Heuristic Solution for the p-hub Problem

IVAN DERPICH
U.of Santiago of Chile
Dep. of Industrial Engineering
3769 Ecuador, E.Central
Santiago, Chile
ivan.derpich@usach.cl

REMIGIO CONTRERAS
Chilean Commission for Nuclear
Energy - CCHEN, Chile
12501 Nueva Bilbao, Las Condes
Santiago, Chile
rcontrerascamus@gmail.com

JUAN SEPULVEDA
U.of Santiago of Chile
Dep. of Industrial Engineering
3769 Ecuador, E.Central
Santiago, Chile
juan.sepulveda@usach.cl

Abstract: The hub location problem is important in the selection of technological networks, such as computer, cellular, or wireless sensor networks. These modern communication networks must be dynamically set as triggered by changes in external conditions; the nodes deplete their batteries and go out of service. For this reason, it is necessary to update the available data in order to determine which nodes can be used as hubs. The dynamic location problem requires a short solution time in despite of optimality. Heuristic methods are used for their simplicity and they are easy to package in the firmware. The central aim of this work is to design a heuristic method that will obtain a good feasible solution in a reasonable amount of time. The methodology proposed for the heuristic method consists of obtaining the optimum solution of the relaxed problem, followed by rounding this solution to a 0 or 1 value. The strategy developed for rounding the calculations is to first use a measure, called attractive force, for each node and then to define those nodes more attractive as hubs. Finally, an integer programming model is solved for assigning the nodes to the selected hubs. An interesting result is that the hubs selected by the optimal solution of the relaxed problem are always between the nodes that have the major attractive force. The heuristic algorithm is well established for problems with 10, 20, 25, 50 and 100 nodes. So, mixing two levels of difficulty we obtain four problems.

Key–Words: Hubs, location, networks, integer programming, heuristics.

1 Introduction

The problem of designing communications networks requires in some applications the solution of an optimal facility location problem; for instance, the ideal location of communication equipment and the location of transmission lines. Let us consider, for example, a network in which the nodes are computers and the edges are physical lines or wireless signal transmissions. For this case, it is necessary to decide how many and which of the computers will be connected to each hub. Another example is the location of antennas for cellular telephones, in which the goal is to optimize signal traffic for the user. In this paper, we discuss a subclass of problems in network design, where the nodes act as consolidation points for flows between such nodes. This type of facility is known as a hub. Hubs are places where flows, such as air cargo or communications, are concentrated. All of the traffic that is exchanged between nodes must be routed through one o more hubs which can be completely or partially interconnected. There are several variants of hub location problems, see for instance Campbell [1] for a comprehensive review of hub location problems.

1.1 Preliminaries

The exact number of required hubs, the location of hub and nodes, and the use of simple or multiple connections are just few examples of the variables that must be determined. In addition, the hubs may have capacity, cost, or location constraints that must be satisfied for a feasible network design. In this work, we focus on a particular variant of the hub location problem, known as the simple hub location problem with a node capacity constraint. We will refer to this problem as the Capacitated Single Allocation Hub Location Problem, denoted by CSAHLP. The choice of the problem originates from a design application involving a digital wireless equipment network.

Let us consider a network of antennas as shown in Figure 1. In this network, the antennas need to send to, and receive flows from other nodes in the network. In the figure, the problem of communication has been solved by connecting all nodes. This solution is known as a completely connected network; and, as it is shown in the figure, this solution requires dedicated communication arcs to all of the combinations.
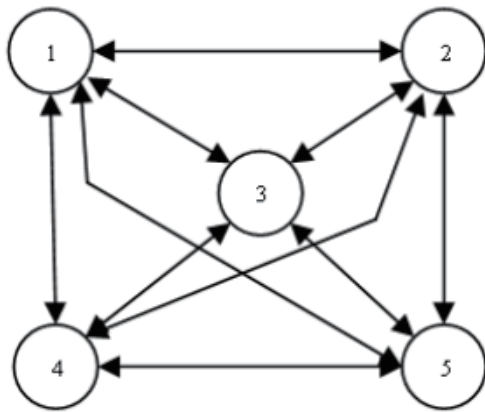
Figure 1. A network full connected

A more efficient way from the point of view of the number of arcs of communication, is using a flow concentrator node in which case the communication arc are greatly reduced. This concept is shown in Figure 2. In a network that uses concentrators for conveying the communication flows, these concentrator nodes should have the capability to concentrate and re-transmit. In the case shown in Figure 2, for example, all flows requiring to be transmitted pass through node 3, so this node needs to have a greater capacity than normal. This generates the idea of networks with limited capacity. The transmission costs also may vary when using a node or another as a hub. Thus, in an antenna array where there are physical difficulties associated with the transmission, for instance derived from interferences, this makes that certain nodes are preferable to others for use as hubs. This is modeled as the cost of using or not a node as a hub.
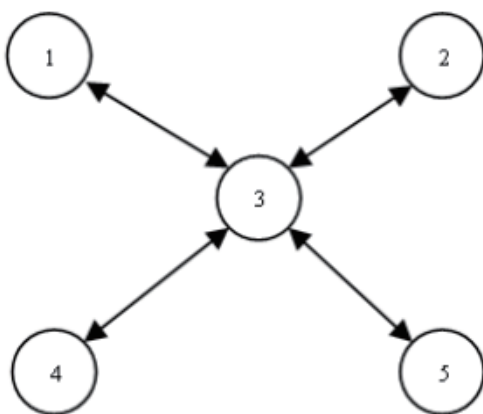

Figure 2. A network configured using hubs

Moreover, if the costs of using a hub are not too high in relation to the communication costs, there may be gains in the total cost if hubs are used. The decision of how many hubs must be used can also use included in the modeling and resolution of the problem.

Given the limited capacity of concentration of the antennas, it is necessary to introduce constraints on the node capacity and to determine the quantity of required antennas since there is a cost associated with the antenna use to be considered. In order to complete the communication between antennas, the model must decide what node will be use as a hub. Additionally, the model must assign the connections of nodes to hubs and account for the cost of transmission in each case. The cost of transmission between hub nodes is called the transmission cost, and the cost of transmission between a hub node and a target node is called the distribution cost. In this paper we are interested in the p-hub median problem. In this, there are p nodes that are selected as hubs while the remaining nodes are connected to the latter, such that the total cost of flow transportation is minimized.

1.2 Review of the literature

If the nodes need to connect to one or more hubs, we speak of "single allocation" (SA) or "multiple allocation" (MA), respectively. On the other hand, if there are constraints that indicate that the transport flow between nodes or hubs is limited to certain maximum or minimum flows, we say that the problem is "Capacitated" (C), otherwise, it is said to be "uncapacitated" (U). The uncapacitated single allocation p-hub median problem (USApHMP) was first formulated by O'Kelly [2]. Whereas, Campbell [3] formulated the first mixed integer linear programming (MILP) model for the uncapacitated multiple allocation p-hub median problem (UMApHMP).

By using a similar version, Campbell [4] formulated the USApHMP as a mixed integer programming (MIP) problem and he also made versions of the "single allocation" and "multiple allocation" of the "uncapacitated hub location problem". Ernst and Krishnamoorthy [5] developed a new MIP model for the USApHMP using less variables and constraints. With a similar approach the same authors [6] made the "multiple allocation" version of the Hub Median Problem (HMP). Some studies restrict the amount of flow through the connections between hubs. Campbell [4] introduced capacities to the HLP with "single allocation" and "multiple allocation", framing the problems of CSAHLP and USAHLP. Ernst and Krishnamoorthy [7] proposed two MIP models for the CSAHLP problem, where the second model required fewer constraints and variables.

There are several recent works on the location problem of simple hub assignments that are worth mentioning. Marcus Randall discussed solutions for the CSAHLP problem by using ant-colony metaheuristics [7]. Chen Chen et al. [9] developed a

heuristic for CSAHLP and compared it with simulated annealing (SA). Costa [10] produced a unique and interesting approach for CSAHLP problems by using bi-criteria. The approach used in these work does not use capacity constraints on the flow; instead, it uses a second objective function to minimize the CPU time. In this paper, we work with the CSAHLP-N model, which corresponds to the model reformulated by Ernst and Krishnamoorthy [4]. The original CSAHLP-C model is modified by changing the main variable X, which shows the percentage of flow from an origin node to a final node, to the variable Y. This new variable represents the amount of flow from the origin and over the route through the hubs.

Some additional equations were modified from the original formulation. The most important advantage of the new CSAHLP-N model with respect to CSAHLP-C is the reduction in the problem size and the CPU time. In 2008, Chen [11] developed a heuristic algorithm based on simulated annealing. Computational results indicate that the present heuristic outperforms the simulated annealing method. An Ant Colony System Hybridized with a Genetic Algorithm for the Capacitated Hub Location Problem was presented by Sun et al. [12], they deal with a capacitated asymmetric allocation hub location problem (CAAHLP). A review of the state of the art published by Alumur and Kara [13] showed the increasing interest of the OR community for improving the power of the algorithms in order to solve big problems in few seconds. Contreras et al.[14] presented the Tree of Hubs Location Problem. They propose an integer programming formulation for the problem and present some families of valid inequalities that reinforce the formulation; in this paper we give an exact separation procedure for them.

2 Mathematical Formulation

The following is the formulation used in this paper of the p-hub Problem in CSAHLP-N version.

2.1 Formulation CSAHLP-N

Minimize

$$\sum_{i\in\mathbb{N}}\sum_{k\in\mathbb{N}} d_{i,k} Z_{i,k}(\chi O_i + \delta D_i)$$
$$+ \sum_{i\in\mathbb{N}}\sum_{k\in\mathbb{N}}\sum_{l\in N} \alpha d_{k,l} Y_{k,l}^i + \sum_k F_k Z_{k,k}$$

subject to

1) $\sum_{k\in\mathbb{N}} Z_{k,k} = 1 \,\forall i \in \mathbb{N}$
2) $Z_{i,k} \leq Z_{k,k} \quad \forall i,k \in \mathbb{N}$

3) $\sum_{i\in\mathbb{N}} O_i Z_{i,k} \leq \tau_k Z_{k,k} \,\forall k \in \mathbb{N}$

4) $\sum_{l\in\mathbb{N}} Y_{k,l}^i - \sum_{l\in N} Y_{l,k}^i = O_i Z_{i,k} - \sum_{j\in N} W_{i,j} Z_{j,k} \forall i,k \in \mathbb{N}$

5) $Z_{i,k} \in \{0,1\} \,\forall i,k \in \mathbb{N}$

6) $Y_{k,l}^i \geq 0 \,\forall i,k,l \in \mathbb{N}$

Where the variables of the model are:

$Y_{k,l}^i$: is the flow per unit time unit from node i through hubs k y and l

$Z_{i,k}$: is a binary variable equal to 1 if the optimal route includes nodes i and k; 0 otherwise.

The data of the model are:

$d_{k,l}$: distance between nodes i and k;

τ_k: flow capacity of the node k;

χ: unitary recollection cost;

α : unitary transportation cost;

δ : unitary distribution cost;

O_i: outgoing material flow from node i;

D_i: incoming material flow to node i;

F_k: cost of using node k as a hub;

$W_{i,j}$: flow of material from node i to node j.

N: set of all nodes.

Equation (1) forces the assignment of a unique route from any node i is used to constraint the decision variable Z; equation (2) is used to select node hubs for each flow; . Equation (3) bounds restricts the use of nodes to their capacity limitations. Equation (4) is the difference equation for node i on node k, where the demand and the supply of the nodes are determined by the location $Z_{j,k}$; . Equation (5) defines $Z_{j,k}$ as binary, and Equation (6) defines Y as a real positive variable that includes 0.

The proposed model has an integer variable $Z_{i,k}$ and a real variable called $Y_{k,l}^i$ that correspond to the flow from node i to hubs k and l, respectively. $Z_{i,k} = 1$ indicates that node i is connected by node k, and $Z_{i,k} = 1$ indicates that node i is a hub. This formulation therefore corresponds to an mixed integer programming mixed problem (MIP) problem.

2.2 Relaxing the CSAHLP-N formulation

If the variables $Z_{i,k}$ are relaxed such that they take real values between zero and one, this model becomes a problem of linear programming (LPPPL) problem. In this case, the variable $Z_{i,i}$ measures the degree to which node i has the potential to be a hub. In this paper, it is assumed that this potential is a measure of the attraction of the node to be used as hub. For this reason, it is designated as the "attractive force". Now, we will focus on the relaxed problem CSAHLP-N, letting P_R) be the relaxed problem:

P_R)
Minimize

$$\sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} d_{i,k} Z_{i,k} (\chi O_i + \delta D_i)$$
$$+ \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} \alpha d_{k,l} Y^i_{k,l} + \sum_k F_k Z_{k,k}$$

With the following: (1), (2), (3), (4) and (6) of the original problem as expressed in 2.1.

5') $0 \leq Z_{i,k} \leq 1 \quad \forall i, k \in \mathbb{N}$

2.3 Formulation CSAHLP-C

Minimize

$$\sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \sum_{j \in N} W_{i,j} C_{i,j,k,l} X_{i,j,k,l} + \sum_{k \in N} Z_{k,k}$$

subject to

1) $\sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} X_{i,j,k,l} = 1 \quad \forall i, j \in \mathbb{N}$

2) $Z_{i,k} \leq Z_{k,k} \quad \forall i, k \in \mathbb{N}$

3) $\sum_{j \in N} \sum_{l \in N} (W_{i,j} X_{i,j,k,l} + W_{j,i} X_{j,i,l,k}) = (O_i + D_i) Z_{i,k}, \forall i, k \in \mathbb{N}$

4) $\sum_{i \in \mathbb{N}} O_i Z_{i,k} \leq \tau_j Z_{j,k} \forall i, k \in \mathbb{N}$

5) $Z_{i,k} \in \{0, 1\} \quad \forall i, k \in \mathbb{N}$

6) $0 \leq X_{i,j,k,l} \leq 1 \quad \forall i, j, k, l \in \mathbb{N}$

Skorin-Kapov et al. [10] note that constraint (3) is very weak. Hence, to obtain useful lower bounds from the LP relaxations, they replace (3) by the pair of constraints :

7) $\sum_{l \in \mathbb{N}} X_{i,j,k,l} = Z_{i,k} \quad \forall i, j, k \in \mathbb{N}$

8) $\sum_{k \in \mathbb{N}} X_{i,j,k,l} = Z_{j,l}, \forall i, j, l \in \mathbb{N}$

The new formulation is denoted as CSAHLP-LP. Unfortunately these constraints make the formulation very large. In practical terms, it means that solving the LP relaxation for problems with more than 20 nodes becomes too slow.

3 Heuristic attraction force algorithm (HAFA)

3.1 Attraction force of a node

The force of attraction of a node represents a measure of the incoming and outgoing flows of a node. It was made a linear correlation study between these

Table 1: Correlation of evaluation 9)

Problem	Coefficient of correlation
20LL	0.415
20LT	0.536
20TL	-0.249
20TT	-0.267
25LL	0.432
25LT	-0.383
25TL	-0.253
25TT	-0.307

flows and the variable Zi,j obtained from the resolution of the relaxed problem, in order to obtain a related variable that was accurate and easily obtainable. The model used is as follows:

9) $Z_{i,i} = \alpha_0 + \alpha_1 * (\sum_{j \in \mathbb{N}} W_{i,j} * d_{i,j} + \sum_{j \in \mathbb{N}} W_{j,i} * d_{j,i})$

This model is based on the idea that the amount of flow, weighted by the distance, of each flow, is associated with the force of attraction of the node and this node has its related variable in the diagonal of the matrix Z. For validating this intuitive idea, we conducted experiments with problems of reduced size, namely problems with 20 and 25 nodes. We run a linear correlation model as the one shown above. This was estimated for the problems of 20 and 25 nodes from the test data 'AP data' from the public OR-library uploaded by Beasley [15]. The results are shown in Table 1.

The results showed in the table 1 suggest that there is a correlation between the attraction force of the hubs over the rest of the nodes and the value of the relaxation problem. The attraction force of a node is a measure of the proportion of the total flow that would be assigned to this node if it was defined as a hub. This measure is calculated in the different iterations of the heuristic algorithm with results of the relaxed problems. The intuitive idea, it is that the node that receives more flow is a good candidate to be a hub. In addition we use another intuitive idea, that a good solution is one which minimizes the numbers of hubs serving all of the flows. In order to consider this fact we divided the first measure by the number of hubs.

The attraction force of a node $a(i)$, is the average between the pure attraction force and relative attraction force. The pure attraction force is the sum of all proportions of the flow that would be assigned to this node using $Z_{i,j}$ of the optimalum solution of the relaxed problem. The normalized attraction force is the

pure attraction force divided by the numbers of utilized hubs used. The flow chart and pseudo code of the heuristic are shown below. The inputs are the classic input of the hub and spoke problem. The outputs correspond to the optimalum solution.

3.2 Flowchart of the HAFA

The algorithm has three stages perfectly defined, in the first problem three relaxed problems are solved, in the first, which we call the base case, we solve the relaxed problem without adding additional constraints and the attraction force of the problem is measured, which is used to calculate the number of hubs that to meet the requirements of the problem. After two subproblems are solved, one with an additional hub and other with a hub less than in the base case. In the second stage, the problem with smaller value of the objective function,is selected and the number of hubs of this problem will be the number of hub to use in the solution of the problem. In the third stage, fixed number of hubs in the amount defined in the second stage and the original problem is solved fixing the number of hubs.

First Stage : Solving three relaxing problem
Input: d, O, D, F and τ
1. Solve the linear problem P_R and two new linear problems that we call P_R^+ and P_R^-
2. In the problems P_R , P_R^+ and P_R^- we define the normalized attractive force
3. In the problems P_R , P_R^+ and P_R^- wedefine the value of the normalized relative attractive
4. Let a_i^0,a_i^+ and a_i^- the averages between f_i and r_i
5. Let P^* be the problem that shows the minimum value of the optimal solution Let v^* be this value and let $x^* = (Z, Y)$ be the optimum solution

Second Stage : Finding the Optimum Solution
1. If $P^* = P_R$ then we make $N = N^0$ and $a_i = a_i^0$
2. If $P^* = P_R^+$ then we make $N = N^0 + 1$ and $a_i = a_i^0 + 1$
3. If $P^* = P_R^-$ then we make $N = N^0 - 1$ and $a_i = a_i^0 - 1$

Third Stage : Solving an integer problem
1. We define the set I with the N nodes that have the greatest value of the variable a_i
2. Solve the original MIP problem adding the following constraints : $\forall i \in I$: $Z_{i,i} = 1$ and $\forall i \notin I$ $Z_{i,i} = 0$

3.3 Heuristic algorithm

HAFA ROUTINE
INPUT: $W, d, F, \alpha, \chi, \delta, \tau$
OUTPUT:X_{int}^*

1) Solve the linear problem P_R. Let $Z^o \in R^{nxn}$ be the matrix of the optimalum solution of this problem and let N^0 be the number of hubs a priori, which we will improve within the heuristic algorithm.

$$N_0 = \lceil \sum_{i \in \mathbb{N}} Z_{i,i}^0 \rceil$$

Let Z^0 be the matrix Z obtained of the optimalum solution of the problem P_R).

2) Then we solve two new linear problems, that we call P_R^+ and P_R^- respectively. The problem P_R^+ is formed with the problem P_R adding the constraint

$$\sum_{i \in \mathbb{N}} Z_{i,i} = N^0 + 1.$$

By doing this we force that the model to assigns $N^0 + 1$ nodes.

The problem P_R^- is formed with the P_R by adding the constraint $\sum_{i \in \mathbb{N}} Z_{i,i} = N^0 - 1$. And now we force the model such in order that it assigns $N^0 - 1$ nodes.

Let Z_+^* and Z_-^* be the matrix Z obtained of the optimalum solution of the problems P_R^+ and P_R^- respectively.

3) In the problems P_R , P_R^+ and P_R^- we define f_0, f_i^+ and f_i^- as the value of the normalized attractive force for the node i.

$$f_i^0 = \frac{\sum_{j \in \mathbb{N}} Z_{j,i}^0}{n}$$

$$f_i^+ = \frac{\sum_{j \in \mathbb{N}} Z_{j,i_+}^*}{n}$$

$$f_i^- = \frac{\sum\limits_{j \in \mathbb{N}} Z_{j,i_-}^*}{n}$$

4) In the problems P_R, P_R^+ and P_R^- we define r_i^0, r_i^+ and r_i^- as the value of the normalized relative attractive force for the node i.

$$r_i^0 = \frac{\sum\limits_{j \in \mathbb{N}} Z_{j,i}^0}{N^0}$$

$$r_i^+ = \frac{\sum\limits_{j \in \mathbb{N}} Z_{j,i_+}^*}{N^0 + 1}$$

$$r_i^- = \frac{\sum\limits_{j \in \mathbb{N}} Z_{j,i_-}^*}{N^0 - 1}$$

5) Let a_i^0 , a_i^+ and a_i^- be the averages between f_i and r_i :

$$a_i^0 = \frac{f_i^0 + r_i^0}{2}$$

$$a_i^+ = \frac{f_i^+ + r_i^+}{2}$$

$$a_i^- = \frac{f_i^- + r_i^-}{2}$$

We will call these forces, normalized attractive forces average.

6) Let P^* be the problem that show the minimum value of the optimalum solution between problems P_R, $P_R^+ + P_R^-$.

Let v^* be this value. Let $x^* = (Z, Y)$ be the optimum solution of the problem P^*.

If $P^* = P_R$ then we make $N = N^0$ and $a_i = a_i^0$, if $P^* = P_R^+$ then we make $N = N^0 + 1$ and $a_i = a_i^+$ if $P^* = P_R^-$ then we make $N = N^0 - 1$ and $a_i = a_i^-$

Then we define the set I with the N nodes that have the greatest value of the variable a_i.

Solve the original MIP problem adding the next constraints : $\forall i \in I$, $Z_{i,i} = 1$ $\forall i \notin I$, $Z_{i,i} = 0$.

Let $x_{int}^* = (Z, Y)$ be the optimum solution of this problem.
END SUBROUTINE

3.4 A bound for the integrality gap

The following theorem provides a bound for the integrality gap, which is a measure of the maximum errors that can be committed by using the proposed heuristic.

Theorem 1 *Let v^* be the optimum solution of the integer problem and let v_R^* be the optimum value of the relaxed problem. Then we define the integrality gap.*
The integrality gap of the CSAHLP-N is:

$$gap = v^* - v_R^*$$

Then

$$gap \leq N^2 \max_{i,k}\{d_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ N^3 \, \alpha \max_{k,l}\{d_{k,l}\,\tau_k + \tau_k\}$$
$$+ \sum_k F_k - 2\,(N^0 - 1)\,\min_k\{F_k\}$$

Proof: In the optimal solution of the relaxed problem P_R the cost of transport is equal to the cost of operation of the hubs, that is:

$$\sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}}\{d_{i,k}Z_{i,k}\{\chi O_i + \delta D_i\}\} + \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} \alpha d_{k,l}Y_{k,l}^i$$
$$= \sum_k\{F_k Z_{k,k}\}$$

The left hand side of the equality is the cost of transport and the right hand side corresponds to the operation cost. Then the value of the objective function at the optimal solution is:

$$v_R^* = 2 \sum_k F_k Z_{k,k}$$

with $0 \leq Z_{i,k_-} \leq 1 \; \forall i, k \in N$.
Lower bounding:

$$v_R^* \geq 2 \, \min_k\{F_k\} \sum_k Z_{k,k}$$

The lowest value of $\sum\limits_{k \in \mathbb{N}} Z_{k,k} = N^0 - 1$
Then $v_R^* \geq \min_k\{F_k\}(N^0 - 1)$
For other size:

$$v^* = \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}}\{d_{i,k}Z_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} \alpha d_{k,l}Y_{k,l}^i + \sum_k F_k Z_{k,k}$$

and

$Z_{i,k} \leq 1 \; \forall i, k \in \mathbb{N}$
$Y_{k,l}^i \leq \tau_k \; \forall i, k, l \in \mathbb{N}$,
$Z_{k,k} \leq 1 \; \forall k \in \mathbb{N}$,
bounding

$$v^* \leq \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}}\{d_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ \alpha N \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} d_{k,l}\tau_k + \sum_k F_k$$

and bounding

$$d_{i,k}\{\chi O_i + \delta D_i\} \le \max\{d_{i,k}\{\chi O_i + \delta D_i\}\} \quad \forall i, k \in \mathbb{N}$$

and

$$d_{k,l}\tau_k \le \max\{d_{k,l}\tau_k\} \quad \forall k, l \in \mathbb{N},$$

we have

$$v^* \le \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} \max\{d_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ \alpha N \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} \max\{d_{k,l}\tau_k\} + \sum_k F_k$$

then

$$v^* \le N^2 \max_{i,k}\{d_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ \alpha N^3 \max_{k,l}\{d_{k,l}\tau_k\} + \sum_k F_k$$

and finally

$$gap \le N^2 \max_{i,k}\{d_{i,k}\{\chi O_i + \delta D_i\}\}$$
$$+ \alpha N^3 \max_{k,l}\{d_{k,l}\tau_k\}$$
$$+ \sum_k F_k - 2(N^0 - 1)\min_k\{F_k\}$$

end proof.

It is possible that this bound is was not a tight bound, but it shows that the gap decreases as when N^0 grows.

Conversely as N^0 is related to the attraction force of all nodes, we can establish that the gap decreases when the attraction force of all nodes grows.

Since the integrality gap relates with the complexity of solving the integer linear programming problem, then we can conjecture that the attraction force of all nodes is a measure of the complexity of solving the integer linear programming problem.

4 Computational experiments

In order to test the heuristic algorithm, a data set denoted 'AP data'was used. The data set belongs to the public library in the OR-library posted by Beasley [15]. The authors Ernest and Krishnamoorthy [5] posted this collection. The data include capacity constraints and costs on nodes. We show the characteristics of every instance of the utilized test problem.

The name of each instance consists of the number of nodes followed by two characters: the first is the type of cost and the second is the capacity type. The characters LL designates low cost and relaxed capacity, the characters LT designates low costs and tight capacity, the characters TL designates high costs and relaxed capacity while the characters TT represents high cost and tight capacity.

Table 2 shows the results of the HAFA for CSAHLP-N.During the benchmarking, we solved two

Table 2: Results of the HAFA for CSAHLP-N

Code pro-blem	CPU Time m:s	Object. function dollars	Hubs selected	Error %
10LL	00:00	224,250	4,5,10	0.00
10LT	00:00	250,992	1,4,5,10	0.00
10TL	00:00	263,399	4,5,10	0.00
10TT	00:00	263,399	4,5,10	0.00
20LL	00:00	$250,022^1$	7,14	6.53
20LT	00:01	259,755	$4, 6, 14^2$	2.46
20TL	00:00	$271,892^1$	7,19	2.82
20TT	$00:05^3$	296,035	1,10,19	0.00
25LL	00:15	238,978	8,18	0.00
25LT	00:26	286,200	$9, 14, 16, 25^2$	3.56
25TL	00:01	310,317	9,23	0.00
25TT	$00:20^3$	352,070	9,12,14	1.06
40LL	00:20	242.167	14, 29	0.09
40LT	$06:55^3$	272.218	14,26,30	0.00
40TL	03:36	298,919	14,19	0.00
40TT	$05:27^3$	354,874	14, 19, 40	0.00
50LL	00:35	$238,573^1$	15, 35	0.02
50LT	00:24	273,382	6,26,32,46	0.18
50TL	00:34	331,511	3,27,45	3.92
50TT	13:13	424,125	12, 25, 26	1.60
100LL	14:52	246,714	29,64,73	0.00
100LT	29:12	265,932	29, 71,95	3.62
100TL	16:44	362,950	44, 52	0.00
100TT	40:56	465,213	5, 34, 86, 95	0.00

Notes: (1) The allocation is not optimal (2) The hub assigned is not optimal (3) CPU time becomes

versions of the problem: CSAHLP-N version and CSAHLP-C version.

Table 3 shows the results of HAFA for the CSAHLP-C. First we used the CSAHLP-N version and solved problems containing up to 100 nodes. The obtained solution and the selected hubs correspond to the optimal solution for the instances from 10LL to 100LL.

The instances that are not optimal are marked with an asterisk. Then we used the CSAHLP-C version of the problem and solved problems containing up to 25 nodes. As said before, this formulation is very large and solving problems with more than 25 nodes was impossible.

As shown in Figure 3, we can see that the CPU time rapidly increases for the problems with formulation CSAHLP-C, while the problems with formulation CSAHLP-N increase slowly. This exponential

Table 3: Results of the HAFA for CSAHLP-C

Code problem	CPU Time m:s	Object. function dollars	Hubs selected	Error %
10LL	00:20	224,250	4,5,10	0.00
10LT	$01:49^{1}$	250,992	1,4,5,10	0.00
10TL	00:21	263,399	4,5,10	0.00
10TT	00:18	263,399	4,5,10	0.00
20LL	04:08	234,690	7,14	6.53
20LT	03:48	253,517	10, 14	2.46
20TL	04:08	271,128	7,19	2.82
20TT	06:11	296,035	1,10,19	0.00
25LL	20:41	238,978	8,18	0.00
25LT	22:37	279,991	$9, 12, 25^{2}$	1.31
25TL	25:05	316,327	$14, 23^{2}$	1.94
25TT	16:14	359,043	$8, 14, 25^{2}$	3.06

Notes: (1) An unfeasible case, (2) The selected nodes are not optimal.

growth in the CPU time in the formulation CSAHLP-C is caused by the explosive growth in the number of variables and in the number of constraints.
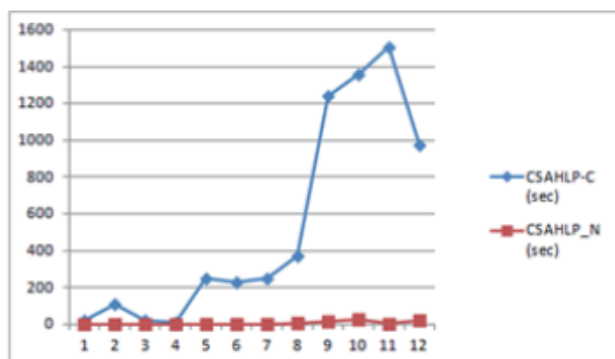

Figure 3. CPU-time of CSAHLP-C vs vs CSAHLP-N in seconds

In this formulation every flow between two nodes uses four indexes in the variable X. For this reason the numbers of variables grows exponentially with the numbers of nodes. Thus, we did not prove the set problem for 30 o more nodes with the formulation CSAHLP-C.

As mentioned earlier, this heuristic is based on the relaxation of the integer variables for finding a solution by linear programming for the original linear IP problem and then for determining a number of hubs to be used in the final solution. Subsequently, this amount of hubs is used for solving the associated IP problem. The structure of the problems for which the method was applied, corresponds to the type of lo-

cation problems. A proof of this is that the objective function is made out of two terms: one that represents the cost of transport between nodes and the other representing the cost of using a certain number of nodes as concentrators. For this reason, it is very reasonable to think that this algorithm can be applied to the resolution of location problems.

5 Conclusion

This work developed a heuristic algorithm to find a solution for the CSAHLP problem. Two formulations were proved CSAHLP-C y CSAHLP-N. For the CSAHLP-C only three size of nodes were proved: 10, 20 and 25 nodes. For problems with more nodes, the CPU time was very large. For the CSAHLP-N, six size of nodes were proved: 10, 20, 25, 40, 50 and 100 nodes. The obtained CPU times are interesting and the gaps are small in most of the cases. The heuristic is fast and it approaches the optimal solution for most types of problems. The heuristic is very simple because it only needs mathematical operations and could be packaged in the firmware. This work introduced a new concept, called the attractive force. This is a measure of the likelihood of a node to become a candidate hub. This measure is made by using a mix of characteristic nodes that are placed into the linear programming problem.

This kind of heuristic is based on relaxing some constraints and reformulating the original problem in such a manner that it is easier to solve than the original one. The decision variables of this problem are then transformed into new equations that constrain the solution space for the new problem. Again, we must make the decision variable and transform it to a new constraint for the next problem. We can state that it takes a circular form to obtain the result, and in most cases, it is possible to find the optimal solution.

A comparison of our heuristic algorithm with the meta-heuristic approach, shows that our approach sometimes finds the optimal solution, while the optimal solution is never found in the meta-heuristic approach, only a quasi-optimal solution is found. The useful concept of 'Total Attractive Force'measures the necessary capacity to cover the flow demand at a minimum cost. This allows us to use the 'Total Attractive'as a first approach to find the optimal number of hubs, since the cost function is quadratic and convex.

The complexity of an algorithm is measured by the number of iterations that uses in the worst case to solve any problem. The iterations can be elementals operations, matricial operations or others algorithms. The complexity of this algorithm we will measure using the simplex algorithm for linear programming

and the Branch and Bound algoritm (B& B) for integer programming. For this reason, the complexity of this algorithm is bounded by a term with three components, associated with the three problems solved. First, a linear problem (LP) is solved. Second, two linear problems fixing the number of hubs are solved. Finally, one MIP problems (MIP) with fixed numbers of hubs are solved. Then the number of iterations of the heuristic algorithm proposed is in the order of $3LP + 1B\&B$. Although mixed integer programs are not polynomial problems, in practical terms, the computational complexity and CPU time is reduced by fixing the nodes that will be assigned as hubs.Although is not possible to sure that the complexity of this heuristic procedure is lower than the complexity of the original problem, is possible to show that the number of integer variables that match the heuristic proposed in the CSAHLP-N formulation is lower than the integer variables of the original problem. In the original problem there are $n^2 + n^3$ integer variables while in the heuristic proposed, the algorithm uses a matrix of $n X n$ where only a few variables are not zero, then number of integer variables is $h X n + h X n^2$ with $h \leq n$. h corresponds to the number of hub in the final solution. So, the complexity is reduced strongly.

This methodology can be extended to other problems with similar characteristics, such as location, set covering, and network problems. In the heuristic, the resolution of a LP problem is necessary, when we relax the integer variables. This is done using the simplex algorithm, which is not very efficient for big problems. Thus, a future research direction is to reduce the times of resolution of the LP problem by using some heuristic algorithm of faster convergence.

Another area of improvement of the heuristic is the step seven. In this step a linear integer programming problem is solved. This problem correspond to the original problem in which there have been fixed the variable corresponding to the values of the diagonal of the matrix Z, with ones or zeros.

References:

[1] J. F. Campbell, Hub location and the p-hub median problem, *Operations Research* , Vol. 44, 6, 1994 pp. 923–935 .

[2] M. E. Ó Kelly, A quadratic Integer Program for the Location of Interacting Hub Facilities, *European Journal of Operational Research*, Vol. 32, 3, 1987, pp. 393–404.

[3] J. F. Campbell, Location and Allocation for Distribution System with Transshipments and Operations Research, Vol. 72, 1, 1992, pp 77–99.

[4] J. F. Campbell, Integer programming formulations of discrete hub location problems, *European Journal of Operational Research*, 72, 1994b, pp 387–405.

[5] A. T. Ernst, M., Krishnamoorthy, Efficient algorithm for the Uncapacitated Single Allocations p-Hub Median Problem, *Location science*, Vol.4, 3, 1996, pp. 582–593.

[6] A. T. Ernst, M. Krishnamoorthy, Exact and heuristic Algorithm for the Uncapacitated Multiple Allocation p-Hub Median Problem. *European Journal of Operational Research* , Vol. 104, 1, 1998, pp. 100-112.

[7] A. T Ernst, M. Krishnamoorthy, Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86, 1999, pp. 141-159.

[8] M. Randall. (2008). Solution approaches for the capacitated single allocation hub location problem using ant colony, *Computational Optimization and Application*,2008, pp. 239–261.

[9] J. Chen, X. Cao, Y. Xiao and Y. Sun, Simulated annealing for optimization with wireless sensor and actuator networks, *Electronics Letters*, Vol.44, 20, 2008, pp. 1208-1209.

[10] M. Costa, M. E. Captivo, J. Climaco, Capacitated single allocation hub location problem - A bi-criteria approach, *Computers and Operations Research*, 35, 2008 pp. 3671–3695.

[11] J. F. Chen, , (2008). A Heuristic for the Capacitated Single Allocation Hub Location Problem, *Lecture Notes in Electrical Engineering, Advances in Industrial Engineering and Operations Research* , Vol.5, 2008, pp. 185–196.

[12] J. U. Sun, D. Park, An Ant Colony System Hybridized with a Genetic Algorithm for the Capacitated Hub Location Problem, *Lecture Notes in Computer Science*, Vol. 7425, 2012, pp 173–181.

[13] S. Alumur, B. Y. Kara, Network hub location problems: The state of the art, *European Journal of Operational Research* , Vol.190, 1, 2008 pp. 1–21.

[14] I. Contreras, E. Fermández, A. Marin, The Tree of Hubs Location Problem, *European Journal of Operational Research* Vol.202, 1,2010 pp.390–400.

[15] J. Beasley, OR-Library distributing test problems by electronic mail,*Journal of the Operational Research Society* , Vol.41, 11, 1990, pp.1069–1072.