www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 10(2), May-Aug 2019, 17-25. ISSN: 2007-1558.

Solution Search for the Capacitated P-Median Problem using Tabu Search

Mauricio Romero Montoya[1], Rogelio González Velázquez[2], Martín Estrada Analco[2], José Luis Martínez Flores[1], María Beatriz Bernábe Loranca[2]
[1] Universidad Popular Autónoma del Estado de Puebla, México.
[2] Benemérita Universidad Autónoma de Puebla, México.
mauricio.romero@upaep.edu.mx,rgonzalez@cs.buap.mx,mestrada@cs.buap.mx,
joseluis.martinez01@upaep.mx, beatriz.bernabe@gmail.com

Abstract. Capacitated p-median problem (CPMP) is an important facility location problem, where Capacitated p-medians are economically selected in order to serve a set of demand vertexes in such a way that the total demand placed on each of the candidate medians does not exceed their capacity. This work presents an efficient Tabu-search based metaheuristic to solve the CPMP. Said metaheuristic is tested using computational experiments aided with the OR-Library test instances, and in most cases, the optimum value is reached in reasonable computing times. The obtained results are compared and evaluated with the aid of Lingo 16.0. Lastly, an actual case is solved in which the objective is presenting new locations for a meat-product company's distribution centers.
Keywords: Tabu Search, Metaheuristic, Capacited p-median, Np-Hard.

Article Info
Received Dec 26, 2016
Accepted Dec 11, 2017

1 Introduction

Location problems arise from the need to define the most convenient location to place facilities such as distribution centers (DC), production plants, commercial warehouses, service centers, trash dumps, fire or police stations, hospitals or supermarkets [20]. How can one ensure that the DC is placed at the best location? By the best location, we mean that both customers and suppliers have the best logistic advantages, that is closest to the optimum and that costs, in general, are minimized.

The p-median problem (PMP) and the Capacitated p-median problem (CPMP) are excellent options to define the best location for these centers. The Capacitated p-median basically considers a situation in which it is required to partition a set of clients with demand into exactly p groups. Each group is defined not only by the set of clients in it, but also by the location of their median and each p's capacity. A clear representation of this problem is presented in fig. 1.

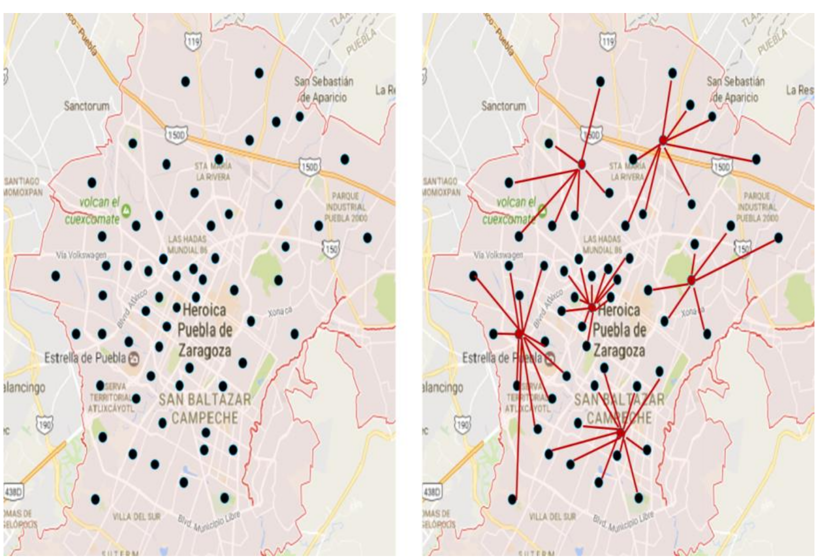

Fig. 1. Capacitated P-Median Problem (CPMP)

The present work is organized as follows: Section 1 presents the introduction. Section 2 includes a brief description of the Capacitated P-Median Problem and its mathematical formulation, In section 3 the approximation method and the algorithm designed under Tabu search are presented, Section 4 shows the result of the computational experiments, section 5 shows the algorithm applied to our life case. Lastly, in section 6 the conclusions from this work are presented.

2 Problem description

The CPMP problem is an adaptation of the PMP which was developed by Hakimi (1946). It is not easy to solve the CPMP problems with any of the linear programming traditional methods when some applications are considered and at the same time are compound of a large number of variables and, therefore, they have large solution spaces. Due to this, it is necessary to consider other possible solution methods such as heuristic ones. The combinatorial nature of the p-median problem and the advancements which were added to the solution techniques, both allow the promotion of excellent outcomes which are very close to the optimum result. These outcomes are very often calculated from the combination of mathematical programming and some others searching and heuristic techniques such as, the tabu search and the Lagrangian relaxation [22].
Informally, the objective of the p-median problem is the determination of P locations for facilities for a set of *n* (nsp) previously defined candidates in such a way that the demand set is met and the sum of the distances between the demand points and the selected locations is minimized.

Formally, it is assumed that each vertex in a potential graphic median. The p-median problem may be formally defined as follows:

Let G=(V, A) be a simple directed graph where V represents the vertexes, and A represents the arcs. The objective is to find the vertex set vpCV (set of medians) with cardinality P, such that: a) the sum of the distances between the remaining vertexes and the selected vertex Vp in {V-Vp} (the demand set) is minimized, and b) all demand points are met without violating the capacity restrictions for the located median facilities.

When compared with the p-median problem, the Capacitated P-median has the added restrictions that: (1) each facility placed can only satisfy a limited demand (capacity restriction) and (2) each demand point should have its demand met by its respective median- selected facility.

The capacitated p-median problem is presented next, regarding integer programming [15].

The capacitated p-median problem (CPMP) aims to solve the optimal location for p facilities, considering both distances and service capacities offered by each facility. Total service demanded by the vertexes identified in each p-median cluster may not exceed the facility's capacity. The mathematical formulation follows:

$$y_{ij} = \begin{cases} 1 & \text{if the median is placed at location } \; j \in J \\ 0 & \text{is not assigned in other cases} \end{cases}$$

$$X_{ij} = \begin{cases} 1 & \text{if client } i \text{ is assigned to location } j \\ 0 & \text{is not assigned in other cases} \end{cases}$$

$$Min \sum_{i \in N} \sum_{j \in J} w_i C_{ij} X_{ij} \tag{1}$$

Subject to:

$$\sum_{j \in J} X_{ij} = 1 \qquad \forall i \in N \tag{2}$$

$$\sum_{j \in J} Y_j = P \tag{3}$$

$$\sum_{i \in J} w_i X_{ij} \leq q_j Y_j \qquad \forall i \in N, j \in J \tag{4}$$

$$X_{ij} \in \{0,1\}, Y_j \in \{0,1\} \forall i \in N, j \in J \tag{5}$$

Target function (1) minimizes the sum of distances between all demand nodes. Constraint (2) ensures each node is assigned to exactly one facility. (3) restricts demand nodes to be assigned. (4) ensures that one client is assigned to only one selected median. (5) are binary restrictions.

As can be seen, the difference between the classical P-median problem and the CPMP lies in the fact that CPMP sets a fixed amount of service or product that facility *j* provides, which is very important from the economic point of view. On the one hand, CPMP allows the decision maker to locate facilities regardless of their capacity, while covering all or part of the demand that corresponds to the capacity of each facility.

Much research has been done to solve the p-median and CPMP problems, yet one of the biggest challenges is dealing with the great amounts of data needed to provide good quality solutions within reasonable computing time. A proposal to solve the CPMP for large instances is the inclusion of Tabu search embedded with dual relaxation, with the added characteristic that a limit on the quality of the solutions while the dual relaxation of the problem is solved heuristically. [14]

2.1 Metaheuristics for Capacitated P-Median CPMP

The combinatorial nature of the p-median problem caused advances in solution techniques to be often made using combinations of mathematical programming and heuristic search techniques such as branch and bound and Lagrangian relaxation [18]. Several efforts have been made in order to solve the p-median and capacitated P-median problems. However, two articles are of the greatest interest when Tabu search is involved, because the methods developed to solve the p-median problem are solved using Tabu search (TS). A TS procedure for CPMP was developed by Sörensen (2008) [21] that achieved optimal solutions for all the small and medium-sized instances in the OR-Library, except for two. Even though the paper fails to identify said instances, it is established that the relative errors were under 0.1%, which represents an improvement over all previous work for the CPMP. This work backs the conclusions by Arostegui et al. (2006) [3] and establishes that TS may well be a dominant methodology to use when solutions for the CPMP are sought.

There are algorithms of solution that go from the search of the optimum making simplifications of the model as it can be seen in COBRA [5], to the application of metaheuristics, some examples of these are genetic algorithms, simulated annealing, clustering tools, among others. It is important to bring out that all these algorithms search the optimum. Taking into account the clustering method, the methods of the p-median have been created within the data mining area to detect patterns in the large scale of data as it can be seen in regions which are densely populated in a Euclidean space. One of the challenges is to use those large amounts of data properly to obtain quality in the solutions in a reasonable computing time [9].

In this context, it is possible to see that most of those models are formulated with Mixed Integer Linear Programming (MILP); this is done because the exact ways to solve this type of problems in medium sizes need powerful computers; so the heuristic approaches are used to solve the real problems. The main techniques used are patterns of decomposition, hierarchical, Lagrangian relaxation, branching and dimensioning models and some metaheuristics such as simulated annealing and genetic algorithms [19]. Hindi and Basta in 1994 used the branching and dimensioning techniques to work the same Geoffrion & Graves model; they also added the limited capacity constraint and calculated the lower limits through structured transformations (Hindi and Basta, 1994).In the same area, there were great results when p-median and simulated annealing were mixed. The effectiveness of this type of methods can be seen clearly through the tests done with the instances of OR-Library [2]. Lately, it can be found an excellent review of this literature on p-median [14]. About partitioning, considered as a p-median problem, the correct grouping is given when the elements of the partition meet the geometric criterion of compactness, which it is achieved when the distance is minimized between the distribution centers and the clients (centroids and geographic objects) [4]. The Tabu Search (TS) meta-heuristic has proved highly successful in solving combinatorial and nonlinear problems. A key aspect of TS consists of using adaptive forms of memory to forbid the search process to revisit solutions already examined unless the trajectory to reach it is different.[10]. Herda, M., & Haviar, M. suggests two ways of combining a genetic algorithm (GA) with integer programming to improve the quality solution of the CPMP [12]. For another side, Comber & Sasaki describes the development and application of a modified grouping genetic algorithm (GGA) used to identify sets of optimal ambulance locations [6]. Osman and Christofides [17] propose simulated annealing and tabu search method. Finally, good results are obtained when using three different types of search for the CPMP, In the first one solutions are obtained with Scatter Search whereas the second algorithm uses Path Relinking. A third approach that combines Path Relinking with Scatter Search is also analyzed. The GRASP methodology is used to generate the initial Reference Set both for Scatter Search and Path Relinking [7].

3 Tabu Search for Capacitated P-Median

Tabu Search emerged from diverse works published in the late 70's. Even though the main concepts and strategies behind it already existed, it was in 1989 when Fred Glover and Manuel Laguna formally established the name and methodology in the homonymous book Tabu Search [8].

This approximation relies on traditional Tabu Search, which generates a random initial solution. The parameters here have been defined as: 1) number of iterations (global iterations to be performed by the algorithm) 2) perturbationIterations, (determines the number of times a solution may be accepted according to the solution of the previous iteration. When the counter perturbationCounter crosses this limit, a new random solution is generated, and search is restarted).

The module parameter determines the frequency with which in Vicinity 1 (Algorithm 2) is to be used instead of Vicinity 2 (Algorithm 3), as can be seen in the first conditional within the main while loop. When the cost given by the vicinity functions improves on the best-known cost (bestCost), this will be replaced, along with the solution configuration (best solution). The last step updates the tabu lists and global iteration counter.

Algorithm 1. Tabu Search (PSEUDOCODE):

```
READ numberOfIterations
READ perturbationIterations
READ module
SET iterationCounter and perturbationCounter to 0
SET initialSolution and bestSolution to currentSolution

WHILE iterationCounter < numberOfIterations
      SET currentCost to cost of currentSolution
      IF iterationCounter % module = 0
      CALL Neighbor1 with currentSolution RETURNING newCost
      ELSE
            CALL Neighbor2 with currentSolution RETURNING newCost
      END IF
      IF newCost > currentCost
            INCREMENT perturbationCounter
      END IF
      IF newCost < bestCost
            SET bestSolution to newSolution
            SET bestCost to newCost
      END IF
      SET currentSolution to newSolution
      SET currentCost to newCost
      IF perturbationCounter > perturbationIterations
            CALL RestartSolution RETURNING newSolution
            SET currentSolution to newSolution
            SET perturbationCounter to 0
      END IF
      CALL UpdateTabuLists RETURNING updatedTabuLists
      INCREMENT iterationCounter
```

```
END WHILE
RETURN bestSolution
```

The CPMP is a combinatorial optimization problem belonging to the NP-Hard category [13,15]. The Tabu search metaheuristic has proven effective in searching for solutions for a great variety of problems of this kind, such as the quadratic assignment problem, the traveling salesman problem or the p-median problem, amongst others. The Tabu search metaheuristic uses different intelligent search schemes such as the use of short and long-term memories, keeping a list of Tabu solutions to be avoided by their leading to trajectories where it has been proven there are no good solutions; these are search tools of artificial intelligence.

4 Results

The algorithm here presented was validated using the instances drawn from the O-Lib [16] for CPMP problems, where the range of instances runs from n=100 to n=900, and p=5 to p=20. The files of the instances, which were downloaded directly from the O-Lib files, are encoded in order to decode the information. It was also necessary to apply the Floyd algorithm. As a result of this activity, the necessary dissimilarity matrices were obtained to solve the CPMP instances. The code was developed in Java and tested with a PC with 6GB RAM and a second generation i5 processor. Each instance from the O-Lib has executed a total of 50 times in order to obtain registers for the best solution and computing times. The O-Lib instances were also solved with the Lingo 16.0 optimization software, and results were later compared, which showed that the solution times for each of the instances was short. However, the solutions obtained by the algorithm were not always near the optimum.
On table 1, the notation is the next: I=Instance, C=(n) are the Customers, R COST are the values in the experiment when parameters varied (random), T (seg), is the execution time for each test, OPT are the values obtained with Lingo 16.0, T OPT is the run time for each instance in Lingo 16.0, D COST are the default parameters with TS, and GAP is the difference between the best cost and the obtained value by the approximation method.

Table 1. Results for Capacited P-Median with TS and Lingo 16.0

I	Capacited Q=120 Random						Default			Lingo 16	
	BV	C	P	RC	T	GAP	DC	T	GAP	OPT	T OPT
1	713	50	5	717	48340	0.56	722	10040	1.26	713	7.77
2	740	50	5	740	50933	0	740	9950	0	740	1.05
3	751	50	5	**752**	72445	0.13	**772**	9810	2.8	751	2.34
4	651	50	5	653	53402	0.31	654	10030	0.46	651	1.18
5	664	50	5	664	50073	0	664	9950	0	664	2.80
6	778	50	5	778	24426	0	778	9930	0	778	1.55
7	787	50	5	796	52371	1.14	810	9820	2.92	787	63.02
8	820	50	5	826	54607	0.73	851	9690	3.78	820	236.96
9	715	50	5	718	23938	0.42	721	9840	0.84	715	9.95
10	829	50	5	844	47455	1.81	844	9810	1.81	829	129.25
11	1006	100	10	1054	3E+05	4.77	1074	11900	6.76	1006	119.65
12	966	100	10	978	2E+05	1.24	983	11900	1.76	966	219.48
13	1026	100	10	1032	1E+05	0.58	1055	12050	2.83	1026	27.22
14	982	100	10	1012	3E+05	3.05	1027	12110	4.58	982	1190.93
15	1091	100	10	1133	7E+05	3.85	1160	12010	6.32	1091	711.64
16	954	100	10	975	2E+05	2.2	985	12120	3.25	954	108.53
17	1034	100	10	1077	64331	4.16	1072	11960	3.68	1034	112.21
18	1043	100	10	1086	67329	4.12	1088	12040	4.31	1043	543.21
19	1031	100	10	1057	63633	2.52	1049	11780	1.75	1031	380.67
20	1005	100	10	1133	66159	12.7	1146	11850	14.03	1005	48439.28

5 Test with a real instance

The studied company has two main business areas, production and commercialization of meat products. All raw materials arrive at the distribution center located in the Mexican city of Puebla, in the state of the same name. From this location, they are re distributed and / or are processed to be later distributed to each of 17 different sale points, other than this distribution center, which implies major costs in transportation as well as some other problems. Table 2 shows the capacity and demand for each business unit. A fixed operation cost of $30000.00 is considered for all company's facilities. Figure 2 shows the distribution network originally operated by the company with only one distribution center.

Table 2: Current status of the company's business units.

Location	Capacity (tons)	Demand (tons)
Puebla, Puebla.	60	7
San Martin Texmelucan, Puebla.	30	7
Ciudad de México.	30	5
Guadalajara, Jalisco.	30	4
Aguascalientes, Aguascalientes.	30	4
Tampico, Tamaulipas.	15	2
Actopan, Hidalgo.	18	4
Tierra Blanca, Veracruz.	30	2
Acayucan, Veracruz.	30	2
Tulancingo, Hidalgo.	17	3
Tonalá, Chiapas.	30	4
Villaflores, Chiapas.	30	3
Pijijiapan, Chiapas.	30	2
Arriaga, Chiapas.	30	6
Mapastepec, Chiapas.	30	2
Monterrey, Nuevo León.	30	4
Reynosa, Tamaulipas.	30	5
Torreón, Coahuila.	30	4

The current monthly total operation cost for the distribution system of the company is $160,066.00MXN, including the fixed monthly operation cost for the distribution center in Puebla, of $30,000.00MXN.

The distance matrix (in km) between all 18 cities that could potentially be distribution centers is considered. Distances were found using Google Maps, considering round trips. The corresponding files with all data (capacity, demand, fixed costs and distance tables). To generate a proposal for a new distribution strategy for the company, two P values are considered: P=2 and P=3. Therefore, two instances for the Capacitated P-Median problem a are used to run both the proposed algorithm and Lingo 16.0.

Fig. 2. Current distribution network

First instance, 2 Distribution Center proposal

In this first instance, two distribution centers, P=2, are proposed, instead of the current one center system. The obtained results, and the computing run times for both solution methods are presented in Table 3. The generated distribution strategy thanks to the results derived from the Tabu search algorithm is shown in Figure 3.

Table 3 Results for real data with two DC

P=2	Sol.	T (sec.)
Tabu Search	138,198	7.43
Lingo 16.0	138,198.3	0.39

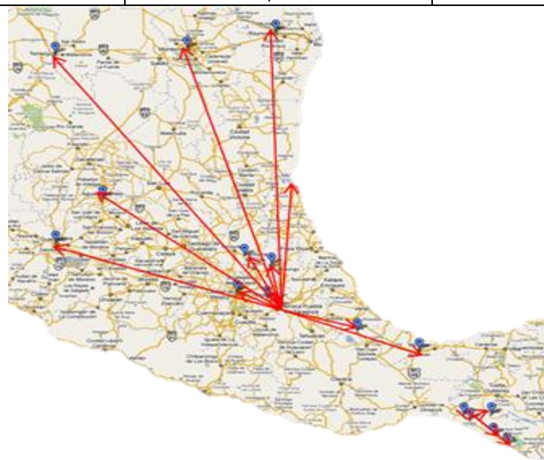

Fig. 3. Proposed distribution for p=2

The study of the obtained data from the Tabu search algorithm is of great interest on its own, and for this particular company, for when total transportation costs are analyzed, for one distribution center results in costs of $160,066.00, versus total costs of $138,198.3 for two distribution centers, representing a reduction of 13.7% in transportation costs, In this proposal, both Puebla and the city of Arriaga are enabled as distribution centers, the latter servicing mostly the south-east of the country, while Puebla serves the center and north of the country. This initial solution is good for it implies considerable cost reductions. However it is plain to see that distances between the distribution center and locations such as Monterrey are still very large, so a second instance is tried, contemplating the possibility of two distribution centers in addition to the one already in operation.

Second instance, 3 Distribution Center proposal

In a second instance, the impact of opening three operation centers (p=3) in the distribution-related costs are analyzed for the company's products. The obtained results and computing run times for both models are shown in Table 4, while the three distribution centers assigned by the Tabu search algorithm to serve the different sales points are shown in Figure 4.

Table 4 Real data results with 3 DC

P=3	Sol.	T (sec)
Tabu Search	133,824	7,345
Lingo 16.0	133,824.7	0.23

Fig. 4. Proposed distribution network for p=3

Analysis of the resulting data shows total transportation costs of $133,824.7, that is, a reduction of 16.4% when compared to the current monthly costs, and 3% when compared to the first instance. This seemingly small cost reduction comes with other substantial benefits, such as an improvement in customer service, for the previous delivery time is 4 to 6 days, while the new estimate is 3 to 4 days, which will benefit the perception of the company by the customer. It is also estimated that the product availability will increase, as the demand projections will be made in function of the sector to be supplied with these products, so demand will be disaggregated amongst them, while each distribution center will be able to place orders from suppliers, eliminating duplicated buying orders and the delay in their processing.

6 Conclusions

The research described in the present work has been motivated by the need to more efficiently find solutions to the CPMP, as shown by the robustness of the designed algorithm, validated by means of the 20 OR-Library instances, which have been compared with the results obtained with Lingo, with an acceptable GAP, as shown in Table 1. It is well known that CPMP is part of the location-allocation problems, which define the strategy for an organization, that is, they are long-term decisions, and a good decision on the distribution and placement of facilities has positive implications on the efficiency of the supply chains, as well as in the cost competitivity of the products or services reaching customers, thus generally benefiting the population sector that demands the product or service.

Given the results for the real instance, the implementation of these strategies in the industrial sector becomes fundamental in ensuring competitivity, and with it, the survival of the organizations.

Lastly, it has been shown that PMP and CPMP are the same when for CPMP an unlimited capacity is considered in the restrictions.

References

1. Alegre, J., Aragón, A., Casado, S., Delgado, Cristina, & Joaquín, P. (2010). Resolución de 2 modelos de localización mediante Búsqueda Dispersa. Universidad de Burgos.
2. Al-khedhairi 2008. Simulated Annealing Metaheuristic for Solving P-Median. Int. J. Contemp. Math. Sciences, 3, 1357- 1365
3. Arostegui, M. J., S. Kadipasaogul, and B. Khumawala. An empirical comparison of tabu search, simulated annealing and genetic algorithms for facilities location problems. International Journal of Production Economics 103, 742–754 (2006)
4. Bernábe B,. Gonzalez V. 2014 P-Median: a performance analysis, 2do International Symposium on Language a Knowledge Engeeniring LKE 2014
5. Church, L. Richard, Cobra: A New Formulation of the Classic P-Median Location Problem, Annals of Operation Research, p-103.
6. Comber, A. J., Sasaki, S., Suzuki, H., & Brunsdon, C. (2011). A modified grouping genetic algorithm to select ambulance site locations. International Journal of Geographical Information Science, 25(5), 807-823.

7. Díaz, J. A., & Fernandez, E. (2006). Hybrid scatter search and path relinking for the capacitated p-median problem. European Journal of Operational Research, 169(2), 570-585.
8. Glover, F., Laguna, M.: Tabu Search. Kluwer Academic Publishers, Norwell, MA, USA (1997)
9. Hansen P, GERARD, and CRT. Data Clustering using Large p-Median Models and Primal Dual Variable Neighborhood Search. Montreal Canada.
10. Hanafi, S. (2001). On the convergence of tabu search. Journal of Heuristics, 7(1), 47-58.
11. Hakimi, S.: An algorithmic approach to network location problems. part i: The p-centres, SIAM Journal of Applied Mathematics, 37, pp. 513–538 (1979)
12. Herda, M., & Haviar, M. Hybrid genetic algorithms with selective crossover for the capacitated p-median problem. Central European Journal of Operations Research, 1-14 (2017).
13. J. Resse. 2006. Methods for Solving the p-Median Problem: An Annotated Bibliography. Wiley Periodicals, Inc. Networks 48(3), 125-142.. 8
14. Kariv O., & Hakimi, S.L. 1979. An algorithmic approach to network location problems: The p-medians. SIAM Journal on Applied Mathematics, 37(3), 539-560.
15. M, Daskin, M., 1995. Network and Discrete Location: Models, Algorithms and Applications. New York: John Wiley and Sons, Inc. 2
16. OR-library: http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html, Retrieved (2017)
17. Osman, I. H., Christofides, N.: Capacitated clustering problems by hybrid simulated annealing and tabu search. Intern. Trans. in Operational Research, Vol.1 (3), pp. 317–336
18. Prikul, H., & Jayaraman, V. (1996). Production, Transportation, and Distribution Planning in a Multi-Commodity Tri-Echelon System. T
19. Reese, J.: Methods for solving the p-median problem: An annotated bibliography. Networks, 48(3), pp. 125–142 (2006)
20. Revelle, C.: A perspective of location Science. Location Science, 5, pp. 3–13 (1997)
21. Sörensen, K. Investigation of practical, robust and flexible decisions for facility location problems using tabu search and simulation. J Operations Research Society 59(5), 624–636 (2008, May)
22. Venables, H.: Ant Colony Optimization: A Proposed Solution Framework for the Capacitated Facility Location Problem. http://sure.sunderland.ac.uk/policies.html (2011)