $E_h(Q)=L$. Consider the set M_q with $q \in Q\setminus R$ and similarly for $q' \in R\setminus Q$ consider $M_{q'}$, as defined before. Let's define $M_{q_0}=\emptyset$. Then

$$
\begin{aligned}
&z((S,S',H,A,R)\cup(T,T',L,B,Q))-z(S,S',H,A,R)= \qquad (3.3)\\
&z((S,S',H,A,R)\cup M_{q_1})-z(S,S',H,A,R)\\
&+z((S,S',H,A,R)\cup M_{q_1}\cup M_{q_2})-z((S,S',H,A,R)\cup M_{q_1})+...\\
&+z((S,S',H,A,R)\cup M_{q_1}\cup M_{q_2}\cup...\cup M_{q_\alpha})-z((S,S',H,A,R)\cup M_{q_1}\cup M_{q_2}\cup\\
&...\cup M_{q_{\alpha-1}})=\sum_{i=1}^{\alpha}\rho_{M_{q_i}}\left((S,S',H,A,R)\bigcup_{j=0}^{i-1}M_{q_j}\right)\\
&\leq\sum_{t=1}^{\alpha}\rho_{M_{q_t}}(S,S',H,A,R)=\sum_{q\in Q\setminus R}\rho_{M_q}(S,S',H,A,R),
\end{aligned}
$$

which shows

$$
\begin{aligned}
&z((S,S',H,A,R)\cup(T,T',L,B,Q))-z(S,S',H,A,R)\leq \qquad (3.4)\\
&\sum_{q\in Q\setminus R}\rho_{M_q}(S,S',H,A,R)
\end{aligned}
$$

and we obtain similarly

$$
\begin{aligned}
&z((S,S',H,A,R)\cup(T,T',L,B,Q))-z(T,T',L,B,Q)\geq \qquad (3.5)\\
&\sum_{q'\in R\setminus Q}\rho_{M_{q'}}((T,T',L,B,Q)\cup(S,S',H,A,R)\setminus M_{q'}).
\end{aligned}
$$

Subtracting the above inequalities results in the following:

$$
\begin{aligned}
&z(T,T',L,B,Q)\leq z(S,S',H,A,R)+\sum_{q\in Q\setminus R}\rho_{M_q}(S,S',H,A,R) \qquad (3.6)\\
&-\sum_{q'\in R\setminus Q}\rho_{M_{q'}}((T,T',L,B,Q)\cup(S,S',H,A,R)\setminus M_{q'}).
\end{aligned}
$$

Then,

$$
z(T,T',L,B,Q)\leq z(S,S',H,A,R)+\sum_{q\in Q\setminus R}\rho_{M_q}z(S,S',H,A,R)+\beta\theta \qquad (3.7)
$$

for $\rho_{M_{q'}}((T,T',L,B,Q)\cup(S,S',H,A,R)\setminus M_{q'})\geq-\theta$. □

Proposition 5 *Suppose $z\in C(\theta),\theta\geq 0$, and the greedy heuristic stops after k^* iterations, then the corresponding $\{\rho_i\}_{i=0}^{k^*-1}$ satisfy*

$$
z^*\leq z(\emptyset)+\sum_{i=0}^{t-1}\rho_i+|K|\rho_t+t\theta, \qquad t=0,...,k^*-1 \qquad (3.8)
$$

and also

$$
z^*\leq z(\emptyset)+\sum_{i=0}^{k^*-1}\rho_i+k^*\theta, \qquad if \quad k^*<|K| \qquad (3.9)
$$

Proof : From 4 we have

$$z(T,T',L,B,Q) \le z(S,S',H,A,R) + \sum_{q\in Q\setminus R} \rho_{M_q}(S,S',H,A,R) + |R\setminus Q|\theta. \tag{3.10}$$

Let (T,T',L,B,Q) be the optimal solution set of the problem 2.6 and (S,S',H,A,R) be the set $(S,S',H,A,R)^t$ constructed in t-th iteration of the greedy heuristic. We can show

$$z^* = z(T,T',L,B,Q), \qquad \rho_{M_q}(S,S',H,A,R)^t \le \rho_t, \quad \rho_t \ge 0, \tag{3.11}$$

and

$$\begin{aligned} &|R^t - Q| \le t, \quad \theta \ge 0, \\ &|Q - R^t| \le |K|, \end{aligned} \tag{3.12}$$

so,

$$\sum_{q\in Q\setminus R} \rho_{M_q}(S,S',H,A,R) \le \sum_{q\in Q\setminus R} \rho_t \tag{3.13}$$

this summation has an upper bound of

$$\sum_{q\in Q\setminus R} \rho_t \le |K|\rho_t \tag{3.14}$$

and

$$z(S,S',H,A,R)^t = z(\emptyset) + \sum_{i=0}^{t-1} \rho_i, \tag{3.15}$$

we obtain

$$z^* \le z(\emptyset) + \sum_{i=0}^{t-1} \rho_i + |K|\rho_t + t\theta \qquad t = 0,\ldots,k^*-1 \tag{3.16}$$

If $k^* \le |K|$, taking $(S,S',H,A,R) = (S,S',H,A,R)^{k^*}$ yields

$$z^* \le z(\emptyset) + \sum_{i=0}^{k^*-1} \rho_i + k^*\theta \tag{3.17}$$

as $\rho_{k^*} \le 0$ $\square$

Proposition 6 *For nondecreasing z, if the greedy heuristic is applied to problem 2.6, and the greedy heuristic stops after $k^* < |K|$ steps, the greedy solution is optimal.*

Proof : from Proposition 5 we have

$$z^* \leq z(\emptyset) + \sum_{i=0}^{k^*-1} \rho_i + k^*\theta, \qquad if \quad k^* < |K| \tag{3.18}$$

also, we can show

$$z^* \geq z^G = z(\emptyset) + \sum_{i=0}^{k^*-1} \rho_i \tag{3.19}$$

and using the fact that z is non-decreasing, $\theta = 0$. Therefore we have $z^* = z^G$. □

With Proposition 6 and assuming there is no setup cost since $h(R)$ is non-decreasing, the solution of the greedy heuristic is optimal.

Proposition 7 *By applying greedy heuristic to problem 2.6, we obtain*

$$\frac{z^* - z^G}{z^* - z(\emptyset)} \leq \frac{|K|-1}{|K|} \tag{3.20}$$

Proof : For $t = 0$, the inequality 3.8 yields $z^* - z(\emptyset) \leq |K|\rho_0 \leq |K|(z^G - z(\emptyset))$ or, equivalently,

$$\frac{z^* - z^G}{z^* - z(\emptyset)} \leq \frac{|K|-1}{|K|} \tag{3.21}$$

□

Now in the case with no set-up cost, we recall that z becomes non-decreasing ($\theta = 0$), so we can have a much sharper bound.

Proposition 8 *If the greedy heuristic terminates after k^* iterations, then*

$$\frac{z^* - z^G}{z^* - z(\emptyset) + |K|\theta} \leq \left(\frac{|K|-1}{|K|}\right)^{|K|} \leq \frac{1}{e} \tag{3.22}$$

Proof : See (21), Theorem 4.1 and 4.2. □

Now considering $z(\emptyset)$=0 and $\theta = 0$, we can claim,

$$\frac{z^G}{z^*} \leq 1 - \left(\frac{|K|-1}{|K|}\right)^{|K|} \leq \frac{1-e}{e} \tag{3.23}$$

4. Conclusion

We have studied a class of hub location problems called hub location problems with a single assignment. We modelled this problem as a combinatorial optimization problem. We proposed a greedy heuristic algorithm for our model, and it was derived the worst-case performance results of the algorithm. We proved a more robust bound for the case that has no set-up cost. We have shown the algorithm runs in a polynomial amount of time, which is $O(|K|^2|N|^2)$ when K shows the set of commodities and N denotes the set of nodes of the original graph.

Our contribution is to provide a combinatorial representation of hub location problems with a single assignment with profit. This is the first time that submodularity for HNDPSA is presented in the literature. By designing a polynomial greedy heuristic for such a submodular combinatorial optimization problem, we can reach a good approximation of the optimal solution in a short amount of time. This is very reasonable in the industry from the economic point of view instead of using expensive solvers and a lot of time. There is a theory that says if a combinatorial optimization satisfies submodularity, we can present a mixed-integer model for such a problem such that the number of decision variables will be the same as the number of elements of the combinatorial solution set (see (28)). In future works, we can apply the mentioned theory, and by achieving the optimal solution of the mixed-integer problem, we can compare it with the results of the greedy heuristic of this work.

REFERENCES

1. M. E. O'Kelly, "Location of hubs in a competitive environment," *Transportation Science*, vol. 20, no. 2, pp. 92–107, 1986.
2. M. E. O'Kelly, "A quadratic integer programming model for the location of interacting hub facilities," *Transportation Science*, vol. 21, no. 4, pp. 273–286, 1987.
3. J. F. Campbell and M. E. O'Kelly, "The integer solution to the single allocation p-hub median problem," *European Journal of Operational Research*, vol. 79, no. 3, pp. 390–399, 1994.
4. J. F. Campbell, "Hub location and the p-hub median problem," *Operations research*, vol. 44, no. 6, pp. 923–935, 1996.
5. A. T. Ernst and M. Krishnamoorthy, "Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem," *European Journal of Operational Research*, vol. 104, no. 1, pp. 100–112, 1998a.
6. S. García, M. Landete, and A. Marín, "New formulation and a branch-and-cut algorithm for the multiple allocation p-hub median problem," *European Journal of Operational Research*, vol. 220, no. 1, pp. 48–57, 2012.
7. B. Y. Kara and B. C. Tansel, "On the single-assignment p-hub center problem," *European Journal of Operational Research*, vol. 125, no. 3, pp. 648–655, 2000.
8. A. T. Ernst, H. Hamacher, H. Jiang, M. Krishnamoorthy, and G. Woeginger, "Uncapacitated single and multiple allocation p-hub center problems," *Computers & Operations Research*, vol. 36, no. 7, pp. 2230–2241, 2009.
9. B. Y. Kara and B. C. Tansel, "The single-assignment hub covering problem: Models and linearizations," *Journal of the Operational Research Society*, vol. 54, no. 1, pp. 59–64, 2003.
10. H. Calık, S. A. Alumur, B. Y. Kara, and O. E. Karasan, "A tabu-search based heuristic for the hub covering problem over incomplete hub networks," *Computers & Operations Research*, vol. 36, no. 12, pp. 3088–3096, 2009.
11. A. T. Ernst and M. Krishnamoorthy, "Solution algorithms for the capacitated single allocation hub location problem," *Annals of operations Research*, vol. 86, pp. 141–159, 1999.
12. M. Labbé and H. Yaman, "Projecting the flow variables for hub location problems," *Networks: An International Journal*, vol. 44, no. 2, pp. 84–93, 2004.

13. J. F. Campbell, A. T. Ernst, and M. Krishnamoorthy, “Hub arc location problems: part i—introduction and results,” *Management Science*, vol. 51, no. 10, pp. 1540–1555, 2005 a.
14. J. F. Campbell, A. T. Ernst, and M. Krishnamoorthy, “Hub arc location problems: part ii—formulations and optimal algorithms,” *Management Science*, vol. 51, no. 10, pp. 1556–1571, 2005b.
15. M. E. O’Kelly, “Hub facility location with fixed costs,” *Papers in Regional Science*, vol. 71, no. 3, pp. 293–306, 1992.
16. D. Skorin-Kapov, J. Skorin-Kapov, and M. O’Kelly, “Tight linear programming relaxations of uncapacitated p-hub median problems,” *European journal of operational research*, vol. 94, no. 3, pp. 582–593, 1996.
17. A. T. Ernst and M. Krishnamoorthy, “Efficient algorithms for the uncapacitated single allocation p-hub median problem,” *Location science*, vol. 4, no. 3, pp. 139–154, 1996.
18. D. Skorin-Kapov, “Genetic algorithm for the single allocation p-hub center problem,” *European Journal of Operational Research*, vol. 95, no. 1, pp. 61–70, 1996.
19. H. Pirkul and D. A. Schilling, “Scatter search for the single allocation p-hub median problem,” *European Journal of Operational Research*, vol. 98, no. 1, pp. 20–29, 1997.
20. D. Song, “A tabu search algorithm for the single allocation p-hub median problem,” *Computers & Operations Research*, vol. 29, no. 1, pp. 67–83, 2002.
21. G. L. Nemhauser and L. A. Wolsey, “Analysis of linear production programming,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 26, no. 2, pp. 418–444, 1978.
22. M. Grötschel, L. Lovász, and A. Schrijver, “Ellipsoid methods for integer programming,” *Combinatorica*, vol. 1, no. 2, pp. 169–197, 1981.
23. I. Contreras and E. Fernández, “Hub location as the minimization of a supermodular set function,” *Operations Research*, vol. 62, no. 3, pp. 557–570, 2014.
24. I. Ljubić and E. Moreno, “Outer approximation and submodular cuts for maximum capture facility location problems with random utilities,” *European Journal of Operational Research*, vol. 266, no. 1, pp. 46–56, 2018.
25. C. Ortiz-Astorquiza, I. Contreras, and G. Laporte, “Formulations and approximation algorithms for multilevel uncapacitated facility location,” *INFORMS Journal on Computing*, vol. 29, no. 4, pp. 767–779, 2017.
26. C. Ortiz-Astorquiza, I. Contreras, and G. Laporte, “An exact algorithm for multilevel uncapacitated facility location,” *Transportation Science*, vol. 53, no. 4, pp. 1085–1106, 2019.
27. D. Ortiz and I. Contreras, “A multi-level model and exact algorithm for the uncapacitated single allocation hierarchical hub location problem,” *Computers & Operations Research*, vol. 55, pp. 152–161, 2015.
28. G. Nemhauser and L. Wolsey, “The scope of integer and combinatorial optimization,” *Integer and combinatorial optimization*, pp. 1–26, 1988.