

An Efficient Modified Greedy Algorithm for the P -Median Problem

M. Dzator^a and J.Dzator^b

^a *ALSU, Central Queensland University, Mackay;* ^b *The University of Newcastle, NSW*
Email: m.dzator@cqu.edu.au

Abstract: The fundamental objectives of locating facilities can be summarized into three categories. The first category refers to those designed to cover demand within a specified time or distance. This objective gives rise to location problems which are known as the Location Set Covering Problem (LSCP) and the Maximal Covering Location Problem (MCLP). The LSCP seeks to locate the minimum number of facilities required to 'cover' all demand or population in an area. The MCLP is to locate a predetermined number of facilities to maximize the demand or population that is covered. The second category refers to those designed to minimize maximum distance. This results in a location problem known as the p -center problem which addresses the difficulty of minimizing the maximum distance that a demand or population is from its closet facility given that p facilities are to be located. The third category refers to those designed to minimize the average weighted distance or time. This objective leads to a location problem known as the p -median problem. The p -median problem finds the location of p facilities to minimize the demand weighted average or total distance between demand or population and their closest facility. The p -median problem is a typical combinatorial optimization problem with many practical applications such as location of warehouses, schools, health centers, shops etc.

Greedy algorithms are the simplest algorithms to design however it is not easy to understand its capability and limitations. A greedy algorithm solves a global optimization problem by making a sequence of locally optimal decisions. That is a greedy algorithm always chooses the next step of an algorithm that is locally optimal. For example for Facility Location Problem we will consider the facilities for which decisions regarding locally optimal locations will be made. The decisions that are made regarding where to locate successive facilities by a greedy method are permanent. That is the greedy algorithms make permanent decisions about the construction of a solution, based on the restricted consideration such as choosing a location that gives a minimum cost. Greedy algorithms for facility location problems are constructive in principle. They are designed to give solutions of fairly good quality without using much time that is needed to compute better quality solutions by other algorithms. The most natural and simple heuristic for the p -median problem is the greedy algorithm. For the p -median problem to locate facilities, this algorithm picks a most 'cost-effective' facility until every required number of facilities p is located.

We propose a modified form of the myopic (greedy) algorithm for the p -median problem. The new algorithm is simple and it gives relatively quality solutions. We demonstrated the importance of the removal of extreme values from a distance matrix before locating the first facility. The modification of the algorithm involves the removal of the extreme or large values from each column of the distance matrix. We then determine the first facility (1-median) after the removal of the extreme values. We revert to the original distance matrix after the first facility (1-median) is located. We then determine the additional facilities using the original distance matrix. We compare the results obtained by the original Myopic algorithm with the modified version using the 400 random problems. The results demonstrate the efficiency and superiority of our new method.

Keywords: *Facilities, p -median, greedy algorithm, modified*

1. INTRODUCTION

The term greedy algorithms have been first used by Edmonds (1970) in a published paper on Submodular Functions, Matroids and Polyhedra. A matroid is a finite set for which a greedy algorithm is guaranteed to be optimal. In the paper by Edmonds (1970) the term greedy algorithm was used when the theorem involving the maximization of a linear program by vector over a set was proposed. Hoffman (1985) describes the greedy algorithms in the context of Linear Programming for which the algorithm maximizes the optimization problem. He also stated that the greedy algorithms were known as Myopic ‘in ancient times’. Kuehn and Hamburger (1963) introduces the Myopic algorithm which is similar to the idea of greedy algorithms described earlier. The proposed algorithm which is greedy is used for locating warehouses and which has been applied to the p -median problem. The location that gives the minimum cost is considered for each stage of the location procedure for the algorithm.

Cornuejols et al. (1977) discussed the problem of optimally locating bank accounts to maximize clearing times. The authors present a solution technique for the p -median problem which involves a greedy interchange heuristic. This greedy heuristic is to determine the lower bound of the value of the optimal solution. The upper bounds of the worst case of the performance of the greedy interchange heuristic were also presented. Captivo (1991) combines the greedy algorithms of Cornuejols et al. (1977) and Kuehn and Hamburger (1963) with the Neighborhood heuristic of Maranzana (1964) resulting in a new heuristic known as *GreedyG*. Hribar and Daskin (1997) combined Kuehn and Hamburger (1963) heuristic with a dynamic programming algorithm to solve the p -median problem. Hassin and Levin (2005) also applied the idea of the greedy algorithm to the set covering problem by allowing the reversal of an earlier decision to include a given set in the solution.

The recent papers on greedy algorithms include Turner et al. (2015); Ding et al. (2015); Turner et al. (2015) discussed greedy algorithms for solving matroid base problems. The authors showed that the proposed greedy algorithm is applicable to a larger class of objective functions than commonly known. The potential and the limits of the greedy algorithm were also illustrated. Ding et al. (2015) proposes a Tabu-mechanism improved iterated greedy algorithm to solve a scheduling problem. The greedy algorithm was shown to be simple but able to produce promising computational results. The good performance of the algorithm depends on the Tabu-based reconstruction strategy in the algorithm.

The other heuristics of the p -median problem include the hybrid heuristic of Resende and Werneck (2004) and that of Ren and Jian (2012). Dzator and Dzator (2013) also proposed a reduction based heuristic of the p -median problem which was applied to the location of ambulances in the Perth metropolitan area.

In this paper we present the modified version of the myopic (greedy) algorithm for the p -median problem. The modification is based on using a reduced problem to locate the 1-median facility. The subsequent facilities are then located using the original matrix. Our method is a modification that allows the new algorithm to choose the 1-median that might be far from the optimal with regard to the original matrix. The ‘sacrifice’ of locating the first facility non-optimally with regard to the original matrix has resulted in having better solutions when two or more facilities are located. The objective of the paper is to demonstrate the effectiveness of our method over the original algorithm regarding the quality of the solution.

The remainder of this paper is organized as follows. In Section 2 we present the p -median problem. The Myopic algorithm of the p -median problem is presented in Section 3. Section 4 discusses the modification of the Myopic algorithm with an example. Section 5 presents computational results and conclusions are briefly presented in Section 6.

2. THE P-MEDIAN PROBLEM

Hakimi (1964, 1965) introduced the p -median problem when he considered the optimal location of switching centers in a communication network. The objective of this problem is to find the locations of p facilities to minimize the demand weighted total distance (total cost) between each demand node and the nearest facility. For the p -median problem the cost of serving demands at node i is the product of the demand at node i and the distance between demand node i and the nearest facility to node i .

The p -median problem consists of determining the location of facilities to minimize the weighted distance between demand points and their closest new emergency facility.

We now present the well known mathematical model for the p -median problem by defining the following notations:-

$I = \{1, \dots, m\}$, the set of demand locations,

$J = \{1, \dots, n\}$, candidates sites for facilities,

d_{ij} = the shortest distance between location i and location j ,

x_{ij} = 1 if the customer at location i is allocated to facility at location j , 0 otherwise,

y_j = 1 if a facility is established at location j , 0 otherwise,

p = the number of facilities to be established,

a_i = the population at the demand node i .

The mathematical formulation of a p -median problem can be specified as follows,

$$\text{Min } \sum_{i=1}^m \sum_{j=1}^n a_i d_{ij} X_{ij}, \quad (1)$$

subject to

$$\sum_{j \in J} x_{ij} = 1, \quad \forall i \in I \quad (2)$$

$$\sum_{j \in J} y_j = p \quad (3)$$

$$x_{ij} \leq y_j \quad \forall i \in I, \forall j \in J \quad (4)$$

$$y_j \in \{0,1\}, \quad x_{ij} \in \{0,1\} \quad (5)$$

The objective (1) is to minimize the total distance from customers or clients to their nearest facility. Constraint (2) shows that the demand of each customer or client must be met. From constraint (3), the number of facilities to be located is p . Constraint (4) shows that customers must be supplied from an open facility, and constraint (5) presents the problem as a binary integer programming.

For uniform demand, we replace the objective function (1) by

$$\sum_i \sum_j d_{ij} Y_{ij}$$

The above formulation assumes that the potential facility sites are nodes on the network. Hakimi (1964) however showed that allowing facilities to be located on the arcs of the network instead of the nodes would not reduce total travel cost. Garey and Johnson (1979) showed that the p -median problem is NP -hard for variable values of p .

3. THE MYOPIC ALGORITHM (MP)

The myopic algorithm is a greedy type which works in the following way. Firstly, a facility is located in such a way as to minimize the total cost for all customers. Facilities are then added one by one until p is reached. For this algorithm, the location that gives the minimum cost is selected. The main problem with this approach is that once a facility is selected it stays in all subsequent solutions. Consequently, the final solution attained may be far from optimal. This algorithm is specifically known as Greedy-Add since facilities are

added one-by-one to attain the required number of facilities. The reverse approach is known as Greedy-Drop which starts with facilities located at all potential facility sites and then eliminate (drop) the facility that has the least impact on the objective function. This method eliminates the facilities one-by-one until the required number of facilities p remains.

The outline of the Myopic algorithm as indicated by Daskin (1995) is presented below as follows:-

Step 1: Initialize $k = 0$ and $X_k = \{ \}$, the empty set.

Step 2: Increase k , the counter on the number of facilities located.

Step 3: Compute $Z_j^k = \sum_i h_i d(i, j \cup X_{k-1})$ for each node j , which is not in the set X_{k-1} , where h_i is the demand at node i . Note that Z_j^k gives the value of the p -median of the objective function if we locate the k^{th} facility at node j , given that the first $k-1$ facilities are at the locations given in the set X_{k-1}

Step 4: Find the node $j^*(k)$ that minimizes Z_j^k . Add node $j^*(k)$ to the set X_{k-1} to obtain the set X_k . Note that $j^*(k)$ gives the best location for the k^{th} facility, given the location of the first $k-1$ facilities.

Step 5: If $k = p$ stop. Go to step 2 if $k < p$.

We present results of the example for the original Myopic algorithm (MP) and the modified Myopic algorithm (MMP) in the tables in Section 4.

4. MODIFICATION AND ILLUSTRATION OF THE MODIFIED MYOPIC ALGORITHM (MMP)

This can be demonstrated with the Myopic algorithm using the 5-node weighted distance matrix in Table 1. To locate two facilities using the Myopic algorithm to serve all the customers (demands) at nodes 1, 2, 3, 4 and 5 we initially sum all columns in Table 1. The Myopic algorithm then locates one facility at node 3 since the minimum column total of 193 corresponds to node 3 giving rise to Table 2 since for the p -median problem a customer (demand) is allocated to the nearest facility. We note also that a customer (demand) located at node 3 will use the facility located at node 3 at a zero cost. Summing all the columns of Table 2 the minimum total of 101 corresponds to node 2 therefore the Myopic algorithm locates the second facility at node 2. The Myopic algorithm locates two facilities at nodes 2 and 3 and the cost of using the two facilities by all customers at nodes 1, 2, 3, 4 and 5 is 101.

In considering column totals we may not select a location because of a very large cost entry. Since in practice one would never allocate a customer (demand) to very distant facility, it would make sense to eliminate large entries before taking column sums. For example we eliminate two greatest values from each facility node, then the column sums are shown in the Table 3. The totals for the potential facilities from Table 3 are as follows. Facility 1 is 61, Facility 2 is 55, Facility 3 is 64, Facility 4 is 62 and Facility 5 is 105.

We demonstrate below the advantage of using the reduced matrix to locate two facilities by the Myopic algorithm. After the removal of the two greatest values from each column in Table 1, the Myopic algorithm then locate the first facility at node 2, since the minimum total column of 55 corresponds to node 2. Using similar arguments as already stated in this section Table 1 is again adjusted to give Table 4. The minimum total of 74 corresponds to node 1 when the values of each column are summed in Table 4. The Myopic algorithm therefore locates the second facility at node 1. The Myopic algorithm locates two facilities at node 1 and 2 after the removal of extreme values. The cost of using facilities located at nodes 1 and 2 is 74 which is lower than facilities located at nodes 2 and 3 when large distance values are not eliminated.

Table 1. Weighted Distance Matrix.

	Facility 1	Facility 2	Facility 3	Facility 4	Facility 5
Demand 1	0	82	37	42	100
Demand 2	67	0	78	93	97
Demand 3	74	18	0	20	49
Demand 4	19	87	27	0	66
Demand 5	42	37	51	87	0

Table 2. Adjusted Weighted Distance Matrix.

	Facility 1	Facility 2	Facility 3	Facility 4	Facility 5
Demand 1	0	37	37	37	37
Demand 2	67	0	78	78	78
Demand 3	0	0	0	0	0
Demand 4	19	27	27	0	27
Demand 5	42	37	51	51	0

Table 3. Reduced Weighted Distance Matrix.

Facility 1	Facility 2	Facility 3	Facility 4	Facility 5
0	0	0	0	0
19	18	27	20	49
42	37	37	42	66

Table 4. Adjusted Weighted Distance Matrix.

	Facility 1	Facility 2	Facility 3	Facility 4	Facility 5
Demand 1	0	82	37	42	82
Demand 2	0	0	0	0	0
Demand 3	18	18	0	18	18
Demand 4	19	87	27	0	66
Demand 5	37	37	37	37	0

The steps for the determination of the 1-median are presented below.

Step 1: Arrange the columns of the distance matrix in ascending order.

Step 2: Delete the multiples of p largest values from each column where p is the number of facilities to be located. The coefficient of p depends on the problem size. For example for 10 to 29 nodes we delete p largest values and for 30 to 39 nodes we delete $2p$ largest values.

Step 3: Add the columns and arrange the totals in ascending order.

Step 4: Choose the column corresponding to the minimum as the first facility (1-median)

We then use the original Myopic algorithm to determine the additional facilities (that is 2-median, 3-median, 4-median,....., p -median))

5. COMPUTATIONAL RESULTS

We determine the first facility (1-median) using a program written in C++. The additional facilities are determine using the Location software known as SITUATION (Daskin; 1995). To determine the additional facilities we assume that the first facility is the one obtained from the reduced matrix. The optimal solutions for the comparison were determined by the implementation of the Lagrangian relaxation in the Location software (Daskin; 1995). The Myopic Algorithm and the Modified Myopic Algorithm are tested on 400 randomly generated sets of data for a [10, 100] matrix with n ranging from 10 to 100 in steps of ten and p ranging from 2 to 5. That is, for each problem size n and for locating 2, 3, 4 or 5 facilities, 40 uniformly

distributed random problems were generated. The computations were carried on a personal computer with an Intel Pentium 4 processor, 2.8GHZ and 448 of RAM.

Table 5. Percentage Improvement of Modified Myopic Algorithm.

Number of Nodes (<i>n</i>)	Average Value-MP	Average Value-MMP	Percentage Improvement
10	2.66	0.95	64.29
20	2.97	2.75	7.41
30	2.76	2.41	12.68
40	1.92	1.87	2.60
50	1.78	1.84	-3.37
60	3.07	2.07	32.57
70	2.40	2.15	10.42
80	3.01	2.34	22.26
90	1.98	2.13	-7.57
100	2.04	1.71	16.18

Table 6. Average Values for *RRH* and *MMP* using 20 Uniformly Distributed Random Data.

RRH Deviations from Optimal	MMP Deviations from Optimal	CPU time (sec) RRH	CPU time (sec) MMP
1.22	4.51	21.05	3.89

Table 5 shows the comparison of the original Myopic algorithm (MP) and the modified Myopic algorithm (MMP) with the optimal value. We note from the table that the modified version of the Myopic algorithm performs better for 10, 20, 30, 40, 60, 70, 80 and 100 nodes. That is the modified Myopic algorithm performs better for the 80% of the different problems. The best improvement occurs for a problem with size is 10 nodes where there is improvement of 64.29%. The next best improvement is with the problem size of 60 of an improvement of 32.57%. The problem sizes of 80, 100 have given percentage improvement of 22.26% and 16.18% respectively. Figure 1 which is shown below also confirmed the better performance of our modified algorithm in most cases. Table 6 shows the performance of the Modified Myopic when compared to a *p*-median Reduction Heuristic (*RRH*) from the literature (Dzator and Dzator; 2013). The result shows that the improvement heuristic from the literature is better but the computational time of *MMP* is 18.5% of the literature heuristic.

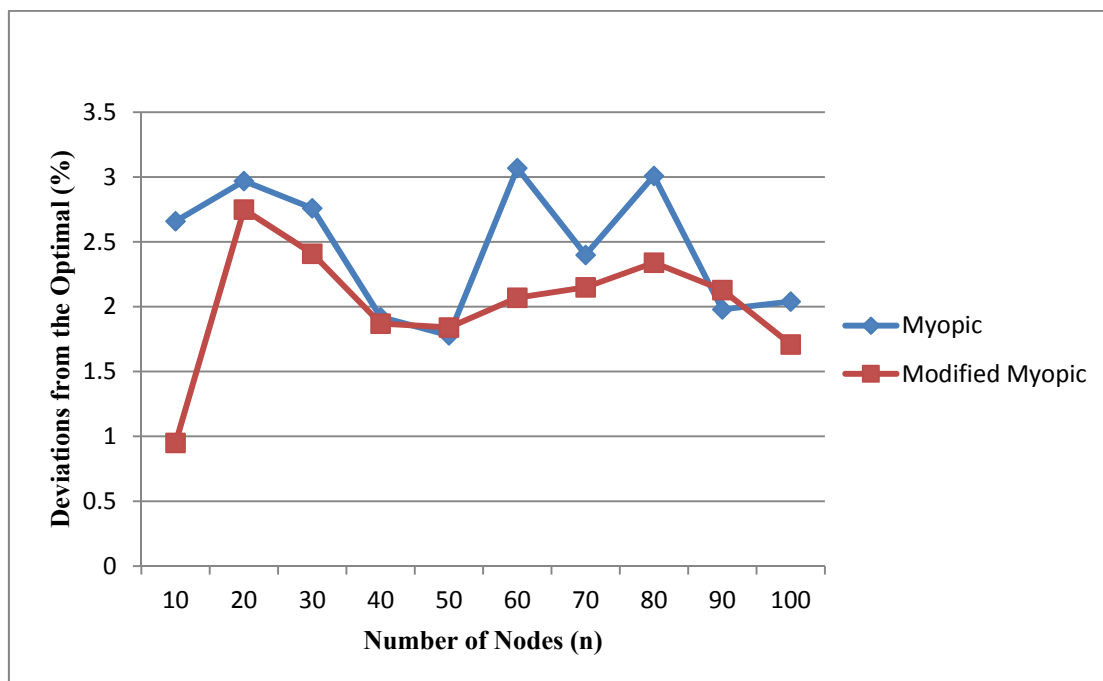


Figure 1. Performance of Myopic and Modified Myopic Algorithm.

6. CONCLUSIONS

In this paper we presented the modified version of the Myopic Algorithm. The modification is based on the determination of the 1-median solution for the p -median problem by the reduced distance matrix. We then used the original matrix to determine the subsequent facilities. The application of the new algorithm to locate facilities for the problem size of up to 100 nodes gives better solutions in most cases when compared with the original algorithm. The new algorithm is also compared to an improvement heuristic from the literature with promising results

This new modified Myopic Algorithm performs better than the original Myopic Algorithm when tested on 400 random data sets so will be better starting solutions for improvement heuristics of the p -median problem.

The new modified algorithm will be tested against other greedy algorithms for the p -median problem for future research using a larger problem size and more facilities. We will examine both Myopic Algorithm and the Modified Myopic Algorithm as starting solutions for other algorithms. The new algorithm will also be applied to real problems.

REFERENCES

- Captivo, M.L. (1991), Fast primal and dual heuristics for the p -median location problem, *European Journal of Operational Research*, 52, 65-74.
- Cornuejols, G., M. L. Fisher and G. L. Nemhauser (1977), Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms, *Management Science*, 23, 789-810.
- Daskin, M.S. (1995), *Network and Discrete Location: Models, Algorithms and Applications*, John Wiley and Sons, Inc., 498 pp., New York.
- Ding, J., S. Song, N.D Gupta, R. Zhang, R. Chiong and C. Wu (2015) An improved iterated greedy algorithm with a Tabu-based reconstruction strategy for the no-wait flowshop scheduling problem, *Applied Soft Computing*, 30, 604-613
- Dzator, M., and J. Dzator (2013). An effective heuristic for the p -median problem with application to ambulance location. *Opsearch* 50 (1), 60-74.
- Edmonds, J. (1970) Submodular functions, matroids and certain polyhedral, *Synopsis for the Instructional Series of Lectures*, 11-26.
- Hakimi, S.L. (1964), Optimisation locations of switching centres and the absolute centres and medians of a graph, *Operations Research*, 12, 450-459.
- Hakimi, S.L. (1965), Optimum distribution of switching centers in a communication network and some related graph theoretic problems, *Operations Research*, 13, 462-475.
- Hassin and Levin (2005), A better-than-greedy approximation algorithm for the minimum set cover problem, *SIAM Journal of Computing*, 35, 189-200.
- Hoffman, A. (1985), On greedy algorithms that succeed, *Surveys in combinatorics*, 92-112.
- Hribar, M. and M.S. Daskin (1997), A dynamic programming heuristic for the p -median problem, *European Journal of Operational Research*, 101, 499-508.
- Kuehn, A.A. and M.J. Hamburger (1963), A heuristic program for locating warehouses, *Management Science*, 9, 643-666.
- Maranzana, F.E. (1964), On the location of supply points to minimize transport costs, *Operational Research Quarterly*, 15, 261-270.
- Ren, Z. and H. Jiang (2012), An accelerated-limit-crossing-based multilevel algorithm for the p -median problem, *IEEE Transactions*, 42, 1187-1202.
- Resende, M. and R. Werneck (2004), A hybrid heuristic for the p -median problem, *Journal of Heuristics*, 10, 59-88.
- Turner, L., M. Ehrgott and H. W. Hamacher (2015), On the generality of the greedy algorithm for solving matroid base problems, *Discrete Applied Mathematics*, 195, 114-128.